Slice Sampling - a Binary Implementation

John Skilling

Department of Applied Maths and Theoretical Physics,
Wilberforce Road, Cambridge CB3 0WA. United Kingdom.

David J.C. MacKay

Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE. United Kingdom.
mackay@mrao.cam.ac.uk

The real variables of a probabilistic model will always be represented in a computer using a finite number of bits. We describe an implementation of slice sampling in which the stepping-out, randomization, and shrinking operations, described by Neal (2003) in terms of floating-point operations, are replaced by binary and integer operations.

We assume that the variable x that is being slice-sampled is represented by a b-bit integer X taking on one of $B = 2^b$ values, $0, 1, 2, \ldots B-1$, many or all of which correspond to valid values of x. We often take these points to have equal prior measure, so that the prior becomes flat over X and all points are automatically a-priori-equivalent. Floating-point numbers, by contrast, are not equivalent, because of their variable rounding. Using an integer grid eliminates any errors in detailed balance that might thus ensue. We denote by $F(X)$ the appropriately transformed version of the unnormalized density $f(x(X))$.

We assume the following operators on b-bit integers are available:

$X + N$	arithmetic sum, modulo B, of X and N.
$X - N$	difference, modulo B, of X and N.
$X \oplus N$	bitwise exclusive-or of X and N.
$N \leftarrow \texttt{randbits}(l)$	sets N to a random l-bit integer.

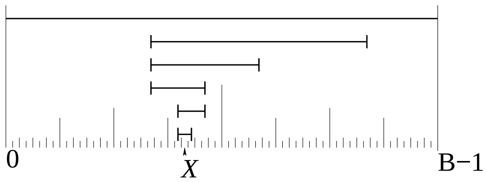
A slice-sampling procedure for integers is then as follows:

Given: a current point X and a height $Y = F(X) \times \text{Uniform}(0, 1) < F(X)$

$U \leftarrow \texttt{randbits}(b)$	Define a random translation U of the binary coordinate system.
set l to a value $l \leq b$	Set initial l-bit sampling range
do {	
$N \leftarrow \texttt{randbits}(l)$	Define a random move within the current interval of width 2^l.
$X' \leftarrow ((X - U) \oplus N) + U$	Randomize the lowest l bits of X (in the translated coordinate system).
$l \leftarrow l - 1$	If X' is not acceptable, decrease l and try again
} until $(X' = X)$ or $(F(X') > Y)$	with a smaller perturbation of X; termination at or before $l = 0$ is assured.

The translation U is introduced to avoid permanent sharp edges, where for example the adjacent binary integers `0111111111` and `1000000000` would otherwise be permanently in different sectors, making it difficult for X to move from one to the other.

Pictorially, the sequence of intervals from which the new candidate points are drawn are like the sequence of intervals in Neal's doubling procedure (Neal, 2001, figure 2).



If preliminary stepping-out from the initial range is required, step 2 above can be replaced by the following similar procedure:

2a	set l to a value $l < b$	(l sets the initial width)
2b	do {	
2c	$N \leftarrow$ `randbits`(l)	
2d	$X' \leftarrow ((X - U) \oplus N) + U$	
2e	$l \leftarrow l + 1$	
2f	} until $(l = b)$ or $(F(X') < Y)$	

These shrinking and stepping out methods shrink and expand by a factor of two per evaluation. A variant is to shrink or expand by more than one bit each time, setting $l \leftarrow l \pm \Delta l$ with $\Delta l > 1$. Provided the initial sampling range is well chosen (*i.e.,* of the same order of magnitude as the acceptable range), we found experimentally that the mean diffusion rate of X per evaluation when $\Delta l = 1$ is at most 25% slower than for Neal's method of shrinking to the rejected point. If the initial sampling range is not well chosen, the faster shrinking allowed here by setting $\Delta l > 1$ enables more rapid diffusion because an admittedly poorer acceptable jump is found more quickly.

A feature of using the integer representation is that, with a suitably extended number of bits, the single integer X can represent two or more real parameters – for example, by mapping X to (x_1, x_2, x_3) through a space-filling curve. Thus multi-dimensional slice sampling can be performed using the same software as for one dimension. Peano curves are useful here because they relate conveniently to a rectangular grid and they have the best possible locality properties: nearby points on the curve are close in space (though not the converse, which is unattainable). In this case, each successive bit of X represents a factor of 2 in volume. Because we are likely to be uncertain about the optimal sampling volume in several dimensions, it may be helpful to set Δl to the dimensionality. Taking Δl at each step from any pre-assigned distribution (which may include $\Delta l = 0$) allows extra flexibility.

References

Neal, R. M. (2003) Slice sampling. *Annals of Statistics* **31** (3): 705–767.

Published in the discussion of Neal (2003) in *Annals of Statistics*.

August 24, 2004 – Version 3.5