Lecture 3, September 29th, 2025

1. Markov chains and ergodic theorem
2. Markov chain Monte Carlo

Q: Why aren't we happy with Monte Carlo?
A: Monte Carlo algorithms often don't scale to high dimensional problems, so we cannot perform Bayesian inference with Monte Carlo when we have many model parameters.

It turns out that relaxing the independence assumption in Monte Carlo helps. The easiest way to construct dependent samples is using Markov chains.

Def. An infinite sequence of discrete random variables $X_1, X_2, X_3, \ldots$ ($\{X_n\}$) is called a Markov chain if for all $n \geq 0$ and for all non-negative integers $i_0, i_1, \ldots, i_{n-1}, i, j$

$$\Pr(X_{n+1}=j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \Pr(X_{n+1}=j \mid X_n = i)$$

Def. If $\Pr(X_{n+1}=j \mid X_n=i) = \Pr(X_{m+1}=j \mid X_m = i)$ for all $m, n \geq 0$, then the Markov chain is called homogeneous.

Note 1: A homogeneous Markov chain on a finite state space $S=\{1,\dots,s\}$ is defined by its initial distribution $\bar{\nu}\begin{pmatrix}\nu_1\\ \vdots\\ \nu_s\end{pmatrix}$, $\nu = \nu u$

$P \cdot \underline{1} = \underline{1}$

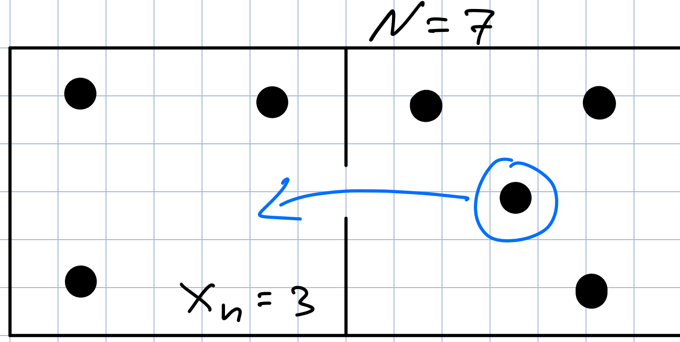
and transition probability matrix $P=\{p_{ij}\}$, where

$\nu_i = Pr(X_0=i)$ and $p_{ij} = Pr(X_1=j \mid X_0=i) = Pr(X_5=j / X_4=i)$

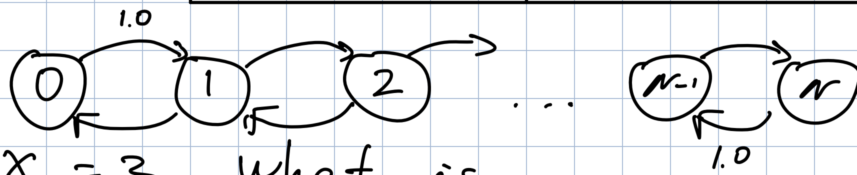
Example: Ehrenfest model of diffusion

Consider a 2-dimensional rectangular box with a divider in the middle. The box contains N balls (gas molecules) distributed somehow between the 2 parts of the box. The divider has a small gap, through which the balls can go one at a time. At each time step we select a ball uniformly at random and force it to go through the gap to the opposite side of the box. Let X_n be the number of balls in the left half of the box. By construction $\{X_n\}$ is a Markov chain. Let's work out its transition probabilities.

Q: What is our state space (a set of all possible values of X_n)?



A: $S = \{0, 1, \dots, N\}$



Q: Say $N = 7$ and $X_n = 3$. What is

$$\Pr(X_{n+1} = 4 \mid X_n = 3)?$$

$$2$$

$$5$$

A: $\Pr(X_{n+1} = 4 \mid X_n = 3) = \frac{4}{7}$

$$2 \qquad = \frac{3}{7}$$

$$5 \qquad = 0$$

In general,

\# of molecules on the right

$$p_{ij} = \begin{cases} \frac{N-i}{N} = 1 - \frac{i}{N}, & \text{if } j = i+1 \\ \frac{i}{N} & \text{if } j = i-1 \\ 0 & \text{otherwise} \end{cases}$$

<u>Def.</u> Probability vector $\underline{\pi}$ is called a stationary distribution with respect to a transition probability matrix P if $\underline{\pi}^T P = \underline{\pi}^T$ (global balance) – difficult to check or solve for $\underline{\pi}$ when state space S is large

<u>Def</u>. A probability vector $\underline{\pi}$ is said to satisfy detailed balance equations with respect to transition probability matrix P if

