Aula 12

Roteiro

- Caminho amostral
- Teorema Ergódico
- Simulação de CM (eficiente)
- Gerando amostras
- Simulando CM enormes

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?

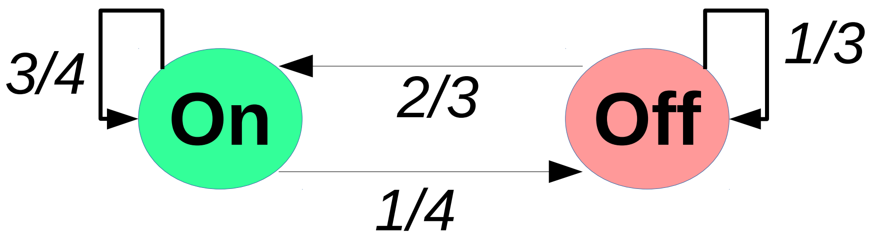
- **Nenhum deles!**
- Então o que é aleatório na CM?

Cadeia de Markov

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



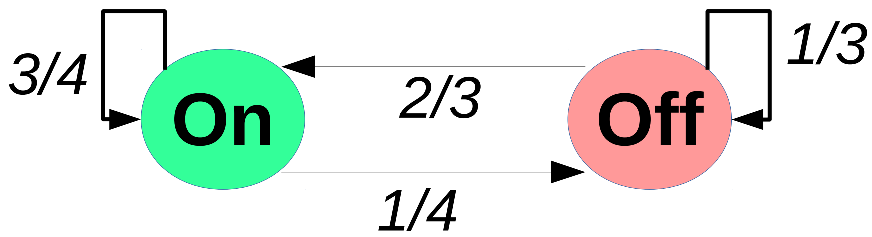
- Realização: X_0= 1, X_1= 2, X_2= 2, X_3= 1, X_4= 2, X_5= 1, X_6= 1, …
- Realização: X_0= 2, X_1= 2, X_2= 1, X_3= 1, X_4= 2, X_5= 2, X_6= 1, …
- Realização: X_0= 1, X_1= 1, X_2= 2, X_3= 1, X_4= 1, X_5= 1, X_6= 2, ...

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 = 1,1,1,1,1,1,1,1,1
  - ω = 2,1,2,1,2,1,2,1,2 $\quad P[\omega] = ?$



Figueiredo 2021

O que ocorre com X_t?

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

Usar 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 para um caminho amostral muito longo (*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 para um caminho amostral muito longo (*k* muito grande)?

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

Função indicadora:
vale 1 quando $X_t = s$

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

Figueiredo 2021

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 possui viés para tempo *k*, mas é consistente (viés vai a zero com *k*)
- Outro método para calcular π

Simulando uma CM

- Como simular uma CM?
  - entrada: matriz P, distribuição inicial $\pi(0)$

Simular é gerar um caminho amostral!

Ideia 0: usar diretamente os caminhos amostrais

1) enumerar todos os caminhos amostrais de tamanho k

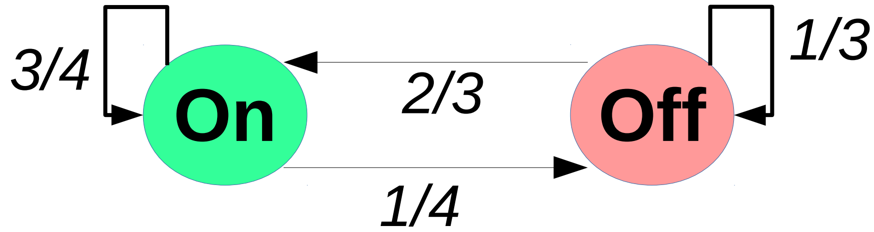
2) determinar probabilidade de cada caminho

3) gerar amostras deste conjunto

- **Intuição:** cada caminho amostral é uma face de um dado enviesado
  - sabemos gerar amostras de dado enviesado!

