ORF524 - Problem Set 3

Bachir EL KHADIR

October 19, 2015

Question 1

1. for $p > 0$, let's denote $f_p : x \to x^p$ for $x > 0$. f_p is convexe because $f_p''(x) = p(1-p)x^{p-2}$.
   $\mathcal{L}a = \sum_i (a_i - \theta_i)^p = \sum_i f_p(a_i)$
2. $\mathcal{L}a = f_q(f_p(a))$ is convexe as the composition of two convexe functions.

Question 2

1. The X_i have the same distribution and play symetric roles, so:

$$\tilde{p} = E[\hat{p}|T(X)] = E[X_1|\sum_i X_i] = \frac{1}{n}E[\sum X_i|\sum_i X_i] = \frac{T(X)}{n}$$

2. ...

Question 3

$$Var(E[X|Y]) = E[(E[X|Y] - E[E[X|Y]])^2] = E[(E[X|Y] - E[X])^2]$$

$$\begin{aligned} E[Var(X|Y)] &= E[E[(X - E[X|Y])^2|Y]] \\ &= E[E[(X - E[X] + E[X] - E[X|Y])^2] \\ &= E[E[(X - E[X])^2]] - E[(E[X] - E[X|Y])^2] \quad \text{By Pythagor, because } E[X] - E[X|Y] \perp X - E[X] \end{aligned}$$

By summing:

$$E[Var(X|Y)] + Var(E[X|Y]) = E[E[(X - E[X])^2]] = Var(X)$$

Question 4

Let's prove that $\phi(\{c_j\}^l$ is non-increasing.

$$\phi(\{c_j\}^l = \sum_j \sum_{x_i \in C_j^l} ||x_i - c_j^l||^2 \leq \sum_{x_i \in C_j^l} ||x_i - c_j^{l+1}||^2$$

For every $j = 1..K$, $\sum_{x_i \in C_j^l} ||x_i - c_j^l||^2 \leq \sum_{x_i \in C_j^l} ||x_i - c_j^{l+1}||^2$ because the mean of the point $x_i \in C_j^l$, minimizes the quantity $\mu \to \sum_{x_i \in C_j^l} ||x_i - \mu||^2$.

So: $\phi(\{c_j\}^l = \sum_j \sum_{x_i \in C_j^l} ||x_i - c_j^l||^2 \leq \sum_i \sum_{x_i \in C_j^l} ||x_i - c_j^{l+1}||^2$

By assigning each x_i to the nearest c_j^{l+1}, each quantity $||x_i - c_j^l||^2$ in the sum above is replaced by a smaller (or equal) quantity $||x_i - c_j^{l+1}||^2$

Therefore $\phi(\{c_j\}^l$ is non-increasing, and the limit exsits.

Question 5

1.

$$
\begin{aligned}
Cov((T,a)^T) &= \mathbb{E}(T,a)(T,a)^T - (\mathbb{E}(T,a))(\mathbb{E}(T,a))^T \\
&= \mathbb{E}\begin{bmatrix} TT^T & T^T a \\ aT^T & aa^T \end{bmatrix} - \begin{bmatrix} \mathbb{E}(T)\mathbb{E}(T)^T & \mathbb{E}(T)^T\mathbb{E}(a) \\ \mathbb{E}(a)\mathbb{E}(T)^T & \mathbb{E}(a)\mathbb{E}(a)^T \end{bmatrix} \\
&= \begin{bmatrix} Cov(T) & Cov(T,a) \\ Cov(a,T) & Cov(a) \end{bmatrix} \\
&= \begin{bmatrix} Cov(T) & \nabla_\theta g(\theta) \\ \nabla_\theta g(\theta)^T & I(\theta) \end{bmatrix}
\end{aligned}
$$

Because:

- $\mathbb{E}(a) = \int \nabla_\theta \log f_\theta(x) f_\theta(x) \mathrm{dx} = \int \frac{\nabla_\theta \mathrm{f}_\theta(\mathrm{x})}{\mathrm{f}_\theta(\mathrm{x})} \mathrm{f}_\theta(\mathrm{x})\mathrm{dx} = \nabla_\theta 1 = 0$
- $Cov(a) = \mathbb{E}(aa^T) = I(\theta)$
-

$$
\begin{aligned}
Cov(T,a) &= \mathbb{E}(T^T a) \\
&= \int T(x) \nabla_\theta \log f_\theta(x) f_\theta(x) \mathrm{dx} \\
&= \int T(x) \frac{\nabla_\theta f_\theta(x)}{f_\theta(x)} f_\theta(x) \mathrm{dx} \\
&= \nabla_\theta \int T(x) f_\theta(x) \mathrm{dx} && \text{(By regularity condition)} \\
&= \nabla_\theta g(\theta)
\end{aligned}
$$

