Disciplina: Estatística Computacional **(Entrega: 11/11/19)**

Lista de Exercícios 3

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

Exercise 1 (Gibbs Sampler)

1.

Segundo o enunciado, temos $K^S_{X,Y}\left((x,y),(x',y')\right)$ como a matriz de transição (kernel) tomando como ponto de partida o par (x,y) e ponto de chegada (x',y') em S passos. Isto é, $K^S_{X,Y}\left((x,y),(x',y')\right)$ denota a probabilidade de chegarmos em (x',y') a partir de (x,y), tomando apenas S passos.

Considerando que, no *Gibbs Sampler* sistemático, amostramos primeiro de $\pi_{Y|X}(\cdot \mid X^{(t-1)}$ e depois de $\pi_{X|Y}(\cdot \mid Y^{(t)}$, temos que

$$K^S_{X,Y}\left((x,y),(x',y')\right) = P\left((x',y') \mid (x,y), S\right) = \pi_{Y|X}(y' \mid x) \cdot \pi_{X|Y}(x' \mid y') \tag{1}$$

Agora, suponha, por contradição, que $K^S_{X,Y}$ é reversível em relação a $\pi_{X,Y}$. Sob esse pressuposto, temos que:

$$\pi_{X,Y}(x,y)K^S_{X,Y}\left((x,y),(x',y')\right) = \pi_{X,Y}(x',y')K^S_{X,Y}\left((x',y'),(x,y)\right) \tag{2}$$

No entanto, note que, se reescrevermos 2 e usarmos a informação em 1,

$$\frac{\pi_{X,Y}(x,y) \cdot K^S_{X,Y}\left((x,y),(x',y')\right)}{\pi_{X,Y}(x',y') \cdot K^S_{X,Y}\left((x',y'),(x,y)\right)} = \frac{\pi_{X,Y}(x,y) \cdot \pi_{Y|X}(y' \mid x) \cdot \pi_{X|Y}(x' \mid y')}{\pi_{X,Y}(x',y') \cdot \pi_{Y|X}(y \mid x') \cdot \pi_{X|Y}(x \mid y)} \tag{3}$$

Como

$$\pi_{X|Y}(x \mid y) = \frac{\pi_{X,Y}(x,y)}{\pi_Y(y)} \quad \text{e} \quad \pi_{X|Y}(x' \mid y') = \frac{\pi_{X,Y}(x',y')}{\pi_Y(y')},$$

temos que

$$\frac{\pi_{X,Y}(x,y) \cdot K^S_{X,Y}\left((x,y),(x',y')\right)}{\pi_{X,Y}(x',y') \cdot K^S_{X,Y}\left((x',y'),(x,y)\right)} = \frac{\pi_Y(y) \cdot \pi_{Y|X}(y' \mid x)}{\pi_Y(y') \cdot \pi_{Y|X}(y \mid x')}. \tag{4}$$

A razão em 4, não necessariamente, é igual a 1 $\forall x, y \in \mathbb{X}$, onde $\mathbb{X}$ é o espaço de estados (para ver isso, note que, se $y \neq y'$, então $\pi_Y(y) \neq \pi_Y(y')$). Com isso, concluímos que $K^S_{X,Y}$ não é reversível (i.e., é irreversível) com relação a $\pi_{X,Y}$. $\square$

2.

Do item 1 equação 1, sabemos que $K^S_{X,Y}\left((x,y),(x',y')\right) = \pi_{Y|X}(y' \mid x) \cdot \pi_{X|Y}(x' \mid y')$. Partindo dessa relação e integrando ambos os lados com relação a y', obtemos:

$$\int K_{X,Y}\left((x,y),(x',y')\right)\, dy' = \int \pi_{Y|X}(y' \mid x) \cdot \pi_{X|Y}(x' \mid y')\, dy' \tag{5}$$

Analisando o lado direito, vemos que 5 independe do ponto de partida y e do ponto de chegada y'. Assim, temos que $\left(X^{(t)}\right)_{t\geq 1}$ é um processo de Markov, cujo Kernel é caracterizado como:

$$K_X(x, x') = \int \pi_{Y|X}(y' \mid x) \cdot \pi_{X|Y}(x' \mid y') \, dy' \tag{6}$$

Ainda resta mostrar que $\left(X^{(t)}\right)_{t\geq 1}$ é π_X-reversível, i.e., que $\pi_X(x)K_X(x,x') = \pi_X(x')K(x',x)$. De maneira esperta, vamos multiplicar ambos os lados por $\pi_X(x) > 0$. Assim,

$$\pi_X(x) \cdot K_X(x, x') = \pi_X(x) \cdot \int \pi_{Y|X}(y' \mid x) \cdot \pi_{X|Y}(x' \mid y') \, dy' \tag{7}$$

Ainda, pela regra de Bayes, podemos escrever:

$$\pi_X(x) \cdot K_X(x, x') = \pi_X(x) \cdot \int \frac{\pi_{X|Y}(x \mid y') \cdot \pi_Y(y')}{\pi_X(x)} \cdot \frac{\pi_{Y|X}(y' \mid x') \cdot \pi_X(x')}{\pi_Y(y')} \, dy' =$$
$$\pi_X(x') \cdot \int \pi_{X|Y}(x \mid y') \cdot \cdot \pi_{Y|X}(y' \mid x') \, dy' = \pi_X(x') \cdot K_X(x', x) \tag{8}$$

Logo, $\left(X^{(t)}\right)_{t\geq 1}$ é π_X-reversível, como queríamos demonstrar. □

3.

