Perfect sampling, old and new

Mark Jerrum

Queen Mary, University of London

Edinburgh Mathematical Society meeting, St Andrews
22nd March 2024

Motivation

Problem: produce a realisation of a random variable with a specified probability distribution.

Example: Given an undirected graph G, sample, uniformly at random, a spanning tree in G.

The sample must exactly match the desired probability distribution. This rules out direct Markov chain simulation, a common approach to sampling.

Why study perfect sampling?

- Theoretical appeal.
- Perfect samplers are 'self clocking'.

Scope

- Perfect sampling is now quite a wide a topic.

  Mark Huber, *Perfect simulation*. Monographs on Statistics and Applied Probability, 148. CRC Press, 2016.

  . . . and quite a lot has happened since 2016.
- Today, we concentrate on samplers that can, in principle, work in time linear in the input size (or in constant time for each 'piece' of the output).

  When this happens, it also opens up the possibility of sampling (portions of) infinite objects.
- We'll look at a couple of sampling algorithms that achieve this.

Spin systems: a playground

A spin model is defined by a set of spins Q, a 'field' $b : Q \to \mathbb{R}$, and symmetric pairwise interactions $A : Q \times Q \to \mathbb{R}$. In our case, Q is finite.

The spin system associated with a graph $G = (V, E)$ has as its set of *configurations* $\Omega = Q^V$. For $\sigma \in \Omega$ and $W \subseteq V$ we denote the restriction of σ to W by σ_W. Initially, we restrict attention to finite graphs.

The *Gibbs distribution* assigns to each configuration σ the weight

$$\mathrm{wt}(\sigma) = \prod_{v \in V} b(\sigma_v) \prod_{(u,v) \in E} A(\sigma_u, \sigma_v).$$

Spin systems (continued)

Define the *partition function* $Z(G)$ of the system to be the sum of all possible weights

$$Z(G) := \sum_{\sigma \in \Omega} \mathrm{wt}(\sigma).$$

Then we define the *Gibbs measure* by $\mu(\sigma) = \mathrm{wt}(\sigma)/Z(G)$, for all $\sigma \in \Omega$.

The main computational problems are:

- Compute (exactly or approximately) the partition function $Z(G)$.
- Sample (perfectly or near-perfectly) from the Gibbs distribution.

We focus here on perfect sampling.

An example: the Ising model

This is a two-spin system; conventionally, the two spins are written $Q = \{-, +\}$ or $Q = \{-1, +1\}$. Then

$$b = \begin{matrix} - \\ + \end{matrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad A = \begin{matrix} \\ - \\ + \end{matrix} \begin{matrix} \begin{matrix} - & \quad + \end{matrix} \\ \begin{pmatrix} \lambda & \lambda^{-1} \\ \lambda^{-1} & \lambda \end{pmatrix} \end{matrix}.$$

When $\lambda > 1$, the model is ferromagnetic; adjacent spins prefer to be alike.

By altering b we can bias the spins by introducing an 'external field'.

The goal

Realise a configuration from the Gibbs distribution. Desiderata:

- Generate perfect samples.
- Handle infinite graph instances (finite windows onto an infinite configuration).
- Work in time linear in the size if the output.

References:

- Van den Berg & Steif, On the existence and nonexistence of finitary codings for a class of random fields. *Ann. Probab.*, 1999.
- Anand & Jerrum, Perfect sampling in infinite spin systems via strong spatial mixing, SIAM J. Comput., 2022.

Coupling From The Past (CFTP)

Suppose we want to select a number uniformly a random from $\{0, 1, \dots, n-1\}$. (Not a hard problem!)

