



HW 6

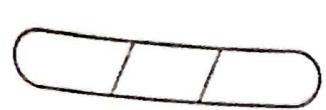
① Considere uma distribuição alvo em $\mathbb{X} = \mathbb{R}^d$ de densidade $\bar{\pi}(x)$. Preparamos para uma amostragem usando um algoritmo Metropolis-Hastings baseado em uma proposta aleatória. Dado $x^{(t-1)}$, a proposta para X é amostrada de forma: $x^{(t-1)}$ é a entrada de um processo de otimização estocástica que retorna um estimador $\theta | \mathcal{O} \in \mathbb{X}$ de maximizadores locais de $\bar{\pi}(x)$ na vizinhança de $x^{(t-1)}$. A v.a. \mathcal{O} segue uma função de densidade de prob. desconhecida $f(\theta | x^{(t-1)})$. Então amostramos $x^t \sim g(\cdot | \mathcal{O})$ onde $g(\cdot | \mathcal{O})$ é uma normal multivariada de média \mathcal{O} e covariância fixa Σ .

Exprese a proposta $g(x^t | x)$ como uma função de f e g . Se é possível avaliar a probabilidade de aceite $\alpha_{\text{MH}}(x^t | x) = \min\left\{1, \frac{\bar{\pi}(x^t)g(x^t)}{\bar{\pi}(x)g(x | x)}\right\}$ associada a proposta?

Temos que

$$g(x^t | x) = \int g(x^t | \theta) f(\theta | x) d\theta.$$

A distribuição proposta não pode ser avaliada ponto a ponto pois $f(\theta | x)$ é desconhecida.



(2) Do problema anterior. Considere agora o seguinte algoritmo Metropolis - Hastings alternado. Tome $X^{(0)} = x$, $\theta^{(0)} = \theta$ então itere em $t = 2, 3, \dots$

(a) Amortize $X \sim g(\cdot | \theta^{(t-1)})$. Dica: $f(\cdot | X)$

(b) Compute

$$\alpha_{RHM}((x, \theta) | (x^{(t-1)}, \theta^{(t-1)})) = \min \left\{ 1, \frac{\pi(x) g(x^{(t-1)} | \theta)}{\pi(x^{(t-1)}) g(x | \theta^{(t-1)})} \right\}.$$

(c) Com probabilidade $\alpha_{RHM}((x, \theta) | (x^{(t-1)}, \theta^{(t-1)}))$, tomamos $(X^{(t)}, \theta^{(t)}) = (x, \theta)$ e caso contrário $(X^{(t)}, \theta^{(t)}) = (X^{(t-1)}, \theta^{(t-1)})$

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

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

$$\Rightarrow \bar{\pi}(x, \theta) = \pi(x) f(\theta | x)$$

e

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

$$\begin{aligned} \text{Assim, } \bar{\pi}(x, \theta) k(x', \theta' | x, \theta) &= \pi(x) f(\theta | x) g(x' | \theta) f(\theta' | x') \min \left\{ 1, \frac{\pi(x) g(x' | \theta)}{\pi(x') f(\theta' | x) g(x | \theta)} \right\} \\ &\quad \pi(x') f(\theta' | x) g(x | \theta) f(\theta | x') \min \left\{ 1, \frac{\pi(x) g(x | \theta)}{\pi(x') g(x' | \theta)} \right\} \\ \text{usando que } &\rightarrow = \frac{\pi(x) g(x' | \theta)}{\pi(x') g(x | \theta)} \cdot \frac{\pi(x') g(x | \theta)}{\pi(x) g(x' | \theta)} \\ \min(1, a) = a \min(1, a') &\quad = 1. \end{aligned}$$

Logo é invariante



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

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

Temos que

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

Daí, tomando o valor esperado.

