Information bounds for Gibbs samplers

Priscilla E. Greenwood* — University of British Columbia

Ian W. McKeague † — Florida State University

Wolfgang Wefelmeyer — University of Siegen

Abstract

If we wish to efficiently estimate the expectation of an arbitrary function on the basis of the output of a Gibbs sampler, which is better: deterministic or random sweep? In each case we calculate the asymptotic variance of the empirical estimator, the average of the function over the output, and determine the minimal asymptotic variance for estimators that use no information about the underlying distribution. The empirical estimator has noticeably smaller variance for deterministic sweep. The variance bound for random sweep is in general smaller than for deterministic sweep, but the two are equal if the target distribution is continuous. If the components of the target distribution are not strongly dependent, the empirical estimator is close to efficient under deterministic sweep, and its asymptotic variance approximately doubles under random sweep.

1 Introduction

The Gibbs sampler is a widely used Markov chain Monte Carlo (MCMC) method for estimating analytically intractable features of multi-dimensional distributions. Interest in the method has been spurred by the resurgence of Bayesian statistics in recent years. While other MCMC methods can be more suitable in some Bayesian applications, the Gibbs sampler is often seen as the default option, e.g., in the `BUGS` software package, see Spiegelhalter et al. (1996).

*Partially supported by the Deutsche Forschungsgemeinschaft.

†Partially supported by NSF Grant ATM-9417528.

[1] Work supported by NSERC, Canada.

[2] *AMS* 1991 *subject classifications.* Primary 62M05, 62G20.

[3] *Key words and phrases.* Efficient estimator, empirical estimator, Markov chain Monte Carlo, variance bound.

A given MCMC method may be judged by various criteria. One is the speed at which the Markov chain converges to its target distribution, the stationary distribution of the chain. This is a well-studied problem; recent references are Frigessi, Hwang, Sheu and Di Stefano (1993), Ingrassia (1994), Meyn and Tweedie (1994), Roberts and Polson (1994), Rosenthal (1995), Mengersen and Tweedie (1996), Roberts and Tweedie (1996) and Roberts and Sahu (1997).

It is common to discard the initial observations (burn-in), until the sampler is thought to be close to its stationary distribution π. Then, to extract information from the remaining observations, $X^0, \dots, X^n$ say, the empirical estimator $E_n f = \frac{1}{n}\sum_{i=1}^n f(X^i)$ is used to approximate the expectation πf of any given function of interest f. Provided the burn-in period is relatively short, it is reasonable to judge the sampler by the asymptotic variance of the empirical estimator. This criterion is utilized by Peskun (1973), Frigessi, Hwang and Younes (1992), Green and Han (1992), Liu, Wong and Kong (1994, 1995), Clifford and Nicholls (1995), Liu (1996), Fishman (1996) and Frigessi and Rue (1998).

Here we consider a third criterion by which MCMC methods can be judged: How much information about πf is contained in the sample $X^0, \dots, X^n$? In particular: What fraction of the information is exploited by the empirical estimator? Is it worthwhile to construct improved estimators?

While the first two criteria are essentially probabilistic, the third is statistical and appears to have received comparatively little attention. The statistical approach is to view π as an infinite-dimensional parameter of the transition distribution driving the sampler. To study the above questions, one needs to determine the minimal asymptotic variance of estimators of πf in the sense of an infinite-dimensional version of Hájek's convolution theorem. The variance bound is also called the *information bound* for estimating πf.

We are led to a class of statistical problems on which some progress has been made recently: Given a family of transition distributions, how well can one estimate the invariant law on the basis of realizations of the corresponding Markov chain? The answer depends very much on the type of model. If nothing is known about the transition distribution, then the empirical estimator is efficient for πf; see Penev (1991), Bickel (1993) and Greenwood and Wefelmeyer (1995). It does not help to know that the chain is reversible; see Greenwood and Wefelmeyer (1998). If one has a parametric model for π, one can use what would be the maximum likelihood estimator if the observations were independent; Kessler, Schick and Wefelmeyer (1998) show that one can do even better.

In the present paper, we are concerned with the information in the simulated values $X^0, \dots, X^n$, given the knowledge that a Gibbs sampler was used to generate them. It is assumed that no information about π itself is made available to the statistician, apart from the link between π and the transition distribution of the observed Markov chain. Of course, π is known in principle, and part of that knowledge can sometimes be exploited to improve upon the empirical estimator. An example is Rao–Blackwellization, which consists in taking an appropriate conditional expectation of the empirical estimator; see Gelfand and Smith (1990, 1991), Schmeiser and Chen (1991), Liu, Wong and Kong (1994), Geyer (1995) and Casella and Robert (1996). Indeed, McKeague and Wefelmeyer (1998) show that an estimator of πf with arbitrarily small asymptotic variance is obtained, at least theoretically, by repeated Rao–Blackwellization of the empirical estimator. Symmetries of π are exploited in Greenwood, McKeague and Wefelmeyer (1996). If π is a random field on a lattice and the interactions between the sites are known to be local and relatively weak, the so-called *von Mises type statistics* of Greenwood, McKeague and Wefelmeyer (1998) also lead to considerable variance reduction over the empirical estimator. In the present setting, however, we avoid the use of any specific structural knowledge about π, so our conclusions apply to generic Gibbs samplers.

We consider Gibbs samplers with deterministic and random sweeps. The chain $X^0, X^1, \ldots$ is formed by updating a single component at each step. Under deterministic sweep, the sampler cycles through the components in some fixed order. Under random sweep, each step consists of choosing a component at random and then updating it. In the literature, a deterministic sweep sampler with k components is usually taken to be the *subchain* $X^0, X^k, X^{2k}, \ldots$, which changes only when a full cycle of updates is completed. The empirical estimator based on this 'coarse' chain is expected to have larger asymptotic variance than $E_n f$, at least in balanced situations, see Greenwood, McKeague and Wefelmeyer (1996). Subsampling of the coarse chain further increases the asymptotic variance, as observed by Geyer (1992, Theorem 3.3) and McEachern and Berliner (1994).

Our principal objective is to assess the efficiency of the empirical estimator $E_n f$ based on the 'fine' chain observations $X^0, X^1, \ldots, X^n$ under both deterministic and random sweeps. Here, efficiency is defined as the ratio of the information bound to the asymptotic variance of the estimator. The main points are as follows:

- If π has only two components, then the empirical estimator is fully efficient under deterministic sweep.
- For random sweep, the efficiency of the empirical estimator is at best only slightly more than 50%, but close to 50% if π is continuous.
- The information bound is smaller for random sweep than for deterministic sweep except when π is continuous, in which case the bounds coincide. The information bound for deterministic sweep does not depend on the order of the sweep.
- The asymptotic variance of the empirical estimator under random sweep is no more than twice that under deterministic sweep.

If the components of π are not strongly dependent:

- The empirical estimator is close to efficient under (any) deterministic sweep.
- The asymptotic variance of the empirical estimator under random sweep is close to twice that under deterministic sweep.

The paper is organized as follows. Section 2 collects various facts about Gibbs samplers used in the sequel. In Section 3 we calculate the asymptotic variance of $E_n f$ and the information bound of regular estimators of πf under deterministic sweep. Section 4 develops similar results for random sweep. The main points above are discussed in Section 5. Two simulation examples are presented in Section 6. Proofs are collected in Section 7.

2 Preliminaries

In this section we describe Gibbs samplers with deterministic and random sweep and collect some well-known properties of their transition distributions which will be used later. The presentation is rather detailed because our proofs require a careful description of the transition distributions.

Let $E = E_1 \times \ldots \times E_k$ be a product of measurable spaces, with product σ-field, and π a distribution on E. For each $j = 1, \ldots, k$, we can express $x \in E$ by separating out the jth component: $x = (x_j, x_{-j})$, where x_{-j} is obtained from x by omitting the j-th component x_j. Factor π in k different ways,

$$\pi(dx) = m_{-j}(dx_{-j})p_j(x_{-j}, dx_j), \quad j = 1, \ldots, k, \tag{2.1}$$

with $p_j(x_{-j}, dx_j)$ the one-dimensional conditional distribution under π of x_j given x_{-j}, and $m_{-j}(dx_{-j})$ the $(k-1)$-dimensional marginal distribution of x_{-j}.

Gibbs samplers successively use the transition distributions

$$Q_j(x, dy) = p_j(x_{-j}, dy_j)\varepsilon_{x_{-j}}(dy_{-j})$$

which change only the j-th component of x. Such samplers are defined inductively as follows. At time 0, choose an initial value $X^0 = (X_1^0, \ldots, X_k^0)$. At time i, pick an index j and replace X^{i-1} by a value X^i generated from $Q_j(X^{i-1}, dy)$. It follows from (2.1) that each Q_j has invariant law π. At time n we have created $n+1$ realizations $X^0, \ldots, X^n$ of a Markov chain with invariant law π. A specific sampler corresponds to a specific sampling scheme for picking the indices.

Before describing specific sampling schemes, we recall some properties of Q_j. Let $\langle \cdot, \cdot \rangle_2$ denote the inner product, and $\| \cdot \|_2$ the norm (or, depending on the context, the operator norm) on $L_2(\pi)$. By definition of Q_j,

$$Q_j(x, h) = \int Q_j(x, dy)h(y) = \int p_j(x_{-j}, dx_j)h(x_{-j}, x_j) = p_j(x_{-j}, h). \tag{2.2}$$

In particular, $Q_j(x, dy)$ does not depend on x_j. Hence Q_j is *idempotent*,

$$Q_j^2 = Q_j. \tag{2.3}$$

This means that Q_j is a *projection operator* on $L_2(\pi)$. Indeed, we can write $L_2(\pi)$ as the orthogonal sum of two subspaces, one consisting of functions h with $Q_j h = 0$, the other of functions $h(x)$ not depending on x_j, and Q_j is the projection on the second subspace along the first. Therefore,

$$\langle h,\, Q_j h'\rangle_2 = \langle Q_j h,\, Q_j h'\rangle_2 \quad \text{for } h, h' \in L_2(\pi). \tag{2.4}$$

Relation (2.4) implies that Q_j, as an operator on $L_2(\pi)$, is *positive*,

$$\langle h,\, Q_j h\rangle_2 = \langle Q_j h,\, Q_j h\rangle_2 \geq 0 \quad \text{for } h \in L_2(\pi), \tag{2.5}$$

and *selfadjoint*,

$$\langle h,\, Q_j h'\rangle_2 = \langle Q_j h,\, Q_j h'\rangle_2 = \langle Q_j h,\, h'\rangle_2 \quad \text{for } h, h' \in L_2(\pi). \tag{2.6}$$

The last relation is seen to be equivalent to *detailed balance*,

$$\pi(dx) Q_j(x, dy) = \pi(dy) Q_j(y, dx). \tag{2.7}$$

This, in turn, implies again that Q_j has invariant law π. From (2.2) we obtain

$$(Q_{j_1} \cdots Q_{j_s})(x, h) = (p_{j_1} \cdots p_{j_s})(x_{-j_1}, h) \quad \text{for } h \in L_2(\pi).$$

We will occasionally use this identity.

