Disciplina: Estatística Computacional **(Entrega: 28/10/19)**

Lista de Exercícios 2

Professor(a): Eduardo Mendes **Aluno:** Franklin Alves de Oliveira

Exercise 1 (Monte Carlo for Gaussians)

1.

Conforme sugerido pelo enunciado, vamos calcular $\mathbb{E}_\pi[\phi(x)]$:

$$\mathbb{E}_\pi[\phi(x)] = \int_{\mathbb{X}} \phi(x)\pi(x)\ dx = \int_{\mathbb{X}} \phi(x)\frac{1}{(2\pi)^{d/2}} \exp\left\{-\frac{1}{2}x^T x\right\} dx$$

Tomando $x = y + \theta$, temos:

$$\begin{aligned}
\mathbb{E}_\pi[\phi(x)] = \mathbb{E}_\pi[\phi(y+\theta)] &= \int_{\mathbb{X}} \phi(y+\theta)\frac{1}{(2\pi)^{d/2}} \exp\left\{-\frac{1}{2}(y+\theta)^T(y+\theta)\right\} dy = \\
&= \int_{\mathbb{X}} \phi(y+\theta)\frac{1}{(2\pi)^{d/2}} \exp\left\{-\frac{1}{2}(y+\theta)^T(z+\theta)\right\} dy = \\
&= \int_{\mathbb{X}} \phi(y+\theta)\frac{1}{(2\pi)^{d/2}} \exp\left\{-\frac{1}{2}(y^T y + 2\theta^T y + \theta^T\theta)\right\} dy = \\
&= \int_{\mathbb{X}} \phi(y+\theta)\frac{1}{(2\pi)^{d/2}} \exp\left\{-\frac{1}{2}y^T y\right\} \exp\left\{-\frac{1}{2}\theta^T\theta - \theta^T y\right\} dy = \\
&= \int_{\mathbb{X}} \phi(y+\theta)\pi(y) \exp\left\{-\frac{1}{2}\theta^T\theta - \theta^T y\right\} dy = \\
&= \mathbb{E}_\pi\left[\phi(y+\theta)\exp\left\{-\frac{1}{2}\theta^T\theta - \theta^T y\right\}\right]
\end{aligned}$$

Fazendo uma troca de variáveis, i.e., $y = X$, temos:

$$\mathbb{E}_\pi[\phi(x)] = \mathbb{E}_\pi\left[\phi(X+\theta)\exp\left\{-\frac{1}{2}\theta^T\theta - \theta^T X\right\}\right]$$

□

2.

Tomando a fórmula da variância, temos:

$$\sigma^2(\theta) = E_\pi\left[\phi^2(X+\theta)\exp\left\{-\theta^T\theta - 2\theta^T X\right\}\right] - E_\pi\left[\phi(x)\right]^2 \tag{1}$$

Note que:

$$E_\pi\left[\phi^2(x)\right] = E_\pi\left[\phi^2(X+\theta)\exp\left\{-\theta^T\theta - 2\theta^T X\right\}\right] =$$

$$= \int \phi^2(x+\theta)\frac{1}{(2\pi)^{d/2}}\exp\left\{-\theta^T\theta - 2\theta^T x\right\}\exp\left\{-\frac{1}{2}x^T x\right\}dx =$$

$$= \frac{1}{(2\pi)^{d/2}}\int \phi^2(x+\theta)\exp\left\{-\theta^T\theta - 2\theta^T x\right\}\exp\left\{-\frac{1}{2}x^T x\right\}dx \tag{2}$$

Trocando a variável x por $y-\theta$, é fácil ver de 2 que:

$$E_\pi\left[\phi(x)^2\right] = \frac{1}{(2\pi)^{d/2}}\int \phi^2(y)\exp\left\{-\theta^T\theta - 2\theta^T(y-\theta)\right\}\exp\left\{-\frac{1}{2}(y-\theta)^T(y-\theta)\right\}dy =$$

$$= \frac{1}{(2\pi)^{d/2}}\int \phi^2(y)\exp\left\{-\theta^T\theta - 2\theta^T(y-\theta) - \frac{1}{2}(y-\theta)^T(y-\theta)\right\}dy =$$

$$= \frac{1}{(2\pi)^{d/2}}\int \phi^2(y)\exp\left\{-\theta^T\theta - 2\theta^T(y-\theta) - \frac{1}{2}\left(y^T y + y^T\theta + \theta^T y - \theta^T\theta\right)\right\}dy =$$

$$= \frac{1}{(2\pi)^{d/2}}\int \phi^2(y)\exp\left\{-\theta^T\theta - 2\theta^T(y-\theta) - \frac{1}{2}y^T y + y^T\theta - \frac{1}{2}\theta^T\theta\right\}dy =$$

$$= \frac{1}{(2\pi)^{d/2}}\int \phi^2(y)\exp\left\{-\theta^T\theta - 2\theta^T y + 2\theta^T\theta - \frac{1}{2}y^T y + y^T\theta - \frac{1}{2}\theta^T\theta\right\}dy =$$

$$= \frac{1}{(2\pi)^{d/2}}\int \phi^2(y)\exp\left\{\frac{1}{2}\theta^T\theta - \theta^T y - \frac{1}{2}y^T y\right\}dy =$$

$$= \frac{1}{(2\pi)^{d/2}}\int \phi^2(y)\exp\left\{\frac{1}{2}\theta^T\theta - \theta^T y\right\}\exp\left\{-\frac{1}{2}y^T y\right\}dy =$$

$$= \mathbb{E}\left[\phi^2(X)\exp\left\{\frac{1}{2}\theta^T\theta - \theta^T X\right\}\right], \tag{3}$$

Uma vez que $-\frac{1}{2}x^T x + \frac{1}{2}(x-\theta)^T(x-\theta) = \frac{1}{2}\theta^T\theta - \theta^T x$. Sendo assim, é fácil ver que

$$\mathbb{E}_\pi\left[\phi^2(X)\exp\left\{-\frac{1}{2}X^T X + \frac{1}{2}(X-\theta)^T(X-\theta)\right\}\right] = \mathbb{E}_\pi\left[\phi^2(X)\exp\left\{\frac{1}{2}\theta^T\theta - \theta^T X\right\}\right] \tag{4}$$

Substituindo 4 em 1:

$$\sigma^2(\theta) = \mathbb{E}_\pi\left[\phi^2(X)\exp\left\{\frac{1}{2}\theta^T\theta - \theta^T X\right\}\right] - \mathbb{E}_\pi\left[\phi(x)\right]^2 \tag{5}$$

Como queríamos demonstrar. □

3.

