Coupling from the Past: Introduction

Aidan Jameson

University of Utah

May 26, 2024

CTFP

Recall. Given an ergodic Markov chain on Ω with transition matrix $\mathcal{P}$ and stationary distribution π. A coupling from the past (CTFP) is the process (F^0_{-t}) with coalescing time T_0 where:

- $F^0_{-t} = f_{-1} \circ f_{-2} \circ \cdots \circ f_{-t}$
  - where $f_{-1}, f_{-2}, \ldots$ are i.i.d. random variables taking values in the space of functions from Ω to itself (Ω^Ω)
  - and such that $\mathbb{P}\{f_{-i}(x) = y\} = \mathcal{P}(x, y)\ \forall x, y \in \Omega$
  - If there exists partial order $\preceq$ on Ω with unique maximum and minimum elements, then a monotone CTFP holds if $f_i(x) \preceq f_i(y)\ \forall i$ whenever $x \preceq y$
- $T_0 = \min\left\{t > 0 : \left|F^0_{-t}(\Omega)\right| = 1\right\}$ with $T_0 < \infty$ *a.s.*

Then $F^0_{T_0}(x) \stackrel{d}{=} \pi\ \forall x \in \Omega$. Thus, sampling $F^0_{T_0}(x)$ samples π.

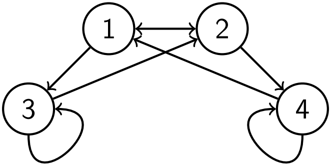
Coupling via Random Mapping Representation

Counter Example

Let $\Omega = \{1, 2, 3, 4\}$. Define an ergodic Markov chain by transition matrix

$$\mathcal{P} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{pmatrix}$$

Then the random mapping representation of $\mathcal{P}$ will not couple if the coupling is started from 1 and 2 (or 3 and 4) given the ordering in the representation preserves the ordering of Ω.



Notation

- $(\Omega, \preceq)$ denotes a partially ordered state space where it is assumed that there exists unique maximum and minimum elements $\hat{1}$ and $\hat{0}$ resp.
    - We also let L denote the length of the longest ordered sequence under $\preceq$ of elements in Ω
    - If Ω is partially ordered we will usually assume any CFTP is a monotone CTFP
- The notation surrounding any CTFP will be the same as in the previous slide

Bounds On Coupling Time

Theorem

Given $(\Omega, \preceq)$ and a Markov chain $\mathcal{P}$ with stationary distribution π. If there exists a Monotone CFTP then

$$\frac{\mathbb{P}\{T_0 > t\}}{L} \leq \max_{x,y\in\Omega} \left\|\mathcal{P}^t(x,\cdot) - \mathcal{P}^t(y,\cdot)\right\|_{TV} \leq \mathbb{P}\{T_0 > t\}$$

Bounds On Coupling Time

Theorem

If K_1 and K_2 are integer-valued random variables, then for any CTFP

$$\mathbb{P}\{T_0 > K_1 + K_2\} \leq \mathbb{P}\{T_0 > K_1\} \cdot \mathbb{P}\{T_0 > K_2\}$$

Bounds On Coupling Time

Lemma

$$t\mathbb{P}\{T_0 > t\} \leq \mathbb{E}T_0 \leq \frac{t}{\mathbb{P}\{T_0 \leq t\}}$$

Lemma

If there is a monotone CFTP on $(\Omega, \preceq)$

$$\mathbb{E}T_0 \leq 2t_{mix}\left(e^{-1}\right)(1 + \log(L))$$

Bounds On Coupling Time

Theorem

Let $T_1, \ldots, T_k, T_0$ be independent samples of the coupling time. Then

$$\mathbb{P}\{T_0 > j \max(T_1, \ldots, T_k)\} \leq \frac{j!k!}{(j+k)!}$$

Efficient Implementation

Proposition

If we take test coupling times $T'_0, T'_1, \ldots$ *where* $T'_n = rT'_{n-1}$. *Then*

- $r = 2$ *minimizes the worst-case number of computation steps*
- $r = e$ *minimizes the expected number of computation steps*

Efficient Implementation

Remark

The benefits of running with exponential growth $r = e$ is minimal compared to $r = 2$. As such, $r = 2$ is often used instead to slightly simplify computations but also potentially reduce impatience bias.