We focus on two particular sampling schemes.

Deterministic sweep. For the Gibbs sampler with *deterministic* (and *cyclic*) sweep, the indices $j = 1, \ldots, k$ are numbered in a fixed way, and then the Q_j are applied cyclically according to this numbering. We will not compare different deterministic sweeps, but the numbering of the indices is arbitrary. The transition distribution of the corresponding Markov chain at time $i = (q-1)k + j$ is Q_j.

The chain is not homogeneous, but the blocks $(X^{(q-1)k+1}, \ldots, X^{qk})$, with $q = 1, 2, \ldots$, form a homogeneous Markov chain, with transition distribution $Q_1 \otimes \cdots \otimes Q_k$. For $j = 1, \ldots, k-1$, the realization $X^{(q-1)k+j}$ is determined by $X^{(q-1)k}$ and X^{qk} as $(X^{qk}_{\leq j}, X^{(q-1)k}_{>j})$, where

$$x_{\leq j} = (x_1, \ldots, x_j), \quad x_{>j} = (x_{j+1}, \ldots, x_k).$$

Hence nothing is lost if we observe only the chain $X^{(q-1)k}$, $q = 1, 2, \ldots$. By Gibbs sampler one often means this subchain. The subchain is homogeneous, with transition distribution

$$Q_{(d)}(x, dy) = (Q_1 \cdots Q_k)(x, dy) = \prod_{j=1}^{k} p_j(y_{<j}, x_{>j}, dy_j). \tag{2.8}$$

The subscript (d) stands for *deterministic*. The transition distribution $Q_{(d)}$ is in general neither positive nor selfadjoint. The adjoint $Q^*_{(d)}$ of $Q_{(d)}$ is obtained by reversing the order of the sweep, $Q^*_{(d)} = Q_k \cdots Q_1$. This follows by repeated application of the basic relation (2.4). For $k = 2$,

$$\langle h,\, Q_i Q_j h' \rangle_2 = \langle Q_i h,\, Q_j h' \rangle_2 = \langle Q_j Q_i h,\, h' \rangle_2.$$

Similarly, for $j = 1, \ldots, k-1$, the subchain $(X^{qk}_{\le j}, X^{(q-1)k}_{>j})$, $q = 1, 2, \ldots$, is homogeneous, with transition distribution $Q_{j+1} \cdots Q_k Q_1 \cdots Q_j$.

Random sweep. For the Gibbs sampler with *random* sweep (with *equal* probabilities), each index j is picked according to the uniform distribution on $1, \ldots, k$, independently at successive time steps. The transition distribution of the corresponding Markov chain at each time is

$$Q_{(r)}(x, dy) = \frac{1}{k} \sum_{j=1}^{k} Q_j(x, dy) = \frac{1}{k} \sum_{j=1}^{k} p_j(x_{-j}, dy_j) \varepsilon_{x_{-j}}(dy_{-j}). \tag{2.9}$$

The subscript (r) stands for *random*. This chain is homogeneous. Since the Q_j are positive by (2.5), so is their average, $Q_{(r)}$,

$$\langle h,\, Q_{(r)} h \rangle_2 \ge 0 \quad \text{for } h \in L_2(\pi). \tag{2.10}$$

Since the Q_j are selfadjoint by (2.6), so is $Q_{(r)}$,

$$\langle h,\, Q_{(r)} h' \rangle_2 = \langle Q_{(r)} h,\, h' \rangle_2 \quad \text{for } h, h' \in L_2(\pi). \tag{2.11}$$

For a different argument see Liu et al. (1995, Lemma 3). Since powers of positive and selfadjoint operators are positive and selfadjoint, we obtain that $Q^s_{(r)}$ is positive and selfadjoint for all s.

3 Deterministic sweep

In this section we find, for the Gibbs sampler with deterministic sweep, the asymptotic variance of the empirical estimator, in two versions, and the minimal asymptotic variance of regular estimators of πf.

Let $X^0, \ldots, X^n$ be realizations from the Gibbs sampler for π with deterministic sweep, with n a multiple of k, say $n = pk$. We want to estimate the expectation πf of a function $f \in L_2(\pi)$. The most common estimator for πf is the empirical estimator based on the *subchain* $X^0, X^k, \ldots, X^{pk}$,

$$E^k_n f = \frac{1}{p} \sum_{q=1}^{p} f(X^{qk}).$$

The empirical estimator based on the *full* chain $X^0, \ldots, X^n$ is

$$E_n f = \frac{1}{n} \sum_{i=1}^{n} f(X^i) = \frac{1}{k} \sum_{j=1}^{k} E^j_n f$$

with

$$E_n^j f = \frac{1}{p}\sum_{q=1}^{p} f_{\le j}(X^{(q-1)k}, X^{qk})$$

and

$$f_{\le j}(x,y) = f(y_{\le j}, x_{>j}).$$

To fix things, by *asymptotic distribution* of an estimator T_n we will mean the asymptotic distribution of $n^{1/2}(T_n - \pi f)$, even though standardizing by $p^{1/2}$ rather than $n^{1/2}$ is more common for the empirical estimator $E_n^j f$.

Meyn and Tweedie (1993, p. 382) say that a Markov chain with transition distribution Q and invariant law π is *V-uniformly ergodic* if $V : E \to \mathbf{R}$ with $V \ge 1$ and

$$\sup_x V(x)^{-1} \sup_{|v|\le V} |Q^n(x,v) - \pi v| \to 0 \quad \text{for } n \to \infty.$$

The following propositions are consequences of their version of the central limit theorem for Markov chains (1993, p. 411, Theorem 17.0.1). Proofs are in Section 7.

Proposition 1. *Assume that the Gibbs sampler for π with deterministic sweep is positive Harris recurrent and the subchains are V-uniformly ergodic, and that $f^2 \le V$. Then the empirical estimator $E_n^j f$ is asymptotically normal with variance*

$$\sigma_j^2 = k\sigma_f^2 + 2k\sum_{s=1}^{\infty}\langle f - \pi f,\, p_j^{\mathrm{cycl}\, sk}(f - \pi f)\rangle_2,$$

where $p_j^{\mathrm{cycl}\, s} = p_j p_{j+1}\cdots p_k p_1 p_2 \cdots$ with s terms, and σ_f^2 is the variance of f under π.

Proposition 2. *Under the assumptions of Proposition 1, the empirical estimator $E_n f$ is asymptotically normal with variance*

$$\sigma_{(d)}^2 = \sigma_f^2 + 2\sum_{s=1}^{\infty}\frac{1}{k}\sum_{j=1}^{k}\langle f - \pi f,\, p_j^{\mathrm{cycl}\, s}(f - \pi f)\rangle_2. \tag{3.1}$$

Because the empirical estimator $E_n^k f$ based on the subchain is often used in practice, we have included the description of its asymptotic variance in Proposition 1. However, we do not recommend this estimator; the simulations in Section 6 show that $E_n^k f$ can be considerably worse than $E_n f$. This is true even when π has only two components; see Greenwood et al. (1996).

To determine the information bound of (regular) estimators of πf, we must prove that the model is locally asymptotically normal. A local model around π is obtained by perturbing π,

$$\pi_{nh}(dx) = \pi(dx)(1 + n^{-1/2}h(x)), \tag{3.2}$$

with *local parameter* h running through

$$H = \{h : E \to \mathbf{R} \text{ measurable, bounded, } \ \pi h = 0\}. \tag{3.3}$$

Write $p_{j,nh}(x_{-j}, dx_j)$ for the one-dimensional conditional distribution under $\pi_{nh}(dx)$ of x_j given x_{-j}. The effect of the perturbation of π on p_j can be described as follows.

Lemma 1. *For* $h \in H$,

$$p_{j,nh}(x_{-j}, dx_j) = p_j(x_{-j}, dx_j)(1 + n^{-1/2} h_j(x) + s_{j,nh}(x)) \tag{3.4}$$

with

$$h_j(x) = h(x) - Q_j(x, h) = h(x) - p_j(x_{-j}, h) \tag{3.5}$$

and $s_{j,nh}(x)$ *of order* $O(n^{-1})$ *uniformly in* x.

Write $Q_{(d)nh}$ for the transition distribution (2.8) of the Gibbs sampler for π_{nh} with deterministic sweep,

$$Q_{(d)nh}(x, dy) = (Q_{1,nh} \cdots Q_{k,nh})(x, dy) = \prod_{j=1}^{k} p_{j,nh}(y_{<j}, x_{>j}, dy_j).$$

It follows easily from Lemma 1 that $Q_{(d)nh}$ is obtained by perturbing $Q_{(d)}$ as follows: Uniformly in x and y,

$$Q_{(d)nh}(x, dy) = Q_{(d)}(x, dy)(1 + n^{-1/2}(K_{(d)}h)(x, y) + O(n^{-1})) \tag{3.6}$$

with

$$(K_{(d)}h)(x, y) = \sum_{j=1}^{k} h_j(y_{\le j}, x_{>j}). \tag{3.7}$$

In short: If π is perturbed by h, then $Q_{(d)}$ is perturbed by $K_{(d)}h$ up to $O(n^{-1/2})$. Of course, for $Q_{(d)nh}(x, dy)$ to be a transition distribution, we must have $Q_{(d)}(x, K_{(d)}h) = 0$. Indeed,

$$\int Q_{(d)}(x, dy) h_j(y_{\le j}, x_{>j}) = (Q_1 \cdots Q_j)(x, h_j) = 0$$

and hence $Q_{(d)}(x, K_{(d)}h) = 0$.

Our statement of local asymptotic normality will involve a new inner product. Note first that the closure of the local parameter space H in $L_2(\pi)$ is

$$L_{2,0}(\pi) = \{h \in L_2(\pi) : \pi h = 0\}.$$

The operator $K_{(d)}$ maps $L_{2,0}(\pi)$ into $L_2(\pi \otimes Q_{(d)})$. It induces a new inner product on $L_{2,0}(\pi)$,

$$\langle h,\, h' \rangle_{(d)} = \frac{1}{k} \pi \otimes Q_{(d)}(K_{(d)}h \cdot K_{(d)}h') \quad \text{for } h, h' \in L_{2,0}(\pi). \tag{3.8}$$

The corresponding norm is denoted $\|h\|_{(d)}$. The factor $1/k$ is included here to avoid its repeated appearance later.

