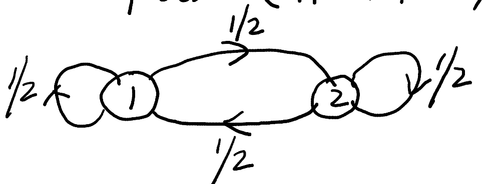
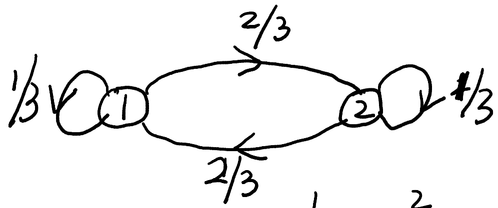
Introduction

Monday, January 10, 2011
10:29 PM

The following example illustrates the difference in rates at which two similar Markov chains converges to the same steady-state probability distribution $\vec{\pi}$.

$\vec{P}(t) = (P_1(t), P_2(t))$





$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$\vec{P}(0) = (1, 0)$$

$$\vec{P}(1) = \vec{P}(0)P = \left(\tfrac{1}{2}, \tfrac{1}{2}\right)$$

$$\vec{P}(2) = \vec{P}(0)P = \left(\tfrac{1}{2}, \tfrac{1}{2}\right)$$

......

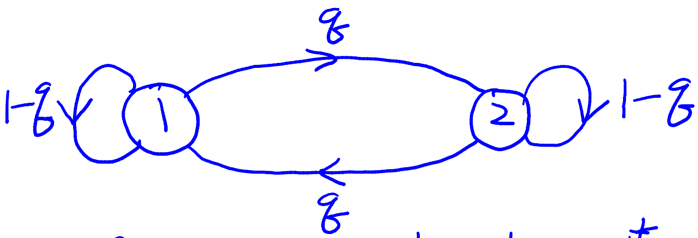
$$P = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

$$\vec{P}(0) = (1, 0)$$

$$\vec{P}(1) = \vec{P}(0)P = \left(\tfrac{1}{3}, \tfrac{2}{3}\right)$$

$$\vec{P}(2) = \vec{P}(0)P = \left(\tfrac{5}{9}, \tfrac{4}{9}\right)$$

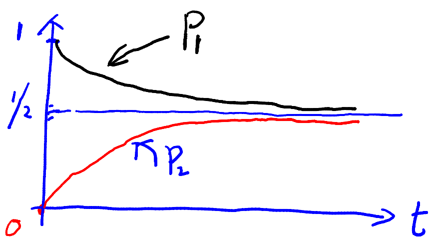
....



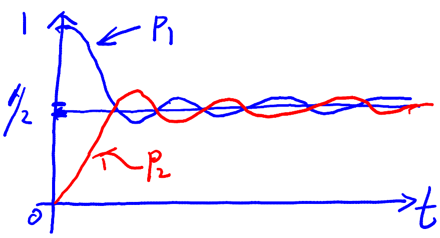
$$P = \begin{bmatrix} 1-q & q \\ q & 1-q \end{bmatrix} \qquad P^t = \begin{bmatrix} \frac{1}{2} + \frac{1}{2}(1-2q)^t & \frac{1}{2} - \frac{1}{2}(1-2q)^t \\ \frac{1}{2} - \frac{1}{2}(1-2q)^t & \frac{1}{2} + \frac{1}{2}(1-2q)^t \end{bmatrix}$$

$$\vec{P}(t) = \vec{P}(0)P^t = \left(\tfrac{1}{2} + \tfrac{1}{2}(1-2q)^t,\ \tfrac{1}{2} - \tfrac{1}{2}(1-2q)^t\right)$$

If $0 < q < \frac{1}{2}$



If $1 > q > \frac{1}{2}$



Concepts

Monday, January 10, 2011
10:52 PM

1. Total variation distance.

Given two probability distributions μ and η on Ω, the total variation distance is

$$\|\mu-\eta\| := \frac{1}{2}\sum_{x\in\Omega} |\mu(x)-\eta(x)|$$

$$\boxed{0 \leq \frac{1}{2}\sum_{x\in\Omega} |\mu(x)-\eta(x)| \leq \frac{1}{2}\left[\sum_{x\in\Omega}\mu(x) + \sum_{x\in\Omega}\eta(x)\right] = 1}$$

example:

Consider n cards.

