8

The Slice Sampler

> He'd heard stories that shredded documents could be reconstructed. All it took was patience: colossal patience.
> —Ian Rankin, *Let it Bleed*

While many of the MCMC algorithms presented in the previous chapter are both generic and universal, there exists a special class of MCMC algorithms that are more model dependent in that they exploit the local conditional features of the distributions to simulate. Before starting the general description of such algorithms, gathered under the (somewhat inappropriate) name of *Gibbs sampling*, we provide in this chapter a simpler introduction to these special kind of MCMC algorithms. We reconsider the *fundamental theorem of simulation* (Theorem 2.15) in light of the possibilities opened by MCMC methodology and construct the corresponding *slice sampler*.

The previous chapter developed simulation techniques that could be called "generic," since they require only a limited amount of information about the distribution to be simulated. For example, the generic algorithm ARMS (Note 7.4.2) aims at reproducing the density f of this distribution in an automatic manner where only the numerical value of f at given points matters. However, Metropolis–Hastings algorithms can achieve higher levels of efficiency when they take into account the specifics of the target distribution f, in particular through the calibration of the acceptance rate (see Section 7.6.1). Moving even further in this direction, the properties and performance of the method presented in this chapter are closely tied to the distribution f.

8.1 Another Look at the Fundamental Theorem

Recall from Section 2.3.1 that the generation from a distribution with density $f(x)$ is equivalent to uniform generation on the subgraph of f,

$$\mathscr{S}(f) = \{(x,u);\, 0 \le u \le f(x)\}\ ,$$

whatever the dimension of x, and f need only be known up to a normalizing constant.

From the development at the beginning of Chapter 7, we can consider the possibility of using a Markov chain with stationary distribution equal to this uniform distribution on $\mathscr{S}(f)$ as an approximate way to simulate from f. A natural solution is to use a *random walk* on $\mathscr{S}(f)$, since a random walk on a set $\mathscr{A}$ usually results in a stationary distribution that is the uniform distribution on $\mathscr{A}$ (see Examples 6.39, 6.40, and 6.73).

There are many ways of implementing a random walk on this set, but a natural solution is to go one direction at a time, that is, to move iteratively along the u-axis and then along the x-axis. Furthermore, we can use uniform moves on both directions, since, as formally shown below, the associated Markov chain on $\mathscr{S}(f)$ does not require a Metropolis–Hastings correction to have the uniform distribution on $\mathscr{S}(f)$ as stationary distribution. Starting from a point (x,u) in $\{(x,u) : 0 < u < f(x)\}$, the move along the u-axis will correspond to the conditional distribution

$$U|X = x \sim \mathcal{U}(\{u : u \le f(x)\})\,, \tag{8.1}$$

resulting in a change from point (x,u) to point (x,u'), still in $\mathscr{S}(f)$, and then the move along the x-axis to the conditional distribution

$$X|U = u' \sim \mathcal{U}(\{x : u' \le f(x)\}), \tag{8.2}$$

resulting in a change from point (x,u') to point (x',u').

This set of proposals is the basis chosen for the original *slice sampler* of Neal (1997) (published as Neal 2003) and Damien et al. (1999), which thus uses a 2-step uniform random walk over the subgraph. We inaccurately call it the *2D slice sampler* to distinguish it from the general slice sampler defined in Section 8.2, even though the dimension of x is arbitrary.

Algorithm A.31 –2D slice sampler–

At iteration t, simulate

1. $u^{(t+1)} \sim \mathcal{U}_{[0,f(x^{(t)})]}$; [A.31]
2. $x^{(t+1)} \sim \mathcal{U}_{A^{(t+1)}}$, with

$$A^{(t+1)} = \{x :\ f(x) \ge u^{(t+1)}\}.$$

From (8.1) it is also clear that $x^{(t)}$ is always part of the set $A^{(t+1)}$, which is thus nonempty. Moreover, the algorithm remains valid if $f(x) = Cf_1(x)$, and we use f_1 instead of f (see Problem 8.1). This is quite advantageous in settings where f is an unnormalized density like a posterior density.

