Convergence of slice sampler Markov chains

by

Gareth O. Roberts
Statistical Laboratory
University of Cambridge

and

Jeffrey S. Rosenthal
Department of Statistics
University of Toronto

Convergence of slice sampler Markov chains

by

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

(July 2, 1997.)

1. Introduction.

This paper considers the use of slice samplers to sample from a complicated \(d\)-dimensional probability distribution. Slice samplers are a form of auxiliary variable technique, which introduces auxiliary random variables \(Y_1, \ldots, Y_k\) to facilitate the design of an improved Markov chain Monte Carlo (MCMC) sampling algorithm.

The idea of using of auxiliary variables for improving MCMC was introduced for the Ising model by Swendsen and Wang (1987). Edwards and Sokal (1988) generalised the Swendsen-Wang technique, and since then, their use in statistical problems has gradually increased, partly as a result of Besag and Green (1993). In recent years there has been a large amount of activity on this topic, including a very clear discussion of auxiliary variable techniques by Higdon (1996), a variety of examples of uses of auxiliary variable techniques in statistical problems by Damien and Walker (1997), and some theoretical progress by Mira and Tierney (1997).

However, except for the original Swendsen-Wang method (which has been shown to be superior to more naive Gibbs methods for sub-critical Ising models), rather little is known about the theoretical properties of auxiliary variable algorithms.

In this paper, we concentrate on the slice sampler (defined in Section 2), demonstrating that it has a number of good theoretical properties. In particular, we shall show that it is

* 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.
usually geometrically ergodic (Section 3), and that it often converges quickly with useful rigorous quantitative bounds available (Section 4). We also consider (Section 5) further properties of the so-called product slice sampler, corresponding to \( k \geq 2 \) above.

2. Slice samplers: definitions and preliminaries.

Suppose that \( \pi : \mathbb{R}^d \rightarrow [0, \infty) \) is a density (i.e., a non-negative measurable function which is not a.e. 0) with respect to \( d \)-dimensional Lebesgue measure. Such a density gives rise to a probability measure \( \nu_\pi(\cdot) \), by

\[
\nu_\pi(A) = \frac{\int_A \pi(x) \, dx}{\int_{\mathbb{R}^d} \pi(x) \, dx}, \quad A \subseteq \mathbb{R}^d.
\]

Typically, \( \pi \) is a complicated function, and \( d \) is reasonably large. The slice sampler then provides a Markov chain algorithm which can be used to sample from \( \nu_\pi(\cdot) \).

Specifically, suppose \( \pi \) can be written as \( \pi(x) = \prod_{i=0}^k f_i(x) \), for some densities \( f_i : \mathbb{R}^d \rightarrow [0, \infty) \). The \( f_0 \)-slice sampler, \( P_{f_0} \), proceeds as follows. Given \( X_n \), we sample \( k \) independent uniform random variables \( Y_{n+1,1}, Y_{n+1,2}, \ldots, Y_{n+1,k} \), with \( Y_{n+1,i} \sim U(0, f_i(x)) \). We then sample \( X_{n+1} \) from the truncated probability distribution having density proportional to \( f_0(\cdot)1_{L(Y_{n+1})}(\cdot) \), where

\[
L(y) = \{ x \in \mathbb{R}^d : f_i(x) \geq y_i, \ i = 1, 2, \ldots, k \}.
\]

This algorithm gives rise to a Markov chain \( \{X_n\}_{n=0}^\infty \), having transition probabilities \( P_{f_0}(x, A) \equiv P(X_{n+1} \in A \mid X_n = x) \). This Markov chain is easily seen to have \( \nu_\pi(\cdot) \) as a stationary distribution. Furthermore, it is easily seen to be \( \nu_\pi \)-irreducible and aperiodic. Thus, from standard Markov chain theory (see e.g. Tierney, 1994) it follows that from \( \nu_\pi \)-almost every starting point, the law of \( X_n \) will converge to \( \nu_\pi(\cdot) \) as \( n \to \infty \). This algorithm appears to be extremely useful in practice; see e.g. Besag and Green (1993), Higdon (1996), Damien and Walker (1997), and Mira and Tierney (1997).

The following proposition shows that, for theoretical purposes at least, it suffices to consider the case where \( f_0 \) is the indicator function of a (possibly infinite) subset of \( \mathbb{R}^d \).
Proposition 1. Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be a differentiable injective transformation. Let $J$ be its Jacobian (assumed to be positive everywhere). Then $P_{f_0}(x, A) = P_{f_0/J}(T(x), T(A))$. That is, the $f_0$-slice sampler on $f_0(x)f_1(x) \ldots f_k(x)$ behaves identically to the $(f_0 \circ T^{-1} / J \circ T^{-1})$-slice sampler on $(f_0(T^{-1}(x))/J(T^{-1}(x)))f_1(T^{-1}(x)) \ldots f_k(T^{-1}(x))$ (where we take $f_0(T^{-1}(x)) = 0$ if $x$ is not in the range of $T$). Furthermore, it is always possible to find such a transformation $T$ for which the quotient $f_0/J$ is equal to the indicator function of a (possibly infinite) subset of $\mathbb{R}^d$.

Proof. The first statement follows directly from the multi-dimensional change of variable formula (see e.g. Marsden, 1974, Section 9.3); specifically, sampling $T(x)$ from the density $(f_0(T^{-1}(\cdot))/J(T^{-1}(\cdot))) 1_{T(L(Y))}(\cdot)$ is equivalent to sampling $x$ from the density $f_0(\cdot)1_{L(Y)}$.

For the final statement, we define $T_1(x) = \int_0^{x_1} f_0(t, x_2, \ldots, x_d)dt$. We further define $T_i(x) = x_i$ for $i \geq 2$. We then set $T = (T_1, T_2, \ldots, T_k)$. It is easily verified that this gives $J(x) = f_0(x)$, so that $f_0(T^{-1}(x))/J(T^{-1}(x))$ is equal to the indicator function of the range of $T$.

Remark. This implies that it is sufficient to consider the case where the sampling in the $X$ component is with respect to an appropriate uniform distribution. We further note that $T(\mathbb{R}^d)$ has finite Lebesgue measure if and only if $\int f_0 < \infty$; in fact, in that case the Lebesgue measure of $T(\mathbb{R}^d)$ is precisely equal to $\int f_0(x)dx$.

As a result of this remark, we concentrate from now on on the uniform slice sampler, i.e. on the case when $f_0$ takes on only the values 0 and 1. In this case, we shall write the slice sampler Markov chain transition probabilities as $P_{ssl}$ (for “simple slice”) when $k = 1$, and as $P_{psl}$ (for “product slice”) when $k \geq 2$.