Let μ be the uniform distribution over all permutations of the n cards. and η be the same distribution except that the bottom card is fixed.

$\|\mu-\eta\| = ?$ $\Omega = \{$ permutations of the n cards $\}$

$A = \{$ pemutations of the n cards except that the bottom card is fixed $\}$

For any $x\in A$, $\eta(x) = \frac{1}{(n-1)!}$

For any $x\in\Omega$, $\mu(x) = \frac{1}{n!}$

$$\begin{aligned}
\|\mu-\eta\| &= \frac{1}{2}\sum_{x\in\Omega} |\mu(x)-\eta(x)| \\
&= \frac{1}{2}\sum_{x\in A} |\mu(x)-\eta(x)| + \frac{1}{2}\sum_{x\in\Omega\setminus A} |\mu(x)-\eta(x)| \\
&= \frac{1}{2}\sum_{x\in A} \left|\frac{1}{(n-1)!}-\frac{1}{n!}\right| + \frac{1}{2}\sum_{x\in\Omega\setminus A}\frac{1}{n!} \\
&= \frac{1}{2}(n-1)!\left(\frac{1}{(n-1)!}-\frac{1}{n!}\right) + \frac{1}{2}\left(n!-(n-1)!\right)\frac{1}{n!} \\
&= \frac{1}{2}\left(1-\frac{1}{n}\right) + \frac{1}{2}\left(1-\frac{1}{n}\right) \\
&= 1-\frac{1}{n}.
\end{aligned}$$

Exercise: prove that $\frac{1}{2}\sum_{x\in\Omega} |\mu(x)-\eta(x)| = \max_{A\subseteq\Omega} |\mu(A)-\eta(A)|$.

$S^+ = \{x: \mu(x)\geq\eta(x)\}$, $S^- = \{x: \mu(x)<\eta(x)\}$

$\max_{A\subseteq\Omega} \mu(A)-\eta(A) = \mu(S^+)-\eta(S^+)$

$$\max_{A \subseteq \Omega} \eta(A) - \mu(A) = \eta(S^-) - \mu(S^-).$$

Moreover, $\mu(S^+) + \mu(S^-) = \eta(S^+) + \eta(S^-) = 1.$

$$\Rightarrow \quad \mu(S^+) - \eta(S^+) = \eta(S^-) - \mu(S^-)$$

hence $\max_{A \subseteq \Omega} |\mu(A) - \eta(A)| = |\mu(S^+) - \eta(S^+)| = |\eta(S^-) - \mu(S^-)|$

Since $|\mu(S^+) - \eta(S^+)| + |\eta(S^-) - \mu(S^-)|$

$$= \mu(S^+) - \eta(S^+) + \eta(S^-) - \mu(S^-)$$

$$= \sum_{x \in S} |\mu(x) - \eta(x)|$$

$$\Rightarrow \max_{A \subseteq \Omega} |\mu(A) - \eta(A)| = \frac{1}{2} \sum_{x \in S} |\mu(x) - \eta(x)|.$$

Go back to previous example.

The probability of the bottom card of the first deck being the same as the fixed card of the second deck is $1/n$.
but for the second deck is 1.

2. (Coupling) Let μ and η be any two probability distributions over Ω. A probability distribution ω over $\Omega \times \Omega$ is said to be a coupling of μ and η if its marginals are μ and η; that is,

$$\mu(x) = \sum_{y \in \Omega} \omega(x, y)$$

$$\eta(x) = \sum_{y \in \Omega} \omega(y, x)$$

3. (Coupling Lemma.) Let μ and η be probability distributions on Ω, and let X and Y be random variables with distributions μ and η, respectively. Then,

① $\|\mu - \eta\| \leq \Pr\{X \neq Y\}$

② There exists a coupling of (μ, η) such that

② There exists a coupling of (μ, η) such that
$pr\{X \neq Y\} = \|\mu - \eta\|$.

