

Fundação Getúlio Vargas  
Escola de Matemática Aplicada

Trabalho da A2  
Disciplina de Processos Estocásticos

Daniel Falqueto  
Ezequiel Braga  
Lucas Westfal  
Pedro Lima Garcia

Rio de Janeiro  
2023

# Sumário

<b>1</b>	<b>Cadeias de Markov aplicadas em Monte Carlo</b>	<b>2</b>
1.1	Método de Monte Carlo . . . . .	2
1.2	O princípio do MCMC . . . . .	5
1.3	O algoritmo de Metropolis-Hastings . . . . .	6
1.4	O amostrador de Gibbs de dois estágios . . . . .	6
1.5	O amostrador de Gibbs de múltiplos estágios . . . . .	6
<b>2</b>	<b>Aplicações</b>	<b>7</b>
2.1	Amostragem de distribuições . . . . .	7
2.2	Cálculo de distribuição a posteriori aproximada por MCMC comparado com posteriori conjugada . . . . .	9
2.3	Diaconis e a carta criptografada . . . . .	12
2.3.1	O problema a ser resolvido . . . . .	12
2.3.2	O método . . . . .	12
2.3.3	Resultado . . . . .	13
2.3.4	Implementação . . . . .	13
	<b>Referências</b>	<b>15</b>

# 1 Cadeias de Markov aplicadas em Monte Carlo

Aqui exploramos Monte Carlo via Cadeias de Markov (MCMC) em seus aspectos teórico e computacional. Exemplos são apresentados no corpo do texto e códigos das implementações se encontram no repositório referenciado.

## 1.1 Método de Monte Carlo

Na matemática em geral, em diversas situações nos deparamos com cálculos de integrais dificilmente calculadas a mão ou até mesmo utilizando sistemas computacionais. Nesse caso, utilizar o método Monte Carlo pode ser uma solução ou uma ótima aproximação.

Antes de aplicar nossas técnicas de simulação a problemas mais práticos, é necessário desenvolver suas propriedades com algum detalhe. Isto é mais facilmente conseguido olhando para o problema genérico de avaliar a integral

$$E_f[h(X)] = \int_{\chi} h(x)f(x) dx, \quad (1.1)$$

é natural propor usar uma amostra  $(X_1, \dots, X_m)$  gerada da densidade  $f$  para aproximar a integral pela média empírica

$$\bar{h}_m = \frac{1}{m} \sum_{j=1}^m h(x_j),$$

onde  $\bar{h}_m$  converge quase certamente para  $E_f[h(X)]$  pela lei forte dos grandes números. Além disso, quando  $h^2$  tem valor esperado finito em  $f$ , a velocidade de convergência de  $\bar{h}_m$  pode ser avaliada uma vez que a variância

$$\text{var}(\bar{h}_m) = \frac{1}{m} \int_{\chi} (h(x) - E_f[h(X)])^2 f(x) dx,$$

pode ser estimada pela amostra  $(X_1, \dots, X_m)$  como

$$v_m = \frac{1}{m^2} \sum_{j=1}^m [h(x_j) - \bar{h}_m]^2.$$

Para  $m$  grande

$$\frac{\bar{h}_m - E_f[h(X)]}{\sqrt{v_m}}$$

é aproximadamente distribuído como uma variável  $N(0,1)$ .

**Exemplo 1.1.1** (Integração de Monte Carlo).

$$h(x) = [\cos(50x) + \sin(20x)]^2.$$

Para calcular a integral via Monte Carlo geramos  $U_1, U_2, \dots, U_n$  variáveis aleatórias iid  $U(0,1)$  e aproximamos  $\int h(x)dx$  com  $\sum h(U_i)/n$ . Vemos que o método converge pois com 10.000 iterações chegamos ao valor de 0,963 onde o valor exato é 0,965.

**Exemplo 1.1.2** (Testando o número de componentes). Uma situação em que as condições de regularidade da  $\chi_r^2$  padrão não se aplicam ao teste de razão de verossimilhança é o da mistura normal

$$pN(\mu, 1) + (1 - p)N(\mu + \theta, 1)$$

onde  $\theta > 0$  garante identificabilidade. Um teste sobre a existência de uma mistura não pode ser facilmente representada em um teste de hipótese, uma vez que  $H_0 : p = 0$  elimina uma parte do problema. A incapacidade de estimar o parâmetro  $p$  sob  $H_0$  resulta na probabilidade de não satisfazer as condições de regularidade necessárias.

Contudo, se a restrição de identificabilidade for considerada  $p \geq 1/2$  em vez de  $\theta > 0$ , então  $H_0$  pode ser representado como

$$H_0 : p = 1 \quad \text{ou} \quad \theta = 0$$

Queremos, portanto, determinar a distribuição limite sob esta hipótese e sob uma alternativa local. Para isso estimamos utilizando o método de Monte Carlo.

Pode parecer que o método proposto acima seja suficiente para aproximar integrais como (1.1) de forma controlada. No entanto, existem alternativas mais eficientes que não só evitam uma simulação direta de  $f$ , mas também podem ser usada repetidamente para várias integrais do formato (1.1). Em particular, a utilidade desta flexibilidade é bastante evidente nas análises bayesianas de robustez, de sensibilidade, ou para o cálculo de funções de poder de testes específicos.

