Laboratório Nacional de Computação Científica
UNIDADE DE PESQUISA DO MCTI

Introdução a Cadeias de Markov e MCMC

Modelos Compartimentais em Epidemiologia e Inferência Bayesiana

Gustavo Libotte e Regina Almeida

21 de janeiro de 2022

Noções básicas sobre cadeias de Markov

- Um **processo estocástico** é qualquer coleção de variáveis aleatórias definidas sobre o mesmo espaço de probabilidade, $\{U(t), t \in T\}$, onde T é um subconjunto de $\mathbb{R}$ que, por comodidade, é entendido como uma classe de instantes de tempo.
- Quando esta classe é o conjunto discreto de inteiros positivos $T = \{0, 1, 2, \ldots\}$, o processo estocástico dito a **tempo discreto** é usualmente denotado por $\{U_n, n \geq 0\}$, sendo esta a situação típica no contexto de um esquema de simulação estocástica.
- O conjunto $\mathcal{U}$ de valores das variáveis é denominado espaço de estados.
- Num processo o conhecimento de estados do passado e do presente influencia geralmente a plausibilidade de ocorrência dos estados futuros.
- Quando fixado o estado do presente os estados do passado deixam de ter influência no futuro, diz-se que o processo goza da propriedade de dependência de Markov.

Noções básicas sobre cadeias de Markov

- O processo $\{U_n, n \geq 0\}$ satisfazendo tal propriedade de independência condicional é denominado **cadeia de Markov**, podendo ser definido através de

$$U_{n+1} \perp\!\!\!\perp (U_0, \ldots, U_{n-1}) \mid U_n \Leftrightarrow$$
$$P\left(U_{n+1} \in A \mid U_0 = u_0, \ldots, U_n = u\right) = P\left(U_{n+1} \in A \mid U_n = u\right) \equiv P_n(u, A),$$

  para todo evento A e $n \geq 0$, onde o símbolo $P_n(u, A)$ denota a chamada função de transição (em um passo) quando parte do instante n.

- Equivalentemente, tomando $A =]-\infty, v]$ a cadeia de Markov pode ser definida através das funções de distribuição condicionais por

$$F_{U_{n+1}}\left(v \mid U_0 = u_0, \ldots, U_n = u\right) = F_{U_{n+1}}\left(v \mid U_n = u\right) \equiv F_n(u, v),$$

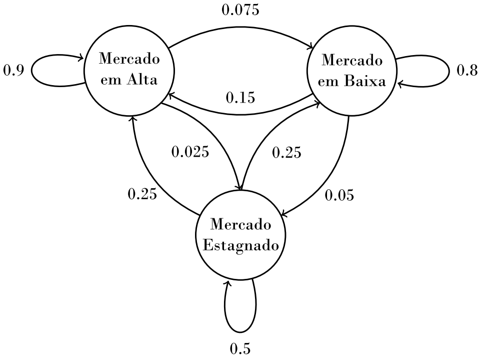
  para todo o $v, u \in \mathcal{U}$.

- Quando a função de transição é invariável com n, sendo então denotada por $P(u, A)$ (ou $F(u, v)$), a cadeia de Markov diz-se **homogênea**.

Noções básicas sobre cadeias de Markov

- Quando o espaço de estados é um conjunto discreto, a cadeia de Markov (implicitamente homogênea) fica definida através das funções de probabilidade condicionais $P(u, \{v\})$,

$$P\left(U_{n+1} = v \mid U_0 = u_0, \ldots, U_n = u\right) = P\left(U_{n+1} = v \mid U_n = u\right) \equiv p(u, v), \forall n \geq 0, u, v \in \mathcal{U}.$$



- A **função de transição** $p(\cdot, \cdot)$ é apresentável na forma de uma matriz P de probabilidades no caso de um número finito de estados.

- **Exemplo**: dados os estados {1 = alta, 2 = baixa, 3 = estagnado}, a matriz de transição para este exemplo é

$$P = \begin{bmatrix} 0,9 & 0,075 & 0,025 \\ 0,15 & 0,8 & 0,05 \\ 0,25 & 0,25 & 0,5 \end{bmatrix}$$