Exemplo com Modelo On-Off

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



- Todas as sequências binárias com k dígitos
- k = 5: $\omega^0 = 0,0,0,0,0$ $\omega^1 = 0,0,0,0,1$, …, $\omega^{31} = 1,1,1,1,1$
- $P[\omega^0] = 0.2*(1/3)^4$, $P[\omega^1] = 0.2*(1/3)^3*2/3$, ...
- Dado enviesado com 32 faces
- Gerar amostras do dado
  - tempo constante, usando Método Alias

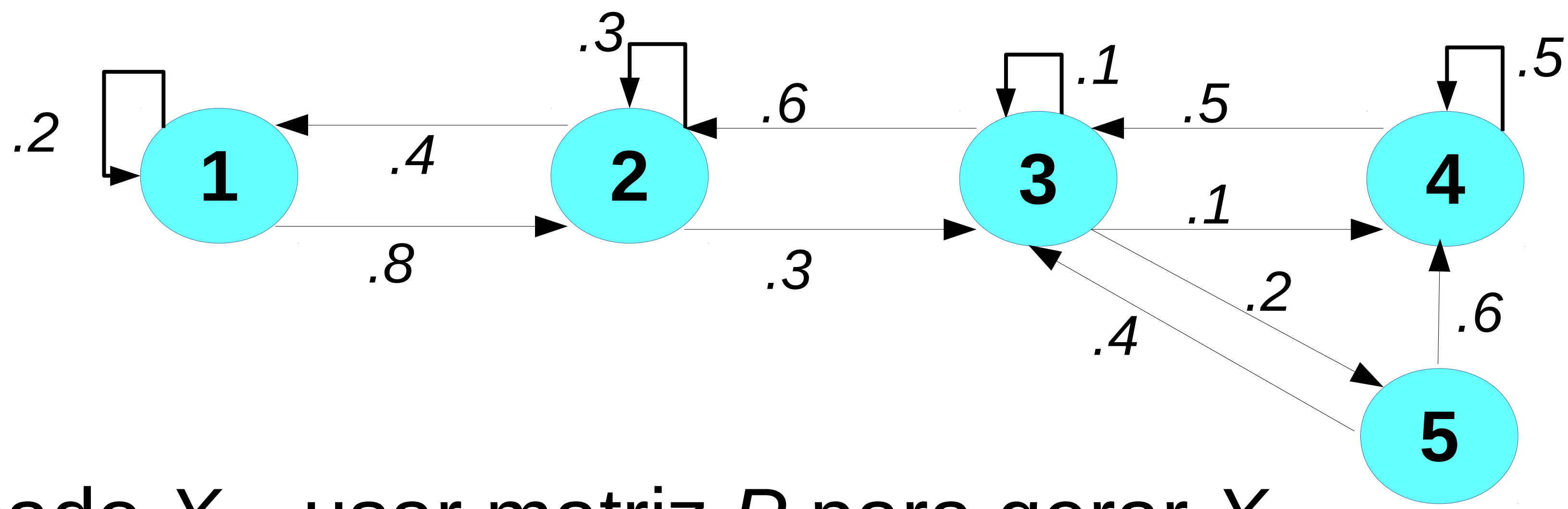
Problema?

- Número de caminhos é exponencial em *k*
- Tempo exponencial para construir o dado!

Simulando uma CM

- Outra abordagem: amostrar a sequência X_t de forma iterativa, para t=0,1,2, …
  - construir dinamicamente o caminho amostral
- Usar $\pi(0)$ para gerar X_0
- Dado X_0, usar P (matriz de transição) para gerar X_1
  - $\mathrm{P}[X_1 = s_1 \mid X_0 = s_0] = P(s_0,s_1)$
- Dado X_1, usar P (matriz de transição) para gerar X_2
  - $\mathrm{P}[X_2 = s_2 \mid X_1 = s_1] = P(s_1,s_2)$
- De forma geral, dado X_k, usar P para gerar X_{k+1}

