Exact Mixing and the Halting State Theorem

Kassie Archer

Math 100 – Winter 2013

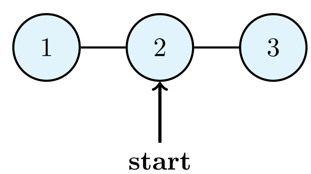
Everything in this paper comes from [1], [2], and [3]. Respectively, these are 74, 77, and 91 in Peter Winkler's publication list.

1 Exact Mixing

The goal of exact mixing is to achieve a target distribution, τ, exactly given a starting distribution σ. We accomplish this through the use of a *stopping rule*, which is denoted here by Γ. A stopping rule will tell us the probability that we stop at a given vertex depending on how the walk has progressed so far, but independent of the future. Often, our target distribution τ is the stationary distribution π and our starting distribution σ will be concentrated at one vertex j, in which case we write $\sigma = j$. (Here, we really mean that $\sigma_j = 1$ and $\sigma_i = 0$ for all $i \neq j$.)

Example 1.1. Consider the simple example of the walk on the graph G pictured to the right. This is the line on 3 vertices. Suppose our starting distribution σ is concentrated at vertex 2 and our target distribution is the stationary distribution

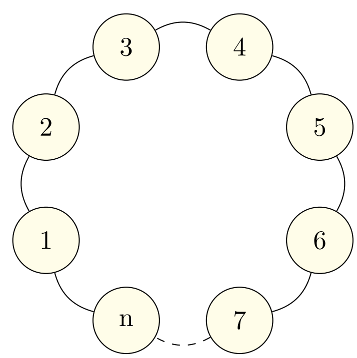
$$\pi = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}.$$



Then we can follow the stopping rule Γ which tells us to stop where we started (at 2) with probability $\frac{1}{2}$ and otherwise to stop after taking one step. With this stopping rule we attain our stationary distribution exactly.

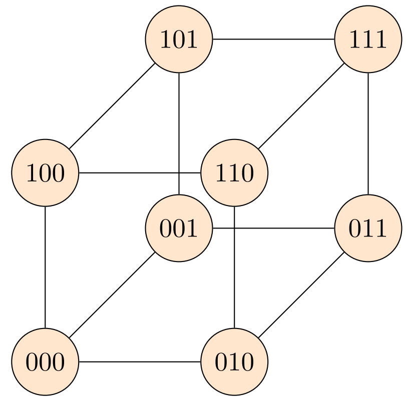
Example 1.2. Suppose G is the cycle on n vertices. Recall that the cycle has the property that if you start at a vertex j and walk until you have hit every vertex at least once, every state except for j has an equal likelihood of being the last vertex hit. We can use this fact to find an interesting stopping rule for this graph. As usual, suppose our starting distribution is $\sigma = j$ for some vertex j and our target distribution is $\tau = \pi$, the stationary distribution. Because of symmetry, the stopping rule is the same for whichever j we choose. Additionally, the stationary distribution is

$$\pi = \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix}.$$



Our stopping rule Γ says to stop with probability $\frac{1}{n}$ at vertex j where you started and with probability $\frac{n-1}{n}$, walk until you have covered the entire graph. Then you have a $\frac{1}{n}$ chance of stopping at vertex j and an equal chance of stopping at every other vertex and so you have attained the stationary distribution.

Example 1.3. Suppose G is the n-cube. The 3-cube is pictured to the right. Suppose you start at the origin. Since the cube is symmetric, it doesn't actually matter where you start. As usual, suppose $\tau = \pi$, the uniform distribution. Our stopping rule says to choose a direction at random. For the 3-cube, this means choose the up-down direction, the left-right direction, or the forward-backwards direction each with probability 1/3. Next, flip a coin. If it comes up heads, move in that direction. If it comes up tails, don't move. Once you've chosen every direction at least once, stop. Since this is a lazy walk instead of an actual random walk, just say that no time passes if the coin comes up tails and you don't move. This results in the uniform distribution π since each coordinate has an equal chance of being 0 or 1.



