MTH 565

Lectures 9 - 19

Yevgeniy Kovchegov
Oregon State University

Topics:

- Recurrent and transient states.
- Reversible Markov chains.
- Martingales.
- Stopping time.
- Optional Stopping Theorem.
- Harmonic functions.

- First passage probability.
- Markov Chain Monte Carlo (MCMC).
- Mixing times.

Recurrent and transient states.

We will use the following notations:

$P_x(A) = P(A \,|\, X_0 = x)$ and $E_x[Y] = E[Y \,|\, X_0 = x]$.

For $x \in S$, consider the first hitting time

$$T_x = \min\{t \geq 1 \,:\, X_t = x\}.$$

Definition. A state $x \in S$ is said to be recurrent if

$$P_x(T_x < \infty) = 1.$$

A recurrent state $x \in S$ is positive recurrent if

$$E_x[T_x] < \infty.$$

Otherwise it is null recurrent.

Definition. A state $x \in S$ is said to be transient if

$$P_x(T_x < \infty) < 1.$$

Recurrent and transient states.

Example. All states in an irreducible Markov chain over a finite state space S are positive recurrent.

Example. Let $S = \{1, 2, 3\}$ and $P = \begin{bmatrix} 0.2 & 0.8 & 0 \\ 0 & 0.6 & 0.4 \\ 0 & 0.5 & 0.5 \end{bmatrix}$

Then, 1 is a transient state, while 2 and 3 are positive recurrent states.

Example (Simple Random Walk on $\mathbb{Z}^d$). Consider a simple nearest-neighbor random walk on $S = \mathbb{Z}^d$ with transition probabilities equal $\frac{1}{2d}$ for each neighbor state. Then,

- all vertices of $S = \mathbb{Z}^d$ are null recurrent if $d = 1, 2$;
- all vertices of $S = \mathbb{Z}^d$ are transient if $d \geq 3$.

Stationary distribution.

Definition. A state $x \in S$ is said to be positive recurrent if

$$E_x[T_x] < \infty$$

The following is a version of ergodicity theorem for a general discrete state space S.

Theorem (Ergodicity). Consider an irreducible homogeneous Markov chain over a discrete state space S. If all of its states are positive recurrent, then there exists a unique stationary distribution π such that

$$\pi(x) = \frac{1}{E_x[T_x]}.$$

Furthermore, if the Markov chain is aperiodic,

$$\lim_{t \to \infty} p_t(x, y) = \pi(y) \qquad \forall x, y \in S.$$

Stationary distribution.

Theorem (Ergodicity). Consider an irreducible homogeneous Markov chain over a discrete state space S. If all of its states are positive recurrent, then there exists a unique stationary distribution π such that

$$\pi(x) = \frac{1}{E_x[T_x]}.$$

Furthermore, if the Markov chain is aperiodic,

$$\lim_{t\to\infty} p_t(x,y) = \pi(y) \qquad \forall x, y \in S.$$

The **probabilistic proof** can be done in two steps. First, proving existence/uniqueness, and then establishing convergence.

Lemma (Existence/Uniqueness). Consider an irreducible homogeneous Markov chain over a discrete state space S, all of whose states are positive recurrent. Then

$$\pi(x) = \frac{1}{E_x[T_x]}$$

is the unique stationary distribution.

Lemma (Convergence). Consider an irreducible aperiodic homogeneous Markov chain over a discrete state space S. If all of its states are positive recurrent, then

$$\lim_{t\to\infty} p_t(x,y) = \pi(y) \qquad \forall x, y \in S,$$

where π is the unique stationary distribution.

Stationary distribution.

Lemma (Existence/Uniqueness). Consider an irreducible homogeneous Markov chain over a discrete state space S, all of whose states are positive recurrent. Then

$$\pi(x) = \frac{1}{E_x[T_x]}$$

is the unique stationary distribution.

Proof. For a given $x \in S$, let

$$\nu_x(y) = \sum_{n=0}^{\infty} P_x(X_n = y,\, T_x > n) = E_x\left[\sum_{n=0}^{\infty} \mathbf{1}_{\{X_n=y,\, T_x>n\}}\right]$$

be the mean number of visits to state $y \in S$ between the times 0 and T_x. Then,

$$\sum_{z\in S} \nu_x(z)p(z,y) = \nu_x(y), \quad \text{where} \quad \nu_x(x) = 1.$$

For a given $x \in S$, let $\nu_x(y) = \sum_{n=0}^{\infty} P_x(X_n = y,\, T_x > n)$.

Then,

$$\sum_{z \in S} \nu_x(z) p(z,y) = \nu_x(y), \quad \text{where} \quad \nu_x(x) = 1.$$

Indeed, for all $y \neq x$, we have

$$\sum_{z \in S} \nu_x(z) p(z,y) = \sum_{z \in S} \sum_{n=0}^{\infty} P_x(X_n = z,\, T_x > n) P_x(X_{n+1} = y \,|\, X_n = z)$$

$$= \sum_{n=0}^{\infty} \sum_{z \neq x} P_x(X_n = z,\, T_x > n) P_x(X_{n+1} = y \,|\, X_n = z,\, T_x > n)$$

$$= \sum_{n=0}^{\infty} P_x(X_{n+1} = y \,|\, T_x > n) = \sum_{n=0}^{\infty} P_x(X_{n+1} = y \,|\, T_x > n+1) = \nu_x(y)$$

by Markov property, as $\{T_x > n-1\}$ is determined by $X_0, \ldots, X_{n-1}$, and for $z \neq x$,

$$\{X_n = z,\, T_x > n\} = \{X_n = z,\, T_x > n-1\}.$$

For a given $x \in S$, let $\nu_x(y) = \sum_{n=0}^{\infty} P_x(X_n = y, T_x > n)$.

Then,

$$\sum_{z \in S} \nu_x(z) p(z,y) = \nu_x(y), \quad \text{where} \quad \nu_x(x) = 1.$$

In case of $y = x$, we have

$$\sum_{z \in S} \nu_x(z) p(z,x) = \sum_{z \in S} \sum_{n=0}^{\infty} P_x(X_n = z, T_x > n) P_x(X_{n+1} = x \,|\, X_n = z)$$

$$= P_x(X_1 = x) + \sum_{n=1}^{\infty} \sum_{z \neq x} P_x(X_n = z, T_x > n) P_x(X_{n+1} = x \,|\, X_n = z, T_x > n)$$

$$= P_x(X_1 = x) + \sum_{n=1}^{\infty} P_x(X_{n+1} = x, T_x > n) = P_x(T_x < \infty) = 1 = \nu_x(x)$$

as x is a recurrent state.

Stationary distribution.