Como no *Random Scan Gibbs Sampler* temos 1/2 de chance de amostrarmos de $\pi_{Y|X}\left(\cdot \mid X^{(t-1)}\right)$ ou de $\pi_{X|Y}\left(\cdot \mid Y^{(t-1)}\right)$, e também setamos $X^{(t)} = X^{(t-1)}$ e $Y^{(t)} = Y^{(t-1)}$ respectivamente, temos:

$$K^R_{X,Y}\left((x,y),(x',y')\right) = \frac{1}{2}\pi_{Y|X}(y' \mid x) \; \delta_X(x') + \frac{1}{2}\pi_{X|Y}(x' \mid y) \; \delta_Y(y'), \tag{9}$$

onde $\delta_X(x')$ é a função massa de *Dirac*[1] aplicada no ponto x' e a mesma coisa vale para $\delta_Y(y')$, isto é,

$$\delta_X(x') = \begin{cases} 1, & \text{se } x = x' \\ 0, & \text{caso contrário} \end{cases} \quad \text{e} \quad \delta_Y(y') = \begin{cases} 1, & \text{se } y = y' \\ 0, & \text{caso contrário} \end{cases}$$

Agora, vamos mostrar que $K^R_{X,Y}$ é $\pi_{X,Y}$-reversível. Para isso, vamos multiplicar ambos os lados de 9 por $\pi_{X,Y}(x,y)$. Assim,

$$\pi_{X,Y}(x,y) \; K^R_{X,Y}\left((x,y),(x',y')\right) = \pi_{X,Y}(x,y)\left[\frac{1}{2}\pi_{Y|X}(y' \mid x) \; \delta_X(x') + \frac{1}{2}\pi_{X|Y}(x' \mid y) \; \delta_Y(y')\right] \tag{10}$$

Agora, perceba que

$$\begin{aligned} \pi_{X,Y}(x,y) \; \pi_{Y|X}(y' \mid x) \; \delta_X(x') &= \pi_{X,Y}(x',y) \; \pi_{Y|X}(y' \mid x) \; \delta_X(x') \\ &= \pi_X(x') \; \pi_{Y|X}(y \mid x') \; \pi_{Y|X}(y' \mid x) \; \delta_X(x') \\ &= \pi_{X,Y}(x',y') \; \pi_{Y|X}(y \mid x') \; \delta_X(x') \end{aligned} \tag{11}$$

e, de maneira análoga,

$$\begin{aligned} \pi_{X,Y}(x,y) \; \pi_{X|Y}(x' \mid y) \; \delta_Y(y') &= \pi_{X,Y}(x,y') \; \pi_{X|Y}(x' \mid y) \; \delta_Y(y') \\ &= \pi_Y(y') \; \pi_{X|Y}(x \mid y') \; \pi_{X|Y}(x' \mid y) \; \delta_Y(y') \\ &= \pi_{X,Y}(x',y') \; \pi_{X|Y}(x \mid y') \; \delta_Y(y') \end{aligned} \tag{12}$$

Substituindo 11 e 12 em 10,

[1]tomada com base nas notas de aula

$$\begin{aligned}
\pi_{X,Y}(x,y)\,K^R_{X,Y}\big((x,y),(x',y')\big) &= \frac{1}{2}\,\pi_{X,Y}(x',y')\,\pi_{Y|X}(y\mid x')\,\delta_X(x') + \frac{1}{2}\,\pi_{X,Y}(x',y')\,\pi_{X|Y}(x\mid y')\,\delta_Y(y') = \\
&= \pi_{X,Y}(x',y')\left[\frac{1}{2}\pi_{Y|X}(y\mid x')\;\delta_X(x') + \frac{1}{2}\pi_{X|Y}(x\mid y')\;\delta_Y(y')\right] \\
&= \pi_{X,Y}(x',y')K^R_{X,Y}\big((x',y'),(x,y)\big)\,, \qquad (13)
\end{aligned}$$

exatamente como queríamos demonstrar. □

Exercise 2 (Metropolis-within-Gibbs)

1.

