HW 6

(1) Considere uma distribuição alvo em $\mathcal{X} = \mathbb{R}^d$ de densidade $\pi(x)$. Propomos fazer uma amostragem usando um algoritmo Metropolis-Hastings baseado em uma proposta aleatória. Dado $X^{(t-1)}$, a proposta para X é amostrado de forma: $X^{(t-1)}$ é a entrada de um processo de otimização estocástica que retorna um estimador $\Theta \in \mathcal{X}$ do maximizador local de $\pi(x)$ na vizinhança de $X^{(t-1)}$. A v.a. Θ segue uma função de densidade de prob. desconhecida $f(\Theta \mid X^{(t-1)})$. Então amostramos $X \sim g(\cdot \mid \Theta)$ onde $g(\cdot \mid \Theta)$ é uma normal multivariada de media Θ e covariancia fixa Σ.

Expresse a proposta $g(x' \mid x)$ como uma função de f e g. É possível avaliar a probabilidade de aceite $\alpha_{MH}(x' \mid x) = \min\left\{1, \frac{\pi(x')\, g(x \mid x')}{\pi(x)\, g(x' \mid x)}\right\}$ associada a proposta?

Temos que

$$g(x' \mid x) = \int g(x' \mid \theta)\, f(\theta \mid x)\, d\theta .$$

A distribuição proposta não pode ser avaliada pontualmente pois $f(\theta \mid x)$ é desconhecida.

(2) Do problema anterior. Considere agora o seguinte algoritmo Metropolis-Hastings aleatório. Tome $X^{(1)} = x$, $\vartheta^{(1)} = \theta$ então itere em $t = 2, 3, \ldots$

(a) Amostre $X \sim g(\cdot \mid \vartheta^{(t-1)})$ e $\vartheta \sim f(\cdot \mid X)$

(b) Compute

$$\alpha_{RMH}((X,\vartheta) \mid (X^{(t-1)}, \vartheta^{(t-1)})) = \min\left\{1, \frac{\pi(X)\, g(X^{(t-1)} \mid \vartheta)}{\pi(X^{(t-1)})\, g(X \mid \vartheta^{(t-1)})}\right\}.$$

(c) Com probabilidade $\alpha_{RMH}((X,\theta) \mid (X^{(t-1)}, \theta^{(t-1)})$, tomamos $(X^{(t)}, \vartheta^{(t)}) = (X, \vartheta)$ e caso contrario $(X^{(t)}, \vartheta^{(t)}) = (X^{(t-1)}, \theta^{(t-1)})$

Mostre que o Kernel de transição associado com o algoritmo acima admite uma distribuição invariante $\overline{\pi}(x,\theta)$ tal que $\overline{\pi}(x) = \pi(x)$.

Como queremos que $\overline{\pi}(x,\theta)$ seja de tal forma que $\overline{\pi}(x) = \pi(x)$, basta supor

$$\Rightarrow \overline{\pi}(x,\vartheta) = \pi(x)\, f(\vartheta \mid x)$$

e

$$g((x',\theta') \mid (x,\vartheta)) = g(x' \mid \theta)\, f(\theta' \mid x').$$

Assim, $$\frac{\overline{\pi}(x,\theta)\, K(x',\theta' \mid x,\theta)}{\overline{\pi}(x',\theta')\, K(x,\theta \mid x',\theta')} = \frac{\pi(x)\, \cancel{f(\theta \mid x)}\, g(x' \mid \theta)\, \cancel{f(\theta' \mid x')} \min\left\{1, \frac{\pi(x')\, g(x \mid \theta)}{\pi(x)\, g(x' \mid \theta')}\right\}}{\pi(x')\, \cancel{f(\theta' \mid x')}\, g(x \mid \theta')\, \cancel{f(\theta \mid x)} \min\left\{1, \frac{\pi(x)\, g(x' \mid \theta)}{\pi(x')\, g(x \mid \theta')}\right\}}$$

usando que $\min(1, a) = a \min(1, a^{-1})$

$$\Rightarrow = \frac{\pi(x)\, g(x' \mid \theta)}{\pi(x')\, g(x \mid \theta')} \cdot \frac{\pi(x')\, g(x \mid \theta')}{\pi(x)\, g(x' \mid \theta)}$$

$$= 1.$$

Logo é invariante.