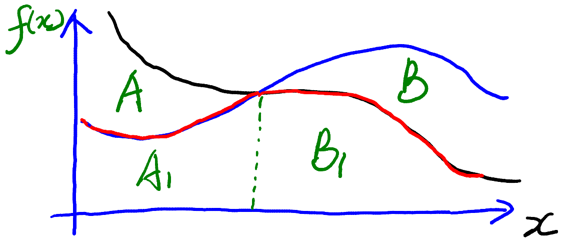
proof: ① $P\{X \in A\} = P\{X \in A, Y \notin A\} + P\{X \in A, Y \in A\}$
$\leq P\{X \in A, Y \notin A\} + P\{Y \in A\}$

$\Rightarrow \mu(A) - \eta(A) = P\{X \in A\} - P\{Y \in A\}$
$\leq P\{X \in A, Y \notin A\} \leq P\{X \neq Y\}$

$\|\mu - \eta\| = \max_{A \subseteq \Omega} |\mu(A) - \eta(A)| \leq P\{X \neq Y\}$.

② Want to mimize $P\{X \neq Y\}$
$\Leftrightarrow$ maximize $P\{X = Y\}$.

Clearly, the best we can do is to make $X = Y = z$, with probability $\min\{Pr\{X = z\}, Pr\{Y = z\}\}$ for $\forall z \in \Omega$.



$Pr\{X = Y\} = A_1 + B_1$

but $\begin{cases} A + A_1 + B_1 = 1 \\ A_1 + B_1 + B = 1 \end{cases} \Rightarrow A + B + 2(A_1 + B_1) = 2$
$\Rightarrow \frac{A+B}{2} + (A_1 + B_1) = 1$.

$P\{X \neq Y\} = 1 - P\{X = Y\} = \frac{A+B}{2}$

$A + B = \sum_{x \in \Omega} |\mu(x) - \eta(x)|$

$P\{X \neq Y\} = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \eta(x)| = \|\mu - \eta\|$.

Concepts_2

Tuesday, January 11, 2011
7:40 PM

Definition

① For any $x \in \Omega$, we define $\Delta_x(t) = \|P_x^{(t)} - \pi\|$. $P_x^{(t)} = P^t(x,\cdot)$

② $\Delta(t) = \max_{x\in\Omega} \Delta_x(t)$ is the maximum possible distance from π after t time steps.

③ $\tau_x(\varepsilon) = \min\{t : \Delta_x(t) \le \varepsilon\}$ is the first time step t at which the distance $\|P_x^{(t)} - \pi\|$ drops to ε.

④ $\tau(\varepsilon) = \max_{x\in\Omega} \tau_x(\varepsilon)$.

⑤ The mixing time τ_{mix} of a Markov chain is $\tau(\frac{1}{2e})$.

Claim: $\Delta_x(t)$ is non-increasing in t.

Proof: Let $X_0 = x$ and Y_0 have the stationary distribution π. We fix t and couple the distributions of the random variables X_t and Y_t such that

$$Pr\{X_t \neq Y_t\} = \|P_x^{(t)} - \pi\| = \Delta_x(t)$$

which is possible because of the coupling Lemma.

We now use this coupling of the distributions of X_{t+1} and Y_{t+1}, as follows:

- If $X_t = Y_t$, then set $X_{t+1} = Y_{t+1}$
- otherwise, Let $X_t \to X_{t+1}$, and $Y_t \to Y_{t+1}$ independently.

Then, we have

$$\Delta_x(t+1) = \|P_x^{(t+1)} - \pi\| \le \underbrace{Pr\{X_{t+1} \neq Y_{t+1}\} \le Pr\{X_t \neq Y_t\}} = \Delta_x(t)$$

Note $\{X_t = Y_t\} \subseteq \{X_{t+1} = Y_{t+1}\}$

$\Leftrightarrow \{X_{t+1} \neq Y_{t+1}\} \subseteq \{X_t \neq Y_t\}$ ∎

$$\Leftrightarrow \{X_{t+1} \neq Y_{t+1}\} \subseteq \{X_t \neq Y_t\}. \quad \blacksquare$$

We now define more general quantities which capture the evolution of distance between corresponding distance for arbitrary initial configurations.