**Definição 1.1.1** (Amostragem de Importância). O método de amostragem de importância é uma avaliação de (1.1) com base na geração de uma amostra  $X_1, \dots, X_n$  de uma determinada distribuição  $g$  e aproximando

$$E_f[h(X)] \approx \frac{1}{m} \sum_{j=1}^m \frac{f(X_j)}{g(X_j)} h(X_j) \quad (1.2)$$

Este método é baseado na representação alternativa de (1.1) como:

$$E_f[h(X)] = \int_{\mathcal{X}} h(x) \frac{f(x)}{g(x)} g(x) dx,$$

e o estimador converge pela mesma razão que o estimador Monte Carlo normal  $\bar{h}_m$  converge, qualquer que seja a escolha da distribuição  $g$  (desde que  $\text{supp}(g) \supset \text{supp}(f)$ ).

**Exemplo 1.1.3** (Probabilidades de cauda pequena). Se  $Z \sim N(0, 1)$  e estamos interessados na probabilidade  $P(Z > 4.5)$  (que sabemos que é muito pequena), nós podemos simular  $Z^{(i)} \sim N(0, 1)$  para  $i = 1, \dots, M$  e calcular

$$P(Z > 4.5) \approx \frac{1}{M} \sum_{i=1}^M \mathbf{1}(Z^{(i)} > 4.5).$$

É claro que o problema é que estamos calculando a probabilidade de uma ocorrência de um evento muito raro e uma simulação ingênua precisará de muitas iterações para obter uma

resposta razoável. Por isso tomamos  $Y$  uma distribuição exponencial truncada a esquerda em 4,5 com escala 1 e densidade

$$f_Y(y) = \frac{e^{-(y-4,5)}}{\int_{4,5}^{\infty} e^{-x} dx}.$$

Se agora simulamos de  $f_Y$  e usamos Amostragem de Importância nós obtemos

$$P(Z > 4,5) \approx \frac{1}{M} \sum_{i=1}^M \frac{\phi(Y^{(i)})}{f_Y(Y^{(i)})} \mathbb{1}(Y^{(i)} > 4,5) = 0.000003377.$$

**Teorema 1.1.1.** *A escolha de  $g$  que minimiza a variância do estimador (1.2) é*

$$g * (x) = \frac{|h(x)|f(x)}{\int_{\mathcal{X}} |h(z)|f(z)dz}$$

**Exemplo 1.1.4** (Estimação de matriz de transição). Considere uma cadeia de Markov com dois estados, 1 e 2, em que a matriz de transição é

$$T = \begin{bmatrix} p_1 & 1 - p_1 \\ 1 - p_2 & p_2 \end{bmatrix}$$

assuma, além disso, que  $p_1 + p_2 < 1$ . Se a amostra é  $X_1, \dots, X_m$  e a distribuição a priori é

$$\pi(p_1, p_2) = 2\mathbb{1}_{p_1+p_2<1}$$

a distribuição a posteriori de  $(p_1, p_2)$  é

$$\pi(p_1, p_2 | m_{11}, m_{12}, m_{21}, m_{22}) \propto p_1^{m_{11}} (1 - p_1)^{m_{12}} (1 - p_2)^{m_{21}} p_2^{m_{22}} \mathbb{1}_{p_1+p_2<1}$$

onde,  $m_{ij}$  é o número de passagens de  $i$  para  $j$ , isto é,

$$m_{ij} = \sum_{t=2}^m \mathbb{1}_{x_t=i} \mathbb{1}_{x_{t+1}=j}.$$

e com isso  $D = (m_{11}, \dots, m_{22})$  é estatística suficiente.

Suponha agora que as quantidades de interesse sejam as posteriores das probabilidades associadas:

$$h_1(p_1, p_2) = p_1, h_2(p_1, p_2) = p_2, h_3(p_1, p_2) = \frac{p_1}{1 - p_1}$$

e

$$h_4(p_1, p_2) = \frac{p_2}{1 - p_2}, h_5(p_1, p_2) = \log\left(\frac{p_1(1 - p_2)}{p_2(1 - p_1)}\right),$$

respectivamente.

Agora olhamos para várias maneiras de encontrar a posteriori.

A tabela abaixo fornece os estimadores das posteriores das funções  $h_j$  avaliadas para a distribuição verdadeira  $\pi$  e para as três distribuições instrumentais  $\pi_1, \pi_2$  e  $\pi_3$ . A distribuição  $\pi_3$  é claramente preferível às outras duas distribuições, uma vez que fornece a mesma estimativa que a distribuição verdadeira, com um menor custo computacional.

Distribuição	$h_1$	$h_2$	$h_3$	$h_4$	$h_5$
$\pi_1$	0,748	0,139	3,184	0,163	2,957
$\pi_2$	0,689	0,210	2,319	0,283	2,957
$\pi_3$	0,697	0,189	2,373	0,241	2,358
$\pi$	0,697	0,189	2,373	0,240	2,358

Tabela 1: Comparação das avaliações para os estimadores correspondentes a três distribuições  $\pi_i$  e a verdadeira distribuição  $\pi$  (10.000 simulações)

Vemos, portanto, que a amostragem de importância não pode ser aplicada cegamente. Em vez disso, deve-se ter cuidado ao escolher uma densidade, pois a convergência quase certa de (1.2) é apenas formal (no sentido de que pode exigir um enorme número de simulações para produzir uma aproximação precisa da quantidade de interesse). Se finitude da variância for ignorada e não detectada, isso pode resultar em um viés muito forte.

## 1.2 O princípio do MCMC

Como visto anteriormente, há várias maneiras de aproximar a integral

$$\mathcal{J} = \int h(x)f(x) dx$$

por métodos de Monte Carlo. Analogamente, os métodos de MCMC buscam encontrar maneiras de obter uma amostra  $X_1, \dots, X_n$  com distribuição aproximada  $f$ , sem precisar simular diretamente  $f$ . Assim, o princípio desses métodos consiste em usar uma cadeia de Markov ergódica com distribuição estacionária  $f$ .