Exemplo



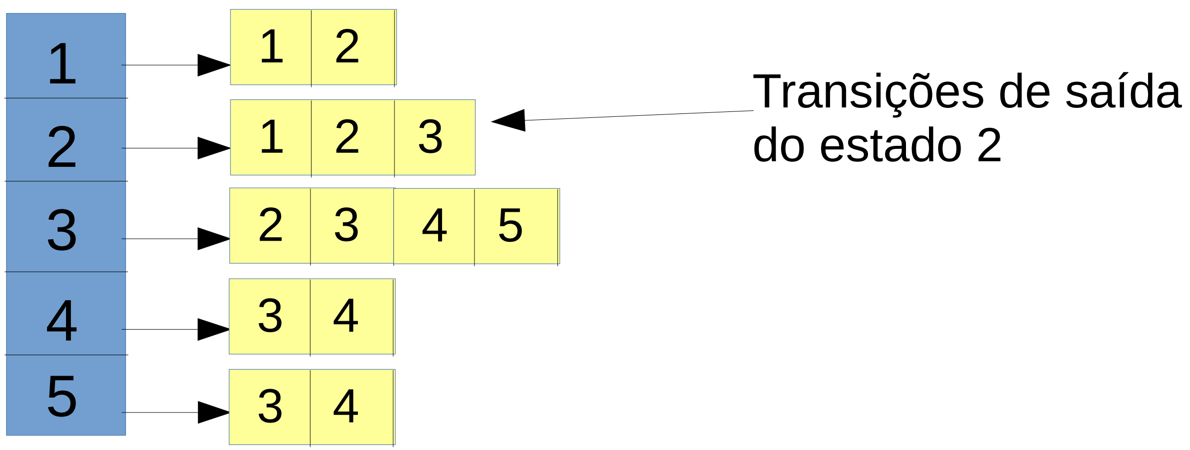
- Dado X_k, usar matriz P para gerar X_{k+1}
- Ex. dado $X_4 = 3$, como gerar X_5?
- X_5 é um dos vizinhos de saída do estado 3
  - 4 estados possíveis {2, 3, 4, 5}, cada um possui uma probabilidade, dado por $P_{3,2}$, $P_{3,3}$, $P_{3,4}$, $P_{3,5}$
  - gerar amostra para X_5 de acordo com essas probabilidades

Simulando uma CM

- Para gerar X_{k+1} , usamos a linha X_k da matriz P
- Temos um dado enviesado para cada linha de P
  - cada face corresponde a um estado da CM
- Precisamos escolher uma face
- **Ideia 1**: usar diretamente a linha da matriz P
- Linha i da matrix P
  - $P(i,*) = p_{i,1}, p_{i,2}, p_{i,3}, \ldots, p_{i,n}$
- Algoritmo básico para gerar amostra de dado enviesado
  - tempo médio: $n/2$
- **Problema:** muito ineficiente se matriz P é esparsa

Simulando CM Eficientemente

- Representar matriz *P* como vetor de adjacência
  - apenas entradas não-nulas são representadas
- Exemplo da CM anterior

