$P_{t+1}(x) \approx P_t(x)$ for large ts

At convergence as t $\to \infty$

$P_{t+1}(x) = P_t(x) = \pi(x) \leftarrow$ Stationary Distribution

We are interested in $\mathrm{T}(x|x')$ for which as t $\to \infty$

$$\pi(x) = \sum_{x'} \pi(x')T(x|x')$$

has a unique solution for a given $\mathrm{T}(x|x')$ and $\pi(x)$ should be reachable from any initial state x^0 via Markov walks using $\mathrm{T}(x|x')$

$|x_0$

$|x_0 \to$ probability of your being in state x after T MCMC steps-starting from x_0

x^t

m^n

$x = [x_1 \cdots x_n]$

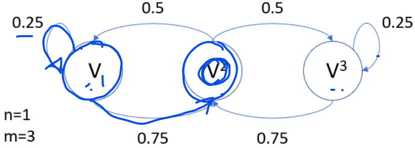
$x_i \in \{1, \cdots m\}$

Example

$|X| = 3$

$m=3, n=1$

For any $T(x|x')$ there always exists a solution to $\pi(x)$



$$\begin{array}{c|c} & x' \rightarrow \\ \hline x \downarrow & \begin{bmatrix} 0.25 & 0.5 & 0 \\ 0.75 & 0 & 0.75 \\ 0 & 0.5 & 0.25 \end{bmatrix} \end{array}$$

$\Sigma = 1$

$$\pi(x) = [\pi_1, \pi_2, \pi_3]$$

$$\pi(x) = \sum_{x'} \pi(x')\ T(x|x')$$

$$\boxed{\begin{aligned} \pi_1 &= 0.25\pi_1 + 0.5\pi_2 \\ \pi_2 &= 0.75\pi_1 + 0.75\pi_3 \\ \pi_3 &= 0.5\pi_2 + 0.25\pi_3 \\ \pi_1 + \pi_2 + \pi_3 &= 1 \end{aligned}}$$

$$\pi_1 = \pi_3 = 2/7,\ \pi_2 = 3/7$$

$$\pi = [2/7\ ,\ 3/7\ ,\ 2/7]$$

Example - with multiple solutions

There might exist $T(x|x')$ for which $\boxed{\pi(x) = \sum_{x'} \pi(x')\, T(x|x')}$ has multiple solutions

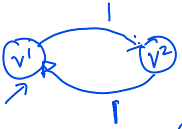
$T(x|x') = 1$ if $x = x'$
$T(x|x') = 0$ otherwise

$\pi(x)$ can be any distribution
Gives an Identity matrix
Hence Infinite $\#$ of solutions

$P_t(x) = \delta(x^0)$

Example - unreachable solution

Even when $T(x|x')$ has a unique stationary distribution $\pi_T(x)$, this distribution $\pi_T(x)$ may not be the one that we can reach or converge to via MCMC walks.



$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$
$T(x|x')$

$T(x|x') = 1$ if $x \neq x'$
$T(x|x') = 0$ otherwise

$\cancel{\pi_T(0) = \pi_T(1)} = 1/2 \leftarrow$ *unique*

$\pi_1 = \pi_2$

$$\begin{cases} \pi_1 = \pi_2 \\ \pi_2 = \pi_1 \\ \pi_1 + \pi_2 = 1 \end{cases}$$

$\sum_{x'} P_t(x') T(x|x') = P_{t+1}(x)$

Even as $t \to \infty$ $P_t(x) \neq P_{t+1}(x)$

t=0 $\quad x^0 = v_1 \qquad$ t is even $x^t = v^1$
t=1 $\quad x^1 = v_2 \qquad$ t is odd $x^t = v^2$

Not converging to $\pi_T(x)$

Ergodicity

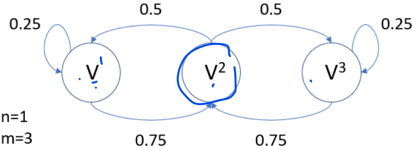
Definition: Markov chains with a unique $\pi(x)$ that can be reached via MCMC steps irrespective of starting point are called ergodic.

Theorem

When number of states is finite, then a MC is ergodic iff it is regular

Definition: A MC is regular if $\exists$k such that any x to x' can be reached in exactly k steps.

