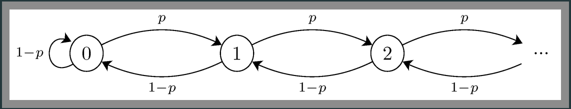
Markov chains: the ergodic theorem

'averages over space = averages over time'



- Steady-state distn:

$$\pi(i) = \frac{1}{2}\left(\frac{p}{1-p}\right)^i$$

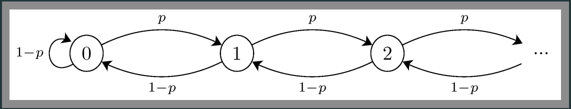
- Want - $\mathbb{E}_{X\sim\pi}[\phi(X)] = \sum_{i=0}^{\infty} \phi(i)\,\pi(i)$ ('space' average)

- Easier to compute - $\frac{1}{T}\sum_{t=1}^{T} \phi(X_t)$ (time average)

Ergodic Thm (for positive recurrent Markov Chains)

$$\boxed{\mathbb{E}_{X\sim\pi}[\phi(X)] = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \phi(X_t)}$$

(for any X_0)

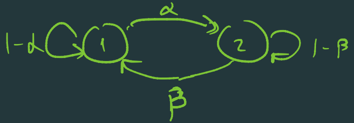


Idea - Want MC $X_1, X_2, \ldots X_n$ such that $X_n X_{n-1} \ldots X_1$ is a MC with the same transition probability matrix P (reversible MC)

'Magical'

- Thm - A MC P is reversible iff stationary dist^n π obeys $\underbrace{\pi(i) P_{ij} = \pi(j) P_{ji}}_{\text{Local balance}} \quad \forall\, i, j$

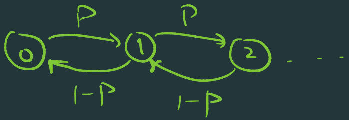
• Eg -



$$\Pi = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$$

$$1\to 2 : \underbrace{\frac{\beta}{\alpha+\beta}}_{\pi(1)} \cdot \underbrace{\alpha}_{P_{12}} = \underbrace{\frac{\alpha}{\alpha+\beta}}_{\pi(2)} \cdot \underbrace{\beta}_{P_{21}} : 2 \to 1$$

• Eg -



$$\pi(i) = \frac{1}{Z}\left(\frac{p}{1-p}\right)^i$$

- For $i, j \in \{i-1, i+1\}$, $P_{ij} = P_{ji} = 0$

- $i \to i+1 : \frac{1}{Z}\left(\frac{p}{1-p}\right)^i \cdot p = \frac{1}{Z}\left(\frac{p}{1-p}\right)^{i-1} \cdot (1-p) : (i+1 \to i)$

Markov-chain monte carlo

- Given a target distribution π over state-space $\mathcal{X}$
- Given a 'base' Markov chain $\tilde{P}$ with stationary dist $\tilde{\pi}$
  (Usually, $\tilde{\pi} \equiv \text{Unif}(\mathcal{X})$)
- MCMC
  - start at X_0
  - At time t, generate $Y_t \equiv \underbrace{\tilde{P}}_{\text{Proposal distribution (transition kernel)}}(\cdot \mid X_{t-1})$
  - Set $X_t = \begin{cases} Y_t & \text{with prob } \underbrace{A(X_{t-1}, Y_t)}_{\text{Acceptance kernel}} \\ X_{t-1} & \text{with prob } 1 - A(X_{t-1}, Y_t) \end{cases}$

Claim - For any π, given $\tilde{P}$, can design $A(\cdot,\cdot)$ s.t. this MC is reversible and has steady-state dist π

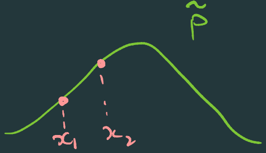
the Metropolis algorithm

- target distribution $P(x) = \widetilde{P}(x)/Z$ $\Rightarrow Q$ is reversible, with $\pi = \frac{1}{|X|}$
- proposal distribution(s) $Q(x|y)$, with $Q(x|y) = Q(y|x) \, \forall \, x, y$

Metropolis sampling

1. choose initial Z_0
2. to obtain sample t, generate $Y_t \sim Q(\cdot|Z_{t-1})$
3. accept $Z_t = Y_t$ with probability $A(Y_t, Z_{t-1}) = \underbrace{\max\left\{1, \frac{\widetilde{P}(Y_t)}{\widetilde{P}(Z_{t-1})}\right\}}_{\in[0,1]}$
   else reject and set $Z_t = Z_{t-1}$

- Always go from lower $\tilde{P}(x)$ to higher $\tilde{P}(y)$
- Sometimes go from higher to lower

