Monte Carlo

Monte Carlo methods aim to solve numerical integration problems

Why do we care?

Because Bayes!

Posterior distribution: $\Pr(\overset{\text{parameters}}{\bar{\theta}} \mid \overset{\text{data}}{\bar{y}}) = \dfrac{\Pr(\bar{y} \mid \bar{\theta})\Pr(\bar{\theta})}{\Pr(\bar{y})}$

Bayesian point estimate:

$$\underbrace{E(\bar{\theta} \mid \bar{y})}_{\text{posterior mean}} = \int \bar{\theta} \Pr(\bar{\theta} \mid \bar{y}) \, d\bar{\theta}$$

This $\bar{\theta}$ could be $\bar{\theta} = (\theta_1, \theta_2, \ldots, \theta_{1000})$

Classical Monte Carlo

Objective: $E(h(X)) = \int h(x) f(x) \, dx$

e.g. $h(x) = x^2$

r.v. with prob density $f(x)$

Use law of large numbers by generating $\underline{X_1, \ldots, X_n} \overset{iid}{\sim} f(x)$

$$\bar{h}_n = \frac{1}{n} \sum_{i=1}^{n} h(x_i) \xrightarrow[n \to \infty]{a.s.} E(h(X))$$

$$\mathrm{Var}(\bar{h}_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} h(X_i)\right) =$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \underbrace{\mathrm{Var}(h(X_i))}_{\mathrm{Var}(h(X_1))} = [\text{iid}] = \frac{1}{n^2} \times n \,\mathrm{Var}(h(X_1))$$

$$= \frac{1}{n} \mathrm{Var}(h(X_1)) \approx \frac{1}{n} \underbrace{\frac{1}{n-1} \sum_{i=1}^{n} [h(X_i) - \bar{h}_n]^2}_{v_n}$$

can be computed from the samples

95% confidence interval for $E(h(X))$

is $\quad \bar{h}_n \pm 1.96 \sqrt{\frac{v_n}{n}} \xrightarrow[n \to \infty]{} 0 \qquad v_n \approx \mathrm{Var}(h(X))$

$\frac{1}{\sqrt{n}} \to 0$ slowly

Examples of $h(x)$: $\quad h(x) = x, \quad h(x) = x^2$

$$h(x) = \begin{cases} 1, & x > 2.3 \\ 0, & x \le 2.3 \end{cases}$$

$$E(h(X)) = E(1_{\{X > 2.3\}}) =$$

$$= \Pr(X > 2.3)$$

<u>Example</u>: $\quad X \sim \mathrm{Beta}(2,2)$

$E(X)$ – first moment

Objective: $\quad E(X^2)$ – second moment

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) = \mathrm{Var}(X) + [E(X)]^2$$

$= 0.3$ – suppose we don't know this.

because wikipedia is down

In R session:

$\bar{x}$ = rbeta (1000, 2, 2)

$\bar{y} = \bar{x}^2$

mean ($\bar{y}$)

var ($\bar{y}$)

Importance Sampling

Objective: $E_f(h(x))$, where $X \sim f(x)$

But suppose generating samples from $f(x)$ is hard. However, generating samples from a distribution with density $g(x)$ is easy.

$$E_f[h(x)] = \int h(x) f(x) dx =$$

$$= \int h(x) f(x) \frac{g(x)}{g(x)} dx = \int \underbrace{\boxed{h(x) \frac{f(x)}{g(x)}}}_{z(x)} g(x) dx$$

$$= E_g\left[h(x) \frac{f(x)}{g(x)}\right] \approx \frac{1}{n} \sum \frac{h(y_i) f(y_i)}{g(y_i)},$$

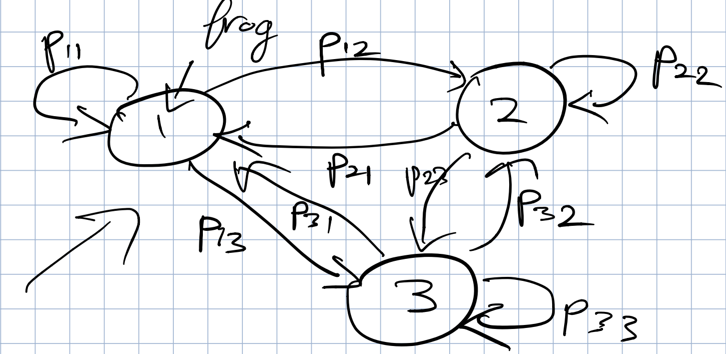
where $y_1, \ldots, y_n \overset{iid}{\sim} g(x)$

Markov chain theory

Markov property: $Pr(X_{n+1}=j \mid X_n=i, \cancel{\dots, X_0=i_0}) =$

$$= Pr(X_{n+1}=j \mid X_n=i)$$

Example:



$X_0 = 1$

$X_1 = 2$

$X_2 = 2$

$X_3 = 2$

$X_4 = 3$

$X_5 = 1$

$X_6 = 1$

$$Pr(X_7 = 3 \mid X_6 = 1, \cancel{X_5 = 1, \dots, X_1 = 1})$$

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix} \quad \begin{array}{l} p_{11}+p_{12}+p_{13}=1 \\ \dots = 1 \\ = 1 \end{array}$$

$P = \{p_{ij}\}$ – transition probability matrix

$p_{ij} = Pr(X_{n+1}=j \mid X_n = i) =$ [homogeneous] $=$

$= Pr(X_1 = j \mid X_0 = i)$ for all n.