Examples of Ergodicity



n=1
m=3

$$\begin{bmatrix} 0.25 & 0.5 & 0 \\ 0.75 & 0 & 0.75 \\ 0 & 0.5 & 0.25 \end{bmatrix}$$

$\pi(x) = \left[\frac{2}{7} \quad \frac{3}{7} \quad \frac{2}{7}\right]$

$k=1$: $v^1 \to v^1$ - Yes ; $v^1 \to v^2 \to$ yes. $v^1 \to v^3$ - No

$k > 1$

$k=2$: $v^1 \to v^1 \to v^1$. Yes; $v^1 \to v^1 \to v^2 \to$ yes $v^1 \to v^2 \to v^3$ - Yes.

$v^2 \to v^1 \to v^1$ Yes $v^2 \to v^1 \to v^2$ Yes $v^2 \to v^3 \to v^3$ Yes

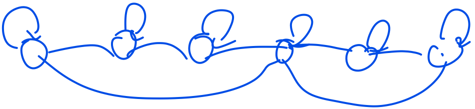
Yes Ergodic with $k=2$

Simpler Sufficient condition for Ergocity

Lemma

A finite state MC is regular when [not iff]:

1. *every state has a self loop with non-zero probability*
2. *between any two states $\exists$ a path of non-zero probability*



Gibbs sampling and Ergodicity

MC chain defined by Gibbs sampling is ergodic and has stationary distribution $P(\mathbf{x} = x_1 \ldots, x_n)$ when $P(\mathbf{x})$ is positive

Proof:

1. Self loop probability $> 0\ \forall \mathbf{x}$ $\quad T(\mathbf{x}|\mathbf{x}) > 0$ if $P(\mathbf{x})$ is a positive distribution
2. $\mathbf{x}' \to \mathbf{x}$ can be reached with non-zero probability [with n steps can always go from any x to x'].
3. We will next show that $\pi(\mathbf{x}) = P(\mathbf{x})$ when $T(\mathbf{x}|\mathbf{x}')$ as per *page 3 definition of G.S.*

$$\pi(\mathbf{x}) = \sum_{\mathbf{x}'} \pi(\mathbf{x}') \boxed{T(\mathbf{x}|\mathbf{x}')}$$
$$= \sum_{\mathbf{x}'} \boxed{P(\mathbf{x}')} \boxed{T(\mathbf{x}|\mathbf{x}')}$$
$$= \sum_{x_1'} \boxed{P(x_1', x_2, .., x_n)} \frac{1}{n} \boxed{P(x_1|x_2, .., x_n)} + \ldots$$
$$+ \frac{1}{n} \sum_{x_n'} P(x_1, .., x_{n-1}, x_n') P(x_n|x_1, .., x_{n-1})$$

$$
\begin{aligned}
&= \boxed{P(x_2, .., x_n)\frac{1}{n}P(x_1|x_2, .., x_n)} + \ldots \\
&\qquad + \frac{1}{n}P(x_1, .., x_{n-1})P(x_n|x_1, .., x_{n-1}) \\
&= \frac{1}{n}P(x_1, x_2.., x_n) + \ldots + \frac{1}{n}P(x_1, .., x_{n-1}, x_n) \\
&= P(x_1, x_2.., x_n) \equiv P(\mathbf{x}) = \pi(x)
\end{aligned}
$$

This shows that $P(x_1, .., x_n)$ is a S.D. of $T(\mathbf{x}|\mathbf{x}')$ defined as in Gibbs sampling

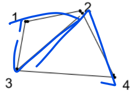
Calculating $P(x_i|\mathbf{x}_{-i})$ for undirected graphical models

$P(x_1, x_2.., x_n) = \frac{1}{Z}\prod_c \psi_c(\mathbf{x}_c)$

$$P(x_i|\mathbf{x}_{-i}) = \frac{P(x_1, x_2.., x_n)}{\sum_{x_i} P(x_1, x_2.., x_n)} = \frac{\frac{1}{z}\prod_c \psi_c(x_c)}{\frac{1}{z}\sum_{x_i}\prod_c \psi_c(x_c)}$$

$$= \frac{\prod_{c:i\in C}\psi_c(x_c)}{\sum_{x_i}\prod_{c:i\in C}\psi_c(x_c)}$$

Example-



$P(x_1, x_2, x_3, x_4)$

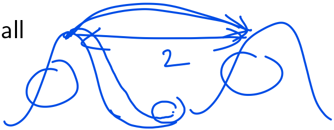
$$P(x_1 \mid \mathbf{x}_{-1}) = \frac{\Psi_{123}(x_1, x_2, x_3)\,\Psi_{234}(x_2, x_3, x_4)}{\sum_{x_1}\psi_{123}(x_1, x_2, x_3)\,\Psi_{234}(x_2, x_3, x_4)} = \frac{\Psi_{123}(x_1, x_2, x_3)}{\sum_{x_1}\psi_{123}(x_1, x_2, x_3)}$$

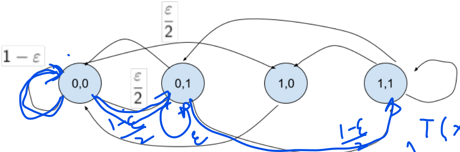
Limitations of Gibbs

1. For continuous distribution $P(x_i|x_{-i})$ may not be easy to obtain.
2. When variables are highly correlated Gibbs sampling which was only local moves might have high probability states whose neighbours are mostly low probability leading to poor mixing.