Write $P_{(d)}$ for the joint distribution of $X^0, X^k, \dots, X^{pk}$ if π is true, and $P_{(d)nh}$ if π_{nh} is true. Relation (3.6) implies that $Q_{(d)nh}$ is Hellinger differentiable:

$$\int Q_{(d)}(x,dy)\left(\left(\frac{dQ_{(d)nh}}{dQ_{(d)}}(x,y)\right)^{1/2} - 1 - \frac{1}{2}n^{-1/2}(K_{(d)}h)(x,y)\right)^2 \le n^{-1}r_n(x)$$

for $h \in H$, where r_n decreases to 0 pointwise and is π-integrable for large n. This version of Hellinger differentiability is adapted from Höpfner, Jacod and Ladelli (1990). From Höpfner (1993) we obtain a nonparametric version of local asymptotic normality: If the Gibbs sampler for π with deterministic sweep is positive Harris recurrent, then we have for $h \in H$,

$$\log dP_{(d)nh}/dP_{(d)} = n^{-1/2}\sum_{q=1}^{p}(K_{(d)}h)(X^{(q-1)k}, X^{qk}) - \frac{1}{2}\|h\|_{(d)}^2 + o_{P_{(d)}}(1) \qquad (3.9)$$

with

$$n^{-1/2}\sum_{q=1}^{p}(K_{(d)}h)(X^{(q-1)k}, X^{qk}) \Rightarrow N_{\|h\|_{(d)}} \quad \text{under } P_{(d)},$$

where N_σ denotes a normal random variable with mean 0 and variance σ^2.

Call an estimator T_n *regular* for πf with *limit* L if

$$n^{1/2}(T_n - \pi_{nh}f) \Rightarrow L \quad \text{under } P_{(d)nh} \text{ for } h \in H.$$

The desired minimal asymptotic variance of regular estimators of πf is the squared length of the gradient of πf with respect to the new inner product (3.8). To define this gradient, we note that by definition (3.2) of π_{nh}, and since $\pi h = 0$,

$$n^{1/2}(\pi_{nh}f - \pi f) = \pi h f = \langle h,\, f - \pi f\rangle_2 \quad \text{for } h \in H. \qquad (3.10)$$

Hence $f - \pi f \in L_{2,0}(\pi)$ is the gradient of πf with respect to the usual inner product $\langle h,\, h'\rangle_2$ on $L_{2,0}(\pi)$. The gradient $g_{(d)} \in L_{2,0}(\pi)$ with respect to the new inner product (3.8) is now implicitly defined by

$$\langle h,\, f - \pi f\rangle_2 = \langle h,\, g_{(d)}\rangle_{(d)} \quad \text{for } h \in H. \qquad (3.11)$$

By a nonparametric version of Hájek's (1970) convolution theorem, the limit L of a regular estimator has the form $L = N_{\|g_{(d)}\|_{(d)}} + M$ with M independent of $N_{\|g_{(d)}\|_{(d)}}$. It follows that the *information bound* $B_{(d)}$ (the minimal asymptotic variance of regular estimators of πf) is given by $\|g_{(d)}\|_{(d)}^2$. A convenient reference for the appropriate convolution theorem is Greenwood and Wefelmeyer (1990).

We wish to find an explicit form of $g_{(d)}$ and $B_{(d)}$. Our main tool is the following lemma. It expresses the new inner product (3.8) in terms of the natural inner product on $L_{2,0}(\pi)$.

Lemma 2. *The inner product (3.8) can be written*

$$\langle h,\, h'\rangle_{(d)} = \langle h,\, (I - Q_{(r)})h'\rangle_2 \quad \text{for } h, h' \in L_{2,0}(\pi).$$

Surprisingly, this inner product for *deterministic* sweep involves the transition distribution for *random* sweep. To calculate the gradient, we need the inverse of $I - Q_{(r)}$.

Lemma 3. *If $\|Q_{(r)}^t\|_2 < 1$ for some t, then the operator $I - Q_{(r)}$ has a bounded inverse on $L_{2,0}(\pi)$,*

$$(I - Q_{(r)})^{-1} = \sum_{s=0}^{\infty} Q_{(r)}^s = I + \sum_{s=1}^{\infty} \frac{1}{(k-1)^s} \sum_{\substack{j_1,\dots,j_s=1 \\ j_r \neq j_{r+1}}}^{k} Q_{j_1} \cdots Q_{j_s}.$$

The main result of this section now follows easily.

Theorem 1. *If $\|Q_{(r)}^t\|_2 < 1$ for some t, then the functional πf has gradient*

$$\begin{aligned} g_{(d)} &= (I - Q_{(r)})^{-1}(f - \pi f) \\ &= f - \pi f + \sum_{s=1}^{\infty} \frac{1}{(k-1)^s} \sum_{\substack{j_1,\dots,j_s=1 \\ j_r \neq j_{r+1}}}^{k} p_{j_1} \cdots p_{j_s}(f - \pi f), \end{aligned}$$

and the information bound is

$$\begin{aligned} B_{(d)} &= \langle f - \pi f,\, (I - Q_{(r)})^{-1}(f - \pi f)\rangle_2 \\ &= \sigma_f^2 + \sum_{s=1}^{\infty} \frac{1}{(k-1)^s} \sum_{\substack{j_1,\dots,j_s=1 \\ j_r \neq j_{r+1}}}^{k} \langle f - \pi f,\, p_{j_1} \cdots p_{j_s}(f - \pi f)\rangle_2. \end{aligned} \qquad (3.12)$$

If the Gibbs sampler for π with deterministic sweep is positive Harris recurrent, then local asymptotic normality (3.9) holds, and (3.12) is the minimal asymptotic variance of regular estimators of πf. Note that $B_{(d)}$ does not depend on the order of the deterministic sweep; it only depends on π and f.

4 Random sweep

In this section we determine the asymptotic variance of the empirical estimator for the Gibbs sampler with random sweep, and the minimal asymptotic variance of regular estimators of πf under this sweep.

Let $X^0, \dots, X^n$ be realizations from the Gibbs sampler for π with random sweep. The usual estimator for the expectation πf of a function $f \in L_2(\pi)$ is the empirical estimator $E_n f = \frac{1}{n}\sum_{i=1}^n f(X^i)$.

Proposition 3. *Assume that the Gibbs sampler for π with random sweep is positive Harris recurrent and V-uniformly ergodic. Then for $f \in L_2(\pi)$ the empirical estimator $E_n f$ is asymptotically normal with variance*

$$\begin{aligned}
\sigma^2_{(r)} &= \sigma_f^2 + 2\sum_{s=1}^{\infty}\langle f-\pi f,\, Q^s_{(r)}(f-\pi f)\rangle_2 \\
&= \sigma_f^2 + 2\sum_{s=1}^{\infty}\frac{1}{(k-1)^s}\sum_{\substack{j_1,\dots,j_s=1\\ j_r\neq j_{r+1}}}^{k}\langle f-\pi f,\, p_{j_1}\cdots p_{j_s}(f-\pi f)\rangle_2.
\end{aligned}$$

Notice that the last term in the deterministic sweep information bound $B_{(d)}$, given in Theorem 1, appears in $\sigma^2_{(r)}$, so we find that $B_{(d)} = \frac{1}{2}(\sigma_f^2 + \sigma^2_{(r)})$.

As in the previous section, we want to determine the minimal asymptotic variance of regular estimators of πf, with $f \in L_2(\pi)$. To introduce a local model, we define sequences π_{nh} as in (3.2), where h is in the space H defined in (3.3). Then the corresponding $p_{j,nh}$ are obtained by perturbing p_j as in Lemma 1. Write $Q_{(r)nh}$ for the transition distribution of the Gibbs sampler with random sweep for π_{nh},

$$Q_{(r)nh}(x, dy) = \frac{1}{k}\sum_{j=1}^{k} Q_{j,nh}(x, dy) = \frac{1}{k}\sum_{j=1}^{k} p_{j,nh}(x_{-j}, dy_j)\varepsilon_{x_{-j}}(dy_{-j}).$$

The following lemma describes how $Q_{(r)nh}$ is obtained by perturbing $Q_{(r)}$. This is less straightforward than the corresponding perturbation (3.6) for *deterministic* sweep. The perturbation now involves the probabilities of not changing the value when updating a component. The reason is that the transition distribution $Q_j(x, dy)$ is supported by the line through x parallel to the j-th coordinate axis, $\{y : y_{-j} = x_{-j}\}$. Hence the support of $Q_{(r)}(x, dy)$ is contained in the union of the k lines. The supports of the $Q_j(x, dy)$ are disjoint except for the point x, which may be charged by some or all of them. Therefore, to calculate the $Q_{(r)}(x, dy)$-density of $Q_{(r)nh}(x, dy)$, we must treat x separately. We assume that the σ-field on each E_j contains the one-point sets, which will be the case in all applications.

Lemma 4. *For $h \in H$, and uniformly in x and y,*

$$Q_{(r)nh}(x, dy) = Q_{(r)}(x, dy)(1 + n^{-1/2}(K_{(r)}h)(x, y) + O(n^{-1})) \tag{4.1}$$

with

$$(K_{(r)}h)(x,y) = \sum_{j=1}^{k}(K_jh)(x,y), \tag{4.2}$$

$$\begin{aligned}(K_jh)(x,y) &= \Big(1(y_{-j}=x_{-j}, y_j\neq x_j) + \frac{r_j(x)}{r(x)}1(y=x)\Big)h_j(x_{-j},y_j) \\ &= \Big(1(y_{-j}=x_{-j}) - \Big(1-\frac{r_j(x)}{r(x)}\Big)1(y=x)\Big)h_j(x_{-j},y_j),\end{aligned} \tag{4.3}$$

$$r_j(x) = p_j(x_{-j},\{x_j\}), \quad r(x) = \sum_{j=1}^{k} r_j(x), \tag{4.4}$$

where $h_j(x) = h(x) - p_j(x_{-j},h)$ *as in (3.5), and* $0/0=0$.

In short: If π is perturbed by h, then $Q_{(r)}$ is perturbed by $K_{(r)}h$ up to $O(n^{-1/2})$. For $Q_{(r)nh}(x,dy)$ to be a transition distribution, we must have $Q_{(r)}(x, K_{(r)}h) = 0$. Indeed,

$$Q_j(x, K_jh) = -(1-r_j(x)/r(x))r_j(x)h_j(x),$$

and for $i\neq j$,

$$Q_i(x, K_jh) = r_i(x) - (1-r_j(x)/r(x))r_i(x)h_j(x),$$

so that

$$Q_{(r)}(x, K_jh) = \frac{1}{k}\sum_{i=1}^{k} Q_i(x, K_jh) = 0$$

and hence $Q_{(r)}(x, K_{(r)}h) = 0$.

The operator $K_{(r)}$ maps $L_{2,0}(\pi)$ into $L_2(\pi\otimes Q_{(r)})$. It induces a new inner product on $L_{2,0}(\pi)$,

$$\langle h,\, h'\rangle_{(r)} = \pi\otimes Q_{(r)}(K_{(r)}h\cdot K_{(r)}h') \quad \text{for } h,h'\in L_{2,0}(\pi). \tag{4.5}$$

The corresponding norm is denoted $\|h\|_{(r)}$.