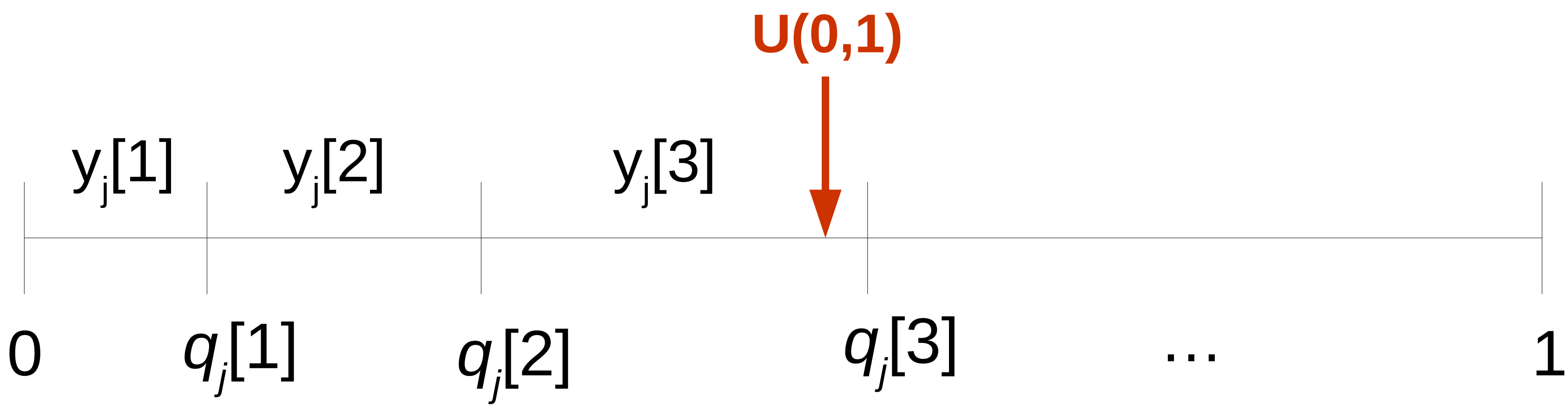


- Dois vetores para cada estado j
- $y_j[i]$ = estado destino da i-ésima transição não-nula de j
- $q_j[i]$ = probabilidade de transição acumulada pelas primeiras i-ésimas transições de saída de j
- Ex: $y_2[2] = 2$, $q_2[2] = .7$, $y_5[2] = 4$, $q_5[2] = 1$

Figueiredo 2021

Simulando CM Eficientemente

- Usar vetores *y* e *q* para simular o dado enviesado
- Sendo *j* o estado atual, temos:



- No exemplo, $y_j[3]$ é o próximo estado!
- Complexidade: O(*d*), onde *d* é grau de saída do estado
- Método Alias: complexidade O(1) + custo inicial
  - vantagem depende do retorno ao mesmo estado e tamanho do caminho amostral

Gerando Amostras

- Como gerar amostras de X_t?
  - caminho amostral não interessa, apenas amostra no tempo t
- **Ideia 1:**

  1) Calcular $\pi(t)$ (distribuição no tempo t)

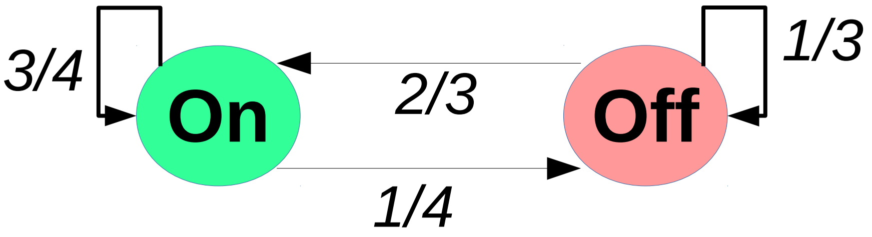
  2) Gerar amostras desta distribuição

  - lembrando: $\pi(t) = \pi(0)P^t$
- **Ideia 2:**

  1) Simular CM até X_t, gerando caminho amostral

  2) Retornar a amostra gerada para X_t

Exemplo



- Exemplo: $\pi(0) = (0.8 \;\; 0.2)$
- Gerar amostras de X_{10}?

- **Ideia 1:** calcular $\pi(10) = (0.727 \;\; 0.273)$
- Gerar amostra para X_{10} usando $\pi(10)$

- **Ideia 2:** gerar caminho amostral $X_0, X_1, X_2, \ldots, X_{10}$
  - ex. ω = 1,2,1,1,2,2,1,2,1,2,2
- Retornar X_{10} = 2 (Off)

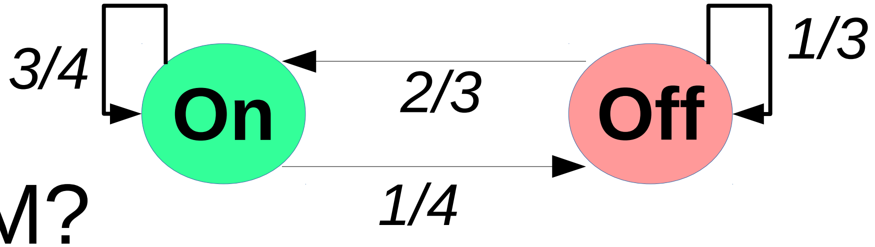
Gerando Amostras

Qual abordagem é mais eficiente?