Noções básicas sobre cadeias de Markov

- Restringindo-nos a uma cadeia de Markov com espaço de estados discreto, tem-se
$$P\left(U_{n+1}=v\right)=\sum_u P\left(U_n=u\right)p(u,v)=\sum_u P\left(U_0=u\right)p^n(u,v)$$
onde $p^n(u,v)=P\left(U_n=v \mid U_0=u\right)=\sum_u p^{n-1}(u,z)p(z,v), n\geq 1$ traduz a função de transição em n passos.
- A construção de uma cadeia de Markov é assim completamente determinada pela sua função de transição, **desde que se conheça a distribuição inicial**.
- O estudo do **comportamento assintótico** $(n \to \infty)$ das cadeias é fundamental para os métodos MCMC e nele desempenha um papel crucial o seguinte conceito.
- Diz-se que uma distribuição de probabilidade $\pi(u), u\in\mathcal{U}$ é **estacionária** se
$$\pi(v)=\sum_u \pi(u)p(u,v)$$
- Em particular, a distribuição inicial $P\left(U_0=u\right)=\pi(u)$ é estacionária, se e somente se, a distribuição de U_n é invariante com n, i.e. $P\left(U_n=u\right)=\pi(u), \forall n\geq 0$.

Noções básicas sobre cadeias de Markov

- A ideia de estacionariedade traduz a **invariância distribucional** da cadeia no sentido de que, uma vez atingido um estado caraterizado pela distribuição marginal π, os estados seguintes da cadeia **não envolvem qualquer alteração** a π.
- Em outras palavras, é como se a cadeia atingisse uma **situação de equilíbrio** em termos de estabilidade distribucional.
- **Exemplo**: A distribuição por estados pode ser escrito como um vetor de linha estocástico x, com $x^{(n+1)} = x^{(n)}P$. Assim, se no tempo n o sistema está no estado $x^{(n)}$, e em seguida, três períodos de tempo mais tarde, no tempo $n+3$ a distribuição é

$$\begin{aligned} x^{(n+3)} &= x^{(n+2)}P = \left(x^{(n+1)}P\right)P \\ &= x^{(n+1)}P^2 = \left(x^{(n)}P\right)P^2 \\ &= x^{(n)}P^3 \end{aligned}$$

Noções básicas sobre cadeias de Markov

- Em particular, se num momento n o sistema está no estado 2 (baixa), então no tempo $n+3$, a distribuição é

$$
\begin{aligned}
x^{(n+3)} &= \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0,9 & 0,075 & 0,025 \\ 0,15 & 0,8 & 0,05 \\ 0,25 & 0,25 & 0,5 \end{bmatrix}^3 \\
&= \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0,7745 & 0,17875 & 0,04675 \\ 0,3575 & 0,56825 & 0,07425 \\ 0,4675 & 0,37125 & 0,16125 \end{bmatrix} \\
&= \begin{bmatrix} 0,3575 & 0,56825 & 0,07425 \end{bmatrix},
\end{aligned}
$$

- Utilizando a matriz de transição, é possível calcular, por exemplo, a fração de **longo prazo** de semanas durante o qual o mercado é estagnado, ou o número médio de semanas que será necessário para passar de uma estagnada a um mercado de touro.

$$
\lim_{N \to \infty} P^N = \begin{bmatrix} 0,625 & 0,3125 & 0,0625 \\ 0,625 & 0,3125 & 0,0625 \\ 0,625 & 0,3125 & 0,0625 \end{bmatrix}
$$

Noções básicas sobre cadeias de Markov

- A **existência** e **unicidade** de distribuições estacionárias depende da cadeia possuir algumas propriedades de estabilidade conhecidas como irredutibilidade e recorrência.
- Uma cadeia é **irredutível** quando pode atingir qualquer estado quando parte dele ou de outro estado qualquer em um número finito de transições.
- Diz-se **recorrente** se volta infinitas vezes ao estado de onde parte com probabilidade 1, qualquer que ele seja.
- Como caso particular, diz-se **recorrente positiva** se é finito o valor médio do tempo de retorno a $u, T_u = \min\{n \geq 1 : U_n = u\}$ quando parte dele, $\forall u$.
- Note-se que a irredutibilidade implica recorrência positiva se $\mathcal{U}$ é finito.
- Toda a cadeia de Markov (com $\mathcal{U}$ discreto) irredutível e recorrente positiva apresenta uma **distribuição estacionária única**.
- Por outro lado, existindo uma distribuição estacionária $\pi(v)$ tal que $\lim_{n\to\infty} p^n(u,v) = \pi(v)$, então a distribuição estacionária é **única** satisfazendo $\lim_{n\to\infty} P(U_n = v) = \pi(v)$.
- Sendo assim, independentemente da distribuição inicial, para valores suficientemente grandes de n a distribuição de U_n é aproximadamente dada por π.