2.

$$B = \left(\; -I_p \quad , \nabla_\theta g(\theta) I(\theta)^{-1} \; \right)^T$$

$B^T Cov(T,a)^T B = ...$

3. $Cov(T) - \nabla_\theta g(\theta) I(\theta) \nabla_\theta g(\theta) = B^T Cov(T,a)^T B = Cov(B(T,a)^T) \geq 0$

Question 6

In the following we write f instead of $f_\theta(x)$ or $f_\theta(X)$.

$$\nabla_\theta^2 \log f = \nabla_\theta \left(\frac{\nabla_\theta f}{f} \right) = \frac{\nabla_\theta^2 f}{f} - \frac{\nabla_\theta f \nabla_\theta f^T}{f^2} = \frac{\nabla_\theta^2 f}{f} - \nabla_\theta \log f \nabla_\theta \log f^T$$

But $\mathbb{E}(\frac{\nabla_\theta^2 f}{f}) = \int \frac{\nabla_\theta^2 f}{f} f \mathrm{dx} = \nabla_\theta^2 \int \mathrm{fdx} = 0$, so

$$I(\theta) = \mathbb{E}(\nabla_\theta f_\theta(x) \nabla_\theta f_\theta(x)^T) = -\mathbb{E}(\nabla_\theta^2 f)$$

Question 7

1. By the series expansion of exponential:

$$\left|\frac{e^{az}-1}{z}\right| = \left|\sum_{k=1}^{\infty} \frac{a^k z^{k-1}}{k!}\right| \leq \sum_{k=1}^{\infty} \frac{|a|^k |z|^{k-1}}{k!} \leq \sum_{k=1}^{\infty} \frac{1}{\delta} \frac{|a|^k |\delta|^k}{k!} \frac{e^{a\delta}}{\delta}$$

2. Let $\alpha_n \Rightarrow \alpha_n$

$$\frac{g(x)f_{\alpha_n}(x) - g(x)f_\alpha(x)}{\alpha_n - \alpha} = g(x)h(x)$$

Question 8

1.

$$\hat{\beta} = (X^TX)^{-1}X^Ty = (X^TX)^{-1}X^TX\beta + (X^TX)^{-1}X^T\eta = \beta + (X^TX)^{-1}X^T\eta$$

$\mathbb{E}(\hat{\beta}) = \beta + (X^TX)^{-1}X^T\mathbb{E}(\eta) = \beta$. So $\hat{\beta}$ is unbiased.

best var

2. $R_2(\hat{\beta}) = Var(\hat{\beta}) = Var((X^TX)^{-1}X^T\eta) = (X^TX)^{-1}X^TVar(\eta)((X^TX)^{-1}X^T)^T = \sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1}$ $\sigma^2(X^TX)^{-1}$

If $X^TX = I_p$, $R_2(\hat{\beta}) = \sigma^2 I_p$

Question 9

1.

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{dc}}\mathbb{E}(|X-c|) &= \frac{\mathrm{d}}{\mathrm{dc}}\int^c (c-x)f(x)\mathrm{dx} + \int_{\mathrm{c}} \mathrm{(x-c)f(x)dx} \\
&= \frac{\mathrm{d}}{\mathrm{dc}}c(F(c) - (1-F(c)) + \int^c -xf(x)\mathrm{dx} + \int_{\mathrm{c}} \mathrm{xf(x)dx} \\
&= \frac{\mathrm{d}}{\mathrm{dc}}c(2F(c)-1) - 2\int^c xf(x)\mathrm{dx} + \int_{\mathbb{R}} \mathrm{xf(x)dx} \\
&= 2F(c) - 1 + 2cf(c) - 2cf(c) \\
&= 2F(c) - 1
\end{aligned}$$

The derivative is increasing so the function is strictly convexe, and therefore it attains its minimum when the derivative is 0, or $F(c) = \frac{1}{2}$, or $c = \text{median}(P_X)$