Aula 15

Aula passada

- Autovalores, autovetores, decomposição
- Convergência para estacionaridade
- Tempo de mistura
- Spectral gap
- Tempo de mistura de passeios aleatórios

Aula de hoje

- Caminho amostral
- Teorema Ergódico
- Estimando π
- Simulação de CM
- Gerando amostras
- Markov Chain Monte Carlo (caso simétrico)
- Exemplo

Aleatoriedade de CM

- Até agora vimos os seguintes objetos matemáticos
- P: matriz de transição de estados da CM
- $\pi(0)$: distribuição inicial da CM
- $\pi(t)$: distribuição da CM no tempo t, dada por $\pi(t) = \pi(0)P^t$
- π: distribuição estacionária da CM, dada por $\pi = \pi P$
- τ_ε: tempo de mistura da CM, dado ε

Qual destes objetos é aleatório?

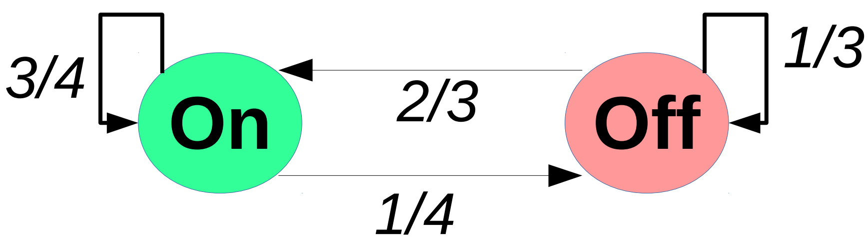
- **Nenhum deles!**
- Então o que é aleatório na CM?

Cadeia de Markov

- Seja P a matriz de transição de estados de uma CM
- Seja X_t uma v.a. que determina o estado da cadeia no instante de tempo t, para t = 0,1,2,…
  - P[$X_t = s$] para todo s em S

X_t é aleatório (é uma v.a.)

- Exemplo: $\pi(0) = (0.8 \ \ 0.2)$



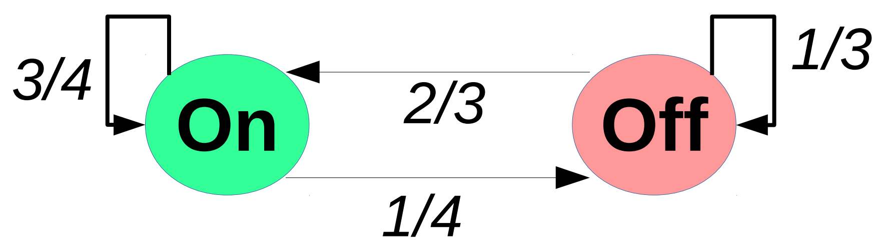
- Realização: X_0= 1, X_1= 0, X_2= 0, X_3= 1, X_4= 0, X_5= 1, X_6= 1, …
- Realização: X_0= 0, X_1= 0, X_2= 1, X_3= 1, X_4= 0, X_5= 0, X_6= 1, …
- Realização: X_0= 1, X_1= 1, X_2= 0, X_3= 1, X_4= 1, X_5= 1, X_6= 0, ...

Caminho Amostral

- Uma realização da sequência de v.a. X_t para t = 0,1,…
- Probabilidade de um caminho amostral $\omega = (\omega_0, \omega_1, \omega_2, \ldots)$
  - prob. da CM realizar exatamente ω

$$P[\omega] = P[X_0 = \omega_{0,} X_1 = \omega_{1,} \ldots] =$$
$$= \pi_{\omega_0}(0) P_{\omega_{0,}\omega_1} P_{\omega_{1,}\omega_2} P_{\omega_{2,}\omega_3} \ldots$$

- Todo caminho amostral ω tem uma probabilidade
  - que vai a zero com o comprimento do caminho!
- Exemplo: $\pi(0)$ = (0.8 0.2)



  - ω = 1,1,1,1,1,1,1,1
  - ω = 0,1,0,1,0,1,0,1

$$P[\omega] = ?$$

O que ocorre com X_t?