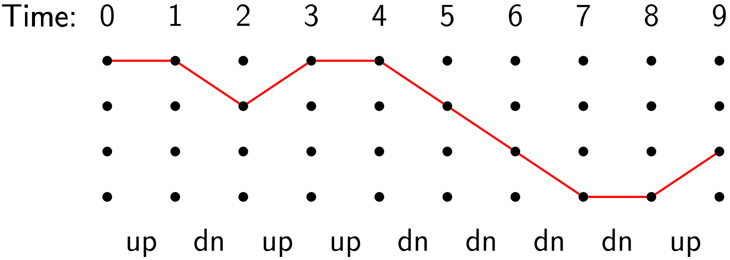
We could do this approximately by simulating a symmetric random walk $\{0, 1, \dots, n-1\}$. Our convention at the boundaries is to set $p_{00} = p_{01} = \frac{1}{2}$ and $p_{n-1,n-1} = p_{n-1,n-2} = \frac{1}{2}$.

After $C(\varepsilon)n^2$ steps, this random walk will be within variation distance ε of uniform.

Now suppose that we start n random walks from all possible starting points, *using the same random coin flips for each*.

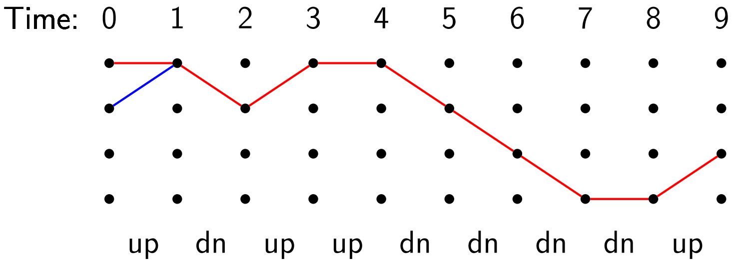
Coupling random walks

After a while, the walks coalesce.



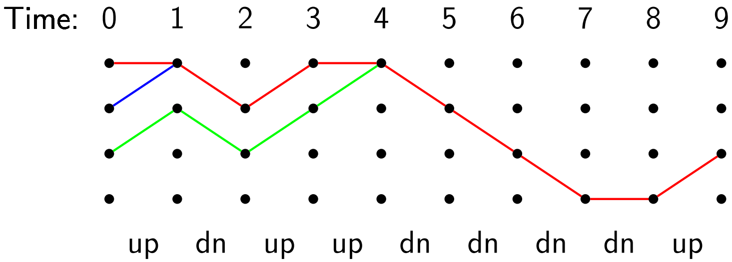
Coupling random walks

After a while, the walks coalesce.



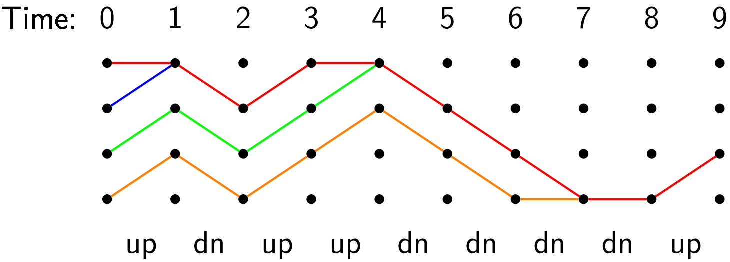
Coupling random walks

After a while, the walks coalesce.



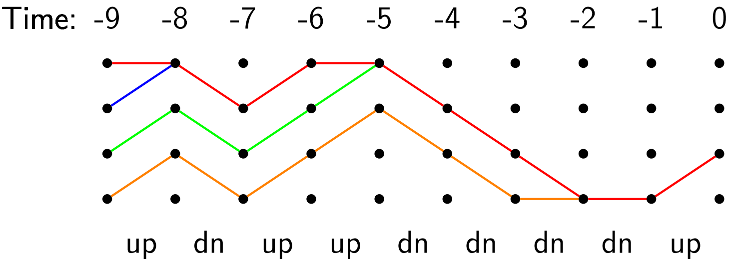
Coupling random walks

After a while, the walks coalesce.



Coupling random walks

After a while, the walks coalesce.



