The Annals of Applied Probability
1993, Vol. 3, No. 3, 914–933

IMPORTANCE SAMPLING FOR GIBBS RANDOM FIELDS

By Paolo Baldi,[1] Arnoldo Frigessi and Mauro Piccioni

Il Università di Roma, Istituto per le Applicazioni del Calcolo, CNR and Università dell'Aquila

The existence of an asymptotically efficient importance sampling distribution for estimating small probabilities of statistics of Gibbs random fields is proved, together with its uniqueness, in an appropriate sense. This distribution is also a Gibbs random field associated with an interaction potential that is explicitly given. The particular case of Markov chains is treated. The practical relevance of the results in applications is discussed.

1. Introduction. Recently the role of large deviations results in the development of importance sampling methods for stochastic processes has been investigated. See, for example, the book by Bucklew [1] and the papers by Glynn and Iglehart [11] and Bucklew, Ney and Sadowsky [2]. Assume that the probability

$$p_n(A) = P_n\{T_n \in A\} \tag{1.1}$$

has to be evaluated, where P_n is a probability measure on $\mathbb{R}^n$ and $T_n\colon \mathbb{R}^n \to \mathbb{R}$ is a random variable. Especially when n is large, $p_n(A)$ cannot, in general, be computed analytically. Furthermore, if $p_n(A)$ is very small, direct Monte Carlo sampling from the measure P_n is not efficient because a very large sample would be required to reach reasonable accuracy. This situation is often encountered in different contexts, from hypothesis testing to reliability evaluation of stochastic systems. Here importance sampling methods are useful.

An importance sampling estimate of $p_n(A)$ is given by

$$\hat{p}_n(A) = \frac{1}{N}\sum_{k=1}^{N} 1_{\{T_n \in A\}}(X^{(k)})\frac{dP_n}{dQ_n}(X^{(k)}), \tag{1.2}$$

where $X^{(k)} = (X_1^{(k)}, \dots, X_n^{(k)})$, $k = 1, \dots, N$, is an i.i.d. sample from the distribution Q_n (the importance sampling distribution), which is supposed to be absolutely continuous with respect to P_n. Under Q_n, $\hat{p}_n(A)$ is unbiased and

Received April 1992; revised January 1993.

[1]Partially supported by a research appointment with I.A.C.–C.N.R. Roma, Italy.

AMS 1991 *subject classifications.* Primary 60G60, 65C05; secondary 60F10, 62G20.

Key words and phrases. Large deviations for Gibbs random fields, Monte Carlo methods, rare events, asymptotic efficiency, variance reduction techniques.

its variance is

$$\frac{1}{N}\left\{E_{Q_n}\left(1_{\{T_n\in A\}}\left(\frac{dP_n}{dQ_n}\right)^2\right)-p_n^2(A)\right\}. \tag{1.3}$$

Therefore, the best importance sampling distribution Q_n minimizes the second moment

$$v_{Q_n}(A)=E_{Q_n}\left(1_{\{T_n\in A\}}\left(\frac{dP_n}{dQ_n}\right)^2\right).$$

The class $\mathscr{C}_n$ of admissible distributions Q_n should be selected on the basis of computational convenience. For example, if all distributions on $\mathbb{R}^n$ are included in $\mathscr{C}_n$, then $v_{Q_n}(A)$ can be reduced to $p_n^2(A)$ by choosing Q_n as the restriction of P_n to the set $\{T_n \in A\}$. This solution is obviously impractical.

The explicit minimization of $v_{Q_n}(A)$ is, in general, impossible for large n. Bucklew, Ney and Sadowsky [2] proposed to look at the minimization asymptotically in n. Assume that it is a priori known that

$$p_n(A) \sim e^{-ni(A)} \tag{1.4}$$

for $n \to \infty$, where $i(A) > 0$, and that for all measures in $\mathscr{C}_n$,

$$v_{Q_n}(A) \sim e^{-ni_Q(A)}, \tag{1.5}$$

where Q is any weak limit of Q_n as $n \to \infty$. For large n, the minimization of $v_{Q_n}(A)$ may be replaced by the maximization of the rate $i_Q(A)$. Moreover, because the quantity in (1.3) is nonnegative, one always has $i_Q(A) \le 2i(A)$. Whenever

$$i_Q(A) = 2i(A), \tag{1.6}$$

$\{Q_n\}$ is said to be *asymptotically efficient.* The meaning of (1.4) and (1.5) is that the probability measure $p_n(\cdot)$ and the finite measure $v_{Q_n}(\cdot)$, for any $Q_n \in \mathscr{C}_n$, satisfy a *large deviations principle* with rate functions i and i_Q, respectively. The minimization of these rate functions over A (assuming suitable regularity conditions on this set) yields the exponents $i(A)$ and $i_Q(A)$.

Bucklew, Ney and Sadowsky [2] investigated the existence and uniqueness of an asymptotically efficient sampling distribution for empirical averages of Markov chains. The goal of this paper is to extend their results to Gibbs random fields in any dimension. As a special case, the Markov chain result is viewed in a new light.

Consider a Gibbs random field P_n, with values on a compact metric space S, on a d-dimensional finite box $\Lambda_n \uparrow \mathbb{Z}^d$, $n \to \infty$ (hence the number of sites $|\Lambda_n|$ replaces n from now on), that is associated with a continuous, translation-invariant and summable interaction potential $\mathscr{U}$. Fix any kind of boundary conditions, possibly free. Let $\mathscr{S}$ be another interaction potential as before and define T_n as the corresponding average energy over the box Λ_n.

Under natural conditions on A, we determine the asymptotically efficient importance sampling distribution in the class of finite volume Gibbs random fields with the given boundary condition. We prove that it is associated with the interaction potential $\mathscr{U} - t^*\mathscr{G}$, where t^* depends on A.

Precise definitions are given in Section 2, where we recall the main definitions and results concerning Gibbs random fields. In Section 3 we prove the importance sampling result and give an easy two-dimensional example. Section 4 is devoted to the special case of Markov chains. Finally, in Section 5 we collect some comments about the practical relevance of our result in applications.

2. Gibbs random fields and large deviations. We recall the basic ingredients for the specification of large deviations estimates for Gibbs random fields. It is convenient to proceed first in a more abstract setting. Because of our specific needs, we will consider general finite measures rather than just probabilities.

A real extended function $I: \mathscr{X} \to \mathbb{R} \cup \{+\infty\}$ on a topological space $\mathscr{X}$ is a *rate function* if it is bounded from below and its level sets $\{I \le l\}$ are compact for any $l \in \mathbb{R}$. Let now $\mathscr{X}$ be a Polish space and P_n be a sequence of finite measures on the Borel σ-algebra $\mathscr{B}(\mathscr{X})$. Let $\{c_n\}$ be a sequence such that $c_n \nearrow +\infty$. The sequence $\{P_n\}$ is said to satisfy a *large deviations principle* with rate function I and speed rate $\{c_n\}$ if:

(a) for every closed set $C \subset \mathscr{X}$,

$$\lim_{n\to\infty} \sup \frac{1}{c_n} \log P_n(C) \le - \inf_{x \in C} I(x);$$

(b) for every open set $O \subset \mathscr{X}$,

$$\lim_{n\to\infty} \inf \frac{1}{c_n} \log P_n(O) \ge - \inf_{x \in O} I(x).$$

Let $T: \mathscr{X} \to \mathbb{R}$ be a continuous function and let $R_n = P_n \circ T^{-1}$. Then the following *contraction principle* holds. See Varadhan [17] for a proof.

PROPOSITION 2.1. *If $\{P_n\}$ satisfies a large deviations principle with rate function I, then R_n satisfies a large deviations principle with the same speed rate and rate function given by*

$$i(u) = \inf_{x:\, T(x) = u} I(x).$$

In many applications, including ours, the rate function turns out to be convex. The next lemma, which is inspired by Comets [3], gives the general form of convex conjugate functions of rate functions contracted by a linear functional (we refer to the book by Ekeland and Temam [5] for general definitions).