Nesse caso, vamos calcular a probabilidade α de aceitação nesse algoritmo.

$$\alpha\left(X_1 \mid X_1^{(t-1)}, X_2^{(t-1)}\right) = \min\left\{1, \frac{\pi\left(X_1, X_2^{(t-1)}\right) q\left(X_1^{(t-1)} \mid X_1, X_2^{(t-1)}\right)}{\pi\left(X_1^{(t-1)}, X_2^{(t-1)}\right) q\left(X_1 \mid X_1^{(t-1)}, X_2^{(t-1)}\right)}\right\} \tag{14}$$

Se tomarmos $q\left(x_1' \mid x1, x2\right) = \pi_{X_1|X_2}(x_1' \mid x_2)$ em 14, temos:

$$\begin{aligned}\alpha\left(X_1 \mid X_1^{(t-1)}, X_2^{(t-1)}\right) &= \min\left\{1, \frac{\pi\left(X_1, X_2^{(t-1)}\right) \pi_{X_1|X_2}\left(X_1^{(t-1)} \mid X_1, X_2^{(t-1)}\right)}{\pi\left(X_1^{(t-1)}, X_2^{(t-1)}\right) \pi_{X_1|X_2}\left(X_1 \mid X_1^{(t-1)}, X_2^{(t-1)}\right)}\right\} = \\ &= \min\left\{1, \frac{\pi\left(X_1, X_2^{(t-1)}\right) \pi_{X_1|X_2}\left(X_1^{(t-1)} \mid X_2^{(t-1)}\right)}{\pi\left(X_1^{(t-1)}, X_2^{(t-1)}\right) \pi_{X_1|X_2}\left(X_1 \mid X_2^{(t-1)}\right)}\right\} \end{aligned} \tag{15}$$

Pela regra de Bayes, podemos reescrever 15 como:

$$\begin{aligned}\alpha\left(X_1 \mid X_1^{(t-1)}, X_2^{(t-1)}\right) &= \min\left\{1, \frac{\pi_{X_2}\left(X_2^{(t-1)}\right) \pi_{X_1|X_2}\left(X_1 \mid X_2^{(t-1)}\right) \pi_{X_1|X_2}\left(X_1^{(t-1)} \mid X_2^{(t-1)}\right)}{\pi_{X_2}\left(X_2^{(t-1)}\right) \pi_{X_1|X_2}\left(X_1^{(t-1)} \mid X_2^{(t-1)}\right) \pi_{X_1|X_2}\left(X_1 \mid X_2^{(t-1)}\right)}\right\} = \\ &= \min\{1, 1\} = 1. \end{aligned} \tag{16}$$

2.

O kernel, nesse caso (tomando como base as notas de aula), é dado por:

$$K\left((x_1, x_2), (x_1', x_2')\right) = \left[\alpha\left(x_1' \mid x_1, x_2\right) q\left(x_1' \mid x_1, x_2\right) + \left(1 - I(x_1, x_2)\right) \delta_{x_1}(x_1')\right] \pi_{X_2|X_1}\left(x_2' \mid x_1'\right), \tag{17}$$

com $I(x_1, x_2) = \int \alpha(y \mid x_1, x_2) \cdot q(y \mid x_1, x_2)\ dy$.

Logo,

$$\begin{aligned} &\int\int \pi(x_1, x_2) K\left((x_1, x_2), (x_1', x_2')\right)\ dx_1 dx_2 = \\ &\int\int \pi_{X_2}(x_2)\pi_{X_1|X_2}(x_1 \mid x_2)\{\alpha(x_1' \mid x_1, x_2) q(x_1' \mid x_1, x_2) + \left(1 - I(x_1, x_2)\right) \delta_{x_1}(x_1')\}\pi_{X_2|X_1}\left(x_2' \mid x_1'\right)\ dx_1 dx_2 = \\ &= \int \pi_{X_2}(x_2)\Bigg[\int \pi_{X_1|X_2}(x_1 \mid x_2)\{\alpha(x_1' \mid x_1, x_2) q(x_1' \mid x_1, x_2) + \\ &+\ \left(1 - I(x_1, x_2)\right) \delta_{x_1}(x_1')\}\ dx_1\Bigg] \pi_{X_2|X_1}\left(x_2' \mid x_1'\right)\ dx_2 = \\ &\int \pi_{X_2}(x_2) \pi_{X_1|X_2}\left(x_1' \mid x_2\right) \pi_{X_2|X_1}(x_2' \mid x_1')\ dx_2 = \\ &\int \pi\left(x_1', x_2\right) \pi_{X_2|X_1}\left(x_2' \mid x_1'\right)\ dx_2 = \\ &\pi_{X_1}\left(x_1'\right) \pi_{X_2|X_1}\left(x_2' \mid x_1'\right) = \pi\left(x_1', x_2'\right). \end{aligned} \tag{18}$$

$\square$

Exercise 3 (Metropolis-Hastings and Gibbs Sampler)

1.