As we will see in more generality in Chapter 10, the validity of [A.31] as an MCMC algorithm associated with f_1 stems from the fact that both steps 1. and 2. in [A.31] successively preserve the uniform distribution on the subgraph of f: first, if $x^{(t)} \sim f(x)$ and $u^{(t+1)} \sim \mathcal{U}_{[0,f_1(x^{(t)})]}$, then

$$(x^{(t)}, u^{(t+1)}) \sim f(x) \frac{\mathbb{I}_{[0,f_1(x)]}(u)}{f_1(x)} \propto \mathbb{I}_{0 \le u \le f_1(x)} .$$

Second, if $x^{(t+1)} \sim \mathcal{U}_{A^{(t+1)}}$, then

$$(x^{(t)}, u^{(t+1)}, x^{(t+1)}) \sim f(x^{(t)}) \frac{\mathbb{I}_{[0,f_1(x^{(t)})]}(u^{(t+1)})}{f_1(x^{(t)})} \frac{\mathbb{I}_{A^{(t+1)}}(x^{(t+1)})}{\text{mes}(A^{(t+1)})} ,$$

where $\text{mes}(A^{(t+1)})$ denotes the (generally Lebesgue) measure of the set $A^{(t+1)}$. Thus

$$\begin{aligned}(x^{(t+1)}, u^{(t+1)}) &\sim C \int \mathbb{I}_{0 \le u \le f_1(x)} \frac{\mathbb{I}_{f_1(x^{(t+1)}) \ge u}}{\text{mes}(A^{(t+1)})} \, dx \\ &= C \mathbb{I}_{0 \le u \le f_1(x^{(t+1)})} \int \frac{\mathbb{I}_{u \le f_1(x)}}{\text{mes}(A^{(t+1)})} \, dx \\ &\propto \mathbb{I}_{0 \le u \le f_1(x^{(t+1)})}\end{aligned}$$

and the uniform distribution on $\mathcal{S}(f)$ is indeed stationary for both steps.

Example 8.1. Simple slice sampler. Consider the density $f(x) = \frac{1}{2} e^{-\sqrt{x}}$ for $x > 0$. While it can be directly simulated from (Problem 8.2), it also yields easily to the slice sampler. Indeed, applying (8.1) and (8.2), we have

$$U|x \sim \mathcal{U}\left(0, \frac{1}{2} e^{-\sqrt{x}}\right), \quad X|u \sim \mathcal{U}\left(0, [\log(2u)]^2\right) .$$

We implement the sampler to generate 50,000 variates, and plot them along with the density in Figure 8.1, which shows that the agreement is very good. The performances of the slice sampler may however deteriorate when $\sqrt{x}$ is replaced with $x^{1/d}$ for d large enough, as shown in Roberts and Tweedie (2004). ||

Example 8.2. Truncated normal distribution The distribution represented in Figure 8.2 is a truncated normal distribution $\mathcal{N}_-^+(3,1)$, restricted to the interval $[0,1]$,

$$f(x) \propto f_1(x) = \exp\{-(x+3)^2/2\} \, \mathbb{I}_{[0,1]}(x) .$$

As mentioned previously, the naïve simulation of a normal $\mathcal{N}(3,1)$ random variable until the outcome is in $[0,1]$ is suboptimal because there is only a 2% probability of this happening. However, devising and optimizing the

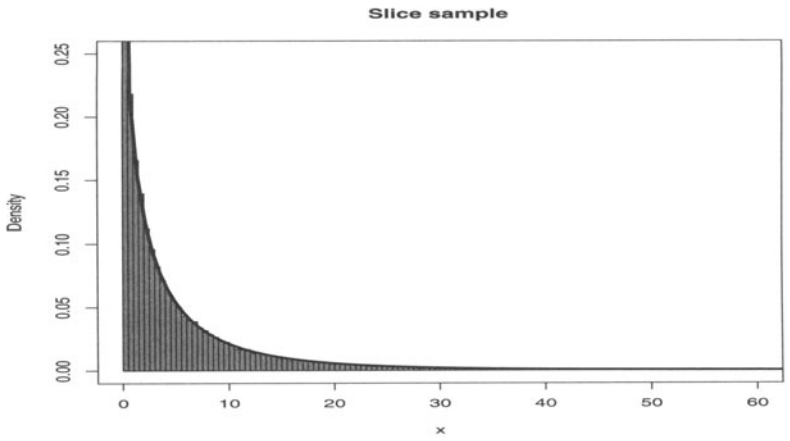

Fig. 8.1. The slice sampler histogram and density for Example 8.1.

Accept–Reject algorithm of Example 2.20 can be costly if the algorithm is to be used only a few times. The slice sampler [*A*.31] applied to this problem is then associated with the horizontal slice

$$A^{(t+1)} = \{y;\ \exp\{-(y+3)^2/2\} \geq u\, f_1(x^{(t)})\}\,.$$

If we denote $\omega^{(t)}$ the value $u\, f_1(x^{(t)})$, the slice is also given by

$$A^{(t+1)} = \{y \in [0,1];\ (y+3)^2 \leq -2\,\log(\omega^{(t)})\}\,,$$

which is an interval of the form $[0, \gamma^{(t)}]$. Figure 8.2 shows the first ten steps of [*A*.31] started from $x^{(0)} = 0.25$. ‖

This algorithm will work well only if the exploration of the subgraph of f_1 by the corresponding random walk is fast enough; if we take the above example of the truncated normal, this is the case. Whatever the value of $x^{(t)}$, the next value $x^{(t+1)}$ can be anything in $[0,1]$. This property actually holds for all f's: given that, when $\omega^{(t+1)}$ is close to 0, the set $A^{(t+1)}$ is close to the entire support of f and, formally, every value $x^{(t+1)}$ in the support of f can be simulated in one step. (We detail below some more advanced results about convergence properties of the slice sampler.)

