On convergence rates of Gibbs samplers for uniform distributions

by

Gareth O. Roberts
Statistical Laboratory
University of Cambridge

and

Jeffrey S. Rosenthal
Department of Statistics
University of Toronto

Technical Report No. 9711, June 23, (1997)

TECHNICAL REPORT SERIES

University of Toronto
Department of Statistics

On convergence rates of Gibbs samplers for uniform distributions

by

Gareth O. Roberts
Statistical Laboratory
University of Cambridge

and

Jeffrey S. Rosenthal
Department of Statistics
University of Toronto

Technical Report No. 9711, June 23, (1997)

On convergence rates of Gibbs samplers for uniform distributions

by

Gareth O. Roberts* and Jeffrey S. Rosenthal**

(June 1997.)

1. Introduction.

This paper considers the use of Gibbs samplers applied to the uniform distribution on a bounded open region $R \subseteq \mathbf{R}^d$. We shall show that, subject to C^2 smoothness of the boundary of R, such Gibbs samplers are always uniformly ergodic. We shall also show that, even with certain types "pointy" boundaries, the Gibbs samplers are still geometrically ergodic.

By way of contrast, it has recently been shown by Bélisle (1997) that if the boundary of R is sufficiently *irregular*, then the Gibbs sampler can converge arbitrarily slowly. Our results thus complement those of Bélisle.

We note that our interest in Gibbs samplers arises partially from our interest in "slice sampler" or "auxiliary variable" algorithms, whereby sampling from a complicated $(d-1)$-dimensional density f is achieved by applying the Gibbs sampler to the uniform distribution on the d-dimensional region underneath the graph of f. Thus, Gibbs samplers for uniform distributions promise to be a very important subject in the future. For further details, see Higdon (1997), Damien et al. (1997), Mira and Tierney (1997), and Roberts and Rosenthal (1997b).

We begin with some definitions. Let $R \subseteq \mathbf{R}^d$ be a bounded open connected region in d-dimensional Euclidean space, and let $\pi(\cdot)$ be the uniform distribution on R (i.e.,

* Statistical Laboratory, University of Cambridge, Cambridge CB2 1SB, U.K. Internet: `G.O.Roberts@statslab.cam.ac.uk`.. Supported in part by EPSRC of the U.K.

** Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3. Internet: `jeff@utstat.toronto.edu`. Supported in part by NSERC of Canada.

$\pi(A) = \lambda(A \cap R)/\lambda(R)$ for Borel sets $A \subseteq \mathbf{R}^d$, where λ is d-dimensional Lebesgue measure). Let $\mathbf{X}^{(0)}$ be some random variable taking values in R. The *random-scan Gibbs sampler* proceeds as follows. Given a point $\mathbf{X}^{(n)} \in \mathbf{R}^d$, it chooses $I_{n+1} \in \{1, 2, \ldots, d\}$ uniformly at random. It then chooses $\mathbf{X}^{(n+1)}$ uniformly from the one-dimensional set

$$\left\{(X_1^{(n)}, \ldots, X_{I-1}^{(n)}, y, X_{I+1}^{(n)}, \ldots, X_d^{(n)}) \; ; \; y \in \mathbf{R}\right\} \; \cap \; R,$$

i.e. from the intersection of R with a line through $\mathbf{X}^{(n)}$ parallel to the i^{th} coordinate axis. This process is repeated for $n = 0, 1, 2, \ldots$.

Remark. Other versions of this algorithm are available. For example, instead of choosing a single coordinate I_{n+1} to update, it is possible to update all d coordinates in sequence, one at a time; this is the *deterministic-scan Gibbs sampler.* Also the Gibbs sampler may be defined for non-uniform distributions, by sampling from the full conditional distributions on the one-dimensional sets instead of sampling uniformly. For further details, see e.g. Gelfand and Smith, 1990; Smith and Roberts, 1993; Tierney, 1994).