- Depende de *n* (tamanho da CM), de *t*, do número de amostras de X_t (supor *r* amostras)
- Calcular $\pi(t)$ tem complexidade $O(t\, n^2)$
  - *t* multiplicações de vetor por matriz
  - gerar uma amostra tem complexidade $O(n)$
  - mas se *r* for grande podemos usar Método Alias em π(t)
- Gerar X_t tem complexidade $O(t\, d)$, para cada amostra
  - se r for grande podemos usar Método Alias nas transições da CM, ao custo amortizado de $O(t)$ por amostra

Amostras Estacionárias

- Como gerar amostras da distribuição estacionária da CM?



Mesma abordagem!

- **Ideia 1:**

1) Calcular π (distribuição estacionária)

2) Gerar amostras desta distribuição

- **Ideia 2:**

1) Simular CM até X_t para t suficiente grande

  - tempo de mistura τ_ε

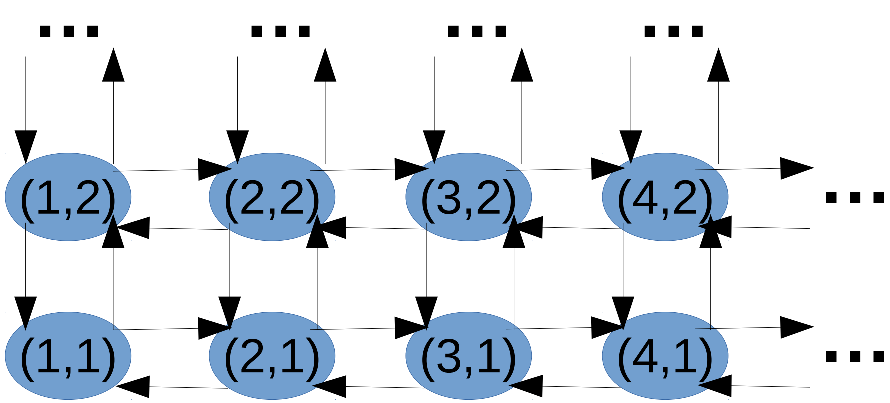
2) Retornar a amostra gerada para X_t

Simulando CM Enormes

- Como simular CM com muitos estados?
  - até mesmo com espaço de estado infinito
- **Problema:** impossível representar CM na memória (mesmo com representação esparsa)
- **Ideia:** gerar apenas possíveis próximos estados a partir do estado atual
  - determinar as transições de saída do estado atual
  - construir dado enviesado
  - escolher próximo estado de acordo
- **Restrição:** gerar transições de saída a partir do estado atual (sem conhecer a CM)
  - regras que definem possíveis transições

Exemplo 1

- Látice de duas dimensões nos naturais
  - probabilidades de transição para norte, sul, leste, oeste: $p_n + p_s + p_l + p_o = 1$
  - self-loops nos estados de borda



- Representar estado como (i, j)
- Dado estado atual, sabemos os possíveis próximos estados e respectivas probabilidades de transição

- Dado estado (i, j), escolher:
  - (i+1,j) com p_o, (i-1,j) com p_l, (i,j+1) com p_n, (i,j-1) com p_s
  - atenção com estados de borda (1, j) ou (i, 1)
- Muito eficiente: O(1) de memória para simular CM infinita!

Exemplo 2

- Passeio aleatório em Hipercubo de 100 dimensões
  - CM tem 2^{100} estados, impossível representar em qualquer computador
- Estado da CM é um vetor binário com 100 bits
- Estados vizinhos são os vetores que diferem do atual em exatamente 1 bit
  - definição de hipercubo
- Passeio aleatório escolhe vizinho uniformemente
  - escolher um bit uniformemente, inverter o bit (não precisa nem enumerar os vizinhos)
  - O(1) para fazer uma transição