Figure 8.3 illustrates the very limited dependence on the starting value for the truncated normal example: the three graphs represent a sample of values of $x^{(10)}$ when $(x^{(0)}, \omega^{(1)})$ is in the upper left hand corner, the lower right hand corner and the middle of the subgraph, respectively. To study the effect of the starting value, we always used the same sequence of uniforms for the three starting points. As can be seen from the plot, the samples are uniformly spread out over the subgraph of f and, more importantly, almost identical!

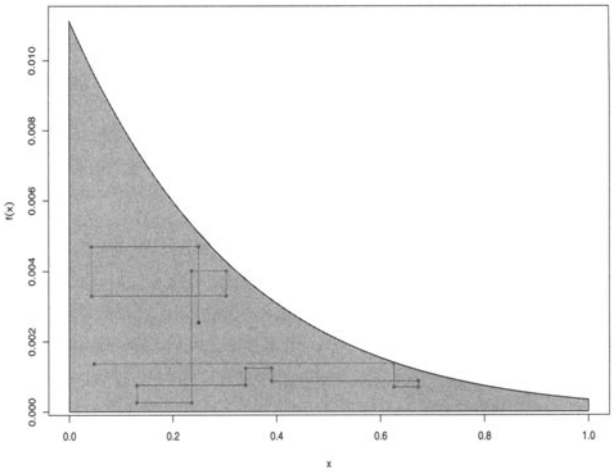

Fig. 8.2. First ten steps of the slice sampler for the truncated normal distribution $\mathcal{N}(3,1)$, restricted to the interval $[0,1]$.

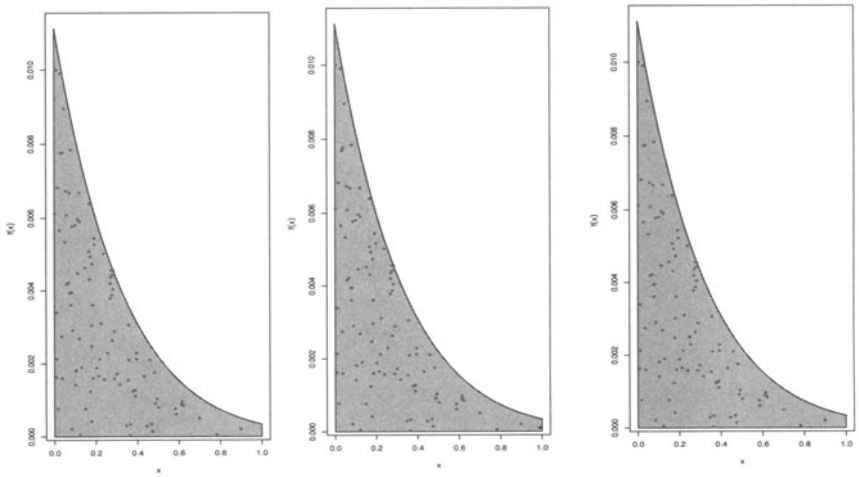

Fig. 8.3. Comparison of three samples obtained by ten iterations of the slice sampler starting from $(.01,.01)$ *(left)*, $(.99,.001)$ *(center)* and $(.25,.025)$ *(right)* and based on the same pool of uniform variates.

In this simple example, ten iterations are thus enough to cancel the effect of the starting value.

The major (practical) difficulty with the slice sampler is with the simulation of the uniform distribution $\mathcal{U}_{A^{(t+1)}}$, since the determination of the set of y's such that $f_1(y) \geq \omega$ can be intractable if f_1 is complex enough. We thus consider an extension to [*A*.31] in the next section, but call attention to an extension by Neal (2003) covered in Note 8.5.1.

8.2 The General Slice Sampler

As indicated above, sampling uniformly from the slice $A^{(t)} = \{x; f_1(x) \geq \omega^{(t)}\}$ may be completely intractable, even with the extension of Note 8.5.1. This difficulty persists as the dimension of x gets larger. However, there exists a generalization of the 2D slice sampler [*A*.31] that partially alleviates this difficulty by introducing multiple slices.