Write $P_{(r)}$ for the joint distribution of $X^0,\dots,X^n$ if π is true, and $P_{(r)nh}$ if π_{nh} is true. Relation (4.1) implies that $Q_{(r)nh}$ is Hellinger differentiable. As in Section 3 we obtain a nonparametric version of local asymptotic normality: If the Gibbs sampler for π with random sweep is positive Harris recurrent, then we have for $h\in H$,

$$\log dP_{(r)nh}/dP_{(r)} = n^{-1/2}\sum_{i=1}^{n}(K_{(r)}h)(X^{i-1},X^i) - \frac{1}{2}\|h\|_{(r)}^2 + o_{P_{(r)}}(1) \tag{4.6}$$

with

$$n^{-1/2}\sum_{i=1}^{n}(K_{(r)}h)(X^{i-1},X^i) \Rightarrow N_{\|h\|_{(r)}} \quad \text{under } P_{(r)}.$$

Exactly as in Section 3, the desired minimal asymptotic variance $B_{(r)}$ of regular estimators of πf is the squared length $\|g_{(r)}\|_{(r)}^2$ of the gradient $g_{(r)} \in L_{2,0}(\pi)$ of πf with respect to the new inner product (4.5). This gradient is now (compare (3.11)) defined by

$$\langle h,\, f - \pi f\rangle_2 = \langle h,\, g_{(r)}\rangle_{(r)} \quad \text{for } h \in H. \tag{4.7}$$

To calculate $g_{(r)}$ and $\|g_{(r)}\|_{(r)}$, we use the following lemma (compare Lemma 2), which expresses the new inner product (4.5) in terms of the natural inner product on $L_{2,0}(\pi)$.

Lemma 5. *The inner product (4.5) can be written*

$$\langle h,\, h'\rangle_{(r)} = \langle h,\, (I - Q_{(r)} + S)h'\rangle_2 \quad \text{for } h, h' \in L_{2,0}(\pi)$$

with

$$Sh = \frac{1}{k}\sum_{i,j=1}^{k} Q_i(R_{ij}Q_j h), \tag{4.8}$$

$$R_{ij}(x) = \delta_{ij} r_j(x) - r_i(x) r_j(x)/r(x). \tag{4.9}$$

The matrix $R(x) = (R_{ij}(x))_{i,j=1,\dots,k}$ is symmetric, and both $R(x)$ and $I - R(x)$ are positive semidefinite, where $I = (\delta_{ij})_{i,j=1,\dots,k}$ is the identity matrix. The operator S is selfadjoint on $L_2(\pi)$, and both S and $Q_{(r)} - S$ are positive. The operators $I - Q_{(r)}$ and $I - Q_{(r)} + S$ are positive.

The properties of R and S will be used in Section 5. The main result of this section follows easily from Lemma 5.

Theorem 2. *If $\|(Q_{(r)} - S)^t\|_2 < 1$ for some t, then the functional πf has gradient*

$$\begin{aligned} g_{(r)} &= (I - Q_{(r)} + S)^{-1}(f - \pi f) \\ &= f - \pi f + \sum_{s=1}^{\infty}(Q_{(r)} - S)^s (f - \pi f), \end{aligned}$$

and the information bound is

$$\begin{aligned} B_{(r)} &= \langle f - \pi f,\, (I - Q_{(r)} + S)^{-1}(f - \pi f)\rangle_2 \\ &= \sigma_f^2 + \sum_{s=1}^{\infty}\langle f - \pi f,\, (Q_{(r)} - S)^s (f - \pi f)\rangle_2. \end{aligned} \tag{4.10}$$

If the random sweep Gibbs sampler for π is positive Harris recurrent, then local asymptotic normality (4.6) holds, and (4.10) is the minimal asymptotic variance of regular estimators of πf.

5 Discussion

In this section we explain how the main points in the Introduction are a consequence of the results in Sections 3 and 4. We first compare the information bounds under deterministic and random sweeps. Then we compare the efficiencies of the empirical estimator under the two sweeps.

Information bounds for deterministic and random sweep. We first show that the information bound is no larger for random sweep than for deterministic sweep.

To compare the bounds given by Theorems 1 and 2, consider transition kernels $K(x, dy)$ as operators on $L_2(\pi)$ and write $K \geq L$ if $K - L$ is positive, i.e., $\langle h, (K-L)h\rangle_2 \geq 0$ for $h \in L_2(\pi)$. The operator S defined in (4.8) is positive by Lemma 5. Hence $I - Q_{(r)} + S \geq I - Q_{(r)}$ and therefore $(I - Q_{(r)} + S)^{-1} \leq (I - Q_{(r)})^{-1}$. Thus the variance bound is no larger for random sweep than for deterministic sweep: $B_{(r)} \leq B_{(d)}$.

Suppose that π is continuous in the sense that it is absolutely continuous with respect to the product of its marginals, and the marginals have no atoms. Then $p_j(x_{-j}, dx_j)$ has no atoms for $m_{-j}(dx_{-j})$-a.s. x_{-j}. Hence $r_j(x) = p_j(x_{-j}, \{x_j\}) = 0$ for π-a.s. x, and therefore $R(x) = 0$ for π-a.s. x, and the operator S reduces to 0. This implies that information bound for random sweep coincides with the information bound for deterministic sweep: $B_{(r)} = B_{(d)}$.

Two-step samplers. Of particular interest is the case of a two-dimensional π. In this case, if we are interested in only one of the two marginal distributions of π, the corresponding Gibbs sampler is called an auxiliary variable method (Swendsen and Wang, 1987), data augmentation algorithm (Tanner and Wong, 1987) or successive substitution sampler (Gelfand and Smith, 1990). See also Higdon (1998).

For this case, $k = 2$, the variances and information bounds simplify considerably. The asymptotic variance of the empirical estimator $E_n f$ for deterministic sweep is, by Proposition 2,

$$\sigma_{(d)}^2 = \sigma_f^2 + \sum_{s=1}^{\infty} \langle f - \pi f,\, (p_1^{\mathrm{cycl}\, s} + p_2^{\mathrm{cycl}\, s})(f - \pi f)\rangle_2,$$

and for random sweep,

$$\sigma_{(r)}^2 = \sigma_f^2 + 2\sum_{s=1}^{\infty} \langle f - \pi f,\, (p_1^{\mathrm{cycl}\, s} + p_2^{\mathrm{cycl}\, s})(f - \pi f)\rangle_2$$

by Proposition 3, where now $p_1^{\mathrm{cycl}\, s} = p_1 p_2 p_1 \cdots$ and $p_2^{\mathrm{cycl}\, s} = p_2 p_1 p_2 \cdots$ with s terms. Hence, if the leading term σ_f^2 of the variances is relatively small (as in most applications), then the variance $\sigma_{(r)}^2$ for random sweep is nearly twice as large as the variance $\sigma_{(d)}^2$ for deterministic sweep.

For $k = 2$, the deterministic sweep information bound (3.12) equals $\sigma_{(d)}^2$. Hence, for the two-step Gibbs sampler with deterministic sweep, the empirical estimator $E_n f$

is *efficient*. This result was mentioned in Greenwood et al. (1996, Section 1). If π is continuous, then the information bound for random sweep equals the information bound for deterministic sweep. Since $\sigma^2_{(r)} > \sigma^2_{(d)}$, as seen above, it follows that the empirical estimator for random sweep is *not* efficient. Indeed, $\sigma^2_{(r)} \approx 2\sigma^2_{(d)} = 2B_{(d)}$, so the efficiency of the empirical estimator for random sweep is close to 50%.

Variance of the empirical estimator. From Proposition 3, the second summation in $\sigma^2_{(r)}$ contains $k(k-1)^{s-1}$ terms, each being an s-order autocovariance of the form $\langle f - \pi f,\, p_{j_1} \cdots p_{j_s}(f - \pi f)\rangle_2 \equiv \sigma^2_{j_1,\dots,j_s}$. From Proposition 2, the s-order term in $\sigma^2_{(d)}$ is an average of k of these s-order autocovariances, those of the form $\langle f - \pi f,\, p_j^{\text{cycl}\, s}(f - \pi f)\rangle_2 \equiv \sigma^2_{j,\text{cycl}\, s}$. Thus, unless the averages of the s-order autocovariances in $\sigma^2_{(r)}$ and $\sigma^2_{(d)}$ are systematically different from one another, we expect $\sigma^2_{(r)} \approx [k/(k-1)]\sigma^2_{(d)}$, or that $\sigma^2_{(r)}$ is slightly larger than $\sigma^2_{(d)}$. Such a result holds if one considers a random sweep without repetition, see Fishman (1996, Theorem 8). The simulations reported in the next section show, however, that $\sigma^2_{(r)}$ can be up to twice as large as $\sigma^2_{(d)}$, even if k is large. The reason is that the higher-order terms in $\sigma^2_{(r)}$ can decay more slowly than those in $\sigma^2_{(d)}$. This is seen clearly in the following special case.

Independent components. Relatively simple formulae for $\sigma^2_{(r)}$ and $\sigma^2_{(d)}$ can be obtained when π has independent components, which we now assume. Now $\sigma^2_{j,\text{cycl}\, s}$ vanishes for $s \geq k$, because integration of $f(x)$ cyclically over k components gives πf. However, $\sigma^2_{j_1,\dots,j_s}$ vanishes only if all k components are present among $j_1 \dots, j_s$. Also, if some of the j_r are equal, fewer than s components are integrated out, so $\sigma^2_{j_1,\dots,j_s}$ is larger than any $\sigma^2_{j,\text{cycl}\, s}$ 'covering' $j_1 \dots, j_s$.

To simplify the expression for $\sigma^2_{(r)}$, note that the Q_j are now not only idempotent but also commute, so we have

$$Q^t_{(r)} = \sum_{s=1}^{k} \frac{a_{ts}}{k^t} \sum_{\{j_1,\dots,j_s\} \subset \{1,\dots,k\}} Q_{j_1} \cdots Q_{j_s},$$

where a_{ts} is the number of ways of placing t different objects into s different cells, with no cell empty; $a_{ts}/s!$ is a Stirling number of the second kind. We may interpret a_{ts}/k^t as the probability that, over t steps of the random sweep sampler, each member of a given set of s components is updated at least once and none of the remaining components are updated. Now we find that

$$\sigma^2_{(r)} = \sigma^2_f + 2 \sum_{s=1}^{k-1} b_s \binom{k}{s} \overline{\sigma}^2_s$$

where

$$b_s = \sum_{t=s}^{\infty} \frac{a_{ts}}{k^t} = \binom{k-1}{s}^{-1}$$

and $\overline{\sigma}_s^2$ is the average of the s-order autocovariances over distinct components. The last identity above is proved by relating b_s to certain expected values arising in a classical occupancy problem, see Lemma 6 at the end of Section 7. We may interpret b_s as the expected number of times, during an infinite run of the random sweep sampler, at which a given set of s components have been updated at least once and the remaining components have not yet been updated.

