Coupling from the Past: Introduction

Aidan Jameson

University of Utah

May 19, 2020

Notation

- $(X_t) \cong \mathcal{P} \implies (X_t)$ is a Markov chain with transition matrix $\mathcal{P}$

- $(X, Y) \propto (\mu, \nu) \implies (X, Y)$ is a coupling of (μ, ν)
- $(X_t, Y_t) \propto \mathcal{P} \implies (X_t, Y_t)$ is a coupling of Markov chains with transition matrix $\mathcal{P}$

- Generally, unless otherwise stated, the state space of any Markov chains will be denoted Ω (finite)
- Markov chains will also occasionally only be referred to by their transition matrix

Markov Chain Convergence Theorem

Theorem

Given any $\mathcal{P}$ ergodic Markov Chain transition matrix with stationary distribution π. Then there exists $\alpha \in (0,1)$ and $C > 0$ such that

$$\max_{x \in \Omega} \left\| \mathcal{P}^t(x, \cdot) - \pi \right\|_{TV} \leq C\alpha^t$$

As such, from here on out we will assume all Markov chains are ergodic unless explicitly stated

Metropolis Chains

Proposition

For any probability distribution π on Ω, given any irreducible transition matrix Φ the Markov chain with transition matrix

$$\mathcal{P}(x,y) = \begin{cases} \Phi(x,y)a(x,y) & y \neq x \\ 1 - \sum_{z \in \Omega \setminus \{x\}} \Phi(x,z)a(x,z) & y = x \end{cases}$$

$$\textit{where } a(x,y) = \min\left(1, \frac{\pi(y)\Phi(y,x)}{\pi(x)\Phi(x,y)}\right)$$

has stationary distribution π.

In practice $a(x,y)$ is known as the acceptance probability so the chain can be simulated by running the initial chain and accepting transitions with probability according to $a(x,y)$.

Coupling

Proposition

For any μ and ν probability distributions on state space Ω (finite):

$$\|\mu - \nu\|_{TV} = \inf \{\mathbb{P}\{X \neq Y\} : (X, Y) \propto (\mu, \nu)\}$$

Proposition

Further $\exists(X^*, Y^*) \propto (\mu, \nu)$ such that

$$\|\mu - \nu\|_{TV} = \mathbb{P}\{X^* \neq Y^*\}$$

Coupling

Theorem

Let $(X_t, Y_t) \propto \mathcal{P}$ *such that* $X_0 = x$ *and* $Y_0 = y$ *(we may assume that* $X_t = Y_t \implies X_{t+1} = Y_{t+1}$ *as any coupling may be modified to do so). Let* $\tau_{couple} = \min\{t : X_t = Y_t\}$*. Then*

$$\left\|\mathcal{P}^t(x,\cdot) - \mathcal{P}^t(y,\cdot)\right\|_{TV} \leq \mathbb{P}_{x,y}\{\tau_{couple} > t\}$$

Random Mapping Representation

Theorem

For any Markov chain on states space Ω with transition matrix $\mathcal{P}$ there exists $f : \Omega \times \Lambda \to \Omega$ and a random variable Z on Λ such that

$$\mathbb{P}\{f(x, Z) = y\} = \mathcal{P}(x, y)$$

Observation

With f and Z as above let $Z_1, Z_2, \ldots$ be i.i.d. with the same distribution as Z. For any $x \in \Omega$ define inductively the Markov chain $(X_t^x) \cong \mathcal{P}$ by $X_0^x = x$ and $X_t = f(X_{t-1}^x, Z_t)$. Then $(X_t^x : x \in \Omega)$ is a grand coupling over Ω and $\mathcal{P}$

Coupling From The Past

Given an ergodic Markov chain on Ω with transition matrix $\mathcal{P}$ and stationary distribution π.

- Let $\mathbb{P}$ be a probability measure on the space of functions from Ω to itself (Ω^{Ω}) such that $\mathbb{P}\{f(x) = y\} = \mathcal{P}(x, y)$.
- We operate under the assumption that the chain has been run long enough to reach its stationary distribution and declare this time $t = 0$.
- Starting at some time $-T$ in the past we create a grand coupling on Ω beginning at $-T$ via inductively defining $X^x_{t+1} = f_t(X^x_t)$ ($X^x_{-T} = x$) $\forall x \in \Omega$ where $f_{-T}, \ldots, f_{-1}$ is an i.i.d sequence of random maps from Ω to itself distributed according to $\mathbb{P}$

Sample one such i.i.d. sequence $f_{-T}, \ldots, f_{-1}$ according to $\mathbb{P}$.

Coupling From The Past

Theorem

If $(f_{-1} \circ f_{-2} \circ \cdots \circ f_{-T})(\Omega) = \{x\}$ *(for some* $x \in \Omega$*). Then* x *is a sample from the stationary distribution* π

Coupling From The Past

Theorem

If $(f_{-1} \circ f_{-2} \circ \cdots \circ f_{-T_0})(\Omega) = \{x\}$ (for some $x \in \Omega$). Then x is a sample from the stationary distribution π

Proof.

Take $\left(f_t \in \Omega^\Omega : t \in \mathbb{Z}\right)$ to be an i.i.d. sequence of random maps with distribution $\mathbb{P}$. For any $T_1 < T_2$ let $F_{T_1}^{T_2} = f_{T_2-1} \circ f_{T_2-2} \circ \cdots \circ f_{T_1}$. By construction and the Markov Chain Convergence Theorem $\lim_{T\to\infty} \mathbb{P}\{F^0_{-T}(x) = y\} = \lim_{T\to\infty} \mathcal{P}^T(x,y) = \pi(y)$. Let $T_0 = \inf\{t > 0 : |F^0_{-t}(\Omega)| = 1\}$ (Unless f and $\mathbb{P}$ are poorly chosen, $T_0 < \infty$ *a.s.* although this needs to be checked). Now since $F^0_{-T_1} = F^0_{-T_2} \circ F^{-T_2}_{-T_1}$ for T large enough $F^0_{-T} = F^0_{-T_0} \circ F^{-T_0}_{-T} = F^0_{-T_0}$ and thus $\mathbb{P}\{F^0_{-T_0}(x) = y\} = \lim_{T\to\infty} \mathbb{P}\{F^0_{-T}(x) = y\} = \pi(y)$ $\square$

Monotone Coupling From The Past

Corollary

If Ω is equipped with partial order $\preceq$ such that there exists $\hat{1}, \hat{0} \in \Omega : \hat{0} \preceq x \preceq \hat{1}\ \forall x \in \Omega$. Then given that the coupling preserves the partial order (ie. $f_t(x) \preceq f_t(y)\ \forall t$ whenever $x \preceq y$) it suffices to find $-T_0$ such that $\left(f_{-1} \circ f_{-2} \circ \cdots \circ f_{-T_0}\right)\left(\hat{1}\right) = \left(f_{-1} \circ f_{-2} \circ \cdots \circ f_{-T_0}\right)\left(\hat{0}\right) = x$ for some $x \in \Omega$.

Applications

- Determining the stationary distribution when it is not known
- Perfectly sampling from the stationary distribution

The Ising Model

- Equipped with a partial order
- Glaubner dynamics has the Gibbs stationary distribution
- $\mathbb{E}[T_0] = O(t_{\text{mix}}(\frac{1}{4}) \ln |V|)$