6.896 Probability and Computation February 28, 2011

Lecture 8

Lecturer: Constantinos Daskalakis *Scribe: Maciej Pacula, Andreas Stuhlmüller*

NOTE: The content of these notes has not been formally reviewed by the lecturer. It is recommended that they are read critically.

1 Coupling from the Past

So far we have developed techniques that enable us to sample approximately from the stationary distribution. The goal of Coupling from the Past is to sample **exactly** from the stationary distribution. We already know how to do this naively ($\pi(x) = w(x)/Z$), but we are aiming for something less computationally expensive. The main idea behind Coupling from the Past is to try to simulate the following scenario: the Markov Chain has been running since time $-\infty$, and we are about to sample its state at present time. If this scenario were true, we would be sampling from the stationary distribution of the chain, since the chain would have already mixed by now.

2 Preliminaries: Random Function Representation of a MC

Definition 1. *Let $P(.;.)$ define an ergodic Markov Chain on Ω. A distribution $\mathcal{F}$ over functions $f : \Omega \to \Omega$ is called a Random Function Representation of P iff:*

$$Pr_{\mathcal{F}}[f(x) = y] = P(x, y)$$

In other words, the probability of mapping x to y is the transition probability $P(x, y)$ in the Markov Chain.

Example 1. *Walk on the n-cycle. Define:*

$$P(i,j) = \begin{cases} 1/2, & j = i + 1 \ (mod\ n); \\ 1/2, & j = j - 1 \ (mod\ n); \\ 0, & otherwise; \end{cases}$$

i.e. the chain goes right with probability 1/2 and left with probability 1/2.

To obtain a random function representation of the Markov Chain consider the following functions:

$$f(i) = i + 1 \ (mod\ n);$$
$$\text{and } f'(i) = i - 1 \ (mod\ n).$$

Now consider the distribution $\mathcal{F}$ that samples f with probability 1/2, and f' with probability 1/2.

Remark 1. *Observe that $\mathcal{F}$ defines a valid coupling step for every pair of states. Indeed, consider the random pair obtained through $\mathcal{F}$ as follows:*

$$(x, y) \overset{f \sim \mathcal{F}}{\to} (f(x), f(y)).$$

*It is not hard to check that the random pair obtained as above is a valid coupling of one step of the Markov Chain from states x and y. In fact, $\mathcal{F}$ defines a **grand** (or **complete**) coupling of the Markov Chain in the following sense: $\mathcal{F}$ simultaneously couples a one-step transition of the Markov Chain starting from all states in Ω (see Figure 1).*

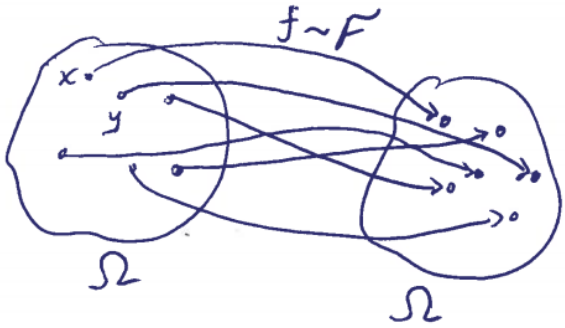

Figure 1: $\mathcal{F}$ defines a grand coupling of the Markov Chain, since it couples a one-step transition of the Markov Chain starting from all states in Ω.

Proposition 1. *Every transition matrix $P(.,.)$ has a Random Function representation.*

Proof: Let $\Omega = \{x_1, x_2, \ldots, x_n\}$ and define $F_{j,k} \stackrel{def}{=} \sum_{i=1}^{k} P(x_j, x_i)$. $F_{j,k}$ is the probability that x_j transitions to a state indexed by an index less than or equal to k. We can then define a grand coupling $\mathcal{F}$ as follows:

- pick $r \in [0,1]$ uniformly at random;
- $\forall x_j \in \Omega$, set $f(x_j) = x_k$ if $F_{j,k-1} < r \leq F_{j,k}$.

Hence, $Pr_{\mathcal{F}}[j \text{ is mapped to } k] = F_{j,k} - F_{j,k-1} = P(x_j, x_k)$.

□

Notice that, given the Random Function representation of a Markov Chain, we can figure out the Markov Chain as follows:

$$P(j,k) = Pr_{\mathcal{F}}[f(j) = k].$$

The converse is not necessarily true - a Markov Chain may have many Random Function representations. For our use of such representations in coupling from the past (discussed next), it will be crucial to choose a good one.