The random-scan Gibbs sampler algorithm thus implicitly defines Markov chain transition probabilities $\mathcal{L}(\mathbf{X}^{(n+1)} \,|\, \mathbf{X}^{(n)})$. It is easily checked that the resulting Markov chain is reversible with respect to $\pi(\cdot)$. Furthermore the Markov chain is easily seen to be π-irreducible and aperiodic. Thus, from the general theory of Markov chains on general state spaces (see e.g. Nummelin, 1984; Meyn and Tweedie, 1993; Tierney, 1994, Section 3), we will have that

$$\|\mathcal{L}(\mathbf{X}^{(n)}) - \pi(\cdot)\| \;\equiv\; \sup_{A \subseteq \mathbf{R}^d} |\mathbf{P}(\mathbf{X}^{(n)} \in A) - \pi(A)| \;\to\; 0, \qquad n \to \infty.$$

(Here $\| \cdots \|$ is the *total variation distance* metric.)

A natural question is the rate at which this convergence takes place. It is shown by Bélisle (1997) that, without further restrictions on R, this convergence can be arbitrarily slow: for any sequence $\{b_n\}$ converging to 0, Bélisle shows that R and $\mathbf{X}^{(0)}$ can be chosen so that $\|\mathcal{L}(\mathbf{X}^{(n)}) - \pi(\cdot)\| \geq b_n$ for all sufficiently large n. However, it is reasonable to expect that if regularity conditions are imposed on R, then convergence will be faster.

Recall (cf. Meyn and Tweedie, 1993; Tierney, 1994) that a Markov chain with state space $\mathcal{X}$ and stationary distribution $\pi(\cdot)$ is *geometrically ergodic* if there is $\rho < 1$, a subset $\mathcal{X}_0 \subseteq \mathcal{X}$ with $\pi(\mathcal{X}_0) = 1$, and $M : \mathcal{X}_0 \to \mathbf{R}$ such that

$$\|\mathcal{L}(\mathbf{X}^{(n)} \,|\, \mathbf{X}^{(0)} = x_0) - \pi(\cdot)\| \;\leq\; M(x)\rho^n\,, \qquad n \in \mathbf{N}, \quad x_0 \in \mathcal{X}_0\,.$$

The chain is *uniformly ergodic* if it is geometrically ergodic with M constant (or, equivalently, with M bounded above). We note that geometric or uniform ergodicity ensures that the chain does *not* converge arbitrarily slowly in the sense of Bélisle.

In this paper, we shall show that for certain regions R (for example, if the boundary of R is C^2), the corresponding Gibbs sampler is uniformly ergodic (Section 2). For slightly less regular regions R, the Gibbs sampler is still geometrically ergodic (Section 3).

2. Uniform ergodicity.

In this section we shall derive conditions on R which ensure uniform ergodicity of the corresponding random-scan Gibbs sampler for the uniform distribution on R.

We recall (see e.g. Nummelin, 1984; Meyn and Tweedie, 1993) that, given a Markov chain on a state space $\mathcal{X}$, a subset $C \subseteq \mathcal{X}$ is *small* (or, (n_0, a, ν)-*small*) if for some $n_0 \in \mathbf{N}$, $a > 0$, and probability distribution $\nu(\cdot)$ on $\mathcal{X}$, we have

$$P^{n_0}(x, \cdot) \;\geq\; a\,\nu(\cdot)\,, \qquad x \in C\,.$$

We note that if $B \subseteq C$ and C is small, then B is also small (with the same n_0, a, and ν). We further recall (cf. Meyn and Tweedie, 1993, Theorem 16.0.1) that a Markov chain is uniformly ergodic *if and only if* the entire state space $\mathcal{X}$ is small, i.e. if and only if the above condition is satisfied with $C = \mathcal{X}$.

We begin with a simple lemma.

Lemma 1. *Let R be a bounded region in $\mathbf{R}^d$, and let C be a d-dimensional rectangle which lies entirely inside R. Then C is small for the Gibbs sampler on the uniform distribution on R (with either random- or deterministic-scan).*

Proof. If C has widths $a_1, a_2, \ldots, a_d$, and if R is bounded by a rectangle with widths $A_1, A_2, \ldots, A_d$, then the deterministic-scan Gibbs sampler starting inside C is clearly at least $\prod_i \frac{a_i}{A_i}$ times the uniform measure on C. For random-scan, we just need an extra factor of $d!/d^d$, which is the probability that the first d directions chosen include each direction precisely once. We thus obtain that

$$P_{DS}(x,\cdot) \;\geq\; \left(\prod_{i=1}^{d} \frac{a_i}{A_i}\right) \mathcal{U}_C(\cdot)\,; \qquad \text{and} \qquad P_{RS}(x,\cdot) \;\geq\; (d!/d^d) \left(\prod_{i=1}^{d} \frac{a_i}{A_i}\right) \mathcal{U}_C(\cdot)\,,$$

where P_{DS} and P_{RS} and the deterministic-scan and random-scan Gibbs samplers, respectively, and where $\mathcal{U}_C$ is the uniform distribution on C. ∎

To make use of this lemma, we require a general result about small sets. (A similar result is presented in Meyn and Tweedie, 1993, Proposition 5.5.5 (ii).)