LEMMA 2.2. *Let $\mathscr{X}$ be a locally convex Hausdorff space and let $\mathscr{X}^*$ be its topological dual. Let $I: \mathscr{X} \to \mathbb{R} \cup \{\infty\}$ be a rate function and $I^*: \mathscr{X}^* \to \mathbb{R} \cup \{\infty\}$ its conjugate; that is, $I^*(x^*) = \sup_x[\langle x, x^* \rangle - I(x)]$. For $y \in \mathscr{X}^*$ define*

$$i(u) = \inf_{x:\langle x, y\rangle = u} I(x).$$

Then i is a rate function and its conjugate is the function $i^*: \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ defined as

$$i^*(t) = I^*(ty).$$

Moreover, the following two facts are equivalent:

(i) $u \in \partial i^*(t)$;
(ii) *there exists $x' \in \mathscr{X}$ with $\langle x', y \rangle = u$ such that $x' \in \partial I^*(ty)$.*

PROOF. Because I is a rate function and $y \in \mathscr{X}^*$ defines a continuous function on $\mathscr{X}$, it is easy to see that i is still a rate function. Then

$$i^*(t) = \sup_v \{tv - i(v)\} = \sup_x \{t\langle x, y\rangle - I(x)\} = I^*(ty).$$

Next assume (i). Because $i^*(t) = \sup_v[tv - i(v)]$, the fact that $u \in \partial i^*(t)$ means that u is the point at which the supremum is attained so that

$$I^*(ty) = i^*(t) = tu - i(u) = tu - \inf_{x:\langle x, y\rangle = u} I(x).$$

Because I is a rate function, the preceding infimum is attained at x', which implies that

$$I^*(ty) = tu - I(x') = \langle x', ty\rangle - I(x'),$$

which means that $x' \in \partial I(ty)$. The converse is proved similarly. □

We now turn to Gibbs random fields. A general reference for most of the material that follows is Georgii's book [9]. Let S be a compact metric space. An *interaction potential* is a collection $\mathscr{U}$ of continuous functions $U_V: S^V \to \mathbb{R}$, indexed by V varying on all finite subsets of $\mathbb{Z}^d$. $\mathscr{U}$ is said to be *translation invariant* if

$$U_V = U_{V+i}$$

for all $i \in \mathbb{Z}^d$. $\mathscr{U}$ is said to be *summable* if

$$\|\mathscr{U}\| := \sum_{V: 0 \in V} \|U_V\|_\infty < +\infty.$$

We denote by $\mathscr{B}_s$ the Banach space of such summable and translation-invariant interactions, endowed with the norm $\|\mathscr{U}\|$. An interaction potential $\mathscr{U}$ has *finite range* if $U_V = 0$ for all sets V whose diameter is larger than some integer m. Finite-range interaction potentials are dense in $\mathscr{B}_s$.

For any $\mathscr{U} \in \mathscr{B}_s$, for any finite $\Lambda \subset \mathbb{Z}^d$ and any configuration x_{Λ^c} on S^{Λ^c} (called a boundary condition, abbreviated b.c.), define the *energy*

$$\mathscr{U}_\Lambda(x_\Lambda; x_{\Lambda^c}) = \sum_{V:\, V \cap \Lambda \neq \varnothing} U_V(x_V), \tag{2.1}$$

where x_V denotes the restriction of the configuration $x = (x_\Lambda, x_{\Lambda^c}) \in S^{\mathbb{Z}^d}$ on V and $x_\Lambda \in S^\Lambda$.

Let ρ denote a fixed probability distribution on S and let ρ_Λ denote the corresponding product measure on S^Λ for $\Lambda \in \mathbb{Z}^d$. We shall suppose that the support of ρ is S.

The *finite-volume $\mathscr{U}$-Gibbs distribution* on S^Λ given the b.c. x_{Λ^c} is defined as

$$\Pi^{\mathscr{U}}_{\Lambda, x_{\Lambda^c}}(ds_\Lambda) = \frac{1}{Z^{\mathscr{U}}_{\Lambda, x_{\Lambda^c}}} \exp\{-\mathscr{U}_\Lambda(x_\Lambda; x_{\Lambda^c})\} \rho_\Lambda(dx_\Lambda). \tag{2.2}$$

We consider a sequence of finite boxes $\Lambda_n \subset \mathbb{Z}^d$ increasing to $\mathbb{Z}^d$ as $n \to \infty$ and a sequence $x_{\Lambda_n^c}$ of b.c.'s (possibly free). None of the limits below depends on the particular choice of Λ_n and the convergence is always uniform in $x_{\Lambda_n^c}$. For the sake of simplicity, we assume the boxes to be symmetric around the origin. The limit

$$p(\mathscr{U}) := \lim_{n \to \infty} \frac{1}{|\Lambda_n|} \log Z^{\mathscr{U}}_{\Lambda_n, x_{\Lambda_n^c}} \tag{2.3}$$

exists and is called the *pressure*; it is independent of the given b.c. and finite for all $\mathscr{U} \in \mathscr{B}_s$. Moreover, the pressure p is convex and continuous over $\mathscr{B}_s$.

Let $\Omega = S^{\mathbb{Z}^d}$ be equipped with the product topology. Denote by $\mathscr{M}_s$ the family of stationary (translation-invariant) probability distributions on the product σ-algebra of Ω that, endowed with the weak topology, is a compact metric space.

An *infinite-volume $\mathscr{U}$-Gibbs distribution* corresponding to the potential $\mathscr{U}$ is any probability measure on Ω whose conditional distributions on S^Λ given x_{Λ^c} are almost surely equal to (2.2) for any Λ. The set of infinite volume translation-invariant $\mathscr{U}$-Gibbs distributions on Ω is denoted by $\Gamma_s(\mathscr{U})$ and is not empty. Two interaction potentials $\mathscr{U}$ and $\mathscr{V}$ are said to be *equivalent* if $\Gamma_s(\mathscr{U}) = \Gamma_s(\mathscr{V})$. It can be shown that $\Gamma_s(\mathscr{U}) \cap \Gamma_s(\mathscr{V}) \neq \varnothing$ implies equivalence among $\mathscr{U}$ and $\mathscr{V}$.

If $Q \in \mathscr{M}_s$, we denote by Q_{Λ_n} the restriction of Q to S^{Λ_n}. The *specific entropy* of Q is defined as

$$H(Q) := \lim_{n \to \infty} \frac{1}{|\Lambda_n|} E_{Q_{\Lambda_n}}\left(\log \frac{dQ_{\Lambda_n}}{d\rho_{\Lambda_n}}\right),$$

where H is set equal to $+\infty$ if Q_{Λ_n} is not absolutely continuous w.r.t. ρ_{Λ_n} for some n.

It is well known that the specific entropy, as a function on $\mathscr{M}_s$, is a rate function. Defining the pairing

$$\langle Q, \mathscr{U} \rangle := -E_Q(\mathscr{U}_0), \qquad Q \in \mathscr{M}_s, \mathscr{U} \in \mathscr{B}_s, \tag{2.4}$$

where

$$\mathscr{U}_0 = \sum_{V:\,0 \in V} \frac{1}{|V|} U_V,$$

$\mathscr{M}_s$ can be shown to be homeomorphic to a subset of $\mathscr{B}_s^*$. Thus we can extend H over all $\mathscr{B}_s^*$ assigning entropy $+\infty$ to each functional that cannot be represented as in (2.4). This extended entropy is clearly still a rate function on $\mathscr{B}_s^*$.

The *Gibbs variational principle* states that the pressure $p(\mathscr{U})$ and the entropy $H(Q)$ are convex conjugate functions. Moreover, the subgradient of p at $\mathscr{U}$ is equal to $\Gamma_s(\mathscr{U})$. This means that

$$p(\mathscr{U}) = -\inf_{Q \in \mathscr{M}_s} \{E_Q(\mathscr{U}_0) + H(Q)\} = \sup_{Q \in \mathscr{M}_s} \{\langle Q, \mathscr{U} \rangle - H(Q)\} \tag{2.5}$$

and the infimum in (2.5) is attained exactly at all the elements in $\Gamma_s(\mathscr{U})$.

Let

$$P_n := \Pi^{\mathscr{U}}_{\Lambda_n, x_{\Lambda_n^c}}. \tag{2.6}$$

Here is the large deviations results of interest for P_n.

