Markov Chains:

Stochastic Process \( \{ X_t : t \in T \} \) (assume that \( T = \{1, 2, 3, \ldots \} \)) with the property that \( f(x_t | x_{t+1}) = f(x_t | x_{t+1}) \)

\[ f(x_1, x_2, \ldots, x_n) = f(x_1) f(x_2 | x_1) f(x_3 | x_2) \cdots f(x_n | x_{n-1}) . \]

State space: \( X \) (define for now)

"States": 1, 2, ...

![Diagram of Markov chain]

The matrix \( P \) of transition probabilities is the transition matrix. \( P_{ij} = p_{ij} \).

Two properties:
1. \( p_{ij} > 0 \)
2. \( \sum_j p_{ij} = 1 \) (Typo in book.)

In each row, \( i \), in \( P \) is a probability mass function.
n-step transition probability: \( P(X_{m+n} = j \mid X_m = i) = p_{ij}(n) \)

**Theorem (Chapman-Kolmogorov)**: The n-step transition probabilities satisfy:
\[
p_{ij}(m+n) = \sum_k p_{ik}(m) p_{kj}(n) = (P(m) P(n))_{ij}
\]

\[\Rightarrow P(2) = P \cdot P = P^2\]
\[\Rightarrow P(3) = P^3\]
\[\Rightarrow P(n) = P^n\]

This means that if at time 0, my probability of being in state \( i \) is \( \mu_i \), and define
\[
\mu(0) = (\mu_0, \mu_0, \ldots, \mu_0)
\]
\[\Rightarrow \mu(n) = \mu(0) P^n \]

**Question**: As \( n \to \infty \), is \( \mu_i(n) > 0 \)? Or is \( p_{ij} > 0 \) for all \( i, j \)?

**Def**: state \( i \) reaches state \( j \) (\( j \) is accessible from \( i \)) if \( p_{ij}(n) > 0 \) for some \( n \)
\[\Rightarrow i \rightarrow j\]
\[\Rightarrow \text{if } i \rightarrow j \text{ and } j \rightarrow i, \text{ then } i \leftrightarrow j \text{ is ''communicate''} \]
Thus

1. \( i \leftrightarrow i \)
2. \( i \leftrightarrow j \Rightarrow j \leftrightarrow i \)
3. \( i \leftrightarrow j \) and \( j \leftrightarrow k \) then \( i \leftrightarrow k \).

4. The state space \( X \) can be written as a disjoint union of classes

\[ X = X_1 \cup X_2 \cup \ldots \]

where \( i, j \) communicate iff \( i, j \in X_k \).

**Def:** If all states communicate, then the chain is **irreducible**.

**Closed:** set of states is closed if the chain enters but never leaves.

**Closed set with a single state:** an absorbing state.

**Recurrent/persistent:** \( P(X_n = i \text{ for some } n \geq 1 \mid X_0 = i) = 1 \)

**Transient:** \( \text{else} \).

**Stationarity:** \( \pi \) is a stationary (or invariant) distribution if

\[ \pi = \pi P. \]

\[ \Rightarrow \pi \text{ is a row eigenvector of } P \]

\[ \Rightarrow P^T \pi = \pi \]

\[ \Rightarrow \text{with eigenvalue } 1. \]
Idea: Draw $X_0$ from $\pi$, a stationary distribution of $P$.

Next, draw $X_1 \sim \pi P$.

Notationally: $X_1 \sim \mu_1 = \mu_0 P = \pi P = \pi$

$\Rightarrow$ If $X_2 \sim \mu_2 = \mu_1 P = \mu_0 P^2 = \pi P = \pi$

$\Rightarrow$ that $X_2 \sim \pi$

When a chain has distribution $\pi$, it will forever.

Def A Markov Chain has limiting distribution $\pi$ if $P^n \to \left(\begin{array}{c} \pi \\ \pi \\ \vdots \\ \pi \end{array}\right) = \left(\begin{array}{c} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_N \end{array}\right)$

$\Rightarrow \mu_0 P^n = \pi$

$\left(\begin{array}{c} \pi_i \\ \pi_i \\ \vdots \\ \pi_i \end{array}\right) = \left(\begin{array}{c} \pi_1 \sum_{M} \pi_j \\ \pi_2 \sum_{M} \pi_j \\ \vdots \\ \pi_N \sum_{M} \pi_j \end{array}\right)$

Detailed Balance $\pi$ satisfies detailed balance if

for all $i, j$$\pi_i P_{ij} = \pi_j P_{ji}$

$\frac{P(X_n = i) P(X_{n+1} = j \mid X_n = i)}{P(X_{n+1} = j, X_n = i)}$
Thus if \( \Pi \) satisfies detailed balance, then 
\( \Pi \) is a stationary distribution.

**Proof:** Detailed balance says \( \Pi_i P_{ij} = \Pi_j P_{ji} \)

We need to show that \( \Pi P = \Pi \). The \( j \)th element 
of \( \Pi P = (\Pi P)_j = \sum_{k=1}^{N} \Pi_k P_{kj} \).

\( \Pi P = \sum_{k=1}^{N} \Pi_k P_{kj} = \sum_{k=1}^{N} \frac{\tau_k}{\sum_{j=1}^{N} \tau_j} \pi_j P_{kj} = \frac{\sum_{j=1}^{N} \tau_j P_{jk} \pi_j}{\sum_{j=1}^{N} \tau_j} \pi_j = \pi_j \). \( \checkmark \)

**Markov Chain Monte Carlo (MCMC)**

**Goal:** Estimate an integral \( E(h(X)) = \int h(x) f(x) \, dx \).

**Idea:** Construct a Markov Chain \( X_1, X_2, \ldots \)
whose stationary distribution is \( \Pi \)

\[ \Rightarrow X_n \sim F = \int f \]

We're specifying \( \Pi_i \) and trying to find \( P \)
such that \( \Pi = \Pi P \).