Proposition 2. *For an irreducible aperiodic Markov chain, the finite union of small sets (each of positive stationary measure) is small.*

Proof. By induction, it suffices to consider just two small sets. Suppose that C_1 is (n_1, ϵ_1, ν_1)-small, and that C_2 is (n_2, ϵ_2, ν_2)-small.

By irreducibility, since $\pi(C_2) > 0$, there is $m \in \mathbf{N}$ and $\delta > 0$, such that $\nu_1 P^m(C_2) \equiv \int_R P^m(x, C_2)\nu_1(dx) \geq \delta$. It follows that $P^{n_1+m+n_2}(x,\cdot) \geq \epsilon_1 \delta \epsilon_2 \nu_2(\cdot)$ for $x \in C_1$. Also $P^{n_2}(x,\cdot) \geq \epsilon_2 \nu_2(\cdot) \geq \epsilon_1 \delta \epsilon_2 \nu_2(\cdot)$ for $x \in C_2$. Thus, $\sum_{n=1}^{\infty} P^n(x,\cdot) \geq \epsilon_1 \delta \epsilon_2 \nu_2(\cdot)$ for $x \in C_1 \cup C_2$. Hence, $C_1 \cup C_2$ is "petite" in the sense of Meyn and Tweedie (1993, p. 121).

But then by irreducibility and aperiodicity, it follows (cf. Meyn and Tweedie, 1993, Theorem 5.5.7) that $C_1 \cup C_2$ must be small. ∎

We now put these results together. For $\mathbf{x} \in \mathbf{R}^d$, we shall write $B(x, \epsilon)$ for the open L^1 cube centered at $\mathbf{x}$ of radius ϵ, i.e.

$$B(\mathbf{x}, \epsilon) \;=\; \left\{\mathbf{y} \in \mathbf{R}^d\,;\, x_i - \epsilon < y_i < x_i + \epsilon\,,\ i = 1, 2, \ldots, d\right\}\,.$$

We begin with the case where R is a triangle. We recall from the previous section that, if the triangle is such that all vertices have apex which contains a coordinate direction, then the associated Gibbs sampler is uniformly ergodic. Thus, we instead consider the case where one of the vertices is "tilted" and does not contain a coordinate direction.

Proposition 8. *Let $R \subseteq \mathbf{R}^2$ be the width-1 triangle with lower angle θ, and upper angle ϕ, i.e.*

$$R = \left\{(x,y) \in \mathbf{R}^2;\ 0 < x < 1,\ x\tan(\theta) < y < x\tan(\phi)\right\},$$

where $0 < \theta < \phi < \pi/2$. Then the Gibbs sampler (with either random- or deterministic-scan) for the uniform distribution on R is geometrically ergodic.

Proof. We recall from the previous section that the subset $C = \{(x,y) \in R; y > \tan(\phi)\}$ (say) is small for the Gibbs sampler. Thus, by standard Markov chain theory (see e.g. Nummelin, 1984; Meyn and Tweedie, 1993, Theorem 15.0.1), we will be done if we can find a *drift function* $V : R \to [1,\infty)$ and $\lambda < 1$ such that

$$PV(x,y) \equiv \int_R V(\mathbf{z})P\left((x,y), d\mathbf{z}\right) \leq \lambda V(x,y), \qquad (x,y) \in R, \quad y < \tan(\phi).$$

To continue, we consider the drift function $V(x,y) = 1/x$. To compute $PV(x,y)$, for ease of computation we shall focus on the deterministic-scan Gibbs sampler on R which updates first the y coordinate and then the x coordinate, rather than on the random-scan Gibbs sampler. This is not a restriction since it is known (see e.g. Roberts and Rosenthal, 1997a, Proposition 5) that if the deterministic-scan Gibbs sampler is geometrically ergodic, then so is the random-scan Gibbs sampler.