For $P_{ssl}$ (i.e. in the case where $f_0$ is an indicator function, and $k = 1$), we shall write $L(y) = \{x \in \mathbb{R}^d; \pi(x) \geq y\}$, and shall write $Q(y)$ for $m(L(y))$, where $m$ is $d$-dimensional Lebesgue measure. The algorithm then proceeds by alternately updating $Y_{n+1} \sim U[0, \pi(X_n)]$, and $X_{n+1} \sim U(L(Y_{n+1}))$. We note that the behaviour of the simple
slice sampler is \textit{completely determined} by the function $Q$; indeed, two different densities which gave rise to the same function $Q$ would have identical simple-slice-sampler convergence properties. This is also true for constant scaling, as the following proposition records.

\textbf{Proposition 2.} Let $\pi$ and $\pi'$ be two different densities, of dimension $d$ and $d'$ respectively. Suppose there exists $a > 0$ such that their corresponding functions $Q$ and $\tilde{Q}$ satisfy $Q(y) = \tilde{Q}(ay)$, for all $y > 0$. Then the convergence properties of the (uniform) simple slice sampler $P_{sel}$ for $\pi$ and for $\pi'$ are identical. Specifically, we have

$$P(\pi(X_{n+1}) \leq z \mid \pi(X_n) = y) = P(\pi'(X_{n+1}) \leq az \mid \pi'(X_n) = ay), \quad y, z > 0.$$ 

\textbf{Proof.} We have that

$$P(\pi(X_{n+1}) \leq z \mid \pi(X_n) = y) = \frac{1}{y} \int_0^y \max\left(\frac{Q(z)}{Q(y)}, 1\right) dy,$$

and this expression is clearly unchanged upon replacing $y$ by $ay$, and $z$ by $az$, and $Q$ by $\tilde{Q}$. The result follows. \qed

\textbf{Remarks.}

1. This proposition shows that, for theoretical purposes, an arbitrary simple slice sampler is equivalent to the one-dimensional simple slice sampler on the density $f(x) = \inf\{w > 0; Q(w) \geq x\}$ for $x > 0$ (with $f(x) = 0$ for $x \leq 0$), since such a density has the appropriate value for $Q(y)$. This is often a helpful way to think about slice samplers.

2. This proposition clearly also applies if $f_0$ is \textit{not} uniform, provided we use the more general definition $Q(y) = \int_{L(y)} f_0(z)dz$ instead of the uniform-specific definition $Q(y) = m(L(y))$. However, it does require that we are in the simple slice sampler case $k = 1$; in general we would need a $k$-dimensional function $Q$ to completely specify the slice sampler convergence properties.
To continue, we define a partial ordering on \( \mathbb{R}^d \) based on values of \( \pi \). That is, we say that \( x_1 \preceq x_2 \) if and only if \( \pi(x_1) \leq \pi(x_2) \). Now, recall that a Markov chain \( X \) on a totally ordered space is said to be stochastically monotone if for all fixed \( z \), we have that 
\[
P(X_1 \preceq z | X_0 = x_1) \geq P(X_1 \preceq z | X_0 = x_2)
\]
whenever \( x_1 \preceq x_2 \). (Stochastically monotone chains are usually easier to analyse than more general classes of chains.) We have

**Proposition 3.** With the ordering on \( \mathbb{R}^d \) given above, \( P_{sel} \) is stochastically monotone.

**Proof.** We see (as in the previous proof) that for \( i = 1, 2 \), setting \( w = \pi(z) \), we have

\[
P(X_1 \preceq z | X_0 = x_i) = P(\pi(X_1) \leq w | X_0 = x_i)
\]

\[
= \frac{1}{\pi(x_i)} \int_0^{\pi(x_i)} \max \left( \frac{Q(y)}{Q(w)}, 1 \right) dy,
\]

i.e. is an average of the function \( f(y) = \max \left( \frac{Q(y)}{Q(w)}, 1 \right) \), averaged over the interval \([0, \pi(x_i)]\). But clearly \( f \) is non-increasing. Hence, if \( \pi(x_1) \leq \pi(x_2) \), then \( P(X_1 \preceq z | X_0 = x_1) \geq P(X_1 \preceq z | X_0 = x_2) \), as required.

Although the Markov chain \( \{\pi(X_n), n \in \mathbb{N}\} \) is a non-trivial simplification of \( \{X_n, n \in \mathbb{N}\} \), the convergence properties of the two chains are identical, since by the construction of the algorithm, the conditional distribution of \( X_n \) given that \( \pi(X_n) = y \) is uniformly distributed, for all \( n \geq 1 \).

To end this section, we mention a result which is presented in Mira and Tierney (1997), using a theorem of Peskun (1973; see also Tierney, 1995, Section 3). We therefore omit the proof.

**Proposition 4.** Suppose \( \pi \) is bounded and \( \text{supp}(\pi) \) has finite Lebesgue measure. Then \( P_{sel} \) is uniformly ergodic, with principle eigenvalue being bounded above by the rate of convergence of the independence sampler with uniform proposal distribution.

In this section, we consider the geometric ergodicity of slice samplers. We concentrate on the case \( P_{ssl} \), i.e. on the case where \( f_0 \) is an indicator function and \( k = 1 \). For some of our results, we shall further assume that \( \pi \) is a bounded function. In that case, since the slice-sampler is scale-invariant (Proposition 2), it suffices to assume that \( \pi \leq 1 \), i.e. that \( \pi \) is bounded by 1.

Recall (see e.g. Nummelin, 1984; Meyn and Tweedie, 1993) that a Markov chain \( P(x, \cdot) \) on a state space \( \mathcal{X} \), having stationary distribution \( \nu(\cdot) \), is geometrically ergodic if there is \( \rho < 1 \) and a \( \nu \)-a.e.-finite function \( V : \mathcal{X} \to [1, \infty] \), such that

\[
||P(X_n \in \cdot | X_0 = x) - \nu(\cdot)|| \equiv \sup_{A \subseteq \mathcal{X}} |P(X_n \in A | X_0 = x) - \nu(A)| \leq V(x)\rho^n, \quad x \in \mathcal{X}.
\]

Recall further that this is equivalent to the existence of a \( \pi \)-a.e. finite function \( V : \mathcal{X} \to [1, \infty] \), a subset \( C \subseteq \mathcal{X} \), a probability measure \( \mu(\cdot) \) on \( \mathcal{X} \), and constants \( \epsilon > 0 \), \( \lambda < 1 \), and \( b < \infty \), such that (a) \( P(x, \cdot) \geq \epsilon \mu(\cdot) \) for all \( x \in C \) (i.e., the set \( C \) is small); and (b) \( PV(x) \leq \lambda V(x) + b1_C(x) \) for all \( x \in \mathcal{X} \) (i.e., \( V \) satisfies a drift condition). We shall examine these two conditions separately.