$Pr(X_0 = i) = \nu_i$, $\bar{\nu} = (\nu_1, \nu_2, \nu_3)$

e.g., $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$

$\bar{\nu}, P$ – defines a Markov chain

$Pr(X_n = j \mid X_0 = i) = p_{ij}^{(n)}$ – n-step transition probability

$P^{(n)} = \{p_{ij}^{(n)}\}$ - n-step transition prob. matrix

$P^{(n)} = P^n \leftarrow$ matrix raised into power n

$$P^n = \underbrace{P \cdot P \cdot \ldots P}_{n \text{ times}}$$

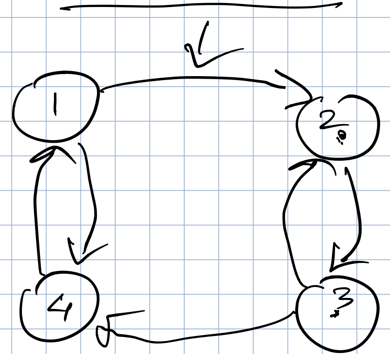
$Pr(X_n = i) = \nu_i^{(n)}$, $\bar{\nu}^{(n)} = (\nu_1^{(n)}, \ldots, \nu_S^{(n)})$

S ← # of states

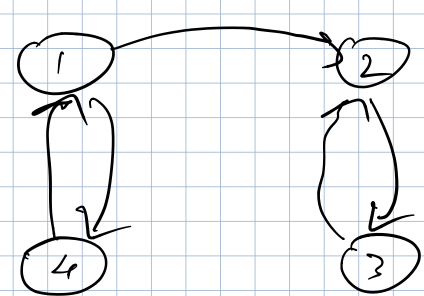
$\bar{\nu}^{(n)} = \bar{\nu} \cdot P^n$

Def. Markov chain is called irreducible if you can from anywhere to anywhere in a finite number of steps.

<u>Irreducible:</u>

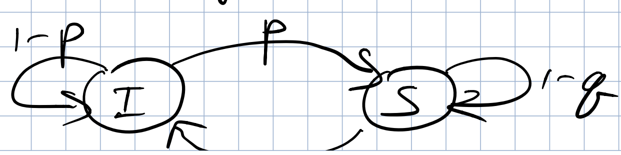


non-irreducible



impossible to get from 3 to 1

<u>Example:</u>



q

$$\begin{array}{c|cc} & I & S \\ \hline I & 1-p & p \\ S & q & 1-q \end{array} \qquad \begin{array}{l} 1<p<0 \quad \text{irreducible} \\ 1<q<0 \end{array}$$

Stationary and long term behavior

$\bar{\nu}^{(0)} = \bar{\nu}$ - initial distribution

$\bar{\nu}^{(1)} = (\nu_1^{(1)}, \dots, \nu_S^{(1)})$, $\nu_i^{(1)} = \Pr(X_1 = i)$

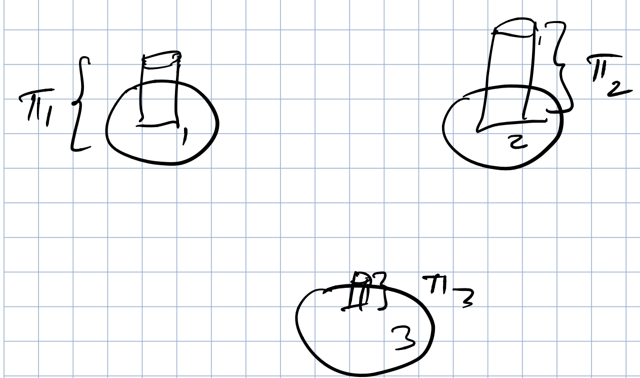
$\bar{\nu}^{(1)} = \bar{\nu} \cdot P$

$\bar{\nu}^{(2)} = \bar{\nu}^{(1)} \cdot P = \bar{\nu} \cdot P^2 \dots$

What if we have a special vector

$\bar{\pi}^{T} = (\pi_1, \dots, \pi_S)$ ← transpose $\qquad \pi_1 + \dots + \pi_S = 1$

$$\underline{\bar{\pi}}^{T} = \underline{\bar{\pi}}^{T} \cdot P$$



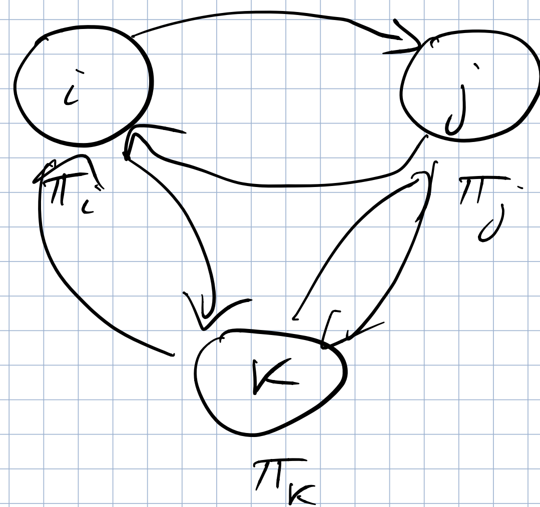
Reversible Markov chains + detailed balance

If we have a vector $\bar{\pi}^{T} = (\pi_1, \dots, \pi_S)$

and we can show that

detailed balance $\longrightarrow$ $\pi_i p_{ij} = \pi_j p_{ji}$ for all pairs (i, j)

then π - stationary distribution



$$\sum_{i=1}^{S} \pi_i p_{ij} = \sum_{i=1}^{S} \pi_j p_{ji} = \pi_j \underbrace{\sum_{i=1}^{n} p_{ji}}_{=1.0}$$

$$\boxed{\pi_j = \sum_{i=1}^{S} \pi_i p_{ij}} \Longleftrightarrow \pi^T = \pi^T P$$

detailed balance $\Rightarrow$ global balance