Proof (continued). Notice that

$$\sum_{y\in S} \nu_x(y) = \sum_{y\in S}\sum_{n=0}^{\infty} P_x(X_n = y,\, T_x > n) = \sum_{n=0}^{\infty} P_x(T_x > n) = E_x[T_x].$$

Thus $\displaystyle\sum_{z\in S} \nu_x(z)p(z,y) = \nu_x(y)$ yields $\displaystyle\pi(y) = \frac{\nu_x(y)}{E_x[T_x]}$

is a stationary measure $\forall x \in S$. The existence of stationary distribution follows.

Since we have shown the existence of a stationary distribution, our next goal is to show it is unique. Consider a stationary distribution π.

Stationary distribution.

Proof (continued). For any pair $a, b \in S$, a stationary distribution π should satisfy

$$\pi(b) = \sum_{x_1 \in S} \pi(x_1) p(x_1, b) = \pi(a) p(a, b) + \sum_{x_1 \neq a} \pi(x_1) p(x_1, b)$$

Similarly, $$\pi(x_1) = \pi(a) p(a, x_1) + \sum_{x_2 \neq a} \pi(x_2) p(x_2, x_1)$$

Substituting, we have

$$\pi(b) = \pi(a) p(a, b) + \pi(a) \sum_{x_1 \neq a} p(a, x_1) p(x_1, b) + \sum_{\substack{x_1 \neq a \\ x_2 \neq a}} \pi(x_2) p(x_2, x_1) p(x_1, b),$$

where $$\pi(x_2) = \pi(a) p(a, x_2) + \sum_{x_3 \neq a} \pi(x_3) p(x_3, x_2)$$

Stationary distribution.

Proof (continued). We have

$$\pi(b) = \pi(a)p(a,b) + \pi(a)\sum_{x_1\neq a} p(a,x_1)p(x_1,b) + \sum_{\substack{x_1\neq a\\ x_2\neq a}} \pi(x_2)p(x_2,x_1)p(x_1,b),$$

where $\quad \pi(x_2) = \pi(a)p(a,x_2) + \sum_{x_3\neq a} \pi(x_3)p(x_3,x_2)$

Substituting, we have

$$\pi(b) = \pi(a)p(a,b) + \pi(a)\sum_{x_1\neq a} p(a,x_1)p(x_1,b) + \pi(a)\sum_{\substack{x_1\neq a\\ x_2\neq a}} p(a,x_2)p(x_2,x_1)p(x_1,b)$$

$$+ \sum_{\substack{x_1\neq a\\ x_2\neq a\\ x_3\neq a}} \pi(x_3)p(x_3,x_2)p(x_2,x_1)p(x_1,b) \quad \text{and so on} \ldots$$

Proof (continued). After n iterations we have

$$\pi(b) = \pi(a)p(a,b) + \pi(a)\sum_{x_1\neq a} p(a,x_1)p(x_1,b)$$

$$+\ldots+\pi(a)\sum_{\substack{x_1\neq a\\ \cdots\\ x_{n-1}\neq a}} p(a,x_{n-1})p(x_{n-1},x_{n-2})\ldots p(x_1,b)$$

$$+\sum_{\substack{x_1\neq a\\ \cdots\\ x_n\neq a}} \pi(x_n)p(x_n,x_{n-1})p(x_{n-1},x_{n-2})\ldots p(x_1,b)$$

which rewrites as

$$\pi(b) = \pi(a)P_a(T_a \geq 1,\, X_1 = b) + \pi(a)P_a(T_a \geq 2,\, X_2 = b) + \ldots$$

$$\ldots+\pi(a)P_a(T_a \geq n,\, X_n = b)+\sum_{y\neq a}\pi(y)P_y(T_a \geq n,\, X_n = b)$$

Proof (continued). After n iterations we have

$$\pi(b) = \pi(a)\sum_{k=1}^{n} P_a(T_a \geq k,\, X_k = b) + \sum_{y\neq a}\pi(y)P_y(T_a \geq n,\, X_n = b)$$

Summing over all $b \in S$, we obtain

$$1 = \sum_{b\in S}\pi(b) = \pi(a)\sum_{k=1}^{n} P_a(T_a \geq k) + \sum_{y\neq a}\pi(y)P_y(T_a \geq n)$$

Since for all $y \in S$ and $E_y[T_a] < \infty$ (see next slide),

$$P_y(T_a \geq n) \;\leq \frac{E_y[T_a]}{n} \;\to 0$$

as $n \to \infty$, we have

$$1 = \pi(a)\sum_{k=1}^{\infty} P_a(T_a \geq k) = \pi(a)E_a[T_a]. \quad \text{Hence,} \quad \pi(a) = \frac{1}{E_a[T_a]}. \quad \square$$

Stationary distribution.

We used the following result.

Proposition. Consider a homogeneous Markov chain over a discrete state space S. Then

$$E_x[T_x] \geq P_x(T_y < T_x) E_y[T_x] \qquad \forall x, y \in S.$$

Proof. For $x \neq y$, we have

$$E_x[T_x] \geq E_x[T_x \mathbf{1}_{\{T_y<T_x\}}] \geq E_x[(T_x - T_y)\mathbf{1}_{\{T_y<T_x\}}]$$

$$\sum_{t=1}^{\infty} E_x[(T_x - T_y)\mathbf{1}_{\{T_y<T_x\}}\mathbf{1}_{\{T_y=t\}}] = \sum_{t=1}^{\infty} E_x[(T_x - t)\mathbf{1}_{\{t<T_x\}}\mathbf{1}_{\{T_y=t\}}]$$

$$= \sum_{t=1}^{\infty} E_x\big[T_x - t \,|\, X_t = y,\ X_{t-1} \in S\backslash\{x,y\}, \ldots, X_1 \in S\backslash\{x,y\}\big] P_x\big(T_y = t < T_x\big)$$

$$= \sum_{t=1}^{\infty} E_x\big[T_x - t \,|\, X_t = y\big] P_x\big(T_y = t < T_x\big) = E_y[T_x] \sum_{t=1}^{\infty} P_x\big(T_y = t < T_x\big) \quad \square$$

Stationary distribution.

Next, we will need the following result.

Proposition. Consider an irreducible aperiodic homogeneous Markov chain over a discrete state space S with transition probabilities $P = \Big(p(i,j)\Big)_{i,j\in S}$. Let (X_n, Y_n) be a homogeneous Markov chain on $S \times S$ with transition probabilities

$$\mathbf{p}\big((i_1,i_2),(j_1,j_2)\big) = p(i_1,j_1)p(i_2,j_2),$$

i.e., X_n and Y_n are independent Markov chains over S with transition probabilities $P = \Big(p(i,j)\Big)_{i,j\in S}$. Then, (X_n, Y_n) is an irreducible Markov chain over $S \times S$.

Corollary. If all states in S are positive recurrent, then $\pi(x,y) = \pi(x)\pi(y)$ is the unique stationary distribution of (X_n, Y_n). Consequently, $E_{(x,y)}[T_{(x,y)}] = E_x[T_x]E_y[T_y]$.

