

# Perfect Simple Slice Sampler

Antonietta Mira

*University of Insubria, Department of Economics*

*Via Ravasi 2, 21100 Varese, Italy*

*anto@aim.unipv.it*

Jesper Møller

*Aalborg University, Department of Mathematical Sciences*

*Fredrik Bajers Vej 7G, DK-9220 Aalborg, Denmark*

*jm@math.auc.dk*

Gareth Roberts

*Lancaster University, Department of Mathematics and Statistics*

*Lancaster, LA1 4AF, UK*

*G.O.Roberts@lancaster.ac.uk*

## Abstract

Perfect sampling allows exact simulation of random variables from the stationary measure of a Markov chain. By exploiting monotonicity properties of the simple slice sampler we show that a perfect version of the algorithm can be easily implemented, at least when the target distribution is bounded. If this is not the case we suggest a way to construct an upper and/or lower bounding process.

*Keywords:* Auxiliary variables; Coupling from the past; Perfect simulation.

# 1 Introduction

The goal is to sample a complicated target distribution. To this aim we combine two quite powerful ideas that have recently appeared in the Markov chain Monte Carlo (MCMC) literature: the slice sampler (SS) and perfect simulation (PS).

The SS is a method of constructing, with the aid of auxiliary variables, a reversible Markov chain with a specified stationary distribution (Swendsen and Wong, 1987). The *simple SS* (SSS) is a special case where a single auxiliary variable is used. As for every MCMC method, draws from the stationary distribution, the target distribution, are obtained only after a "sufficiently long" run of the simulation. It is typically impossible to determine how long is sufficiently long.

Perfect simulation is a clever way of running a Markov chain which ensures that the terminal value of the implementation is an exact draw from the stationary distribution of the chain (Propp and Wilson, 1996).

By exploiting monotonicity properties of the SS we show that a perfect version of the algorithm can be easily implemented. This eliminates the problem of determining how long the SS Markov chain must be run before it has reached stationarity.

Numerous practical extensions and real applications of the perfect SSS are presented in Mira et al. (2001).

## 2 Slice Sampler

Here we will briefly introduce the SS. More extensive descriptions of the algorithm can be found in Mira and Tierney (1998) and Roberts and Rosenthal (1999).

Suppose  $\pi(x)$ ,  $x \in \mathcal{X}$  is an un-normalised integrable density with respect to the measure  $\mu$  and let  $\nu_\pi$  be the corresponding probability measure. In the SSS we introduce an auxiliary variable,  $u \in \mathcal{U}$ , and construct the joint distribution of  $u$  and  $x$  by taking the marginal for  $x$  unchanged and defining the conditional distribution of  $u$  given  $x$  to be uniform on  $(0, \pi(x))$ .

An irreducible and aperiodic Markov chain  $\{(X_n, U_n)\}_{n=0}^\infty$  is then set up over the enlarged state space  $\mathcal{X} \times \mathcal{U}$  having the probability measure corresponding to  $\pi(x, u) \propto I_{\{u < \pi(x)\}}(x, u)$ , as its unique stationary distribution, where  $I_A(x)$  is the indicator function of the set  $A$ . In particular we will perform a Gibbs sampler: in the vertical update  $u|x$  is sampled uniformly on  $(0, \pi(x))$ ; in the horizontal update  $x|u$  is sampled from the normalisation of the restriction of  $\mu$  to the set  $A_\pi(u) = \{x : \pi(x) > u\}$ . The marginal chain  $\{X_n\}_{n=0}^\infty$  has  $\nu_\pi$  as its stationary distribution and can thus be used to estimate integrals with respect to  $\pi$ .

When  $\mathcal{X} = \mathbf{R}^d$  and  $\mu$  is the  $d$ -dimensional Lebesgue measure, the SSS is uniformly ergodic if  $\pi$  is bounded and the support of  $\pi$  has finite Lebesgue measure (Mira and Tierney, 2001).

