The Randomness Recycler Approach to Perfect Sampling

James Allen Fill
The Johns Hopkins University, Dept. of Mathematical Sciences
Whitehead Hall 306-F
Baltimore, MD 21218-2682, USA
jimfill@jhu.edu

Mark Huber
Stanford University, Dept. of Statistics
390 Serra Mall//Sequoia Hall
Stanford, CA 94305-4065, USA
mhuber@orie.cornell.edu

1. The Randomness Recycler versus Markov chains

At the heart of the Monte Carlo approach is the ability to sample from distributions that are in general very difficult to describe completely. For instance, the distribution might have an unknown normalizing constant which might require exponential time to compute. In these situations, in lieu of an exact approach, Markov chains are often employed to obtain approximately random samples. The primary drawback to Markov chain methods is that the mixing time of the chain is usually unknown, which makes it impossible to determine how close the output samples are to the target distribution.

Here we present the randomness recycler (RR) protocol, which overcomes this difficulty for several problems of interest. In contrast to traditional Markov chain approaches, an RR-based algorithm creates samples that are drawn exactly from the desired distribution. Other perfect sampling methods such as coupling from the past use existing Markov chains, but RR does not use the traditional Markov chain at all. While not universal, RR does apply to a wide variety of problems. In restricted instances of these problems, it gives the first expected linear time algorithms for generating samples. Here we present RR-type algorithms for self-organizing lists, the Ising model, random independent sets, random colorings, and the random cluster model.

In Markov chain approaches, small random changes are made in the observation until the entire observation has nearly the stationary distribution of the chain. The Metropolis [4] and heat bath algorithms utilize the idea of reversibility to design chains with a stationary distribution matching the desired distribution. Unfortunately, samples from the Markov chain approach are only approximately, not exactly, drawn from the stationary distribution of the chain. Moreover, they will not be close to the stationary distribution until a number of steps larger than the mixing time of the chain have been taken. Often the mixing time is unknown, and so the quality of the sample is suspect.

Propp and Wilson have shown how to avoid these problems using techniques such as coupling from the past (CFTP) [5]. For some chains, CFTP provides a procedure that allows perfect samples to be drawn from the stationary distribution of the chain, without knowledge of the mixing time. However, CFTP and related approaches have drawbacks of their own. These algorithms are noninterruptible, which means that the user must commit to running such an

algorithm for its entire (random) running time even though that time is not known ahead of time. Failure to do so can introduce bias into the sample. Other algorithms, such as FMMR [1], are interruptible, but require storage of random bits used by the algorithm. Because FMMR needs to read these bits twice, it is a read-twice algorithm. The method we present will be both interruptible and read-once, with no storage of random bits needed.

In addition, algorithms like CFTP and FMMR require an underlying Markov chain, and can never be faster than the mixing time of this underlying chain. Often these chains make changes to the state where the sample has already been randomized. This leads to wasted effort when running the algorithm that often adds a log factor to the running time of the algorithm.

The randomness recycler (RR) is not like any of these perfect sampling algorithms. In fact, the RR approach abandons the traditional Markov chain entirely. This is what allows the algorithm in restricted cases to reach an expected running time that is linear, the first such interruptible algorithm for several problems of interest.

2. The problems

In situations where Markov chains are commonly used, the state space is often of the form $\Omega \subseteq C^D$ and so consists of colorings of the elements of D with colors in C. For example, permutations form a subset of $\{1,\ldots,n\}^{\{1,\ldots,n\}}$. We will call a coloring $x \in \Omega$ a *configuration*, and we assign a weight $w(x)$ to each configuration. Let $Z := \sum_{x\in\Omega} w(x)$; then our goal is to sample from Ω, where the probability of choosing x is $w(x)/Z$. The table below lists several such examples where the Randomness Recycler can be employed instead of Markov chains. The first four models of the table use a graph (V, E); the fifth deals with permutations x.