- Se todo caminho amostral tem probabilidade que vai a zero, o que podemos dizer sobre sequência X_t ?

Usar a média sobre valores da sequência!

- Ex. média amostral dos valores de estado observados

$$S_k = \frac{1}{k} \sum_{t=0}^{k-1} X_t$$

- Ex. fração de vezes que um estado *s* é visitado

$$f_k(s) = \frac{1}{k} \sum_{t=0}^{k-1} I(X_t = s)$$

Convergência Intuitiva

- Para onde converge a média amostral de X_t, com um grande número de amostras (*k* muito grande)?
  - caminho amostral muito longo pela CM

$$S_k = \frac{1}{k} \sum_{t=0}^{k-1} X_t \longrightarrow E_\pi[X] = \sum_s s\,\pi_s$$

- Para onde converge a fração de visitas a um estado, com um grande número de amostras (*k* muito grande)?

$$f_k(s) = \frac{1}{k} \sum_{t=0}^{k-1} I(X_t = s) \longrightarrow \pi_s$$

Teorema Ergódico

- Seja *f* uma função sobre o espaço de estados da CM
  - mapeia cada estado da CM em um valor real
- Se CM é irredutível e aperiódica, com distribuição estacionária π, temos

$$P\left[\lim_{k\to\infty}\frac{1}{k}\sum_{t=0}^{k-1}f(X_t)=E_\pi[f(X)]\right]=1$$

Sobre distribuição estacionária

Como lei forte dos grandes números!

- Teorema fundamental: Média no espaço = média no tempo
  - valor esperado de qualquer função pode ser aproximado usando caminho amostral
- Exemplos anteriores são casos especiais

Estimando π

- Teorema ergódico garante que método de Monte Carlo funciona também em CM
  - conexão entre teoria (equações) e prática (simulação)
- Exemplo: Como estimar π ?
- Usar a CM para gerar um caminho amostral ω bem longo e calcular a fração de visitas a cada estado

$$\hat{\pi}_s(k) = \frac{1}{k} \sum_{t=0}^{k-1} I(\omega_t = s)$$

- Teorema ergódico garante convergência
  - estimador consistente (possui viés para tempo *k*)
- Outro método para encontrar π

Simulando uma CM

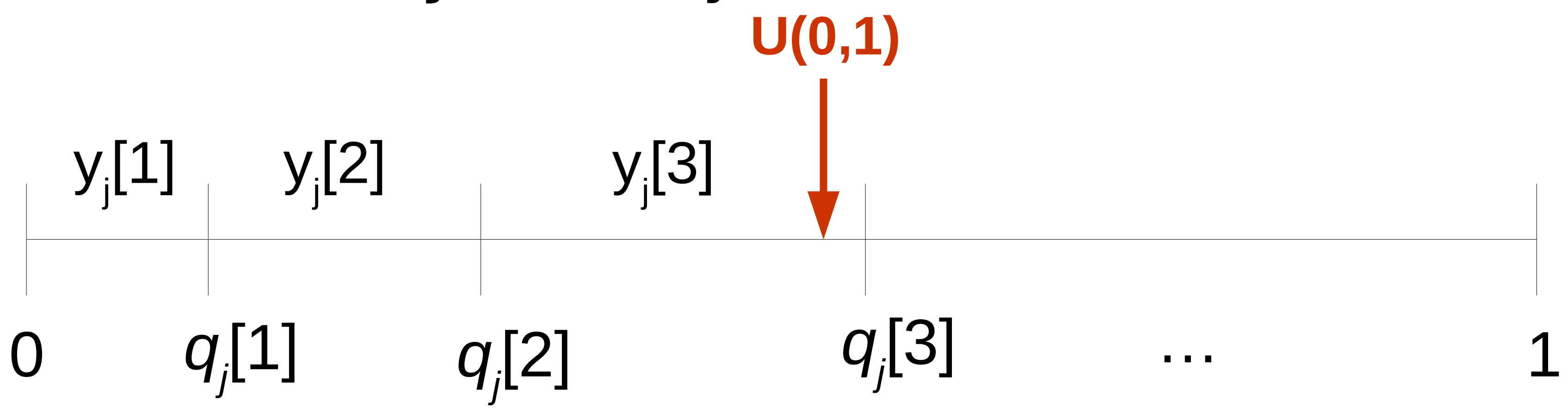
- Como simular uma CM?
  - entrada: matriz P, distribuição inicial $\pi(0)$
- Simular cada valor X_t, para t=0,1, …
- Usar $\pi(0)$ para gerar X_0
- Usar P (matriz de transição) para gerar X_1, X_2, …
- Como gerar X_k dado X_{k-1} ?

