HW 3

(1) ~~Seja~~ Vamos considerar a densidade da normal multivariada em $\mathbb{R}^d$ com covariancia de identidade, ou seja

$$\pi(x) = \frac{1}{(2\pi)^{d/2}} \exp\left\{-\frac{1}{2}x^T x\right\}.$$

(Cameron-Martin formula) Mostre que para qualquer $\theta \in \mathbb{R}^d$ e função $\phi : \mathbb{R}^d \to \mathbb{R}$

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

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

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

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

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

② Do exercício anterior. Segue diretamente da fórmula de Cameron-Martin e da Lei dos grandes números que, para $X_1, \dots, X_n \sim \pi$ independentes, o estimador

$$\hat{I}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \phi(X_i+\theta) \exp\left\{-\frac{1}{2}\theta^T\theta - \theta^T X_i\right\}$$

de $\mathbb{E}(\phi(X))$ é fortemente consistente para algum $\theta \in \mathbb{R}^d$ tal que $\mathbb{E}\left[\left|\phi(X+\theta)\exp\{-\frac{1}{2}\theta^T\theta - \theta^T X\}\right|\right] < \infty$.

O caso $\theta = (0,0,\dots,0)^T$ corresponde ao estimador usual de Monte Carlo. A variância de $\hat{I}_n(\theta)$ é dada por $\sigma^2(\theta)/n$ onde

$$\sigma^2(\theta) = \mathrm{Var}\left[\phi(X+\theta)\exp\left\{-\frac{1}{2}\theta^T\theta - \theta^T X\right\}\right].$$

Assumindo que $\sigma^2(\theta) < \infty$ para qualquer θ. Mostre que

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

Temos que

$$\sigma^2(\theta) = \mathbb{E}\left[\phi^2(X+\theta)\exp\{-\theta^T\theta - 2\theta^T X\}\right] - \left(\mathbb{E}[\phi(X)]\right)^2.$$

Daí,

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

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

$$= \frac{1}{(2\pi)^{d/2}}\int_{\mathcal{X}} \phi^2(z)\exp\left\{-\theta^T\theta - 2\theta^T(z-\theta) - \frac{1}{2}(z-\theta)^T(z-\theta)\right\} dz$$

$$= \frac{1}{(2\pi)^{d/2}}\int_{\mathcal{X}} \phi^2(z)\exp\left\{-\theta^T\theta - 2\theta^T z + 2\theta^T\theta + \theta^T z - \frac{1}{2}z^Tz - \frac{1}{2}\theta^T\theta\right\} dz$$

tilibra

$$= \frac{1}{(2\pi)^{d/2}} \int_{\mathcal{X}} \phi^2(z) \exp\left\{-\frac{1}{2} z^T z - \theta^T z + \frac{1}{2}\theta^T\theta\right\} dz$$

$$= \frac{1}{(2\pi)^{d/2}} \int_{\mathcal{X}} \phi^2(z) \exp\left\{-\frac{1}{2} z^T z\right\} \exp\left\{\frac{1}{2}\theta^T\theta - \theta^T z\right\} dz$$

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

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

~~que segue do~~ que finaliza a prova.

③ Uma função duas vezes diferenciável $f:\mathbb{R}^d \to \mathbb{R}$ é estritamente convexa se $\nabla^2 f(\theta)$ (a hessiana de f) é uma matriz positiva definida para todo $\theta \in \mathbb{R}^d$. Deduza da expressão de $\sigma^2(\theta)$ da questão ② e ① que a função $\theta \to \sigma^2(\theta)$ é estritamente convexa

Temos

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

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

Sendo $H_{ij}(\theta)$ o elemento i,j de $\nabla^2\sigma^2(\theta)$:

$$H_{ij}(\theta) = \left[\nabla^2\sigma^2(\theta)\right]_{ij} = \frac{\partial^2\sigma^2(\theta)}{\partial\theta_i\partial\theta_j}$$

$$= \mathbb{E}\left[\phi^2(X)\,\delta_{ij}\exp\left\{-\tfrac{1}{2}X^TX + \tfrac{1}{2}(X-\theta)^T(X-\theta)\right\}\right] +$$

$$+ \mathbb{E}\left[\phi^2(X)(\theta_i - X_i)(\theta_j - X_j)\exp\left\{-\tfrac{1}{2}X^TX + \tfrac{1}{2}(X-\theta)^T(X-\theta)\right\}\right]$$

onde $\delta_{ij} = 1$ se $i=j$ e 0 caso contrário. Daí,

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

$$+ \mathbb{E}\left[\phi^2(X)(\theta - X)(\theta - X)^T\exp\left\{-\tfrac{1}{2}X^TX + \tfrac{1}{2}(X-\theta)^T(X-\theta)\right\}\right].$$