We may ask whether there exists a stopping rule for any graph G, any starting distribution σ, and any target distribution τ. In fact, there are many such stopping rules. The most simple one is called the *naive stopping rule*.

Example 1.4. For any graph G, any starting distribution σ, and any target distribution τ, there is a stopping rule denoted by $\Omega_{\sigma,\tau}$ called the *naive stopping rule*. This rule says to choose some vertex i according to τ, that is, choose i with probability τ_i. Then walk until you hit i. This way, no matter what σ is, you have a τ_i chance of stopping at i and so you do achieve the target distribution.

2 Mean Length and the Halting State Theorem

The *mean length* of a stopping rule Γ, denoted $\mathbb{E}\Gamma$ is the expected number of steps you take in the random walk before stopping according to Γ. A stopping rule is said to be *mean-optimal* if $\mathbb{E}\Gamma$ is minimal over all possible stopping rules for the given G, σ, and τ. In general $\Omega_{\sigma,\tau}$ is not mean optimal.

The naive stopping rule always has mean length

$$\mathbb{E}\Omega_{\sigma,\tau} = \sum_{i,j} \sigma_i \tau_j \mathcal{H}(i,j)$$

where $\mathcal{H}(i,j)$ is the expected hitting time from i to j. This formula follows from the fact that for each pair of vertices i and j, there is probability σ_i of starting at i, probability τ_j of choosing vertex j to stop, and expected time $\mathcal{H}(i,j)$ until you stop.

In general $\Omega_{\sigma,\tau}$ is not a mean optimal stopping rule. For example, if $\sigma = \tau$, the stopping rule Γ which tells you to stop where you start has mean length $\mathbb{E}\Gamma = 0$, but as defined above $\mathbb{E}\Omega_{\sigma,\tau}$ is clearly not zero.

We can recognize when a stopping rule is optimal by means of the *Halting State Theorem.*

Theorem 2.1. *(Halting State Theorem) A stopping rule* Γ *is mean-optimal if and only if* Γ *has a halting state.*

A *halting state* for a stopping rule Γ is a state (vertex) i where if you enter i you stop with probability one according to Γ.

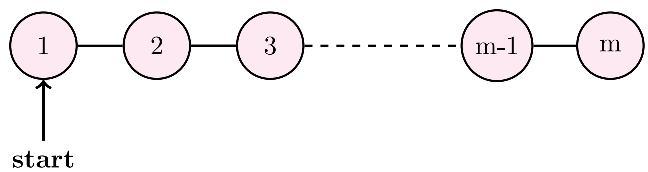
Example 2.2. In example 1.1, the graph has two halting states, 1 and 3. You may stop at vertex 2, but if you do take a step to 1 or 3, you will certainly stop. Therefore the stopping rule described is mean-optimal.

Example 2.3. In example 1.2, there is no halting state. If there were, you would not be able to cover the graph because you would always stop at the halting state. So the stopping rule Γ described in this example is not mean-optimal. In fact, the naive stopping rule does better. Certainly, in general you don't have to cover the cycle every time to complete the naive stopping rule like you do for Γ.

Example 2.4. In example 1.3, there is one halting state, the point with all coordinates equal to one. In the 3-cube example, it is 111. By the time you reach the point with all coordinates equal to one, you must have chosen every direction at least once, so you stop. Therefore, this stopping rule is optimal.

As mentioned, the naive stopping rule is not in general mean-optimal, but for certain examples, it may be.

Example 2.5. Consider the walk on a line of m vertices where our starting distribution is $\sigma = 1$ and our target distribution is $\tau = \pi$. Then the naive stopping rule $\Omega_{1,\pi}$ is mean-optimal because it has a halting state m. Since the naive stopping rule says to stop once you hit your target vertex i chosen at random from τ, then if we make it to vertex m, we must have passed through every other vertex. Therefore, our target vertex must have been $i = m$ and so we stop.