Noções básicas sobre cadeias de Markov

- A convergência para a distribuição estacionária π **não é garantida** por uma cadeia irredutível e recorrente positiva.
- Todavia, se for imposto a essa cadeia a condição adicional de ser **aperiódica**, traduzida por

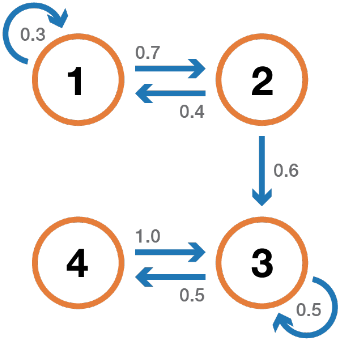
$$\min\{n \geq 1 : p^n(u,u) > 0\} = 1$$

basta exigir que $\exists u, p(u,u) > 0$, tal cadeia denominada então **ergódica** apresenta o comportamento limite

$$p^n(u,v) \underset{n\to\infty}{\longrightarrow} \pi(v), \forall u, v \in \mathcal{U}$$

assegurando assim a convergência de $P(U_n = u)$ para $\pi(u), \forall u$.

Noções básicas sobre cadeias de Markov
Propriedade de irredutibilidade



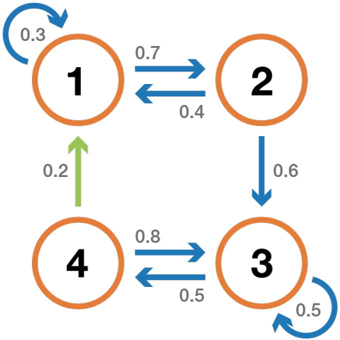


Ilustração da propriedade de **irredutibilidade**. A cadeia da esquerda **não** é irredutível: de 3 ou 4 não podemos chegar a 1 ou 2. A cadeia da direita (uma aresta foi adicionada) **é** irredutível: cada estado pode ser alcançado a partir de qualquer outro estado.

Noções básicas sobre cadeias de Markov
Propriedade de periodicidade

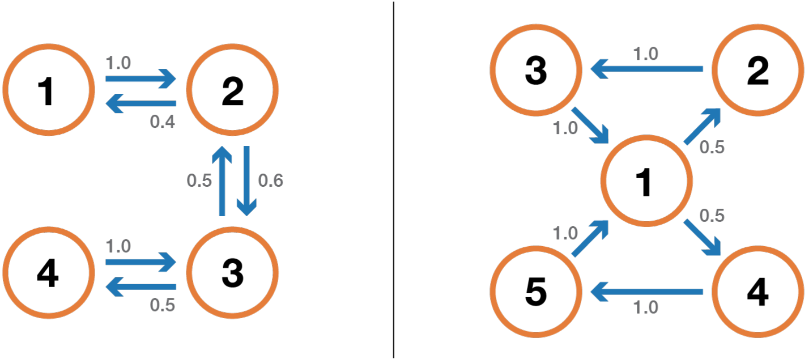


Ilustração da propriedade de **periodicidade**. A cadeia à esquerda é **2-periódica**: ao sair de qualquer estado, sempre leva um múltiplo de 2 passos para voltar a ele. A cadeia à direita é **3-periódica**.

Noções básicas sobre cadeias de Markov
Propriedade de recorrência

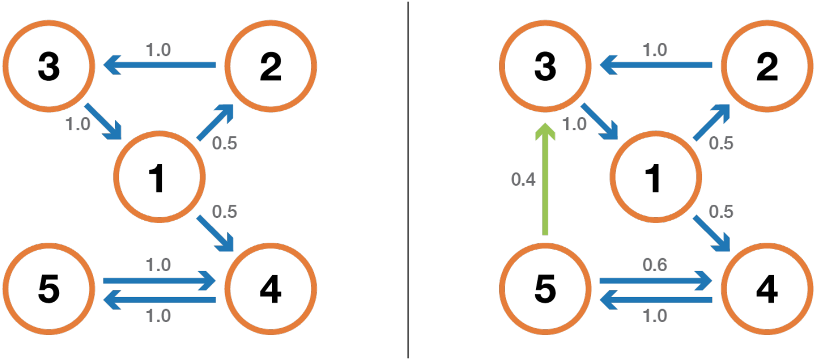


Ilustração da propriedade de **recorrência/transitoriedade**. A cadeia da esquerda é tal que: 1, 2 e 3 são **transitórios** (ao sair desses pontos não podemos ter certeza absoluta de que voltaremos a eles) e 3-periódicos enquanto 4 e 5 são **recorrentes** (ao sair destes pontos estamos absolutamente certos de que voltaremos a eles em algum momento) e 2-periódicos. A corrente à direita tem mais uma aresta que torna a corrente completa recorrente e aperiódica.