Do item 2, temos que

$$\sigma^2(\theta) = \mathbb{E}_\pi\left[\phi^2(X)\exp\left\{-\frac{1}{2}X^{\mathrm{T}}X + \frac{1}{2}(X-\theta)^{\mathrm{T}}(X-\theta)\right\}\right] - \mathbb{E}_\pi[\phi(X)]^2 \tag{6}$$

Assim, derivando 6 em θ:

$$\nabla\sigma^2(\theta) = \nabla\mathbb{E}_\pi\left[\phi^2(X)\exp\left\{-\frac{1}{2}X^{\mathrm{T}}X + \frac{1}{2}(X-\theta)^{\mathrm{T}}(X-\theta)\right\}\right] - \nabla\mathbb{E}_\pi[\phi(X)]^2 =$$

$$= \nabla\mathbb{E}_\pi\left[\phi^2(X)\exp\left\{-\frac{1}{2}X^{\mathrm{T}}X + \frac{1}{2}(X-\theta)^{\mathrm{T}}(X-\theta)\right\}\right] + 0 =$$

$$\mathbb{E}_\pi\left[\phi^2(X)(\theta - X)\exp\left\{-\frac{1}{2}X^TX + \frac{1}{2}(X-\theta)^T(X-\theta)\right\}\right] \tag{7}$$

Agora, vamos derivar 7 para obter a derivada segunda de $\sigma^2(\theta)$. Como $\theta \in \mathbb{R}^d$, temos que tomar as derivadas componente a componente, i.e., calcular $\mathbb{H}_{i,j}(\theta) = \partial\sigma^2(\theta_i)/\partial\theta_j,\ \forall\ i,j \in \{1,\dots,d\}$ onde, por sua vez, $\mathbb{H}_{i,j}$ denota o elemento da i-ésima linha e j-ésima coluna da matriz Hessiana de σ^2. Sendo assim, temos:

(1) $\forall\ i \in \{1,\dots,d\},\ \mathbb{H}_{i,i}(\theta) = \mathbb{E}_\pi\left[\phi^2(X)\exp\left\{-\frac{1}{2}X^TX + \frac{1}{2}(X-\theta)^T(X-\theta)\right\}\right]$

(2) $\forall\ i,j \in \{1,\dots,d\}, i \neq j,\ \mathbb{H}_{i,j}(\theta) = \mathbb{E}_\pi\left[\phi^2(X)(\theta_i - X)(\theta_j - X)^T\exp\left\{-\frac{1}{2}X^TX + \frac{1}{2}(X-\theta)^T(X-\theta)\right\}\right]$

Não obstante, podemos escrever a matriz Hessiana como:

$$\mathbb{H}(\theta) = \mathbb{I}_d \cdot \mathbb{E}_\pi\left[\phi^2(X)\exp\left\{-\frac{1}{2}X^TX + \frac{1}{2}(X-\theta)^T(X-\theta)\right\}\right] +$$
$$+ \mathbb{E}_\pi\left[\phi^2(X)(\theta_i - X)(\theta_j - X)^T\exp\left\{-\frac{1}{2}X^TX + \frac{1}{2}(X-\theta)^T(X-\theta)\right\}\right]$$

É fácil perceber que essa matriz é positiva semi-definida (todos os termos são estritamente positivos, pois são exponenciais ou quadrados). Isto é, se tomarmos um vetor não nulo u arbitrário tal que $u \in \mathbb{R}^d$, temos que $u^T \cdot \mathbb{H} \cdot u > 0$. O que prova que $\sigma^2(\theta)$ é estritamente convexa.

4.

No item anterior, mostramos que $\sigma^2(\theta)$ é estritamente convexa. A partir dessa informação, resta apenas encontrar o ponto crítico de $\sigma^2(\theta)$, i.e., encontrar um ponto θ^* tal que $\nabla\sigma^2(\theta^*) = 0$. Isto posto, usando o resultado do item 3,

$$\nabla\sigma^2(\theta) = \mathbb{E}_\pi\left[\phi^2(X)(\theta - X)\exp\left\{-\frac{1}{2}X^TX + \frac{1}{2}(X-\theta)^T(X-\theta)\right\}\right] =$$
$$= \mathbb{E}_\pi\left[\phi^2(X)(\theta - X)\exp\left\{-\frac{1}{2}X^TX + \frac{1}{2}(X^TX - X^T\theta - \theta^TX + \theta^T\theta)\right\}\right] =$$
$$= \mathbb{E}_\pi\left[\phi^2(X)(\theta - X)\exp\left\{\frac{1}{2}\theta^T\theta - X^T\theta\right\}\right]$$

Rearranjando os termos, encontramos:

$$\nabla\sigma^2(\theta) = \exp\left\{\frac{1}{2}\theta^T\theta\right\}\mathbb{E}_\pi\left[\phi^2(X)(\theta - X)\exp\left\{-X^T\theta\right\}\right] \tag{8}$$

Como $\exp\{\theta^T\theta\} > 0,\ \forall\ \theta$, vale que:

$$\nabla\sigma^2(\theta^*) = 0 \iff \mathbb{E}_\pi\left[\phi^2(X)(\theta^* - X)\exp\left\{-X^T\theta^*\right\}\right] = 0.$$

Temos, então, exatamente o resultado que desejamos.

5.

Dos item anterior, sabemos que

$$\nabla\sigma^2(\theta) = \exp\left\{\frac{1}{2}\theta^T\theta\right\}\mathbb{E}_\pi\left[\phi^2(X)(\theta - X)\exp\left\{-X^T\theta\right\}\right]$$

Convenientemente, vamos definir $\phi(x)$ da seguinte forma:

$$\phi(x) = \max\{0, \lambda \exp\{\sigma X\} - K\} \tag{9}$$

Podemos, ainda, reescrever $\nabla\sigma^2(\theta)$ para obter:

$$\nabla\sigma^2(\theta) = \exp\left\{\frac{1}{2}\theta^T\theta\right\}\mathbb{E}_\pi\left[\theta\ \phi^2(X)\exp\left\{-\theta^T X\right\}\right] - \exp\left\{\frac{1}{2}\theta^T\theta\right\}\mathbb{E}_\pi\left[X\ \phi^2(X)\exp\left\{-\theta^T X\right\}\right] \tag{10}$$

Se tomarmos os casos:

(a) $\theta \leq 0 \;\Rightarrow\; \theta \cdot \mathbb{E}_\pi\left[\theta\ \phi^2(X)\exp\left\{-\theta^T X\right\}\right] \leq 0$ e $\mathbb{E}_\pi\left[X\ \phi^2(X)\exp\left\{-\theta^T X\right\}\right] > 0 \Rightarrow \nabla\sigma^2(\theta) < 0$.

(b) Se $\theta \in \left(0,\ \sigma^{-1}\log(K/\lambda)\right) \;\Rightarrow\; X \geq \theta \;\Rightarrow\; \nabla\sigma^2(\theta) < 0$

Logo, para os dois casos, vemos que $\sigma^2(\theta)$ é decrescente.

Exercise 2 (Metropolis-Hastings)

1.

(1) Amostramos X de $q(X_{t-1})$, i.e., $X \sim q(X_{t-1})$.

(2) Amostramos $U \sim U[0,1]$.

(3) Se $U \leq \alpha(X_{t-1}, X)$ então tomamos $X_t \coloneqq X$. Caso contrário, tomamos $X_{t-1} \coloneqq X_t$.

2.

Ora,

$$\alpha(y,x) = \frac{\gamma(y,x)}{\pi(y)q(y,x)} \implies \alpha(y,x)\pi(y)q(y,x) = \gamma(y,x) \tag{11}$$

e

$$\alpha(x,y) = \frac{\gamma(x,y)}{\pi(x)q(x,y)} \implies \alpha(x,y)\pi(x)q(x,y) = \gamma(x,y) \tag{12}$$

Combinando 11 e 12, dado $\gamma(x,y) = \gamma(y,x)$, temos:

$$\alpha(x,y)\pi(x)q(x,y) = \alpha(y,x)\pi(y)q(y,x) \tag{13}$$

Exatamente como queríamos demonstrar.

3.

Tomando a probabilidade de aceitação do MCMC,

$$\alpha(x,y) = \min\left\{1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right\} \iff \frac{\gamma(x,y)}{\pi(x)q(x,y)} = \min\left\{1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right\} \iff$$
$$\iff \gamma(x,y) = \pi(x)q(x,y)\ \min\left\{1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right\} \iff \gamma(x,y) = \min\left\{\pi(x)q(x,y), \pi(y)q(y,x)\right\} \tag{14}$$

Com isso, é fácil ver que, ao aplicar o algoritmo MCMC, estamos, nesse caso, escolhendo valores convenientes para $\gamma(x,y)$.