Fill's Algorithm

We are given $(\Omega, \preceq)$ (which possesses unique maximum and minimum elements) and an ergodic markov chain on Ω with transition matrix $\mathcal{P}$ and stationary distribution π

Recall. The time reversal of $\mathcal{P}$ is the markov chain governed by the transition matrix $\tilde{\mathcal{P}}(x, y) = \frac{\pi(y)\mathcal{P}(y,x)}{\pi(x)}$

Fill's Algorithm

Definition. μ and ν probability measures on Ω, then $\mu \preceq \nu$ stochastically if $\mu(I) \geq \nu(I)$ for all order ideals I (i.e. $y \preceq x \in I \implies y \in I$).

Definition. A transition matrix $\mathcal{P}$ is monotone if $\mathcal{P}(x, \cdot) \preceq \mathcal{P}(y, \cdot)$ (stochastically) whenever $x \preceq y$.

Fill's Algorithm

Definition. μ and ν probability measures on Ω, then $\mu \preceq \nu$ stochastically if $\mu(I) \geq \nu(I)$ for all order ideals I (i.e. $y \preceq x \in I \implies y \in I$).

Definition. A transition matrix $\mathcal{P}$ is monotone if $\mathcal{P}(x, \cdot) \preceq \mathcal{P}(y, \cdot)$ (stochastically) whenever $x \preceq y$.

Lemma

$\mu \preceq \nu$ stochastically if and only if there exists an upward kernel K such that $\nu = \mu K$.
In this case, K is a transition matrix such that for all $x \in \Omega$, the measures $K(x, \cdot)$ are supported on $\{y \in \Omega : y \succeq x\}$

Fill's Algorithm

It is assumed that $\mathcal{P}$ is ergodic and $\tilde{\mathcal{P}}$ is monotone on $(\Omega, \preceq)$ (where there is a unique maximum and minimum element). We let $K_{x,y}$ denote the upward kernel between $\mathcal{P}(x, \cdot) = \mathcal{P}(y, \cdot)K_{x,y}$ where $x \preceq y$.

Independently for $T = 1, 2, 4, 8, \ldots$, let $(X_t) \cong \mathcal{P}$ be the chain beginning at $X_0 = \hat{0}$. If we let $X_T = z$, consider the time reversal $(\tilde{X}_0, \ldots, \tilde{X}_T) = (X_0, \ldots, X_T)$. Build $(\tilde{Y}_t)$ such that $\tilde{Y}_0 = \hat{1}$ and where if $\tilde{X}_{t+1} = x'$, $\tilde{X}_t = x$, then transition from $\tilde{Y}_t = y$ to $\tilde{Y}_{t+1} = y'$ with probability $K_{x,y}(x', y')$. If $\tilde{Y}_T = \hat{0}$ accept z, else reject the sample.

Fill's Algorithm

Lemma

$$\mathbb{P}\{\tilde{Y}_T = \hat{0} | \tilde{X}_0 = z, \tilde{X}_T = \hat{0}, \tilde{Y}_0 = \hat{1}\} = \frac{\tilde{\mathcal{P}}^T(\hat{1}, \hat{0})}{\tilde{\mathcal{P}}^T(z, \hat{0})}$$

What we are doing is rejection sampling on $\tilde{\mathcal{P}}^T(\hat{0}, \cdot)$ with $c = \frac{\pi(\hat{0})}{\tilde{\mathcal{P}}^T(\hat{1}, \hat{0})}$ (which is a bound since the chain is monotone).