If this can be done, then under certain assumptions

\[ \frac{1}{N} \sum_{i=1}^{N} h(X_i) \xrightarrow{P} E(h(X)) \].

For example: Draw from posterior in Bayesian calculation:

\[ f(\theta | x) = \frac{L(\theta) f(\theta)}{\int L(\theta) f(\theta) \, d\theta} \]
Specific Algorithm: Metropolis - Hastings.

Listed as one of top 10 algorithms of 20th century (along with FFT, FMM, QR, Fortran).

Goal: Draw samples from $X$ with density $f$.

**M-H Algorithm**

1. Choose $X_0$ arbitrarily. Assuming that we have generated $X_0, \ldots, X_i$:
   - **Generate $Y$ from density $q(y|X_i)$**
     - $q$ is a density that is easy to draw from: proposal distribution.
     - Example: $q(y|x) \sim N(x, \sigma^2)$.

2. Evaluate $r = r(X_i, Y)$ where
   $$r(x,y) = \min \left\{ \frac{f(y)}{f(x)} \frac{q(x|y)}{q(y|x)}, 1 \right\}$$

3. Set $X_{i+1} = \begin{cases} Y & \text{with probability } r \\ X_i & \text{with probability } 1-r \end{cases}$

Completely opaque algorithm: look at specific example first before understanding why it works.
Ex: Draw from Cauchy distribution $f(x) = \frac{1}{\pi(1+x^2)}$. Take $q(y|x) = \frac{1}{\sqrt{2\pi}b} e^{-|y-x|^2/2b^2}$. So then $r(x,y) = \min\left\{ \frac{f(y)q(x|y)}{f(x)q(y|x)}, 1 \right\}$

$$= \min\left\{ \frac{1+x^2}{1+y^2} \frac{e^{-(x-y)^2/2b^2}}{e^{-(y-x)^2/2b^2}}, 1 \right\}$$

$$= \min\left\{ \frac{1+x^2}{1+y^2}, 1 \right\}$$

So the algorithm reduces to following:

$$X_{c+1} = \begin{cases} Y \sim N(X_c, b^2) \text{ with probability } r(x,y) \\ X_c \text{ with prob. } 1 - r(x,y) \end{cases}$$

**Note:**

Why does this algorithm work at all? Short answer: we enforce detailed balance in the chain, therefore guaranteeing the existence of a stationary distribution.
Recall: \( p_{ij} \pi_i = p_{ji} \pi_j \)

Continuous version of detailed balance:
\[
p_{ij} \to p(x,y) \propto \Pi(x_{n+1} = y \mid x_n = x) \]
\[
\pi_i \to f(x) \propto \Pi(x_n=x) .
\]

The function \( f \) is a stationary distribution if
\[
f(y) = \int p(x,y) f(x) \, dx
\]

\[\Rightarrow \text{ Detailed Balance then means that} \]
\[
f(x) p(x,y) = f(y) p(y,x)
\]

If this equation holds, then just integrate each side to show that \( f \) is a stationary distribution.

Using the construction of the M-H algorithm, show that detailed balance is satisfied, and therefore \( f \) is the stationary distribution.

Consider \( x, y \) (i.e. \( x = X_i \), and \( y = Y \), the proposal value).

Either
\[
f(x) q(y \mid x) < f(y) q(x \mid y)
\]
or
\[
f(x) q(y \mid x) > f(y) q(x \mid y)
\]
Without loss of generality, assume that (*) holds, and we then have:

\[
\frac{f(y) \cdot q(x \mid y)}{f(x) \cdot q(y \mid x)} > 1
\]

and therefore

\[
r(x \mid y) = \frac{f(y) \cdot q(x \mid y)}{f(x) \cdot q(y \mid x)}.
\]

(And obviously \( r(y, x) = \min \left\{ \frac{f(x) \cdot q(y \mid x)}{f(y) \cdot q(x \mid y)} \right\} = 1 \).

Next, compute the transition probabilities:

\[
p(x \mid y) = P(x \rightarrow y) \quad \text{and require that}
\]

\begin{enumerate}
  
\item generate \( y \)
\item accept \( y \)
\end{enumerate}

\[\Rightarrow p(x \mid y) = q(y \mid x) \cdot r(x \mid y) = q(y \mid x) \cdot \frac{f(y) \cdot q(x \mid y)}{f(x) \cdot q(y \mid x)} = \frac{f(y) \cdot q(x \mid y)}{f(x)}\]

\[\Rightarrow f(x) \cdot p(x \mid y) = f(y) \cdot q(x \mid y)\]

On the other hand, \( p(y, x) = P(y \rightarrow x) \) and require:

\begin{enumerate}
  \item generate \( x \)
  \item accept \( x \)
\end{enumerate}
\[ p(y|x) = \frac{q(x|y)}{r(y,x)} = q(x|y) \]

And then for \( f(x) p(x|y) = f(y) p(y,x) \). \( \checkmark \)

This is detailed balance.

Monte Carlo methods:

\[ \int h(x) f(x) \, dx \approx \frac{1}{N} \sum_{j=1}^{N} h(X_i) \quad \text{when } X_i \sim \text{samp\ from } f \]

\[ I \]

\[ E(I) = \int h f \]

\[ \text{Var}(I) \propto \frac{1}{N} \quad \Rightarrow \quad \text{std}(I) \sim \frac{1}{\sqrt{N}} \]