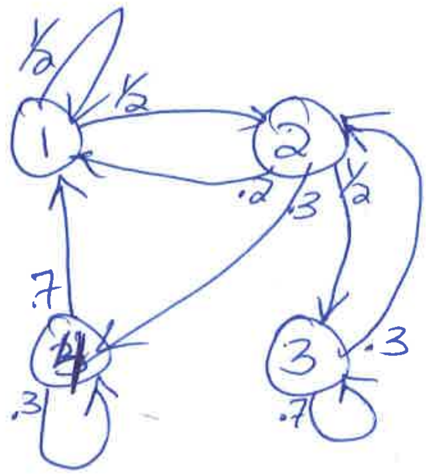
Markov chains:

Example:

States $\{1,2,3,4\}$



$$P=\begin{bmatrix} .5 & .5 & 0 & 0 \\ .2 & 0 & .5 & .3 \\ 0 & .3 & .7 & 0 \\ .7 & 0 & 0 & .3 \end{bmatrix}$$

In General: State space Ω

Random variable $X_t \in \Omega$ = state at time t

Discrete time $t \geq 0$.

Transition matrix P is a $N \times N$ matrix
with non-negative entries
& is stochastic (rows sum to 1)

For states $i,j \in \Omega$, $\Pr(X_{t+1}=j \mid X_t=i) = P(i,j)$

Markovian Property:

for states $k_0, k_1, \ldots, k_{t-1}, i, j \in \Omega$

$$\Pr(X_{t+1}=j \mid X_0=k_0, X_1=k_1, \ldots, X_{t-1}=k_{t-1}, X_t=i)$$
$$= \Pr(X_{t+1}=j \mid X_t=i) = P(i,j)$$

2-step transition probabilities:

$$\Pr(X_{t+2}=j \mid X_t=i) = \sum_{k\in\Omega} \Pr(X_{t+2}=j \mid X_{t+1}=k)\Pr(X_{t+1}=k \mid X_t=i)$$

$$= \sum_k P(i,k)P(k,j)$$

$$= P^2(i,j).$$

for $\ell > 1$, by induction,

$$\Pr(X_{t+\ell}=j \mid X_t=i)$$

$$= \sum_k \Pr(X_{t+1}=k \mid X_t=i)\Pr(X_{t+\ell}=j \mid X_{t+1}=k)$$

$$= \sum_k P(i,k)P^{t-1}(k,j)$$

$$= P^t(i,j)$$

If $X_0 \sim \mu$ (so X_0 is sampled from distribution μ_0)

then $X_1 \sim \mu_1$ where $\mu_1 = \mu_0 P$

& $X_t \sim \mu_t$ where $\mu_t = \mu_0 P^t$

Earlier example:

$$P^{20} = \begin{bmatrix} .244190 & .244187 & .406971 & .104652 \\ .244187 & .244186 & .406975 & .104651 \\ .244181 & .244185 & .406984 & .104650 \\ .244195 & .244188 & .406966 & .104652 \end{bmatrix}$$

Distribution $\pi \approx [.2442, .2442, .4070, .10465]$

A stationary distribution is a distribution invariant with respect to the transition matrix:

$$\pi = \pi P.$$

(in other words, for all $j \in \Omega$,

$$\pi(j) = \sum_{i \in \Omega} \pi(i) P(i,j))$$

This π is an eigenvector with eigenvalue 1.

Ergodic if $\exists t$ s.t. $\forall i,j \in \Omega$, $P^t(i,j) > 0$
(the graph defined by P^t is fully-connected)

Irreducible if $\forall i,j \in \Omega$, $\exists t$ s.t. $P^t(i,j) > 0$
(graph defined by P is 1 SCC)

Aperiodic: For $i \in \Omega$, period of i is $T_i = \{t : P^t(i,i) > 0\}$

Aperiodic if $\forall i \in \Omega$, $\gcd(T_i) = 1$.

Ergodic $\iff$ Irreducible & Aperiodic.

see [LWP].

Theorem: For a finite, ergodic MC, there is a unique stationary distribution π and for all $i,j \in \Omega$,

$$\lim_{t \to \infty} P^t(i,j) = \pi(j).$$

No matter initial state, we converge to π.

What is π?

In general, need to do Gaussian elimination to find it, & usually $|\Omega|$ is HUGE.