The SSS is stochastically monotone with respect to this ordering (Roberts and Rosenthal, 1999):

$$x \prec x' \text{ if and only if } \pi(x) \leq \pi(x'). \quad (1)$$

## 3 Perfect Simple Slice Sampler

Consider a positive recurrent Markov chain with invariant distribution  $\nu_\pi$  specified by a stochastic recursive sequence (SRS):  $X_{n+1} = f(X_n, \gamma_n)$ ,  $n > 0$ , where  $\{\gamma_n\}_{n=-\infty}^\infty$  is a sequence of iid random variables,  $n \in \mathbf{Z}$ . Below we briefly describe the Propp and Wilson (1996)

*coupling from the past* (CFTP) algorithm. Let  $X_n^{(x,t)}$  be the value at time  $n$  of the chain started in  $x$  at time  $-t$ . The CFTP algorithm applies, in theory, provided the *vertical backward coupling time*,  $T = \inf\{t > 0 : X_0^{(x,t)} = X_0^{(y,t)} \text{ for all } x, y\}$ , is almost surely finite. In fact  $X_0^{(x,t)} = X_0^{(y,t)} = X_0^{(\cdot, T)} \sim \nu_\pi$  for all states  $x, y$  and times  $t \geq T$ . PS becomes feasible, in practice, if there is a partial ordering  $\prec$  on the state space such that  $f(x, \gamma) \prec f(y, \gamma)$  if  $x \prec y$  and if there exist a maximal,  $x_{max}$ , and a minimal,  $x_{min}$ , state (i.e.  $x_{max} \prec x \prec x_{max}, \forall x$ ). Usually these states are assumed to be unique, but for the perfect SSS more than one maximal or minimal state can exist, furthermore the existence of  $x_{max}$  can be eliminated (Mira et al. 2001). CFTP works as follows. Choose a time  $T_1 > 0$ , generate  $\{\gamma_n\}_{n=-T_1}^{-1}$  and set

$$\begin{aligned} X_{-T_1}^{(x_{max}, T_1)} &= x_{max}, & X_{n+1}^{(x_{max}, T_1)} &= f(X_n^{(x_{max}, T_1)}, \gamma_n), & -T_1 \leq n < 0, \\ X_{-T_1}^{(x_{min}, T_1)} &= x_{min}, & X_{n+1}^{(x_{min}, T_1)} &= f(X_n^{(x_{min}, T_1)}, \gamma_n), & -T_1 \leq n < 0. \end{aligned}$$

If  $X_0^{(x_{max}, T_1)} = X_0^{(x_{min}, T_1)}$ , that is if coalescence occurs, then  $X_0^{(x_{min}, T_1)} = X_0^{(x, T_1)}$  for all  $x$ , and the common value,  $X_0^{(\cdot, T_1)}$ , is necessarily distributed as  $\nu_\pi$ . Otherwise choose a new value  $T_2 > T_1 > 0$  and restart the backward simulation from time  $-T_2$ . When running the simulation over the time range  $[-T_1, 0]$ , we need to reuse the same random numbers,  $\{\gamma_n\}_{n=-T_1}^{-1}$ , used in the first stage of the simulation. The procedure is repeated for  $k = 1, 2, \dots$  until  $X_0^{(x_{max}, T_k)} = X_0^{(x_{min}, T_k)}$ , whereby  $X_0^{(x_{min}, T_k)} = X_0^{(\cdot, T_k)}$  and so we return  $X_0^{(x_{min}, T_k)} \sim \nu_\pi$ . Notice that sample paths of maximal and minimal chains started at time point further back will be sandwiched in between paths started at earlier times (funneling property).

We now give an explicit SRS for the SSS, which preserves monotonicity with respect to the order given in (1). We assume for simplicity that maximal and minimal states exist and are unique; in fact, as explained in Mira et al. (2001), all we need to assume is that  $\mu(\mathcal{X})$  and  $\sup \pi$  are both finite. The SRS construction allows the continuum of chains implicitly defined in the PS, to be mapped to a countable collection of images in any particular iteration, of which only a finite number need ever be explicitly calculated. The construction is described in Figure 1. We shall carry out the vertical slice first, followed by the horizontal slice. For all  $t < 0$  define a vertical slice variable,  $\epsilon_t \sim U[(0, 1)]$ . Then, for the chain that, at time  $t$ , is in state  $x$ , set  $U_t(x) = \epsilon_t \pi(x)$ . The horizontal slice is more complicated. At each time  $t < 0$  construct an infinite sequence of random variables,  $\mathbf{W}_t = \{W_{t,j} : j = 1, 2, \dots\}$  by  $W_{t,1} \sim U[A_\pi(U_t(x_{min}))]$  and  $W_{t,j} \sim U[A_\pi(\pi(W_{t,j-1}))]$ . Let  $\sigma_t(x) = \inf\{j \geq 1 : \pi(W_{t,j}) \geq U_t(x)\}$ , and set