4.

Temos $X^{\tau_k} = Y^t$, onde Y^t denota todas as propostas aceitas em um dado tempo t. No entanto, nem todo valor de Y é aceito. Logo, podemos denotar o tempo entre propostas aceitas por:

$$\tau_{k+1} - \tau_k \tag{15}$$

Assim, é fácil ver que

$$\frac{1}{\tau_k - 1}\sum_{j=1}^{k-1}(\tau_{j+1} - \tau_j)\cdot\phi(Y^j). \tag{16}$$

Como o *Kernel* é a probabilidade de, dado que estamos em um estado t, passarmos para $t+1$, basta encontrarmos a probabilidade de aceitarmos um novo valor para Y, isto é, sairmos de Y^t para Y^{t+1}. Ora, isso equivale a aceitar um valor proveniente da distribuição $q(x,y)$. Sendo assim,

$$P(Y^{t+1} \text{ aceito} \mid Y^t \text{ aceito}) = \alpha(x,y)\cdot q(x,y) \tag{17}$$

Note que, não necessariamente, essa probabilidade integra 1. Para garantir isso, podemos normalizar os valores de 17 obtidos.

5.

Ora,

$$\tilde{\pi}(x)K(x,y) = \frac{\pi(x)m(x)}{\sum_{z\in X}\pi(z)m(z)}\frac{\alpha(x,y)q(x,y)}{m(x)} = \frac{\pi(x)\alpha(x,y)q(x,y)}{\sum_{z\in X}\pi(z)m(z)} \tag{18}$$

Se aplicarmos o resultado do item 3, obtemos:

$$\tilde{\pi}(x)K(x,y) = \frac{\pi(x)\alpha(x,y)q(x,y)}{\sum_{z\in X}\pi(z)m(z)} = \frac{\pi(y)\alpha(y,x)q(y,x)}{\sum_{z\in X}\pi(z)m(z)} \tag{19}$$

Multiplicando e dividindo por $m(y)$, chegamos a:

$$\tilde{\pi}(x)K(x,y) = \frac{\pi(x)\alpha(x,y)q(x,y)}{\sum_{z\in X}\pi(z)m(z)}\frac{m(y)}{m(y)} = \frac{\pi(y)\alpha(y,x)q(y,x)}{\sum_{z\in X}\pi(z)m(z)}\frac{m(y)}{m(y)} = \tilde{\pi}(y)K(y,x) \tag{20}$$

e, portanto, K é reversível em $\tilde{\pi}$.

6.

(Não consegui fazer)

Exercise 3 (Metropolis-Hastings)

1.

Ora,

$$\begin{aligned}\int_{\mathbb{X}} \pi(x)K(x,y)dx &= \int_{\mathbb{X}} \pi(x)\left[\alpha(x)q(y) + (1-\alpha(x))\delta_x(y)\right]\, dx = \\ &= \int_{\mathbb{X}} \pi(x)\alpha(x)q(y)\, dx + \int_{\mathbb{X}} (1-\alpha(x))\delta_x(y)]\, dx \propto \\ &\propto q(y)\int_{\mathbb{X}} \frac{q(x)}{\alpha(x)}\alpha(x)\, dx + \int_{\mathbb{X}} (1-\alpha(y))\frac{q(y)}{\alpha(y)}\, dx \propto \\ &\propto q(y) + (1-\alpha(y))\frac{q(y)}{\alpha(y)} \propto \pi(y)\end{aligned}$$

□

2.

Já sabemos que $\pi(x) \propto \frac{q(x)}{\alpha(x)}$. Assumindo que $0 \le \alpha(x) = \alpha < 1$, temos que

$$\pi(x) \propto \frac{q(x)}{\alpha(x)} \propto q(x). \tag{21}$$

Adicionalmente, vamos assumir que $\mathbb{V}_q(X_1) = \sigma^2 < \infty$. Para calcularmos a variância assintótica, vamos primeiro determinar $\mathbb{C}ov(X_1, X_k)$. Assim,

$$\mathbb{C}ov(X_1, X_k) = \mathbb{E}[X_1X_k] - \mathbb{E}[X_1]\mathbb{E}[X_k] \tag{22}$$

No regime estacionário, temos que $\mathbb{E}[X_1] = \mathbb{E}[X_k]$. Assim, prosseguindo com os cálculos em 22, temos:

$$\mathbb{C}ov(X_1, X_k) = \mathbb{E}[X_1X_k \mid X_1, X_{k-1}] - \mathbb{E}[X_1]^2 \tag{23}$$

Note que $P(X_k = X_{k-1}) = 1-\alpha$ e $P(X_k \sim q) = \alpha$. Com isso, podemos destacar que X_k é independente de X_1 e de X_{k-1}. Portanto, podemos reescrever a esperança condicional em 23 da seguinte forma:

$$\begin{aligned}\mathbb{C}ov(X_1, X_k) &= \mathbb{E}[X_1X_k \mid X_1, X_{k-1}] - \mathbb{E}[X_1]^2 = (1-\alpha)\mathbb{E}[X_1X_{k-1}] + \alpha\mathbb{E}[X_1]\mathbb{E}[X_k] - \mathbb{E}[X_1]^2 = \\ &= (1-\alpha)\mathbb{E}[X_1X_{k-1}] + \alpha\mathbb{E}[X_1]^2 - \mathbb{E}[X_1]^2 = (1-\alpha)\mathbb{E}[X_1X_{k-1}] - \left[(1-\alpha)\mathbb{E}[X_1]^2\right] = \\ &= (1-\alpha)\ \mathbb{C}ov(X_1, X_{k-1}) = (1-\alpha)^{k-1}\sigma^2 \end{aligned}\tag{24}$$

Onde a última igualdade vem do fato que estamos suponto $\mathbb{V}_q(X_1) = \sigma^2$, e resolvendo recursivamente a equação para a covariância.

Agora estamos prontos para expressar a variância assintótica como:

$$\begin{aligned}\sigma_X^2 &= \mathbb{V}_q(X_1) + 2\sum_{k=2}^{\infty}\mathbb{C}ov(X_1, X_k) = \sigma^2 + 2\sigma^2\sum_{k=2}^{\infty}(1-\alpha)^{k-1} = \sigma^2\left[1 + \sigma^2\sum_{k=2}^{\infty}(1-\alpha)^{k-1}\right] = \\ &\sigma^2\left[1 + \sigma^2\sum_{k=1}^{\infty}(1-\alpha)^{k}\right] = \sigma^2\left[\frac{2-\alpha}{\alpha}\right]\end{aligned}\tag{25}$$

Assim fica fácil ver que

$$\lim_{\alpha\to 0}\sigma_X^2 = \lim_{\alpha\to 0}\sigma^2\left[\frac{2-\alpha}{\alpha}\right] = \infty. \tag{26}$$

Exercise 4 (Gibbs Sampler)

1.

Sabemos, trivialmente, que

$$\pi(x \mid y) = \frac{\pi(x,y)}{\pi(y)} = \frac{\pi(x,y)}{\int \pi(x,y)\ dx} \propto \pi(x,y) \propto \exp\left\{-\frac{1}{2}(x-1)^2(y-2)^2\right\} = $$

$$= \exp\left\{\frac{-1}{2}\frac{(x-1)^2}{(y-2)^{-2}}\right\} \tag{27}$$

Analogamente,

$$\pi(y \mid x) = \frac{\pi(y,x)}{\pi(x)} = \frac{\pi(x,y)}{\int \pi(x,y)\ dy} \propto \pi(x,y) \propto \exp\left\{-\frac{1}{2}(x-1)^2(y-2)^2\right\} = $$

$$= \exp\left\{\frac{-1}{2}\frac{(y-2)^2}{(x-1)^{-2}}\right\} \tag{28}$$

2.