3 Main: Coupling from the Past

Given a Random Function Representation (RFR) of a Markov Chain, we can simulate the forward evolution of the chain for t steps as follows: pick a sequence $(f_\tau)_{\tau=0}^{t-1}$ where f_τ are i.i.d samples from the RFR $\mathcal{F}$. The t-th step of the Markov Chain is then $X_t = f_{t-1} \circ f_{t-2} \circ \ldots \circ f_0(X_0)$. Since applying a random f to the current state defines a valid step of the Markov Chain, X_t is a valid sample from the t-step evolution of the Markov Chain.

Now let us associate a random sample $f_t \sim \mathcal{F}$ to all times $t = -\infty, \ldots, +\infty$. We can then define the quantity:

$$F_i^j = f_{j-1} \circ f_{j-2} \circ \ldots \circ f_{i+1} \circ f_i, \text{ for } j \geq i.$$

F_0^t is then a forward simulation of the Markov Chain for t steps, in the sense that

$$p_{x_0}^t(x) = Pr[X_t = x | X_0 = x_0] = Pr[F_0^t(x) = x].$$

Similarly, F_{-t}^0 captures the evolution of the chain from time $-t$ to 0 (the present). We call F_{-t}^0 *simulation from the past.*

Definition 2. *The coalescence time of a Markov Chain is the following stopping time:*

$$T_c \stackrel{def}{=} \min\left\{t : F_0^t \text{ is the constant function}\right\}.$$

If F_0^t has become constant at time $t = T_c$, all states are mapped to a single state through $F_0^{T_c}$. In other words, under the grand coupling defined by $\mathcal{F}$, copies of the chain starting at time 0 from all states in Ω will meet at time T_c. Given our intuition about coupling, it is reasonable to expect that the unique element in the range of $F_0^{T_c}(\cdot)$ is distributed according to the stationary distribution π. This intuition fails miserably, as our example (discussed shortly) shows. Quite remarkably though the intuition *is true* if we consider the simulation from the past F_{-t}^0. This motivates the definition of the following stopping time.

Definition 3. *Similarly to the coalescence time for the forward simulation, define the stopping time for simulation from the past as follows*

$$T \stackrel{def}{=} \min\left\{t : F_{-t}^0 \text{ is the constant function}\right\}.$$

The main theorem of today's lecture is the following.

Theorem 1. [*Propp & Wilson '96*] *Suppose that the random function representation $\mathcal{F}$ of a Markov Chain guarantees that T is finite with probability* 1. *Then $Z_{-\infty}^0 \stackrel{def}{=} F_{-T}^0(x)$ is distributed according to π.*

Before we attempt a proof, let's gain some intuition about the difference between Coupling from the Past and the forward simulation of a Markov Chain. Consider the Markov Chain with two states α and b shown in Figure 2. This Markov Chain has the following stationary distribution:

$$\pi(\alpha) = 1/3;$$
$$\pi(b) = 2/3.$$

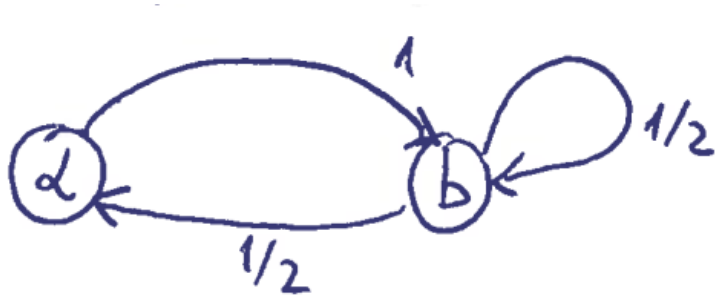

Figure 2: A simple Markov Chain with two states.

A random function representation of this chain is shown in Figure 3. Figure 4 shows a forward simulation that results in coalescence time $T_c = 3$ and the constant function $F_0^{T_c}(x) = b$. In particular, it is not hard to check the following.

Claim 1. *If $F_0^t(\cdot)$ is constant, then its range is $\{b\}$.*

On the other hand, Figure 5 shows two simulations from the past, which result in $T = 2$ and $T = 1$ and converge to the constant functions $F_{-T}^0(x) = \alpha$ and $F_{-T}^0(x) = b$, respectively. It is not hard to check that the probability that $F_{-T}^0(x) = \alpha$ and $F_{-T}^0(x) = b$ is precisely 1/3 and 2/3 respectively.

Figure 3: A random function representation of the Markov Chain from Figure 2. Each function is chosen with probability 1/2.

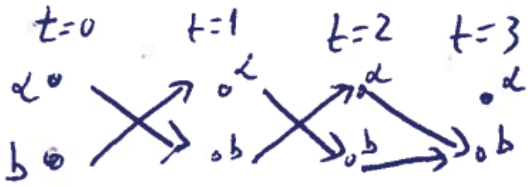

Figure 4: An example forward simulation of the Markov Chain from Figure 2, resulting in $T_c = 3$.