The average $\overline{\sigma}_s^2$ of the s-order autocovariances takes the value σ_f^2 at $s = 0$, and it decreases to zero at $s = k$. Linear interpolation between these two values gives the approximation

$$\sigma_{(r)}^2 \approx \sigma_f^2 + 2\sum_{s=1}^{k-1} b_s \binom{k}{s}\left(1 - \frac{s}{k}\right)\sigma_f^2 = (2k-1)\sigma_f^2.$$

A similar approximation involving the autocovariances appearing in the first k terms of (3.1) gives

$$\sigma_{(d)}^2 \approx \sigma_f^2 + 2\sum_{s=1}^{k-1}\left(1 - \frac{s}{k}\right)\sigma_f^2 = k\sigma_f^2,$$

so $\sigma_{(r)}^2$ is close to twice as large as $\sigma_{(d)}^2$. Also note that (in general) $\sigma_{(r)}^2$ can never be *more* than twice as large as $\sigma_{(d)}^2$, because $\sigma_{(r)}^2 = 2B_{(d)} - \sigma_f^2 \leq 2\sigma_{(d)}^2 - \sigma_f^2$.

If f depends on only *one* of the k components (and π has independent components), the above approximations for $\sigma_{(d)}^2$ and $\sigma_{(r)}^2$ are exact and the deterministic sweep information bound $B_{(d)}$ can be computed explicitly: $B_{(d)} = \frac{1}{2}(\sigma_f^2 + \sigma_{(r)}^2) = k\sigma_f^2$. In this case the empirical estimator is efficient under deterministic sweep, but has an efficiency of no more than $1/(2 - k^{-1})$ under random sweep.

Efficiency of the empirical estimator. Under deterministic sweep, the empirical estimator is close to efficient when the components of π are not strongly dependent. Indeed, π can then be considered as a perturbation of a distribution having independent components, so the above approximations and a continuity argument give $\sigma_{(d)}^2 \approx \sigma_{(r)}^2/2 = B_{(d)} - \sigma_f^2/2$, showing that the empirical estimator comes close to attaining the information bound.

Under random sweep, the efficiency of the empirical estimator is at best only slightly more than 50%, because $\sigma_{(r)}^2 = 2B_{(d)} - \sigma_f^2 \geq 2B_{(r)} - \sigma_f^2$. However, its efficiency is close to 50% when π is continuous, as we have $\sigma_{(r)}^2 = 2B_{(r)} - \sigma_f^2$ in that case.

More work is needed to determine the efficiency of the empirical estimator under deterministic sweep when π has strongly dependent components. Then we are not able to exclude the possibility that there is a poor choice of deterministic sweep for which the empirical estimator is far from efficient, or with an asymptotic variance *larger* than under random sweep. Random sweep is a reasonable but conservative alternative: we are guaranteed a relative efficiency no worse than 50% of the best deterministic sweep, and its overall efficiency is roughly 50% if π is continuous. Based on the simulations in Section 5, we suspect that the empirical estimator is always close to being efficient under deterministic sweep, but we have only proved this in the case of two component Gibbs samplers (where we have full efficiency).

Variance reduction under deterministic sweep. We have seen that the empirical estimator has a much smaller asymptotic variance for deterministic than for random sweep when the components of π are not strongly dependent. This is due to a particular feature of the Gibbs sampler: the transition distribution Q_j used to update component j does not use the present value of that component. Hence Q_j is idempotent (2.3), and an s-order autocovariance $\sigma^2_{j_1,\dots,j_s}$ reduces to a lower-order autocovariance whenever two adjacent indices j_r and j_{r+1} are equal. Other samplers use x_j to update component j. Their transition distributions are of the form

$$Q_j(x, dy) = q_j(x, dy_j)\varepsilon_{x_{-j}}(dy_{-j}), \tag{5.1}$$

with $q_j(x, dy_j)$ depending on x_j. The proof of Proposition 3 shows that the asymptotic variance of $E_n f$ for the corresponding sampler with random sweep is

$$\sigma_f^2 + 2\sum_{s=1}^{\infty}\frac{1}{k^s}\sum_{j_1,\dots,j_s=1}^{k}\langle f - \pi f,\, q_{j_1}\cdots q_{j_s}(f-\pi f)\rangle_2.$$

This is about the size of the asymptotic variance for deterministic sweep. The $s = 1$ term is exactly the same as the corresponding term for deterministic sweep.

Among transition distributions Q_j of the form (5.1), with Q_j in detailed balance with π, the transition distribution of the Gibbs sampler is the only one that does not use the present x_j. To see this, consider $q_j(x, dy_j) = q_j(x_{-j}, dy_j)$ not depending on x_j. Write

$$\pi(dx)Q_j(x, dy) = m_{-j}(dx_{-j})p_j(x_{-j}, dx_j)q_j(x_{-j}, dy_j)\varepsilon_{x_{-j}}(dy_{-j})$$

and

$$\begin{aligned}\pi(dy)Q_j(y, dx) &= m_{-j}(dy_{-j})p_j(y_{-j}, dy_j)q_j(y_{-j}, dx_j)\varepsilon_{y_{-j}}(dx_{-j})\\ &= m_{-j}(dx_{-j})p_j(x_{-j}, dy_j)q_j(x_{-j}, dx_j)\varepsilon_{x_{-j}}(dy_{-j}).\end{aligned}$$

Detailed balance (2.7) implies

$$p_j(x_{-j}, dx_j)q_j(x_{-j}, dy_j) = p_j(x_{-j}, dy_j)q_j(x_{-j}, dx_j) \quad m_{-j}(dx_{-j})\text{-a.s.}$$

Now integrate on y_j to obtain

$$p_j(x_{-j}, dx_j) = q_j(x_{-j}, dx_j) \quad m_{-j}(dx_{-j})\text{-a.s.}$$

6 Simulation examples

In this section we show how simulation can be used to estimate information bounds for Gibbs samplers, and we demonstrate the method in two simple examples. The examples are designed to compare the performance of the various empirical estimators with the corresponding information bounds and to study how they are affected by parameters in the model.

The second term in $B_{(d)}$, as given by (3.12), is difficult to evaluate directly, but since this term appears doubled in $\sigma^2_{(r)}$, we have $B_{(d)} = \frac{1}{2}(\sigma^2_f + \sigma^2_{(r)})$. Thus, $B_{(d)}$ can be estimated by running the corresponding random sweep sampler and estimating the asymptotic variance of the empirical estimator. If π does not have atoms (as in the examples), then this method also gives the random sweep information bound, since $B_{(r)} = B_{(d)}$ in that case.

The function f is taken to be the indicator that the random field exceeds a unit threshold: $f(x) = 1\{\max_j x_j > 1\}$. The asymptotic variances and information bounds are found over a fine grid of values of a parameter that controls the correlation structure of π; they are divided by $k\sigma^2_f$ and then smoothed for display in the plots.

Exchangeable normal variables. Let π be an exchangeable k-dimensional multivariate normal distribution in which each component has zero mean and unit variance, and all the pairwise correlations are identical. This example has been widely used in the literature for studying convergence rates of Gibbs samplers, see, e.g., Raftery and Lewis (1992, Example 3) and Roberts and Sahu (1997). The results for 10 and 20 dimensions are shown in Figure 1.

The asymptotic variance of $E_n f$ is close to the deterministic sweep information bound $B_{(d)}$, indicating that $E_n f$ is close to being efficient. Under random sweep, the asymptotic variance of $E_n f$ almost doubles (as the discussion in Section 5 led us to expect), despite the fact that $B_{(r)} = B_{(d)}$. On the other hand, random sweep can be shown to have a better *convergence rate* than deterministic sweep when the pairwise correlation is negative, see Roberts and Sahu (1997, Section 3.2).

For moderate correlations, $B_{(d)}$ is considerably less than the asymptotic variance of the 'common' empirical estimator $E^k_n f$. For high positive correlation, the curves for deterministic sweep in each plot increase exponentially and come closer together, leaving less room for potential improvements over $E_n f$. Increasing the dimension k has the effect of increasing the information bound under positive correlation, but decreasing it under negative correlation.

Gaussian conditional autoregressions. View each component x_j of x as a being located at a site j on square lattice. Suppose that the local characteristics of π are Gaussian with conditional mean and variance at each site j given by $\lambda \bar{x}_j$ and κ/n_j, respectively, where $\lambda \in (0, 1)$, $\bar{x}_j$ is the mean of the components of x that are neighbors of site j, $\kappa > 0$ and n_j is the number of neighbors of site j. This is the simplest example of a Gaussian conditional autoregression with some practical value in spatial statistics and image processing, see Besag and Kooperberg (1995) for discussion and references. We use the (standard) deterministic sweep strategy known as checkerboard updating, which updates each site of even parity and then each site of odd parity.

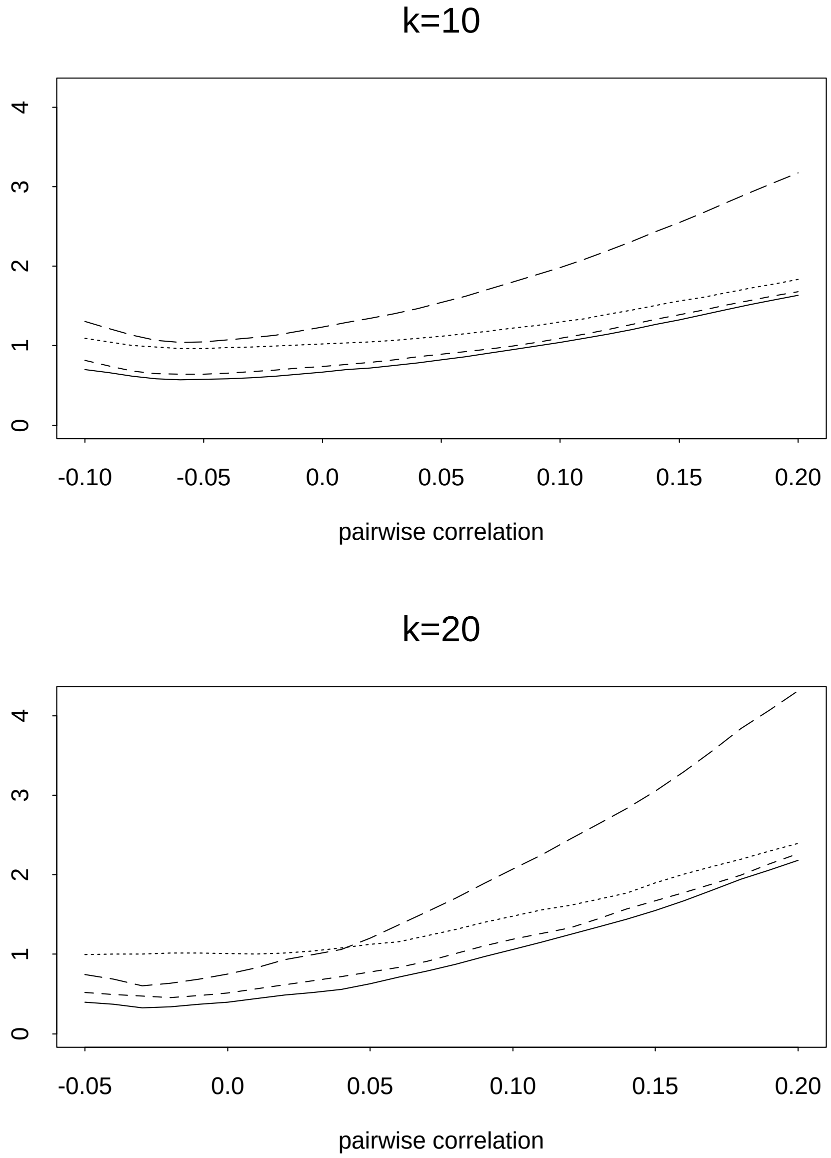

Figure 1: Exchangeable k-dimensional normal example. The information bounds for both random and deterministic sweep (solid line) and the asymptotic variances (in units of $k\sigma_f^2$) of the 'common' empirical estimator $E_n^k f$ under deterministic sweep (dotted line), and the 'full chain' empirical estimator $E_n f$ under deterministic sweep (short dashed line) and random sweep (long dashed line).