Given $x_{\Lambda_n} \in S^{\Lambda_n}$, we extend this configuration periodically to the whole $\mathbb{Z}^d$ and we denote such a configuration by $x^{(n)}$. Define

$$D_{n, x^{(n)}} := \frac{1}{|\Lambda_n|} \sum_{i \in \Lambda_n} \delta_{\tau_i x^{(n)}}, \tag{2.7}$$

where τ_i is the shift operator on Ω. Clearly $D_{n, x^{(n)}}$ is an element of $\mathscr{M}_s$. Then the following result holds. (See Comets [3], Föllmer and Orey [6] and Olla [13].)

Proposition 2.3. *The distribution of $D_{n, x^{(n)}}$ under P_n satisfies a large deviations principle with speed rate $c_n = |\Lambda_n|$ and rate function*

$$I(Q) = H(Q) + E_Q(\mathscr{U}_0) + p(\mathscr{U}). \tag{2.8}$$

Next we specify the class of statistics T_n we are interested in. Let $\mathscr{G}$ be an interaction potential in $\mathscr{B}_s$ and denote by

$$T_n(x_{\Lambda_n}) := \frac{1}{|\Lambda_n|} \mathscr{G}_{\Lambda_n}(x_{\Lambda_n}; x_{\Lambda_n^c}) \tag{2.9}$$

the corresponding average energy per site of the configuration x_{Λ_n} over the box Λ_n. Also free b.c.'s are possible, in which case T_n becomes

$$\frac{1}{|\Lambda_n|}\sum_{V:\,V\subset\Lambda_n} G_V(x_V). \tag{2.10}$$

In particular, this covers the case of empirical averages of some local continuous function $g\colon S^V\to\mathbb{R}$, where V is a finite subset of $\mathbb{Z}^d$. It suffices to define the interaction potential as $\mathscr{G}=\{G_{V+i}:=g,\ \forall i,\ G_W:=0 \text{ otherwise}\}$. (2.9) and (2.10) differ by $o(|\Lambda_n|)$ from the quantity

$$\frac{1}{|\Lambda_n|}\sum_{i\in\Lambda_n}\mathscr{G}_0\big(\tau_i x^{(n)}\big) = -\langle D_{n,x^{(n)}},\mathscr{G}\rangle, \tag{2.11}$$

uniformly in x_{Λ_n} and in the b.c. $x_{\Lambda_n^c}$. This is because the three quantities (2.9), (2.10) and (2.11) only differ for the terms in which the boundary is involved (Georgii [9], Theorem 15.23). This fact easily enables us to derive a large deviations principle for the distributions of T_n directly as a consequence of the contraction principle. In fact, the final approximation argument in the proof of Theorem 6.1 in Olla [13] can be easily extended to non-finite-range interactions.

PROPOSITION 2.4. *The distribution of T_n under P_n satisfies a large deviations principle with speed rate $|\Lambda_n|$ and rate function*

$$i(u) = \inf_{\substack{Q\in\mathscr{M}_s\\ E_Q(\mathscr{G}_0)=u}} I(Q). \tag{2.12}$$

To compute the contracted rate function i, Lemma 2.2 turns out to be useful. First note that by the previous considerations, one easily has

$$\begin{aligned}
&\liminf_{n\to\infty}\left[\inf_{x_{\Lambda_n^c}}\inf_{x_{\Lambda_n}}\frac{1}{|\Lambda_n|}\mathscr{G}_{\Lambda_n}(x_{\Lambda_n};x_{\Lambda_n^c})\right]\\
&\qquad=\liminf_{n\to\infty}\left[\sup_{x_{\Lambda_n^c}}\inf_{x_{\Lambda_n}}\frac{1}{|\Lambda_n|}\mathscr{G}_{\Lambda_n}(x_{\Lambda_n};x_{\Lambda_n^c})\right]=:c_-,\\
&\limsup_{n\to\infty}\left[\inf_{x_{\Lambda_n^c}}\sup_{x_{\Lambda_n}}\frac{1}{|\Lambda_n|}\mathscr{G}_{\Lambda_n}(x_{\Lambda_n};x_{\Lambda_n^c})\right]\\
&\qquad=\limsup_{n\to\infty}\left[\sup_{x_{\Lambda_n^c}}\sup_{x_{\Lambda_n}}\frac{1}{|\Lambda_n|}\mathscr{G}_{\Lambda_n}(x_{\Lambda_n};x_{\Lambda_n^c})\right]=:c_+.
\end{aligned} \tag{2.13}$$

THEOREM 2.5.

(i) *The function $F(Q)=H(Q)-\langle Q,\mathscr{U}\rangle$, defined on $\mathscr{B}_s^*$, has the convex conjugate $F^*(\mathscr{V})=p(\mathscr{U}+\mathscr{V}),\ \mathscr{V}\in\mathscr{B}_s$.*

(ii) *The real function*

$$f(u) := \inf_{\substack{Q\in\mathscr{M}_s\\ E_Q(\mathscr{G}_0)=u}} F(Q) \tag{2.14}$$

has the convex conjugate $f^*(t)=p(\mathscr{U}-t\mathscr{G}),\ t\in\mathbb{R}$.

(iii) *If*

$$\mathrm{dom}(f) := \{v\colon E_Q(\mathscr{G}_0) = v,\, Q \in \mathscr{M}_s,\, H(Q) < +\infty\}, \tag{2.15}$$

then

$$\mathrm{d\delta m}(f) = (c_-, c_+).$$

In particular $\mathrm{d\delta m}(f)$ *is independent of* $\mathscr{U}$.

(iv) *For all* $u \in (c_-, c_+)$ *there exists* $t \in \mathbb{R}$ *such that the infimum in* (2.14) *is attained at all the elements* $Q \in \Gamma_s(\mathscr{U} - t\mathscr{G})$ *with* $E_Q(\mathscr{G}_0) = u$.

PROOF. (i) is an immediate consequence of the Gibbs variational principle. As for (ii), observe that

$$\begin{aligned} p(\mathscr{U} - t\mathscr{G}) &= -\inf_{Q \in \mathscr{M}_s} \{E_Q(\mathscr{U}_0 - t\mathscr{G}_0) + H(Q)\} \\ &= -\inf_u \left\{ \inf_{\substack{Q \in \mathscr{M}_s \\ E_Q(\mathscr{G}_0) = u}} \{E_Q(\mathscr{U}_0) + H(Q)\} - tu \right\} \\ &= -\inf_u \{f(u) - tu\} \\ &= f^*(t). \end{aligned}$$

To prove (iii), first observe that dom(f) is convex. Hence one has to show that for any $\varepsilon > 0$, there exist M and N in $\mathscr{M}_s$ with $E_M(\mathscr{G}_0) < c_- + \varepsilon$ and $E_N(\mathscr{G}_0) > c_+ - \varepsilon$, both with finite specific entropy. We prove the existence of M because the existence of N is obtained by the same argument. Choose Λ_n so that

$$\inf_{x_{\Lambda_n}} \mathscr{G}_{\Lambda_n}(x_{\Lambda_n}; x_{\Lambda_n^c}) < (c_- + \varepsilon/2)|\Lambda_n|.$$

Because of the following Lemma 2.6,

$$\lim_{t \to +\infty} E_{\Pi^{\mathscr{U} - t\mathscr{G}}_{\Lambda_n, x_{\Lambda_n^c}}} \left(\frac{1}{|\Lambda_n|} \mathscr{G}_{\Lambda_n} \right) \le c_- + \frac{\varepsilon}{2}$$

and the convergence is uniform in $x_{\Lambda_n^c}$. Thus for t and n large enough, if $M \in \Gamma_s(\mathscr{U} - t\mathscr{G})$,

$$\begin{aligned} E_M(\mathscr{G}_0) &= E_M\left(\frac{1}{|\Lambda_n|} \sum_{i \in \Lambda_n} \mathscr{G}_0 \circ \tau_i \right) \\ &\le E_M\left(\frac{1}{|\Lambda_n|} \mathscr{G}_{\Lambda_n} \right) + \frac{\varepsilon}{2} \\ &= E_M\left(E_{\Pi^{\mathscr{U} - t\mathscr{G}}_{\Lambda_n, x_{\Lambda_n^c}}} \left(\frac{1}{|\Lambda_n|} \mathscr{G}_{\Lambda_n} \right) \right) + \frac{\varepsilon}{2} \\ &\le c_- + \varepsilon. \end{aligned}$$