Condition (a) above is fairly straightforward. Indeed, we have the following.

**Proposition 5.** Consider the slice sampler \( P_{ssl} \) on a density \( \pi \). For any fixed \( y^* > y_* > 0 \), define the subset \( C \subseteq \mathbb{R}^d \) by

\[
C = \{ x \in \mathbb{R}^d ; y_* \leq \pi(x) \leq y^* \}.
\]

Then we have

\[
P_{ssl}(x, \cdot) \geq \frac{y_*}{y^*} \mu(\cdot), \quad x \in C,
\]

where \( \mu(A) = y_*^{-1} \int_0^{y_*} \mathcal{U}(L(z)) dz \). That is, the set \( C \) is small with \( \epsilon = y_* / y^* \). In particular, of \( \pi \) is bounded (without loss of generality by 1 say), then \( L(y_*) \) is small with \( \epsilon = y_* \).
Proof. If we start the slice sampler at some $X_n \in C$, then we clearly have

$$\mathcal{L}(Y_{n+1} | X_n) \geq \frac{y^*}{y^n} \mathcal{U}([0, y^*]).$$

But since $\mathcal{L}(X_{n+1} | Y_{n+1}) = \mathcal{U}(L(Y_{n+1}))$, the result follows immediately. $\blacksquare$

To continue, we need to establish a drift condition (i.e., condition (b) above) for $P_{ssl}$. This is somewhat more difficult. We shall need the following well known stochastic approximation result (the “FKG inequality”), which we state in a way relevant to our current context. Briefly, it states that if a measure has a non-decreasing Radon-Nikodym derivative with respect to another measure, then it will have larger conditional expected value of any non-decreasing function. For a similar application to conditional expectations, and discussion of the result, see Roberts (1991).

Lemma 6. Suppose that $\mathcal{M}_1$ and $\mathcal{M}_2$ are two probability measures on $\mathbb{R}$, such that there is a version of the Radon-Nikodym derivative $R(x) = \mathcal{M}_2(dx)/\mathcal{M}_1(dx)$, which is a non-decreasing function. Suppose also that $f$ is a non-decreasing function from $\mathbb{R}$ into $\mathbb{R}^+$. Let $E_i$, $i = 1, 2$ denote expectations with respect to the two measures $\mathcal{M}_i$, $i = 1, 2$. Then for any set $A$ for which the following conditional expectations exist,

$$E_1[f(X)|X \in A] \leq E_2[f(X)|X \in A].$$

Using this lemma, we are now able to establish a drift condition for $P_{ssl}$.

Proposition 7. Consider the slice sampler $P_{ssl}$ on a density $\pi \leq 1$. Suppose its corresponding function $Q(y) = m(L(y))$ is differentiable, and that there exists a constant $\alpha > 1$ such that $Q'(y)y^{1+\frac{1}{\alpha}}$ is non-increasing, at least for $y \leq Y$. Then, for any $\beta$ with $0 < \beta < \min\left(\frac{\alpha-1}{\alpha}, \frac{1}{\alpha}\right)$, and for any $y^* \in (0, Y)$, we have

$$P_{ssl}V(x) \leq \lambda V(x) + b1_{L(y^*)}(x), x \in \mathbb{R}^d,$$

where $V(x) = \pi(x)^{-\beta}$, and where

$$\lambda \equiv \frac{1}{(1-\beta)(1+\alpha\beta)} + \frac{\alpha\beta(y^*/Y)^{\beta}}{1+\alpha\beta}$$
and
\[ b = \max \left( \frac{Y^{-\beta}(1 + \alpha \beta(1 - \beta))}{(1 - \beta)(1 + \alpha \beta)} - \lambda, \frac{2\alpha \beta}{1 + \alpha \beta} \right). \]
(Since \((1 - \beta)(1 + \alpha \beta) > 1\) for \(0 < \beta < \frac{\alpha - 1}{\alpha}\), it follows that by choosing \(y_* > 0\) sufficiently small, we can insure that \(\lambda < 1\).) Furthermore, if \(Y = 1\) then the formula for \(b\) may be simplified to
\[ b = \frac{1 + \alpha \beta(1 - \beta)}{(1 - \beta)(1 + \alpha \beta)} - \lambda. \]

**Proof.** We note that if \(x \in \mathbb{R}^d\) is such that \(\pi(x) \leq Y\), then
\[
P_{x \cdot t} V(x) = \frac{1}{\pi(x)} \int_0^{\pi(x)} \frac{1}{Q(y)} \int_{L(y)} \pi(z)^{-\beta} dz \ dy
\]
\[= \frac{1}{\pi(x)} \int_0^{\pi(x)} \frac{1}{Q(y)} \int_y^{\infty} w^{-\beta}(-Q'(w)) dw \ dy
\]
\[= \frac{1}{\pi(x)} \int_0^{\pi(x)} \left( \int_y^{\infty} + \int_y^{Y} \right) w^{-\beta}(-Q'(w)) dw \ dy
\]
\[\leq \frac{1}{\pi(x)} \int_0^{\pi(x)} \int_y^{Y} w^{-\beta}(-Q'(w)) dw \ dy
\]
\[\leq \frac{1}{\pi(x)} \int_0^{\pi(x)} \int_y^{Y} w^{-(1+\beta+\alpha^{-1})} dw \ dy
\]
\[= \frac{1}{1 + \alpha \beta} \left( \int_0^{\pi(x)} y^{-\beta} dy + \int_0^{\pi(x)} \frac{\pi(x) Y^{-\alpha^{-1}}(y^{-\beta} - Y^{-\beta})}{y^{-\alpha^{-1}} - Y^{-\alpha^{-1}}} dy \right)
\]
\[\leq \frac{V(x)}{(1 - \beta)(1 + \alpha \beta)} + \frac{\alpha \beta Y^{-\beta}}{1 + \alpha \beta}.
\]
Here the first equality follows simply from writing out the definition of \(PV(x)\), and the second equality then follows from rewriting the inner integral with respect to the measure \(-Q'(w) dw\). The first inequality follows from the fact that \(w^{-\beta}\) is a non-increasing function. The second inequality follows from Lemma 3 with \(M_2(dy) \propto y^{-(1+\alpha^{-1})} dy\) and \(M_1(dy) \propto (-Q'(y)) dy\). The third inequality follows from the fact that \((y^{-\beta} - Y^{-\beta})/(y^{-\alpha^{-1}} - Y^{-\alpha^{-1}})\) is a non-decreasing function of \(y \in (0, Y)\), at least when \(\beta \alpha < 1\) as we’ve stipulated; this can be checked by differentiating with respect to \(y\) and then maximising over \(Y\). Hence an upper bound for this function is obtained by taking the limit as \(y \to Y\).
For \( \pi(x) \geq Y \), we note that by stochastic monotonicity (Proposition 3), it follows that \( P_{\text{sel}}V(x) \) is non-increasing according to the ordering \( \preceq \) on \( \mathbb{R}^d \). Therefore, if \( x \) is such that \( \pi(x) \geq Y \), then we must have \( P_{\text{sel}}V(x) \leq P_{\text{sel}}V(x') \) where \( \pi(x') = Y \). Hence, from the above bound on \( P_{\text{sel}}V(x') \), we have that
\[
P_{\text{sel}}V(x) \leq Y^{-\beta} \frac{1 + \alpha \beta (1 - \beta)}{(1 + \alpha \beta)(1 - \beta)}, \quad \pi(x) \geq Y.
\]