Método da transformada inversa!

- Gerar U ~ uniforme(0,1), usar (X_{k-1})-ésima linha da matriz P para escolher vizinho, valor de X_k
- Problema: ineficiente (se matriz P é esparsa)

Simulando CM Eficientemente

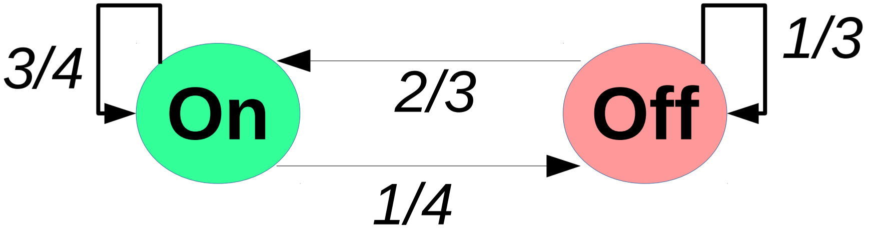
- Representar matriz *P* como vetor de adjacência
  - apenas entradas não-nulas são representadas
- $y_j[i]$ = estado da i-ésima transição não-nula do de j
- $q_j[i]$ = prob. de transição acumulada das primeiras i-ésimo transições de j



- Complexidade: O(# transições não-nulas)
- Método Alias: complexidade O(1) + custo inicial

Gerando Amostras

- Considere uma CM com matriz de transição P
- Como gerar amostras de X_t ?
- Exemplo: $\pi(0) = (0.8 \;\; 0.2)$
- Gerar amostras de X_{10} ?



- **Ideias**

  1) Calcular $\pi(t)$, gerar amostras desta distribuição

  2) Simular CM até X_t , para cada amostra

Qual é a melhor abordagem?

- Depende de n (tamanho da CM), e t

Espaços Grandes e Complicados

- Considere um espaço amostral grande e complicado
  - grande = muitos elementos
  - complicado = não é fácil enumerar os elementos
- Todos os grafos conexos com *n* vértices
  - não é fácil enumerar grafos conexos
- Todos os percursos por *n* cidades de comprimento *L* ou menor
  - não é fácil enumerar percursos de comp. *L* ou menor
- Em geral: espaço amostral combinatorial com restrições que dificultam enumeração

Como gerar amostras destes espaços?

Gerando amostras uniformes

- Como gerar amostras uniformes destes espaços?

Ideia preliminar

  - 1) Aumentar espaço de estado amostral (ex. removendo restrição) – facilitar geração
  - 2) Usar método da rejeição – rejeitar amostras que não atendem a restrição, remover viés

- Exemplo: Gerar grafos conexos com *n* vértices?
  - modelo G(*n*, *p*) com *p*=½ todos os grafos são equiprováveis (não temos viés)
  - gerar grafo de *G'* ~ G(*n*, ½)
  - retorna se *G'* é conexo, cc gera outro *G'*

Problema?

- Pode ser muito ineficiente!

Markov Chain to the Rescue

- Como usar CM para resolver este problema?
- Construir CM cujos estados correspondem aos elementos de *S* (espaço amostral)
- Construir uma matriz de transição *P* tal que distribuição estacionária π seja uniforme
- Simular CM para gerar amostras
  - dar τ_ε passos para gerar uma amostra

Markov Chain Monte Carlo (MCMC)

- Método baseado em CM para gerar amostras de uma distribuição qualquer
  - não precisa ser uniforme

MCMC

Como construir CM, se *S* é muito grande?