Definition:

① $D_{xy}(t) = \| P_x^{(t)} - P_y^{(t)} \|$

② $D(t) = \max_{x,y\in\Omega} D_{xy}(t)$.

Claim: $\Delta(t) \leq D(t) \leq 2\Delta(t)$.

$$\Delta(t) = \max_{x\in\Omega} \Delta_x(t) = \max_{x\in\Omega} \| P_x^{(t)} - \pi \|$$

①
$$\| P_x^{(t)} - P_y^{(t)} \| = \| P_x^{(t)} - \pi + \pi - P_y^{(t)} \|$$
$$\leq \| P_x^{(t)} - \pi \| + \| P_y^{(t)} - \pi \|$$
$$\max_{x,y\in\Omega} \| P_x^{(t)} - P_y^{(t)} \| \leq \max_{x\in\Omega} \| P_x^{(t)} - \pi \| + \max_{y\in\Omega} \| P_y^{(t)} - \pi \|$$
$$\Leftrightarrow \quad D(t) \leq 2\Delta(t).$$

②
$$\| P_x^{(t)} - \pi \| = \max_{A\subset\Omega} | P^{(t)}(x,A) - \pi(A) |$$
$$= \max_{A\subset\Omega} \left| \sum_{y\in\Omega} \pi(y) [P^t(x,A) - P^t(y,A)] \right|$$
$$\leq \sum_{y\in\Omega} \pi(y) \max_{A\subset\Omega} | P^t(x,A) - P^t(y,A) |$$
$$= \sum_{y\in\Omega} \pi(y) \| P_x^{(t)} - P_y^{(t)} \|$$
$$\max_{x\in\Omega} \| P_x^{(t)} - \pi \| \leq \max_{x,y\in\Omega} \sum_{y\in\Omega} \pi(y) \| P_x^{(t)} - P_y^{(t)} \|$$
$$\leq \sum_{y\in\Omega} \pi(y) \max_{x,y\in\Omega} \| P_x^{(t)} - P_y^{(t)} \|$$
$$= \max_{x,y\in\Omega} \| P_x^{(t)} - P_y^{(t)} \|.$$

claim. $\Delta(t) \leq \exp(-\lfloor \frac{t}{\tau_{mix}} \rfloor)$

proof: <u>claim.</u> $D(t+s) \leq D(t) D(s)$ ($D(t)$ is submultiplicative)

It follows that $D(kt) \leq D(t)^k$ for all positive integers k.

Consequently,

$$\Delta(k\tau_{mix}) \leq D(k\tau_{mix}) \leq D(\tau_{mix})^k \leq (2\Delta(\tau_{mix}))^k = e^{-k}$$

$\Rightarrow$ Let $t = k\tau_{mix}$, we have

$$\Delta(t) \leq \exp(-\lfloor \frac{t}{\tau_{mix}} \rfloor)$$

<u>proof of claim.</u>

Let $X_0 = x$ and $Y_0 = y$. We use the coupling Lemma to couple the distributions of X_t and Y_t so that

$$D_{xy}(t) \equiv \| P_x^{(t)} - P_y^{(t)} \| = P\{X_t \neq Y_t\}.$$

We then construct a coupling of X_{t+s} and Y_{t+s} as follows:

- If $X_t = Y_t$, then set $X_{t+i} = Y_{t+i}$ for $i = 1, 2, \cdots, s$.
- otherwise, Let $X_t = x'$ and $Y_t = y' \neq x'$. Use the coupling Lemma to couple the distributions of X_{t+s} and Y_{t+s}, conditioned on $X_t = x'$ and $Y_t = y'$, such that

$$\Pr\{X_{t+s} \neq Y_{t+s} \mid X_t = x', Y_t = y'\} = \| P_{x'}^{(s)} - P_{y'}^{(s)} \| = D_{x'y'}(s) \leq D(s)$$