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \text{for all } i, j$$

— easier to check than global balance

Proposition: detailed balance implies global balance

Proof: $\pi_i p_{ij} = \pi_j p_{ji} \Rightarrow \sum_j \pi_i p_{ij} = \sum_j \pi_j p_{ji} \Rightarrow$

$$\Rightarrow \pi_i \underbrace{\sum_j p_{ij}}_{\text{law of total probability}} = \sum_j \pi_j p_{ji} \Rightarrow \pi_i = \sum_j \pi_j p_{ji} \Rightarrow$$

$$\Rightarrow \underline{\pi}^T = \underline{\pi}^T P \quad \blacksquare$$

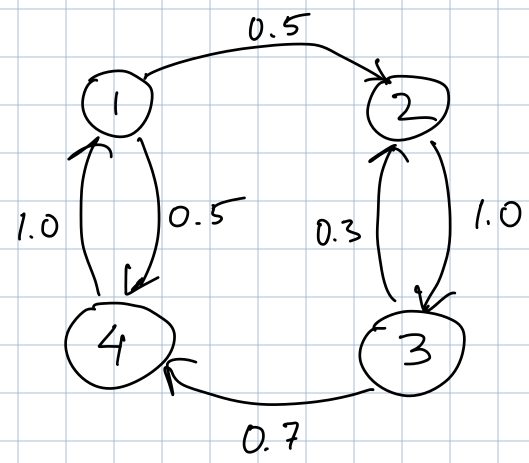
Example: Ehrenfest model of diffusion

Binomial $(N, \frac{1}{2})$ is the stationary distribution. Proof of this will be your homework exercise

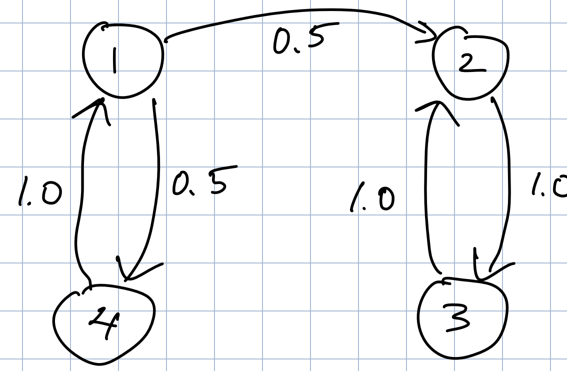
<u>Def</u> A Markov chain is called irreducible if it can get from any state to any other state in a finite number of steps with positive probability.

Examples

Irreducible



Non-irreducible



Def An irreducible Markov chain is called **recurrent** if starting from any state the chain returns to this state eventually with probability one. The recurrent chain is called **positive recurrent** if all expected return times are finite.

Ergodic Theorem

Let $\{X_n\}$ be an irreducible and positive recurrent Markov chain and let $f(\cdot)$ be a function such that $\sum f(i)\pi_i < \infty$. Then

$$\underbrace{\frac{1}{N}\sum_{k=1}^{N} f(X_k)}_{\text{time average}} \longrightarrow \underbrace{\sum_{i\in S} f(i)\pi_i}_{\text{space average}} = E_{\pi}[f(X)]$$

Example: $f = 1_{\{X=2\}}$ $\rightarrow$ $\frac{1}{N}\sum 1_{\{X_k=2\}} \rightarrow 0\cdot\pi_1 + 1\cdot\pi_2 + \dots + = \pi_2$

computer demo here

Markov chain Monte Carlo

Big idea: suppose we want to study some distribution; we want to compute its mean, variance, quantiles, etc. We will design/engineer (completely artificially) a Markov chain in such a way that the distribution of interest is the stationary distribution of this Markov chain. Then, we will simulate a Markov chain path and use the ergodic theorem.

Rosenbluth-Hastings Algorithm (aka Metropolis-Hastings)

Objective: $E_\pi[h(X)] = \sum_{x \in E} \pi_x h(x)$

$\pi^T = (\pi_1, \dots, \pi_s)$ – target distribution

(In Bayesian inference, $\pi = P(\bar{\theta} | \bar{y}) = \frac{1}{c} P(\bar{y} | \bar{\theta}) P(\bar{\theta})$, c-?)

1. Start with value $X_0 = x_0$

for $n=0$ to N do

simulate $y \sim q(j \mid X_n = i)$, suppose $y=j$

compute R-H acceptance probability

$$a_{ij} = \min \left\{ \frac{\cancel{\frac{1}{c}} \pi_j \, q(i|j)}{\cancel{\frac{1}{c}} \pi_i \, q(j|i)}, 1 \right\}$$

generate $U \sim \mathrm{Unif}(0,1)$

accept $y=j$ with probability a_{ij}, which means that

$$X_{n+1} = \begin{cases} j \ (y) & \text{if } U \le a_{ij} \\ i \ (X_n) & \text{if } U > a_{ij} \end{cases}$$

Proposition: Rosenbluth-Hastings algorithm generates a Markov chain with stationary distribution π.

Pf/ This will be your homework.

Hint: transition probabilities of the R-H Markov chain: $p_{ij} = g(j|i) \times a_{ij}$

If detailed balance works, then we are done.

computer demo here