We compute that, for the deterministic-scan Gibbs sampler,

$$\begin{aligned} PV(x,y) &= \frac{1}{x\tan(\phi) - x\tan(\theta)} \int_{x\tan(\theta)}^{x\tan(\phi)} \frac{1}{w\cot(\theta) - w\cot(\phi)} \int_{w\cot(\phi)}^{w\cot(\theta)} V(z,w)\,dz\,dw \\ &= \lambda V(x,y), \end{aligned}$$

where

$$\lambda = \lambda(\theta,\phi) = [\log(\cot(\theta)/\cot(\phi))]^2 \;/\; [(\tan(\phi) - \tan(\theta)(\cot(\theta) - \cot(\phi))].$$

(Note that we actually have equality here, even though we only require an inequality.) Now, we have $\lambda(\theta, \phi) < 1$ whenever $0 < \theta < \phi < \pi/2$; indeed, if we set $f(\epsilon) = \lambda(\theta, \theta + \epsilon)$, then to second order in ϵ, as $\epsilon \to 0^+$, we have

$$f(\epsilon) \approx 1 - \epsilon^2/(3\sin^2(2\theta)) < 1.$$

The geometric ergodicity follows. ∎

It is possible to combine Proposition 8 with the results of Section 2. For example, we have

Theorem 9. *Suppose R is a region whose boundary is a $(d-1)$-dimensional C^2 manifold except at a finite number of points. Suppose further that in a neighbourhood of each of these exceptional points, R coincides with a triangle (as in Proposition 8). Then the random-scan Gibbs sampler for the uniform distribution on R is geometrically ergodic.*

Proof. (Outline.) As noted at the end of Section 2, the Gibbs sampler is uniformly ergodic except near those exceptional points whose vertices are "tilted", i.e. have apexes which do not contain any coordinate direction. For such tilted vertices, it is possible to choose $\epsilon > 0$ small enough that $R \setminus R_\epsilon$ breaks up into a finite number of connected components, one near each exceptional point, such that it is impossible to get from one of these components to another in a single step. Once we have done that, then we define a drift function V to be equal to 1 on R_ϵ, and equal to the appropriate drift function (as in the proof of Proposition 8) on each of the different connected components of $R \setminus R_\epsilon$. Then, separately from each connected component, the Gibbs sampler has geometric drift back to the small set R_ϵ. Hence, as in Proposition 8, the result follows. ∎

Similar results are available for higher-dimensional regions R having "vertices" on the boundary. We illustrate this with a particular example, a "tilted cone" with base at the origin, tilted so that it does not contain any coordinate direction.

Proposition 10. *Suppose $R \subseteq \mathbf{R}^3$ is the tilted cone*

$$R = \left\{(x,y,z) \in \mathbf{R}^3; 0 < x < 1,\ z^2 + \frac{(\alpha x - y)^2}{1+\alpha^2} < c\frac{(x+\alpha y)^2}{1+\alpha^2}\right\},$$

for some $\alpha > 0$ and $0 < c < 1$. Then the Gibbs sampler (with either random- or deterministic-scan) for the uniform distribution on R is geometrically ergodic.

Proof. We use the same drift function $V(x,y,z) = 1/x$ as before. We consider the deterministic-scan Gibbs sampler which updates first z, then y, and then x. (The corresponding result for the random-scan Gibbs sampler then follows once again from Roberts and Rosenthal, 1997a, Proposition 5.) Clearly updating z does not change the value of V, so it suffices to consider the effect of updating x and y conditional on a fixed value of z.

Now, conditional on $z = 0$, the point (x,y) is restricted to the *triangle*

$$R \cap \{z = 0\} = \left\{(x,y,0) \in \mathbf{R}^3;\ x\tan(\theta) < y < x\tan(\phi)\right\},$$

for some $0 < \theta < \phi < \pi/2$. Furthermore, conditional on a particular value of $z \neq 0$, the point (x,y) is restricted to a *hyperbola* lying inside (and asymptotic to) the triangle $R \cap \{z = 0\}$, whose proximity to this triangle depends on z.

To proceed, let P_{z_0} be the two-dimensional random-scan Gibbs sampler for the uniform distribution on $R \cap \{z = z_0\}$, i.e. which acts on the coordinates x and y while leaving the value of z fixed at $z = z_0$. Then P_0 is the usual two-dimensional random-scan Gibbs sampler on the triangle $R \cap \{z = 0\}$, and hence by Proposition 8, P_0 is geometrically ergodic with $P_0 V(x,y,z) \leq \lambda V(x,y,z)$ for some $\lambda < 1$.