③ Mostre que para qualquer v.a. U e V temos que

$$E[\min\{U,V\}] \le \min\{E[U], E[V]\}$$

Temos que

$$\min\{U,V\} \le U \text{ e } \min\{U,V\} \le V.$$

Daí, tomando o valor esperado

$$E[\min\{U,V\}] \le E[U] \text{ e } E[\min\{U,V\}] \le E[V]$$

Logo, $E[\min\{U,V\}] \le \min\{E[U], E[V]\}$.

④ Considerando os problemas anteriores. Prove que $\alpha_{RMH}\{(X,\theta)|(X^{(t-1)},\theta^{(t-1)})\}$ tem valor esperado estacionário menor que o valor esperado de $\alpha_{MH}\{X|X^{(t-1)}\}$, o algoritmo Metropolis-Hasting padrão.

~~A~~ O valor esperado da probabilidade de aceitação estacionária é dado por

$$\iiiint \pi(x,\theta)\, g((x',\theta')|(x,\theta))\, \min\left\{1, \frac{\pi(x',\theta')\, g((x,\theta)|(x',\theta'))}{\pi(x,\theta)\, g((x',\theta')|(x,\theta))}\right\} dx\, d\theta\, dx'\, d\theta'$$

$$\overset{(2)}{=} \iiiint \pi(x) f(\theta|x)\, g(x'|\theta) f(\theta'|x) \min\left\{1, \frac{\pi(x') f(\theta'|x')\, g(x|\theta')\, f(\theta|x)}{\pi(x) f(\theta|x)\, g(x'|\theta) f(\theta'|x)}\right\} dx\, d\theta\, dx'\, d\theta'$$

$$= \iiiint \pi(x) \min\left\{f(\theta|x)\, g(x'|\theta) f(\theta'|x), \frac{\pi(x') f(\theta'|x') g(x|\theta') f(\theta|x)}{\pi(x)}\right\} dx\, d\theta\, dx'\, d\theta'$$

$$\overset{(3)}{\leq} \iint \pi(x) \min\left\{\iint f(\theta|x) g(x'|\theta) f(\theta'|x)\, d\theta\, d\theta', \frac{\pi(x') \iint f(\theta'|x') g(x|\theta') f(\theta|x)\, d\theta\, d\theta'}{\pi(x)}\right\} dx\, dx'$$

$$= \iint \pi(x) \min\left\{g(x'|x), \frac{\pi(x')}{\pi(x)} g(x|x')\right\} dx\, dx'$$

$$= \iint \pi(x) \min\left\{1, \frac{\pi(x')\, g(x|x')}{\pi(x)\, g(x'|x)}\right\} dx\, dx'$$

tilibra

5) Prove a desigualdade de Cauchy-Schwarz, para quaisquer variáveis aleatórias reais X e Z vale que

$$(\mathbb{E}[XZ])^2 \leq \mathbb{E}[X^2]\mathbb{E}[Z^2].$$

Temos que

$$\mathbb{E}[(X+\alpha Z)^2] \geq 0 \quad \forall \alpha \in \mathbb{R}$$

$$\Leftrightarrow \quad \mathbb{E}[X^2] - 2\alpha \mathbb{E}[XZ] + \alpha^2 \mathbb{E}[Z^2] \geq 0 \quad \forall \alpha \in \mathbb{R}$$

$$\Leftrightarrow (\mathbb{E}[Z^2])\alpha^2 + (-2\mathbb{E}[XZ])\alpha + \mathbb{E}[X^2] \geq 0$$