Não. O Gibbs Sampler não faz sentido nesse caso, pois

$$\int\int \pi(x,y)\ dxdy = \int\int \exp\left\{\frac{-1}{2}\frac{(x-1)^2}{(y-2)^{-2}}\right\}\ dxdy = \int \sqrt{2\pi(y-2)^{-2}}\ dy = +\infty, \tag{29}$$

isto é, $\pi(x,y)$ não é uma função de densidade de probabilidade pois não integra 1.

Exercise 5 (Gibbs Sampler)

1.

Considere duas v.a.'s independentes $X_i \sim \mathcal{B}inomial(m_i, \theta_1)$ e $Y_i \sim \mathcal{B}inomial(n_i, \theta_2)$, i.e.,

$$P(X_i = x_i) = \binom{m_i}{x_i}\theta_1^{x_i}(1-\theta_1)^{m_i - x_i}$$

e

$$P(Y_i = y_i) = \binom{n_i}{y_i}\theta_2^{y_i}(1-\theta_2)^{n_i - y_i}$$

Considerando, agora, uma terceira v.a. dada por $Z_i = X_i + Y_i$, temos

$$P(Z_i = z_i \mid \theta_1, \theta_2) = P(X_i + Y_i = z_i \mid \theta_1, \theta_2) = \sum_{x_i=0}^{z_i} P(X_i = x_i \mid \theta_1)P(Y_i = z_i - x_i \mid \theta_2) =$$

$$= \sum_{x_i=0}^{z_i}\binom{m_i}{x_i}\theta_1^{x_i}(1-\theta_1)^{m_i-x_i}\binom{n_i}{z_i - x_i}\theta_2^{z_i-x_i}(1-\theta_2)^{n_i - z_i + x_i} \tag{30}$$

que é, de fato, a função verossimilhança para uma observação de Z_i. Tomando uma amostra *iid* $\{z_1, \ldots, z_T\}$, temos:

$$P(z_1, \dots, z_T \mid \theta_1, \theta_2) = \prod_{i=1}^{T} \left[\sum_{x_i=0}^{z_i} \binom{m_i}{x_i} \theta_1^{x_i} (1-\theta_1)^{m_i - x_i} \binom{n_i}{z_i - x_i} \theta_2^{z_i - x_i} (1-\theta_2)^{n_i - z_i + x_i} \right] \tag{31}$$

2.

Assuma, inicialmente, $\vartheta_1, \vartheta_2 \sim \mathcal{U}_{[0,1]}$ e note que

$$\begin{aligned} P(\vartheta_1 = \theta_1 \mid X_1 = x_1, \dots, X_T = x_T, Y_1 = y_1, \dots, Y_T = y_t, Z_1 = z_1, \dots Z_T = z_T) \\ = P(\vartheta_1 = \theta_1 \mid X_1 = x_1, \dots, X_T = x_T) \end{aligned} \tag{32}$$

Pois $X_i \perp\!\!\!\perp Y_i, \ \forall \ i = 1, \dots, T$; e, por sua vez,

$$P(\vartheta_1 = \theta_1 \mid X_1 = x_1, \dots, X_T = x_T) \propto P(X_1 = x_1, \dots, X_T = x_T \mid \theta_1) \cdot P(\theta_1) \tag{33}$$

Pela independência dos X_i's, temos:

$$\begin{aligned} P(\vartheta_1 = \theta_1 \mid X_1 = x_1, \dots, X_T = x_T) = P(X_1 = x_1 \mid \theta_1) \dots P(X_T = x_T \mid \theta_1) \cdot P(\theta_1) \propto \\ \propto P(x_1 \mid \theta_1) \dots P(x_T \mid \theta_1), \end{aligned} \tag{34}$$

Dado que estamos assumindo $\vartheta_1 \sim \mathcal{U}_{[0,1]}$. Como cada um dos $X_i \sim \mathcal{B}inomial(m_i, \theta_1)$, podemos reescrever 34 da seguinte forma:

$$P(\theta_1 \mid x_1, \dots, x_T) \propto P(x_1 \mid \theta_1) \dots P(x_T \mid \theta_1) = \binom{m_i}{x_1} \theta_1^{x_1} (1-\theta_1)^{m_i - x_1} \dots \binom{m_i}{x_T} \theta_1^{x_T} (1-\theta_1)^{m_i - x_T}$$

Como estamos olhando para a distribuição dos θ's (os x's são realizações e, portanto, conhecidos), podemos escrever:

$$P(\theta_1 \mid x_1, \dots, x_T) \ \propto \ \prod_{j=1}^{T} \binom{m_i}{x_j} \theta_1^{x_j} (1-\theta_1)^{m_i - x_j} \ \propto \ \prod_{j=1}^{T} \theta_1^{x_j} (1-\theta_1)^{m_i - x_j} \tag{35}$$

Assim, podemos concluir que $\vartheta_1 \mid x_1, \dots, x_T \sim \mathcal{B}eta\left(1 + \sum_{j=1}^{T} x_j, \ 1 + \sum_{j=1}^{T} (m_i - x_i)\right)$.

Por analogia, temos que $\vartheta_2 \mid x_1, \dots, x_T \sim \mathcal{B}eta\left(1 + \sum_{j=1}^{T} y_j, \ 1 + \sum_{j=1}^{T} (n_i - y_i)\right)$.

Com isso, definimos completamente as probabilidades condicionais para ϑ_1 e ϑ_2. Contudo, para aplicar o Gibbs Sampler, ainda precisamos ser capazes de amostrar de $P(x_i \mid y_1, \dots, y_T, z_1, \dots, z_T, \theta_1, \theta_2)$ e $P(y_i \mid X_1, \dots, x_T, z_1, \dots, z_T, \theta_1, \theta_2)$. Como $X_i \perp\!\!\!\perp Y_i$, isso equivale a amostrar de

$$P(X_1 = x_1, \dots, X_T = x_T, Y_1 = y_1, \dots, Y_T = y_T \mid Z_1 = z_1, \dots, Z_T = z_T, \vartheta_1 = \theta_1, \vartheta_2 = \theta_2)$$

Note que, saber $Z_i = z_i$ não nos diz muito sobre X_i e Y_i individualmente (saber a soma de X e Y não nos diz muito a respeito dos seus valores) além de limitar seu suporte. Ou seja,

$$P(x_1, \ldots, x_T, y_1, \ldots, y_T \mid z_1, \ldots, z_T, \theta_1, \theta_2) = \prod_{i=1}^{T} P(x_i, y_i \mid z_i, \theta_1, \theta_2) \ \propto$$

$$\propto \prod_{i=1}^{T} \left[\binom{m_i}{x_i} \theta_1^{x_i} (1-\theta_1)^{m_i - x_i} \binom{n_i}{z_i - x_i} \theta_2^{z_i - x_i} (1-\theta_2)^{n_i - z_i + x_i} \right] \mathbb{1}_{\{x_i + y_i = z_i\}} \tag{36}$$

Esse passo, na prática, constitui um problema para a aplicação direta do *Gibbs Sampling* a esse problema pois não conseguimos amostrar diretamente da distribuição apresentada em 36. Para contornar esse problema, podemos propor um *Rejection Sampling*. Apesar de não sabermos amostrar dessa distribuição, sabemos como avaliá-la para realizações $(X_i = x_i, Y_i = y_i \mid Z_i = z_i, \theta_1, \theta_2)$.