Precisamos garantir que $\pi(x)T(x,y) = \pi(y)T(y,x)$, com $x \neq y$. Ou seja, isso equivale a mostrar que

$$\gamma(x,y) = \alpha(x,y)\ \pi(x)\ q(x,y) = \alpha(y,x)\ \pi(y)\ q(y,x) = \gamma(y,x) \tag{19}$$

Ora, mas essa relação é trivialmente satisfeita. Sendo assim, basta notar que 19 implica que

$$T(x,y) = \alpha(x,y)q(x,y) + \left(1 - \sum \alpha(x,z)q(x,z)\right)\delta_x(y) \Rightarrow$$
$$T(x,y) = \frac{\gamma(x,y)}{\pi(x)} + \left(1 - \sum \frac{\gamma(x,z)}{\pi(x)}\right)\delta_x(y). \tag{20}$$

Analogamente,

$$T(y,x) = \frac{\gamma(y,x)}{\pi(y)} + \left(1 - \sum \frac{\gamma(y,z)}{\pi(y)}\right)\delta_y(x). \tag{21}$$

Agora, é fácil ver que

$$\pi(x)T(x,y) = \gamma(x,y) + \left(1 - \sum \gamma(x,z)\right)\delta_x(y) \tag{22}$$

e

$$\pi(y)T(y,x) = \gamma(y,x) + \left(1 - \sum \gamma(y,z)\right)\delta_y(x) \tag{23}$$

Como estamos interessados no caso em que $x \neq y$[2], temos que $\delta_y(x) = \delta_x(y) = 0$. Logo, temos que (22) = (23), i.e.,

$$\pi(x)T(x,y) = \pi(y)T(y,x), \text{ com } x \neq y \tag{24}$$

E, portanto, T é π-reversível. □

2.

A probabilidade de aceitação para o algoritmo de Metropolis-Hastings é dada por:

$$\alpha(x,y) = \min\left\{1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right\} \tag{25}$$

Se multiplicarmos ambos os lados por $\pi(x)q(x,y)$, obtemos:

$$\pi(x)q(x,y)\alpha(x,y) = \pi(x)q(x,y)\min\left\{1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right\} = \min\left\{\pi(x)q(x,y), \pi(y)q(y,x)\right\} \tag{26}$$

Logo,

$$\gamma(x,y) = \min\left\{\pi(x)q(x,y), \pi(y)q(y,x)\right\} \tag{27}$$

Para o algoritmo de Baker, vamos partir da relação

$$\gamma(x,y) = \frac{\pi(x)q(x,y)\pi(y)q(y,x)}{\pi(x)q(x,y) + \pi(x)q(y,x)} \tag{28}$$

e, dividindo ambos os lados por $\pi(x)q(x,y)$, é trivial ver que

[2]O caso em ue $x = y$ é trivialmente satisfeito

$$\frac{\gamma(x,y)}{\pi(x)q(x,y)} = \frac{\pi(x)q(x,y)\pi(y)q(y,x)}{\pi(x)q(x,y)+\pi(y)q(y,x)} \cdot \frac{1}{\pi(x)q(x,y)} = \\ = \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)+\pi(x)q(y,x)} = \frac{1}{1+\frac{\pi(x)q(x,y)}{\pi(y)q(y,x)}}. \tag{29}$$

□

3.

Tome $x \neq y$. Vamos comparar o kernel do algoritmo de Metropolis-Hastings (T_{M-H}) com o kernel do algoritmo de Baker (T_{Baker}).

$$T_{M-H} = \alpha(x,y)q(x,y) = \frac{\gamma(x,y)}{\pi(x)} \tag{30}$$

Usando um dos resultados do item 2,

$$T_{M-H}(x,y) = \frac{\min\{\pi(x)q(x,y), \pi(y)q(y,x)\}}{\pi(x)} \tag{31}$$

Agora, para o algoritmo de Baker:

$$T_{Baker}(x,y) = \alpha(x,y)q(x,y) = \frac{\gamma(x,y)}{\pi(x)} = \frac{1}{\pi(x)}\frac{\pi(x)q(x,y)\pi(y)q(y,x)}{\pi(x)q(x,y)+\pi(y)q(y,x)} \tag{32}$$

Sem perda de generalidade, vamos supor que $\pi(x)q(x,y) \leq \pi(y)q(y,x)$.[3] Nesse caso,

$$T_{M-H}(x,y) = \frac{\pi(x)q(x,y)}{\pi(x)} = q(x,y) \tag{33}$$

e

$$T_{Baker} = \frac{q(x,y)\pi(y)q(y,x)}{\pi(x)q(x,y)+\pi(y)q(y,x)} = T_{M-H} \cdot \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)+\pi(y)q(y,x)} \tag{34}$$

Agora, denote $c = \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)+\pi(y)q(y,x)} \leq 1$. Assim, é fácil ver que

$$T_{Baker} = c \cdot T_{M-H} \leq T_{M-H} \tag{35}$$

Portanto, por 35 e pelo teorema enunciado nesse item, temos que o T_{M-H} possui menor variância assintótica. □

4.