**Definição 1.2.1** (Estado ergódico). Um estado  $i$  é ergódico se é aperiódico e recorrente positivo. Se todos os estados de uma cadeia de Markov irredutível são ergódicos, então, a cadeia é ergódica.

**Definição 1.2.2** (MCMC). Um método de Monte Carlo via cadeias de Markov (Markov chain Monte Carlo - MCMC) para simulação da distribuição de  $f$  é qualquer método produzindo uma cadeia de Markov ergódica  $(X_t)$  cuja distribuição estacionária é  $f$ .

Para métodos de Monte Carlo via cadeias de Markov, o uso de uma cadeia  $(X_t)$  produzida por um algoritmo de MCMC com distribuição estacionária  $f$  é idêntico ao uso de uma amostra iid de  $f$ , no sentido de que o Teorema Ergódico garante a convergência da média empírica

$$\mathcal{J}_T = \frac{1}{T} \sum_{t=1}^T h(X(t))$$

para  $\mathbb{E}[h(X)]$ .

### 1.3 O algoritmo de Metropolis-Hastings

O algoritmo de Metropolis-Hastings inicia com uma densidade objetivo  $f$ . Escolhe-se uma densidade condicional  $q(y|x)$  (formalmente,  $q$  é escolhida com respeito a uma medida de probabilidade que domina o modelo) que, na prática, seja possível simular valores dela. Uma condição geral é que a razão  $\frac{f(y)}{q(y|x)}$  seja, pelo menos, uma constante independente de  $x$ . Este processo produz uma cadeia de Markov através da seguinte transição:

Dado  $x_t$ ,

1. Gere  $Y_t \sim q(y|x_t)$ ;
2. Tome  $X_{t+1} = \begin{cases} Y_t, & \text{com probabilidade } \rho(x_t, Y_t) \\ x_t, & \text{com probabilidade } 1 - \rho(x_t, Y_t) \end{cases}$  onde  $\rho(x, y) = \min \left\{ \frac{f(y)q(x|y)}{f(x)q(y|x)}, 1 \right\}$

A função  $q$  é dita densidade proposta e  $\rho$ , a probabilidade de aceitação de Metropolis-Hastings.

**Teorema 1.3.1.** *Sejam  $X_t$ ,  $f$  e  $q$  definidas anteriormente. Para toda distribuição  $q$  cujo suporte inclui o suporte de  $f$ , a cadeia de Markov  $X_t$  tem distribuição estacionária  $f$ .*

### 1.4 O amostrador de Gibbs de dois estágios

Na seção anterior, foi discutido um método de amostragem de uma distribuição univariada  $f(x)$ . Considere agora variáveis  $X$  e  $Y$  com distribuição conjunta  $f(x, y)$  e distribuições condicionais  $f_{X|Y}(x|Y=y)$  e  $f_{Y|X}(y|X=x)$ . O algoritmo de Gibbs consiste em gerar uma cadeia de Markov  $(X_t, Y_t)$  com distribuição estacionária  $f(x, y)$  da seguinte forma:

Dado  $X_t = x_t$ , para  $t = 1, 2, \dots$ , gere

1.  $Y_t \sim f_{Y|X}(\cdot, x_{t-1})$ ;
2.  $X_t \sim f_{X|Y}(\cdot, y_t)$ .

### 1.5 O amostrador de Gibbs de múltiplos estágios

Como pode-se imaginar, a amostragem em múltiplos estágios simplesmente é uma generalização do algoritmo discutido na seção anterior. Para isso, considere  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$  com distribuição conjunta  $f_{\mathbf{X}}(\mathbf{x})$  com densidades condicionais

$$X_i | x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n \sim f_i(x_i | x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$$

para  $i = 1, 2, \dots, n$ . O amostrador de Gibbs correspondente é construído da seguinte forma:

Dado  $x^{(t)} = (x_1^{(t)}, \dots, x_n^{(t)})$ , para  $t = 1, 2, \dots$ , gere

1.  $X_1^{(t+1)} \sim f_1(x_1 | x_2^{(t)}, \dots, x_n^{(t)})$ ;
2.  $X_2^{(t+1)} \sim f_2(x_2 | x_1^{(t+1)}, x_3^{(t)}, \dots, x_n^{(t)})$ ;
- $\vdots$
- n.  $X_n^{(t+1)} \sim f_n(x_n | x_1^{(t+1)}, \dots, x_{n-1}^{(t+1)})$ ;

## 2 Aplicações

### 2.1 Amostragem de distribuições

**Exemplo 2.1.1** (Exponencial). Suponha que estamos interessados em gerar valores  $X_1, \dots, X_n$  da distribuição exponencial com parâmetro  $\lambda$ . Para isso, considere

$$f(x) = \lambda \exp\{-\lambda x\}, \quad x \geq 0$$

e seja

$$q(y|x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(y-x)^2}{2}\right\}$$

ou seja,  $Y|X = x \sim N(x, 1)$ . Observe que  $q(x|y) = q(y|x)$  e, então,

$$\rho(x, y) = \min\left\{\frac{f(y)}{f(x)}, 1\right\} = \min\{\exp\{-\lambda(y-x)\}, 1\}$$

Daí, chutamos um valor inicial  $x_0$  e aplicamos o procedimento descrito anteriormente, de modo que geramos  $U \sim U[0, 1]$  e definimos

$$X_{t+1} = \begin{cases} Y_t, & \text{se } U \leq \rho(x_t, Y_t) \\ x_t, & \text{c.c.} \end{cases}$$