O amostrador de Gibbs

- A construção da cadeia de Markov que se pretende que convirja para a distribuição $\pi(u), u = (u_1, u_2, \ldots, u_k) \in \mathcal{U}$ é no algoritmo de Gibbs feita sequencialmente por amostragem das distribuições condicionais (tipicamente univariadas) dado todos os outros componentes.
- O algoritmo vai sucessivamente substituindo num ciclo de k passos os elementos do vetor corrente u de modo que no passo j **apenas** u_j é trocado pelo valor amostrado da densidade condicional $\pi\left(v_j \mid \{u_i, i \notin j\}\right), j = 1, 2, \ldots, k$.
- Por exemplo, tomando $k = 3$, o ciclo de 3 passos visando substituir o valor corrente $u = (u_1, u_2, u_3)$ é representável pelo esquema

$$(u_1, u_2, u_3) \xrightarrow{it.1} (v_1, u_2, u_3) \xrightarrow{it.2} (u_1, v_2, u_3) \xrightarrow{it.3} (u_1, u_2, v_3)$$

- Como cada etapa percorre os parâmetros e os atualiza de acordo com os valores atuais de todos os outros parâmetros, as amostras **não são independentes**.
- Toda a amostragem é realizada condicionalmente ao valor da iteração anterior e, portanto, as amostras formam uma **cadeia de Markov**.

O amostrador de Gibbs
Descrição formal

- A amostragem de Gibbs é projetada para **distribuições *a posteriori* multivariadas**.
- Seja $U = (U_1, U_2, \ldots, U_k)$ o vetor aleatório com densidade $\pi(u)$ e denote-se por U_{-j} o vetor U **sem** o j-ésimo componente, $U_{-j} = \left(U_1, \ldots, U_{j-1}, U_{j+1}, \ldots, U_k\right), j = 1, 2, \ldots, k$.

  1. Dado $u^{(t)} = \left(u_1^{(t)}, \ldots, u_k^{(t)}\right)$, iniciado em $t = 0$, amostre cada um dos componentes, $v_j^{(t)}$, do próximo vetor da cadeia a partir da distribuição $V_j^{(t)} \sim \pi\left(v_j^{(t)} \mid u_{-j}^{(t)}\right)$, para $j = 1, 2, \ldots, k$.
  2. Finalizado o ciclo de k iterações, tome $u^{(t+1)} = \left(v_1^{(t)}, \ldots, v_k^{(t)}\right)$ e repita-se o ciclo 1, substituindo t por $t+1$.

- Tudo se passa como se no j-ésimo passo Gibbs do ciclo assente em $u^{(t)}$, para todo o j, se obtivesse um vetor $v^{(t)} = \left(u_1^{(t)}, \ldots, u_{j-1}^{(t)}, v_j^{(t)}, u_{j+1}^{(t)}, \ldots, u_k^{(t)}\right)$ tal que

$$V^{(t)} \mid u^{(t)} \sim q_j\left(v^{(t)} \mid u^{(t)}\right) = \begin{cases} \pi\left(v_j^{(t)} \mid u_{-j}^{(t)}\right), & \text{se } v_{-j}^{(t)} = u_{-j}^{(t)} \\ 0, & \text{caso contrário} \end{cases}$$

O amostrador de Gibbs
Exemplo

- Para este exemplo, considere uma distribuição normal bivariada.
- A distribuição normal bivariada para X e Y, e a distribuição normal multivariada para $X_1, \ldots, X_m$, são **generalizações** da distribuição normal para dimensões superiores.
- A distribuição normal **bivariada** é especificada por cinco parâmetros: $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho$. Essas são as médias e variâncias de X e Y, e suas correlações.
- Se $\mu_X = \mu_Y = 0$ e $\sigma_X^2 = \sigma_Y^2 = 1$, temos a distribuição normal padrão bivariada com densidade conjunta
$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2\left(1-\rho^2\right)}\right)$$
para $-\infty < x, y < \infty$ e $-1 < \rho < 1$.
- As distribuições marginais e condicionais de uma distribuição normal bivariada são normais.
- Em particular, a distribuição condicional de X dado $Y = y$ é normal com **média** ρy e **variância** $1 - \rho^2$. Da mesma forma, a distribuição condicional de Y dado $X = x$ é normal com **média** ρx e **variância** $1 - \rho^2$.