This general slice sampler can be traced back to the auxiliary variable algorithm of Edwards and Sokal (1988) applied to the Ising model of Example 5.8 (see also Swendson and Wang 1987, Wakefield et al. 1991, Besag and Green 1993, Damien and Walker 1996, Higdon 1996, Neal 1997, Damien et al. 1999, and Tierney and Mira 1999). It relies upon the decomposition of the density $f(x)$ as

$$f(x) \propto \prod_{i=1}^{k} f_i(x)\,,$$

where the f_i's are positive functions, but *not necessarily densities*. For instance, in a Bayesian framework with a flat prior, the $f_i(x)$ may be chosen as the individual likelihoods.

This decomposition can then be associated with k auxiliary variables ω_i, rather than one as in the fundamental theorem, in the sense that each $f_i(x)$ can be written as an integral

$$f_i(x) = \int \mathbb{I}_{0 \leq \omega_i \leq f_i(x)}\, d\omega_i\,,$$

and that f is the marginal distribution of the joint distribution

$$(x, \omega_1, \ldots, \omega_k) \sim p(x, \omega_1, \ldots, \omega_k) \propto \prod_{i=1}^{k} \mathbb{I}_{0 \leq \omega_i \leq f_i(x)}\,. \tag{8.3}$$

This particular *demarginalization* of f (Section 5.3.1) introduces a larger dimensionality to the problem and induces a generalization of the random walk of Section 8.1 which is to have uniform proposals one direction at a time. The corresponding generalization of [*A*.31] is thus as follows.

Algorithm A.32 –Slice Sampler–

At iteration $t+1$, simulate

1. $\omega_1^{(t+1)} \sim \mathcal{U}_{[0, f_1(x^{(t)})]}$;

$\vdots$

k. $\omega_k^{(t+1)} \sim \mathcal{U}_{[0, f_k(x^{(t)})]}$;

k+1. $x^{(t+1)} \sim \mathcal{U}_{A^{(t+1)}}$, with

$$A^{(t+1)} = \{y;\ f_i(y) \geq \omega_i^{(t+1)},\ i = 1, \ldots, k\}.$$

[*A*.32]

Example 8.3. A 3D slice sampler. Consider the density proportional to

$$(1+\sin^2(3x))\,(1+\cos^4(5x))\,\exp\{-x^2/2\}.$$

The corresponding functions are, for instance, $f_1(x) = (1+\sin^2(3x))$, $f_2(x) = (1+\cos^2(5x))$, and $f_3(x) = \exp\{-x^2/2\}$. In an iteration of the slice sampler, three uniform $\mathcal{U}([0,1])$ u_1, u_2, u_3 are generated and the new value of x is uniformly distributed over the set

$$\left\{x : |x| \le \sqrt{-2\log\omega_3}\right\} \cap \left\{x : \sin^2(3x) \ge 1-\omega_1\right\} \cap \left\{x : \cos^4(5x) \ge 1-\omega_2\right\},$$

which is made of one or several intervals depending on whether $\omega_1 = u_1 f_1(x)$ and $\omega_2 = u_2 f_2(x)$ are larger than 1. Figure 8.4 shows how the sample produced by the algorithm fits the target density. ‖

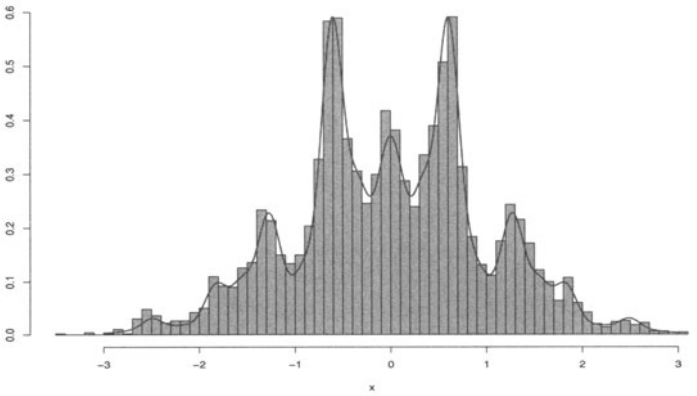

Fig. 8.4. Histogram of a sample produced by 5,000 iterations of a 3D slice sampler and superposition of the target density proportional to $(1+\sin^2(3x))\,(1+\cos^4(5x))\,\exp\{-x^2/2\}$.

In Section 5.3.1, the completion was justified by a *latent variable* representation of the model. Here, the auxiliary variables are more artificial, but there often is an obvious connection in (natural) latent variable models.