Foram realizadas 10.000 iterações, com  $\lambda = 1$  e  $x_0 = 2$ , resultando no seguinte histograma:

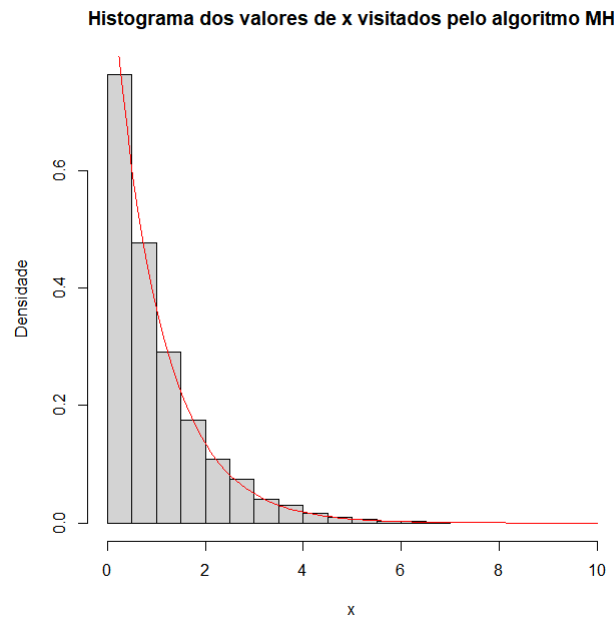


Figura 1: Histograma dos valores gerados por Metropolis-Hastings comparados com a densidade da exponencial com parâmetro 1 (em vermelho).



**Exemplo 2.1.2** (Normal). Conforme o exemplo anterior, queremos  $X_1, \dots, X_n$  da distribuição  $N(0, 1)$ . Neste caso, tome

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp \left\{ \frac{-x^2}{2} \right\}$$

e

$$q(y|x) = \mathbb{I}_{[x-1, x+1]}$$

isto é,  $Y|X = x \sim U[x-1, x+1]$ . Daí, temos a seguinte função de aceitação:

$$\rho(x, y) = \min \left\{ \frac{f(y)}{f(x)}, 1 \right\} = \min \{ \exp\{-(y^2 - x^2)\}, 1 \}$$

Por fim, realizamos o mesmo procedimento para gerar a cadeia. Foram realizadas 10.000 iterações, resultado no seguinte histograma:

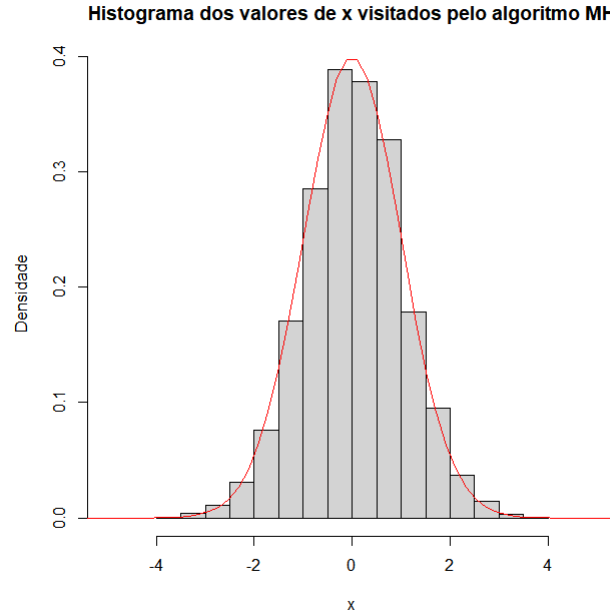


Figura 2: Histograma dos valores gerados por Metropolis-Hastings comparados com a densidade da normal padrão (em vermelho).

**Exemplo 2.1.3** (Normal Bivariada). Suponha que queremos gerar  $(X_1, Y_1), \dots, (X_n, Y_n)$  com distribuição normal bivariada  $N(0, \Sigma)$ , com  $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ . Para isso, usamos os seguintes fatos:

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2) \right\}$$

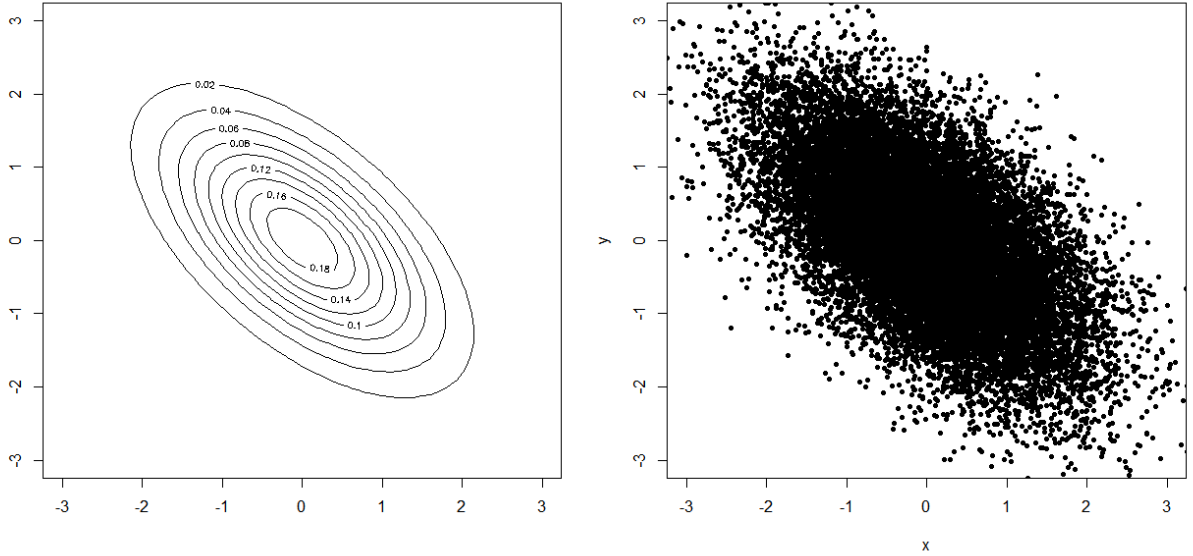
$$Y|X = x \sim N(\rho x, 1 - \rho^2)$$

$$X|Y = y \sim N(\rho y, 1 - \rho^2)$$

Daí, seguimos os passos descritos anteriormente: chutamos  $y_0$  e geramos

1.  $X^{(t+1)} \sim f_{X|Y}(x|Y^{(t)} = y^{(t)});$
2.  $Y^{(t+1)} \sim f_{Y|X}(y|X^{(t+1)} = x^{(t+1)}).$

Foram gerados 20.000 valores com correlação  $-0,6$ , resultando nas figuras a seguir:



(a) Curvas de nível da normal bivariada

(b) Valores gerados pelo amostrador de Gibbs

Figura 3: Amostragem da normal bivariada

## 2.2 Cálculo de distribuição a posteriori aproximada por MCMC comparado com posteriori conjugada

No paradigma Bayesiano, associamos amostras aleatórias a famílias paramétricas e quantificamos a incerteza dos parâmetros em variáveis aleatórias. Considere o modelo com independência:

$$X_1, \dots, X_n \sim \text{Normal}(\theta, 1)$$

$$\theta \sim N(\mu, v_0^2)$$

Pelo teorema de Bayes, temos que

$$P(\theta|x) = \frac{P(x|\theta)\pi(\theta)}{P(x)}$$

Então podemos estimar a distribuição de  $\theta$  usando uma distribuição a priori, a verossimilhança e a distribuição marginal da amostra. Os dois primeiros são simples de avaliar, enquanto o terceiro envolve uma integral que na maioria dos casos não é trivial. No exemplo, seja  $\pi(\theta)$  a distribuição a priori de  $\theta$  e seja  $f$  a função de densidade de probabilidade da  $Normal(\theta, \sigma^2)$  ( $\theta \in \Omega$ ):

$$P(x) = \int_{\Omega} \pi(\theta) \Pi_{i=1}^n f(x_i|\theta) d\theta$$

Pelo Teorema Ergódico, podemos aproximar essa integral simulando uma cadeia de Markov  $(Y_i)_{i=1}^N$  cuja distribuição estacionária seja a distribuição a priori, com  $\frac{1}{N} \sum_{i=1}^N \Pi_{i=1}^n f(y_i|\theta) \rightarrow \mathbb{E}[f(X)] = P(x)$ .

Usando o algoritmo de Metropolis-Hastings, simulam-se 400 dados identicamente distribuídos de uma distribuição  $\mathcal{N}(0, 1)$ , a priori de  $\theta$ .

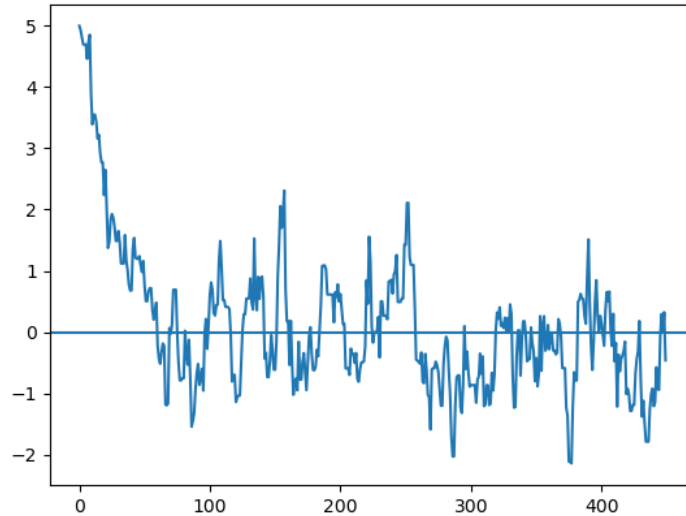


Figura 4: Primeiras iterações de uma cadeia de Markov cuja distribuição estacionária é  $\mathcal{N}(0, 1)$

Truncando a simulação e avaliando a média de cada truncamento, ficamos com

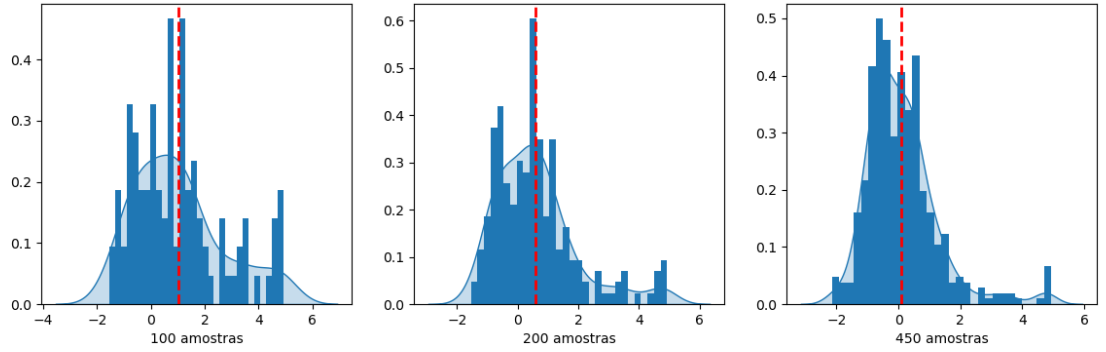


Figura 5: Histograma das simulações

Avaliamos a distribuição a posteriori, explicitamente, por meio de

$$P(\theta|x) = \frac{\pi(\theta)\prod_{i=1}^n f(x_i|\theta)}{\int_{\Omega} \pi(\theta)\prod_{i=1}^n f(x_i|\theta)d\theta}$$

Mas podemos saber imediatamente qual é o resultado assintótico pela teoria de prioris conjugadas, como trata o teorema a seguir.