example	D	C	parameter(s)	$w(x)$
Ising/Potts	V	$\{1,\ldots,Q\}$	T, Q	$\exp(\frac{1}{T}\sum_{\{v,w\}\in E} x(v)x(w))$
Random cluster	E	$\{0,1\}$	p, Q	$p^{\sum_e x(e)}(1-p)^{\lvert E\rvert-\sum_e x(e)}Q^{c(x)}$
Hard core gas	V	$\{0,1\}$	λ	$\left(\lambda^{\sum_{v\in V} x(v)}\right) 1_{(\forall\{v,w\}\in E:\, x(v)x(w)=0)}$
Proper colorings	V	$\{1,\ldots,k\}$	k	$1_{(\forall\{v,w\}\in E:\, x(v)\neq x(w))}$
Move ahead 1	$\{1,\ldots,n\}$	$\{1,\ldots,n\}$	$p_1,\ldots,p_n$	$\left(\prod_{i=1}^n p_{x(i)}^{n-i}\right) 1_{(\forall i\neq j:\, x(i)\neq x(j))}$

Note that the weight of a random cluster configuration includes a factor $Q^{c(x)}$. If we let A be the edge set $A := \{e : x(e) = 1\}$, then $c(x)$ is defined as the number of connected components in the graph (V, A).

3. The Randomness Recycler technique

We now describe the Randomness Recycler technique. At each time step t, we keep track of a configuration X_t and a set D_t. On elements of $D \setminus D_t$, the values of X_t are fixed, but on elements of D_t, the values of X_t are random. Specifically, we let $X_t^* = (D_t, X_t|_{D\setminus D_t})$ and $\pi_{X_t^*}(\cdot) = \pi\left(\cdot \,\middle|\, X_t|_{D\setminus D_t}\right)$. At each time step, we want the observation X_t to come from the restricted distribution, regardless of the history of X^*. More precisely, we want to maintain

$$P(X_t = x_t \mid X_0^* = x_0^*, \ldots, X_t^* = x_t^*) = \pi_{x_t^*}(x_t). \tag{1}$$

In particular, if τ is the (random) first time t that $D_t = D$, then $P(X_\tau = x \,|\tau = t) \equiv \pi(x)$, and our sample X_τ comes exactly from the desired distribution; further, τ and X_τ are independent.

Rather than working with the traditional Markov chain, we rely on a bivariate Markov chain that moves from state (X_t^*, X_t) to (X_{t+1}^*, X_{t+1}) at time $t + 1$. If the bivariate chain preserves (1) from each time t to the next $(t + 1)$, we call it π-*preserving*.

The following lemma provides a framework for creating π-preserving bivariate chains with a minimum of effort.

Lemma 3..1 *Let $P((x_t^*, x_t), (x_{t+1}^*, x_{t+1}))$ denote the probability that the bivariate Markov chain moves in one step from (x_t^*, x_t) to (x_{t+1}^*, x_{t+1}). If for every x_t^*, x_{t+1}^* there exists $C(x_t^*, x_{t+1}^*)$ such that*

$$\sum_{x_t} \pi_{x_t^*}(x_t) P((x_t^*, x_t), (x_{t+1}^*, x_{t+1})) = \pi_{x_{t+1}^*}(x_{t+1}) C(x_t^*, x_{t+1}^*)$$

for every x_{t+1}, then the procedure is π-preserving.

We now apply the lemma to construct an RR-type algorithm for the random cluster model. Here $D = E$, the edge set of a graph, and each edge is colored either 0 or 1. The following pseudocode describes the RR algorithm for this problem; below we discuss how the lemma led us to each key step. Recall that $c(x)$ is the number of connected components in $A := \{e : x(e) = 1\}$.

Randomness Recycler for random cluster model

1: **Set** $t \leftarrow 0$, $d_0 \leftarrow \emptyset$, $x_0(e) \leftarrow 0$ for all $e \in E$, $c(x_0) = |V|$
2: **While** $d_t \neq E$ do
3: **Set** $x_{t+1} \leftarrow x_t$
4: **Choose** an oriented edge $e = (v, w) \in E \setminus d_t$
5: **Set** $x_{t+1}(e) \leftarrow 1$ with probability p, $x_{t+1}(e) \leftarrow 0$ with probability $1 - p$
6: **Set** ACCEPT $\leftarrow$ TRUE
7: **If** $c(x_{t+1}) = c(x_t) - 1$ **then** with probability $1 - (1/Q)$ **set** ACCEPT $\leftarrow$ FALSE
8: **If** ACCEPT = TRUE **then set** $d_{t+1} \leftarrow d_t \cup \{e\}$, $t \leftarrow t + 1$
9: **Else set** $\mathcal{E}_{\text{rej}} \leftarrow$ (component C in x_t containing w) $\cup$ {edges adjacent to C in d_t} and **set** $x_{t+1}(e) \leftarrow 0$ $\forall e \in \mathcal{E}_{\text{rej}}$, $d_{t+1} \leftarrow d_t \setminus \mathcal{E}_{\text{rej}}$, $t \leftarrow t + 1$

Roughly speaking, coloring an edge 1 multiplies the weight by p, while coloring the same edge 0 multiplies the weight by $1 - p$. Hence our initial proposed color for e in line 5 uses these probabilities.