Figure 5: Two "from the past" simulations of the Markov Chain from Figure 2. The one on the left results in $T = 2$ and $F^0_{-T}(x) = \alpha$, while the one on the right results in $T = 1$ and $F^0_{-T}(x) = b$
.

3.1 Forward versus From the Past simulation

The crucial difference between the forward & from the past simulation is that in the forward simulation, were we to continue the simulation beyond time T_c, the unique element in the range of $F^t_0(x)$ would change. This is *not* the case for the simulation from the past, where the range remains constant even if we simulate beyond $-T$. Formally:

$F^0_{-t}(x) = F^0_{-T}(x), \forall t > T;$
$F^t_0(x)$ & $F^{T_c}_0(x)$ are not necessarily the same constant functions for $t > T_c$.

On the other hand,

Exercise (1 pt): Show that T and T_c have the same distribution. In other words, the number of steps we need to simulate forward until F^t_0 becomes a constant and the number of steps we need to simulate backwards until F^0_{-t} becomes a constant have the same distribution.

We are now ready to prove the main theorem of today's lecture, Theorem 1.

3.2 Proof of Theorem 1

Proof: Since T is finite with probability 1, $Z^0_{-\infty}$ is well-defined with probability 1. Similarly define the stopping time:

$$T' = \min\left\{t : F^1_{-t} \text{ is a constant function}\right\},$$

and define the random variable $Z^1_{-\infty} = F^1_{-T'}(x)$.

Claim 2. *T' is finite with probability 1 (and hence the random variable $Z^1_{-\infty}$ is well-defined with probability 1). Moreover, if π_0 is the distribution of $Z^0_{-\infty}$ and π_1 is the distribution of $Z^1_{-\infty}$, then $\pi_0 \equiv \pi_1$.*

Proof: The claim can be easily established via a coupling argument that shifts the time right by 1. Intuitively, both $Z^0_{-\infty}$ and $Z^1_{-\infty}$ are the constant value of a function obtained via from the past simulation of the chain from some past time $-t$ to some fixed present time 0 or 1 respectively. □

We now proceed with the proof of Theorem 1. We couple the sampling of $Z^0_{-\infty}$ and $Z^1_{-\infty}$ as follows:

- pick f_0 randomly from $\mathcal{F}$;
- pick $\{f_t\}_{t<0}$ independently at random from $\mathcal{F}$;
- wait until T' and set $Z^1_{-\infty} = F^1_{-T'}(x)$, where x can be chosen arbitrarily;
- wait until T and set $Z^0_{-\infty} = F^0_{-T}(x)$, where x can be chosen arbitrarily.

It is easy to see that this coupling satisfies the following:

- $T' \leq T$. Indeed, if T has happened, then T' must have also happened (since if F^0_{-t} is a constant function, then F^1_{-t} is also constant function);
- $Z^1_{-\infty} = f_1(Z^0_{-\infty})$.

But Claim 2 implies that $Z^1_{-\infty}$ and $Z^0_{-\infty}$ have the same distribution $\pi_1 = \pi_0$. This common distribution must hence be a fixed point of the "equation" $\pi' = f(\pi')$ for $f \sim \mathcal{F}$. The unique fixed point of this equation is the stationary distribution π of the Markov Chain. Therefore $\pi_0 = \pi_1 = \pi$. □

Remark 2. *Note that our proof crucially depended on the fact that extending the simulation from the past beyond time $-T'$ did not alter the distribution of F^1_{-t}, a property that does not hold for the forward simulation. This property is precisely what enables the theorem for the simulation from the past.*

4 An Exact Sampling Algorithm

We can now use Theorem 1 to sample exactly from the stationary distribution of a Markov Chain. The algorithm is as follows:

```
t ← 0
F_t^0 ← identity function
repeat
        t ← t − 1
        sample random f_t ∼ F
        F_t^0 ← F_{t+1}^0 ∘ f_t
until F_t^0 is a constant function
return the unique state in the range of F_t^0(.)
```

Claim 3. *Suppose $\mathcal{F}$ guarantees that coalescence time is finite with probability 1. Then with probability 1 the procedure shown above terminates and outputs a value that is distributed according to π.*

Proof: Follows immediately from Theorem 1. □

Implementation issues: The algorithm outlined above is not practical. Checking whether the function F_t^0 is constant requires us to apply it to all the states in Ω, and Ω could be prohibitively large. Nevertheless, in many settings we can check if F_t^0 is constant implicitly making the algorithm efficient.

Next time: We will see how to use Theorem 1 constructively for state spaces Ω that are endowed with a partial order. In particular, we will show that if the partial order has a unique minimum and a unique maximum element, we can construct *monotone grand couplings* that allow us to perform the check by simply checking whether the minimum and maximum elements of Ω are mapped to the same state.