Now let \( \lambda \) and \( b \) be as in the statement of the proposition. Then it is easily verified (by considering separately the cases \( \pi(x) < y \), \( y < \pi(x) < Y \), and \( \pi(x) > Y \)) that \( P_{\text{sel}}V(x) \leq \lambda V(x) + b1_{L(y)}(x) \), as required.

The final statement of the proposition follows because, if \( Y = 1 \), then there is no case \( \pi(x) > Y \) to consider. Hence, in this case it is easily verified that we still have \( P_{\text{sel}}V(x) \leq \lambda V(x) + b1_{L(y)}(x) \) with the new, simpler formula for \( b \).

Putting Propositions 5 and 7 together, and using the standard Markov chain theory discussed at the beginning of this section, we obtain

**Theorem 8.** Consider the slice sampler \( P_{\text{sel}} \) on a bounded density \( \pi \). Suppose its corresponding function \( Q(y) = m(L(y)) \) is differentiable, and that there exists a constant \( \alpha > 1 \) such that \( Q'(y)y^{1 + \frac{1}{\alpha}} \) is non-increasing, at least on an open set containing 0. Then \( P_{\text{sel}} \) is geometrically ergodic.

**Remarks.**

(1) These conditions are really rather weak. For instance, for \( \mathcal{X} = \mathbb{R} \) the condition on \( Q'(y)y^{1 + \frac{1}{\alpha}} \) can be loosely stated as saying that \( \pi \) has tails that are at least as light as \( x^{-\alpha} \). A couple of examples illuminate this.

(i) Suppose that \( \mathcal{X} = \mathbb{R}^+ \) and that \( \pi \) is a positive continuous density. Suppose also that \( \pi \propto e^{-\gamma z} \), at least in the right hand tail. Then for small \( y \), \( L(y) = (0, \log(y^{-1})/\gamma + \text{constant}) \). Therefore \( Q'(y)y^{1 + \frac{1}{\alpha}} = -y^{\frac{1}{\alpha}} \) which is non-increasing for all values of \( \alpha \) (because of the minus sign).
(ii) Again suppose $\mathcal{X} = \mathbb{R}^+$ and that $\pi$ is continuous and positive. Now suppose that $\pi \propto x^{-\delta}$, at least in the right hand tail. For small $y$, \( L(y) = (0, y^{-\delta-1} \times \text{constant}) \), $Q'(y)y^{1+\frac{1}{\alpha}} \propto y^{\alpha-1-\delta-1}$, so the condition holds for $\alpha \leq \delta$.

(2) The existence of the derivative of $Q$ has been assumed in this theorem. This condition can certainly be weakened slightly by expressing the key condition on $Q'(y)y^{1+\frac{1}{\alpha}}$ in terms of a suitable Radon-Nikodym derivative for the measure $R$ defined by $R((a, b]) = Q(a) - Q(b)$.

(3) The condition $\beta < 1/\alpha$ will be slightly restrictive for us in Section 4, when we consider quantitative bounds. Indeed, for exponentially-decreasing densities $\pi$ we have that $Q'(y)y^{1+\frac{1}{\alpha}}$ is non-increasing for any $\alpha > 0$, however Proposition 7 unfortunately does not allow us to use $\alpha$ larger than $1/\beta$. Now, it is possible to get around this restriction in the proof of that proposition; for example, if $\alpha \beta = M \in \mathbb{N}$ and $Y = 1$, then we can instead compute the integral $\int_0^1 \frac{y^{-\alpha-1} (y^{-\beta} - y^{-\beta})}{y^{-\alpha-1} - y^{-\alpha-1}} dy$ exactly, by recalling that $\frac{y^{-\beta-1}}{y^{-\alpha-1}} = 1 + y^{-\alpha-1} + y^{-2\alpha-1} + \ldots + y^{-(M-1)\alpha-1}$. It is not difficult to carry out these calculations; however, they do not appear to substantially improve the quantitative bounds that we study in Section 4. Therefore, we do not pursue this idea further.

Finally, we consider the case where $\pi(\cdot)$ is unbounded. In this case, we have $Q(y) > 0$ for arbitrarily large values of $y$, and it is important how quickly $Q(y) \to 0$ as $y \to \infty$. To examine this, we consider the function $Q^{-1}(w) \equiv \inf\{y > 0; Q(y) \geq w\}$. By applying Proposition 7 twice, we obtain the following.

**Theorem 9.** Consider the slice sampler $P_{\text{sel}}$ on a density $\pi$. Suppose $\pi$ is unbounded with infinite support, but that there exists a constant $\alpha > 1$ such that $Q'(y)y^{1+\frac{1}{\alpha}}$ is non-increasing for $y$ in an open set containing 0, and furthermore that $(Q^{-1})'(w)w^{1+\frac{1}{\alpha}}$ is non-increasing for $w$ in an open set containing 0. Then $P_{\text{sel}}$ is geometrically ergodic.