Stationary distribution.

Corollary. If all states in S are positive recurrent, then $\pi(x,y) = \pi(x)\pi(y)$ is the unique stationary distribution of (X_n, Y_n). Consequently, $E_{(x,y)}[T_{(x,y)}] = E_x[T_x]E_y[T_y]$.

Proposition. Consider an irreducible aperiodic homogeneous Markov chain over a discrete state space S whose states are positive recurrent. Let (X_n, Y_n) be a pair of independently evolving Markov chains, then

$$E_{(x_0,y_0)}\Big[\min_{x \in S} T_{(x,x)}\Big] < \infty \qquad \forall x_0, y_0 \in S.$$

Proof. By the preceding results, $\forall x_0, y_0, x \in S$,

$$P_{(x,x)}\big(T_{(x_0,y_0)} < T_{(x,x)}\big)\, E_{(x_0,y_0)}[T_{(x,x)}] \le E_{(x,x)}[T_{(x,x)}] = \big(E_x[T_x]\big)^2 < \infty,$$

where $P_{(x,x)}\big(T_{(x_0,y_0)} < T_{(x,x)}\big) > 0$. Hence,

$$E_{(x_0,y_0)}[T_{(x,x)}] < \infty.$$

□

Stationary distribution.

Lemma (Convergence). Consider an irreducible aperiodic homogeneous Markov chain over a discrete state space S. If all of its states are positive recurrent, then

$$\lim_{t\to\infty} p_t(x,y) = \pi(y) \qquad \forall x,y \in S,$$

where π is the unique stationary distribution.

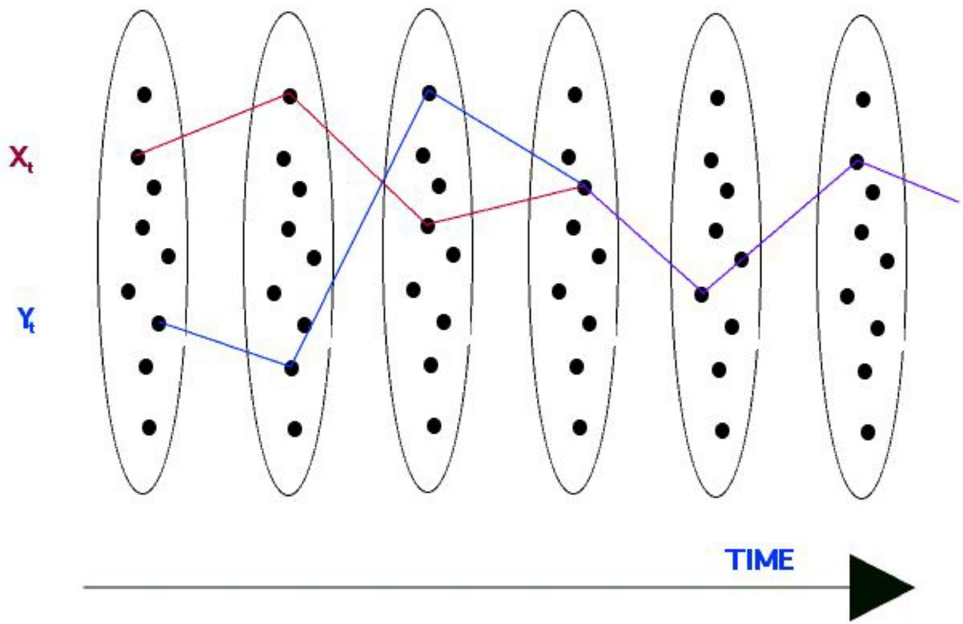
Proof. We use the coupling method. We let (X_n, Y_n) be a homogeneous Markov chain on $S \times S$ with transition probabilities

$$\mathbf{p}\big((i_1,i_2),(j_1,j_2)\big) = \begin{cases} p(i_1,j_1)p(i_2,j_2) & \text{if } i_1 \neq i_2, \\ p(i_1,j_1) & \text{if } i_1 = i_2 \text{ and } j_1 = j_2, \\ 0 & \text{if } i_1 = i_2 \text{ but } j_1 \neq j_2. \end{cases}$$

The process evolves according to

$$P\Big((X_{n+1},Y_{n+1}) = (j_1,j_2) \,\Big|\, (X_n,Y_n) = (i_1,i_2)\Big) = \mathbf{p}\big((i_1,i_2),(j_1,j_2)\big).$$

Proof (continued). We use the coupling method.



Proof (continued). We let (X_n, Y_n) be a homogeneous Markov chain on $S \times S$ with transition probabilities

$$\mathbf{p}\big((i_1, i_2), (j_1, j_2)\big) = \begin{cases} p(i_1, j_1)p(i_2, j_2) & \text{if } i_1 \neq i_2, \\ p(i_1, j_1) & \text{if } i_1 = i_2 \text{ and } j_1 = j_2, \\ 0 & \text{if } i_1 = i_2 \text{ but } j_1 \neq j_2. \end{cases}$$

Notice that each margin, X_n and Y_n, is a Markov chain with transition probabilities $P = \Big(p(i,j)\Big)_{i,j \in S}$.

The (stopping) time

$$\tau = \min\{n \geq 0 \,|\, X_n = Y_n\}$$

is called the coupling time.

For $n \geq \tau$, they X_n and Y_n evolve as a single Markov chain with transition probabilities $p(i, j)$:

$$X_\tau = Y_\tau, \quad X_{\tau+1} = Y_{\tau+1}, \quad X_{\tau+2} = Y_{\tau+2}, \quad X_{\tau+3} = Y_{\tau+3}, \ \ldots$$

Proof (continued). The preceding Proposition implies the coupling time is **finite**, i.e.,

$$P(\tau < \infty) = 1.$$

For any given $x \in S$, let $X_0 = x$, and Y_0 be distributed with probabilities

$$P(Y_0 = y) = \pi(y) \quad \forall y \in S.$$

Then, for all $n \in \mathbb{N}$, $P(Y_n = y) = \pi(y)$ $\forall y \in S$. Hence,

$$\sum_{y \in S} |p_t(x,y) - \pi(y)| = \sum_{y \in S} \Big| P(X_t = y) - P(Y_t = y) \Big|$$

$$= \sum_{y \in S} \Big| P(X_t = y,\, X_t \neq Y_t) - P(Y_t = y\, X_t \neq Y_t) \Big|$$

$$\leq 2P(X_t \neq Y_t) = 2P(\tau > t) \longrightarrow 0 \quad \text{as } t \to \infty. \qquad \square$$

Stationary distribution.