Figure 2 displays the simulation results for a 6×6 lattice with free boundary conditions and $\kappa = 1$. This plot has similar features to Figure 2, but the increase in the asymptotic variances with increasing λ is less pronounced than what it was with the pairwise correlation. The empirical estimator is close to efficient under checkerboard updating, but has an efficiency of only about 50% under random sweep. Our results

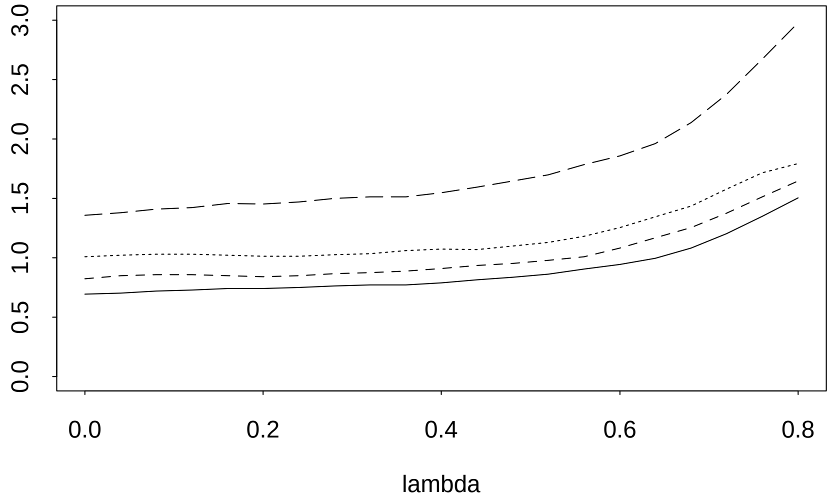

Figure 2: Gaussian conditional autoregression example. See Figure 1 for key.

are consistent with results on convergence rates; Roberts and Sahu (1997, Example 1) show that the Markov chain in this example has a faster rate of convergence under the checkerboard sweep than under random sweep.

7 Proofs

Proof of Proposition 1. By a central limit theorem (Meyn and Tweedie, 1993, p. 411, Theorem 17.01) applied to the subchain $(X^{qk}_{\le j}, X^{(q-1)k}_{>j})$, $q = 1, 2, \dots$, the empirical estimator $E^j_n f$ is asymptotically normal with variance

$$\begin{aligned}\sigma_j^2 &= k\,\pi \otimes Q_{(d)}(f_{\le j} - \pi \otimes Q_{(d)} f_{\le j})^2 \\ &\quad +2k \sum_{s=1}^{\infty} \pi \otimes Q_{(d)}((f_{\le j} - \pi \otimes Q_{(d)} f_{\le j}) \cdot Q^s_{(d)}(f_{\le j} - \pi \otimes Q_{(d)} f_{\le j})).\end{aligned}$$

The function $f_{\le j}(x, y)$ depends only on $(y_{\le j}, x_{>j})$, and the chain $(X^{qk}_{\le j}, X^{(q-1)k}_{>j})$, $q = 1, 2, \dots$, has transition distribution $Q_{j+1} \cdots Q_k Q_1 \cdots Q_j$ and invariant law π. Hence we can write

$$\sigma_j^2 = k\sigma_f^2 + 2k \sum_{s=1}^{\infty} \langle f - \pi f,\, p_j^{\mathrm{cycl}\, sk}(f - \pi f)\rangle_2.$$

Proof of Proposition 2. Write

$$E_n f = \frac{1}{n} \sum_{q=1}^{p} F(X^{(q-1)k}, X^{qk})$$

with

$$F(x, y) = \sum_{j=1}^{k} f_{\le j}(x, y).$$

By the central limit theorem used in the proof of Proposition 1, the empirical estimator $E_n f$ is asymptotically normal with variance

$$\begin{aligned}
\sigma^2_{(d)} &= \frac{1}{k}\pi\otimes Q_{(d)}(F-\pi\otimes Q_{(d)}F)^2 \\
&\quad + \frac{2}{k}\sum_{t=1}^{\infty}\pi\otimes Q_{(d)}((F-\pi\otimes Q_{(d)}F)\cdot Q^t_{(d)}(F-\pi\otimes Q_{(d)}F)).
\end{aligned}$$

We calculate each term on the right side separately. As in the proof of Proposition 1, we make use of the fact that $f_{\le j}(x,y)$ depends only on $(y_{\le j}, x_{>j})$, and that the chain $(X^{qk}_{\le j}, X^{(q-1)k}_{>j})$, $q=1,2,\dots$, has transition distribution $Q_{j+1}\cdots Q_k Q_1\cdots Q_j$ and invariant law π:

$$\begin{aligned}
&\pi\otimes Q_{(d)}(F-\pi\otimes Q_{(d)}F)^2 \\
&= \sum_{i=1}^{k}\sum_{j=1}^{k}\pi\otimes Q_{(d)}((f_{\le i}-\pi f)(f_{\le j}-\pi f)) \\
&= k\sigma_f^2 + 2\sum_{j=1}^{k}\sum_{i=1}^{k-j}\langle f-\pi f,\, p_j^{\mathrm{cycl}\, i}(f-\pi f)\rangle_2
\end{aligned}$$

and, for $t=1,2,\dots$,

$$\begin{aligned}
&\pi\otimes Q_{(d)}((F-\pi\otimes Q_{(d)}F)\cdot Q^t_{(d)}(F-\pi\otimes Q_{(d)}F)) \\
&= \sum_{i=1}^{k}\sum_{j=1}^{k}\pi\otimes Q_{(d)}((f_{\le i}-\pi f)\cdot Q^t_{(d)}((f_{\le j}-\pi f))) \\
&= \sum_{j=1}^{k}\sum_{i=k-j+1}^{k}\langle f-\pi f,\, p_j^{\mathrm{cycl}\ (t-1)k+i}(f-\pi f)\rangle_2 \\
&\quad + \sum_{j=1}^{k}\sum_{i=1}^{k-j}\langle f-\pi f,\, p_j^{\mathrm{cycl}\ tk+i}(f-\pi f)\rangle_2.
\end{aligned}$$

Collecting autocovariances of equal order,

$$\sigma^2_{(d)} = \sigma_f^2 + 2\sum_{s=1}^{\infty}\frac{1}{k}\sum_{j=1}^{k}\langle f-\pi f,\, p_j^{\mathrm{cycl}\, s}(f-\pi f)\rangle_2.$$

Proof of Lemma 1. Factor π and π_{nh} as in (2.1) to obtain

$$\frac{\pi_{nh}(dx)}{\pi(dx)} = \frac{m_{-j,nh}(dx_{-j})p_{j,nh}(x_{-j},dx_j)}{m_{-j}(dx_{-j})p_j(x_{-j},dx_j)}. \tag{7.1}$$

By definition (3.2) of π_{nh}, the left side is $1+n^{-1/2}h(x)$, and

$$\begin{aligned}
m_{-j,nh}(dx_{-j}) &= m_{-j}(dx_{-j})\int p_j(x_{-j},dx_j)(1+n^{-1/2}h(x)) \\
&= m_{-j}(dx_{-j})(1+n^{-1/2}p_j(x_{-j},h)).
\end{aligned}$$

Now solve (7.1) for $p_{j,nh}$.

Proof of Lemma 2. By definition (3.7) of $K_{(d)}$ and definition (2.8) of $Q_{(d)}$, the inner product (3.8) is

$$\begin{aligned}
\langle h,\, h'\rangle_{(d)} &= \frac{1}{k}\pi \otimes Q_{(d)}(K_{(d)}h \cdot K_{(d)}h') \\
&= \frac{1}{k}\sum_{i,j=1}^{k} \iint \pi(dx)(Q_1\cdots Q_k)(x, dy) h_i(y_{\le i}, x_{>i}) h'_j(y_{\le j}, x_{>j}).
\end{aligned}$$

The functions $h_i(y_{\le i}, x_{>i})$ and $h'_j(y_{\le j}, x_{>j})$ are orthogonal with respect to $\langle\cdot,\cdot\rangle_{(d)}$ for $i \neq j$ since for $i < j$ the function $h_i(y_{\le i}, x_{>i})$ does not depend on y_j, and $Q_j h_j = 0$ by definition (3.5) of h_j. Again using $Q_j h_j = 0$, we obtain

$$Q_j h_j h'_j = Q_j h h'_j.$$

Since π is invariant under all Q_i, the distribution of $(y_{\le i}, x_{>i})$ under $\pi \otimes Q_{(d)}$ is π. We arrive at

$$\langle h,\, h'\rangle_{(d)} = \frac{1}{k}\langle h,\, \sum_{j=1}^{k} h'_j\rangle_2 = \langle h,\, (I - Q_{(r)})h'\rangle_2.$$

Proof of Lemma 3. A von Neumann expansion gives, when $\|Q^t_{(r)}\|_2 < 1$ for some t,

$$(I - Q_{(r)})^{-1} = I + \sum_{t=1}^{\infty} Q^t_{(r)}.$$

To calculate $Q^t_{(r)}$, we recall that the Q_j are idempotent, (2.3). Hence a product $Q_{j_1}\cdots Q_{j_t}$ reduces to a product with $s \le t$ factors whenever it consists of s blocks with identical components in each block, but different components in consecutive blocks. We obtain

$$Q^t_{(r)} = \frac{1}{k^t}\sum_{j_1,\ldots,j_t=1}^{k} Q_{j_1}\cdots Q_{j_t} = \frac{1}{k^t}\sum_{s=1}^{t}\binom{t-1}{s-1}\sum_{\substack{j_1,\ldots,j_s=1\\ j_r\neq j_{r+1}}}^{k} Q_{j_1}\cdots Q_{j_s}.$$