We now have

$$D_{xy}(t+s) = \| P_x^{(t+s)} - P_y^{(t+s)} \|$$
$$\leq \Pr\{X_{t+s} \neq Y_{t+s}\}, \text{ (by the coupling Lemma)}$$

$$\begin{aligned}
\Pr\{X_{t+s} \neq Y_{t+s}\} &= \Pr\{X_{t+s} \neq Y_{t+s} \mid X_t \neq Y_t\} \Pr\{X_t \neq Y_t\} \\
&\quad + \Pr\{X_{t+s} \neq Y_{t+s} \mid X_t = Y_t\} \Pr\{X_t = Y_t\} \\
&= \Pr\{X_{t+s} \neq Y_{t+s} \mid X_t \neq Y_t\} \Pr\{X_t \neq Y_t\} \\
&\leq D(s) D_{xy}(t) \leq D(s) D(t)
\end{aligned}$$

hence $D_{xy}(t+s) \leq D(s) D(t)$

$\Rightarrow$ $D(t+s) \leq D(s) D(t)$. ∎

Corollary: $\tau(\varepsilon) \le \tau_{mix} \lceil \log(\frac{1}{\varepsilon}) \rceil$.

$\Delta(t) \le \exp(-\frac{t}{\tau_{mix}})$

Let $\exp(-\frac{t}{\tau_{mix}}) = \varepsilon \Rightarrow \underline{t = \tau_{mix} \log(\frac{1}{\varepsilon})}$.

example

Tuesday, January 11, 2011
8:29 PM

Coupling Techniques:

Definition: A coupling of a Markov chain M_t with state space S is a Markov chain $Z_t=(X_t, Y_t)$ on the state space $S\times S$ such that:

$$Pr(X_{t+1}=x' \mid Z_t=(x,y)) = Pr(M_{t+1}=x' \mid M_t=x)$$
$$Pr(Y_{t+1}=y' \mid Z_t=(x,y)) = Pr(M_{t+1}=y' \mid M_t=y)$$

Definition: A coupling of a Markov chain P is a pair process (X_t, Y_t) such that:

- each of $(X_t, \cdot)$ and $(\cdot, Y_t)$, viewed in isolation, is a faithful copy of the Markov chain; that is,
  $Pr\{X_{t+1}=b \mid X_t=a\} = P(a,b) = Pr\{Y_{t+1}=b \mid Y_t=a\}$
- if $X_t=Y_t$, then $X_{t+1}=Y_{t+1}$.

Now, define the random variable

$$T_{xy}=\min\{t: X_t=Y_t \mid X_0=x, Y_0=y\}$$ to be the (stopping) time until the two processes meet. The following claim gives the desired upper bound on the mixing time:

claim: $\Delta(t) \le \max_{x,y} Pr\{T_{xy}>t\}$.

proof: recall: $\Delta(t) = \max_{x\in\Omega} \| P_x^{(t)} - \pi \|$

$$D(t) = \max_{x,y\in\Omega} \| P_x^{(t)} - P_y^{(t)} \|$$

Then, $\Delta(t) \le D(t)$

$$= \max_{x,y\in\Omega} \| P_x^{(t)} - P_y^{(t)} \|$$

$$< \max Pr\{ X_t \ne Y_t \mid X_0=x, Y_0=y \}$$

$$\leq \max_{x,y\in\Omega} \Pr\{X_t \neq Y_t \mid X_0 = x,\ Y_0 = y\}$$

$$= \max_{x,y} \Pr\{T_{xy} > t \mid X_0 = x,\ Y_0 = y\}$$

random walk

Tuesday, January 11, 2011
10:58 PM

Simple random walk on the hypercube $\{0,1\}^n$.

The n-dimensional cube is a graph with 2^n vertices, each of which can be encoded as an n-bit binary string, $b_1 b_2 \cdots b_n$, whose neighbours are the strings which differ from it by Hamming distance exactly 1. We define a random walk on the cube by the following,

① with prob. $1/2$, do nothing

② else, pick a coordinate $i \in \{1, \cdots, n\}$ uniformly at random and flip coordinate x_i (i.e. $x_i \to 1-x_i$)

This setup is clearly equivalent to the following:

① pick a coordinate $i \in \{1, \cdots, n\}$ uniformly at random and a bit $\{0,1\}$ uniformly at random.