O amostrador de Gibbs
Exemplo

- O amostrador de Gibbs é implementado para simular (X, Y) a partir de uma distribuição normal padrão bivariada com correlação ρ.

$$(X, Y) \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

- Em cada etapa do algoritmo, um componente de um vetor de dois elementos é atualizado por amostragem de sua distribuição condicional, dado o outro componente.

$$p\,(X, Y = y) \sim \mathcal{N}\left(\rho y, 1 - \rho^2\right)$$
$$p\,(Y, X = x) \sim \mathcal{N}\left(\rho x, 1 - \rho^2\right)$$

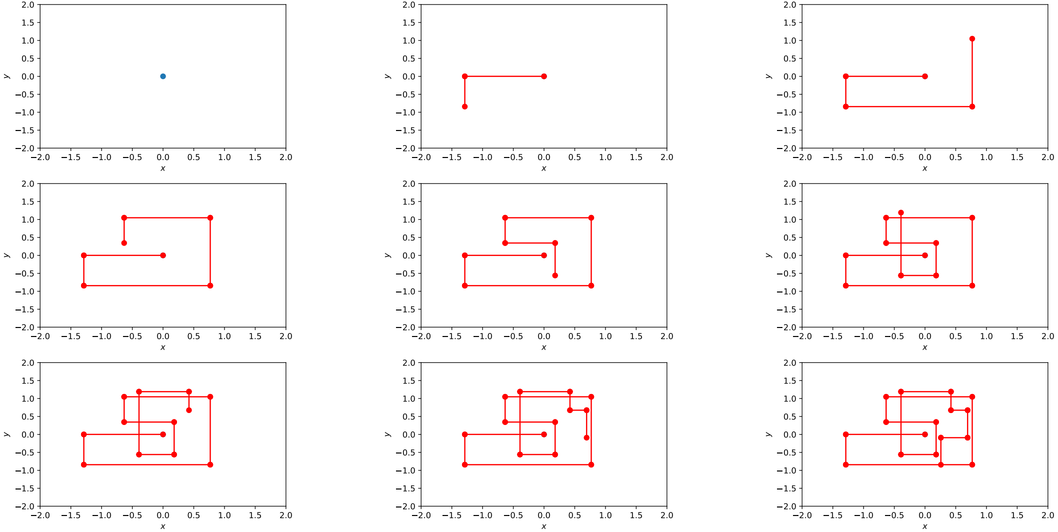
O amostrador de Gibbs
Exemplo

- Para este caso particular, o amostrador de Gibbs procede da seguinte forma:
  1. Inicialize $k = 0$ e $(x_k, y_y) = (0, 0)$.
  2. Gere x_k da distribuição condicional de X, dado $Y = y_{k-1}$, isto é, compute uma amostra da distribuição normal com média ρy_{k-1} e variância $1 - \rho^2$.
  3. Gere y_k da distribuição condicional de Y, dado $X = x_k$, isto é, compute uma amostra da distribuição normal com média ρx_k e variância $1 - \rho^2$.
  4. Faça $k = k + 1$.
  5. Retorne ao passo 2.

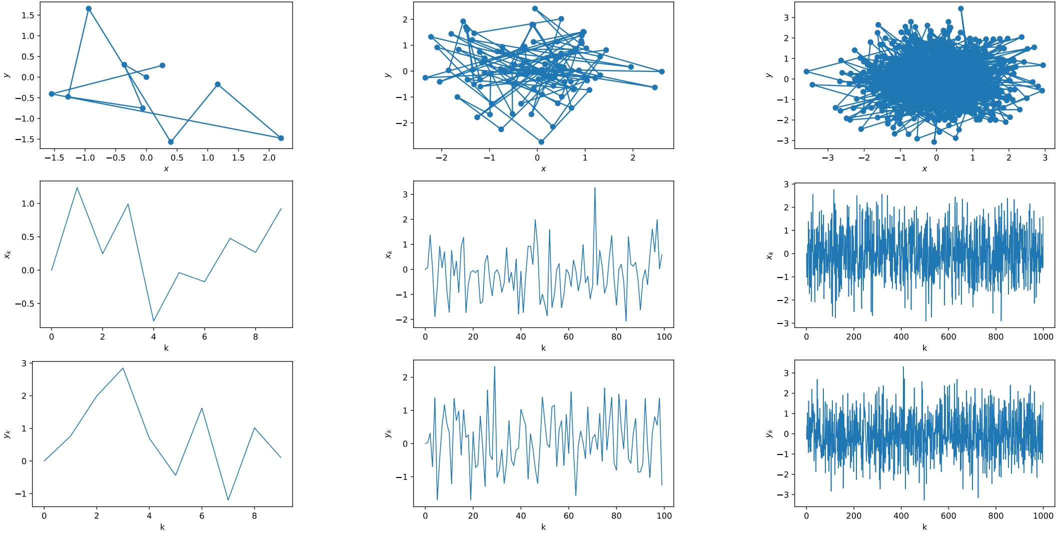
Atividade prática

Reproduza os resultados mostrados nos próximos *slides*, empregando o amostrador de Gibbs para o problema apresentado. Implemente o algoritmo mostrado acima, para algum $-1 \leq \rho \leq 1$. Use quantidade suficiente de amostras.