$$E[\min\{U, V\}] \stackrel{E[U] \leq V}{\leq} E[\min\{U, V\}] \leq E[V].$$

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

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

O valor esperado da probabilidade de aceitação estacionária é dado por

$$\begin{aligned}
 & \int \int \int \bar{\pi}(x, \theta) g((x, \theta) | (x, \theta)) \min \left\{ 1, \frac{\bar{\pi}(x', \theta') g((x, \theta) | (x', \theta'))}{\bar{\pi}(x, \theta) g((x', \theta') | (x, \theta))} \right\} dx dx' d\theta \\
 (2) & = \int \int \int \bar{\pi}(x) f(\theta|x) g(x'| \theta) \min \left\{ 1, \frac{\bar{\pi}(x') f(\theta'|x') g(x| \theta') f(\theta|x)}{\bar{\pi}(x) f(\theta|x) g(x| \theta) f(\theta|x')} \right\} dx d\theta dx' \\
 & = \int \int \int \bar{\pi}(x) \min \left\{ f(\theta|x) g(x| \theta), \frac{\bar{\pi}(x') f(\theta'|x') g(x'| \theta') f(\theta|x)}{\bar{\pi}(x) f(\theta|x) g(x| \theta)} \right\} dx d\theta dx' \\
 (3) & \leq \int \int \bar{\pi}(x) \min \left\{ \int f(\theta|x) g(x| \theta) f(\theta|x) d\theta, \frac{\int f(\theta'|x') g(x'| \theta') f(\theta|x) d\theta'}{\bar{\pi}(x)} \right\} dx \\
 & = \int \int \bar{\pi}(x) \min \left\{ g(x'|x), \frac{\bar{\pi}(x') g(x|x')}{\bar{\pi}(x)} \right\} dx dx' \\
 & = \int \int \bar{\pi}(x) \min \left\{ 1, \frac{\bar{\pi}(x') g(x|x')}{\bar{\pi}(x) g(x'|x)} \right\} dx dx'
 \end{aligned}$$

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

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

Temos que

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

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

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

Resolvendo em α temos

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

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

⑥ Usando a desigualdade de Cauchy-Schwarz, mostra que quando as distribuições marginais de Y e Z não são identicas, temos

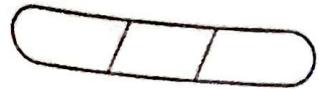
$$\text{Cov}(Y, Z) \leq \text{Var}(Y).$$

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

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

$$\begin{aligned} \text{Pela Cauchy-} \\ \text{-Schwarz} \rightarrow & \leq \sqrt{\mathbb{E}[(Y - \mathbb{E}[Y])^2] \mathbb{E}[(Z - \mathbb{E}[Z])^2]} \\ & = \sqrt{\text{Var}(Y) \text{Var}(Z)} \end{aligned}$$

$$\text{para } \text{Var} Y = \text{Var} Z = \text{Var}(Y)$$



7) O desbaste de uma Cadeia de Markov é uma técnica de iterar uma subseqüência do processo amostrado para fins de cálculo de médias ergódicas. Para algum $m \in \mathbb{N}$ mantemos o "subamostrador" $\{Y^{(t)}\}_{t=0}^{T-1}$ definida por

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

Podemos esperar que $\{Y^{(t)}\}_{t=0}^{T-1}$ tenha menor autocorrelação que a cadeia original $\{X^{(t)}\}_{t=0}^{T-1}$ e assim produzirá médias ergódicas de menor variação.

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

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

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

Temos que,

$$\begin{aligned} \text{Var}\left[\sum_{t=0}^{T-1} X^{(t)}\right] &= \text{Var}\left[\sum_{t=0}^{m-1} \sum_{n=0}^{Tm-1} X^{(2m+n)}\right] \\ &= \sum_{t=0}^{m-1} \text{Var}\left[\sum_{n=0}^{Tm-1} X^{(2m+n)}\right] \\ &\quad + \sum_{t=0}^{m-1} \sum_{n \in \{0, \dots, m-1\} \setminus T} \text{Cov}\left[\sum_{n=0}^{Tm-1} X^{(2m+n)}, \sum_{n=0}^{Tm-1} X^{(2m+n-t)}\right] \\ &\leq m \text{Var}\left[\sum_{n=0}^{Tm-1} X^{(2m)}\right] \\ &\quad + m(m-1) \text{Var}\left[\sum_{n=0}^{Tm-1} X^{(2m)}\right] \\ &= m^2 \text{Var}\left[\sum_{n=0}^{Tm-1} X^{(2m)}\right]. \end{aligned}$$



Daí,

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

$$\leq \frac{m^2}{T^2} \text{Var} \left[\sum_{t=0}^{Tm-1} X^{(tm)} \right]$$

$$= \text{Var} \left[\frac{1}{Tm} \sum_{t=0}^{Tm-1} Y^{(t)} \right].$$

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