Noting that

$$\sum_{t=s}^{\infty}\binom{t-1}{s-1} z^t = \left(\frac{z}{1-z}\right)^s \quad \text{for } 0 \le z < 1, \quad s = 1, 2, \ldots$$

and interchanging sums, we find

$$\begin{aligned}
\sum_{t=1}^{\infty} Q^t_{(r)} &= \sum_{t=1}^{\infty}\frac{1}{k^t}\sum_{s=1}^{t}\binom{t-1}{s-1}\sum_{\substack{j_1,\ldots,j_s=1\\ j_r\neq j_{r+1}}}^{k} Q_{j_1}\cdots Q_{j_s} \qquad (7.2)\\
&= \sum_{s=1}^{\infty}\frac{1}{(k-1)^s}\sum_{\substack{j_1,\ldots,j_s=1\\ j_r\neq j_{r+1}}}^{k} Q_{j_1}\cdots Q_{j_s}.
\end{aligned}$$

Proof of Theorem 1. By definition (3.11) and Lemma 2, the gradient $g_{(d)} \in L_{2,0}(\pi)$ of πf is defined by

$$\langle h,\, f - \pi f\rangle_2 = \langle h,\, g_{(d)}\rangle_{(d)} = \langle h,\, (I - Q_{(r)})g_{(d)}\rangle_2 \quad \text{for } h \in H.$$

Lemma 3 gives us the bounded inverse of the operator $I - Q_{(r)}$. We obtain

$$\begin{aligned} g_{(d)} &= (I - Q_{(r)})^{-1}(f - \pi f) \\ &= f - \pi f + \sum_{s=1}^{\infty} \frac{1}{(k-1)^s} \sum_{\substack{j_1,\ldots,j_s=1 \\ j_r \neq j_{r+1}}}^{k} p_{j_1} \cdots p_{j_s}(f - \pi f). \end{aligned}$$

Again by Lemma 2,

$$\begin{aligned} \|g_{(d)}\|_{(d)}^2 &= \langle g_{(d)},\, (I - Q_{(r)})g_{(d)}\rangle_2 \\ &= \langle f - \pi f,\, (I - Q_{(r)})^{-1}(f - \pi f)\rangle_2 \\ &= \sigma_f^2 + \sum_{s=1}^{\infty} \frac{1}{(k-1)^s} \sum_{\substack{j_1,\ldots,j_s=1 \\ j_r \neq j_{r+1}}}^{k} \langle f - \pi f,\, p_{j_1} \cdots p_{j_s}(f - \pi f)\rangle_2. \end{aligned}$$

Proof of Proposition 3. By Meyn and Tweedie (1993, Theorem 16.0.1), V-uniform ergodicity implies V-uniform ergodicity with an exponential rate: there exist $\rho < 1$ and $C > 0$ such that

$$\sup_x V(x)^{-1} \sup_{|v| \le V} |Q^n(x, v) - \pi v| \le C\rho^{-n}.$$

The Gibbs sampler with random sweep is reversible. Roberts and Rosenthal (1997, Corollary 2.1) prove that these assumptions suffice for the central limit theorem of Kipnis and Varadhan. It follows that the empirical estimator E_n^f is asymptotically normal with variance

$$\sigma_{(r)} = \sigma_f^2 + 2\sum_{t=1}^{\infty} \langle f - \pi f,\, Q_{(r)}^t(f - \pi f)\rangle_2.$$

The power series $\sum_{t=1}^{\infty} Q_{(r)}^t$ was already calculated in (7.2).

Proof of Lemma 4. Recall that $Q_j(x, dy) = p_j(x_{-j}, dy_j)\varepsilon_{x_{-j}}(dy_{-j})$ lives on the line $\{y : y_{-j} = x_{-j}\}$ through x parallel to the j-th coordinate axis. Hence

$$\frac{Q_j(x, dy)}{Q_{(r)}(x, dy)} = k \quad \text{on } \{y : y \neq x\}.$$

By definition (4.4) of r_j and r,

$$\frac{Q_j(x, \{x\})}{Q_{(r)}(x, \{x\})} = k\frac{r_j(x)}{r(x)}.$$

Hence a version of the $Q_{(r)}(x,dy)$-density of $Q_j(x,dy)$ is

$$
\begin{aligned}
&k\Big(1(y_{-j}=x_{-j}, y_j\neq x_j)+\tfrac{r_j(x)}{r(x)}1(y=x)\Big)\\
&=k\Big(1(y_{-j}=x_{-j})-(1-\tfrac{r_j(x)}{r(x)})1(y=x)\Big).
\end{aligned}
$$

By Lemma 1,

$$
Q_{j,nh}(x,dy)=Q_j(x,dy)(1+n^{-1/2}h_j(x_{-j},y_j)+s_{j,nh}(x_{-j},y_j)).
$$

Hence

$$
\begin{aligned}
Q_{(r)nh}(x,dy) &= \frac{1}{k}\sum_{j=1}^{k} Q_{j,nh}(x,dy)\\
&= \frac{1}{k}\sum_{j=1}^{k} Q_j(x,dy)(1+n^{-1/2}h_j(x_{-j},y_j)+s_{j,nh}(x_{-j},y_j))\\
&= Q_{(r)}(x,dy)\Big(1+n^{-1/2}\sum_{j=1}^{k}\Big(1(y_{-j}=x_{-j})-\Big(1-\frac{r_j(x)}{r(x)}\Big)1(y=x)\Big)h_j(x_{-j},y_j)\\
&\qquad +\sum_{j=1}^{k} s_{j,nh}(x_{-j},y_j)\Big)\\
&= Q_{(r)}(x,dy)\Big(1+n^{-1/2}(K_{(r)}h)(x,y)+s_{nh}(x,y)\Big)
\end{aligned}
$$

with

$$
s_{nh}(x,y)=\sum_{j=1}^{k}\Big(1(y_{-j}=x_{-j})-\Big(1-\frac{r_j(x)}{r(x)}\Big)1(y=x)\Big)s_{j,nh}(x_{-j},y_j)
$$

of order $O(n^{-1})$ uniformly in x and y.

Proof of Lemma 5. Since $\{y : y_{-j}=x_{-j}, y_j\neq x_j\}$ and $\{y=x\}$ are disjoint, we have

$$
\begin{aligned}
&(K_ih)(x,y)(K_jh')(x,y)\\
&=\Big(1(y_{-i}=x_{-i}, y_i\neq x_i)+\frac{r_i(x)}{r(x)}1(y=x)\Big)\\
&\quad\Big(1(y_{-j}=x_{-j}, y_j\neq x_j)+\frac{r_j(x)}{r(x)}1(y=x)\Big)h_i(x_{-i},y_i)h'_j(x_{-j},y_j)\\
&=\Big(\delta_{ij}1(y_{-j}=x_{-j}, y_j\neq x_j)+\frac{r_i(x)r_j(x)}{r(x)^2}1(y=x)\Big)h_i(x_{-i},y_i)h'_j(x_{-j},y_j).
\end{aligned}
$$

Hence

$$
\begin{aligned}
Q_{(r)}(x, K_ih\cdot K_jh') &= \frac{1}{k}\delta_{ij}\int Q_j(x,dy)1(y_j\neq x_j)h_j(x_{-j},y_j)h'_j(x_{-j},y_j)\\
&\qquad +\frac{1}{k}\frac{r_i(x)r_j(x)}{r(x)}h_i(x)h'_j(x)\\
&= \frac{1}{k}\delta_{ij}Q_j(x,h_jh'_j)-\frac{1}{k}h_i(x)R_{ij}(x)h'_j(x),
\end{aligned}
$$

with R_{ij} defined in (4.9). We obtain

$$\begin{aligned}\langle h, h'\rangle_{(r)} &= \pi \otimes Q_{(r)}(K_{(r)}h \cdot K_{(r)}h') \\ &= \frac{1}{k}\sum_{j=1}^{k}\langle h_j, h'_j\rangle_2 - \frac{1}{k}\sum_{i,j=1}^{k}\langle h_i, R_{ij}h'_j\rangle_2.\end{aligned} \tag{7.3}$$

Since $h_j = h - Q_j h$, we have $Q_j h_j = 0$ and therefore

$$\langle h_j, h'_j\rangle_2 = \langle h, h'_j\rangle_2. \tag{7.4}$$

The matrix $R(x)$ is symmetric. It has row and column sums 0,

$$\sum_{j=1}^{k} R_{ij} = r_j - \sum_{j=1}^{k}\frac{r_i(x)r_j(x)}{r(x)} = 0.$$

Hence, using selfadjointness (2.6) of Q_i,

$$\begin{aligned}\sum_{i,j=1}^{k}\langle h_i, R_{ij}h'_j\rangle_2 &= \sum_{i,j=1}^{k}\langle h - Q_i h, R_{ij}(h' - Q_j h')\rangle_2 \\ &= \sum_{i,j=1}^{k}\langle Q_i h, R_{ij}Q_j h'\rangle_2 \\ &= \sum_{i,j=1}^{k}\langle h, Q_i(R_{ij}Q_j h')\rangle_2.\end{aligned} \tag{7.5}$$

Applying (7.4) and (7.5) to (7.3), we can write

$$\begin{aligned}\langle h, h'\rangle_{(r)} &= \langle h, \frac{1}{k}\sum_{j=1}^{k} h'_j\rangle_2 - \langle h, \frac{1}{k}\sum_{i,j=1}^{k} Q_i(R_{ij}Q_j h')\rangle_2 \\ &= \langle h, (I - Q_{(r)})h'\rangle_2 - \langle h, Sh'\rangle_2\end{aligned}$$

with S defined in (4.8).

The matrix $R(x)$ is positive semidefinite:

$$\sum_{i,j=1}^{k} a_i R_{ij} a_j = \sum_{j=1}^{k} a_j^2 r_j - \frac{1}{r}\sum_{i,j=1}^{k} a_i r_i r_j a_j,$$

and by the Schwarz inequality,

$$\begin{aligned}\sum_{i,j=1}^{k} a_i r_i r_j a_j &= \sum_{j=1}^{k}(r_j^{1/2} \cdot a_j r_j^{1/2})^2 \\ &\le \sum_{j=1}^{k} r_j \sum_{j=1}^{k} a_j^2 r_j = r\sum_{j=1}^{k} a_j^2 r_j.\end{aligned}$$

The matrix $I - R(x)$ is also positive semidefinite:

$$\begin{aligned}\sum_{i,j=1}^{k} a_i(\delta_{ij} - R_{ij})a_j &= \sum_{j=1}^{k} a_j^2 - \sum_{j=1}^{k} a_j^2 r_j + \frac{1}{r}\sum_{i,j=1}^{k} a_i r_i r_j a_j \\ &= \sum_{j=1}^{k} a_j^2(1 - r_j) + \frac{1}{r}\Big(\sum_{j=1}^{k} a_j r_j\Big)^2 \geq 0.\end{aligned}$$

Since $R(x)$ is positive semidefinite, the operator S is positive: Use selfadjointness (2.6) of Q_i to write

$$\langle h,\, Sh\rangle_2 = \frac{1}{k}\sum_{i,j=1}^{k}\langle h,\, Q_i(R_{ij}Q_j h)\rangle_2 = \frac{1}{k}\sum_{i,j=1}^{k}\langle Q_i h,\, R_{ij}Q_j h\rangle_2 \geq 0.$$

Similarly, use property (2.4) of Q_j to obtain

$$\langle h,\, Q_{(r)}h\rangle_2 = \frac{1}{k}\sum_{j=1}^{k}\langle h,\, Q_j h\rangle_2 = \frac{1}{k}\sum_{j=1}^{k}\langle Q_j h,\, Q_j h\rangle_2$$

and hence

$$\langle h,\, (Q_{(r)} - S)h\rangle_2 = \frac{1}{k}\sum_{i,j=1}^{k}\langle Q_i h,\, (\delta_{ij} - R_{ij})Q_j h\rangle_2.$$

Since the matrix $I - R(x)$ is positive semidefinite, it follows that the operator $Q_{(r)} - S$ is positive.