If P is symmetric then $\pi = \text{uniform}(\Omega)$

<u>Proof:</u> Need to verify $\pi P = \pi$, for $\pi(i) = \frac{1}{N}$, $N = |\Omega|$.

$$(\pi P)(i) = \sum_{k \in \Omega} \pi(k) P(k,i)$$

$$= \frac{1}{N} \sum_{k} P(k,i) = \frac{1}{N} \underbrace{\sum_{k} P(i,k)}_{=1 \text{ since } P \text{ is stochastic}} = \frac{1}{N}. \quad \blacksquare$$

$P(k,i) = P(i,k)$ since symmetric.

Weighted symmetric?

P is reversible with respect to π if:

$$\forall i,j \in \Omega, \quad \pi(i) P(i,j) = \pi(j) P(j,i)$$

then π is a stationary distribution.

<u>Proof:</u> $$(\pi P)(i) = \sum_{k \in \Omega} \pi(k) P(k,i) = \sum_{k} \pi(i) P(i,k) = \pi(i) \sum_{k} P(i,k)$$

$$= \pi(i). \quad \blacksquare$$

Random walk on d-regular, connected undirected $G=(V,E)$.

$$P(i,j)=P(j,i)=1/d$$ So it's symmetric

& $\pi(i)=1/n$.

Non-regular?

Then $\pi(i)=\frac{d(i)}{Z}$ where $d(i)$=degree of i

& $Z=\sum_j d(j)=2m$.

Check: $\pi(i)P(i,j)=\frac{d(i)}{Z}\frac{1}{d(i)}=\frac{1}{Z}=\pi(j)P(j,i)$

What if G is Directed?

Then no idea about π.

Random matchings:

For input graph $G=(V,E)$

let Ω = all matchings of G.

Goal: Sample uniformly from Ω.

Markov chain:

From $X_t \in \Omega$,

1. Choose e u.a.r. from E
2. Let $X' = X_t \oplus e = \begin{cases} X_t \cup e & \text{if } e \notin X_t \\ X_t \setminus e & \text{if } e \in X_t \end{cases}$
3. If $X' \in \Omega$, set $X_{t+1} = X'$ with prob. $1/2$ otherwise set $X_{t+1} = X_t$.

Aperiodic since $P(\sigma,\sigma) > 0$ by

& irreducible: for all $\sigma \in \Omega, \tau \in \Omega$, can go $\sigma \rightsquigarrow \emptyset \rightsquigarrow \tau$.

Thus it's ergodic & symmetric so $\pi = \text{uniform}(\Omega)$.

How fast do we reach π?

Mixing Time:

for $X_0 \in \Omega, \epsilon > 0$, $T_{mix}^{X_0}(\epsilon) = \min\{t : d_{TV}(P^t(X_0, \cdot), \pi) \le \epsilon\}$

& $T_{mix}(\epsilon) = \max_{X_0} \min\{t : d_{TV}(P^t(X_0, \cdot), \pi) \le \epsilon\}$

Suffices to do $\epsilon = 1/4$, $T_{mix} = T_{mix}(1/4)$.

then $T_{mix}(\epsilon) \le T_{mix}(\frac{1}{4}) \log(1/\epsilon)$.

ferromagnetic Ising Model: $L \times L$ box of $\mathbb{Z}^2$ = grid graph, $n = L^2$.

Configurations: $\Omega = \{+1, -1\}^V$

Energy $H(\sigma) = -\sum_{(i,j) \in E} \sigma_i \sigma_j = -(\#\text{monoedges} - \#\text{non-mono})$

$\beta > 0$, $\beta = \frac{1}{T}$ = inverse temperature

$w(\sigma) = e^{-\beta H(\sigma)}$ ← all mono. has most weight $e^{\beta m}$
all diff. has least $e^{-\beta m}$

Gibbs distribution: $\mu(\sigma) = \frac{w(\sigma)}{Z}$

Partition function $Z = \sum_{\sigma} w(\sigma)$.

Metropolis chain:

Metropolis filter

From $X_t \in \Omega$,

1. Choose $v \in V$ u.a.r. & $s \in \{+1, -1\}$ u.a.r.
2. For all $w \neq v$ set $X'(w) = X_t(w)$.
   & set $X'(v) = s$.
3. Set $X_{t+1} = X'$ with prob. $\min\left\{1, \frac{w(X')}{w(X_t)}\right\}$.
   else $X_{t+1} = X_t$.

Check: for $\sigma, \tau \in \Omega$, $\pi(\sigma)P(\sigma,\tau) = \pi(\tau)P(\tau,\sigma)$

let's say $w(\sigma) \leq w(\tau)$ so $P(\sigma,\tau) = 1$, $P(\tau,\sigma) = \frac{w(\sigma)}{w(\tau)}$.