For a homogeneous Markov chain with the transition probability matrix $P = \Big(p(i,j)\Big)_{i,j\in S}$, the stationary distribution (aka 'equilibrium distribution') π is defined as follows:

$$\pi P = \pi \qquad \Leftrightarrow \qquad \sum_{i\in S} \pi(i)p(i,j) = \pi(j) \quad \forall j \in S.$$

Thus $\sum_i \pi(i)p(i,j) = \pi(j)\sum_i p(j,i)$, and for any state $j \in S$,

$$\sum_{i:\ i\neq j} \pi(i)p(i,j) = \sum_{i:\ i\neq j} \pi(j)p(j,i).$$

Thus when restated in terms of traffic flow, the influx to the state j is equal to outflow from j, for each j. Thus the distribution stays unchanged.

Stationary distribution and reversibility.

The following are the detailed balance conditions (d.b.c.) also called time reversibility:

$$\pi(i)p(i,j) = \pi(j)p(j,i) \qquad \forall i,j \in S.$$

Restated in terms of traffic flow: for every pair of states i and j the traffic in between them is balanced (equalized), i.e. the traffic flow from i to j equals to the traffic flow from j to i.

Observe that if **d.b.c.** are satisfied, the distribution will not change with time, i.e. π is stationary;

$$\sum_{i:\ i \neq j} \pi(i)p(i,j) = \sum_{i:\ i \neq j} \pi(j)p(j,i) \qquad \forall j \in S.$$

Birth-and-death chain. Consider state space

$$S = \{0, 1, 2, \ldots\}$$

and a Markov chain $\{X_t\}_{t=0,1,\ldots}$ on S with transition probabilities

$$p(i, i+1) = p_i, \quad p(i, i-1) = q_i, \quad \text{and} \quad p(i,i) = r_i$$

satisfying $q_0 = 0$ and $q_i + r_i + p_i = 1 \quad \forall i$

$$P = \begin{pmatrix} r_0 & p_0 & 0 & 0 & \cdots \\ q_1 & r_1 & p_1 & 0 & \ddots \\ 0 & q_2 & r_2 & p_2 & \ddots \\ 0 & 0 & q_3 & r_3 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

This is a Markov chain with only nearest neighbor transitions.

Stationary distribution and reversibility.

Observe that in the case of a birth-and-death chain, the definition of a stationary distribution

$$\pi P = \pi \qquad \Leftrightarrow \qquad \sum_{i \in S} \pi(i) p(i,j) = \pi(j) \quad \forall j \in S.$$

can be rewritten as $\pi_0 = r_0 \pi_0 + q_1 \pi_1$ and

$$\pi_j = p_{j-1}\pi_{j-1} + r_j \pi_j + q_{j+1}\pi_{j+1} \qquad \text{for } j = 1, 2, \ldots$$

The above equations can be shown to be equivalent to the detailed balance conditions (d.b.c.)

$$p_{k-1}\pi_{k-1} = q_k \pi_k.$$

Stationary distribution and reversibility.

Indeed, since $r_0 = 1 - p_0$ and $r_j = 1 - p_j - q_j$, equations $\pi_0 = r_0\pi_0 + q_1\pi_1$ and

$$\pi_j = p_{j-1}\pi_{j-1} + r_j\pi_j + q_{j+1}\pi_{j+1} \qquad \text{for } j = 1, 2, \ldots$$

are equivalent to $q_1\pi_1 - p_0\pi_0 = 0$ and

$$q_j\pi_j - p_{j-1}\pi_{j-1} = q_{j+1}\pi_{j+1} - p_j\pi_j \qquad \text{for } j = 1, 2, \ldots.$$

Hence,

$$0 = q_1\pi_1 - p_0\pi_0 = q_2\pi_2 - p_1\pi_1 = \ldots = q_j\pi_j - p_{j-1}\pi_{j-1} = \ldots$$

Thus, the detailed balance conditions (d.b.c.) are satisfied

$$p_{k-1}\pi_{k-1} = q_k\pi_k.$$

Hence, $\pi_k = \frac{p_{k-1}}{q_k}\pi_{k-1}$ for $k = 1, 2, \ldots$.

Stationary distribution and reversibility.

In the case of a birth-and-death chain, $\pi_k = \frac{p_{k-1}}{q_k}\pi_{k-1}$ and

$$\pi_k = \frac{p_0 \dots p_{k-1}}{q_1 \dots q_k}\pi_0 \qquad \text{for } k = 1, 2, \dots.$$

Next, $\sum_{k=0}^{\infty} \pi_k = 1$ implies

$$1 = \sum_{j=0}^{\infty} \pi_j = \pi_0 + \sum_{j=1}^{\infty} \frac{p_0 \dots p_{j-1}}{q_1 \dots q_j}\pi_0 = \pi_0 \left(1 + \sum_{j=1}^{\infty} \frac{p_0 \dots p_{j-1}}{q_1 \dots q_j} \right).$$

Hence,

$$\pi_0 = \left(1 + \sum_{j=1}^{\infty} \frac{p_0 \dots p_{j-1}}{q_1 \dots q_j} \right)^{-1} \text{ and } \pi_k = \frac{\frac{p_0 \dots p_{k-1}}{q_1 \dots q_k}}{1 + \sum_{j=1}^{\infty} \frac{p_0 \dots p_{j-1}}{q_1 \dots q_j}} \text{ for } k = 1, 2, \dots.$$

Stationary distribution.

Example (Random walk on weighted graph). Consider a finite simply connected graph $G = (V, E)$ with the weights assigned to all of its edges:

$$W_{x,y} = W_{y,x} > 0$$

for all $x, y \in V$ connected by an edge in E.

Denote by $W_x = \sum_{y \in V} W_{x,y}$ the total weight of the edges adjacent to $x \in V$.

Next, consider a random walk X_n on state space $S = V$ evolving according to the following transition probabilities

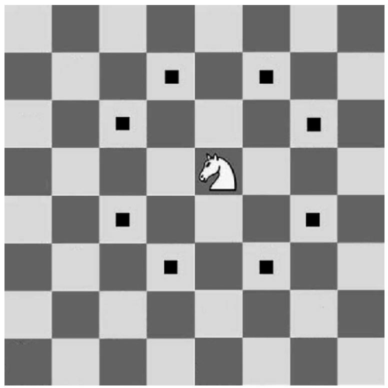
$$p(x, y) = \frac{W_{x,y}}{W_x} \quad \forall x, y \in V.$$

Then, $\pi(x) = W_x / Z_G$ with $Z_G = \sum_{x \in V} W_x$ satisfies the d.b.c.

Stationary distribution.

Example (Knight walk). Here is an example from an unpublished book by Aldous and Fill.

Consider the following random walk: Start with a knight at one of the corner squares of otherwise-empty chessboard. Each step, we move the knight by choosing uniformly from all the possible knight moves. What is the mean number of moves until the knight returns to the starting square?



Martingales.