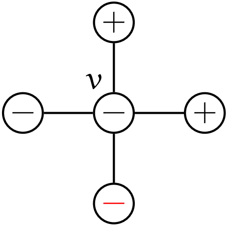
Now think the final time being 0. Can we simulate the Markov chain from time $-\infty$?

Yes! Start at time -9 or even -4 ... but not -3! [Propp and Wilson, 1996]. NB. Need to use the same coin flips each time.

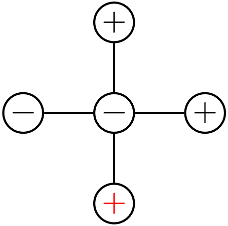
The ferromagnetic Ising model: Glauber dynamics

Random walk on $\{0, \ldots, n-1\} \to$ 'Glauber dynamics' on Ising configurations.

Choose a vertex $v \in V(G)$ uniformly at random and update its spin 'appropriately'.



$\leqslant$
(product order)



$$\sigma(v) \leftarrow \begin{cases} +, & \text{w.p. } \frac{1}{2}; \\ -, & \text{w.p. } \frac{1}{2}. \end{cases} \qquad \sigma(v) \leftarrow \begin{cases} +, & \text{w.p. } \lambda^2/(\lambda^2+\lambda^{-2}); \\ -, & \text{w.p. } \lambda^{-2}/(\lambda^2+\lambda^{-2}). \end{cases}$$

Monotonicity of the Glauber dynamics

The single-vertex updates we have sketched define an ergodic Markov chain with the desired Gibbs distribution.

Furthermore as $\lambda^2/(\lambda^2 + \lambda^{-2}) \geqslant \frac{1}{2}$ (this holds more generally) we can couple two evolutions so that $\leqslant$ is preserved. As with the symmetric random walk, the Glauber dynamics is monotonic. We can apply CFTP to produce exact samples from the Gibbs distribution!

Doesn't work for the antiferromagnetic Ising model, which is antimonotonic. (Can recover something, but with a loss.).

Infinite graphs?

The Gibbs measure is (sometimes) defined for infinite graphs G, e.g., $\mathbb{Z}^2$.

We require that this Gibbs measure agrees with the specification on 'cylinder events' where all but a finite number of vertices have their spins pinned to a certain value.

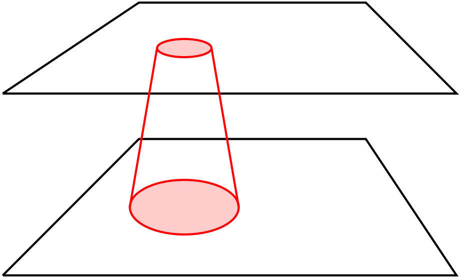
A Gibbs measure on an infinite graph G may or may not exist, and may or may not be unique.

If it is unique, it is reasonable to ask whether we can sample from it.

Coupling from the past on $\mathbb{Z}^2$, pictorially.

Time $= 0$

Time $= -t$



- The further we delve into the past, the further we must look for information.
- The further we delve into the past, the more likely is coalescence.

The second effect may overwhelm the first! [Van den Berg & Steif, 1999; Spinka, 2020].

A second example: the hardcore model (weighted independent sets)

The hardcore model is a distribution over independent sets of a graph, which we model as a spin system with $Q = \{0, 1\}$. The one constraint is that neighbouring vertices are not allowed to both have spin 1.

$$b = \begin{pmatrix} 1 \\ \lambda \end{pmatrix}, \quad A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

When $\lambda = 1$, the Gibbs measure is uniform over all independent sets in G. Varying lambda affects the density of a typical configuration.

A new approach: Lazy depth-first sampling

Suppose $v \in V(G)$ and (Γ, σ_Γ) is a *context*, composed of a set $\Gamma \subseteq V(G) \setminus \{v\}$ and a partial assignment $\sigma_\Gamma : \Gamma \to Q$. We wish to design a perfect sampler that will return a spin $s \in Q$ from the marginal distribution on v, conditioned on vertices in Γ having spins pinned to σ_Γ.