Example of poor mixing

$$P(x_1, x_2) = \frac{1-\varepsilon}{2} \text{ if } x_1 = x_2 \quad x_i \in \{0,1\}$$
$$= \frac{\varepsilon}{2} \quad \text{if } x_1 \neq x_2 \quad \varepsilon \text{ is small}$$





$$T(x \mid x') = \frac{1}{n} \sum_{i=1}^{n} P(x_i \mid x_{-i}) \quad \text{if } x = x'$$

$P(x_1 \mid X_{-1} = [0])$ $\quad$ $P(x_2 \mid x_{-2} = [0])$

$$T([0,0] \| [0,0]) = \frac{1}{2}\left[P(x_1 = 0 \mid x_2 = 0) + P(x_2 = 0 \mid x_1 = 0)\right]$$
$$= \frac{P(x_1 = 0, x_2 = 0)}{P(x_1 = 0, x_2 = 0) + P(x_1 = 1, x_2 = 0)} + \frac{P(x_1 = 0, x_2 = 0)}{P(x_1 = 0, x_2 = 0) + P(x_1 = 0, x_2 = 1)}$$
$$= \frac{(1-\varepsilon)/2}{\left(\frac{1-\varepsilon}{2}\right) + \frac{\varepsilon}{2}} + \frac{(1-\varepsilon)/2}{\left(\frac{1-\varepsilon}{2}\right) + \frac{\varepsilon}{2}} = 1 - \varepsilon$$

A broader class of Markov chains KF 12.3.4: Metropolis Hastings Sampling

Motivated by the need to design moves that go from one high probability state to another without passing through low probability states.

Metropolis Hastings Algorithm

1. Choose any proposal distribution for transferring from x to x' $T^Q(x \to x')$ = $T^Q(x'|x)$
2. Use T^Q to propose a transition from x to x'. We accept the proposal with probability $A(x \to x')$ and transition, or stay in x.

$$\boxed{\begin{aligned} T(x \to x') &= T^Q(x \to x') A(x \to x') \quad x \neq x' \\ T(x \to x) &= T^Q(x \to x) + \sum_{x' \neq x} T^Q(x \to x')(1 - A(x \to x')) \end{aligned}}$$

How to design A?

Reversible Chains

Definition: A finite state Markov chain T is reversible if $\exists$ a unique π such that $\forall x, x' \in \chi$

$$\pi(x')T(x' \to x) = \pi(x)T(x \to x')$$

Above is called the Detailed balance Equation (DBE)

Theorem

If $\pi(x)$ satisfies above then $\pi(x)$ is a stationary distribution of T.

Proof- $\sum_{x'} \pi(x') T(x' \to x) = \sum_{x'} \pi(x) T(x \to x')$

$T(x \to x') = P(x'/x)$

$$\sum_{x'} \pi(x')T(x' \to x) = \pi(x) \underbrace{\sum_{x'} T(x \to x')}_{=1} = \pi(x)$$

$\sum_{x'} \pi(x') T(x|x') = \pi(x) \Rightarrow \pi(x)$ is s.d of T.

Example: Reversibility check for Gibbs

Choosing A

Design A to satisfy detailed balance equation for $x \neq x'$

$$\pi(x) T^Q(x \to x') A(x \to x') = \pi(x') T^Q(x' \to x) A(x' \to x)$$

$$A(x \to x') = \min\left[1, \frac{\pi(x') T^Q(x' \to x)}{\pi(x) T^Q(x \to x')}\right] \text{ satisfies this}$$

Given a desired stationary distribution $P(\mathbf{x})$, designing the $A(\cdot)$ just requires the (user provided) T^Q and the ratio of probabilities $\frac{P(\mathbf{x}')}{P(\mathbf{x})}$.

Example from book

Let us we desire a stationary distribution: $\pi = [2/7 \ , \ 3/7 \ , \ 2/7]$
Earlier we had started with a T that gave this π. Not we choose an arbitrary T^Q and compute A.
Example $T(x \to x') = 1/3$
Compute $A(1 \to 2)$
Compute $A(2 \to 3)$