**Teorema 2.2.1.** *Seja  $X_1, \dots, X_n$  amostra aleatória de uma distribuição  $\mathcal{N}(\theta, \sigma^2)$ , para  $\theta$  desconhecido e  $\sigma^2$  conhecido. Seja também uma distribuição a priori de  $\theta$   $\mathcal{N}(\mu_0, v_0^2)$ . A distribuição a posteriori de  $\theta$  dados  $X_i = x_i (i = 1, 2, \dots, n)$  é  $\mathcal{N}\left(\frac{\sigma^2\mu_0 + nv_0^2\bar{x}_n}{\sigma^2 + nv_0^2}, \frac{\sigma^2v_0^2}{\sigma^2 + nv_0^2}\right)$*

Usando a priori simulada, podemos resolver o problema proposto para  $X_i \sim \mathcal{N}(1, 1)$ . Gerando a amostra e comparando a distribuição a posteriori com a média distribuição conjugada, em vermelho:

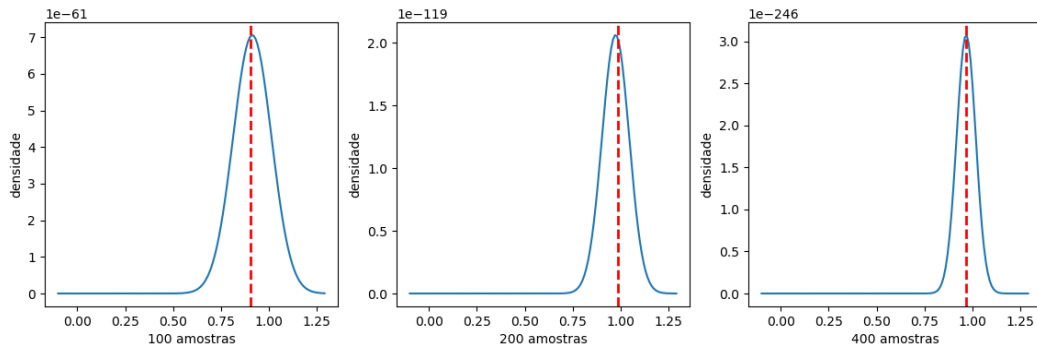


Figura 6: Distribuições a posteriori de  $\theta$

## 2.3 Diaconis e a carta criptografada

### 2.3.1 O problema a ser resolvido

Certo dia, o Departamento de Estatística de Stanford, onde Persi Diaconis trabalhava, recebeu do psicólogo da penitenciária estadual uma carta que havia sido interceptada pelos guardas da prisão. Eis abaixo um trecho da mesma:

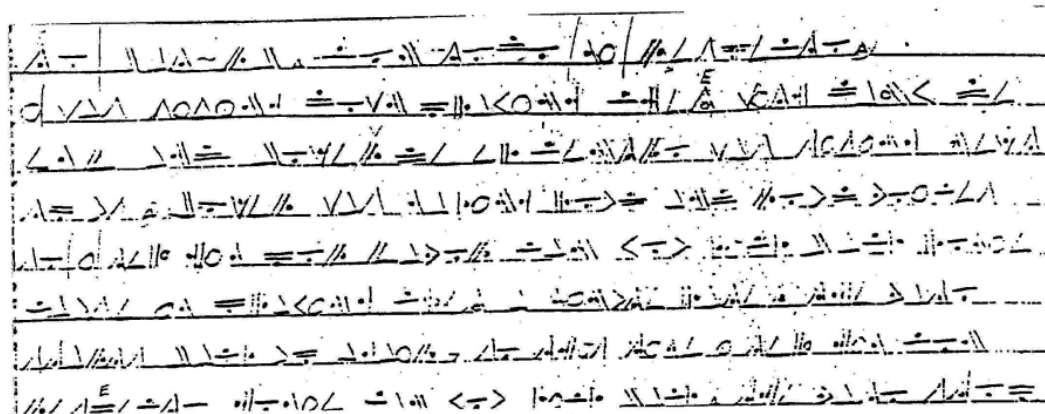


Figura 7: Trecho da carta criptografa.

Diaconis logo notou que se tratava de uma mensagem criptografada. Entretanto, considerando só as letras do alfabeto, existem  $26! = 403.291.461.126.605.635.584.000.000$  maneiras diferentes de se criptografar uma mensagem. Obviamente, tentar todas elas até que alguma fizesse sentido não era uma opção viável, mas Diaconis sabia que esses mapeamentos todos não eram equiprováveis. Elaborou então uma forma de decidir quais deles deveriam ser considerados.

### 2.3.2 O método

A ideia de Diaconis consistia em modelar o idioma como uma Cadeia de Markov  $M$  em que cada estado correspondesse a um caractere e cada entrada  $M_{ij}$  fosse a probabilidade de  $j$  suceder  $i$ . Tais probabilidades seriam calculadas através das frequências de sequências de caracteres dois a dois em um texto de treinamento dado.