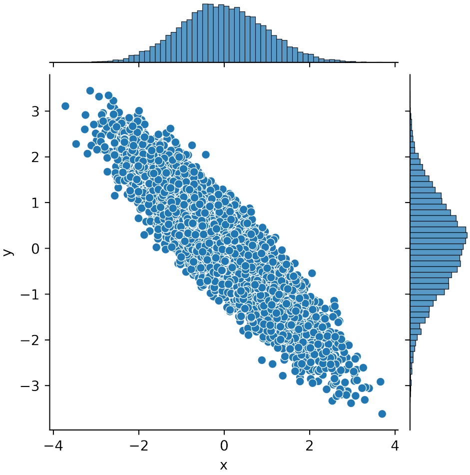
O amostrador de Gibbs
Exemplo



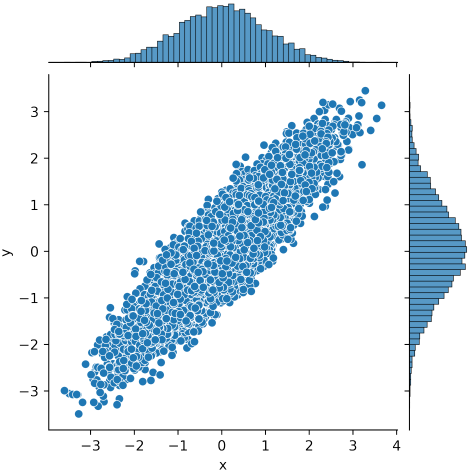
O amostrador de Gibbs
Exemplo



O amostrador de Gibbs
Exemplo

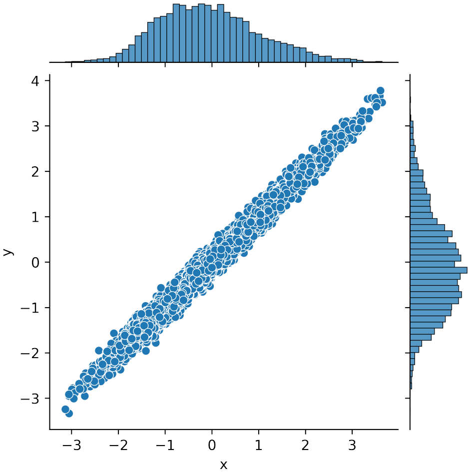


$\rho = -0{,}9$

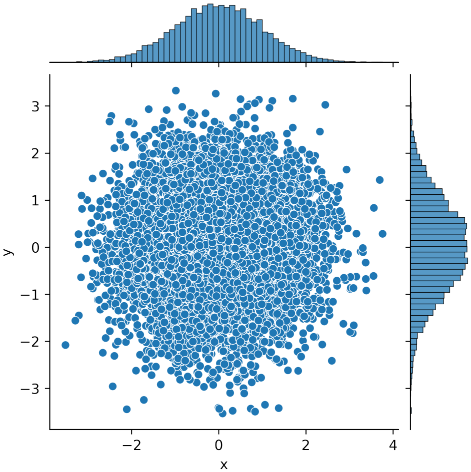


$\rho = 0{,}9$

O amostrador de Gibbs
Exemplo



$\rho = 0{,}99$



$\rho = 0$

O Algoritmo de Metropolis

- Vamos considerar uma situação genérica onde temos um modelo de amostragem $Y \sim p(y \mid \theta)$ e uma distribuição *a priori* $p(\theta)$.
- Embora na maioria dos problemas $p(y \mid \theta)$ e $p(\theta)$ possam ser calculados para quaisquer valores de y e θ,

$$p(\theta \mid y) = \frac{p(\theta)p(y \mid \theta)}{\int p(\theta)\, p(y \mid \theta)\, \mathrm{d}\theta}$$

geralmente é difícil de calcular **devido à integral no denominador**.