Definition. A time homogeneous Markov chain $\{X_t\}$ over a discrete state space $S \subset \mathbb{R}$ is a martingale if

- $E[|X_t|] < \infty$ for all $t \geq 0$,
- $E[X_{t+1} \,|\, X_t] = X_t \quad \Leftrightarrow \quad E[X_{t+1} \,|\, X_t = x] = x$.

Example. Random walk on $\mathbb{Z}$. Take $p \in (0,1)$, and let $\xi_1, \xi_2, \ldots$ be i.i.d. Bernoulli random variables such that

$$\xi_j = \begin{cases} +1 & \text{with probability } p \\ -1 & \text{with probability } q = 1 - p \end{cases}$$

If $p = \frac{1}{2}$, the random walk $X_n = X_0 + \xi_1 + \ldots + \xi_n$ is a martingale.

Martingales.

Recall that $E\big[E[X|Y]\big] = E[X]$. Therefore,

$$E\big[E[X|Y]\,\big|\,Z=z\big] = E\big[X\,\big|\,Z=z\big] \quad \Leftrightarrow \quad E\big[E[X|Y]\,\big|\,Z\big] = E[X|Z]$$

If Markov chain $\{X_t\}$ is a martingale, then $E[X_1|X_0] = X_0$, and

$$E[X_2|X_0] = E\big[E[X_2|X_1]\,\big|\,X_0\big] = E\big[X_1\,\big|\,X_0\big] = X_0$$

Then, recursively, we have

$$E[X_3|X_0] = E\big[E[X_3|X_2]\,\big|\,X_0\big] = E\big[X_2\,\big|\,X_0\big] = X_0$$

$$E[X_4|X_0] = E\big[E[X_4|X_3]\,\big|\,X_0\big] = E\big[X_3\,\big|\,X_0\big] = X_0$$

$$\vdots$$

$$E[X_t|X_0] = X_0 \quad \Leftrightarrow \quad E[X_t\,|\,X_0 = x] = x \quad \text{for all } t \geq 0.$$

Stopping time.

Definition. For a homogeneous Markov chain $\{X_t\}$, a random variable T is a stopping time if for any $t \geq 0$, knowing $X_0, X_1, \ldots, X_t$ is sufficient for determining whether the event $\{T \leq t\}$ occurred or not. In other words, $\mathbf{1}_{T \leq t}$ is a function of $X_0, X_1, \ldots, X_t$.

Important example. For $A \subset S$, the first hitting time

$$T_A = \min\{t \geq 0 : X_t \in A\}$$

is a stopping time.

Optional Stopping Theorem. Suppose a homogeneous Markov chain $\{X_t\}$ is a martingale, and T is a stopping time with respect to X_t. If $P(T < \infty) = 1$ and there is $K > 0$ such that $|X_t| \leq K$ when $t \leq T$, then

$$E[X_T \,|\, X_0] = X_0.$$

First passage probability.

Example. Random walk on $\mathbb{Z}$. Take $p \in (0,1)$, and let $\xi_1, \xi_2, \ldots$ be i.i.d. Bernoulli random variables such that

$$\xi_j = \begin{cases} +1 & \text{with probability } p \\ -1 & \text{with probability } q = 1 - p \end{cases}$$

Consider integers $0 \leq x_0 \leq M$. Let $X_0 = x_0$ and $X_t = X_0 + \xi_1 + \ldots + \xi_t$. Then, the first first hitting time

$$T = \min\{t \geq 0 \,:\, X_t = 0 \text{ or } X_t = M\}$$

is a stopping time.

We want to find $P_{x_0}(X_T = M) = P(X_T = M \,|X_0 = x_0)$.

If $p = \frac{1}{2}$, the random walk $X_t = X_0 + \xi_1 + \ldots + \xi_t$ is a martingale, and by the Optional Stopping Theorem,

$$P_{x_0}(X_T = M) = \frac{x_0}{M}$$

Martingales.

Definition. A sequence of random variables $\{M_t\}$ is a martingale with respect to a homogeneous Markov chain $\{X_t\}$ if

- M_t is a function of $X_t, X_{t-1}, \ldots, X_0$,
- $E[|M_t|] < \infty$ for all $t \geq 0$, and

$$E[M_{t+1} \,|\, X_t, X_{t-1}, \ldots, X_0] = M_t \,.$$

Property: $E[M_t|X_0] = M_0$ for all $t \geq 0$.

Optional Stopping Theorem. Suppose $\{M_t\}$ is a martingale with respect to $\{X_t\}$, and T is a stopping time with respect to X_t. If $P(T < \infty) = 1$ and there is $K > 0$ such that $|M_t| \leq K$ when $t \leq T$, then

$$E[M_T \,|\, X_0] = M_0 \,.$$

Martingales.

Optional Stopping Theorem. Suppose $\{M_t\}$ is a martingale with respect to $\{X_t\}$, and T is a stopping time with respect to X_t. If $P(T < \infty) = 1$ and there is $K > 0$ such that $|M_t| \leq K$ when $t \leq T$, then

$$E[M_T \,|\, X_0] = M_0 \,.$$

Proof. Consider the stopped process $Y_t = M_{t \wedge T}$. Then,

$$Y_t = M_t \mathbf{1}_{T \geq t} + M_T \mathbf{1}_{T < t} = M_t \mathbf{1}_{T > t} + M_T \mathbf{1}_{T \leq t}$$

and

$$E[Y_{t+1} \,|X_t, \ldots, X_0] = \mathbf{1}_{T>t} E[M_{t+1} \,|X_t, \ldots, X_0] + E[M_T \mathbf{1}_{T \leq t} \,|X_t, \ldots, X_0]$$

$$= \mathbf{1}_{T>t} M_t + M_T \mathbf{1}_{T \leq t} = Y_t$$

as $\mathbf{1}_{T>t}$ and $M_T \mathbf{1}_{T \leq t}$ are functions of $X_0, \ldots, X_t$.

Hence, Y_t is a martingale with respect to $\{X_t\}$.

Martingales.

Proof (continued).

$$Y_t = M_t \mathbf{1}_{T \geq t} + M_T \mathbf{1}_{T < t} = M_t \mathbf{1}_{T > t} + M_T \mathbf{1}_{T \leq t}$$

and

$$E[Y_{t+1} \,|\, X_t, \ldots, X_0] = \mathbf{1}_{T>t} E[M_{t+1} \,|\, X_t, \ldots, X_0] + E[M_T \mathbf{1}_{T \leq t} \,|\, X_t, \ldots, X_0]$$

$$= \mathbf{1}_{T>t} M_t + M_T \mathbf{1}_{T \leq t} = Y_t$$

Hence, Y_t is a martingale with respect to $\{X_t\}$.