Resolvendo em α temos

$$(-2\mathbb{E}[XZ])^2 - 4\mathbb{E}[Z^2]\mathbb{E}[X^2] \leq 0$$

$$\Rightarrow \quad (\mathbb{E}[XZ])^2 \leq \mathbb{E}[X^2]\mathbb{E}[Z^2].$$

(6) Usando a desigualdade de Cauchy-Schwarz, mostre que quando as distribuições marginais de Y e Z são idênticas, então

$$\mathrm{Cov}(Y, Z) \leq \mathrm{Var}(Y).$$

$$\mathrm{Cov}(Y,Z) \leq |\mathrm{Cov}(Y,Z)|$$

$$\leq |\mathbb{E}[(Y-\mathbb{E}[Y])(Z-\mathbb{E}[Z])]|$$

por Cauchy-Schwarz $\rightarrow$ $$\leq \sqrt{\mathbb{E}[(Y-\mathbb{E}[Y])^2]\,\mathbb{E}[(Z-\mathbb{E}[Z])^2]}$$

$$= \sqrt{\mathrm{Var}(Y)\,\mathrm{Var}(Z)}$$

para $\mathrm{Var}\,Y = \mathrm{Var}\,Z \Rightarrow = \mathrm{Var}(Y)$

(7) O desbaste de uma Cadeia de Markov $\{X^{(t)}\}_{t\geq 0}$ é uma técnica de reter uma subsequência do processo amostrado para fins de cálculo de médias ergódicas. Para algum $m \in \mathbb{N}$ mantemos o "subamostrador" $\{Y^{(t)}\}_{t\geq 0}$ definida por

$$Y^{(t)} := X^{(m \cdot t)}.$$

Podemos esperar que $\{Y^{(t)}\}_{t\geq 0}$ tenha menor autocorrelação que a cadeia original $\{X^{(t)}\}_{t\geq 0}$ e assim produziria médias ergóticas de menor variância.

Considere uma Cadeia de Markov $\{X^{(t)}\}_{t\geq 0}$. Seja T e m quaisquer 2 inteiros tais que $T \geq m > 1$ e $T/m \in \mathbb{N}$. Mostre que

$$Var\left[\frac{1}{T}\sum_{t=0}^{T-1} X^{(t)}\right] \leq Var\left[\frac{1}{T/m}\sum_{t=0}^{T/m-1} Y^{(t)}\right]$$

e explique brevemente o que esse resultado nos diz sobre o uso do desbaste.

Temos que,

$$Var\left[\sum_{t=0}^{T-1} X^{(t)}\right] = Var\left[\sum_{t=0}^{m-1}\sum_{l=0}^{T/m-1} X^{(lm+t)}\right]$$

$$= \sum_{t=0}^{m-1} Var\left[\sum_{l=0}^{T/m-1} X^{(l m+t)}\right]$$

$$+ \sum_{t=0}^{m-1}\sum_{r\in\{0,\dots,m-1\}\setminus t} Cov\left[\sum_{l=0}^{T/m-1} X^{(lm-t)}, \sum_{l=0}^{T/m-1} X^{(lm-r)}\right]$$

$$\leq m\, Var\left[\sum_{l=0}^{T/m-1} X^{(lm)}\right]$$

$$+ m(m-1) Var\left[\sum_{l=0}^{T/m-1} X^{(lm)}\right]$$

$$= m^2\, Var\left[\sum_{l=0}^{T/m-1} X^{(lm)}\right].$$

Daí,

$$\mathrm{Var}\left[\frac{1}{T}\sum_{t=0}^{T-1} X^{(t)}\right] = \frac{1}{T^2}\mathrm{Var}\left[\sum_{t=0}^{T-1} X^{(t)}\right]$$

$$\leq \frac{m^2}{T^2}\mathrm{Var}\left[\sum_{s=0}^{T/m-1} X^{(sm)}\right]$$

$$= \mathrm{Var}\left[\frac{1}{T/m}\sum_{t=0}^{T/m-1} Y^{(t)}\right].$$

Esse resultado indica que, em termos de variância de média ergódica, não há vantagem no método.