A partir das definições dadas no enunciado, vamos determinar as matrizes de transição para os algoritmos *Random Scan Gibbs Sampler* (RSGS) e *Modified Random Scan Gibbs Sampler* (MRSGS). Primeiramente, para o *Random Scan Gibbs Sampler*, com $x_k \neq y_k$, temos:

$$T_{RSGS}(x,y) = \frac{\pi_{X_K|X_{-K}}(y_K \mid x_{-K})\ \delta_{x-K}(y_{-K})}{d} \tag{36}$$

onde $d \sim \mathcal{DU}nif_{\{1,\ldots,d\}}$. Não obstante, para o *Modified Random Scan Gibbs Sampler*, temos:

$$T_{MRSGS}(x,y) = \frac{q(y_K \mid x_{-K})}{d} \cdot \min\left\{1, \frac{1-\pi_{X_K|X_{-K}}(x_K \mid x_{-K})}{1-\pi_{X_K|X_{-K}}(y_K \mid x_{-K})}\right\} \delta_{x_{-K}}(y_{-K}) =$$

[3] Todas as conclusões a seguir são análogas para o caso contrário.

$$= \frac{\pi_{X_K|X_{-K}}(y_K \mid x_{-K})}{d\left[1 - \pi_{X_K|X_{-K}}(x_K \mid x_{-K})\right]} \cdot \min\left\{1, \frac{1 - \pi_{X_K|X_{-K}}(x_K \mid x_{-K})}{1 - \pi_{X_K|X_{-K}}(y_K \mid x_{-K})}\right\} \delta_{x_{-K}}(y_{-K}) =$$

$$= \frac{1}{d} \cdot \min\left\{\frac{1}{1 - \pi_{X_K|X_{-K}}(x_K \mid x_{-K})}, \frac{1}{1 - \pi_{X_K|X_{-K}}(x_K \mid x_{-K})}\right\} \delta_{x_{-K}}(y_{-K}) \tag{37}$$

Note que, de 36 e 37, temos:

$$\frac{T_{MRSGS}(x,y)}{T_{RSGS}(x,y)} = \min\left\{\frac{1}{1 - \pi_{X_K|X_{-K}}(x_K \mid x_{-K})}, \frac{1}{1 - \pi_{X_K|X_{-K}}(x_K \mid x_{-K})}\right\} \cdot \frac{1}{\pi_{X_K|X_{-K}}(y_K \mid x_{-K})} \Rightarrow$$

$$\Rightarrow \frac{T_{MRSGS}(x,y)}{T_{RSGS}(x,y)} \geq \frac{1}{\pi_{X_K|X_{-K}}(y_K \mid x_{-K})}$$

Logo,

$$T_{MRSGS}(x,y) \geq \frac{T_{RSGS}(x,y)}{\pi_{X_K|X_{-K}}(y_K \mid x_{-K})} \Rightarrow T_{RSGS}(x,y) \tag{38}$$

Uma vez que $0 \leq \pi_{X_K|X_{-K}}(y_K \mid x_{-K}) \leq 1$. Assim,

$$T_{MRSGS}(x,y) \geq T_{RSGS}(x,y) \tag{39}$$

Portanto, pelo teorema enunciado no item 3, o *Modified Random Scan Gibbs Sampler* apresenta a menor variância assintótica. □

Exercise 4 (Metropolis-Hastings)

$\boxed{\textbf{1.}}$

Ora, a distribuição proposta $q(x' \mid x)$ é, simplesmente, dada por:

$$q(x' \mid x) = \int g(x' \mid \theta) f(\theta \mid x)\, d\theta \tag{40}$$

Uma vez que $\vartheta \sim f\left(\theta \mid X^{(t-1)}\right)$ e estamos amostrando $X \sim g(\cdot \mid \vartheta)$. No entanto, a função $f(\cdot \mid x)$ não é conhecida e, portanto, não podemos calcular diretamente a integral em 40. □

$\boxed{\textbf{2}}$

Considere, inicialmente, um algoritmo Metropolis-Hastings tradicional com distribuição *target* dada por

$$\widetilde{\pi}(x,\theta) = \pi(x) f(\theta \mid x)$$

e distribuição proposta

$$q\left((x',\theta') \mid (x,\theta)\right) = g(x' \mid \theta) f(\theta' \mid x')$$

Assim, para mostrar que o kernel é $\widetilde{\pi}$-invariante, temos que mostrar que:

$$\widetilde{\pi}(x',\theta') q\left((x,\theta) \mid (x',\theta')\right) = \pi(x') g(x \mid \theta') \tag{41}$$

e

$$\widetilde{\pi}(x,\theta) q\left((x',\theta') \mid (x,\theta)\right) = \pi(x) g(x' \mid \theta) \tag{42}$$

Para matar dois coelhos com uma só cajadada, vamos dividir 41 por 42, e obter

$$\frac{\widetilde{\pi}(x',\theta')q\left((x,\theta)\mid(x',\theta')\right)}{\widetilde{\pi}(x,\theta)q\left((x',\theta')\mid(x,\theta)\right)}=\frac{\pi(x')f(\theta'\mid x')g(x\mid\theta')f(\theta\mid x)}{\pi(x)f(\theta\mid x)g(x'\mid\theta)f(\theta'\mid x')}=\frac{\pi(x')g(x\mid\theta')}{\pi(x)g(x'\mid\theta)} \tag{43}$$

mostrando, assim, que o kernel é $\widetilde{\pi}$-invariante.