$$f(x, (\epsilon_t, \mathbf{W}_t)) = W_{t, \sigma_t(x)}. \quad (2)$$

It is easy to check that  $\sigma_t(x)$  is almost surely finite for all  $x \in \mathcal{X}$ . Since  $\gamma_t = (\epsilon_t, \mathbf{W}_t)$ ,  $t \in \mathbf{Z}$ , is independent of  $x$ , (2) is indeed an SRS representation for some Markov chain. The chain simulated is in fact a SSS because  $W_{t, \sigma_t(X_{t-1})}$  given  $U_t(X_{t-1}) = u$  is distributed as  $P_u$ : this is just an adaptive rejection sampling scheme where the rejection region becomes more and more refined as the simulation proceeds. The function in (2) is monotone in its first argument since, for all  $t$ ,  $W_{t,\cdot}$  and  $\sigma_t(\cdot)$  are non-decreasing sequences (the  $W$ 's with respect to  $\prec$ ) by construction. Hence, using (2), we have a CFTP algorithm for simulating from  $\nu_\pi$ . Because of the funnelling property and since  $\sigma_t(\cdot)$  is non-decreasing, it will never be necessary to increase the number of simulated  $W_{t,j}$ 's.

Even if an ordering has been defined on the state space it can be hard or impossible to find a maximal and/or a minimal state. Following ideas well summarized in Kendall and Møller

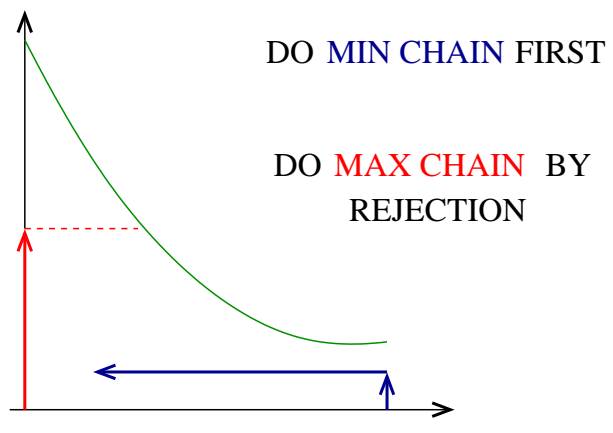
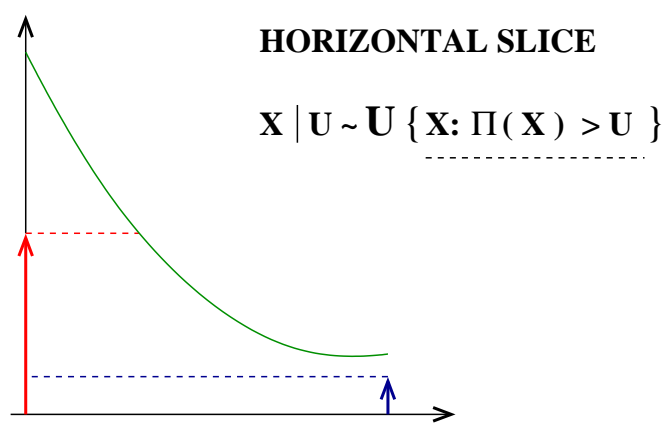
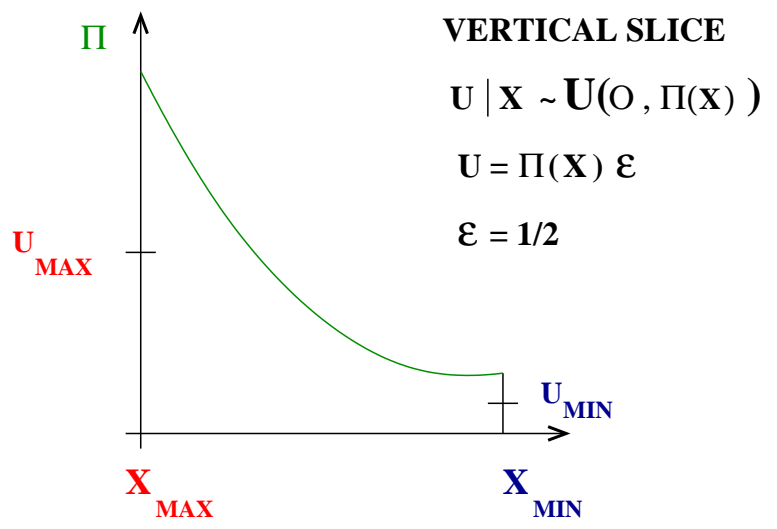


Figure 1: Perfect SSS on a truncated exponential distribution.