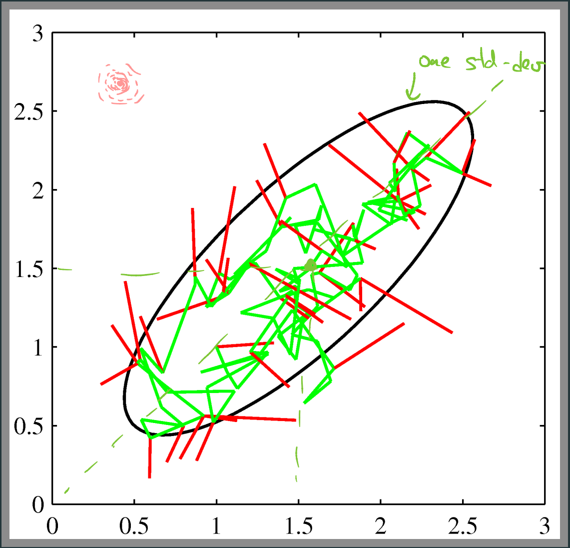


$\mathbb{P}[x_1 \to x_2] = 1$, $\mathbb{P}[x_2 \to x_1] < 1$

Metropolis for 2-d Gaussian

$Q(x|y) = N(0, I)$

$Q(x|y) = Q(y|x)$



Metropolis algorithm: proof of correctness

- Claim - $P = \frac{\tilde{P}}{Z}$ is the stationary dist of the Metropolis chain
- Pf - Reversibility! Suppose claim is true

For every pair $x, y \in \mathcal{X}$

$$(x \to y) \quad P(x) \cdot Q(y|x)\, A(y,x) = \frac{\tilde{P}(x)}{Z} Q(y|x) \min\left(1, \frac{\tilde{P}(y)}{\tilde{P}(x)}\right)$$

equal by assumption

$$(x \leftarrow y) \quad P(y) \cdot Q(x|y)\, A(x,y) = \frac{\tilde{P}(y)}{Z} Q(x|y) \min\left(1, \frac{\tilde{P}(x)}{\tilde{P}(y)}\right)$$

If $\tilde{P}(y) > \tilde{P}(x)$, then $\min\left(1, \frac{\tilde{P}(y)}{\tilde{P}(x)}\right) = 1$, $\min\left(1, \frac{\tilde{P}(x)}{\tilde{P}(y)}\right) = \frac{\tilde{P}(x)}{\tilde{P}(y)}$

(and similarly if $\tilde{P}(x) > \tilde{P}(y)$)

Metropolis-Hastings

- target distribution $P(x) = \widetilde{P}(x)/Z$
- proposal distribution(s) $Q(x|y)$ ← not necessarily symmetric

'inverse propensity score'

Metropolis-Hastings sampling

1. choose initial Z_0
2. to obtain sample t, generate $Y_t \sim Q(\cdot|Z_{t-1}) = \widetilde{Q}_{Z_0}$
3. accept $Z_t = Y_y$ with prob $A(Y_t, Z_{t-1}) = \max\left\{1, \frac{\widetilde{P}(Y_t)Q(Z_{t-1}|Y_t)}{\widetilde{P}(Z_{t-1})Q(Y_t|Z_{t-1})}\right\}$

   else reject and set $Z_t = Z_{t-1}$

$\frac{\widetilde{P}(Y_t)/Q(Y_t|Z_{t-1})}{\widetilde{P}(Z_{t-1})/Q(Z_{t-1}|Y_t)}$

Pf - For $x \to y$: $\frac{\widetilde{P}(x)}{Z} \cdot Q(y|x) \left(\frac{\widetilde{P}(y)}{\widetilde{P}(x)} \cdot \frac{Q(x|y)}{Q(y|x)}\right)$

$=$

$y \to x$: $\frac{\widetilde{P}(y)}{Z} \; Q(x|y) \; (1)$

(Assuming $\widetilde{P}(y)Q(x|y) < \widetilde{P}(x)Q(y|x)$)

Gibbs sampling (for vector valued distns)

- target distribution $P(x(1), x(2), \dots, x(n))$

Gibbs sampling

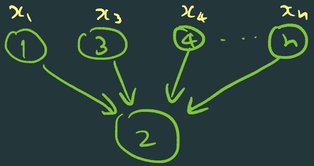
1. choose initial $X_0 = (X_0(1), X_0(2), \dots, X_0(n))$
2. to obtain sample t:

   pick I_t uniformly at random ← (can also do round robin)

   set $X_t(i) = X_{t-1}(i)$ for $i \neq I_t$

   set $X_t(I_t) \sim P(\cdot | X_{t-1} \setminus X_{t-1}(I_t))$ ← the dist $P(x_{I_t} | x_j, j \neq I_t)$

If $I_t = 2$, $X_1 = x_1, X_3 = x_3 \dots$



Gibbs sampling for 2-d Gaussian