**Proof.** It is no longer true that $L(y)$ is small for any $y$, though by Proposition 5, sets on which $\pi$ is bounded above and away from zero are still small. Geometric excursions into either tail ($\pi(X)$ close to 0 or $\infty$) are now possible. The tail $\pi(X) \approx 0$ can be dealt with as in Proposition 7, and an identical calculation deals with the tail $\pi(x) \approx 0$ (using
drift function $Q(\pi(X))^{\beta}$). Therefore, by using Proposition 7 twice, we see that $P_{\text{ssl}}$ has geometric drift away from any fixed neighbourhood of $X = \infty$ and also away from any fixed neighbourhood of $X = 0$. The result now follows similarly to Theorem 8.

4. Quantitative convergence bounds.

In this section we consider quantitative bounds on the convergence of $P_{\text{ssl}}$ to its stationary distribution $\nu_{\pi}(\cdot)$. We recall that we have verified minorisation and drift conditions in the previous section. We further recall that we have verified that $P_{\text{ssl}}$ is stochastically monotone (Proposition 3). From these ingredients, there are well-known quantitative bounds on the distance of $\mathcal{L}(X_n)$ to stationarity. For optimal results, we use the following recent result of Roberts and Tweedie (1997), which builds on the analysis in Rosenthal (1995) and Lund and Tweedie (1996). For notation, we write $\mathbb{E}_{\nu_{\pi} \wedge \delta_{\pi}}(V)$ for the expected value of $V$ under the stochastic minorant (with respect to the ordering $\preceq$) of the stationary distribution $\nu_{\pi}(\cdot)$ and the point mass $\delta_{\pi}(\cdot)$. That is,

$$\mathbb{E}_{\nu_{\pi} \wedge \delta_{\pi}}(V) = V(x)\nu_{\pi}\{x \preceq x\} + \mathbb{E}_{\nu_{\pi}}(V1_{\{x \leq x\}}),$$

so that e.g. (using Meyn and Tweedie, 1993, Proposition 4.3 (i) for the final inequality)

$$\max(\mathbb{E}_{\pi}(V), V(x)) \leq \mathbb{E}_{\nu_{\pi} \wedge \delta_{\pi}}(V) \leq V(x) + \mathbb{E}_{\pi}(V) \leq V(x) + \frac{b}{1 - \lambda}.$$

**Theorem 10.** Consider the slice sampler $P_{\text{ssl}}$ on a density $\pi \leq 1$. Set $V(x) = \pi(x)^{-\beta}$. Then for $n \log(\lambda^{-1}) > \log(\mathbb{E}_{\nu_{\pi} \wedge \delta_{\pi}}(V))$, we have

$$\|P_{\text{ssl}}^n(x, \cdot) - \pi(\cdot)\| \equiv \sup_{A \subseteq \mathbb{R}^d} |P_{\text{ssl}}^n(x, A) - \pi(A)| \leq K(n + \eta - \xi)\rho^n.$$

Here

$$K = \frac{e\epsilon(1 - \epsilon)^{-\xi/\eta}}{\eta},$$

$$\xi = \frac{\log(\mathbb{E}_{\nu_{\pi} \wedge \delta_{\pi}}(V))}{\log(\lambda^{-1})}, \quad \eta = \frac{\log\left(\frac{\lambda \sigma + b - \xi}{\lambda(1 - \epsilon)}\right)}{\log(\lambda^{-1})},$$

$$s = \sup\{V(z); \pi(z) \geq \pi(x)\}$$

and $\rho = (1 - \epsilon)^{\sigma^{-1}}$, where the values of $\epsilon$, $\lambda$, and $b$ are as in Propositions 5 and 7.
Proof. The result follows immediately from Roberts and Tweedie (1997), in light of Proposition 3.

Example. Let \( \pi(x) = e^{-x} 1_{x > 0} \) be the density of the exponential distribution \( \text{Exp}(1) \). We can take \( \alpha \) as large as we like (provided that \( \alpha \beta \leq 1 \)), and can set \( Y = 1 \). Now suppose for illustration that \( \mathbb{E}_{Y^+ \wedge \delta} (V) = 3 \), and that we choose \( \beta = 0.1, \alpha = 1/\beta = 10 \), and \( \epsilon = y_* = 0.1 \). Then from Proposition 7, we have \( \lambda = 0.9527 \) and \( b = 0.10284 \) (so that \( b/(1-\lambda) = 2.175 \)). The bound of Theorem 10 then applies for \( n \geq 23 \) and gives

\[
\|P^n_{sst}(x, \cdot) - \pi(\cdot)\| \leq 0.0324 (0.9897)^n (n - 7.79).
\]

For example, with \( n = 530 \), we obtain

\[
\|P^{530}_{sst}(x, \cdot) - \pi(\cdot)\| < 0.0095.
\]

Hence, for this example, just 530 iterations suffices to make the total variation distance to stationarity provably less than 1% (a convergence criterion suggested in Cowles and Rosenthal, 1996).

Now, it follows immediately from Proposition 2 that this same bound applies when \( \pi(x) = ae^{-ax} \) is the density of the exponential distribution \( \text{Exp}(a) \) for any \( a > 0 \), not just for \( a = 1 \). However, it is surprising that this same bound applies to any density \( \pi \) such that \( yQ'(y) \) is non-increasing, as the following theorem shows. We give the result under the same conditions on initial conditions as in the previous example. Analogous results are clearly possible for all different initial conditions.

Theorem 11. Let \( \pi \) be a bounded density such that its corresponding function \( Q(y) = m(L(y)) \) is differentiable, and satisfies that \( Q'(y)y \) is non-increasing as a function of \( y > 0 \). Assume as in the previous example that \( \mathbb{E}_{Y^+ \wedge \delta} (V) = 3 \). Then the simple slice sampler algorithm for \( \pi \) satisfies

\[
\|P^n_{sst}(x, \cdot) - \pi(\cdot)\| \leq 0.0324 (0.9897)^n (n - 7.79), \quad n \geq 23.
\]
Proof. By renormalising if necessary, we can (and do) assume that \( \pi \leq 1 \). The proof shall proceed by comparing the slice sampler for \( \pi \), i.e. \( P_{\text{sl}} \), to the slice sampler for the \( \text{Exp}(1) \) distribution (as studied in the above example). To that end, let \( y_* \), \( \beta \), and the function \( V(\cdot) \) be exactly as in the \( \text{Exp}(1) \) example above. By Proposition 5, the set \( L(y_*) \) is small for \( P_{\text{sl}} \), with \( \epsilon = y_* \). We will be done if we can show that the drift equation
\[
P_{\text{sl}} V(x) \leq \lambda V(x) + b 1_{L(y_*)}(x)
\]
is satisfied by \( P_{\text{sl}} \), for the same values of \( \lambda \) and \( b \) as in the example.