$\boxed{\textbf{3.}}$

Se decompormos a função $\min\{U,V\}$ como

$$\min\{U,V\}=\frac{U+V}{2}-\frac{|V-U|}{2}=U+\frac{V-U-|V-U|}{2}$$

podemos, então, escrever:

$$\mathbb{E}\left[\min\{U,V\}\right]=\mathbb{E}[U]+\mathbb{E}\left[\frac{V-U-|V-U|}{2}\right]=\mathbb{E}[U]+\frac{\mathbb{E}[V-U]-\mathbb{E}[|V-U|]}{2} \tag{44}$$

De 5, podemos destacar dois possíveis casos:

(1) Se $\mathbb{E}[U]\le\mathbb{E}[V]\Rightarrow\mathbb{E}[V-U]\le\mathbb{E}[|V-U|]\Rightarrow\mathbb{E}\left[\min\{U,V\}\right]=\mathbb{E}[U]$; caso contrário,

(2) Se $\mathbb{E}[U]>\mathbb{E}[V]\Rightarrow\mathbb{E}[V-U]>\mathbb{E}[|V-U|]\Rightarrow\mathbb{E}\left[\min\{U,V\}\right]=\mathbb{E}[V]$.

Com isso, provamos o que foi pedido. $\square$

$\boxed{\textbf{4.}}$

(Não consegui fazer!)

Exercise 5 (Thinning of a Markov Chain)

$\boxed{\textbf{1.}}$

Seguindo a dica sugerida no exercício,

$$\begin{aligned}&(Y-\alpha Z)^2 \geq 0 \;\Rightarrow\; \mathbb{E}\left[(Y-\alpha Z)^2\right] \geq 0 \;\Rightarrow\; \mathbb{E}\left[Y^2 - 2\alpha YZ + (\alpha Z)^2\right] \geq 0 \;\Rightarrow\\ &\mathbb{E}\left[Y^2\right] - \mathbb{E}\left[2\alpha YZ\right] + \mathbb{E}\left[\alpha^2 Z^2\right] \geq 0 \;\Rightarrow\; \mathbb{E}\left[Y^2\right] - 2\alpha\mathbb{E}\left[YZ\right] + \alpha^2\mathbb{E}\left[Z^2\right] \geq 0\end{aligned} \tag{45}$$

Em 45 temos uma função quadrática em α. Como sabemos que essa função é não-negativa, inferimos que:

$$\Delta = \sqrt{4\alpha^2 \cdot \mathbb{E}[YZ]^2 - 4\alpha^2 \cdot \mathbb{E}\left[Z^2\right]\mathbb{E}\left[Y^2\right]} \leq 0 \iff \mathbb{E}[YZ]^2 \leq \mathbb{E}\left[Z^2\right]\mathbb{E}\left[Y^2\right] \tag{46}$$

Exatamente como queríamos encontrar. $\square$

$\boxed{\textbf{2.}}$

Ora,

$$\mathbb{C}ov[Y,Z] \leq |\mathbb{C}ov[Y,Z]| \;\Rightarrow\; \mathbb{C}ov[Y,Z] \leq \Big|\mathbb{E}\left[(Y-\mathbb{E}[Y])(Z-\mathbb{E}[Z])\right]\Big| \tag{47}$$

Aplicando a desigualdade de Cauchy-Schwarz (resultado do item 1), temos:

$$\mathbb{C}ov[Y,Z] \leq \sqrt{\mathbb{E}\Big[(Y-\mathbb{E}[Y])^2\Big] \cdot \mathbb{E}\Big[(Z-\mathbb{E}[Z])^2\Big]} = \mathbb{V}(Y) \tag{48}$$

Onde a última igualdade acontece se Z e Y têm distribuições marginais idênticas. Nesse caso, $\mathbb{V}(Y) = \mathbb{V}(Z)$ e o resultado segue. $\square$

$\boxed{\textbf{3.}}$

(Não consegui fazer!)

Simulation question (Probit model – Gibbs and M-H)

1.

Seguindo a dica dada no exercício, vamos escolher $\beta^T = [3, -1]$, $n = 60$ (pequeno) e $p = 2$. Ainda, geramos a matriz de variáveis explicativas $X_{n\times p}$ proveniente de duas distribuições normais, com médias e desvios-padrão $\mu_1 = \frac{1}{2}$, $\mu_2 = 1.5$, $\sigma_1 = \sigma_2 = 1$. Esses valores foram escolhidos para refletir uma turma de alunos onde um grupo tem média 2, e outro grupo média 5.

Assim, as observações sintéticas de Y geradas aleatoriamente foram:

$Y = 0$	$Y = 1$
32	28

Tabela 1: Amostra sintética de Y. (*seed* = 101)

2.