Concerning (iv), by a well known property of the Legendre–Fenchel transform (see, e.g., Rockafellar [15]), one has

$$\mathrm{d\mathring{o}m}(f) \subset \bigcup_t \partial f^*(t) \subset \mathrm{dom}(f).$$

Thus by Lemma 2.2, $u \in \partial f^*(t)$ if and only if there exists $M \in \mathscr{B}_s^*$ such that $-\langle M, \mathscr{G}\rangle = E_M(\mathscr{G}_0) = u$ with $M \in \partial F^*(-t\mathscr{G}) = \partial p(\mathscr{U} - t\mathscr{G})$. By Theorem 16.14 in Georgii [7], this means that $M \in \Gamma_s(\mathscr{U} - t\mathscr{G})$. □

LEMMA 2.6. *Let $Q_t = \Pi^{\mathscr{U}-t\mathscr{G}}_{\Lambda, x_{\Lambda^c}}$ be the $(\mathscr{U} - t\mathscr{G})$-Gibbs distribution on the finite volume Λ with b.c. x_{Λ^c} and denote by $m(x_{\Lambda^c})$ the minimum value of $\mathscr{G}_\Lambda(\cdot; x_{\Lambda^c})$. Let*

$$A(\varepsilon, x_{\Lambda^c}) = \{x_\Lambda \in S^\Lambda : \mathscr{G}_\Lambda(x_\Lambda; x_{\Lambda^c}) \le m(x_{\Lambda^c}) + \varepsilon\}.$$

Then all the limit points of Q_t as $t \to +\infty$ are concentrated on the set of minima of $\mathscr{G}_\Lambda(\cdot; x_{\Lambda^c})$. Moreover,

$$Q_t(A^c(\varepsilon, x_{\Lambda^c})) \to 0$$

as $t \to +\infty$ uniformly in the b.c. x_{Λ^c}.

PROOF. Because adding a constant to $\mathscr{G}_\Lambda$ does not change Q_t, we can assume that $m(x_{\Lambda^c}) \equiv 0$. Now

$$Q_t(A^c(\varepsilon, x_{\Lambda^c})) = \frac{1}{Z_t}\int_{A^c(\varepsilon, x_{\Lambda^c})} \exp(\mathscr{U}_\Lambda(x_\Lambda; x_{\Lambda^c}) - t\mathscr{G}_\Lambda(x_\Lambda; x_{\Lambda^c}))\, d\rho_\Lambda(x_\Lambda)$$

with

$$\begin{aligned} Z_t &= \int \exp(\mathscr{U}_\Lambda(x_\Lambda; x_{\Lambda^c}) - t\mathscr{G}_\Lambda(x_\Lambda; x_{\Lambda^c}))\, d\rho_\Lambda(x_\Lambda) \\ &\ge \int_{A(\varepsilon/2, x_{\Lambda^c})} \exp(\mathscr{U}_\Lambda(x_\Lambda; x_{\Lambda^c}) - t\mathscr{G}_\Lambda(x_\Lambda; x_{\Lambda^c}))\, d\rho_\Lambda(x_\Lambda) \\ &\ge e^{-t\varepsilon/2} \int_{A(\varepsilon/2, x_{\Lambda^c})} \exp(\mathscr{U}_\Lambda(x_\Lambda; x_{\Lambda^c}))\, d\rho_\Lambda(x_\Lambda) \end{aligned}$$

so that

$$\begin{aligned} Q_t(A^c(\varepsilon, x_{\Lambda^c})) &\le e^{-t\varepsilon/2} \frac{\int_{A^c(\varepsilon, x_{\Lambda^c})} \exp(\mathscr{U}_\Lambda(x_\Lambda; x_{\Lambda^c}))\, d\rho_\Lambda(x_\Lambda)}{\int_{A(\varepsilon/2, x_{\Lambda^c})} \exp(\mathscr{U}_\Lambda(x_\Lambda; x_{\Lambda^c}))\, d\rho_\Lambda(x_\Lambda)} \\ &\le \frac{Ce^{-t\varepsilon/2}}{\rho(A(\varepsilon/2, x_{\Lambda^c}))}, \end{aligned}$$

where $\log C$ is the difference between the maximum and the minimum of $\mathscr{U}_\Lambda$ on the pair $(x_\Lambda; x_{\Lambda^c})$. Finally, to see that the r.h.s. is finite, argue as follows. Because $\mathscr{G}_\Lambda(x_\Lambda; x_{\Lambda^c})$ is uniformly continuous in $(x_\Lambda; x_{\Lambda^c})$, $A(\varepsilon/2, x_{\Lambda^c})$ contains a ball of radius δ (independent of x_{Λ^c}). By compactness, it is easily seen that the measure of any such ball is bounded from below by a positive constant, because the support of ρ is the whole space. □

3. Importance sampling. Now we turn to our original problem. Recall that, by Propositions 1.3 and 2.4, the asymptotic behavior of $p_n(A) = P_n\{T_n \in A\}$, where P_n is defined in (2.6) and T_n is given in (2.9), is governed by a large deviations principle with rate function

$$i(u) = \inf_{\substack{Q \in \mathscr{M}_s \\ E_Q(\mathscr{S}_0) = u}} \left[H(Q) + E_Q(\mathscr{U}_0) + p(\mathscr{U}) \right]. \tag{3.1}$$

Let us choose the set of allowed importance sampling distributions as

$$\mathscr{C}_n = \left\{ Q_n = \Pi^{\mathscr{W}}_{\Lambda_n, x_{\Lambda_n^c}}, \mathscr{W} \in \mathscr{B}_s \right\}.$$

If Q_n is used as the importance sampling distribution, then the second moment of the importance sampling estimator $\hat{p}_n(A)$ is given by

$$\begin{aligned} v_{Q_n}(A) &= E_{Q_n}\left(1_{\{T_n \in A\}} \left(\frac{dP_n}{dQ_n} \right)^2 \right) \\ &= \frac{Z_n^{\mathscr{W}}}{\left[Z_n^{\mathscr{U}} \right]^2} \int_{\{T_n \in A\}} \exp\left(-\left(2\mathscr{U}_{\Lambda_n}(x_{\Lambda_n}; x_{\Lambda_n^c}) - \mathscr{W}_{\Lambda_n}(x_{\Lambda_n}; x_{\Lambda_n^c}) \right) \right) \rho_{\Lambda_n}(dx_{\Lambda_n}) \\ &= \frac{Z_n^{\mathscr{W}} Z_n^{2\mathscr{U} - \mathscr{W}}}{\left[Z_n^{\mathscr{U}} \right]^2} \tilde{Q}_n\{T_n \in A\}, \end{aligned}$$

where $\tilde{Q}_n = \Pi^{2\mathscr{U} - \mathscr{W}}_{\Lambda_n, x_{\Lambda_n^c}}$. Because

$$\lim_{n \to \infty} \frac{1}{|\Lambda_n|} \log \frac{Z_n^{\mathscr{W}} Z^{2\mathscr{U} - \mathscr{W}}}{\left[Z_n^{\mathscr{U}} \right]^2} = -2p(\mathscr{U}) + p(\mathscr{W}) + p(2\mathscr{U} - \mathscr{W}),$$

by applying Proposition 2.4 to $\tilde{Q}_n$, we obtain that the finite measure $v_{Q_n}(\cdot)$ also satisfies a large deviations principle with rate function

$$i_{\mathscr{W}}(u) = 2p(\mathscr{U}) - p(\mathscr{W}) + \inf_{\substack{Q \in \mathscr{M}_s \\ E_Q(\mathscr{S}_0) = u}} \left\{ H(Q) + 2E_Q(\mathscr{U}_0) - E_Q(\mathscr{W}_0) \right\}. \tag{3.2}$$

For any Borel set A define

$$i_{\mathscr{W}}(A) := \inf_{y \in A} i_{\mathscr{W}}(y),$$

$$i(A) := \inf_{y \in A} i(y).$$

Recall that $\{Q_n\}$, or equivalently the interaction potential $\mathscr{W}$, is asymptotically efficient if

$$i_{\mathscr{W}}(A) = i_{\mathscr{W}}(A^0) = 2i(A^0) = 2i(A). \tag{3.3}$$

We are now ready to obtain the asymptotically efficient importance sampling Gibbs random field.

