Coupling From The Past (CFTP)

**Goal:** sample from the stationary distn of a MC exactly!

**How:** assume MC started at $t=-\infty$ and try to understand its value at present time by looking at "properties of recent steps of the chain."

L\rightarrow we'll try to identity such properties.

**Preliminaries:** Random function representation of MC

**Def:** Let $P$ be an ergodic MC on $\Omega$. A probability distribution $\mathcal{P}$ over functions from $\Omega$ to $\Omega$ is a random function representation of $P$ if:

\[ \forall x, y \in \Omega : \mathbb{P}_{\mathcal{P}}[ f(x) = y ] = P(x,y). \]

**Remark:** Observe that $\mathcal{P}$ defines a coupling as follows

\[ (X,Y) \rightarrow (f(X), f(Y)). \]

- In fact, it is a complete coupling since it defines a step for all states in $\Omega$ simultaneously.
e.g. random walk on n-cycle: \( \Omega = \{0, 1, \ldots, n-1\} \)

\[
P(i, j) = \begin{cases} 
\frac{1}{2}, & \text{if } j = (i+1) \text{ (mod } n) \\
\frac{1}{2}, & \text{if } j = (i-1) \text{ (mod } n) \\
0, & \text{otherwise} 
\end{cases}
\]

Random mapping representation:

consider functions: \( f(i) = i + 1 \text{ mod } n \)

\( f'(i) = i - 1 \text{ mod } n \)

\( F \) chooses \( f \) w.p.r. \( \frac{1}{2} \)

\( f' \) w.p.r. \( \frac{1}{2} \)

**Proposition:** Every transition matrix \( P \) has a random mapping representation

**Proof:** Let \( \Omega = \{x_1, x_2, \ldots, x_n\} \) and for all \( j, k \) define \( F_{j, k} = \sum_{i=1}^{k} P(x_i, x_k) \)

- Random mapping representation:
  - pick \( r \in [0, 1] \) uniformly at random
  - define \( f \) as follows:

\[
f(x_j) = x_k \text{ if } F_{j, k-1} < r \leq F_{j, k}
\]

- Clearly:

\[
Pr[f(x_j) = x_k] = \frac{F_{j, k} - F_{j, k-1}}{1} = P(x_j, x_k)
\]

**Main:** Coupling from the past

- Observe that an equivalent way to describe the evolution of the MC is to choose random functions \( f_t \sim F \) i.i.d.
  - for all \( t = 0, 1, \ldots \)
  - then:

\[
X_t = f_t \circ f_{t-2} \circ \ldots \circ f_0(X_0)
\]
Suppose \((f_i)_{i=0}^\infty\) are i.i.d. samples from \(F\).

Define: \(F_i^j = f_{j-1} \circ f_{j-2} \circ \ldots \circ f_i \circ f_{i+1} \circ f_i\), for \(j > i\).

\(F_0^t\): "forwards" simulation of MC for \(t\)-steps
\(F_0^{-t}\): evolution of MC from time \(-t\) to time 0
(i.e., from the past simulation)

**Def:** The coalescence time \(T_c\) is

\[ T_c = \min \{ t : F_0^t \text{ is a constant function} \} \]

- Since \(F_0^{T_c}\) is a constant function, i.e., in the grand coupling defined by \(F\), all paths of the MC starting at all points of \(\mathcal{X}\) have met at time \(T_c\).

- Reasonable, albeit false intuition: \(F_0^{T_c}(x)\) is distributed according to \(\pi\), for all \(x \in \mathcal{X}\).

- Remarkably, intuition is correct for the "from the past" simulation!!
Formally, define the stopping time:

\[ T = \min \{ t : F_{-t}^0 \text{ is a constant function} \} . \]

Theorem [Propp & Wilson '96]: Suppose \( T \) is finite with probability 1.

Then \( F_{-\infty}^0 := F_{-T}^0 (x) \) is distributed according to \( \pi \).

Gaining Intuition: Forward vs from the past simulation.