(2000), it is possible to construct upper and lower bounding stationary processes which allow the identification of upper and lower starting values for our sandwiching algorithm.

Assume that we are in the general setting of a stochastically monotone Markov chain with an SRS function  $f(x, \gamma)$  which is non-decreasing in  $x$ . Suppose we have upper and lower bounding processes with SRS functions  $f^{ub}$  and  $f^{lb}$  respectively, such that  $f^{lb} \prec f \prec f^{ub}$ , where the inequalities are assumed to hold pointwise for all values of the arguments of the functions. Simulate stationary versions of the upper and lower processes,  $S$  and  $L$  respectively, backwards in time till time  $-T_1$ , noting the forward seeds,  $\epsilon_{-T_1}, \dots, \epsilon_{-1}$ . The maximal process  $X^{(x_{max}, T_1)}$  now begins at  $X_{-T_1}^{(x_{max}, T_1)} = S_{-T_1}$ . The minimal process  $X^{(x_{min}, T_1)}$  now begins at  $X_{-T_1}^{(x_{min}, T_1)} = L_{-T_1}$ . If  $X_0^{(x_{max}, -T_1)} = X_0^{(x_{min}, -T_1)}$ , the algorithm ceases outputting the common value. Otherwise, the upper and lower processes are extended, backwards in time, to time  $-T_2$  and the procedure is repeated making sure to retain the seeds from the original time interval. This whole construction is extended until we get a  $k$  such that  $X_0^{(x_{max}, -T_k)} = X_0^{(x_{min}, -T_k)}$  and the common value is output. This value is distributed as  $\pi$  (Kendall and Møller, 2000).

Here we restrict attention to the case where  $\pi$  is bounded. This means that  $x_{max}$  exists, but a lower bounding process construction is needed. The ideas readily extend to the case where  $\pi$  is not bounded. A way of potentially exhibiting a bounding process for the SSS is described in the sequel. This is probabilistically the most natural construction since it is based entirely on  $Q_\pi(u) = \mu[A_\pi(u)]$  a function that completely characterizes the SSS (Roberts and Rosenthal, 1999).

Let  $Q'(u) = \frac{\partial Q(u)}{\partial u}$ . Assume that  $\mathcal{X} = \mathbf{R}^d$  and  $\mu$  is the Lebesgue measure. In order to define a minimal process for the SSS we find an unnormalised density  $\pi_2$  such that

$$\left[ \frac{Q'_\pi(u)}{Q_\pi(u)} - \frac{Q'_{\pi_2}(u)}{Q_{\pi_2}(u)} \right] \leq 0. \quad (3)$$

**Theorem 1.** *Suppose that  $\pi$  and  $\pi_2$  are densities with corresponding  $Q$  functions satisfying (3). Let  $X(2)$  be the simple slice sampler Markov chain induced by  $\pi_2$ . Then the Markov chains  $Y = \pi(X)$  and  $Y(2) = \pi_2(X(2))$ , are stochastically ordered in the sense that, for all  $\lambda, \beta \in \mathbf{R}$ :*

$$P(Y_1 \leq \lambda | Y_0 = \beta) \geq P(Y(2)_1 \leq \lambda | Y(2)_0 = \beta).$$

*Proof.* Writing  $\pi_1$  for  $\pi$ ,  $Y(1)$  for  $Y$ , from Roberts and Rosenthal (1999), for  $i=1,2$

$$P(Y(i)_1 \leq \lambda | Y(i)_0 = \beta) = \frac{1}{\beta} \int_0^\beta \max \left\{ 0, 1 - \frac{Q_{\pi_i}(\lambda)}{Q_{\pi_i}(z)} \right\} dz. \quad (4)$$