THEOREM 3.1. *For any* $\mathscr{W} \in \mathscr{B}_s$,

$$i_{\mathscr{W}}(y) \le 2i(y) \quad \text{for every } y \in \mathbb{R}. \tag{3.4}$$

Moreover, for every $u \in (c_-, c_+)$ *there exists up to equivalence a unique interaction potential for which the equality in* (3.4) *is attained at* u. *It has the form* $\mathscr{W}^* = \mathscr{U} - t^*\mathscr{G}$, *where* t^* *is any solution of the equation in* t,

$$E_{Q_t}(\mathscr{G}_0) = u, \qquad Q_t \in \Gamma_s(\mathscr{U} - t\mathscr{G}). \tag{3.5}$$

PROOF. First note that $\operatorname{dom}(i_{\mathscr{W}}) = \operatorname{dom}(i) = \operatorname{dom}(f)$, according to (2.15). Hence, it suffices to prove (3.4) for $y \in (c_-, c_+)$. Write $\mathscr{W}$ as $\mathscr{U} - \mathscr{V}$. By (2.1), (2.2) and the Gibbs variational principle, we have

$$\begin{aligned} i_{\mathscr{U}-\mathscr{V}}(y) - 2i(y) &= -2 \inf_{\substack{Q \in \mathscr{M}_s \\ E_Q(\mathscr{G}_0) = y}} \{H(Q) + E_Q(\mathscr{U}_0)\} - p(\mathscr{U} - \mathscr{V}) \\ &\quad + \inf_{\substack{Q \in \mathscr{M}_s \\ E_Q(\mathscr{G}_0) = y}} \{H(Q) + E_Q(\mathscr{U}_0) + E_Q(\mathscr{V}_0)\} \\ &= -2 \inf_{\substack{Q \in \mathscr{M}_s \\ E_Q(\mathscr{G}_0) = y}} \{H(Q) + E_Q(\mathscr{U}_0)\} \\ &\quad + \inf_{Q \in \mathscr{M}_s} \{H(Q) + E_Q(\mathscr{U}_0) - E_Q(\mathscr{V}_0)\} \\ &\quad + \inf_{\substack{Q \in \mathscr{M}_s \\ E_Q(\mathscr{G}_0) = y}} \{H(Q) + E_Q(\mathscr{U}_0) + E_Q(\mathscr{V}_0)\}. \end{aligned} \tag{3.6}$$

Define

$$\begin{aligned} A(Q) &:= H(Q) + E_Q(\mathscr{U}_0) - E_Q(\mathscr{V}_0), \\ B(Q) &:= H(Q) + E_Q(\mathscr{U}_0) + E_Q(\mathscr{V}_0) \end{aligned}$$

and observe that

$$\begin{aligned} \inf_{Q \in \mathscr{M}_s} A(Q) + \inf_{\substack{Q \in \mathscr{M}_s \\ E_Q(\mathscr{G}_0) = y}} B(Q) &\le \inf_{\substack{Q \in \mathscr{M}_s \\ E_Q(\mathscr{G}_0) = y}} A(Q) + \inf_{\substack{Q \in \mathscr{M}_s \\ E_Q(\mathscr{G}_0) = y}} B(Q) \\ &\le \inf_{\substack{Q \in \mathscr{M}_s \\ E_Q(\mathscr{G}_0) = y}} [A(Q) + B(Q)]. \end{aligned} \tag{3.7}$$

Substituting this inequality back into (3.6) we get (3.4).

Moreover, if $u \in (c_-, c_+)$, by Theorem 2.5 there exists $t^* \in \mathbb{R}$ such that $E_{Q_{t^*}}(\mathscr{G}_0) = u$, $Q_{t^*} \in \Gamma_s(\mathscr{U} - t^*\mathscr{G})$. By the Gibbs variational principle the measure Q_{t^*} realizes the infimum

$$\inf_{\substack{Q\in\mathscr{M}_s\\ E_Q(\mathscr{G}_0)=u}} [A(Q)+B(Q)].$$

Take $\mathscr{V}=t^*\mathscr{G}$. Then, by the Gibbs variational principle again, the $\inf_{Q\in\mathscr{M}_s} A(Q)$ is attained at Q_{t^*} and, because $E_{Q_{t^*}(\mathscr{G}_0)}=u$, it is obvious that such a Q_{t^*} also realizes the infimum

$$\begin{aligned}\inf_{\substack{Q\in\mathscr{M}_s\\ E_Q(\mathscr{G}_0)=u}} B(Q) &= \inf_{\substack{Q\in\mathscr{M}_s\\ E_Q(\mathscr{G}_0)=u}} \{H(Q)+E_Q(\mathscr{U}_0)+E_Q(\mathscr{V}_0)\}\\ &= \inf_{\substack{Q\in\mathscr{M}_s\\ E_Q(\mathscr{G}_0)=u}} \{H(Q)+E_Q(\mathscr{U}_0)-t^*E_Q(\mathscr{G}_0)\}+2t^*u.\end{aligned}$$

Thus, by (3.6) written for $y=u$, the interaction potential $\Gamma_s(\mathscr{U}-t^*\mathscr{G})$ actually realizes the equality in (3.4).

Conversely, if in (3.4) equality holds for $y=u$, then by the preceding analysis (3.7) also must hold with the $\le$ sign replaced by $=$. If this happens, then all $Q\in\mathscr{M}_s$ that realize the infimum

$$\inf_{\substack{Q\in\mathscr{M}_s\\ E_Q(\mathscr{G}_0)=u}} [A(Q)+B(Q)] \tag{3.8}$$

necessarily also realize both the infima

$$\inf_{Q\in\mathscr{M}_s} A(Q),\qquad \inf_{\substack{Q\in\mathscr{M}_s\\ E_Q(\mathscr{G}_0)=y}} B(Q). \tag{3.9}$$

But we have already proved that the infimum in (3.8) is attained at Q_{t^*}. Thus Q_{t^*} also realizes the first infimum in (3.9), which by the Gibbs variational principle implies $Q_{t^*}\in\Gamma(\mathscr{U}-\mathscr{W})$. Thus the two interaction potentials $\mathscr{U}-\mathscr{W}$ and $t^*\mathscr{G}$ are equivalent. □

As a consequence of the last part in the proof, if $\mathscr{V}=\mathscr{U}-\mathscr{W}$, then $\mathscr{V}$ is equivalent to $t^*\mathscr{G}$. Hence, t^* is unique, except for the trivial case where $\mathscr{G}$ is equivalent to the uniformly zero potential, a situation that we shall exclude from now on.

Theorem 3.1 allows us to obtain the asymptotically efficient importance sampling distribution for $p_n(A_u^+)$ and $p_n(A_u^-)$, where $A_u^+=[u,+\infty)$ and $A_u^-=(-\infty,u]$.

Let

$$\begin{aligned}c_0^- &:= \inf_{Q\in\Gamma_s(\mathscr{U})} E_Q(\mathscr{G}_0),\\ c_0^+ &:= \sup_{Q\in\Gamma_s(\mathscr{U})} E_Q(\mathscr{G}_0).\end{aligned}$$

Because $u<c_0^-$ (resp. $u>c_0^+$) implies $p_n(A_u^+)\to 1$ [resp. $p_n(A_u^-)\to 1$], these cases are not interesting.