Para qualquer vetor não-nulo $u \in \mathbb{R}$:

$$u^T H(\theta) u = \mathbb{E}\left[\phi^2(X)\exp\left\{-\tfrac{1}{2}X^TX + \tfrac{1}{2}(X-\theta)^T(X-\theta)\right\}(\|u\|^2 + (u^T(X-\theta))^2)\right].$$

Como $\|u\|^2 > 0$ e $(u^T(X-\theta))^2 \geq 0$, $u^T H(\theta) u > 0$. Logo positiva definida.

tilibra

(4) Considerando os exercícios anteriores. Mostre que o mínimo de $\theta \mapsto \sigma^2(\theta)$ é atingido com θ^* tal que

$$\mathbb{E}[\phi^2(X)(\theta^* - X)\exp(-\theta^{*T}X)] = 0.$$

Temos que

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

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

$$= \underbrace{\exp\left\{\frac{1}{2}\theta^T\theta\right\}}_{>0}\mathbb{E}\left[\phi^2(X)(\theta - X)\exp\left\{-\theta^TX\right\}\right]$$

Logo, $\nabla\sigma^2(\theta) = 0$ se e só se $\mathbb{E}[\phi^2(X)(\theta - X)\exp\{-\theta^TX\}] = 0$.
Por ser convexa é o que minimiza $\theta \mapsto \sigma^2(\theta)$.

5) Aplicamos os resultados anteriores a um modelo de opções européias no modelo Black-Scholes. Queremos computar

$$I = \exp\{-rT\}\,\mathbb{E}[\max\{0, \lambda \exp\{\sigma X\} - K\}]$$

onde $X \sim N(0,1)$ e r, λ, K, σ, T são constantes reais positivas com $\lambda < K$. Mostre que $\theta \to \sigma^2(\theta)$ é decrescente em $D = (-\infty, \sigma^{-1}\log(K/\lambda))$. Deduza deste resultado que existe um intervalo de valores de θ tais que a variância de $\hat{I}_n(\theta)$ é estritamente menor que a variância do usual estimador Monte Carlo.

Estamos interessados em $I = \exp\{-rT\}\mathbb{E}[\phi(X)]$, onde $\phi(X) = \max\{0, \lambda\exp\{\sigma X\} - K\}$. Queremos mostrar que a derivada de $\sigma^2(\theta)$ é negativa, ou seja, mostrar que

$$\nabla\sigma^2(\theta) = \exp\left(\frac{1}{2}\sigma^2\right)\mathbb{E}[\phi^2(X)(\theta - X)\exp(-\theta X)] \leq 0.$$

$$\iff \theta\,\mathbb{E}[\phi^2(X)\exp\{-\theta X\}] - \mathbb{E}[\phi^2(X)X\exp\{-\theta X\}] \leq 0.$$

$$\iff \theta\,\mathbb{E}[(\lambda\exp\{\sigma X\} - K)^2\exp\{-\theta X\}\mathbb{I}_{D^c}(X)]$$

$$- \mathbb{E}[(\lambda\exp\{\sigma X\} - K)^2 X\exp\{-\theta X\}\mathbb{I}_{D^c}(X)] \leq 0.$$

- Caso $\theta < 0$, temos que a primeira esperança é positiva e a segunda é maior igual a zero, logo vale.
- Caso $\theta \in (0, \sigma^{-1}\log(K/\lambda))$, então $\theta \leq X$ (caso contrário teríamos $X \in D$), daí $(\theta - X) \leq 0 \Rightarrow \nabla\sigma^2(\theta) \leq 0$. Para $\theta = 0$ temos o usual estimador de Monte Carlo, como a variância de $\hat{I}_n(\theta)$ é decrescente em $(0, \sigma^{-1}\log(K/\lambda))$, então é um estimador melhor.