To that end, recall the definition \( Q^{-1}(w) \equiv \inf\{y > 0; Q(y) \geq w\} \). Now, given a value of \( z = X_n \), set \( c = Q^{-1}(z) \), and choose \( a > 0 \) so that the exponential density \( ae^{-aw} \) is exactly tangent to the graph of \( Q^{-1}(w) \), at the point \( w = c \) (i.e., so that \( ae^{-ac} = Q^{-1}(c) \) and \( -a^2 e^{-ac} = (Q^{-1})'(c) \)). Then, since \( yQ'(y) \) is non-increasing for \( \pi \), but is constant for \( \text{Exp}(a) \), we see that the graph of \( ae^{aw} \) lies below \( Q^{-1}(w) \) for \( w > c \), but lies above it for \( w < c \). It follows that \( P_{\text{sl}} V(z) \leq PV(z) \), where \( P \) is the simple slice sampler for \( \text{Exp}(a) \). On the other hand, from Proposition 2, the behaviour of \( P \) does not depend on the value of \( a \). Hence, from Proposition 7 and the above example, we know that \( PV(z) \leq \lambda V(z) + b 1_{L(y_*)}(z) \) for the given values of \( \lambda \) and \( b \). Hence, \( P_{\text{sl}} V(z) \leq \lambda V(z) + b 1_{L(y_*)}(z) \) as desired; that is, the values of \( \epsilon, \lambda, \) and \( b \) are the same for \( P_{\text{sl}} \) as for the \( \text{Exp}(1) \) example. The result now follows from Theorem 10, exactly like for the \( \text{Exp}(1) \) example above.

\[\square\]

This theorem leads to the question of what densities \( \pi \) give rise to functions \( Q(y) \) such that \( Q'(y)y^{1+\frac{1}{2}} \) is non-increasing, for some \( \alpha > 1 \). Note that, since \( Q'(y) < 0 \), if \( yQ'(y) \) is itself non-increasing then this condition is satisfied for every \( \alpha > 1 \).

Observe that if \( Q^{-1}(w) \) is a (one-dimensional) log-concave function, then it is easily checked that \( yQ'(y) \) is in fact non-increasing. Indeed, this follows since \( \frac{d}{dw} \log Q^{-1}(w) \) equals the reciprocal of \( Q'(y)y \) evaluated at \( y = Q^{-1}(w) \). Hence, if the former is non-increasing as a function of \( w \), then the latter is non-decreasing as a function of \( w \) and therefore is non-increasing as a function of \( y \) (since \( Q' < 0 \)).

However connections between the condition and more familiar Euclidean concepts are
more complicated in higher dimensions. We give a condition which relates properties of \( \tau \) along one-dimensional rays from its mode, to the condition on \( yQ'(y) \).

We assume without loss of generality that \( \pi \) has its mode at the origin. We let \( S = \{ x \in \mathbb{R}^d; \| x \| = 1 \} \) be the usual \( L^2 \) unit \((d-1)\)-sphere in \( \mathbb{R}^d \). For \( \theta \in S \) and \( y > 0 \), we let \( D(y; \theta) = \sup \{ t > 0; \pi(t\theta) \leq y \} \). Note that the condition we impose in Proposition 12 is stronger than log-concavity, so that the uniqueness of the mode of \( \pi \) is guaranteed.

**Proposition 12.** Let \( \pi \) be a \( d \)-dimensional density such that for all \( \theta \in S \),

\[
yD(y; \theta)^{d-1} \frac{\partial}{\partial y} D(y; \theta) \text{ is a non-increasing function of } y > 0.
\]

(1)

Then the corresponding \( Q \)-function satisfies that \( yQ'(y) \) is non-increasing.

**Proof.** We can write

\[
Q(y) = \int_S D(y; \theta)^d d\theta,
\]

where \( d\theta \) is \((d-1)\)-dimensional Lebesgue measure on the (curved) space \( S \).

For \( d \geq 2 \), we note that

\[
yQ'(y) = \int_S yD(y; \theta)^{d-1} \frac{\partial}{\partial y} D(y; \theta) d\theta.
\]

(Here the differentiation under the integral sign is justified by e.g. Folland, 1984, Theorem 2.27.) The result follows by the condition imposed on the integrand.

\[ \blacksquare \]

**Remarks.**

1. In one dimension, (1) is weaker than log-concavity. However this is not the case when \( d \geq 2 \). On the other hand, it can be shown that for all log-concave densities, and for all choices of \( \alpha > 0 \), there exists a compact (and therefore small) set outside of which \( \pi \) satisfies that \( y^{1+\frac{1}{\alpha}} Q'(y) \) is a non-decreasing function. The results of Proposition 7 and Theorem 8 therefore apply, and moreover quantitative results analogous to Theorem 11 are available.
2. Since the function $Q$ completely specifies the slice sampler, and since $Q$ is unaffected by isometries, it suffices that $\pi$ be isometric to a function satisfying (1). That is, it suffices that there exists a mapping $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$, which preserves $d$-dimensional Lebesgue measure, such that $\pi \circ T$ satisfies (1).

Putting the previous two results together (and allowing for isometries as in the previous remark), we obtain finally the following.

**Theorem 13.** Suppose $\pi$ is a $d$-dimensional density which is (isometric to) a function satisfying condition (1) above. Let $P_{ssl}$ be the corresponding simple slice sampler for $\pi$. Then $P_{ssl}$ is geometrically ergodic, and in fact

$$
\|P_{ssl}^n(x, \cdot) - \pi(\cdot)\| \leq 0.0324 (0.9897)^n(n - 7.79) , \quad n \geq 23 .
$$

In particular, this Theorem shows that for any density $\pi$ satisfying (1), we have that:

$$
\|P_{ssl}^{530}(x, \cdot) - \pi(\cdot)\| < 0.0095 , \text{ i.e. that the simple slice sampler converges after 530 iterations.}
$$

5. Product slice samplers.

In this section, we shall investigate the geometric ergodicity of product slice samplers. Suppose $\pi(x) = f_1(x)f_2(x)\ldots f_k(x)$. Recall that the product slice sampler $P_{psl}$ on $(X, Y_1, Y_2, \ldots, Y_k) \in \mathbb{R}^d \times \mathbb{R} \times \ldots \times \mathbb{R}$ proceeds, given $X_n$, by updating $Y_{n+1, i} \sim \mathcal{U}(0, f_i(X_n))$ for $1 \leq i \leq k$ conditionally independently, and then updating $X_{n+1} \sim \mathcal{U}(L(Y))$, where $L(Y) = L(Y_1; f_1) \cap \ldots \cap L(Y_k; f_k)$ (here $L(y; f) = \{x \in \mathbb{R}^d; f(x) \geq y\}$).