COROLLARY 3.2. *For $c_0^- \le u < c_+$ (resp. $c_- < u \le c_0^+$), $\mathscr{U} - t^*\mathscr{G}$ is, up to equivalence, the unique asymptotically efficient interaction potential for $p_n(A_u^+)$ [resp. $p_n(A_u^-)$], t^* being the unique real number such that $E_Q(\mathscr{G}_0) = u$ for some $Q \in \Gamma_s(\mathscr{U} - t^*\mathscr{G})$.*

PROOF. We prove the result for $p_n(A_u^+)$. In the sequel we shall denote $\mathscr{W}^* = \mathscr{U} - t^*\mathscr{G}$ to simplify the notations. Because $I(Q) = 0$ only if $Q \in \Gamma_s(\mathscr{U})$ and I is lower semicontinuous, $i(u) = 0$ if and only if $c_0^- \le u \le c_0^+$. Moreover, by convexity, i is nondecreasing and continuous on $[c_0^+, c^+)$. Thus

$$\inf_{y \in A_u^+} i(y) = \inf_{y \in A_u^{+0}} i(y) = i(u).$$

If $\mathscr{W}$ is asymptotically efficient, then $i_{\mathscr{W}}(u) = 2i(u)$, from which $\mathscr{W}$ is equivalent to $\mathscr{W}^*$. Hence, it is enough to prove the foregoing equality for $i_{\mathscr{W}^*}$, which is a consequence of the following elementary convexity argument. The incremental ratio

$$\frac{2i(z) - 2i(y)}{z - y}$$

is a nondecreasing function of z and y (this is true for every convex function). We know that $i(c_0^-) = 0$ and $i_{\mathscr{W}^*}(u) = 2i(u)$. Thus by (3.4) one has for $c_0^- \le y < u$,

$$\frac{i_{\mathscr{W}^*}(u) - i_{\mathscr{W}^*}(y)}{u - y} \ge \frac{2i(u) - 2i(y)}{u - y} \ge \frac{2i(u)}{u - c_0^+} \ge 0.$$

Because $i_{\mathscr{W}^*}$ is itself convex, the ratio

$$\frac{i_{\mathscr{W}^*}(u) - i_{\mathscr{W}^*}(y)}{u - y}$$

is nondecreasing in y and thus nonnegative also for $y > u$, from which it follows that $i_{\mathscr{W}^*}(y) \ge i_{\mathscr{W}^*}(u)$ for $y > u$. The result for $p_n(A_u^-)$ is proven analogously. □

Note that the previous corollary includes $i(A_u^+) = 0$, and in this case the original interaction potential $\mathscr{U}$ is asymptotically efficient. Of course, here the event $T_n \in A_u^+$ is not exponentially rare in the volume $|\Lambda_n|$.

For the sake of simplicity, we assumed the same boundary conditions as in P_n for all members of the class $\mathscr{C}$. It is not hard to see that a different choice of b.c.'s does not change the rate functions $i_{\mathscr{W}}$ and thus does not modify the asymptotically efficient interaction potential. At first sight this may seem surprising when the asymptotically efficient interaction potential has a phase transition, because the choice of b.c.'s effects the infinite volume limit points of the sequence Q_n. To see the effect of b.c.'s in the choice of the importance sampling distribution, one has to take into account in the asymptotic behavior of $v_{Q_n}(A)$ in (1.5) terms besides the leading term $\exp(-|\Lambda_n| i_{\mathscr{W}}(A))$. This is particularly clear in the case $c_0^- \le u \le c_0^+$ and $c_0^- < c_0^+$ [in which case

$i(A_u^+) = 0$ and there is a phase transition]. However, even in this situation, $p_n(A_u^+)$ may still decay exponentially, but at a slower rate than $|\Lambda_n|$. In some cases, the corresponding rate functions have been calculated (see Dobrushin, Kotecky and Shlosman [4]) and depend on the b.c.'s.

Equation (3.5) for t^* involves an infinite volume expectation. The next theorem allows approximation of t^* with t_n^*, computed by means of finite volume expectations.

THEOREM 3.3. *For $c_- < u < c_+$ and n sufficiently large there exists a unique solution t_n^* of*

$$\frac{1}{|\Lambda_n|} E_{\Pi^{\mathscr{U}-t\mathscr{S}}_{\Lambda_n, x_{\Lambda_n^c}}} \big(\mathscr{S}_{\Lambda_n}(x_{\Lambda_n}, x_{\Lambda_n^c}) \big) = u, \tag{3.10}$$

which converges to t^* as $n \to \infty$.

PROOF. Let

$$f_n(t) := \frac{1}{|\Lambda_n|} \log Z^{\mathscr{U}-t\mathscr{S}}_{\Lambda_n, x_{\Lambda_n^c}},$$

$$f^*(t) := p(\mathscr{U} - t\mathscr{S})$$

and recall that, by (2.13), $f_n(t) \to f^*(t)$. For every n the function f_n is smooth, with first derivative

$$f_n'(t) = \frac{1}{|\Lambda_n|} E_{\Pi^{\mathscr{U}-t\mathscr{S}}_{\Lambda_n, x_{\Lambda_n^c}}} \big(\mathscr{S}_{\Lambda_n}(x_{\Lambda_n}, x_{\Lambda_n^c}) \big).$$

Moreover it is strictly convex, because its second derivative

$$f_n''(t) = \frac{1}{|\Lambda_n|} \mathrm{Var}\big(\mathscr{S}_{\Lambda_n}(x_{\Lambda_n}, x_{\Lambda_n^c}) \big)$$

is strictly positive. By the definitions (2.13), if $u \in (c_-, c_+)$, and by Lemma 2.6 for n sufficiently large, there is a unique t_n^* such that $f_n'(t_n^*) = u$.

Suppose that t_n^* does not converge to t^*. Moreover, assume that there is a subsequence (again denoted by t_n^*) that converges to a finite limit $l < t^*$ (the cases $l > t^*$ and $l = +\infty$ can be handled similarly). Choose $\varepsilon > 0$ such that eventually $f_n'(l + \varepsilon) \geq u$. Then

$$\begin{aligned} f^*(t^*) - f^*(l + \varepsilon) &= \lim_{n \to \infty} \big(f_n(t^*) - f_n(l + \varepsilon) \big) \\ &= \lim_{n \to \infty} \int_{l+\varepsilon}^{t^*} f_n'(s)\, ds \geq (t^* - l - \varepsilon) u. \end{aligned}$$

Because a real convex function is absolutely continuous, the r.h.s. of the foregoing equation is equal to

$$\int_{l+\varepsilon}^{t^*} f^{*\prime}(s)\, ds < (t^* - l + \varepsilon) u,$$

the final inequality being obtained because almost everywhere for $s < t^*$,

$$f^{*\prime}(s) < D_- f^*(t^*) \le u \in \partial f^*(t^*). \qquad \square$$

It is also easy to verify that, in the asymptotically efficient importance sampling distribution of Corollary 3.2, one may replace t^* with t_n^*, without affecting the large deviations rate function of v_{Q_n}.

We finish this section with the elementary example of the two-dimensional Ising model, which also shows the effect of a phase transition.

In this case, we take $d = 2$, $S = \{-1, +1\}$ and $\mathscr{U} = \{U_V\}$, where $U_V = 0$ unless

$$U_{\{i\}}(x_i) = hx_i, \qquad i \in \mathbb{Z}^2,$$

$$U_{\{i,j\}}(x_i, x_j) = -\beta x_i x_j, \qquad |i - j| = 1,$$

for $h > 0$ and $\beta > \beta_c = 0.44068\ldots$(critical inverse temperature). The statistic we are interested in is the average magnetization

$$T_n = \frac{1}{|\Lambda_n|} \sum_{i \in \Lambda_n} X_i,$$

which corresponds to the single site interaction potential

$$G_{\{i\}}(x_i) = x_i, \qquad i \in \mathbb{Z}^2.$$

The interaction potential $\mathscr{U} - t\mathscr{G}$ gives rise to an Ising model with inverse temperature β and external field $t - h$. When $t = h$, we have

$$\Gamma_s(\mathscr{U} - h\mathscr{G}) = \left\{\lambda\pi_\beta^+ + (1 - \lambda)\pi_\beta^-, \lambda \in [0, 1]\right\},$$

where π_β^+ and π_β^- are different ergodic measures that are obtained one from the other by reversing the sign of the configurations. Let

$$m_\beta^+ := E_{\pi_\beta^+}(X_0) = -m_\beta^- := -E_{\pi_\beta^-}(X_0) > 0.$$

For $t \neq h$ the infinite volume $(\mathscr{U} - t\mathscr{G})$-Gibbs distribution is unique and denoted by Q_t. The function $E_{Q_t}(X_0)$ is well defined for $t \neq h$ and is strictly increasing from -1 to $+1$. At $t = h$ the function $t \to E_{Q_t}(X_0)$ has a jump from m_β^- to m_β^+. Let us denote by ξ the inverse of $t \to E_{Q_t}(X_0)$.

The asymptotically efficient interaction potential for sampling $\{T_n \in A_u^+\}$, where $u > c_0 = E_{Q_0}(\mathscr{G}_0)$, is given by $\mathscr{U} - \xi(u)\mathscr{G}$. The effect of phase transition at $t = h$ makes the choice of $t^* = h$ optimal for all $m_\beta^- \le u \le m_\beta^+$.