- Não precisamos construir a priori
- Podemos construir e simular CM ao mesmo tempo, um passo de cada vez

Qual CM mais apropriada?

- Baixo tempo de mistura τ_{ε} , parar gerar amostras
- Poucas transições de saída em cada estado (ex. log n)
- Fácil descrição dos estados vizinhos (a partir do estado atual)
- Tema de grande debate e avanço!

MCMC – Caso Simétrico

- Considere uma CM com matriz de transição *P* simétrica
  - $P_{ij} = P_{ji}$ para todo estado i, j
- Distribuição estacionária é uniforme e a CM é reversível
- Queremos modificar *P* para construir uma nova CM cuja distribuição estacionária seja π qualquer
  - π é entrada para o problema

- **Ideia:** não "aceitar" todas as transições da CM original
  - continuar no mesmo estado para induzir π qualquer
  - parecido com método da rejeição
- Rejeição é probabilística, para cada transição não-nula em *P*, temos $a(i, j)$

MCMC – Caso Simétrico

- Considere CM no estado *i* e uma proposta de transição da CM original (ou seja, em *P*) para o estado *j*
  - aceita transição com probabilidade a(*i*, *j*), rejeita com complemento
- Matriz de transição *P’* da nova CM

$$P'_{ij} = \begin{cases} P_{ij}\, a(i,j), & se\, i \neq j \\ 1 - \sum_{k:k\neq i} P_{ik}\, a(i,k), & se\, i = j \end{cases}$$

- Temos que escolher a(i, j) tq *P’* tenha distribuição estacionária dada por π, ou seja:

$$\pi_i P_{ij}\, a(i,j) = \pi_j P_{ji}\, a(j,i)$$

Garantindo também que *P’* é reversível

MCMC – Caso Simétrico

- Como $P_{ij} = P_{ji}$ (por simetria em *P*), temos

$$\pi_i a(i,j) = \pi_j a(j,i)$$

- Como a(i,j) e a(j,i) são probabilidades, temos

$$\pi_i a(i,j) \leq \pi_i \qquad \pi_i a(i,j) = \pi_j a(j,i) \leq \pi_j$$

- Como a(i,j) deve ser o maior possível para evitar desperdício, temos

$$a(i,j) = 1, \quad se\, \pi_i \leq \pi_j$$

$$a(i,j) = \frac{\pi_j}{\pi_i}, \quad se\, \pi_i > \pi_j$$

- Ou seja,

$$a(i,j) = min\left\{1, \frac{\pi_j}{\pi_i}\right\}$$

Exemplo

- Gerar amostras de pares ordenados (x, y) em [1,n] x [1,n]
  - grid 2D discreto e quadrado com *n* pontos

$$P[(x,y)] = \frac{(x+y)^2}{Z} \qquad Z = \sum_{(x,y)\in[1,n]^2} (x+y)^2$$

constante de normalização

- CM original é torus 2D com n x n vértices
  - estado dado por (x, y), com x,y = 1, …, n
  - todo estado tem 4 transições, $P_{ij} = 1/4$, para algum i,j
  - CM é simétrica e consequentemente uniforme
- Como transformar *P* em *P'* para induzir π conforme definição acima?

Exemplo

- Precisamos CM tal que $\pi_{(x,y)} = \frac{(x+y)^2}{Z}$
- Seja estado i = (x,y) e j = (x',y'). Desta forma, podemos definir a(i, j) como

$$a(i,j) = min\{1, \frac{\pi_j}{\pi_i}\} = min\{1, \frac{(x'+y')^2}{(x+y)^2}\}$$

- Por fim, definimos *P'* como anteriormente
  - $P'_{ij} = P_{ij}$ a(i,j) se i != j,   etc.
- **Boa notícia:** não precisamos calcular *Z* para definir a(i, j)
  - seria proibitivo calcular *Z*
- Para gerar uma amostra, simular CM definida por *P'* por τ_ε e retornar o estado atual