(6) Seja $\mathbb{X}$ um espaço de estados finitos. Considere o seguinte Kernel de transição Markov.

$$T(x,y) = \alpha(x,y)\, g(x,y) + \left(1 - \sum_{z\in\mathbb{X}} \alpha(x,z)\, g(x,z)\right) \delta_x(y)$$

onde $g(x,y) \geq 0$, $\sum_{y\in\mathbb{X}} g(x,y) = 1$ e $0 \leq \alpha(x,y) \leq 1$, $\forall x,y \in \mathbb{X}$. $\delta_x(y)$ é o símbolo Kronecker, ou seja, $\delta_x(y) = 1$ se $y = x$ e 0 caso contrário.

Explique como você simularia uma cadeia de Markov com o ~~g~~ Kernel de transição T.

O algoritmo no passo t é dado por:

- $X \sim g(X_{t-1}, \cdot)$
- $U \sim \text{Unif}[0,1]$
- Se $U \leq \alpha(X_{t-1}, X)$, então $X_t := X$
  caso contrário tomamos $X_t := X_{t-1}$.

O termo $\left(1 - \sum_{z\in\mathbb{X}} \alpha(x,z)\, g(x,z)\right)$ explica a probabilidade de permanecer no estado atual. T é um Kernel de transição válido que garante que a cadeia de Markov satisfaça as propriedades de uma distribuição de probabilidade.

(7) Seja π uma função de massa de probabilidade em X. Mostre que se

$$\alpha(x,y) = \frac{\gamma(x,y)}{\pi(x)\, g(x,y)}$$

onde $\gamma(x,y) = \gamma(y,x)$ e $\gamma(x,y)$ é escolhido tal que $0 \leq \alpha(x,y) \leq 1$ $\forall x,y \in X$ então T é reversível sobre π.

Para $x=y$, vale, basta mostrar que

$$\pi(x) T(x,y) = \pi(y) T(y,x), \quad \forall x,y \in X, \; x \neq y$$

$$\iff \pi(x)\, g(x,y)\, \alpha(x,y) = \pi(y)\, g(y,x)\, \alpha(y,x)$$

$$\iff \pi(x)\, g(x,y) \frac{\gamma(x,y)}{\pi(x) g(x,y)} = \pi(y)\, g(y,x) \frac{\gamma(y,x)}{\pi(y) g(y,x)}$$

$$\iff \gamma(x,y) = \gamma(y,x)$$

⑧ Mostre que o algoritmo de Metropolis-Hastings corresponde a uma escolha particular de $\gamma(X,Y)$.

Como

$$\alpha(X,Y) = \min\left\{1, \frac{\pi(Y)\,q(Y,X)}{\pi(X)\,q(X,Y)}\right\}$$

E queremos

$$\frac{\gamma(X,Y)}{\pi(X)\,q(X,Y)} = \min\left\{1, \frac{\pi(Y)\,q(Y,X)}{\pi(X)\,q(X,Y)}\right\},$$

basta tomar $\gamma(X,Y) = \min\{\pi(Y)q(Y,X), \pi(X)q(X,Y)\}$ preservando a reflexibilidade.

(9) Seja π uma função de massa de probabilidade num espaço finito $\mathcal{X}$ tal que $\pi(x) > 0$, $\forall x \in \mathcal{X}$. Para amostrar de π, rodamos um Metropolis-Hasting $(X_t)_{t \geq 1}$ com proposta $q(x,y) \geq 0$, tal que $\sum_{y \in X} q(x,y) = 1$ e $q(x,x) = 0$ $\forall x \in X$. Considere aqui a sequencia $(Y_k)_{k \geq 1}$ de propostas aceitas. $Y_{(1)} = X_{(\tau_1)}$ onde $\tau_1 = 1$ e, para $k \geq 2$, $Y_{(k)} = X_{(\tau_k)}$ onde $\tau_k = \min\{t : t > \tau_{k-1}, X_{(t)} \neq Y_{(\tau_{k-1})}\}$

