COUPLING FROM THE PAST–TAKE HOME ASSIGNMENT

PIERRE BELLEC

Let $\mathcal{X}$ be a finite set and let $P(\cdot,\cdot)$ be an irreducible and aperiodic transition matrix on $\mathcal{X}$, with stationary distribution π.

Random mapping representation. Throughout the exam, we assume that there exists a random function $f : \mathcal{X} \to \mathcal{X}$ such that for any $x, y \in \mathcal{X}$ we have $\mathbf{P}(f(x) = y) = P(x, y)$.

Sequence $(f_t)_{t\geq 0}$ of iid random variables. Throughout the problem, let $(f_t)_{t=0,1,2,...}$ be a countable sequence of iid copies of f, indexed by the set of non-negative integers. Hence for any $t \geq 0$ we have $\mathbf{P}(f_t(x) = y) = P(x, y)$ for all deterministic $x, y \in \mathcal{X}$.

Assumption (C). The coalescence time τ_c, defined by

$$\tau_c = \min\{t \geq 1 : f_t \circ f_{t-1} \circ ... \circ f_1 \text{ is a constant function}\},$$

if this set is nonempty and $\tau_c = +\infty$ if this set is empty, is finite with probability 1.

Part 0.

1. Give an example of a random mapping representation of some aperiodic and irreducible Markov Chain such that Assumption (C) does not hold, that is, an example with $\mathbf{P}(\tau_c = +\infty) > 0$.
2. Prove that if for some integer $t > 0$, $\mathbf{P}(\tau_c \leq t) > 0$ then $\mathbf{P}(\tau_c = +\infty) = 0$.
3. Deduce that $\mathbf{P}(\tau_c = +\infty)$ is equal to either 0 or 1.

In the rest of the exam, we assume that Assumption (C) always holds so that $\mathbf{P}(\tau_c = +\infty) = 0$.

The goal of these questions is to study possible schemes to sample a random variable with distribution π. The first idea that may come to mind is to apply the functions f_t successively until τ_c, then output the current state.

Part I (Forward). Define a grand coupling as follows. For any $x \in \mathcal{X}$, define $X_0^x = x$ and $X_t^x = f_t(X_{t-1})$, so that $X_t^x = f_t \circ f_{t-1} \circ ... \circ f_1(x)$. The coalescence time τ_c is the first time all the Markov Chains $(X_t^x)_{x\in\mathcal{X}}$ have met.

4. Give an example of a random mapping representation of some aperiodic and irreducible Markov Chain on $\mathcal{X} = \{0, 1, 2\}$ such that the random variable $X_{\tau_c}^{x_0}$ for $x_0 \in \mathcal{X}$ is NOT distributed according to π. *Note that $X_{\tau_c}^x$ is the same for any $x \in \mathcal{X}$.*

Part II (Backward). The previous "forward" scheme thus fails. We now study a "backward" scheme, known in the literature as "coupling from the past".

5. *Backward vs. Forward.* If the answer is "always true" prove it, otherwise give a simple counterexample.
    a. Is it always true that if $f_3 \circ f_2 \circ f_1$ is constant, then $f_4 \circ f_3 \circ f_2 \circ f_1 = f_3 \circ f_2 \circ f_1$?
    b. Is it always true that if $f_1 \circ f_2 \circ f_3$ is constant, then $f_1 \circ f_2 \circ f_3 \circ f_4 = f_1 \circ f_2 \circ f_3$?
6. Define $M = \min\{t \geq 1 : f_1 \circ f_2 \circ \cdots \circ f_t$ is a constant function $\}$ if this set is nonempty, and $M = +\infty$ otherwise. Prove that M is finite with probability one under Assumption (C).

7. Prove that $f_1 \circ f_2 \circ \cdots \circ f_M(x) = f_1 \circ f_2 \circ \cdots \circ f_M \circ \cdots \circ f_{M+k}(x)$ for any $x, y \in \mathcal{X}$ and any integer $k \geq 1$.