4. Markov chains. In this section we consider a Markov chain $\{X_n\}_n$ on a compact metric state space S with transition kernel P. We assume that $P(x, \cdot)$ has a derivative $P(x, y)$ w.r.t. a reference probability measure ρ that is jointly continuous and strictly positive [$P(x, y) \ge c > 0$ for some $c > 0$ and for every x, y]. Under these assumptions it is well known that there exists a unique stationary distribution π and that X_n converges in law to π as $n \to \infty$.

Bucklew, Ney and Sadowsky [2] determined the asymptotically efficient importance sampling distribution in the class of uniformly recurrent Markov chains for the estimation of

$$p_n(A) = \{T_n \in A\},$$

where $T_n = (1/n)\sum_{k=1}^n g(X_k)$ or, more generally, $T_n = (1/n)\sum_{k=1}^n g(X_{k-1}, X_k)$.

We now derive their results as a particular application of the theorems in Section 2. In addition, we are able to treat empirical averages of functions of the type $g(x_1, \dots, x_m)$ for any m [however, w.r.t. the results in reference [2] we need the extra assumption of the continuity of g and $P(x, y)$]. We deal with the case $m = 2$ and refer at the end to the more general situation.

First, we describe how the law of a Markov chain can be obtained as a Gibbs random field with an appropriate nearest neighbor interaction. (This material can also be found in Georgii [9].) Indeed, on the space $S^{\mathbb{Z}}$ consider the interaction potential $\mathscr{U}$ given by

$$U_V(x_V) := \begin{cases} -\log P(x_k, x_{k+1}), & \text{if } V = \{k, k+1\}, \\ 0, & \text{otherwise.} \end{cases}$$

The law P on $S^{\mathbb{Z}}$ of the stationary Markov chain with transition kernel P coincides with the infinite volume $\mathscr{U}$-Gibbs distribution, which is unique because in this case there is no phase transition. It is easily checked that the finite volume $\mathscr{U}$-Gibbs distribution on S^{Λ_n}, where $\Lambda_n = \{1, \dots, n\}$, with fixed b.c. x_0 at 0 and free b.c. at the other end, coincides with the distribution of the Markov chain up to time n starting at x_0:

$$P\{T_n \in A\} = \Pi^{\mathscr{U}}_{\Lambda_n, x_0}\{T_n \in A\}.$$

Conversely, a Markov chain can be associated with any translation-invariant interaction potential $\mathscr{U}$ such that $U_V = 0$ unless $V = V_k = \{k, k+1\}$ for some k: Consider the transition kernel associated with the density

$$P(x, y) = \exp\bigl(-U_{V_0}(x, y)\bigr)\frac{r(y)}{\lambda r(x)}, \tag{4.1}$$

where λ is the principal eigenvalue of the operator

$$K_{\mathscr{U}} f(x) = \int \exp\bigl(-U_{V_0}(x, y)\bigr) f(y)\,\rho(dy)$$

and r is the corresponding eigenfunction (which is strictly positive). The unique translation-invariant infinite volume $\mathscr{U}$-Gibbs random field coincides then with the stationary Markov chain associated to P, because it results by computing the density of $X_1, \dots, X_n$, conditioned to $X_0 = x_0$ and $X_{n+1} = x_{n+1}$:

$$\begin{aligned} &\frac{P(x_0, x_1) \cdots P(x_n, x_{n+1})}{P^{n+1}(x_0, x_{n+1})} \\ &\quad = k(x_0, x_{n+1}) \exp\bigl(-U_{V_0}(x_0, x_1)\bigr) \cdots \exp\bigl(-U_{V_0}(x_n, x_{n+1})\bigr) \\ &\quad = \frac{d\Pi^{\mathscr{U}}_{\Lambda_n, (x_0, x_{n+1})}}{d\rho_{\Lambda_n}}(x_1, \dots, x_n). \end{aligned}$$

In the correspondence between Markov chains and Gibbs random fields, of course, equivalent interaction potentials induce the same Markov chain. In particular, if two interaction potentials satisfy the relation

$$\tilde{U}_{V_0}(x, y) = U_{V_0}(x, y) + \psi(y) - \psi(x),$$

where ψ is any continuous function on S, then by (4.1) they induce the same Markov chain.

For a continuous function $g(x, y)$ consider the finite-range interaction potential

$$G_V(x_V) := \begin{cases} g(x_k, x_{k+1}), & \text{if } V = \{k, k+1\}, \\ 0, & \text{otherwise.} \end{cases}$$

Applying Corollary 3.2, the asymptotically efficient importance sampling distribution is the Gibbs random field Q_{t^*} associated with the interaction potential $\mathscr{U} - t^*\mathscr{G}$, where t^* satisfies $E_{Q_{t^*}}(g(X_0, X_1)) = u$. If g is not of the form $g(x, y) = \psi(x) - \psi(y)$ for some ψ (in which case, as we have already remarked, $\mathscr{G}$ is equivalent to the identically zero potential), such a value of t^* is unique again by Theorem 3.1.

Hence by the foregoing equivalence between Markov chains and Gibbs random fields, Q_{t^*} is the law of the stationary Markov chain with transition kernel having density

$$P_{t^*}(x, y) = P(x, y)e^{t^* g(x, y)} \frac{r_{t^*}(y)}{\lambda(t^*) r_{t^*}(x)},$$

where $\lambda(t^*)$ is the principal eigenvalue of the operator associated with the kernel $P(x, y)e^{t^*(x, y)}$ and $r_{t^*}(\cdot)$ the associated eigenfunction. Thus we obtain Theorem 3 in Bucklew, Ney and Sadowsky [2], provided we show that t^* is exactly the solution of

$$\frac{d}{dt} \log \lambda(t) = u.$$

This comes from the fact that $\log \lambda(t)$ is easily seen to be the pressure $p(\mathscr{U} - t\mathscr{G})$. By the Gibbs variational principle, the pressure p is differentiable (in the absence of phase transition) and its Frechet derivative is given by

$$Dp(\mathscr{W})(\mathscr{V}) = \langle Q, \mathscr{V}_0 \rangle,$$

where $Q \in \Gamma_s(\mathscr{W})$ so that, by the chain rule,

$$\begin{aligned} \frac{d}{dt} \log \lambda(t) &= -\langle Q_t, \mathscr{G} \rangle = E_{Q_t}(\mathscr{G}_0) = E_{Q_t}\left(\frac{1}{2}(g(X_0, X_1) + g(X_{-1}, X_0)) \right) \\ &= E_{Q_t}(g(X_0, X_1)). \end{aligned}$$

In case the function g is a function of m variables, with $m > 2$, our theory still can be applied, and we obtain an importance sampling distribution, which, however, is not a standard Markov chain anymore, but an $(m-1)$-dependent Markov chain.

5. Discussion. There are two main objections to the practical relevance of the results presented in this paper. The aim of this final section is to discuss them and point out where some further research is needed.

The first issue concerns the difficulty of getting N independent samples from Gibbs random fields in two or more dimensions. The now standard procedure for sampling from a finite volume $\mathscr{U}$-Gibbs random field is to run an ergodic Markov process on the configuration space S^{Λ_n} having as an invariant measure $\Pi^{\mathscr{U}}_{\Lambda_n, x_{\Lambda_n^c}}$. Several relaxation methods of this type have been proposed, among which are the Metropolis and the Gibbs sampler (see Geman [8] for a review). In this context, sampling from $\Pi^{\mathscr{U}}_{\Lambda_n, x_{\Lambda_n^c}}$ requires running the algorithm until some relaxation time is elapsed, which is needed for the law of the process to get close to the invariant measure.

Once equilibrium is reached, the algorithm produces dependent samples, nevertheless drawn approximately from the correct distribution. It is preferable to take all these samples into account rather than to discard them, until another nearly independent sample is produced. We shall refer to such a procedure as *dynamic sampling*, as opposed to the independent (*static*) sampling we have been considering in this paper.