Thus, $E[Y_t \,|\, X_0] = Y_0 = M_0$ for all $t = 0, 1, 2, \ldots$, and

$$E[M_T \,|\, X_0] + E[Y_t - M_T \,|\, X_0] = M_0,$$

where, by Jensen's inequality,

$$\big| E[Y_t - M_T \,|\, X_0] \big| = \big| E[(M_t - M_T) \mathbf{1}_{T>t} \,|\, X_0] \big| \leq E\Big[|M_t - M_T| \mathbf{1}_{T>t} \,|\, X_0 \Big]$$

$$\leq 2K \, P(T > t \,|\, X_0) \longrightarrow 0 \quad \text{as} \quad t \to \infty.$$

Therefore, $E[M_T \,|\, X_0] = M_0$.

First passage probability.

Example. Random walk on $\mathbb{Z}$. Take $p \in (0,1)$, and let $\xi_1, \xi_2, \ldots$ be i.i.d. Bernoulli random variables such that

$$\xi_j = \begin{cases} 1 & \text{with probability } p \\ -1 & \text{with probability } q = 1 - p \end{cases}$$

Consider $0 \le x_0 \le M$. Let $X_0 = x_0$ and $X_n = X_0 + \xi_1 + \ldots + \xi_n$. Then, the first first hitting time

$$T = \min\{t \ge 0 \, : \, X_t = 0 \ \text{ or } \ X_t = M\}$$

is a stopping time.

We want to find $P_{x_0}(X_T = M) = P(X_T = M \,|\, X_0 = x_0)$.

If $p = \frac{1}{2}$, the random walk X_t is a martingale, and by the Optional Stopping Theorem, $P_{x_0}(X_T = M) = \frac{x_0}{M}$.

If $p \neq \frac{1}{2}$, then X_t is not a martingale.

First passage probability.

Example (continued).

We want to find $P_{x_0}(X_T = M) = P(X_T = M \mid X_0 = x_0)$.

If $p \neq \frac{1}{2}$, then X_t is not a martingale, but $M_t = h(X_t)$ is a martingale when $h(x) = A\left(\frac{q}{p}\right)^x + B$ for any choice of constants A and B.

Taking $A \neq 0$, by the Optional Stopping Theorem, we have

$$h(M)P_{x_0}(X_T = M) + h(0)\big(1 - P_{x_0}(X_T = M)\big) = E_{x_0}[h(X_T)]$$
$$= E_{x_0}[M_T] = h(x_0).$$

Therefore,

$$P_{x_0}(X_T = M) = \frac{h(x_0) - h(0)}{h(M) - h(0)} = \frac{1 - \left(\frac{q}{p}\right)^{x_0}}{1 - \left(\frac{q}{p}\right)^{M}}$$

Martingales and harmonic functions.

Suppose $\{X_t\}$ is a time homogeneous Markov chain (HMC).

We say that $h(\cdot)$ is a harmonic function with respect to the transition probabilities $\{p(x,y)\}$ if h satisfies the averaging property

$$\sum_y p(x,y)h(y) = h(x).$$

Here, $h(X_t)$ is a martingale with respect to $\{X_t\}$:

$$E[h(X_{t+1}) \,|\, X_t = x] = \sum_y p(x,y)h(y) = h(x)$$

and

$$E[h(X_{t+1}) \,|\, X_t] = h(X_t).$$

Martingales and harmonic functions.

For a birth-and-death chain X_t, the probability harmonic function h is the one satisfying the averaging property

$$h(k) = q_k h(k-1) + (1 - q_k - p_k) h(k) + p_k h(k+1)$$

The above recurrence relation, after being simplified as

$$q_k \Big(h(k) - h(k-1) \Big) = p_k \Big(h(k+1) - h(k) \Big)$$

yields $h(0) = A,\ h(1) = A + B$, and

$$h(k) = A + B \left(1 + \sum_{j=2}^{k} \frac{q_1 \ldots q_{j-1}}{p_1 \ldots p_{j-1}} \right) \quad \text{for } k = 2, 3, \ldots$$

Thus $M_t = h(X_t)$ is a martingale with respect to $\{X_t\}$.

Martingales and harmonic functions.

For a birth-and-death chain X_t, the probability harmonic function h is the one satisfying the averaging property

$$h(k) = q_k h(k-1) + (1 - q_k - p_k) h(k) + p_k h(k+1)$$

The above recurrence relation yields $h(0) = A$,

$$h(1) = A + B, \text{ and } \; h(k) = A + B\left(1 + \sum_{j=2}^{k} \frac{q_1 \ldots q_{j-1}}{p_1 \ldots p_{j-1}}\right) \quad \text{for } k = 2, 3, \ldots$$

Thus $M_t = h(X_t)$ is a martingale with respect to $\{X_t\}$. Define the following stopping time with respect to X_t,

$$T = \min\{t \geq 0 \,:\, X_t = 0 \text{ or } m\}.$$

Then, given that $X_0 = x_0$ for $0 \leq x_0 \leq m$,

$$P(X_T = m \,|\, X_0 = x_0) = \frac{h(x_0) - h(0)}{h(m) - h(0)} \qquad (B \neq 0).$$

Martingales and harmonic functions.

Example. For a birth-and-death chain X_t with $p_k = p$ and $q_k = q$ for all k, and $p \neq q$,

$$h(k) = qh(k-1) + (1-q-p)h(k) + ph(k+1)$$

yielding

$$h(k) = A + B\left(\frac{q}{p}\right)^k \quad \text{for } k = 0, 1, 2, 3, \ldots$$

Define the following stopping time with respect to X_t,

$$T = \min\{t \geq 0 \,:\, X_t = 0 \text{ or } m\}.$$

Then, given that $X_0 = x_0$ for $0 \leq x_0 \leq m$,

$$P(X_T = m \,|\, X_0 = x_0) = \frac{h(x_0) - h(0)}{h(m) - h(0)} = \frac{\left(\frac{q}{p}\right)^{x_0} - 1}{\left(\frac{q}{p}\right)^m - 1}.$$

Recurrence: random walk on $\mathbb{Z}$.
For $x \in S$, denote the first hitting time by

$$T_x = \min\{t \geq 1 \,:\, X_t = x\}.$$

Let X_t be a simple random walk on $S = \mathbb{Z}$ $(p_j = q_j = \frac{1}{2})$.
For all $M \in \mathbb{N}$, we have

$$P(T_0 > T_M \,|\, X_0 = x_0) = \frac{x_0}{M} \quad \text{for all} \quad 0 < x_0 < M,$$

and

$$P_0(T_0 < \infty) = P(T_0 < \infty \,|\, X_0 = 1) \geq P(T_0 < T_M \,|\, X_0 = 1) = 1 - \frac{1}{M}$$

Hence, as M is arbitrary, $P_0(T_0 < \infty) = 1$ and, by space homogeneity,

$$P_x(T_x < \infty) = P_0(T_0 < \infty) = 1 \quad \text{for all} \quad x \in \mathbb{Z}.$$

Thus, simple random walk on $\mathbb{Z}$ is recurrent.