Example 8.4. Censored data models. As noted in Section 5.3.1, censored data models (Example 5.14) can be associated with a missing data structure. Consider

$$y^* = (y_1^*, y_2^*) = (y \wedge r, \mathbb{I}_{y<r}), \quad \text{with } y \sim f(y|\theta) \text{ and } r \sim h(r),$$

so the observation y is censored by the random variable r. If we observe $(y_1^*, \ldots, y_n^*)$, the density of $y_i^* = (y_{1i}^*, y_{2i}^*)$ is then

$$\int_{y_{1i}^*}^{+\infty} f(y|\theta)\, dy\, h(y_{1i}^*)\, \mathbb{I}_{y_{2i}^*=0} + \int_{y_{1i}^*}^{+\infty} h(r)\, dr\, f(y_{1i}^*|\theta)\, \mathbb{I}_{y_{2i}^*=1}\,. \tag{8.4}$$

In the cases where (8.4) cannot be explicitly integrated, the likelihood and posterior distribution associated with this model may be too complex to be used. If θ has the prior distribution π, the posterior distribution satisfies

$$\begin{aligned}\pi(\theta|y_1^*,\ldots,y_n^*) &\propto \pi(\theta) \prod_{\{i:y_{2i}^*=0\}} \left\{ h(y_{1i}^*) \int_{y_{1i}^*}^{+\infty} f(y|\theta)dy \right\} \\ &\quad \times \prod_{\{i:y_{2i}^*=1\}} \left\{ f(y_{1i}^*|\theta) \int_{y_{1i}^*}^{+\infty} h(r)dr \right\} \\ &\propto \pi(\theta) \prod_{\{i:y_{2i}^*=0\}} \int_{y_{1i}^*}^{+\infty} f(y|\theta)dy \prod_{\{i:y_{2i}^*=1\}} f(y_{1i}^*|\theta)\,.\end{aligned}$$

If the *uncensored* model leads to explicit computations for the posterior distribution of θ, we will see in the next chapter that a logical completion is to reconstruct the original data, $y_1,\ldots,y_n$, conditionally on the observed y_i^*'s $(i=1,\ldots,n)$, and then implement the Gibbs sampler on the two groups θ and the unobserved y's. At this level, we can push the demarginalization further to first represent the above posterior as the marginal of

$$\pi(\theta) \prod_{\{i:y_{2i}^*=0\}} f(y_{1i}|\theta)\, \mathbb{I}_{y_{1i}>y_{1i}^*} \prod_{\{i:y_{2i}^*=1\}} f(y_{1i}^*|\theta)$$

and write this distribution itself as the marginal of

$$\mathbb{I}_{0\le\omega_0\le\pi(\theta)} \prod_{\{i:y_{2i}^*=0\}} \left\{ \mathbb{I}_{y_{1i}>y_{1i}^*}\, \mathbb{I}_{0\le\omega_i\le f(y_{1i}|\theta)} \right\} \prod_{\{i:y_{2i}^*=1\}} \mathbb{I}_{0\le\omega_i\le f(y_{1i}^*|\theta)}\,,$$

adding to the unobserved y_{1i}'s the basic auxiliary variables ω_i $(0\le i\le n)$. ‖

Although the representation (8.3) of f takes us farther away from the fundamental theorem of simulation, the validation of this algorithm is the same as in Section 8.1: each of the $k+1$ steps in [*A*.32] preserves the distribution (8.3). This will also be the basis of the Gibbs sampler in Chapters 9 and 10.

While the basic appeal in using this generalization is that the set $A^{(t+1)}$ may be easier to compute as the intersection of the slices

$$A_i^{(t+1)} = \{y;\ f_i(y) \ge \omega_i^{(t+1)}\}\,,$$

there may still be implementation problems. As k increases, the determination of the set $A^{(t+1)}$ usually gets increasingly complex. Neal (2003) develops proposals in parallel to improve the slice sampler in this multidimensional

setting, using "over-relaxed" and "reflective slice sampling" procedures. But these proposals are too specialized to be discussed here and, also, they require careful calibration and cannot be expected to automatically handle all settings. Note also that, in missing variable models, the number of auxiliary variables (that is, of slices) increases with the number of observations and may create deadlocks for large data sizes.

8.3 Convergence Properties of the Slice Sampler

Although, in the following chapters, we will discuss in detail the convergence of the Gibbs sampler, of which the slice sampler is a special case, we present in this section some preliminary results. These are mostly due to Roberts and Rosenthal (1998) and and show that the convergence rate of the slice sampler can be evaluated quantitatively.