Thus, once a dynamic sampling algorithm is chosen (e.g., the Gibbs sampler with uniformly chosen single site update [8]), different interaction potentials could be compared, also taking into account the "dynamic" properties induced by such a relaxation method. A natural criterion is the asymptotic variance

$$\lim_{N\to\infty} N \operatorname{Var}\big(\hat{p}_n(A)(X^{(1)},\dots,X^{(N)})\big) =: \sigma^2_{Q_n}(A),$$

where $\{X^{(k)}\}$ are the iterations of the Gibbs sampler for Q_n.

This quantity is certainly more complicated to deal with (see Frigessi, Hwang and Younes [7]) and one might guess that such analysis will favor those potentials for which the corresponding algorithm converges faster to equilibrium. In this context, we observe that the Ising example at the end of Section 2 may give a result that is computationally not very attractive. In fact, for some values of u the asymptotically efficient interaction potential has a phase transition that will slow down the relaxation time in large boxes. On the other hand, when the interaction is low, we can show that the static analysis is actually still the relevant one.

In fact, let $f_n = 1_{\{T_n \in A\}}(dP_n/dQ_n)$ so that, from (1.2),

$$\hat{p}_n(A) = \frac{1}{N}\sum_{k=1}^{N} f_n(X^{(k)}).$$

Hence

$$\sigma^2_{Q_n}(A) = \big(v_{Q_n}(A) - p_n^2(A)\big)\sum_{k=-\infty}^{+\infty} \rho^{(k)}_{Q_n},$$

where $\{\rho_{Q_n}^{(k)}\}$ is the autocorrelation sequence of the process $\{f_n(X^{(k)})\}$. Therefore,

$$\frac{1}{n}\log\frac{\sigma_{Q_n}(A)}{p_n^2(A)} = \frac{1}{n}\log\frac{\left(v_{Q_n}(A) - p_n^2(A)\right)}{p_n^2(A)} + \frac{1}{n}\log\sum_{k=-\infty}^{+\infty}\rho_{Q_n}^{(k)}.$$

By our results, the first term of the r.h.s converges (as $n \to \infty$) to a positive limit if and only if Q_n is not asymptotically efficient. The second term, the *integrated autocorrelation time*, measures the loss in efficiency due to dependent sampling. Its growth can be controlled by the reciprocal of the spectral gap γ_n between 1 and the second largest eigenvalue of the Gibbs sampler (see Frigessi, Hwang and Younes [7]). Hence, whenever the asymptotically efficient importance sampling distribution is such that $\log(\gamma_n) = o(n)$, the effect of dynamic sampling disappears in the limit. A sufficient condition for this is the Dobrushin ergodicity condition (see Georgii [9]), which is stronger than the absence of phase transition.

The second issue that should be taken into account is that two a priori computations are needed to identify the asymptotically efficient importance sampling distribution. Namely, we need (a) to solve (3.10) and (b) to compute the Radon–Nikodym derivative dP_n/dQ_n, which contains as a factor the ratio of the partition functions $Z_n^{\mathscr{U}-t_n^*\mathscr{G}}$ and $Z_n^{\mathscr{U}}$. In general, there is no way to solve these two problems analytically. They are, however, connected by the formula

$$\frac{Z_{\Lambda_n, x_{\Lambda_n}^c}^{\mathscr{U}-t_n^*\mathscr{G}}}{Z_{\Lambda_n, x_{\Lambda_n}^c}^{\mathscr{U}}} = E_{\Lambda_n, x_{\Lambda_n}^c}^{\mathscr{U}-t_n^*\mathscr{G}}\left(e^{t_n^*\mathscr{G}_{\Lambda_n}}\right) = \exp\left(\int_0^{t_n^*} E_{\Lambda_n, x_{\Lambda_n}^c}^{\mathscr{U}-s\mathscr{G}}(\mathscr{G}_{\Lambda_n})\,ds\right),$$

where the expectations are taken in the finite volume Λ_n with the b.c. $x_{\Lambda_n^c}$. Any algorithm that computes $E_{\Lambda_n, x_{\Lambda_n}^c}^{\mathscr{U}-t\mathscr{G}}(\mathscr{G}_{\Lambda_n})$ for various values of t will also serve to solve (3.10). However, this can be done in general again by Monte Carlo computations. This is suggested in Green [12] and Geyer and Thompson [10], for the computation of maximum likelihood estimators for Gibbs fields, which is formally analogous to (3.10). The use of stochastic relaxation algorithms, such as in Younes [18], seems to be quite appropriate in this respect. Because such a priori computations needed to identify the asymptotically efficient importance sampling distributions are again of Monte Carlo type, it seems reasonable to try to use these preliminary runs, where t is varied, immediately for the approximation of $p_n(A)$ itself. The resulting importance sampling estimator will then be an average of the (correlated) samples produced for different values of t. Such an algorithm deserves a computational evaluation, even if it seems hard to analyze it rigorously.

Acknowledgment. The authors wish to thank Professor Peter Glynn and Professor Peter J. Green for helpful discussions during their visits to Rome.

REFERENCES

[1] BUCKLEW, J. A. (1990). *Large Deviations Techniques in Decision, Simulation, and Estimation*. Wiley, New York.

[2] BUCKLEW, J. A., NEY, P. and SADOWSKY, J. S. (1990). Monte Carlo simulation and large deviations theory for uniformly recurrent Markov chains. *J. Appl. Probab.* **27** 44–59.

[3] COMETS, F. (1992). On Consistency of a class of estimators for exponential families of Markov random fields on the lattice. *Ann. Statist.* **20** 455–468.

[4] DOBRUSHIN, R. L., KOTECKY, R. and SHLOSMAN, S. (1992). *Wulff Construction: A Global Shape from Local Interaction*. Amer. Math. Soc., Providence, RI.

[5] EKELAND, I. and TEMAM, R. (1976). *Convex Analysis and Variational Problems*. North-Holland, Amsterdam.

[6] FÖLLMER, H. and OREY, S. (1988). Large deviations for the empirical field of a Gibbs measure. *Ann. Probab.* **16** 961–977.

[7] FRIGESSI, A., HWANG, C. R. and YOUNES, L. (1992). Optimal spectral structure of reversible stochastic matrices, Monte Carlo methods and the simulation of Markov random fields. *Ann. Appl. Probab.* **2** 610–628.

[8] GEMAN, D. (1990). *Random Fields and Inverse Problems in Imaging. Lecture Notes in Math.* **1427**. Springer, Berlin.

[9] GEORGII, H.-O. (1988). *Gibbs Measures and Phase Transitions*. de Gruyter, Berlin.

[10] GEYER, C. J. and THOMPSON, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *J. Roy. Statist. Soc. B* **54** 657–700.

[11] GLYNN, P. W. and IGLEHART, D. L. (1989). Importance sampling for stochastic simulations. *Management Sci.* **35** 1367–1392.

[12] GREEN, P. (1992). Comment on "Constrained Monte Carlo maximum likelihood for dependent data" by C. J. Geyer and E. A. Thompson. *J. Roy. Statist. Soc. B* **54** 683–684.

[13] OLLA, S. (1988). Large deviations for Gibbs random fields. *Probab. Theory Related Fields* **77** 343–357.

[14] ROBERTS, W. A. and VARBERG, D. E. (1973). *Convex Functions*. Academic, New York.

[15] ROCKAFELLAR, R. T. (1970). *Convex Analysis*. Princeton Univ. Press.

[16] VARADHAN, S. R. S. (1966). Asymptotic probabilities and differential equations. *Commun. Pure Appl. Math.* **19** 261–286.

[17] VARADHAN, S. R. S. (1984). *Large Deviations and Applications*. SIAM, Philadelphia.

[18] YOUNES, L. (1988). Estimation and annealing for Gibbsian fields. *Ann. Inst. H. Poincaré* **24** 269–294.

PAOLO BALDI
DIPARTIMENTO DI MATEMATICA
AND CENTRO VITO VOLTERRA
IL UNIVERSITÀ DI ROMA
VIA DEL FONTANILE DI CARCARICOLA
00133 ROMA
ITALY

ARNOLDO FRIGESSI
LABORATORIO DI STATISTICA
DELL'UNIVERSITÀ DI VENEZIA
DORSODURO 3246
30123 VENEZIA
ITALY

MAURO PICCIONI
DIPARTIMENTO DI MATEMATICA
UNIVERSITÀ DI L'AQUILA
67100 L'AQUILA
ITALY