Towards a sampling scheme (for independent sets)

$\mathrm{LDFSample}(v, (\Gamma, \sigma_\Gamma))$

for all vertices u adjacent to v **do**

  if $u \notin \Gamma$ **then**

    $s := \mathrm{LDFSample}(u, (\Gamma, \sigma_\Gamma))$

    $\Gamma := \Gamma \cup \{u\}; \quad \sigma_\Gamma := \sigma_\Gamma \cup \{(u, s)\}$

  end if

end for

if $\sigma_\Gamma(u) = 1$ for some u adjacent to v **then**

  return 0

else

  with probability $\lambda/(\lambda + 1)$ return 1 otherwise 0

end if

Towards a sampling scheme

$\mathtt{LDFSample}(v, (\Gamma, \sigma_\Gamma))$
for all vertices u asjacent to v **do**
  if $u \notin \Gamma$ **then**
    $s := \mathtt{LDFSample}(u, (\Gamma, \sigma_\Gamma))$
    $\Gamma := \Gamma \cup \{u\}; \quad \sigma_\Gamma := \sigma_\Gamma \cup \{(u, s)\}$
  end if
end for
if $\sigma_\Gamma(u) = 1$ for some u adjacent to v **then**
  $I_0 := [0, 1]$ and $I_1 := \emptyset$
else
  $I_0 := [0, 1/(\lambda + 1)]$ and $I_1 := (1/(\lambda + 1), 1]$
end if
Let α be a realisation of a uniform $[0, 1]$ random variable
return s where $\alpha \in I_s$

Towards a sampling scheme

$\mathtt{LDFSample}(v, (\Gamma, \sigma_\Gamma))$
for all vertices u asjacent to v **do**
  if $u \notin \Gamma$ **then**
    $s := \mathtt{LDFSample}(u, (\Gamma, \sigma_\Gamma))$
    $\Gamma := \Gamma \cup \{u\}; \quad \sigma_\Gamma := \sigma_\Gamma \cup \{(u, s)\}$
  end if
end for
if $\sigma_\Gamma(u) = 1$ for some u adjacent to v **then**
  $I_0 := [0, 1]$ and $I_1 := \emptyset$
else
  $I_0 := [0, 1/(\lambda + 1)]$ and $I_1 := (1/(\lambda + 1), 1]$
end if
Let α be a realisation of a uniform $[0, 1]$ random variable
return s where $\alpha \in I_s$

Towards a sampling scheme

$\mathtt{LDFSample}(v, (\Gamma, \sigma_\Gamma))$

Let α be a realisation of a uniform $[0, 1]$ random variable

for all u adjacent to v **do**

  if $u \notin \Gamma$ **then**

    $s := \mathtt{LDFSample}(u, (\Gamma, \sigma_\Gamma))$

    $\Gamma := \Gamma \cup \{u\}; \quad \sigma_\Gamma := \sigma_\Gamma \cup \{(u, s)\}$

  end if

end for

if $\sigma_\Gamma(u) = 1$ for some u adjacent to v **then**

  $I_0 := [0, 1]$ and $I_1 := \emptyset$

else

  $I_0 := [0, 1/(\lambda + 1)]$ and $I_1 := (1/(\lambda + 1), 1]$ // 'Zone of indecision'

end if

return s where $\alpha \in I_s$

Lazy Depth-First Sampler (LDFS)