- Se pudéssemos obter uma amostra de $p(\theta \mid y)$, poderíamos gerar $\theta^{(1)}, \ldots, \theta^{(S)} \overset{i.i.d.}{\sim} p(\theta \mid y)$ e obter aproximações de Monte Carlo para quantidades *a posteriori*, com

$$\mathrm{E}[g(\theta) \mid y] \approx \frac{1}{S} \sum_{s=1}^{S} g\left(\theta^{(s)}\right)$$

- **Mas e se não pudermos amostrar diretamente de $p(\theta \mid y)$?**

O Algoritmo de Metropolis

- Em termos da aproximação da distribuição *a posteriori*, o crítico não é que tenhamos amostras i.i.d. de $p(\theta \mid y)$.
- Mas ao invés disso, somos capazes de construir uma grande coleção de θ-valores, $\left\{\theta^{(1)}, \ldots, \theta^{(S)}\right\}$, cuja distribuição empírica se aproxima de $p(\theta \mid y)$.
- A grosso modo, para quaisquer dois valores diferentes θ_a e θ_b, precisamos

$$\frac{\#\left\{\theta^{(s)} \text{'s na coleção} = \theta_a\right\}}{\#\left\{\theta^{(s)} \text{'s na coleção} = \theta_b\right\}} \approx \frac{p\left(\theta_a \mid y\right)}{p\left(\theta_b \mid y\right)}$$

- Vamos pensar **intuitivamente** sobre como podemos construir essa coleção.
- Suponha que temos uma coleção $\left\{\theta^{(1)}, \ldots, \theta^{(s)}\right\}$ à qual gostaríamos de adicionar um novo valor $\theta^{(s+1)}$.
- Vamos considerar a adição de um valor θ^* que está próximo a $\theta^{(s)}$.

O Algoritmo de Metropolis

- Devemos incluir θ^* no conjunto ou não?
- Se $p(\theta^* \mid y) > p\left(\theta^{(s)} \mid y\right)$ então queremos mais que θ^* esteja no conjunto do que $\theta^{(s)}$. θ^* é **mais verossímil** do que $\theta^{(s)}$.
- Visto que $\theta^{(s)}$ já está no conjunto, então parece que **devemos** incluir θ^* também.
- Por outro lado, se $p(\theta^* \mid y) < p\left(\theta^{(s)} \mid y\right)$ então parece que **não devemos necessariamente** inclui θ^*.
- Portanto, talvez nossa decisão de incluir θ^* ou não deva ser baseada em uma comparação de $p(\theta^* \mid y)$ com $p\left(\theta^{(s)} \mid y\right)$.
- Felizmente, essa comparação **pode ser feita** mesmo que não possamos calcular $p(\theta \mid y)$:

$$r = \frac{p(\theta^* \mid y)}{p\left(\theta^{(s)} \mid y\right)} = \frac{p(y \mid \theta^*)\, p(\theta^*)}{p(y)} \frac{p(y)}{p\left(y \mid \theta^{(s)}\right) p\left(\theta^{(s)}\right)} = \frac{p(y \mid \theta^*)\, p(\theta^*)}{p\left(y \mid \theta^{(s)}\right) p\left(\theta^{(s)}\right)}$$

O Algoritmo de Metropolis

- Tendo calculado r, como devemos **proceder**?
- Se $r > 1$:
  - **Intuição**: Como $\theta^{(s)}$ já está em nosso conjunto, devemos incluir θ^*, pois tem uma probabilidade maior do que $\theta^{(s)}$.
  - **Procedimento**: Aceite θ^* em nosso conjunto, ou seja, defina $\theta^{(s+1)} = \theta^*$.
- Se $r < 1$:
  - **Intuição**: A frequência relativa de θ-valores em nosso conjunto igual a θ^* em comparação com aqueles iguais a $\theta^{(s)}$ deve ser $p\left(\theta^* \mid y\right)/p\left(\theta^{(s)} \mid y\right) = r$. Isso significa que para cada instância de $\theta^{(s)}$, devemos ter apenas uma "fração" de uma instância de um valor θ^*.
  - **Procedimento**: Defina $\theta^{(s+1)}$ igual a θ^* ou $\theta^{(s)}$, com probabilidade r e $1 - r$ respectivamente.
- Esta é a intuição básica por trás do famoso **algoritmo de Metropolis**.