First, note that, for the 2D slice sampler [*A*.31], if we denote by $\mu(\omega)$ the Lebesgue measure of the set

$$\mathscr{S}(\omega) = \{y;\, f_1(y) \geq \omega\}\,,$$

the transition kernel is such that

$$\begin{aligned}\Pr\left(f_1(X^{(t+1)}) \leq \eta | f_1(X^{(t)}) = v\right) &= \int\int \mathbb{I}_{w \leq f_1(x) \leq \eta}\, \mathbb{I}_{0 \leq w \leq v}\, \frac{dw\, dx}{v\, \mu(w)} \\ &= \frac{1}{v} \int_0^{\eta \wedge v} \frac{\mu(w) - \mu(\eta)}{\mu(w)}\, dw \\ &= \frac{1}{v} \int_0^{v} \max\left(1 - \frac{\mu(\eta)}{\mu(w)}, 0\right) dw\,.\end{aligned}$$

The properties of the chain are thus entirely characterized by the measure μ and, moreover, they are equivalent for $(X^{(t+1)})$ and $(f_1(X^{(t+1)}))$ which also is a Markov chain in this case (Problems 8.9 and 8.10).

Under boundedness conditions, Tierney and Mira (1999) established the following uniform ergodicity result.

Lemma 8.5. *If* f_1 *is bounded and* $\text{supp} f_1$ *is bounded, the slice sampler* [*A*.31] *is uniformly ergodic.*

Proof. Without loss of generality, assume that f_1 is bounded by 1 and that the support $\text{supp} f_1$ is equal to $(0,1)$. To establish uniform ergodicity, it is necessary and sufficient to prove that Doeblin's condition (Theorem 6.59) holds, namely that the whole space $\text{supp} f_1$ is a small set. This follows from the fact that

$$\xi(v) = \Pr\left(f_1(X^{(t+1)}) \leq \eta | f_1(X^{(t)}) = v\right)$$

is decreasing in v for all η. In fact, when $v \geq \eta$,

$$\xi(v) = \frac{1}{v} \int_0^\eta \frac{\mu(w) - \mu(\eta)}{\mu(w)}\, dw$$

is clearly decreasing in v. When $v \leq \eta$, $\xi(v)$ is the average of the function

$$\left(1 - \frac{\mu(\eta)}{\mu(W)}\right)$$

when W is uniformly distributed over $(0, v)$. Since $\mu(\omega)$ is decreasing in ω, this average is equally decreasing in v. The maximum of $\xi(v)$ is thus

$$\lim_{v \to 0} \xi(v) = \lim_{v \to 0} \left(1 - \frac{\mu(\eta)}{\mu(v)}\right) = 1 - \mu(\eta)$$

by L'Hospital's rule, while the minimum of $\xi(v)$ is

$$\lim_{v \to 1} \xi(v) = \int_0^\eta \frac{\mu(w) - \mu(\eta)}{\mu(w)}\, dw\,.$$

Since we are able to find nondegenerate upper and lower bounds on the cdf of the transition kernel, we can derive Doeblin's condition and thus uniform ergodicity. □

In a more general setting, Roberts and Rosenthal (1998) exhibit a *small set* associated with the 2D slice sampler.

Lemma 8.6. *For any $\epsilon^\star > \epsilon_\star$, $\{x;\ \epsilon_\star < f_1(x) < \epsilon^\star\}$ is a small set, that is,*

$$P(x, \cdot) \geq \frac{\epsilon_\star}{\epsilon^\star}\, \nu(\cdot) \text{ where } \nu(A) = \frac{1}{\epsilon_\star} \int_0^{\epsilon_\star} \frac{\lambda(A \cap \{y; f_1(y) > \epsilon\})}{\mu(\epsilon)} d\,\epsilon$$

and λ denotes Lebesgue measure.

Proof. For $\epsilon_\star < x < \epsilon^\star$,

$$\begin{aligned} \Pr(x, A) &= \frac{1}{x} \int_0^x \frac{\lambda(A \cap \{y; f_1(y) > \epsilon\})}{\mu(\epsilon)}\, d\epsilon \\ &\geq \frac{1}{\epsilon^\star} \int_0^{\epsilon_\star} \frac{\lambda(A \cap \{y; f_1(y) > \epsilon\})}{\mu(\epsilon)}\, d\epsilon\,. \end{aligned}$$

We thus recover the minorizing measure given in the lemma. □

If f_1 is bounded, $\epsilon^\star$ can be chosen as the maximum of f_1 and the set $\mathcal{S}(\epsilon_\star)$ is a small set. Roberts and Rosenthal (1998) also derive a drift condition (Note 6.9.1), as follows.

Lemma 8.7. *Assume*

(i) f_1 is bounded by 1,
(ii) the function μ is differentiable,

(iii) $\mu'(y)y^{1+1/\alpha}$ *is non-increasing for* $y < \epsilon^\star$ *and an* $\alpha > 1$.

Then, for $0 < \beta < \min(\alpha - 1, 1)/\alpha$,

$$V(x) = f_1(x)^{-\beta}$$

is a drift function, that is, it satisfies the drift condition (6.42) on the sets $\mathcal{S}(\epsilon^\star)$.