Para esse caso, o algoritmo de *Rejection Sampling* funcionaria da seguinte forma:

(1) Dado um valor de Z, digamos $Z = k$, execute os passos a seguir.

(2) Amostre θ_1 e $\theta_2 \sim \mathcal{U}_{[0,1]}$.

(3) Com os valores de θ_1 e θ_2, amostre $X \sim \mathcal{B}inomial(m, \theta_1)$ e $Y \sim \mathcal{B}inomial(n, \theta_2)$.

(4) Se $X + Y \neq k$, rejeitamos X e Y e repetimos o passo (3). Caso contrário, aceitamos X e Y.

(5) Uma vez aceito o par (X, Y), atualizamos θ_1 e θ_2 em 36 e repetimos o processo a partir do passo (3).

Simulation question (Normal mixture model – Gibbs sampling)

1.

Considere um vetor aleatório $Z = \{Z_1, \ldots, Z_N\}$ que assume valores em $\{1, \ldots, K\}^N$ tal que

$$\mathbb{P}(Z = z \mid p) = \prod_{i=1}^{N} \mathbb{P}(Z_i = z_i \mid p) = \prod_{i=1}^{N} p_{z_i}$$

e considere o vetor $X = \{X_1, \ldots, X_N\}$ tal que

$$X_i \mid Z_i = z_i, p, \mu, \sigma^2 \sim \mathcal{N}\left(\mu_{z_i}, \sigma^2_{z_i}\right), \ \forall\ i \in \{1, \ldots, N\}.$$

Primeiramente, vamos determinar a distribuição conjunta de X_i e Z_i, dados os valores dos parâmetros, i.e., encontrar $f\left(x_i, z_i \mid p, \mu, \sigma^2\right)$. Note que:

$$f\left(x_i, z_i \mid p, \mu, \sigma^2\right) = P\left(Z_i = z_i \mid p, \mu, \sigma^2\right) \cdot f\left(x_i \mid Z_i = z_i, p, \mu, \sigma^2\right) \tag{37}$$

considerando $X_i \perp\!\!\!\perp Z_i$ e atentando para o fato de que X_i é uma v.a. contínua, enquanto Z_i é uma v.a. discreta. Como estamos particularmente interessados na distribuição de X_i, vamos integrar 37 com relação a Z_i. Como Z_i é discreta, isso equivale a tomar o somatório em todo o suporte de Z_i. Assim,

$$f\left(x_i \mid p, \mu, \sigma^2\right) = \sum_{j=1}^{K} P\left(Z_i = j \mid p, \mu, \sigma^2\right) f\left(x_i \mid Z_i = z_i, p, \mu, \sigma^2\right) \tag{38}$$

Conforme o enunciado, se fizermos $P(Z_i = k \mid p) = p_k$, temos:

$$f\left(x_i \mid p, \mu, \sigma^2\right) = \sum_{j=1}^{K} p_j \phi\left(x_i; \mu, \sigma^2\right) \tag{39}$$

Onde ϕ é a PDF de uma v.a. Gaussiana. De 39, concluímos que o vetor aleatório X também segue um modelo de misturas de Gaussianas.

2.

Ora,

$$P\left(Z = z \mid x, p, \mu, \sigma^2\right) \propto P\left(Z = z \mid p, \mu, \sigma^2\right) \cdot f\left(x \mid Z = z, p, \mu, \sigma^2\right) \propto$$
$$\propto \prod_{i=1}^{N} P(Z_i = z_i \mid p) \cdot f\left(x_i \mid Z_i = z_i, p, \mu, \sigma^2\right) \propto \prod_{i=1}^{N} p_{z_i} \phi\left(x_i; \mu_{z_i}, \sigma^2_{z_i}\right) \tag{40}$$

Portanto,

$$P\left(Z = z \mid x, p, \mu, \sigma^2\right) \propto p_z \frac{1}{\sqrt{2\pi\sigma_z^2}} \exp\left\{-\frac{1}{2\sigma_z^2}(x_i - \mu_z)^2\right\} \tag{41}$$

Podemos, então, amostrar diretamente da distribuição em 41 para cada Z_i, $i \in \{1, \ldots, N\}$ e, depois, normalizar essas probabilidades (para que a densidade, de fato, integre 1).

3.

Analogamente ao item 2, sabemos que

$$\pi\left(p \mid x, Z = z, \mu, \sigma^2\right) \propto \pi(p) f\left(x, z \mid p, \mu, \sigma^2\right)$$

Logo,

$$\pi\left(p \mid x, z, \mu, \sigma^2\right) \propto \pi(p) \prod_{i=1}^{N} f\left(x_i, z_i \mid p, \mu, \sigma^2\right) \propto \pi(p) \prod_{i=1}^{N} P(Z_i = z_i) \cdot f\left(x_i \mid Z_i = z_i, p, \mu, \sigma^2\right) \propto$$

$$\propto \pi(p) \prod_{i=1}^{N} p_{z_i} \phi\left(x_i; \mu_{z_i}, \sigma_{z_i}^2\right) \propto \pi(p) \prod_{j=1}^{K} p_j^{(\#I)} \tag{42}$$

onde $I = \{i : z_i = j\}$. Como $\pi(p) \propto \prod_{i=1}^{N} p_i^{\gamma_k - 1}$, podemos reescrever 42 como:

$$\pi\left(p \mid x, z, \mu, \sigma^2\right) \propto \prod_{j=1}^{K} p_j^{\gamma_j - 1 + (\#I)} \tag{43}$$

Amostrar dessa distribuição é equivalente a amostrar de uma gamma com parâmetros γ_j e 1, i.e., podemos amostrar v.a.'s $V_j \sim \mathcal{G}amma(\gamma_j, 1)$ (note que os demais termos da PDF Gamma serão eliminados pela constante de proporcionalidade $\propto$).

4.

Como já fizemos nos outros itens, vamos aplicar a fórmula de Bayes

$$\pi\left(\mu \mid x, z, p, \sigma^2\right) \propto \pi(\mu) f\left(x, z \mid p, \mu, \sigma^2\right).$$

Assim, se considerarmos $\pi(\mu) = \mathcal{N}(m, \tau^2)$ (conforme mencionado no item 2), temos

$$\pi\left(\mu \mid x, z, p, \sigma^2\right) \propto \prod_{j=1}^{K} \exp\left\{-\frac{1}{2\tau^2}(\mu_j - m)^2\right\} \prod_{i=1}^{N} p_{z_i} \exp\left\{-\frac{1}{2\sigma_{z_i}^2}(x_i - \mu_{z_i})^2\right\} \tag{44}$$

Portanto, é fácil ver que

$$\pi\left(\mu_j \mid x, z, p, \sigma^2\right) \propto \prod_{j=1}^{K} \exp\left\{-\frac{1}{2\tau^2}(\mu_j - m)^2\right\} \prod_{i=1}^{N} p_{z_i} \exp\left\{-\frac{1}{2\sigma_{z_i}^2}(x_i - \mu_{z_i})^2\right\}\Bigg|_{z_i = j} \tag{45}$$

Como $\pi(\mu)$ é uma priori Gaussiana, sabemos que a posteriori de μ será $\mathcal{N}(\eta, \zeta^2)$ tal que:

$$\eta = \left.\frac{m\tau^{-2} + \sum_{i=1}^{N} x_i \sigma_{z_i}^2}{\tau^{-2} + \sigma_{z_i}(\#I)}\right|_{z_i = j} \quad \text{e } \zeta^2 = \tau^{-2} + (\#I)\sigma_{z_i}^{-2}\big|_{z_i = j}$$

onde $(\#I)$ denota o número de índices para os quais $z_i = j$, i.e., a cardinalidade do conjunto $I = \{i : z_i = j\}$. Assim,

$$\pi\left(\mu_j \mid x, z, p, \sigma^2\right) \propto \exp\left\{-\frac{1}{2}\left(\tau^{-2} + (\#I)\sigma_j^{-2}\right) \cdot \left(\mu_j - \frac{m\tau^{-2} + \sum_{i=1}^{N} x_i \sigma_j^2}{\tau^{-2} + \sigma_j(\#I)}\right)^2\right\}. \tag{46}$$

Trivialmente, podemos amostrar diretamente de $\mathcal{N}(\eta, \zeta^2)$.