Com a Cadeia de Markov gerada, cada mapeamento  $f$  seria associado à sua “plausibilidade”  $Pl(f)$ , definida pela função score

$$Pl(f) = \prod_i M(f(s_i), f(s_{i+1})),$$

onde  $s_i$  percorre os caracteres consecutivos na mensagem criptografada. Com isso, o algoritmo parte de um mapeamento aleatório  $f$  e, a cada iteração,

- 1) estando em  $f$ , calcula  $Pl(f)$ ;
- 2) muda o mapa de um par aleatório de caracteres, definindo um novo mapeamento  $f^*$ ;
- 3) calcula  $Pl(f^*)$ ;

- 4) se  $Pl(f*) \geq Pl(f)$ , vai para  $f^*$ ;
- 5) caso contrário, fica em  $f$  com probabilidade  $1/2$ .

A incerteza presente no passo 5 impede que o algoritmo fique preso a um mapeamento com score máximo local, garantindo que ele sempre possa visitar mapeamentos mais plausíveis a longo prazo.

### 2.3.3 Resultado

Com alguns ajustes manuais, foi possível decifrar a carta - que incluía inglês, espanhol e jargão criminal. Segue um trecho da mensagem original:

to bat-rb. con todo mi respeto. i was sitting down playing chess with danny de emf and boxer de el centro was sitting next to us. boxer was making loud and loud voices so i tell him por favor can you kick back homie cause im playing chess a minute later the vato starts back up again so this time i tell him con respecto homie can you kick back. the vato stop for a minute and he starts up again so i tell him check this out shut the f\*\*k up cause im tired of your voice and if you got a problem with it we can go to celda and handle it. i really felt disrespected thats why i told him. anyways after i tell him that the next thing I know that vato slashes me and leaves. dy the time i figure im hit i try to get away but the c.o. is walking in my direction and he gets me right dy a celda. so i go to the hole. when im in the hole my home boys hit doxer so now "b" is also in the hole. while im in the hole im getting schoold wrong and

Figura 8: Trecho da carta descriptografa.

Essa aplicação é particularmente interessante, pois permitiu que um assalto que estava sendo planejado pelo líder de uma gangue de traficantes através dessa carta fosse evitado.

### 2.3.4 Implementação

Nosso modelo leu o livro “Sob a redoma” (Stephen King) em português, gerando a seguinte matriz de transição:

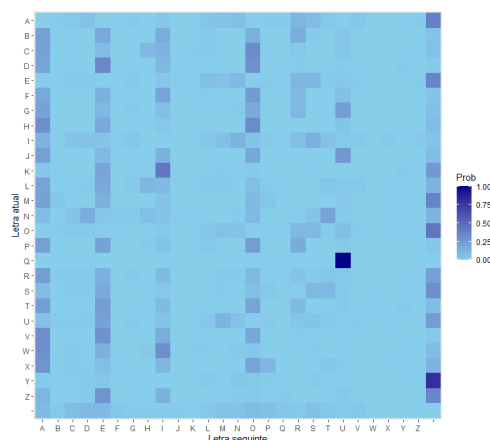
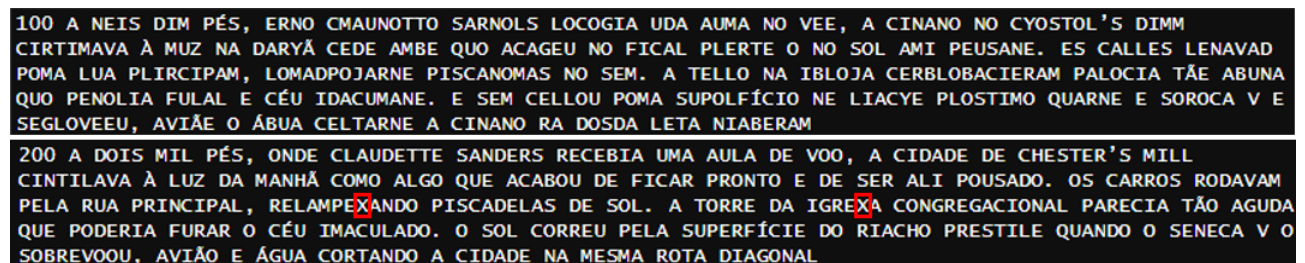


Figura 9: Matriz de transição entre os caracteres.

Nosso objetivo era que, partindo de um mapeamento aleatório, após 500 iterações ele retornasse o seguinte trecho:

“A dois mil pés, onde Claudette Sanders recebia uma aula de voo, a cidade de Chester’s Mill cintilava à luz da manhã como algo que acabou de ficar pronto e de ser ali pousado. Os carros rodavam pela rua principal, relampejando piscadelas de sol. A torre da Igreja Congregacional parecia tão aguda que poderia furar o céu imaculado. O sol correu pela superfície do riacho Prestile quando o Seneca V o sobrevoou, avião e água cortando a cidade na mesma rota diagonal.”

Na centésima iteração, o resultado ainda estava bem ruim - apenas a palavra “pés” estava descriptografada corretamente. Porém, a partir da iteração 200, o algoritmo já aparentou convergir para um mapeamento quase perfeito, trocando apenas a letra *J* pela letra *X*.