Proof. Recall that $\mathcal{S}(\epsilon^\star) = \{y;\ f_1(y) \geq \epsilon^\star\}$. If $f_1(x) \leq \epsilon^\star$, the image of V by the slice sampler transition kernel satisfies

$$\begin{aligned}
KV(x) &= \frac{1}{f_1(x)} \int_0^{f_1(x)} \frac{1}{\mu(\omega)} \int_{\mathcal{S}(\omega)} f_1(y)^{-\beta}\, dy\, d\omega \\
&= \frac{1}{f_1(x)} \int_0^{f_1(x)} \frac{1}{\mu(\omega)} \int_\omega^\infty z^{-\beta}(-\mu'(z))\, dz\, d\omega \\
&\leq \frac{1}{f_1(x)} \int_0^{f_1(x)} \frac{\int_y^{\epsilon^\star} z^{-\beta}(-\mu'(z))\, dz}{\int_y^{\epsilon^\star} (-\mu'(z))\, dz}\, d\omega \\
&\leq \frac{1}{f_1(x)} \int_0^{f_1(x)} \frac{\int_y^{\epsilon^\star} z^{-(1+\beta+1/\alpha}\, dz}{\int_y^{\epsilon^\star} z^{-(1+1/\alpha}\, dz}\, d\omega \\
&= \frac{1}{1+\alpha\beta} \frac{1}{f_1(x)} \left(\int_0^{f_1(x)} \omega^{-\beta}\, d\omega + \int_0^{f_1(x)} \frac{\epsilon^{\star -1/\alpha}(\omega^{-\beta} - \epsilon^{\star -\beta})}{\omega^{-1/\alpha} - \epsilon^{\star -1/\alpha}}\, d\omega \right) \\
&\leq \frac{V(x)}{(1-\beta)(1+\alpha\beta))} + \frac{\alpha\beta\epsilon^{\star -\beta}}{1+\alpha\beta}
\end{aligned}$$

(see Problem 8.12 for details).

When $f_1(x) \geq \epsilon^\star$, the image $KV(x)$ is non-increasing with $f_1(x)$ and, therefore,

$$KV(x) \leq KV(x_0) = -\epsilon^{\star\beta} \frac{1 + \alpha\beta(1-\beta)}{(1+\alpha\beta)(1-\beta)}$$

for x_0 such that $f_1(x_0) = \epsilon^\star$.

Therefore, outside $\mathcal{S}(\epsilon^\star)$,

$$KV(x) \leq \frac{1}{(1-\beta)(1+\alpha\beta))} + \frac{\alpha\beta(\epsilon_\star/\epsilon^\star)^\beta}{1+\alpha\beta} V(x) = \lambda V(x)$$

for any $\epsilon_\star < \epsilon^\star$ and, on $\mathcal{S}(\epsilon^\star)$,

$$KV(x) \leq b = -\epsilon^{\star\beta} \frac{1 + \alpha\beta(1-\beta)}{(1+\alpha\beta)(1-\beta)} - \lambda\,,$$

which establishes the result. □

Therefore, under the conditions of Lemma 8.7, the Markov chain associated with the slice sampler is geometrically ergodic, as follows from Theorem 6.75. In addition, Roberts and Rosenthal (1998) derive explicit bounds on the total variation distance, as in the following example, the proof of which is beyond the scope of this book.

Example 8.8. Exponential $\mathcal{E}xp(1)$ distribution If $f_1(x) = \exp(-x)$, Roberts and Rosenthal (1998) show that, for $n > 23$,

$$||K^n(x,\cdot) - f(\cdot)||_{TV} \leq .054865\,(0.985015)^n\,(n - 15.7043)\,. \tag{8.5}$$

This implies, for instance, that, when $n = 530$, the total variation distance between K^{530} and f is less than 0.0095. While this figure is certainly over-conservative, given the smoothness of f_1, it is nonetheless a very reasonable bound on the convergence time. ||

Roberts and Rosenthal (1998) actually show a much more general result: *For any density such that $\epsilon\mu'(\epsilon)$ is non-increasing, (8.5) holds for all x's such that $f(x) \geq .0025 \sup f(x)$.* In the more general case of the (product) slice sampler [A.32], Roberts and Rosenthal (1998) also establish geometric ergodicity under stronger conditions on the functions f_i.

While unidimensional log-concave densities satisfy the condition on μ (Problem 8.15), Roberts and Rosenthal (2001) explain why multidimensional slice samplers may perform very poorly through the following example.