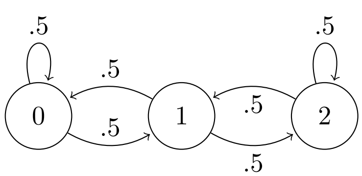
8. *Most important and difficult question of the problem. Make sure to attend this question and be clear and rigourous in your answer.*
   The goal of this question is to prove that $\hat{X} = f_1 \circ f_2 \circ \cdots \circ f_M(x_0)$ is distributed according to the invariant distribution π.
   a. Define a sequence of iid functions $g_1, ..., g_t, ...$ by $g_t = f_{t-1}$ for all $t \geq 1$.
      Define $N = \min\{t \geq 1 : g_1 \circ g_2 \circ \cdots \circ g_t$ is a constant function $\}$ and $\hat{Y} = g_1 \circ ... \circ g_N(x_0)$.
      Prove that $(N, \hat{Y})$ has the same distribution as $(M, \hat{X})$.
   b. Prove that $M + 1 \geq N$ always holds.
   c. Is it always true that $\hat{Y} = f_0(\hat{X})$?
   d. Prove that f_0 is independent of $\hat{X}$.
   e. Prove that $\mathbf{P}(f_0(\hat{X}) = y) = \mathbf{P}(\hat{X} = y)$.
   f. Conclude.

9. Deduce from the previous question an algorithm that outputs a random variable distributed according to π, by sequentially generating iid random functions $f_1, f_2, f_3, ...$

Part III (Another idea). Consider now the following algorithm.

- Algorithm 2:
  a. Set $t = 1$.
  b. Generate $f_1^{(t)}, f_2^{(t)}, f_3^{(t)}, ..., f_t^{(t)}$ iid copies of the random function f independently of all previous iterations of the algorithm
  c. – If $f_1^{(t)} \circ \cdots \circ f_t^{(t)}$ is a constant function, then output its unique value and stop the algorithm.
     – Otherwise, throw away $f_1^{(t)}, f_2^{(t)}, f_3^{(t)}, ..., f_t^{(t)}$, increase t by one, i.e., set $t := t + 1$ and go to step b.

10. By studying the Markov Chain defined on $\mathcal{X} = \{0, 1, 2\}$ with the random function f defined by $\mathbf{P}(f(0) = 1, f(1) = 2, f(2) = 2) = 1/2$, $\mathbf{P}(f(0) = 0, f(1) = 0, f(2) = 1) = 1/2$, show that the algorithm of Algorithm 2 does NOT output a random variable distributed with respect to π. (*Hint: you may, for instance, show that if $\hat{Y}$ is the random variable output by Algorithm 2 then $\mathbf{P}(\hat{Y} \in \{0, 2\})$ is too large.*)



Part IV.

11. Consider a Markov Chain on $\mathcal{X} = \{0, 1, ..., n\}$ with transition probabilities defined by $P(i, \min(i + 1, n)) = 1/2$, $P(i, \max(i - 1, 0)) = 1/2$ and 0 elsewhere, as in the graph below.
    a. Propose a random mapping representation $f : \mathcal{X} \to \mathcal{X}$ such that $\mathbf{P}(f(i) = j) = P(i, j)$ for any $i, j \in \mathcal{X}$. The proposed random mapping representation should satisfy that $f(i) \leq f(j)$ always holds for any $i \leq j$.
    b. Show that the event $\{M \leq t\}$ can be simply expressed in terms of $f_1 \circ f_2 \circ f_3 \circ ... \circ f_t(0)$ and $f_1 \circ f_2 \circ f_3 \circ ... \circ f_t(n)$.

c. Explain why, in this case and thanks to question $f(i) \le f(j)$ always holds for any $i \le j$, the algorithm of Part II that outputs a random variable with distribution π can be greatly simplified. Figure 25.2 in the book illustrates this.