3 Proof of the Halting State Theorem

In order to prove the Halting State Theorem, we need the idea of *exit frequencies*. The exit frequency x_i for the vertex i is the expected number of times you leave the node i during a random walk with stopping rule Γ. If you stop at i, this doesn't count towards the exit frequency. Therefore, if i is a halting state, $x_i = 0$. The following two lemmas will be helpful for the proof of the Halting State Theorem.

Lemma 3.1. $\mathbb{E}\Gamma = \sum_i x_i$.

Proof. The left hand side is the expected number of steps you take before stopping. This must be the sum of the number of times you expect to leave vertex i for each i, which is exactly the right hand side. □

Lemma 3.2. $\sigma_i + \sum_j p_{ji} x_j = x_i + \tau_i$.

Proof. Each side counts the expected number of times you are in vertex i. The left hand side is the expected number of times you start at i (counted by σ_i) or enter i from another vertex i (counted by $\sum_j p_{ji} x_j$) and the right hand side is the expected number of times you leave i (counted by x_i) or stop at i (counted by τ_i). □

With these lemmas in hand, we can now prove the Halting State Theorem.

Proof of the Halting State Theorem. We start by proving the reverse implication, that if Γ has a halting state, it must be mean-optimal. Suppose Γ, Γ' are both stopping rules for graph G with vertices V, starting distribution σ, and target distribution τ. Additionally, suppose Γ has a halting state h and Γ' is arbitrary. We will show that Γ is optimal.

Suppose $\{x_i\}_{i\in V}$ are exit frequencies for Γ and $\{x'_i\}_{i\in V}$ are exit frequencies for Γ'. Define $y_i := x'_i - x_i$. Then we use Lemma 3.2 to write

$$\begin{aligned} y_i &= x'_i - x_i \\ &= \sum_j p_{ji}x'_j + \sigma_i - \tau_i - (\sum_j p_{ji}x_j + \sigma_i - \tau_i) \\ &= \sum_j p_{ji}x'_j - \sum_j p_{ji}x_j \\ &= \sum_j p_{ji}(x'_j - x_j) \\ &= \sum_j p_{ji}y_j. \end{aligned}$$

If you think about y as the vector with entries y_i for $i \in V$, this means that $y = yP$. Therefore y must be a multiple of the stationary distribution π. Suppose D is the constant so that $y = D\pi$. We can solve for this constant D by using the fact that the sum of the entries of π is 1. If we take the sum of the entries of y, we get

$$\sum_i y_i = \sum_i (x'_i - x_i) = \sum_i x'_i - \sum_i x_i = \mathbb{E}\Gamma' - \mathbb{E}\Gamma,$$

where the last equality follows from Lemma 3.1. This implies that $\mathbb{E}\Gamma' - \mathbb{E}\Gamma = D \cdot 1 = D$. Therefore $y = (\mathbb{E}\Gamma' - \mathbb{E}\Gamma)\pi$ and in particular $x'_i - x_i = (\mathbb{E}\Gamma' - \mathbb{E}\Gamma)\pi_i$.

Recall that h was a halting state for Γ and thus $x_h = 0$. This along with the above formula gives us that

$$x'_h = (\mathbb{E}\Gamma' - \mathbb{E}\Gamma)\pi_i.$$

Certainly $x'_h \geq 0$ and $\pi_i > 0$, therefore $\mathbb{E}\Gamma' - \mathbb{E}\Gamma \geq 0$ and equivalently

$$\mathbb{E}\Gamma \leq \mathbb{E}\Gamma'.$$

Since Γ had a halting state and Γ' was an arbitrary stopping rule, this implies Γ is mean-optimal.