Consider the following MC:

\[ \begin{array}{c}
\text{stationary distn.} \\
\pi (a) = \frac{1}{3} \\
\pi (b) = \frac{2}{3}
\end{array} \]

Random function representation:

\[ \begin{array}{c}
\begin{array}{c}
\text{w pr. } \frac{1}{2}
\end{array} \\
\begin{array}{c}
\text{w pr. } \frac{1}{2}
\end{array}
\end{array} \]

\[ \begin{array}{c}
\text{example forward simulation}
\end{array} \]

\[ \begin{array}{c}
t = 0 \quad t = 1 \quad t = 2 \quad t = 3 \\
\begin{array}{c}
\text{w pr. } \frac{1}{2}
\end{array} \\
\begin{array}{c}
\text{w pr. } \frac{1}{2}
\end{array}
\end{array} \]

Claim: \( F_{-T}^0 (x) = b \)

\[ \begin{array}{c}
\text{examples from the past simulations:}
\end{array} \]

\[ \begin{array}{c}
t = 2 \quad t = 1 \quad t = 0 \\
\begin{array}{c}
\text{w pr. } \frac{1}{2}
\end{array} \\
\begin{array}{c}
\text{w pr. } \frac{1}{2}
\end{array}
\end{array} \]

\[ \begin{array}{c}
T = 2 \quad F_{-T}^0 (x) = a
\end{array} \]

\[ \begin{array}{c}
T = 1 \quad F_{-T}^0 (x) = b
\end{array} \]
Crucial difference between Forward & From the past simulation.

\[ F_t^0 (x) = F_{-T}^0 (x), \quad \forall \ t > T \]

while \( F_t^0(x) \) & \( F_{-T}^0(x) \) are not necessarily the same constant functions.

Claim: Show that \( T \) & \( Tc \) have the same distribution.

Proof of theorem: Since \( T \) is finite w/prob 1, \( Z_{-\infty}^0 \) is well-defined w/ probability 1. Similarly, define the stopping time:

\[ T' = \min \{ t : F_{-T}^1 \text{ is a constant fnf} \} \]

and \( Z_{-\infty}' = F_{-T'}^1 (x) \).

Claim 1: If \( \pi_0 \) is the distribution of \( Z_{-\infty}^0 \) and \( \pi_1 \) is \(-1-\)

\[ \pi_0 \equiv \pi_1. \]

Proof: (Shift time to the right by 1)

Both \( Z_{-\infty}^0 \) and \( Z_{-\infty}' \) is the constant value of the function obtained via from the past simulation from some past time \(-t\) to some fixed time \( 0 \) or \( 1 \) respectively.

Can couple the sampling of \( Z_{-\infty}^0 \) and \( Z_{-\infty}' \) by coupling \( F_{-T}^0 \) & \( F_{-T}^1 \) to choose the same \( f_t \) at all \( t < 0 \).

It follows that i. \( T' \leq T \) (since if \( F_{-T}^0 \) is constant, then \( F_{-T}^1 \) is constant)

ii. \( Z_{-\infty}' = f(Z_{-\infty}) \) (\#)

\[ \text{where } f \sim F \text{ is the random function chosen at} \]
An exact Sampling Algorithm

\[
\begin{align*}
& t = 0; \ F^0_t \leftarrow \text{identity function}; \\
& \text{Repeat} \\
& \quad \text{sample random } f_t \sim F \\
& \quad F^0_{t-1} \leftarrow F^0_t \circ f_t \\
& \quad t \leftarrow t - 1 \\
& \text{until } F^0_t \text{ is a constant function.} \\
& \text{return the unique state in the range of } F^0_t(\cdot).
\end{align*}
\]

Claim: Suppose \( F \) guarantees coalescence time is finite w/prob. 1. Then the above procedure terminates w/prob. 1 & the output is distributed according to \( \pi \).

Proof: follows immediately from our theorem.

Implementation Issues: to check if \( F^0_t \) is constant in the end of each iteration need to compute the value of the function for all \( x \in \Omega \), but \( \Omega \) could be huge. However, it can be implemented efficiently in many settings where these can be done implicitly.

Next time: apply coupling from the past to monotone settings.