Example 8.9. A poor slice sampler. Consider the density $f(x) \propto \exp\{-||x||\}$ in $\mathbb{R}^d$. Simulating from this distribution is equivalent to simulating the radius $z = |x|$ from

$$\eta_d(z) \propto z^{d-1}\, e^{-z}\,, \qquad z > 0\,,$$

or, by a change of variable $u = z^d$, from

$$\pi_d(u) \propto e^{-u^{1/d}}\,, \qquad u > 0\,.$$

If we run the slice sampler associated with π_d, the performances degenerate as d increases. This sharp decrease in the performances is illustrated by Figure 8.5: for $d = 1, 5$, the chain mixes well and the autocorrelation function decreases fast enough. This is not the case for $d = 10$ and even less for $d = 50$, where mixing is slowed down so much that convergence does not seem possible! ||

Roberts and Rosenthal (2001) take advantage of this example to propose an alternative to the slice sampler called the *polar slice sample*, which relates more to the general Gibbs sampler of next chapter (see Problem 9.21).

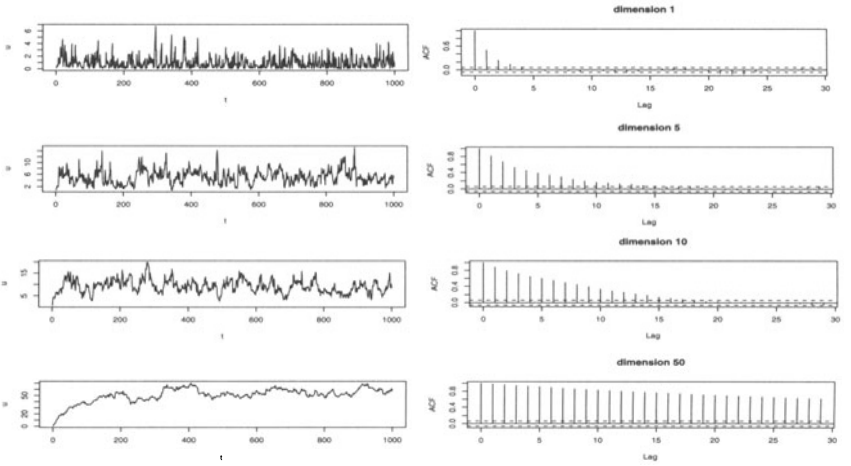

Fig. 8.5. Rawplots and autocorrelation functions for the series $(z^{(t)})$ generated by the slice samplers associated with $f_1(z) = z^{d-1}\, e^{-z}$ for $d = 1, 5, 10, 50$.

8.4 Problems

8.1 Referring to Algorithm [*A*.31] and equations (8.1) and (8.2):
(a) Show that the stationary distribution of the Markov chain [*A*.31] is the uniform distribution on the set $\{(x, u) : 0 < u < f(x)\}$.
(b) Show that the conclusion of part (a) remains the same if we use f_1 in [*A*.31], where $f(x) = Cf_1(x)$.

8.2 In the setup of Example 8.1, show that the cdf associated with the density $\exp(-\sqrt{x})$ on $\mathbb{R}_+$ can be computed in closed form. (*Hint:* Make the change of variable $z = \sqrt{x}$ and do an integration by parts on $z \exp(-z)$.)

8.3 Consider two unnormalized versions of a density f, f_1 and f_2. By implementing the slice sampling algorithm [*A*.31] on both f_1 and f_2, show that the chains $(x^{(t)})_t$ produced by both versions are exactly the same if they both start from the same value $x^{(0)}$ and use the same uniform $u^{(t)} \sim \mathcal{U}([0, 1])$.

8.4 As a generalization of the density in Example 8.1, consider the density $f(x) \propto \exp\{-x^d\}$, for $d < 1$. Write down a slice sampler algorithm for this density, and evaluate its performance for $d = .1, .25, .4$.

8.5 Show that a possible slice sampler associated with the standard normal density, $f(x) \propto \exp(-x^2/2)$, is associated with the two conditional distributions

$$\omega | x \sim \mathcal{U}_{[0, \exp(-x^2/2)]}\,, \qquad X | \omega \sim \mathcal{U}\left([-\sqrt{-2\log(\omega)}, \sqrt{-2\log(\omega)}]\right)\,. \tag{8.6}$$

Compare the performances of this slice sampler with those of an iid sampler from $\mathcal{N}(0, 1)$ by computing the empirical cdf at 0, .67, .84, 1.28, 1.64, 1.96, 2.33, 2.58, 3.09, and 3.72 for two samples of same size produced under both approaches. (Those figures correspond to the .5, .75, .8, .9, .95, .99, .995, .999 and .9999 quantiles, respectively.)

8.6 Reproduce the comparison of Problem 8.5 in the case of (a) the gamma distribution and (b) the Poisson distribution.