However, (3) implies that, for all  $z < \lambda$ ,

$$\int_z^\lambda \frac{Q'_{\pi_1}(w)}{Q_{\pi_1}(w)} dw \leq \int_z^\lambda \frac{Q'_{\pi_2}(w)}{Q_{\pi_2}(w)} dw$$

so that for all  $z < \lambda$

$$1 - \frac{Q_{\pi_1}(\lambda)}{Q_{\pi_1}(z)} \geq 1 - \frac{Q_{\pi_2}(\lambda)}{Q_{\pi_2}(z)}.$$

Hence, the integrand in (4) for  $Y(1)$  dominates that for  $Y(2)$  and the result follows.  $\square$

Although this is a natural way to bound the process in principle, the practical implementation of it is hampered by the fact that explicit information about  $Q'/Q$  is needed, and this is unlikely to be available in real problems. A more practical construction appears in Mira et al. (2001).

## 4 Extensions

Various extensions of the perfect SSS are discussed in Mira et al. (2001), including a construction of upper/lower bounding processes that can be used in real applications when the target is not bounded and perfect product SSs with multiple auxiliary variables. Examples of applications are the Ising model on a two dimensional grid at the critical temperature and various other auto-models.

The perfect SSS we have presented can be easily modified into an interruptible algorithm (Fill, 1998). The advantage of Fill versus Propp and Wilson CFTP is the fact the the former is not biased if the user becomes impatient and interrupts too long runs.

A further modification of the perfect SSS allows to use the idea of read once random numbers presented in Wilson (1999). The advantage of this procedure is to speed up the implementation and to make the coding of the algorithm easier due to the fact that no bookkeeping of the random bits  $\{\gamma_n\}_{n=-\infty}^{-1}$  is needed.

### REFERENCES

Fill, J. (1998). An interruptible algorithm for perfect sampling via Markov chains. *Annals of Applied Probability* **8**, 131–162.

Kendall, W. and Møller, J. (2000). Perfect Metropolis-Hastings simulation of locally stable point processes. *Adv. Appl. Prob.* **32**, to appear.

Mira, A. and Tierney, L. (1998). Efficiency and Convergence Properties of Slice Samplers. *Scandinavian Journal of Statistics*, to appear.

Mira, A., Møller, J. and Roberts, G. (2001). Perfect slice samplers. *J. Royal Stat. Soc. B*. To appear.

Propp, J. and Wilson, D. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms* **9**, 223–252.

Roberts, G. and Rosenthal, J. (1999). Convergence of slice sampler Markov chains. *J of Royal Statistical Society B* **61**, 643–660.

Wilson, D. (1999). How to couple from the past using a read-once source of randomness. *Random Structures and Algorithms*.

## **RESUME of Antonietta Mira**

### **Current position**

Associate Professor, the University of Insubria, Varese, Italy

### **Education**

- Philosophy Doctorate in Statistics, University of Minnesota, Minneapolis, MN, 1998
- Master of Science in Statistics, University of Minnesota, Minneapolis, MN, 1996
- Doctorate, Methodological Statistics, University of Trento, Italy, 1995

### **Honors**

- L.J. Savage Thesis Award 1998 for an outstanding doctoral dissertation in the area of Bayesian Econometrics and Statistics
- Dissertation Fellowship, 1997/98, University of Minnesota
- Research Assistantship, University of Minnesota. Supported by the N.S.F.
- Summa Cum Laude in Economics, University of Pavia, Italy, 1991
- 1991 Prize for the best dissertation in quantitative methods, U. of Pavia, Italy

**Main Research Interests** Statistical theory of Markov chain Monte Carlo methods (MCMC) and other simulation techniques such as Perfect Simulation. Application of statistical technology to problems arising in finance, engineering, industry and medicine.

### **Areas of particular interest are:**

- Convergence and efficiency of MCMC; Reversible Jump MCMC
- Sequential MCMC; Perfect simulation; Resampling techniques
- Auxiliary variables; Slice sampler; Ising model; Graphical models
- Ordering Markov transition kernels; Non reversible Markov chains
- Highly structured stochastic systems; Hidden Markov Models
- Data Mining; Knowledge Data Discovery
- Asymptotic convergence; Measures of skewness; Tests for symmetry
- Bayesian analysis; Non-parametric Bayesian statistics
- Change points problems; Mixture models