Moreover, each state is null recurrent as $E_x[T_x] = E_0[T_0]$ for all $x \in \mathbb{Z}$.

Expected first hitting time.

Consider a birth-and-death chain X_t on $S = \{0, 1, \ldots, M\}$ with forward probabilities and backward probabilities denoted respectively p_j and q_j. Our goal is to find the first hitting time:

$$T = \text{min}\{t \geq 0 \,:\, X_t = 0 \text{ or } M\}.$$

We let

$$\varphi(j) = E[T \,|\, X_0 = j],$$

and write the following recurrence equation:

$$\begin{cases} \varphi(j) = 1 + q_j\varphi(j-1) + r_j\varphi(j) + p_j\varphi(j+1) & \text{for } j = 1, \ldots, M-1 \\ \varphi(0) = \varphi(M) = 0 \end{cases}$$

Let $\Delta\varphi(j) = \varphi(j+1) - \varphi(j)$ denote the forward difference. Then for $0 < j < n$,

$$\Delta\varphi(j) = \frac{q_j}{p_j}\Delta\varphi(j-1) - \frac{1}{p_j}.$$

Expected first hitting time.
Let $\Delta\varphi(j) = \varphi(j+1) - \varphi(j)$ denote the forward difference.
Then for $0 < j < n$,

$$\Delta\varphi(j) = \frac{q_j}{p_j}\Delta\varphi(j-1) - \frac{1}{p_j}.$$

Example (Random Walk on $\mathbb{Z}$). Let $p_j = q_j = \frac{1}{2}$.
Then,

$$\Delta\varphi(j) = \Delta\varphi(j-1) - 2 \quad \text{with} \quad \varphi(0) = \varphi(M) = 0.$$

Hence, $\Delta\varphi(j) = \Delta\varphi(0) - 2j = \varphi(1) - 2j$ and

$$\varphi(j) = \varphi(0) + \sum_{i=0}^{j-1} \Delta\varphi(i) = j\varphi(1) - j(j-1),$$

where for $j = M$, we have $0 = \varphi(M) = M\varphi(1) - M(M-1)$.
Hence, $\varphi(1) = M - 1$ and

$$E[T \,|\, X_0 = j] = \varphi(j) = j\varphi(1) - j(j-1) = j(M-j).$$

Diffusivity of simple random walk.
Consider a simple random walk on $\mathbb{Z}$: let $\xi_1, \xi_2, \ldots$ be i.i.d. Bernoulli random variables such that

$$\xi_j = \begin{cases} 1 & \text{with probability } p = \frac{1}{2} \\ -1 & \text{with probability } q = \frac{1}{2} \end{cases}$$

and let $X_0 = x_0$ and $X_t = X_0 + \xi_1 + \ldots + \xi_t$.

Here, the transition probabilities are

$$p(j, j+1) = p(j, j-1) = \frac{1}{2} \quad \text{for all } j \in \mathbb{Z}.$$

For $x \in \mathbb{Z}$ and $n \in \mathbb{N}$, let

$$T = \min\{t \geq 0 \,:\, X_t = x - n \text{ or } x + n\}.$$

Then, $E[T \,|\, X_0 = x] = n^2$, which can be interpreted as follows:

$$|X_{t+\Delta t} - X_t| \sim \sqrt{\sigma^2 \Delta t}, \quad \text{where } \sigma^2 = Var(\xi_j) = 1 \qquad \text{(diffusivity).}$$

Metropolis-Hastings algorithm.

Goal: simulating an S-valued random variable X, distributed according to a given probability distribution $\pi(z)$, i.e., $P(X = z) \approx \pi(z)$ for all $z \in S$.

MCMC: generating a Markov chain $\{X_t\}$ over S, with distribution $\mu_t(z) = P(X_t = z)$ converging rapidly to its unique stationary distribution, i.e., $\mu_t(z) \to \pi(z)$.

Metropolis-Hastings algorithm: Consider a connected neighborhood network with points in S. Suppose we know the ratios of $\frac{\pi(z')}{\pi(z)}$ for any two neighbor points z and z' on the network.

Let for z and z' connected by an edge of the network, the transition probability be set to

$$p(z, z') = \frac{1}{M} \min\left\{1, \frac{\pi(z')}{\pi(z)}\right\} \qquad \text{for } M \text{ large enough.}$$

Metropolis-Hastings algorithm.

Consider a connected neighborhood network with points in S.

Suppose we know the ratios of $\frac{\pi(z')}{\pi(z)}$ for any two neighbor points z and z' on the network.

Let for z and z' connected by an edge of the network, the transition probability be set to

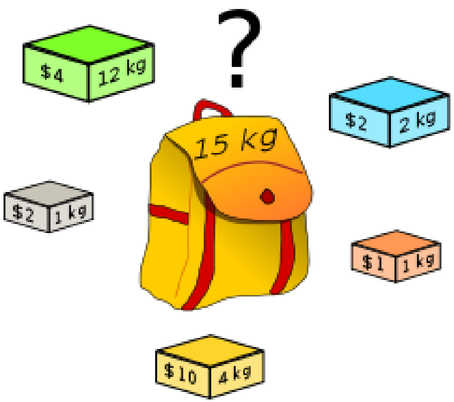
$$p(z,z') = \frac{1}{M}\ \min\left\{1,\ \frac{\pi(z')}{\pi(z)}\right\} \qquad \text{for } M \text{ large enough.}$$

Specifically, M can be any number greater than the maximal degree in the neighborhood network.

Let $p(z,z)$ absorb the rest of the probabilities, i.e.

$$p(z,z) = 1 - \sum_{z':\ z\sim z'} p(z,z')$$

Knapsack problem. The **knapsack problem** is a problem in combinatorial optimization: Given a set of items, each with a mass and a value, determine the number of each item to include in a collection so that the total weight is less than or equal to a given limit and the total value is as large as possible. Knapsack problem is NP complete.



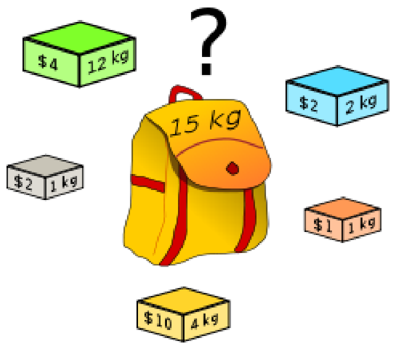
Source: Wikipedia.org

Knapsack problem. Given m items of various weights w_j and value v_j, and a knapsack with a weight limit R. Assuming the volume and shape do not matter, find the most valuable subset of items that can be carried in the knapsack.