5.

Do item 2, temos a priori:

$$\pi(\sigma_k^2) \propto \left(\sigma_k^2\right)^{-\alpha - 1} \exp\left\{-\beta \sigma_k^{-2}\right\} \tag{47}$$

E, da fórmula de Bayes,

$$\pi\left(\sigma^2 \mid x, z, p, \mu\right) \propto \pi\left(\sigma^2\right) \cdot f\left(x, z \mid p, \mu, \sigma^2\right) \tag{48}$$

Substituindo os devidos termos em 48, temos:

$$\pi\left(\sigma^2 \mid x, z, p, \mu\right) \propto \prod_{j=1}^{K}\left(\sigma_j^2\right)^{-\alpha-1} \exp\left\{-\beta\sigma_j^{-2}\right\} \cdot \prod_{i=1}^{N} \frac{p_{z_i}}{\sqrt{2\pi\sigma_{z_i}^2}} \exp\left\{-\frac{1}{2\sigma_{z_i}^2}(x_i-\mu_{z_i})^2\right\} \tag{49}$$

Assim, para cada $j \in \{1, \ldots, N\}$, temos:

$$\begin{aligned}
\pi\left(\sigma_j^2 \mid x, z, p, \mu\right) &\propto \prod_{j=1}^{K}\left(\sigma_j^2\right)^{-\alpha-1} \exp\left\{-\beta\sigma_j^{-2}\right\} \cdot \prod_{i=1}^{N} \frac{p_{z_i}}{\sqrt{2\pi\sigma_{z_i}^2}} \exp\left\{-\frac{1}{2\sigma_{z_i}^2}(x_i-\mu_{z_i})^2\right\}\Bigg|_{z_i=j} \propto \\
&\propto \prod_{j=1}^{K}\left(\sigma_j^2\right)^{-\alpha-1} \exp\left\{-\beta\sigma_j^{-2}\right\} \cdot \prod_{i=1}^{N} \frac{p_j}{\sqrt{2\pi\sigma_j^2}} \exp\left\{-\frac{1}{2\sigma_j^2}(x_i-\mu_j)^2\right\} \propto \\
&\propto \prod_{j=1}^{K}\left(\sigma_j^2\right)^{-\alpha-1-(\#I)/2} \exp\left\{-\beta - \sum_{i=1}^{N}\frac{1}{2\sigma_j^2}(x_i-\mu_j)^2\right\} = \\
&= \prod_{j=1}^{K}\left(\sigma_j^2\right)^{-(\alpha+(\#I)/2)-1} \exp\left\{-\left(\beta + \frac{1}{2}\sum_{i=1}^{N}(x_i-\mu_j)^2\right)\frac{1}{\sigma_j^2}\right\}
\end{aligned} \tag{50}$$

Assim, podemos dizer que $\pi\left(\sigma_j^2 \mid x, z, p, \mu\right) \sim \mathcal{I}nverse\mathcal{G}amma\left(\alpha + (\#I)/2, \beta + \frac{1}{2}\sum_{i=1}^{N}(x_i-\mu_j)^2\right)$. Como o próprio nome da distribuição sugere, podemos amostrar $U \sim \mathcal{G}amma$ com os mesmos parâmetros e, em seguida, tomar o inverso $1/U$.

6.

Estamos interessados em amostrar $(y_1, \ldots, y_N)$ de uma mistura de Gaussianas, com K (número de distribuições) fixo. Isto é, queremos formar uma amostra com base no seguinte modelo:

$$y_i = f(x_i) = \sum_{j=1}^{K} p_j \phi\left(x_i; \mu_j, \sigma_j^2\right), \ i = 1, \ldots, N. \tag{51}$$

A seguir, é apresentado um gráfico para uma mistura de 4 Gaussianas, com parâmetros:

- $p = (0.25, 0.25, 0.25, 0.25)$
- $\mu = (-6, -2, 2, 6)$
- $\sigma^2 = (1, 1, 1, 1)$
- $N = 10000$

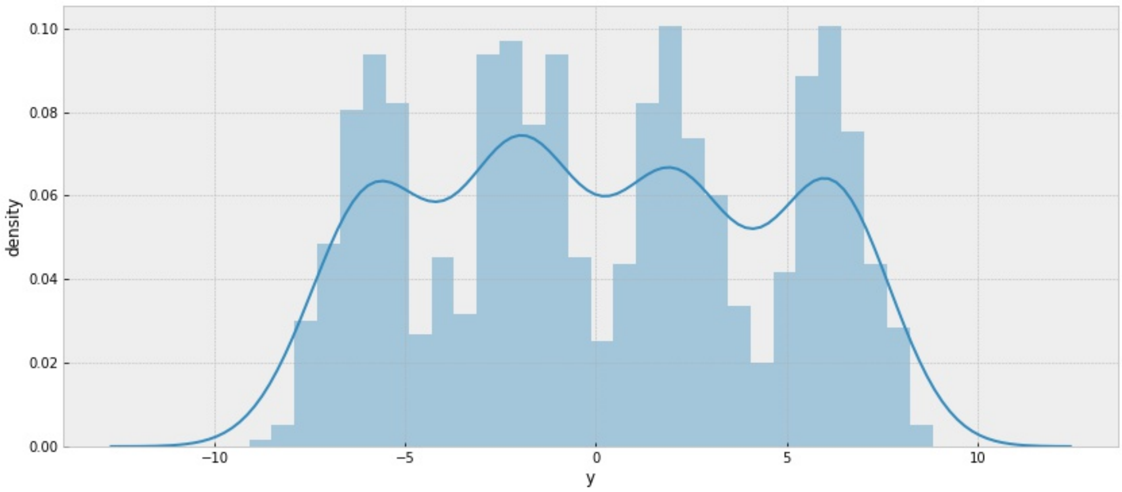

Figura 1: Mistura de 4 gaussianas com parâmetros $p = (0.25, 0.25, 0.25, 0.25)$, $\mu = (-6, -2, 2, 6)$ e $\sigma^2 = (1, 1, 1, 1)$.

7.

Dados os valores de $Y = (y_1, \ldots, y_N)$ do item anterior, estamos agora interessados em amostrar da distribuição posteriori. Para isso, vamos tomar como base as seguintes prioris, com os respectivos valores apontados para seus hiperparâmetros:

- Priori de p: $\pi(p) \propto \prod_{k=1}^{K} p_k^{\gamma_k - 1}$, com $\gamma_k = 1$, $\forall\ k \in \{1, \ldots, K\}$.
- Priori de μ_k: $\pi(\mu_k) \propto \exp\left\{-\frac{1}{2\tau^2}(\mu_k - m)^2\right\}$, com $\mu_k = 0$ e $\tau^2 = 10$, isto é, $\mu_k \sim \mathcal{N}(0, 10)$.
- Priori de σ_k^2: $\pi(\sigma_k^2) = \left(\sigma_k^2\right)^{-\alpha-1} \exp\left\{-\beta\sigma_k^{-2}\right\}$, com $\alpha = 1$ e $\beta = 0.1$.