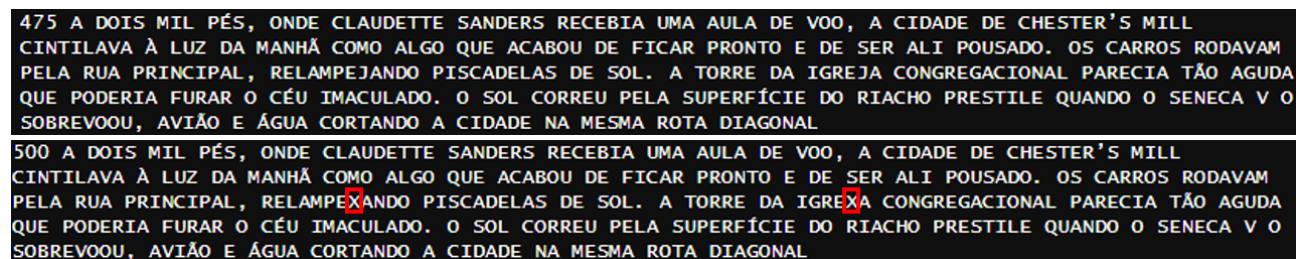


100 A NEIS DIM PÉS, ERNO CMAUNOTTO SARNOLS LOCOGIA UDA AUMA NO VEE, A CINANO NO CYOSTOL'S DIMM CIRTIMAVA À MUZ NA DARYÃ CEDE AMBE QUO ACAGEU NO FICAL PLERTE O NO SOL AMI PEUSANE. ES CALLES LENAVAL POMA LUA PLIRCIPAM, LOMADPOJARNE PISCANOMAS NO SEM. A TELLO NA IBLOJA CERBLOBACIERAM PALOCIA TÂE ABUNA QUO PENOLIA FULAL E CÉU IDACUMANE. E SEM CELLOU POMA SUPOLFÍCIO NE LIACYE PLOSTIMO QUARNE E SOROCA V E SEGLOVEEU, AVIÃO O ÁBUA CELTARNE A CINANO RA DOSDA LETA NIABERAM

200 A DOIS MIL PÉS, ONDE CLAUDETTE SANDERS RECEBIA UMA AULA DE VOO, A CIDADE DE CHESTER'S MILL CINTILAVA À LUZ DA MANHÃ COMO ALGO QUE ACABOU DE FICAR PRONTO E DE SER ALI POUSADO. OS CARROS RODAVAM PELA RUA PRINCIPAL, RELAMPEJANDO PISCAD~~X~~ELAS DE SOL. A TORRE DA IGRE~~X~~A CONGREGACIONAL PARECIA TÃO AGUDA QUE PODERIA FURAR O CÉU IMACULADO. O SOL CORREU PELA SUPERFÍCIE DO RIACHO PRESTILE QUANDO O SENECA V O SOBREVOOU, AVIÃO E ÁGUA CORTANDO A CIDADE NA MESMA ROTA DIAGONAL

Figura 10: Iterações 100 e 200.

Notou-se que, da iteração 473 a 477, o algoritmo retornou a mensagem original, mas depois voltou a convergir pro mapeamento de score dominante:



475 A DOIS MIL PÉS, ONDE CLAUDETTE SANDERS RECEBIA UMA AULA DE VOO, A CIDADE DE CHESTER'S MILL CINTILAVA À LUZ DA MANHÃ COMO ALGO QUE ACABOU DE FICAR PRONTO E DE SER ALI POUSADO. OS CARROS RODAVAM PELA RUA PRINCIPAL, RELAMPEJANDO PISCAD~~X~~ELAS DE SOL. A TORRE DA IGRE~~X~~A CONGREGACIONAL PARECIA TÃO AGUDA QUE PODERIA FURAR O CÉU IMACULADO. O SOL CORREU PELA SUPERFÍCIE DO RIACHO PRESTILE QUANDO O SENECA V O SOBREVOOU, AVIÃO E ÁGUA CORTANDO A CIDADE NA MESMA ROTA DIAGONAL

500 A DOIS MIL PÉS, ONDE CLAUDETTE SANDERS RECEBIA UMA AULA DE VOO, A CIDADE DE CHESTER'S MILL CINTILAVA À LUZ DA MANHÃ COMO ALGO QUE ACABOU DE FICAR PRONTO E DE SER ALI POUSADO. OS CARROS RODAVAM PELA RUA PRINCIPAL, RELAMPEJANDO PISCAD~~X~~ELAS DE SOL. A TORRE DA IGRE~~X~~A CONGREGACIONAL PARECIA TÃO AGUDA QUE PODERIA FURAR O CÉU IMACULADO. O SOL CORREU PELA SUPERFÍCIE DO RIACHO PRESTILE QUANDO O SENECA V O SOBREVOOU, AVIÃO E ÁGUA CORTANDO A CIDADE NA MESMA ROTA DIAGONAL

Figura 11: Iterações 475 e 500.

O método se mostrou eficiente para resolver um problema a princípio muito complicado. Porém, a qualidade do resultado depende fortemente do texto de treinamento e do ponto de partida das iterações.

De todo modo, quanto mais iterações forem feitas, mais certamente o algoritmo convergirá para a mensagem original.

## Referências

- [1] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer Texts in Statistics, 2004.
- [2] Gareth O. Roberts and Jeffrey S. Rosenthal. General State Space Markov Chains and MCMC Algorithms. *Probability Surveys*, 2004.
- [3] Robert P. Dobrow. *Introduction to Stochastic Processes With R*. John Wiley & Sons, 2016.
- [4] Persi Diaconis. The Markov chain Monte Carlo revolution. *Bulletin of the American Mathematical Society*, 2009.
- [5] Morris DeGroot and Mark Schervish. *Probability and Statistics*. 2012.
- [6] Ezequiel. Braga. mcmc. <https://github.com/EzequielEBS/mcmc.git>, 2023.