Continuando o exercício, vamos agora introduzir uma distribuição priori em β:

$$\pi(\beta) = \mathcal{N}(0, B)$$

onde B é uma matriz de covariâncias. Nesse exercício, tomamos

$$B = \begin{bmatrix} 2.5 & 0 \\ 0 & 2.5 \end{bmatrix}$$

Ainda, a função de verossimilhança é dada por:

$$P(Y_1 = y_1, \ldots, Y_n = y_n \mid \beta) = \prod_{i=1}^{n} \Phi\left(X_i^T\beta\right)^{y_i} \left[1 - \Phi\left(X_i^T\beta\right)\right]^{1-y_i}$$

Com essas informações, foi criada uma função que tomava como *input* um vetor β e retornava como *output* o log da distribuição posterior avaliada em β.

3.

Em resposta a esse item, foi executado o algoritmo de Metropolis-Hastings com 150 mil iterações. Abaixo, são apresentados alguns gráficos para visualizarmos a cadeia (idealmente, a cadeia exploraria regiões mais próximas dos verdadeiros valores de $\beta^T = [\beta_1, \beta_2] = [-1, 2.1]$).

A taxa de aceitação desse algoritmo foi de $15,08\%$.

Vemos que, conforme o esperado, esse algoritmo resultou em densidades mais concentradas em torno dos verdadeiros valores de $\beta_1 = 3$ e $\beta_2 = -1$ e que as cadeias convergiram para os mesmos valores.

4.

Tome $Z_i \sim \mathcal{N}(X_i^T\beta, 1)$, $\forall\ i = 1, \ldots, n$. Assim, podemos escrever a condicional

$$\pi(\beta, Z \mid Y) \propto \pi(\beta) \prod_{i=1}^{n} \left[\mathbb{1}_{\{Z_i<0\}}\mathbb{1}_{\{Y_i=0\}} + \mathbb{1}_{\{Z_i\geq 0\}}\mathbb{1}_{\{Y_i=1\}}\right] \cdot \varphi(Z_i; X_i^T\beta, 1)$$

onde $\varphi(Z_i; \cdot)$ denota a densidade Gaussiana avaliada no ponto Z_i. Como a priori de β também é normal, temos que a posteriori de β também é normal.

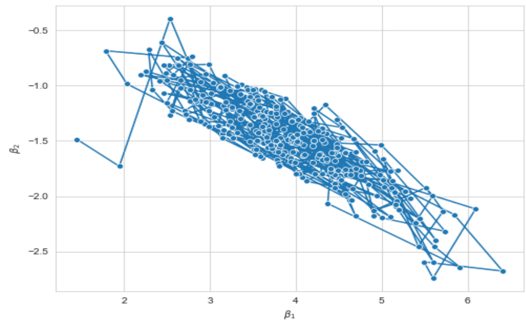

Figura 1: Cadeia

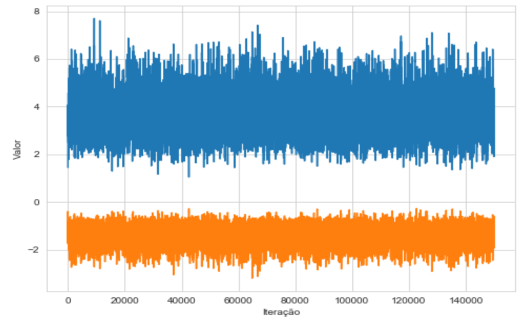

Figura 2: Evolução das cadeias

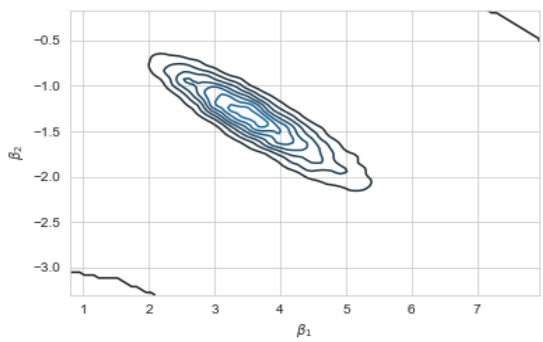

Figura 3: Densidade conjunta

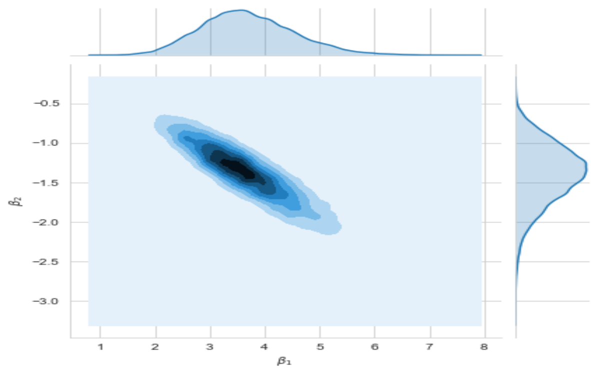

Figura 4: Densidade conjunta + marginais

$\boxed{\textbf{5.}}$

Conforme sugerido na questão, foi feito o mesmo exercício, considerando agora um algoritmo *Gibbs Sampler*, onde amostramos de $P(\beta \mid Z, Y)$ e de $P(Z \mid \beta, Y)$.

Assim como o algoritmo de *Metropolis-Hastings*, o algoritmo *Gibbs Sampler* apresentou maior concentração de massa em torno dos verdadeiros valores de β.