O valor desses hiperparâmetros foi escolhido de maneira tal que tenhamos prioris não informativas em cada caso.

8.

Começamos com valores iniciais para os parâmetros de interesse, com base nas distribuições *priori* e, a cada passo, são atualizadas as distribuições condicionais de Z, p, μ e σ^2.

Para cada parâmetro p, μ e σ, são apresentados os gráficos das distribuições *priori*, *posteriori* e da cadeia de Markov construída por meio do algoritmo de *Gibbs Sampling*. Vemos que, em cada caso, as distribuições posterioris têm centro de massa concentrados em torno dos valores verdadeiros do parâmetro, definidos no item (6).

- Parâmetro p verdadeiro: $(0.25, 0.25, 0.25, 0.25)$. Os resultados da simulação são apresentados a seguir.

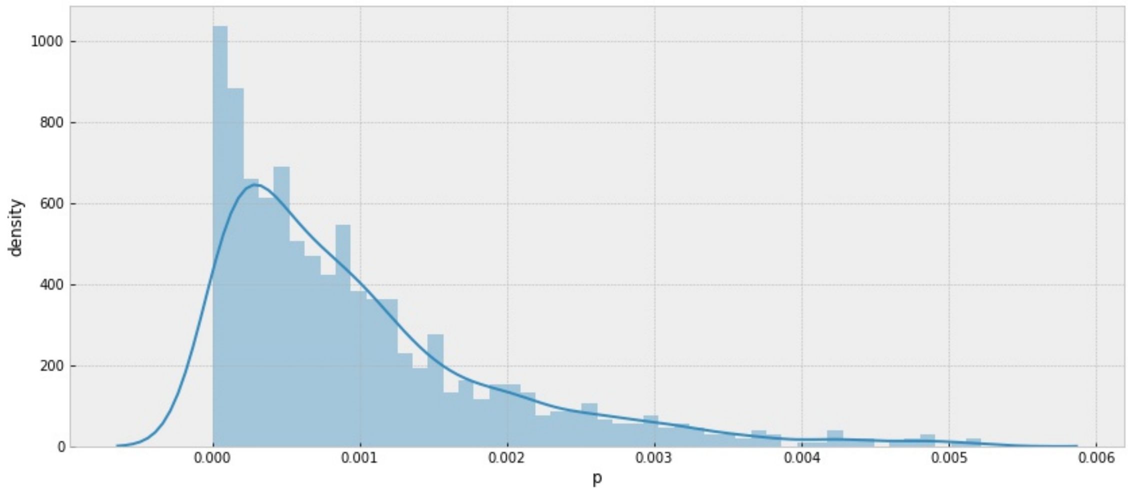

Figura 2: Priori de p, com $\gamma = (1, 1, 1, 1)$

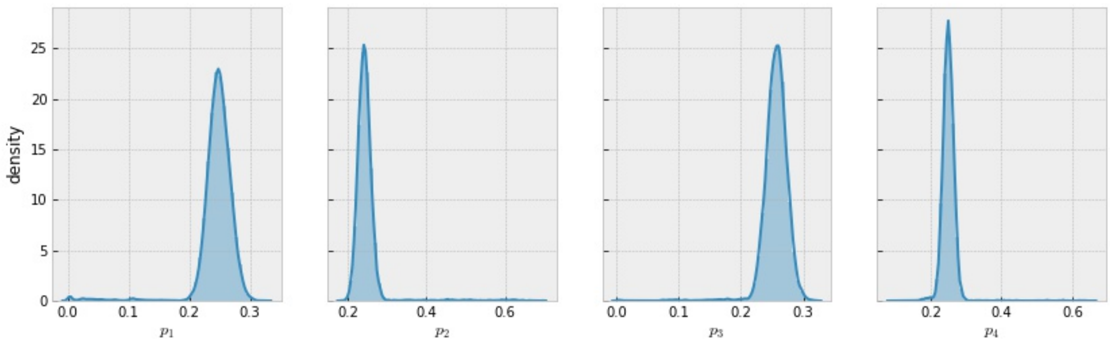

Figura 3: Posteriori de p

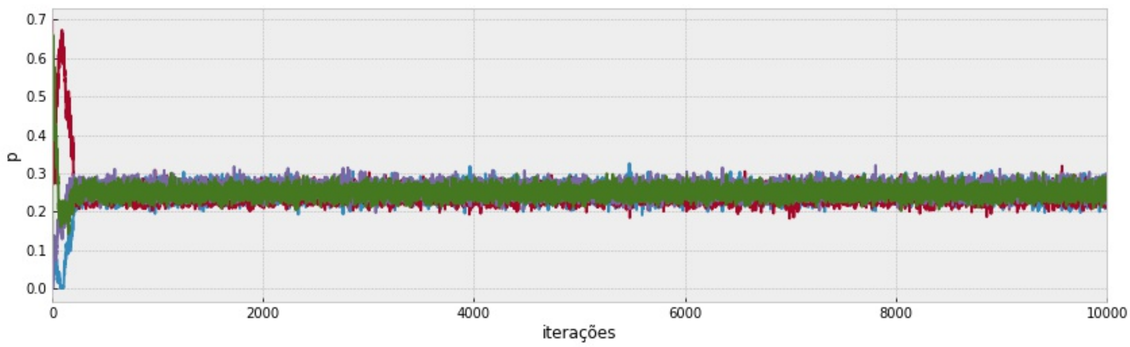

Figura 4: Cadeia de p

- Parâmetros μ verdadeiros: $(-6, -2, 2, 6)$. De forma análoga, os resultados da simulação são apresentados a seguir:

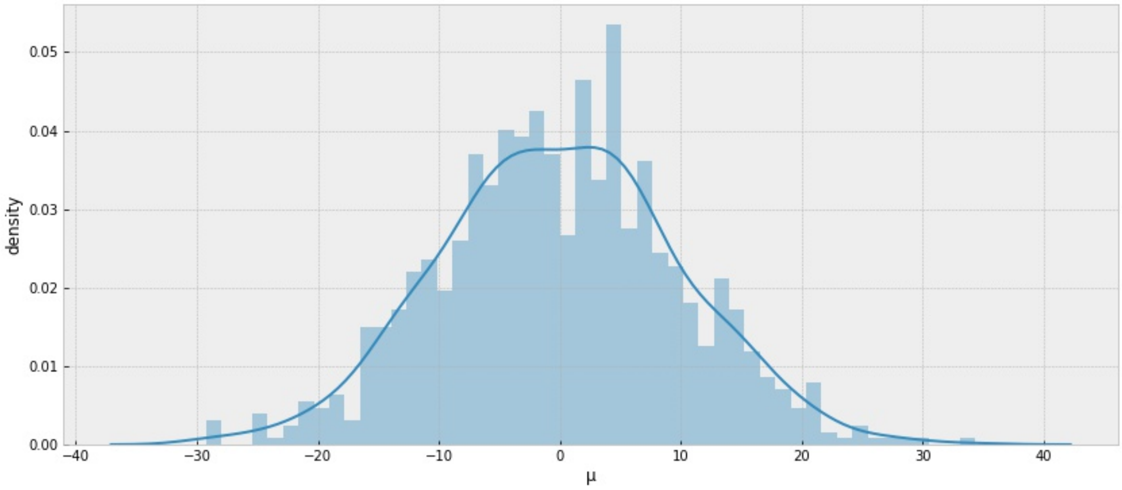

Figura 5: Priori de μ

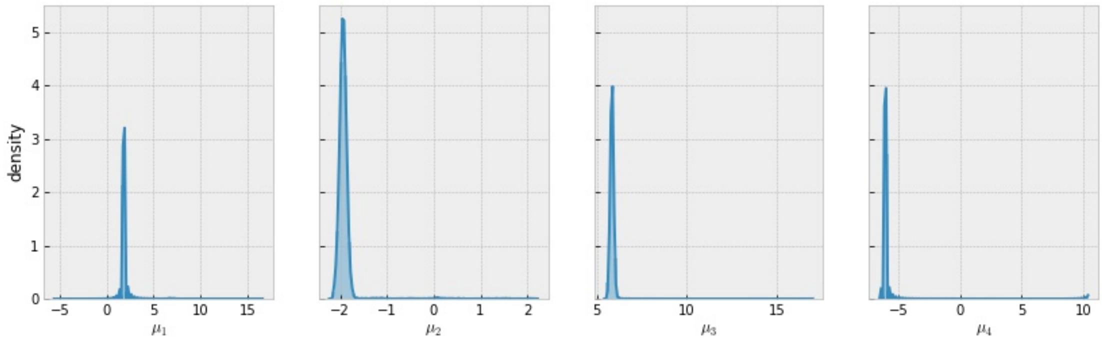

Figura 6: Posteriori de μ, com $(m = 0, \tau = 10)$

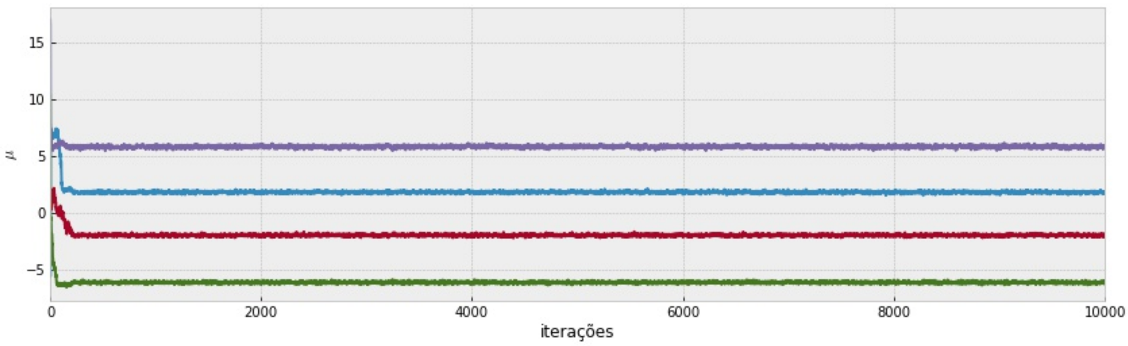

Figura 7: Cadeia de μ

- Por fim, os parâmetros σ^2 verdadeiros são $(1, 1, 1, 1)$ e os resultados da simulação foram:

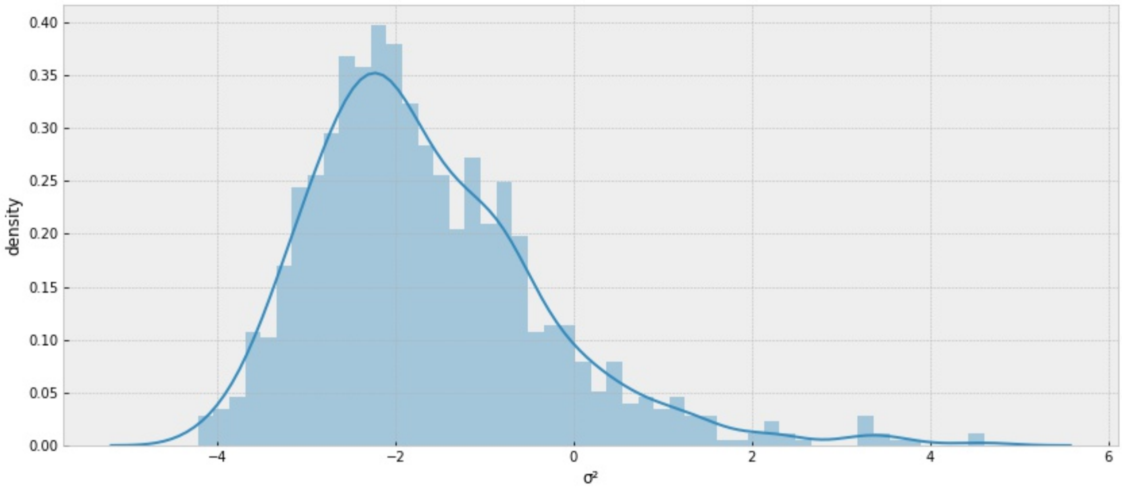

Figura 8: Priori de σ^2, com $(\alpha = 1, \beta = 0.1)$

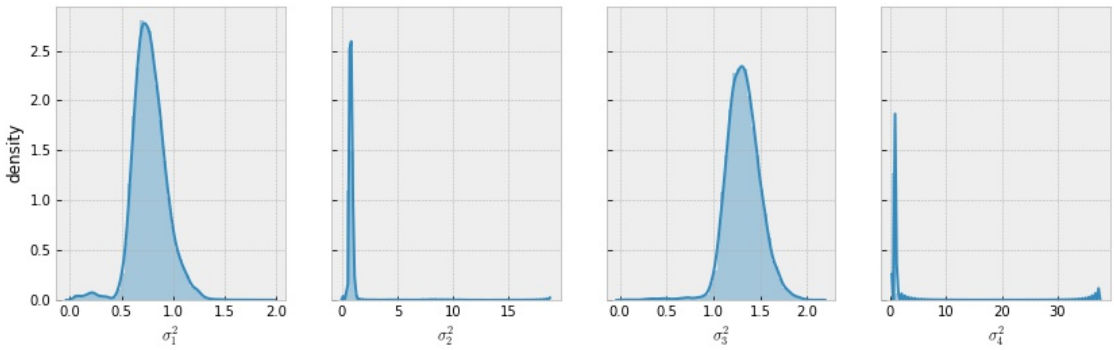

Figura 9: Posteriori de σ^2

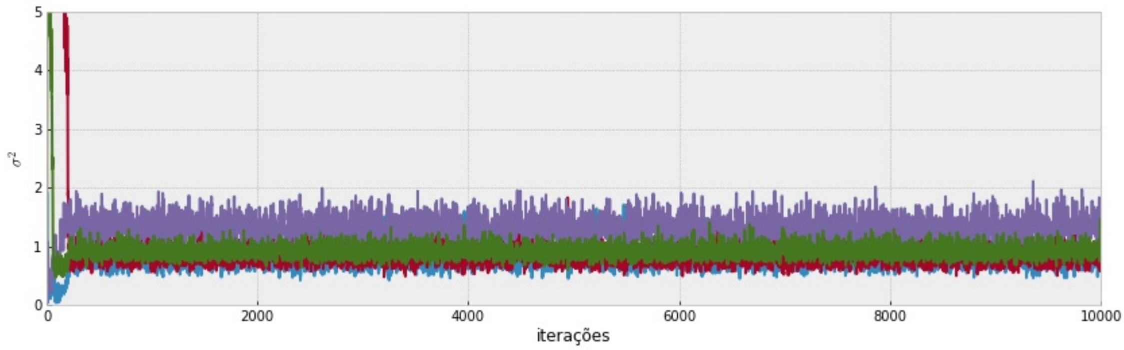

Figura 10: Cadeia de σ^2