O Algoritmo de Metropolis

- O algoritmo de Metropolis procede amostrando um valor de proposta θ^* próximo ao valor atual $\theta^{(s)}$ usando uma **distribuição simétrica de proposta** $J\left(\theta^* \mid \theta^{(s)}\right)$.
- **Simétrico** aqui significa que $J(\theta_b \mid \theta_a) = J(\theta_a \mid \theta_b)$, ou seja, a probabilidade de propondo $\theta^* = \theta_b$ dado que $\theta^{(s)} = \theta_a$ é igual à probabilidade de propor $\theta^* = \theta_a$ dado que $\theta^{(s)} = \theta_b$.
- Normalmente $J\left(\theta^* \mid \theta^{(s)}\right)$ é muito simples, com amostras de $J\left(\theta^* \mid \theta^{(s)}\right)$ estando perto de $\theta^{(s)}$ com alta probabilidade.
- **Exemplos**:
  - $J\left(\theta^* \mid \theta^{(s)}\right) = \text{Uniforme}\left(\theta^{(s)} - \delta, \theta^{(s)} + \delta\right)$
  - $J\left(\theta^* \mid \theta^{(s)}\right) = \text{Normal}\left(\theta^{(s)}, \delta^2\right)$
- O valor do parâmetro δ é geralmente escolhido para fazer o algoritmo de aproximação funcionar de forma **eficiente**.
- **Como δ afeta a eficiência?**

O Algoritmo de Metropolis
Resumo

- Tendo obtido um valor de proposta θ^*, nós o adicionamos ou uma cópia de $\theta^{(s)}$ ao nosso conjunto, dependendo da proporção $r = p\left(\theta^* \mid y\right) / p\left(\theta^{(s)} \mid y\right)$.
- Especificamente, dado $\theta^{(s)}$, o algoritmo de Metropolis gera um valor $\theta^{(s+1)}$ como segue:
  1. Gere uma amostra $\theta^* \sim J\left(\theta \mid \theta^{(s)}\right)$.
  2. Calcule a taxa de aceitação:
  $$r = \frac{p\left(\theta^* \mid y\right)}{p\left(\theta^{(s)} \mid y\right)} = \frac{p\left(y \mid \theta^*\right) p\left(\theta^*\right)}{p\left(y \mid \theta^{(s)}\right) p\left(\theta^{(s)}\right)}$$
  3. Tome
  $$\theta^{(s+1)} = \begin{cases} \theta^*, \text{ com probabilidade } \min(r, 1) \\ \theta^{(s)}, \text{ com probabilidade } 1 - \min(r, 1) \end{cases}$$
- A etapa 3 pode ser realizada amostrando $u \sim \text{Uniforme}(0, 1)$ e definindo $\theta^{(s+1)} = \theta^*$ se $u < r$, ou definindo $\theta^{(s+1)} = \theta^{(s)}$ caso contrário.

O Algoritmo de Metropolis
Exemplo prático

- A NFL tem 32 times e cada time joga 16 jogos da temporada regular por ano, para um total de $N = 256$ jogos.
- De acordo com o site Frontline, houve $Y_1 = 171$ concussões em 2012, $Y_2 = 152$ concussões em 2013, $Y_3 = 123$ concussões em 2014 e $Y_4 = 199$ concussões em 2015.
- O número de concussões é modelado por $Y_i \sim \text{Poisson}\left(N\lambda_i\right)$, onde $\lambda_i = \exp\left(\beta_1 + i\beta_2\right)$ é a taxa no ano i .
- Para completar o modelo Bayesiano, $\beta_1, \beta_2 \sim \text{Normal}\left(0, \tau^2\right)$.
- O logaritmo da taxa média de concussão é linear no tempo, com β_2 determinando a inclinação. O **objetivo** é determinar se a taxa de concussão está aumentando, ou seja, $\beta_2 > 0$.

Atividade prática

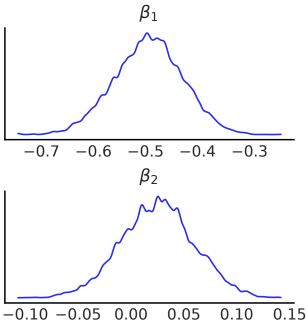
Reproduza os resultados do próximo *slide*, usando as informações apresentadas aqui. Este problema é apresentado em [1], mas **você deve tentar implementar primeiro!**

[1] S. K. Ghosh and B. J. Reich.
Bayesian statistical methods.
Chapman & Hall/CRC, Boca Raton, 1 edition, 2019.

O Algoritmo de Metropolis

Exemplo prático

Distribuição *a posteriori* e traço de β_1, β_2.



Os *boxplots* são a distribuição *a posteriori* de $N\lambda_i = N \exp\left(\beta_1 + i\beta_2\right)$, e os pontos são as amostras observadas.