Mathematically: we need $z = (z_1, \dots, z_m)$ in

$$S = \left\{ z \in \{0,1\}^m \ : \ \sum_{j=1}^{m} w_j z_j \le R \right\}$$

maximizing $U(z) = \sum_{j=1}^{m} v_j z_j$.



Source: Wikipedia.org

Knapsack problem. Find $z = (z_1, \ldots, z_m)$ in

$$S = \left\{ z \in \{0,1\}^m \ : \ \sum_{j=1}^m w_j z_j \le R \right\} \text{ maximizing } \ U(z) = \sum_{j=1}^m v_j z_j.$$

• **MCMC approach:** Assign weights $\pi(z) = \frac{1}{Z_\beta} \exp\left\{\beta\, U(z)\right\}$ to each $z \in S$ with $\beta = \frac{1}{T}$, where

$$Z_\beta = \sum_{z \in S} \exp\left\{\beta \ U(z)\right\}$$

is called partition function. Next, for each $z \in S$ consider a **clique** $\mathcal{C}_z$ of neighbor points in S. Consider a Markov chain over S that jumps from z to a neighbor $z' \in \mathcal{C}_z$ with probability

$$p(z, z') = \frac{1}{m} \ \min\left\{1, \ \frac{\pi(z')}{\pi(z)}\right\}.$$

Knapsack problem. Assign weights $\pi(z) = \frac{1}{Z_\beta} \exp\left\{\beta\, U(z)\right\}$ to each $z \in S$ with $\beta = \frac{1}{T}$, where

$$Z_\beta = \sum_{z \in S} \exp\left\{\beta\ U(z)\right\}$$

is called partition function. Next, for each $z \in S$ consider a **clique** $\mathcal{C}_z$ of neighbor points in S. Consider a Markov chain over S that jumps from z to a neighbor $z' \in \mathcal{C}_z$ with probability

$$p(z, z') = \frac{1}{m}\ \min\left\{1,\ \frac{\pi(z')}{\pi(z)}\right\}.$$

Observe that

$$\frac{\pi(z')}{\pi(z)} = \exp\left\{\beta\ \big(U(z') - U(z)\big) = \exp\left\{\beta\ \big(v \cdot (z' - z)\big)\right\},$$

where $v = (v_1, \ldots, v_m)$ is the values vector.

Mixing times.

Total variation distance:

$$\|\mu - \nu\|_{TV} := \frac{1}{2}\sum_{x\in S} |\mu(x) - \nu(x)| = \sup_{A\subset S} |\mu(A) - \nu(A)|$$

measure the distance between distributions μ and ν on the scale from 0 to 1.

Mixing time: for a given $\varepsilon \in (0,1)$ (e.g. $\varepsilon = 0.1$), let

$$t_{mix}(\varepsilon) := \inf\{t : \|\mu_t - \pi\|_{TV} \le \varepsilon, \quad \text{all } \mu_0\},$$

where $\mu_t = \mu_0 P^t$.

Mixing time is a running time of MCMC algorithm.

Card shuffling.

Problem: We would like to shuffle a deck of n cards so that each of the $n!$ possible configurations is equally likely, i.e., has probability $\frac{1}{n!}$. We use following algorithm: each step, we take the top card and insert it to any of the n slots in the deck (including the top) with equal probability. Each step is performed independently. How soon will the deck be well shuffled?

This card shuffling algorithm is a Markov chain on the space $S = S_n$ of all n-permutations with stationary distribution $\pi(\sigma) = \frac{1}{n!}$ $\forall \sigma \in S_n$. Let μ_t denote the distribution of possible configurations after t shuffles. We need to find the most optimal upper bound on the mixing time $t_{mix}(\varepsilon)$.

Let τ denote the first time all n cards were shuffled in, after ascending to the top of the deck.

Card shuffling. Let τ denote the first time all n cards were shuffled in, after ascending to the top of the deck.

Observe that $P(X_\tau = \sigma) = \pi(\sigma)$, and

$$P(X_t = \sigma \,|\, t \geq \tau) = \pi(\sigma) \quad \forall s \geq 0.$$

Next,

$$\|\mu_t - \pi\|_{TV} = \frac{1}{2}\sum_{\sigma \in S_n} \big|\mu_t(\sigma) - \pi(\sigma)\big| = \frac{1}{2}\sum_{\sigma \in S_n} \big|P(X_t = \sigma) - \pi(\sigma)\big|$$

$$\leq \frac{1}{2}\sum_{\sigma \in S_n} \big|P(X_t = \sigma \,|t < \tau) - \pi(\sigma)\big| P(t < \tau) + \frac{1}{2}\sum_{\sigma \in S_n} \big|P(X_t = \sigma \,|t \geq \tau) - \pi(\sigma)\big| P(t \geq \tau)$$

$$= \frac{1}{2}\sum_{\sigma \in S_n} \big|P(X_t = \sigma \,|\, t < \tau) - \pi(\sigma)\big|\, P(t < \tau) \leq P(t < \tau) \leq \frac{1}{t} E[\tau],$$

where $E[\tau] = n \log n + O(n)$ by the coupon collector problem.

Coupon collector problem.

n types of coupons: $\boxed{1}$, $\boxed{2}$, ..., $\boxed{n}$. Collecting coupons: coupon / unit of time, each coupon type is equally likely. Goal: To collect a coupon of each type.
Question: How much time will it take?

c1	c2	c1	c3	c3	c2	c4	c5	...
↑	↑		↑			↑	↑	
$\tau_1 = 1$	τ_2		τ_3			τ_4	τ_5	

Here, $\tau_1 = 1$, $E[\tau_2 - \tau_1] = \dfrac{n}{n-1}$, $E[\tau_3 - \tau_2] = \dfrac{n}{n-2}, \ldots, E[\tau_n - \tau_{n-1}] = n$.

Hence,

$$E[\tau_n] = n\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right) = n \log n + O(n)$$

Card shuffling.

This card shuffling algorithm is a Markov chain on the space $S = S_n$ of all n-permutations with stationary distribution $\pi(\sigma) = \frac{1}{n!}$ $\forall \sigma \in S_n$. Let μ_t denote the distribution of possible configurations after t shuffles. We need to find the most optimal upper bound on the mixing time $t_{mix}(\varepsilon)$.

Let τ denote the first time all n cards were shuffled in, after ascending to the top of the deck. Then,

$$\|\mu_t - \pi\|_{TV} \le P(t < \tau) \le \frac{1}{t} E[\tau],$$

where $E[\tau] = n \log n + O(n)$ by the coupon collector problem.

Hence, for $t = \frac{E[\tau]}{\varepsilon} = \frac{1}{\varepsilon} n \log n + O(n)$, $\|\mu_t - \pi\|_{TV} \le \varepsilon$ and

$$t_{mix}(\varepsilon) \le Cn \log n \qquad \text{for} \quad C > \frac{1}{\varepsilon}.$$