If the edge is colored 1, then $c(x)$ might be reduced by 1 as two formerly unconnected components are connected. By accepting in that case only with probability $1/Q$, we multiply the weight of x by $1/Q$, exactly making up for the change in the number of connected components.

Line 9 requires some explanation. We begin the **While** loop with an observation x_t distributed according to $\pi_{x_t^*}$. If however, we make it to ACCEPT $\leftarrow$ FALSE in line 7, then we know that v and w are not connected via edges colored 1 in x_t. Thus by this stage x_t is (conditionally) distributed according to $\pi_{x_t^*}(\cdot \mid v, w \text{ unconnected})$. The fact that v and w are not connected can be expressed equivalently as the fact that the component C in x_t containing w does not contain v. To "undo" our knowledge of what constitutes C, in line 9 we remove from d_t all edges in or adjacent to C (and "freeze" their x-values at 0).

These are rough intuitive statements, but use of the lemma can make these arguments precise and show that the bivariate chain actually does preserve π. We do not have space here to present the computations, but they are straightforward.

4. Conclusions

We now tabulate some results on running time for our RR procedure. The column "approximate" contains what is known about the mixing time of the traditional Markov chain Monte Carlo approach. The column "CFTP" refers to the coupling from the past methodology

for obtaining exact samples due to Propp and Wilson, and indicates when CFTP is known to run in time of order $|D| \ln |D|$. The run-time bound can be improved to linear time [2], but the restrictions on the parameters are stronger. The final column refers to the RR procedure, which in cases satisfying the listed parameter restrictions runs in linear time. The value Δ refers to the maximum degree of the graph, and for the move-ahead-1 model we have restricted attention to the geometric case $p_i \equiv r^{i-1}(1-r)/(1-r^n)$. Space limitations prevent us from giving more careful descriptions of the parameter restrictions here.

example	approximate	CFTP [5, 3]	RR
Ising/Potts	$T > T_{\text{crit}}$	$T > T_{\text{crit}}$	$T = \Omega(\Delta)$
Random cluster	unknown	unknown	$p = O(1/\Delta)$
Hard core gas	$\lambda < 2/(\Delta - 2)$	$\lambda < 2/(\Delta - 2)$	$\lambda < 4/(3\Delta - 4)$
Proper colorings	$k > 11\Delta/6$	$k = \Omega(\Delta^2)$	$k = \Omega(\Delta^4)$
Move ahead 1	$r < 0.2$	$r < 0.2$	unknown

One attractive feature of RR procedures is that they may be run even if the order of magnitude of the time until $D_t = D$ is not known in advance. For instance, in the case of the move-ahead-1 permutation chain with $p_i \propto r^i$, experiments indicate that the RR procedure takes linear time for fixed r at least up to $r = 0.99$, although no *a priori* run-time bounds are known.

Again we point out that unlike CFTP, our RR algorithm is interruptible, meaning that the algorithm can be aborted without generating bias in the sample. Therefore our Randomness Recycler approach gives the first interruptible linear time algorithms for several problems of interest.

REFERENCES

[1] James A. Fill, Motoya Machida, Duncan J. Murdoch, and Jeffrey S. Rosenthal. Extension of Fill's perfect rejection sampling algorithm to general chains. *Random Structures & Algorithms*, 17:290–316, 2000.

[2] Olle Häggström and Jeff Steif. Propp-Wilson algorithms and finitary codings for high noise markov random fields. *Combin. Probab. Computing*, 9:425–439, 2000.

[3] Mark L. Huber. *Perfect Sampling with Bounding Chains*. PhD thesis, Cornell University, 1999.

[4] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculation by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.

[5] James G. Propp and David B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures & Algorithms*, 9(1–2):223–252, 1996.

RESUME

L'approche de Recycler d'aspect aléatoire pour obtenir des échantillons provenant des distributions dimensionnelles élevés évite le problème de savoir la période de mélange d'une chaîne de Markov traditionelle, au lieu de cela garantissant que l'échantillon vient exactement de la distribution désirée.