It remains to show the forward direction of the theorem, that if a stopping rule Γ is mean-optimal, then it must have a halting state. In fact, it is sufficient to show there exists some mean-optimal stopping rule with a halting state. This is clear since if Γ and Γ' are both mean-optimal stopping rules, then $\mathbb{E}\Gamma = \mathbb{E}\Gamma'$, and so if $x_h = 0$, then the formula $x'_h - x_h = (\mathbb{E}\Gamma' - \mathbb{E}\Gamma)\pi_i$ implies $x'_h = 0$ as well.

There are actually many examples of stopping rules with halting states that suffice. A few are discussed in the next section. □

Notice that the halting states must be the same for all mean-optimal stopping rules Γ. If your starting distribution $\sigma = j$ and your target distribution is $\tau = \pi$, then one might instead expect that h be the state that maximizes $\mathcal{H}(j, h)$, the hitting time from j to h, since you are starting at j and may stopping at or before hitting h. However, somewhat surprisingly, a halting state h is exactly the state that maximizes $\mathcal{H}(h, j)$, the hitting time from h to j.

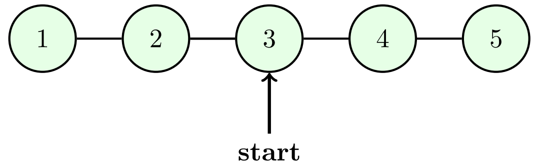
4 Examples of Optimal Stopping Rules

The four optimal stopping rules discussed here are the filling rule, the shopping rule, the threshold rule, and the local rule.

The idea behind the *filling rule* is stop as soon as possible without 'overshooting' your target distribution. (You can't stop too soon, or you'll end up close to σ instead of achieving τ exactly.) This is analogous to the greedy algorithm. An example is below.

Example 4.1. Consider the graph to the right. This is the line on 5 vertices. Suppose $\sigma = 3$ and $\tau = \pi$, the stationary distribution which for this graph is

$$\pi = \left[\begin{array}{ccccc} \frac{1}{8} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{8} \end{array} \right].$$



Then the filling rule says to stop at 3, where you start, with probability $\frac{1}{4}$. If you don't stop, then take one step and stop with probability $\frac{2}{3}$. If you still don't stop, then walk until you hit one of the endpoints.

As you can see with this example, you assign the highest possible probability of stopping without overshooting τ, but this means that once you leave a node, you definitely won't stop there (otherwise you would overshoot τ). Here, we stop at a node distance 0 from the starting point with some probability, then distance 1 from starting point with another probability, and finally stop if we are ever distance 2 from the starting point.

The *shopping rule* assigns a price to each vertex and chooses a starting budget uniformly from $[0, 1]$. The shopping rule says to stop when you find a vertex you can afford. The *threshold rule* assigns to each vertex j a time h_j. We don't stop at j if we hit j at time t before h_j, we definitely stop if we hit j at time t after $h_j + 1$, and we stop with probability $t - h_j$ otherwise. The *local rule* assigns some probability of stopping to each node depending on a set of exit frequencies $\{x_i\}_{i \in V}$ that satisfy the formula in Lemma 3.2. This rule depends only on the node you are at and not on time like the other three rules.

References

[1] L. Lovász & P. Winkler, Efficient Stopping Rules for Markov Chains (Extended Abstract), *Proc. 27th ACM Symp. on the Theory of Computing* (1995), 76–82.

[2] L. Lovász & P. Winkler, Mixing of Random Walks and Other Diffusions on a Graph *Surveys in Combinatorics, 1995*, P.Rowlinson, ed., London Math. Soc. Lecture Note Series 219, Cambridge U. Press (1995), 119–154.

[3] L. Lovász & P. Winkler, Mixing Times, *Microsurveys in Discrete Probability*, D. Aldous and J. Propp, eds., DIMACS Series in Discrete Math. and Theoretical Computer Science 41, Amer. Math. Soc., Providence RI, (1998), 85–134