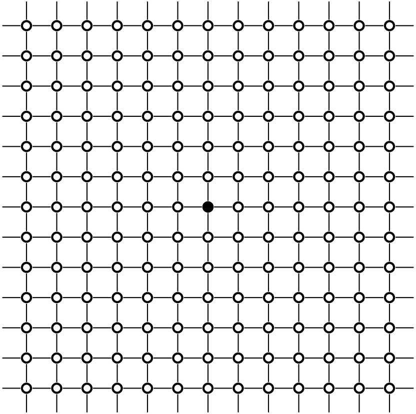
- Now observe that if $\alpha \leqslant 1/(\lambda + 1)$ then the output is necessarily 0... so we don't need to make the recursive calls.
- Suppose we apply $\mathrm{LDFSample}$ to a graph of maximum degree Δ. With probability $1/(\lambda + 1)$ we make no recursive calls. With probability $\lambda/(\lambda + 1)$ we make at most Δ recursive calls. So the expected number of recursive calls is bounded by a branching process. The mean number of offspring at a node of the tree $\lambda\Delta/(\lambda + 1)$. So the expected number of recursive calls is bounded if $\lambda\Delta/(\lambda + 1) < 1$, i.e., $\lambda < 1/(\Delta - 1)$.
- This analysis is crude and can easily be improved. For example, if some recursive call generates a 1 we can omit the succeeding recursive calls.

Radius $r > 1$ Lazy Depth-First Sampler

- The same approach can be applied to many other sampling problems, but there are limits. For example, it does not apply to q-colouring, as any putative colour for v is eliminated by some colouring of its neighbours.
- However, we can generalise the sampling scheme just considered to radius r. In this, we recursively call the procedure on vertices at distance r. What we have seen with independent sets is the special case $r = 1$.
- Even at $r = 2$, the approach has something to say about sampling q-colourings, provided q is sufficiently large in relation to Δ.

Sampling on infinite lattices

Suppose we want to sample a uniform random 6-colouring of an infinite square lattice (or, say, a large $L \times L$ square region, to avoid discussing what we mean by a random colouring of an infinite graph).



Zone of indecision

- When r is large, the influence of the vertices at radius r is small.
- Suppose that, when $r = 6$, the marginal probability at the origin v is such that $\Pr(v \text{ is red}) > 0.165$ and similarly for the other colours.
- The 'zone of indecision' then has length $1 - 6 \times 0.165 = 0.01$.
- There are just 48 vertices in the boundary, so the branching process is subcritical $(48 \times 0.01 < 1)$.
- We thus would have a sampling algorithm that requires expected constant time per site.

Weak spatial mixing

- In fact, in the case of 6-colourings of a square lattice, it *is* the case that the influence of vertices at distance r declines exponentially fast with r. So the zone of indecision contracts exponentially fast with r.
- At the same time, the number of vertices at distance r is $8r$.
- So the branching process will become subcritical at *some* value of r, and we obtain a constant-time per site sampling algorithm, that works even for infinite square lattices.
- But hold on. . . .

Strong spatial mixing

- As the sampling algorithm progresses, the environment of frozen colours grows.
- So to get things to work, we need exponential decay of correlations, *even when some vertices are 'pinned' to certain colours*. This is 'strong spatial mixing' and is a more elusive concept than weak spatial mixing.
- In fact, strong spatial mixing *does* hold for 6-colourings of the square lattice [Achlioptas, Molloy, Moore and Van Bussel, 2005], [Goldberg, Jalsenius, Martin & Paterson, 2006]. (At the time of writing, I don't know the status of 5-colourings.)
- However, it is possible to get away with weak spatial mixing in this and some other similar situations.

The ferromagnetic Ising model

An interesting example is the ferromagnetic Ising model.

Ding, Song and Sun recently (2021) settled a long-standing conjecture in the affirmative: in a strong sense, weak spatial mixing implies strong spatial mixing for the ferromagnetic Ising model. (The ferromagnetic Ising model may be essentially unique in having this strong property.)

As a consequence we can sample Ising configurations on the cubic lattice $\mathbb{Z}^{\mathrm{d}}$ whenever weak spatial mixing holds, i.e., whenever the Ising measure is defined on those lattices.

Work in progress

- When can we weaken the constraint of strong spatial mixing?
- Do we have to pay (in terms of range of validity of the sampling algorithm) for perfect sampling?
- Do we have to pay (in terms of range of validity of the sampling algorithm) for linear time (constant time per site) sampling?