Since $R(x)$ is symmetric and the Q_j are selfadjoint, S is selfadjoint:

$$\begin{aligned}\langle h,\, Sh'\rangle_2 &= \frac{1}{k}\sum_{i,j=1}^{k}\langle h,\, Q_i(R_{ij}Q_j h')\rangle_2 = \frac{1}{k}\sum_{i,j=1}^{k}\langle Q_i h,\, R_{ij}Q_j h'\rangle_2 \\ &= \frac{1}{k}\sum_{i,j=1}^{k}\langle R_{ji}Q_i h,\, Q_j h'\rangle_2 = \frac{1}{k}\sum_{i,j=1}^{k}\langle Q_j(R_{ji}Q_i h),\, h'\rangle_2 = \langle Sh,\, h'\rangle_2.\end{aligned}$$

The operator $I - Q_{(r)}$ is positive:

$$\langle h,\, (I - Q_{(r)})h\rangle_2 = \frac{1}{k}\sum_{j=1}^{k}\langle h,\, h_j\rangle_2 = \frac{1}{k}\sum_{j=1}^{k}\|h_j\|_2^2 \geq 0.$$

The operator $I - Q_{(r)} + S$ is positive as sum of the two positive operators $I - Q_{(r)}$ and S.

Proof of Theorem 2. By definition (4.7) and Lemma 5, the gradient $g_{(r)} \in L_{2,0}(\pi)$ of πf is defined by

$$\langle h,\, f - \pi f\rangle_2 = \langle h,\, g_{(r)}\rangle_{(r)} = \langle h,\, (I - Q_{(r)} + S)g_{(r)}\rangle_2 \quad \text{for } h \in H.$$

Since $\|(Q_{(r)} - S)^t\|_2 < 1$ for some t, the operator $I - Q_{(r)} + S$ has a bounded inverse on $L_{2,0}(\pi)$, and we obtain

$$g_{(r)} = (I - Q_{(r)} + S)^{-1}(f - \pi f).$$

Again by Lemma 5,

$$\|g_{(r)}\|_{(r)}^2 = \langle g_{(r)}, (I - Q_{(r)} + S)g_{(r)}\rangle_2 = \langle f - \pi f, (I - Q_{(r)} + S)^{-1}(f - \pi f)\rangle_2.$$

The alternative forms of $g_{(r)}$ and $\|g_{(r)}\|_{(r)}$ stated in Theorem 2 are obtained by a von Neumann expansion of $(I - Q_{(r)} + S)^{-1}$.

The following lemma was used in Section 5 to argue that the empirical estimator has larger variance under random sweep than under deterministic sweep when π has independent components. We have been unable to find a reference for this result in the literature, so it is included here for completeness.

Lemma 6. *Let a_{ts} be the number of ways of placing t different objects into s different cells, with no cell empty. Then,*

$$\sum_{t=s}^{\infty} \frac{a_{ts}}{k^t} = \binom{k-1}{s}^{-1}$$

for all $k \geq 2$ and $s = 1, \ldots, k-1$.

Proof. Consider k boxes, labeled $1, \ldots, k$, into which balls are thrown successively and at random. Let N_s denote the number of balls thrown from the time at which s of the boxes are occupied until $s+1$ of them are occupied. Note that N_s is a geometric random variable with expectation $k/(k-s)$, for $s = 1, \ldots, k-1$. This expectation can also be written in the form $EN_s = \binom{k}{s} EN_s^*$, where $N_s^* = \sum_{t=s}^{\infty} 1_{A_{ts}}$ and A_{ts} is the event that, immediately after the first t balls have been thrown, the only occupied boxes are $1, \ldots, s$. We conclude that

$$\sum_{t=s}^{\infty} \frac{a_{ts}}{k^t} = EN_s^* = \binom{k}{s}^{-1} EN_s = \binom{k-1}{s}^{-1}.$$

Acknowledgement. We are grateful to the referees for their helpful suggestions.

References

Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika* 82, 733–746.

Bickel, P.J. (1993). Estimation in semiparametric models. In: *Multivariate Analysis: Future Directions* (C.R. Rao, ed.), 55–73, North-Holland, Amsterdam.

Casella, G. and Robert, C. P. (1996). Rao–Blackwell theorem for sampling schemes with rejection. *C. R. Acad. Sci. Paris Sér. I Math.* 322, 571–576.

Clifford, P. and Nicholls, G. (1995). A Metropolis sampler for polygonal image reconstruction. Technical report, Department of Statistics, Oxford University.

Fishman, G. (1996). Coordinate selection rules for Gibbs sampling. *Ann. Appl. Probab.* 6, 444–465.

Frigessi, A., Hwang, C.-R., Sheu, S. J. and Di Stefano, P. (1993). Convergence rates of the Gibbs sampler, the Metropolis algorithm, and other single-site updating dynamics. *J. Roy. Statist. Soc. Ser. B* 55, 205–220.

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.

Frigessi, A. and Rue, H. (1998). Antithetic coupling of two Gibbs sampler chains. Preprint No. 7/98, Department of Mathematical Sciences, Norwegian University of Technology and Science, Trondheim.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* 85, 398–409.

Gelfand, A. E. and Smith, A. F. M. (1991). Gibbs sampling for marginal posterior expectations. *Comm. Statist. Theory Methods* 20, 1747–1766.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo. *Statist. Science* 7, 473–483.

Geyer, C. J. (1995). Conditioning in Markov chain Monte Carlo. *J. Comput. Graph. Statist.* 4, 148–154.

Green, P. J. and Han, X.-l. (1992). Metropolis methods, Gaussian proposals and antithetic variables. In: *Stochastic Models, Statistical Methods, and Algorithms in Image Analysis* (P. Barone, A. Frigessi and M. Piccioni, eds.), 142–164, Lecture Notes in Statistics 74, Springer-Verlag, Berlin.

Greenwood, P. E., McKeague, I. W. and Wefelmeyer, W. (1996). Outperforming the Gibbs sampler empirical estimator for nearest neighbor random fields. *Ann. Statist.* 24, 1433–1456.

Greenwood, P. E., McKeague, I. W. and Wefelmeyer, W. (1998). Von Mises type statistics for single site updated local interaction random fields. *Statistica Sinica*, to appear.

Greenwood, P. E. and Wefelmeyer, W. (1990). Efficiency of estimators for partially specified filtered models. *Stochastic Process. Appl.* 36, 353–370.

Greenwood, P.E. and Wefelmeyer, W. (1995). Efficiency of empirical estimators for

Markov chains. *Ann. Statist.* 23, 132–143.

Greenwood, P.E. and Wefelmeyer, W. (1998). Reversible Markov chains and optimality of symmetrized empirical estimators. *Bernoulli*, to appear.

Hájek, J. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. Verw. Gebiete* 14, 323–330.

Higdon, D. M. (1998). Auxiliary variable methods for Markov chain Monte Carlo with applications. *J. Amer. Statist. Assoc.* 93, 585–595.

Höpfner, R. (1993). On statistics of Markov step processes: representation of log-likelihood ratio processes in filtered local models. *Probab. Theory Related Fields* 94, 375–398.

Höpfner, R., Jacod, J. and Ladelli, L. (1990). Local asymptotic normality and mixed normality for Markov statistical models. *Probab. Theory Related Fields* 86, 105–129.

Ingrassia, S. (1994). On the rate of convergence of the Metropolis algorithm and Gibbs sampler by geometric bounds. *Ann. Appl. Probab.* 4, 347–389.

Kessler, M., Schick, A. and Wefelmeyer, W. (1998). The information in the marginal law of a Markov chain. In preparation.

Kipnis, C. and Varadhan, S. R. S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.* 104, 1–19.

Liu, J. S. (1996). Metropolized Gibbs sampler: an improvement. Technical report, Department of Statistics, Stanford University.

Liu, J. S., Wong, W. H. and Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika* 81, 27–40.

Liu, J. S., Wong, W. H. and Kong, A. (1995). Covariance structure and convergence rate of the Gibbs sampler with various scans. *J. Roy. Statist. Soc. Ser. B* 57, 157–169.

McEachern, S. N. and Berliner, L. M. (1994). Subsampling the Gibbs sampler. *Amer. Statist.* 48, 188–190.

McKeague, I. W. and Wefelmeyer, W. (1998). Markov chain Monte Carlo and Rao–Blackwellization. *J. Statist. Plann. Inference*, in press.

Mengersen, K. L. and Tweedie, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* 24, 101–121.

Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer-Verlag, London.

Meyn, S. P. and Tweedie, R. L. (1994). Computable bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab.* 4, 981–1011.

Penev, S. (1991). Efficient estimation of the stationary distribution for exponentially ergodic Markov chains. *J. Statist. Plann. Inference* 27, 105–123.

Peskun, P. H. (1973). Optimum Monte Carlo sampling using Markov chains. *Biometrika* 60, 607–612.

Raftery, A. E. and Lewis, S. M. (1992). How many iterations in the Gibbs sampler? In: *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid, A. F. M. Smith, eds.), 763–773, Oxford University Press.

Roberts, G. O. and Polson, N. G. (1994). On the geometric convergence of the Gibbs sampler. *J. Roy. Statist. Soc. Ser. B* 56, 377–384.

Roberts, G. O. and Rosenthal, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Comm. Probab.* 2, 13–25.

Roberts, G. O. and Sahu, S. K. (1997). Updating schemes, correlation structure, blocking and parameterisation for the Gibbs sampler. *J. Royal. Statist. Soc. Ser. B* 59, 291–317.

Roberts, G. O. and Tweedie, R. L. (1996). Geometric convergence and central limit theorems for multivariate Hastings and Metropolis algorithms. *Biometrika* 83, 95–110.

Rosenthal, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* 90, 558–566.

Schmeiser, B. and Chen, M.H. (1991). On random-direction Monte Carlo sampling for evaluating multidimensional integrals. Technical report, Department of Statistics, Purdue University.

Spiegelhalter, D. J., Thomas, A. and Best, N. G. (1996). Computation on Bayesian graphical models. In: *Bayesian Statistics 5* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.), 407–426, Oxford University Press.

Swendsen, R. H. and Wang, J.-S. (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.* 58, 86–88.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.* 82, 528–540.