25 Mar 2026

# Mixing Times and Couplings

**Def.** For a Markov chain with stationary distrib $\pi_*$ and approx parameter $\varepsilon > 0$, the <u>$\varepsilon$-mixing time</u> is the least $\tau(\varepsilon)$ s.t.

$\forall t \geq \tau(\varepsilon)$ $\forall \pi_0$ (initial state distrib)

if $\pi_t$ denotes marginal distrib of state at time $t$

then $\| \pi_t - \pi_* \|_1 \leq \varepsilon$.

**DIGRESSION.** $\| \pi_t - \pi_* \|_1$ as a function of $t$.

Assume transition matrix $P$ is <u>ergodic</u>.

$\forall$ states $x, y$, $\forall$ sufficiently large $t$, $\exists$ a $t$-step walk from $x$ to $y$ in state transition graph.

$\exists\ s < \infty$ s.t. $P^s > 0$ ← every entry of matrix is $> 0$.

$P$ is row-stochastic: $P \cdot \vec{1} = \vec{1}$.

$P^s > 0 \Longrightarrow$ let $U = \frac{1}{N} \mathbb{1}\mathbb{1}^T =$

$$N \overset{N}{\begin{bmatrix} \frac{1}{N} & \cdots & \frac{1}{N} \\ \vdots & \ddots & \vdots \\ \frac{1}{N} & \cdots & \frac{1}{N} \end{bmatrix}}$$

$P^s = \delta \cdot U + (1-\delta) \cdot R$ $\quad$ $\delta > 0$

$R$ is non-negative, $R \cdot \mathbb{1} = \mathbb{1}$.

Observations.

Ⓐ $$R \cdot U = U$$

Ⓑ If $\pi_1$ and $\pi_2$ are two distributions,

$$(\pi_1 - \pi_2) \cdot U = 0,$$

$\pi_1$ is a distribution $\Longrightarrow$ $\pi_1 \cdot \vec{1} = 1.$

$\pi_2$ " — — " $\Longrightarrow$ $\pi_2 \cdot \vec{1} = 1.$

$$(\pi_1 - \pi_2) \cdot \vec{1} = 0.$$

If Markov chain with matrix $P$ starts out in distrib $\pi_0$ and ends up in distrib $\pi_t = \pi_0 P^t$ after $t$ steps, how far is $\pi_t$ from $\pi_*$?

$$\pi_t = \pi_0 P^t$$

Assume $t = s \cdot k$, $k \in \mathbb{N}$

$$\pi_* = \pi_* P^t$$

$$(\pi_t - \pi_*) = (\pi_0 - \pi_*) P^t$$

$$= (\pi_0 - \pi_*) \left[ \delta U + (1-\delta) R \right]^k$$

All eigenvals here $|\lambda| \leq 1$.

$$= (1-\delta)^k (\pi_0 - \pi_*) R^k$$

$\Rightarrow \| \pi_t - \pi_* \| \to 0$ exponentially fast

in $k = \frac{t}{s}$.

---

Dependence of $\tau(\varepsilon)$ on $\varepsilon$ is really mild: $\Theta(\log \frac{1}{\varepsilon})$.

Dependence on ~~state~~ state space size varies a lot depending on the transition rule.

**Idea for bounding $\tau(\varepsilon)$: Coupling**

Consider two parallel executions

$(X_0, X_0'), \ldots (X_t, X_t'), \ldots$

$X_0$ drawn from $\pi_0$

$X'_0$ drawn from $\pi_*$

$$\left.\begin{array}{l} X_0, X_1, X_2, \ldots \\ X'_0, X'_1, X'_2, \ldots \end{array}\right\} \text{copies of Markov chain}$$

Correlate state transitions to enhance $\Pr(X_t = X'_t)$ as $t$ grows large.

<u>Ex.</u> Lazy random walk on $\{0,1\}^n$.

States $= \{0,1\}^n$

$$P_{xy} = \begin{cases} 0 & \text{if } \|x-y\|_1 > 1 \\ \frac{1}{2} & \text{if } x = y \\ \frac{1}{2n} & \text{if } \|x-y\|_1 = 1. \end{cases}$$

"Pick random neighbor.
Transition to it with prob $\frac{1}{2}$.
Else stay in place."

"Pick random index, $i_t \in [n]$.
Pick random bit, $b_t \in \{0,1\}$.
Overwrite $(i_t)^{th}$ bit of $x_t$
with $b_t$ to form $x_{t+1}$."

$x_0 \in \{0,1\}^n$ sampled from $\pi_0$

$x'_0 \in \{0,1\}^n$ uniformly random,

Transition from $(x_t, x'_t)$ to
$(x_{t+1}, x'_{t+1})$: choose same $i_t$
and $b_t$ in both chains

00000 11010

$i_0 = 4,\ b_0 = 1$

00010 11010

$i_1 = 1,\ b_1 = 1$

10010 11010

$i_2 = 4,\ b_2 = 0$

10000 11000

Bounded above by $\varepsilon$ for all sufficiently large $n$.

**Lemma.** After $t > 2n\ln(n)$ state transitions $\Pr(X_t \neq X'_t) = o_n(1)$.

**Reason.** Coupon collector problem.

**Lemma.** If $\pi_t$ denotes distrib of $X_t$ and $\pi_*$ is distrib of $X'_t$,

$$\Pr(X_t \neq X'_t) \geq \frac{1}{2} \|\pi_t - \pi_*\|_1 .$$

**Proof.** $\Pr(X_t \neq X'_t)$

$$= \sum_{x \in \mathcal{X}} \sum_{y \neq x} \Pr(X_t = x \wedge X'_t = y)$$

Proof is not hard.

See book, or GPT, or work it out yourself.