Now, we claim that for any choice of $z_0 \in \mathbf{R}$ such that $R \cap \{z = z_0\}$ is non-empty, we have $P_{z_0} V(x,y,z_0) \leq P_0 V(x,y,z_0)$. Indeed, for fixed z_0 we have

$$P_{z_0} V(x,y,z) = \frac{1}{y_2(x) - y_1(x)} \int_{y_1(x)}^{y_2(x)} \frac{1}{x_2(w) - x_1(w)} \int_{x_1(w)}^{x_2(w)} (1/z)\, dz\, dw\,, \qquad (\dagger)$$

where $y_1(x)$, $y_2(x)$, $x_1(w)$, and $x_2(w)$ are defined by

$$R \cap \{(x,t); t \in \mathbf{R}\} = \{(x,t); y_1(x) < t < y_2(x)\};$$

and

$$R \cap \{(t,w); t \in \mathbf{R}\} = \{(t,w); x_1(x) < w < x_2(x)\}.$$

It is furthermore verified that there are functions $d(x)$ and $D(y)$ (which also depend on θ, ϕ, and z_0) such that

$$y_1(x) = x\cot(\theta) + d(x); \quad y_2(x) = x\cot(\phi) - d(x);$$

$$x_1(y) = y\tan(\phi) + D(y); \quad x_2(y) = y\tan(\theta) - D(y);$$

that is, the interval $(x_1(y), x_2(y))$ is symmetrically embedded in the interval $(y\cot(\phi), y\cot(\theta))$ (and similarly for $(y_1(x), y_2(x))$).

To show that $P_{z_0}V \leq P_0 V$, we observe that, for fixed $0 < a < b$ and $0 \leq k < (b-a)/2$, the quantity $\frac{1}{b-a-2k}\int_{a+k}^{b-k}(1/z)dz$ as a function of k is maximised at $k=0$. Applying this observation twice to (†) shows that $P_{z_0}V(x,y,z) \leq P_0 V(x,y,z)$ as desired.

It follows that the deterministic-scan Gibbs sampler on R is again geometrically ergodic, with at least as small a value of λ as the corresponding value from Proposition 8. ∎

Finally, we turn our attention to showing that certain Gibbs samplers are *not* geometrically ergodic. We begin with a result, following Lawler and Sokal (1988), which may be viewed as a generalisation of Roberts and Tweedie (1996, Theorem 5.1).

Lemma 11. *Let $P(x,\cdot)$ be the transition probabilities for a Markov chain on a state space $\mathcal{X}$, having stationary distribution $\pi(\cdot)$. Suppose that, for any $\delta > 0$, there is a subset $A \subseteq \mathcal{X}$ with $0 < \pi(A) < 1$ such that*

$$\frac{\int_A P(\mathbf{x}, A^C)\pi(d\mathbf{x})}{\pi(A)\pi(A^C)} < \delta.$$

Then the Markov chain is not geometrically ergodic.

G.F. Lawler and A.D. Sokal (1988), Bounds on the L^2 spectrum for Markov chains and Markov processes: A generalization of Cheeger's inequality. Trans. Amer. Math. Soc. **309**, 557–580.

S.P. Meyn and R.L. Tweedie (1993), Markov chains and stochastic stability. Springer-Verlag, London.

A. Mira and L. Tierney (1997), Efficient Bayesian inference for nonconjugate models. In preparation, School of Statistics, University of Minnesota.

E. Nummelin (1984), General irreducible Markov chains and non-negative operators. Cambridge University Press.

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

G.O. Roberts and J.S. Rosenthal (1997b), Convergence of slice sampler Markov chains. In preparation.

G.O. Roberts and R.L. Tweedie (1996), Geometric convergence and central limit theorems for multidimensional Hastings Metropolis algorithms. Biometrika **83**, 96–110.

A.F.M. Smith and G.O. Roberts (1993), Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). J. Roy. Stat. Soc. Ser. B **55**, 3–24.

L. Tierney (1994), Markov chains for exploring posterior distributions (with discussion). Ann. Stat. **22**, 1701–1762.

1 SIDE