② set $x_i = b$

This second description of the random walk dynamics suggests the following coupling: make X_t and Y_t choose the same i and b at every step. Clearly this is a valid coupling: obviously each of X_t and Y_t is performing exactly the above random walk.

To analyze the time T_{xy}, notice that once every $i \in \{1, \cdots, n\}$ has been choosen at least once, X_t must equal Y_t.

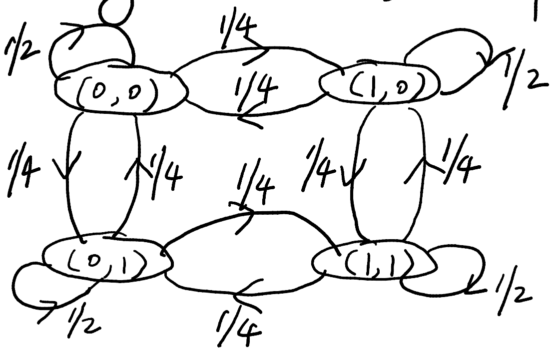
(This is because, once a coordinate i has been choosen, X_t and Y_t agree on that coordinate at all future times).

Thus, for any x and y, T_{xy} is stochastically dominated by the time for a coupon collector to collect all n coupons. Thus $\Pr\{T_{xy} > n\ln n + cn\} < e^{-c}$.

hence $\Delta(n\ln n + cn) \leq e^{-c}$

Let $e^{-c} = \varepsilon$. $\Rightarrow$ $c = \log\frac{1}{\varepsilon}$

More generally $\tau(\varepsilon) \leq n\ln n + \lceil n\ln\frac{1}{\varepsilon}\rceil$



Coupon Collector

there are n types of coupons and at each trial a coupon is chosen at random. Each random coupon is equally likely to be of any of the n types, and the random choice of the coupons are mutually independent.

Let X be the number of trials. $E(X)$?

If X_i is the number of trials needed to get a new coupon while you had exactly $i-1$ different coupons, clearly, $X = \sum_{i=1}^{n} X_i$

The advantage of breaking the random variable X into a sum of n random variables X_i, $i=1,2,\cdots,n$, is that each X_i is a geometric random variable.

when exactly $i-1$ coupons have been found, the probability of obtaining a new coupon is

$$P_i = 1 - \frac{i-1}{n} = \frac{n-i+1}{n}$$

Hence, X_i is a geometric random variable with parameter P_i, and

$$E[X_i] = \frac{1}{P_i} = \frac{n}{n-i+1}$$

hence $E[X] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$

$$= \sum_{i=1}^{n} \frac{n}{n-i+1} = n \sum_{i=1}^{n} \frac{1}{i}$$

$$\sum_{i=1}^{n} \frac{1}{i} \geq \int_1^n \frac{1}{x} dx \geq \sum_{i=2}^{n} \frac{1}{i}$$

$$\log n \leq \sum_{i=1}^{n} \frac{1}{i} \leq \log n + 1$$

$$\Rightarrow \quad \sum_{i=1}^{n} \frac{1}{i} = \log n + \Theta(1)$$

hence $E(X) = n \log n + \Theta(n)$

claim. $P\{X > \lceil n \log n + cn \rceil\} \leq e^{-c}$.

proof. Let A_i be the event that the i-th type does not appear among the first $\lceil n \log n + cn \rceil$ coupons drawn.

Observe first that

$$P\{X > \lceil n \log n + cn \rceil\} = P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$$

Since each trial has prob. $1 - \frac{1}{n}$ of not drawing coupon i and the trials are independent,

$$P(A_i) = \left(1-\frac{1}{n}\right)^{\lceil n\log n + cn\rceil}$$

hence $P\{X > \lceil n\log n + cn\rceil\} \leq \sum_{i=1}^{n} P(A_i)$

$$= \sum_{i=1}^{n} \left(1-\frac{1}{n}\right)^{\lceil n\log n + cn\rceil}$$

$$\leq n\exp\left(-\frac{n\log n + cn}{n}\right) = e^{-c}.$$

(use $1-x \leq e^{-x},\ x \geq 0$)