We let $Q(y)$ denote $m(L(y))$, where $m$ is $d$-dimensional Lebesgue measure.

Before we give our first result about geometric ergodicity of the product slice sampler, we need the following lemma. The hypothesis of this lemma states, roughly, that all of the functions $f_i$ are decreasing in the same direction.

**Lemma 14.** Suppose there exists $Y > 0$ such that for all $x_1$ and $x_2$ such that $f_1(x_1) \leq f_1(x_2) \leq Y$, we have

$$
 f_i(x_1) \leq f_i(x_2) , \quad 1 \leq i \leq k . \quad (2)
$$

Then for $x$ and $y$ such that $f_i(x) \geq y_i$ for $i = 1, \ldots, k$, we have that

$$
 B \equiv L(Y; f_1)^c \cap \{z; f_1(x) \leq f_1(z)\} \subset L(y) .
$$
Proof. Let $x$ and $y$ be such that $f_i(x) \geq y_i$ for $1 \leq i \leq k$, and suppose $z \in B$. Then $f_1(x) \leq f_1(z) \leq Y$, and so by (2), $f_i(x) \leq f_i(z)$ and therefore $y_i \leq f_i(z)$ at least for $i \neq 1$. Since we also have $z \in L(y_1; f_1)$, we must have $f_1(z) \geq y_1$ also. Hence $z \in L(y)$, as required.

We now prove a result about the geometric ergodicity of slice samplers. Like the lemma, it requires that the functions $f_i$ all be decreasing in the same direction.

**Theorem 15.** Suppose that for each $i$, $f_i$ is bounded. Set $Q_1(y) = m(L(y; f_1))$, and suppose that $Q_1$ is differentiable with $Q_1'(y)y^{1+\alpha^{-1}}$ non-increasing, at least in some open set containing 0. Finally, suppose that the hypothesis of the previous lemma holds for the functions $\{f_i\}$. Then the produce slice sampler $P_{x,y}$ is geometrically ergodic.

Proof. We shall assume without loss of generality that we take $Y$ small enough so that $Q_1'(y)y^{1+\alpha^{-1}}$ is non-increasing on $(0, Y)$. Set $V(x) = f_1(x)^{-\beta}$. Choose $x$ such that $f_1(x) < Y$. Then

$$PV(x) = \frac{1}{\prod_{i=1}^k f_i(x)} \int_0^{f_1(x)} \cdots \int_0^{f_k(x)} \frac{1}{Q(y)} \int_{L(y)} f_1(z)^{-\beta} dz \ dy.$$ 

Now partition $L(y) = A(y) \cup B \cup C(y)$, where $B$ is defined in Lemma 5, $A(y) = L(y) \cap L(Y, f_1)$, and $C(y) = L(y) \cap L(Y, f_1)^c \cap \{z; f_1(x) \geq f_1(z)\}$. Therefore we can write

$$PV(x) = \frac{1}{\prod_{i=1}^k f_i(x)} \times$$

$$\int_0^{f_1(x)} \cdots \int_0^{f_k(x)} \left( \frac{m(A(y))}{m(L(y))} E(A(y)) + \frac{m(B(y_1))}{m(L(y))} E(B(y_1)) + \frac{m(C(y))}{m(L(y))} E(C(y)) \right) dy,$$

where $E(S) = E[f_1^{-\beta}(Z)|Z \sim \mathcal{U}(S)]$ for any set $S \in \sigma(X)$.

Now define $D(y_1) = \{z; f_1(x) \geq f_1(z)\}$, so that $C(y) \subseteq D(y_1)$. Moreover, it is clear
that $C(y) = \{ z, c(y) \leq f_1(z) \leq f_1(x) \}$ for a function $c(\cdot)$. Therefore,

$$PV(x) \leq Y^{-\beta} + \frac{1}{\prod_{i=1}^{k} f_i(x)} \int_{0}^{\int_{0}^{f_k(x)}} \cdots \int_{0}^{f_k(x)} E(B \cup C(y)) dy$$

$$\leq Y^{-\beta} \left( \frac{1}{\prod_{i=1}^{k} f_i(x)} \int_{0}^{f_k(x)} \cdots \int_{0}^{f_k(x)} E(B \cup D(y_1)) dy \right)$$

$$\leq Y^{-\beta} \left( \frac{V(x)}{(1-\beta)(1+\alpha \beta)} \right)$$

as required. Here the second inequality follows from Lemma 14, and the third inequality follows from Proposition 7.

Finally, we consider product slice samplers whose component functions $f_i$ are not all decreasing in the same direction. For simplicity, we restrict ourselves to dimension $d = 1$, and to a number of component functions $k = 2$ which are decreasing in opposite directions.

Specifically, let $X \subset \mathbb{R}$, and $\pi(x) = f_1(x)f_2(x)$, where $f_1$ is a non-decreasing function and $f_2$ is a non-increasing function. Then we shall call this special form of the product slice sampler the opposite monotone sampler with transitions $P_{oms}$. We shall assume that $f_1$ and $f_2$ are invertible, so that we can write $P_{oms}$ as follows. Given $X_n$, sample $Y_{n+1,i}$ from $\mathcal{U}(0, f_i(X_n))$ conditionally independently for $i = 1, 2$. $X_{n+1}$ is then sampled from $\mathcal{U}(f_1^{-1}(Y_{n+1,1}), f_2^{-1}(Y_{n+1,2})).$

Although in general the product slice sampler is not stochastically monotone, $P_{oms}$ regains monotonicity properties from the total-orderedness of $\mathbb{R}$. Specifically we have

**Proposition 16.** $P_{oms}$ is stochastically monotone with respect to the usual ordering on $\mathbb{R}$.

**Proof.** Given arbitrary $x_1 \leq x_2$, it is enough to show that there is a joint probability construction of two processes, one started at each of $x_1$ and $x_2$, which almost surely preserves their order. However, given $U_1$, $U_2$ and $U_3$, all independently $\mathcal{U}(0,1)$, we can produce the construction as follows. Start the two processes off at $X_0^j = x_j, j = 1, 2$. Let $Y_i^j = f_i(x_j)U_i, i, j = 1, 2$ (so that $j$ indexes the two processes, and $i$ continues to
index the auxiliary variables). Now set \(X^i_1 = f^{-1}_i(Y^i_1) + (f^{-1}_2(Y^i_2) - f^{-1}_1(Y^i_2))U_3\). Now by the respective monotonicity of \(f_1\) and \(f_2\) it follows that \(f^{-1}_i(Y^i_1) \leq f^{-1}_i(Y^i_2), i = 1, 2\).
Therefore \(X^1_1 \leq X^2_1\) and so the result follows.