Then: $\pi(\sigma)P(\sigma,\tau) = \frac{w(\sigma)}{Z} \times 1 = \frac{w(\sigma)}{Z}$

& $\pi(\tau)P(\tau,\sigma) = \frac{w(\tau)}{Z} \times \frac{w(\sigma)}{w(\tau)} = \frac{w(\sigma)}{Z}$. ✓

Alternative chain:

Gibbs sampler:

From $X_t \in \Omega$,

1. Choose $v \in V$ u.a.r.
2. Set $X_{t+1}(w) = X_t(w)$ $\forall w \neq v$
3. Choose $X_{t+1}(v)$ from $\mu\left(\sigma(v) \mid \sigma(w) = X_t(w) \text{ for all } w \in N(v)\right)$

thus $X_{t+1}(v) = \begin{cases} +1 & \text{with prob. } \frac{e^{-\beta(N^- - N^+)}}{Z_v} \\ -1 & \text{w.p. } \frac{e^{-\beta(N^+ - N^-)}}{Z_v} \end{cases}$

$Z_v = e^{-\beta(N^- - N^+)} + e^{-\beta(N^+ - N^-)}$

Check: Say $\sigma(v) = +$, $\tau(v) = -$, $\sigma(w) = \tau(w)$ $\forall w \neq v$.

$\pi(\sigma)P(\sigma,\tau) \propto \frac{e^{\beta(N^+ - N^-)}}{Z} \frac{e^{-\beta N^+ + \beta N^-}}{Z_v} = \frac{1}{Z Z_v}$

$\pi(\tau)P(\tau,\sigma) \propto \frac{e^{\beta(N^- - N^+)}}{Z} \frac{e^{-\beta N^- + \beta N^+}}{Z_v} = \frac{1}{Z Z_v}$ ✓

Coupling:

For a finite space Ω,
for distributions μ, ν on Ω
a distribution ω on $\Omega \times \Omega$ is
a coupling of μ, ν if:

rows sum to μ. for all $i \in \Omega$, $\sum_{j \in \Omega} \omega(i,j) = \mu(i)$

Columns sum to ν. for all $j \in \Omega$, $\sum_{i \in \Omega} \omega(i,j) = \nu(j)$.

Example:

$\Omega = \{HH, HT, TH, TT\}$

$\mu = (.5, .25, 0, .25)$, $\nu = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0)$

$$\omega = \begin{bmatrix} 1/3 & 1/12 & 1/12 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 0 \end{bmatrix}$$

Sample from ω is pair of $\sigma, \tau \in \Omega$
for $(\sigma, \tau) \sim \omega$,
then $\sigma \sim \mu$
& $\tau \sim \nu$.

Lemma: for coupling ω of μ, ν, let $(\sigma, \tau) \sim \omega$.

a) $d_{TV}(\mu, \nu) \leq \Pr(\sigma \neq \tau)$

b) $\exists$ coupling ω where

$$d_{TV}(\mu, \nu) = \Pr(\sigma \neq \tau).$$

Proof of (a):

Note for $\eta \in \Omega$, $\omega(\eta, \eta) \leq \mu(\eta)$ & $\omega(\eta, \eta) \leq \nu(\eta)$

thus: $\omega(\eta, \eta) \leq \min\{\mu(\eta), \nu(\eta)\}$

~~$\Pr(\sigma \neq \tau)$~~

Thus, $\Pr(\sigma = \tau) = \sum_{\eta} \omega(\eta, \eta) \leq \sum_{\eta} \min\{\mu(\eta), \nu(\eta)\}$

$$\begin{aligned}
\Pr(\sigma \neq \tau) &\geq 1 - \sum_{\eta} \min\{\mu(\eta), \nu(\eta)\} \\
&= \sum_{\eta} \mu(\eta) - \min\{\mu(\eta), \nu(\eta)\} \\
&= \sum_{\eta : \mu(\eta) \geq \nu(\eta)} \mu(\eta) - \nu(\eta) \\
&= \max_{S \subset \Omega} \mu(\eta) - \nu(\eta) \\
&= d_{TV}(\mu, \nu).
\end{aligned}$$

Proof of (b):

Set $\omega(\eta,\eta) = \min\{\mu(\eta), \nu(\eta)\}$

Thus, $d_{TV}(\mu,\nu) = \Pr(\sigma \neq \tau)$

Need to define off-diagonal entries of ω so that it's a valid coupling.

Put product distribution of rest. (HW problem)

$\blacksquare$