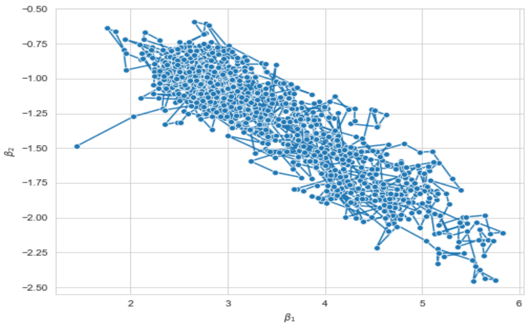

Figura 5: Cadeia

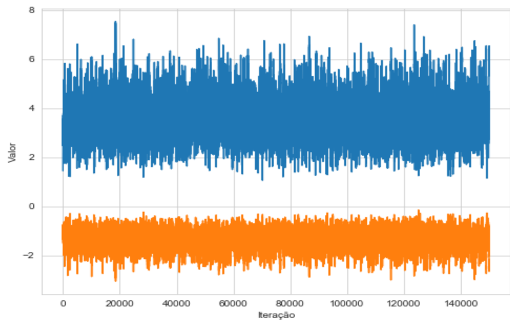

Figura 6: Evolução das cadeias

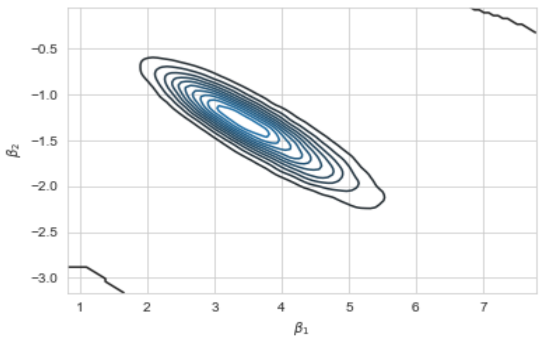

Figura 7: Densidade conjunta

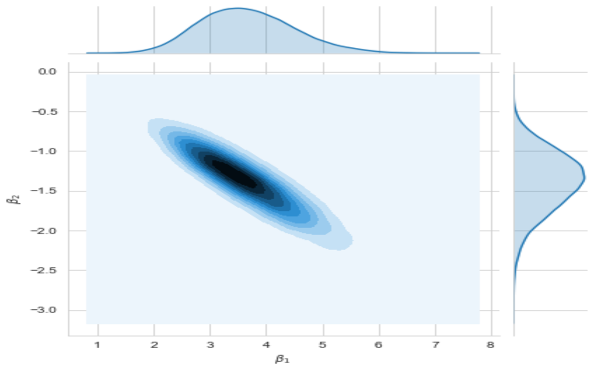

Figura 8: Densidade conjunta + marginais

6.

Agora, vamos comparar os dois algoritmos verificando a autocorrelação serial.

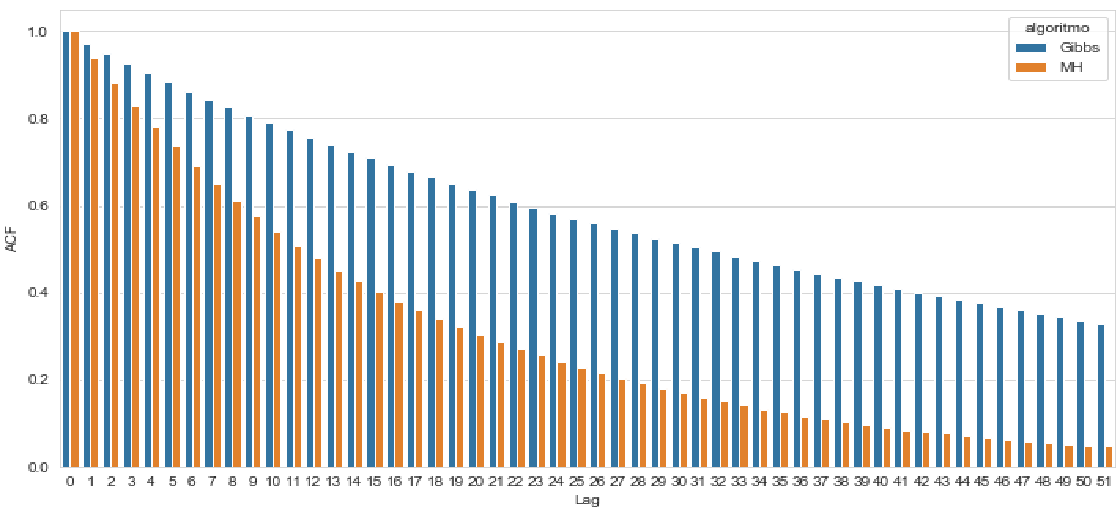

Figura 9: Autocorrelação Serial

Nesse gráfico, é fácil ver que o algoritmo de *Metropolis-Hastings* retornou uma cadeia menos autocorrelacionada, o que é uma propriedade mais desejável.