We turn now to the problem of proving the geometric ergodicity of \(P_{oms}\). The interesting case for \(P_{oms}\) is the case where one (or both) the functions \(f_1, f_2\) are unbounded, though \(\pi\) is still bounded. The case of bounded \(f_i\) is virtually identical to the case of \(P_{ast}\) and we omit any formal statement of the result except that a very weak decay condition on the \(f_i\)'s will be needed as in Proposition 7. Instead we shall deal with the case where both \(f_1\) and \(f_2\) are unbounded and non-zero.

**Theorem 17.** Suppose \(X\) is a (possibly infinite) interval, \((\mathcal{X}_-, \mathcal{X}_+) \subset \mathbb{R}\), and that \(f_1\) and \(f_2\) are unbounded and non-zero, with \(f_1\) non-decreasing and \(f_2\) non-increasing. Let \(\beta\) be a positive constant such that \(f^\beta_1\) and \(f^\beta_2\) are convex functions. Suppose there exists \(0 < \gamma < (1 + 2\beta)^{-1}\) such that \(u^\gamma f_1 f_2^{-1}(u)\) and \(u^\gamma f_2 f_1^{-1}(u)\) are both non-decreasing functions for \(u\) in some neighbourhood of 0. Then \(P_{oms}\) is geometrically ergodic.

**Proof.** Let \(V_1(x) = f_1(x)^\beta\). Suppose \(k_1\) is such that \(f_1 f_2^{-1}(u)\) is non-decreasing for \(u \leq f_2(k_1)\). Then for \(x \geq k_1\),

\[
P_{oms} V_1(x) = \frac{1}{\pi(x)} \int_0^1 \int_0^1 \int_0^1 \int_0^{f_2^{-1}(y_2)} V_1(z) \, dz \, dy_1 \, dy_2
\]

\[
= \int_0^1 \int_0^1 \int_0^1 \int_0^{f_2^{-1}(f_2(x)u_2) - f_1^{-1}(f_1(x)u_1)} f_1(z)^\beta \, dz \, du_1 \, du_2
\]

\[
\leq \frac{1}{2} \int_0^1 \int_0^1 \left[ f_1(x)^\beta u_2 + (f_1 f_2^{-1}(u_2 f_2(x)))^\beta \right] du_1 \, du_2
\]

\[
= \frac{f_1(x)}{2(1 + \beta)} + \int_0^1 (f_1 f_2^{-1}(u_2 f_2(x)))^\beta \, du_2,
\]

the inequality following from the convexity condition on \(f^\beta_1\). Now,

\[
f_1(f_2^{-1}(v)) \leq (f_2(x))^{\alpha} f_1(x)/v^{\alpha}
\]

18
for \( v \leq f_2(x) \), so that the second term may be bounded by

\[
\frac{1}{2} \int_0^1 \frac{f_1(x)^\beta}{u^{\beta \gamma}} \, du = \frac{f_1(x)^\beta}{2(1-\beta \gamma)}.
\]

Hence, for \( x \geq k_1 \),

\[
P_{\text{oms}} V_1(x) \leq \frac{V_1(x)}{2} \left( \frac{1}{1+\beta} + \frac{1}{1-\beta \gamma} \right) = \lambda V_1(x)
\]
say, where \( \lambda < 1 \). Furthermore, by stochastic monotonicity, \( P_{\text{oms}} V_1(x) \leq \lambda V_1(k_1) \) for \( x \leq k_1 \).

Similarly we can prove that if \( V_2(x) = f_2(x)^\beta \), there exists \( k_2 \) such that \( P_{\text{oms}} V_2(x) \leq \lambda V_2(x) \) for \( x \leq k_2 \) with \( P_{\text{oms}} V_2(x) \leq \lambda V_2(k_2) \) for \( x \geq k_2 \).

Geometric drift now follows with drift function \( V(x) = V_1(x) + V_2(x) \). Indeed, from the above bounds on \( P_{\text{oms}} V_1(x) \) and \( P_{\text{oms}} V_2(x) \), it follows that for large enough \( M > 0 \), we will have \( P_{\text{oms}} V(x) \leq \lambda' V(x) \) whenever \( |x| > M \), for some \( \lambda' < 1 \). Furthermore the set \([-M, M]\) is easily seen to be small for \( P_{\text{oms}} \). Hence, the result follows just as in Theorem 8.

Unfortunately, although \( P_{\text{oms}} \) is stochastically monotone, it is not possible to calculate bounds on convergence using Theorem 10 since it is not true that either \((-\infty, x)\) or \((x, \infty)\) are small for any \( x \). Computable bounds are still possible (see Roberts and Tweedie, 1997) but will not be as tight as those in Theorem 11.

6. Discussion and conclusions.

In this paper, we have studied theoretical properties of slice samplers. We have shown that under some rather general hypotheses, these samplers have some very nice convergence properties.

In particular, we have proved geometric ergodicity for all simple slice samplers on densities with asymptotically polynomial tails. This covers virtually all distributions of interest. We have also extended this result to product slice samplers, albeit under more restrictive conditions.
We have also proved quantitative bounds on the convergence of these samplers, for certain classes of densities. In particular, for all multi-dimensional densities satisfying our condition (1) herein, which includes all one-dimensional log-concave densities, we have established a uniform bound of 530 iterations required to achieve 1% accuracy in total variation distance. Previous rigorous quantitative bounds for MCMC samplers have generally been established only for very specific models (Meyn and Tweedie, 1994; Rosenthal, 1995) or have involved large undetermined constants (Polson, 1996). Indeed, we know of no comparable result which gives a reasonable uniform bound on the convergence rate of a realistic sampling algorithm, over such a broad class of distributions.

Of course, it may not always be easy to implement a slice sampler for a particular problem. For example, the sets $L(y)$ and the measures $Q(y)$ may be difficult or impossible to compute. However, the results of this paper suggest that, if it is possible to run a slice sampler algorithm on a given density, then the sampler will probably have excellent convergence properties.

Acknowledgements. We thank Radford Neal for introducing us to slice samplers.

REFERENCES