Seja $\phi : X \to \mathbb{R}$ uma função teste. Mostre que o estimador $\frac{1}{\tau_k - 1} \sum_{t=1}^{\tau_k - 1} \phi(X_{(t)})$ pode ser reescrito como uma função de $(Y_{(k)})_{k \geq 1}$ e $(\tau_k)_{k \geq 1}$ e prove que a sequencia $(Y_{(k)})_{k \geq 1}$ é uma cadeia de Markov com Kernel de transição

$$K(x,y) = \frac{\alpha(x,y)\, q(x,y)}{\sum_{z \in X} \alpha(x,z)\, q(x,z)}.$$

Para a primeira parte temos que

$$\frac{1}{\tau_k - 1} \sum_{t=1}^{\tau_{k-1}} \phi(X_{(t)}) = \frac{1}{\tau_k - 1} \sum_{j=1}^{k-1} \phi(Y_{(j)}) (\tau_{j+1} - \tau_j).$$

Agora para a segunda parte, $Y_{(k)}$ é aceito então

$$P(Y_{(k)} = y \mid Y_{(1)}, \dots, Y_{(k-1)} = x) = P(X_{(\tau_k)} = y \mid X_{(1)}, \dots, X_{(\tau_{k-1})} = x)$$

$$= P(X_{(\tau_k)} = y \mid X_{(\tau_{k-1})} = x)$$

$$= \frac{\alpha(x,y)\, q(x,y)}{\sum_{z \in X} \alpha(x,z)\, q(x,z)}$$

(10) Mostre que o Kernel de transição $K(x,y)$ da cadeia de Markov $(Y_{(k)})_{k\geq 1}$ é $\tilde{\pi}$-reversível onde

$$\tilde{\pi}(x) = \frac{\pi(x)m(x)}{\sum_{z\in\mathcal{X}} \pi(z)\, m(z)}$$

com

$$m(x) := \sum_{z\in\mathcal{X}} \alpha(x,z)\, q(x,z).$$

Temos que

$$\tilde{\pi}(x) K(x,y) = \frac{\pi(x) m(x)}{\underbrace{\sum_{z\in\mathcal{X}} \pi(z) m(z)}_{Z}} \cdot \frac{\alpha(x,y)\, q(x,y)}{\underbrace{\sum_{z\in\mathcal{X}} \alpha(x,z)\, q(x,z)}_{m(x)}}$$

$$= \frac{\pi(x)\cancel{m(x)}}{Z} \cdot \frac{\alpha(x,y)\, q(x,y)}{\cancel{m(x)}}$$

$$= \frac{1}{Z} \cdot \frac{\gamma(x,y)}{\cancel{\alpha(x,y)\, q(x,y)}} \cdot \cancel{\alpha(x,y)\, q(x,y)}$$

$$= \frac{\gamma(y,x)}{Z}$$

$$= \frac{1}{Z} \cdot \frac{\gamma(y,x)}{\alpha(y,x)\, q(y,x)} \cdot \alpha(y,x)\, q(y,x)$$

$$= \frac{\pi(y)\, \alpha(y,x)\, q(y,x)}{Z}$$

$$= \frac{\pi(y)\, m(y)}{Z} \cdot \frac{\alpha(y,x)\, q(y,x)}{m(y)}$$

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

11) Assuma que para algum teste $\phi: \mathbb{X} \to \mathbb{R}$ temos que

$$\frac{1}{k}\sum_{i=1}^{k}\phi(Y_{(i)}) \xrightarrow{q.c.} \sum_{x\in\mathbb{X}} \phi(x)\tilde{\pi}(x)$$

e assuma que $m(x)$ pode ser computado exatamente $\forall x \in \mathbb{X}$. Proponha um estimador fortemente consistente de $\sum_{x\in\mathbb{X}} \phi(x)\pi(x)$ baseado na cadeia de Markov $(Y_{(k)})_{k\geq 1}$ a qual depende de $(\tilde{\pi}_x)_{x \geq 1}$.

Temos que

$$\sum_{x\in\mathbb{X}}\phi(x)\pi(x) = \frac{\sum_{x\in\mathbb{X}}\phi(x)\pi(x)}{\sum_{x\in\mathbb{X}}\pi(x)}$$

$$= \frac{\sum \phi(x)\frac{\pi(x)}{\tilde{\pi}(x)}\tilde{\pi}(x)}{\sum \frac{\pi(x)}{\tilde{\pi}(x)}\cdot\tilde{\pi}(x)}$$

como $\frac{\pi(x)}{\tilde{\pi}(x)} \propto \frac{1}{m(x)}$. Um estimador fortemente consistente é

$$\frac{\sum \phi(Y_{(x)})/m(Y_{(x)})}{\sum 1/m(Y_{(x)})}$$