UGA M1: Économétrie 1

# Problème 3

Michal Urdanivia, Université de Grenoble Alpes, Faculté d'Économie, GAEL
Courriel: michal.wong-urdanivia@univ-grenoble-alpes.fr

11 octobre 2017

## 1. Espérance conditionnelle et meilleure prédiction

Supposons que l'on cherche à prédire une variable $Y \in \mathbb{R}$ par un ensemble de variables représentées par le vecteur $X \in \mathbb{R}^K$. Cette prédiction est une fonction $g(x)$ pour les valeurs de $X = x$. L'erreur de prédiction est $Y - g(x)$ et une mesure de l'ampleur de cette erreur est l'espérance de son carré,

$$\mathbb{E}\left[(Y - g(x))^2 | X = x\right]$$

qui est **l'erreur quadratique moyenne**.

(1) Montrer que,
$$\mathbb{E}\left[(Y - g(x))^2 | X = x\right] = \mathbb{V}(Y|x) + (g(x) - \mathbb{E}(Y|x))^2$$
où $\mathbb{E}(Y|x) = \mathbb{E}(Y|X = x)$, $\mathbb{V}(Y|x) = \mathbb{V}(Y|X = x)$.

(2) En déduire que l'erreur quadratique moyenne est minimisée pour $g(x) = \mathbb{E}(Y|x)$.

## 2. Linéarité de l'espérance conditionnelle

Supposons que $X = (1, W)^\top$ et que,

$$\mathbb{E}(Y|X) = \beta_0 + \beta_1 W$$

avec

$$(\beta_0, \beta_1) = \arg\min_{b_0, b_1} \mathbb{E}\left((Y - b_0 - b_1 W)^2\right)$$

(1) Donnez les conditions du premier ordre qui définissent $\beta_0$ et $\beta_1$.

(2) Montrez que $\beta_0 = \mathbb{E}(Y) - \beta_1 \mathbb{E}(X)$ et $\beta_1 = \frac{\mathbb{C}(X,Y)}{\mathbb{V}(X)}$.

## 3. Régression linéaire simple

*Exercice 1.* Considérons le modèle défini par,

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad i = 1, ..., n$$

pour un échantillon i.i.d., et où $\mathbb{E}[\epsilon_i] = 0$, $\mathbb{V}(\epsilon_i) = \sigma^2$. Soit $\hat{\beta}_0$, et $\hat{\beta}_1$ les estimateurs des moindres carrés de $\beta_0$ et $\beta_1$.

(1) Calculez $\mathbb{V}(\hat{\beta}_0 | X_i)$.

(2) On calcule les résidus $\hat{\epsilon}_i = Y_i - \hat{Y}_i$, où $\hat{Y}_i = (\hat{\beta}_0 + \hat{\beta}_1 X_i)$ sont les valeurs ajustées. Montrez que $\sum_{i=1}^{N} \hat{\epsilon}_i = 0$.

(3) Montrez que $\sum_{i=1}^{N} \hat{Y}_i \hat{\epsilon}_i = 0$. Quelle est l'interprétation de ce résultat.

(4) Supposons calculiez une régression des $Y_i$ sur les résidus $\epsilon_i$(autrement dit, que vous estimiez $Y_i = \alpha_0 + \alpha_1 \hat{\epsilon}_i + v_i$, $v_i$ étant l'erreur du modèle). Quels seraient les paramètres estimés par MCO de $\alpha_0$, et $\alpha_1$ ?

*Exercice 2.* Considérons le modèle,

$$Y_i = \beta_1 X_i + \epsilon_i$$

(1) Calculez l'estimateur des MCO de $\beta_1$, noté $\hat{\beta}_1$.

(2) Montrez que $\hat{\beta}_1$ est sans biais(autrement dit, que $\mathbb{E}[\hat{\beta}_1] = \beta_1$).

(3) Supposons que le vraie modèle soit en fait $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$. Montrez que $\hat{\beta}_1$ calculé dans la première question est biaisé, et donnez une expression de son biais.

## 4. Exercice de simulation

(1) Générez $n = 100$ observations de la façon suivante. D'abord générez une variable $X_i \sim \mathcal{U}(0,1)$ et une autre variable $\epsilon_i \sim \mathcal{N}(0,1)$. Pour $\beta_0 = 5$ et $\beta_1 = 3$ calculez,

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad i = 1, ..., n$$

Représentez alors graphiquement les points $\{(Y_1, X_1), ..., (Y_n, X_n)\}$. Calculez les estimateurs de MCO et ajoutez sur le même graphique la droite $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$.

(2) Répétez l'expérience du point précédent 1000 fois. Pour chaque répétition vous aurez un estimateur $\hat{\beta}_1^{(t)}$, pour $t = 1, ..., 1000$. Calculez la moyenne de la suite des valeurs ainsi obtenues. A quelle valeur de la moyenne devriez vous vous attendre ? Représentez un histogramme de cette suite.

(3) Répétez le point précédent mais avec $\epsilon_i$ suivant une loi de Cauchy. Cette distribution ressemble à la distribution normale mais avec des queues très épaisses. Quels modifications constatez vous au niveau de l'histogramme ?

(4) Étudions à présent ce qu'il arrive quand les $X_i$ sont observés avec des erreurs. Pour cela, générez $n = 100$ comme suit :

$$\begin{aligned} X_i &\sim \mathcal{U}(0,1) \\ W_i &= X_i + \delta_i \\ Y_i &= \beta_0 + \beta_1 X_i + \epsilon_i \end{aligned}$$

où $\beta_0 = 5$, $\beta_1 = 3$, $\epsilon_i \sim \mathcal{N}(0,1)$, et $\delta_i \sim \mathcal{N}(0,2)$. Supposons que nous n'observions que $(Y_1, W_1), ..., (Y_n, W_n)$. Représentez graphiquement ces points. Calculez les estimateur des MCO dans la régression des $Y_i$ sur $W_i$, et représentez la droite $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$. Répétez cette expérience 1000 fois et calculez la moyenne de $\hat{\beta}_1$ sur les 1000 valeurs obtenues. Représentez aussi l'histogramme pour la suite des valeurs obtenues. D'après cet exercice quel est l'effet des erreurs de mesure sur les $X_i$.

 