Rappels : Théorie asymptotique

Cem Ertur *

8 septembre 2014
— Version 1.0 —

1 Motivation

Parfois, il peut arriver qu'on ne puisse pas exprimer analytiquement et démontrer, en échantillon de petite taille, les propriétés d'absence de biais ou de variance minimale d'un estimateur. Mais on peut toutefois dériver des résultats approchés sur le comportement de la distribution de l'estimateur lorsque la taille de l'échantillon devient de plus en plus grand. On s'interesse dans de tels cas aux propriétés pour des échantillons de grande taille : ces propriétés sont appelées propriétés asymptotiques. Nous nous limiterons ici à la convergence en probabilité, la convergence en moyenne quadratique et la convergence en loi d'un estimateur (ou d'une suite de variables aléatoires).

1.1 Le modèle

Considérons le modèle de régression linéaire suivant avec k variables explicatives, si on dispose d'un échantillon de taille n, on écrit pour la ième observation :

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \varepsilon_i \qquad i = 1, ..., n \tag{1}$$

Si dans le modèle il y a un terme constant, cela signifie $x_{i1} = 1, \forall i$. k est alors le nombre de paramètres à estimer et il y à $k-1$ variables explicatives dans le modèle.

Sous forme matricielle, en empilant les n observations, on écrit :

$$y = X\beta + \varepsilon \tag{2}$$

où y est la variable expliquée d'ordre $n \times 1$, X est la matrice des variables explicatives d'ordre $n \times k$, β est le vecteur des coefficients inconnus d'ordre $k \times 1$ et ε est le vecteur d'erreurs d'ordre $n \times 1$.

1.2 Les hypothèses

Sur la nature de la relation entre les variables :

H_1 : forme fonctionnelle linéaire stable.

*Laboratoire d'Economie d'Orléans (UMR 6221 CNRS), UFR Droit Economie Gestion, Université d'Orléans. E-mail : `cem.ertur@univ-orleans.fr`, web site : `http://certur.free.fr/`.

Sur les variables explicatives :

H_2 : les variables explicatives sont supposées non stochastiques (fixes) ou,

$H_{2'}$: indépendantes du terme d'erreur : $cov(X, \varepsilon) = 0$.

On suppose que la matrice X est de rang complet : $rg\,(X) = k < n$ (les colonnes de X sont linéairement indépendantes).

Pour obtenir les propriétés asymptotiques, on note que l'hypothèse d'indépendance entre variables explicatives et terme d'erreur implique plim $\frac{1}{n}X'\varepsilon = 0$. On suppose de plus $\lim_{n\to\infty}\frac{1}{n}X'X = Q$ (ou plim $\frac{1}{n}X'X = Q$), matrice définie positive, En particulier, l'élément sur la diagonale principale de la matrice Q s'écrit : $q_{jj} = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} x_{ij}^2$, et $j = 1, ..., k$.

Sur les erreurs :

H_{3a} : $\begin{cases} \mathrm{E}\,(\varepsilon) = 0 \\ \mathrm{E}\,(\varepsilon\varepsilon') = \sigma^2 \end{cases}$ Les erreurs sont d'espérance mathématique nulle.

H_{3b} : $\varepsilon_i \sim i.i.d.\left(0, \sigma^2\right)$, les erreurs sont indépendamment et identiquement distribuées.

H_{3c} : $\varepsilon_i \sim i.i.d.\left(0, \sigma^2\right)$ et le moment d'ordre 4 existe : $\mathrm{E}(\varepsilon_i^4) = \mu_4 < \infty$.

H_{3d} : $\varepsilon \sim N\left(0, \sigma^2 I\right)$, les erreurs sont normalement distribuées, d'espérance mathématique nulle et de variance constante σ^2.

Ces hypothèses, allant de la moins restrictive à la plus restrictive, sont emboîtées, on a l'implication suivante :

$$H_{3d} \quad \Longrightarrow \quad H_{3c} \quad \Longrightarrow \quad H_{3b} \quad \Longrightarrow \quad H_{3a}$$

1.3 Les estimateurs des MCO

L'estimateurs de β est obtenus en appliquant la méthode des Moindres Carrés qui consiste à minimiser la somme des carrés des erreurs.

Soit le vecteur d'erreur : $\hat{\varepsilon} = y - X\hat{\beta}$. La somme des carrés des erreurs s'écrit :

$$SS = \hat{\varepsilon}'\hat{\varepsilon} = (y - X\hat{\beta})'(y - X\hat{\beta}) \tag{3}$$

Le problème est donc de minimiser cette somme des carrés des erreurs par rapport au vecteur des coefficients inconnus β :

$$\min_{\beta} SS \quad = \min_{\beta}\,(y - X\hat{\beta})'(y - X\hat{\beta}) \tag{4}$$

Développons SS :

$$\begin{aligned} SS &= y'y - \hat{\beta}'X'y - yX\hat{\beta} + \hat{\beta}'X'X\hat{\beta} \\ &= y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta} \end{aligned}$$

En effet $\hat{\beta}'X'y = yX\hat{\beta}$, ces expressions représentant le même scalaire. Notons au passage que $2\hat{\beta}'X'y$ est une forme linéaire en $\hat{\beta}$ et que $\hat{\beta}'X'X\hat{\beta}$ est une forme quadratique en $\hat{\beta}$.

Écrivons les conditions du premier ordre :

$$\frac{\partial SS}{\partial \beta} = -2X'y + 2X'X\hat{\beta} = 0 \tag{5}$$

Si $\mathrm{rg}(X) = k$, $X'X$ est définie positive, donc $X'X$ est non singulière, son inverse existe. On peut donc écrire :

$$\widehat{\beta}_{MCO} = (X'X)^{-1}X'y \tag{6}$$

1.3.1 Propriétés en échantillon de taille finie des estimateurs des MCO

- Considérons l'hypothèse H_{3a}, la moins restrictive sur la spécification des erreurs. On postule l'homoscédasticité et l'absence de corrélation des erreurs. On a vu que l'estimateur des MCO, noté $\widehat{\beta}_{MCO}$ s'écrit $\widehat{\beta}_{MCO} = (X'X)^{-1} X'y$, par conséquent :
  $\widehat{\beta}_{MCO} = Ay$ avec $A = (X'X)^{-1} X'$ et $\widehat{\beta}_{MCO}$ est un estimateur linéaire en y.
  $E(\widehat{\beta}_{MCO}) = AE(y) = AX\beta = (X'X)^{-1} X'X\beta = \beta$ donc $\widehat{\beta}_{MCO}$ est un estimateur centré de β.
  $V(\widehat{\beta}_{MCO}) = AV(y)A' = \sigma^2 AA' = \sigma^2 (X'X)^{-1} X'X (X'X)^{-1} = \sigma^2 (X'X)^{-1}$ est la matrice des variances-covariances d'ordre $k \times k$ de l'estimateur $\widehat{\beta}_{MCO}$.
  Si on appelle x^{ij} l'élément d'ordre (i,j) de la matrice $(X'X)^{-1}$:

$$\begin{aligned} V(\widehat{\beta}_i) &= \sigma^2 x^{ii} \\ cov(\widehat{\beta}_i, \widehat{\beta}_j) &= \sigma^2 x^{ij} \end{aligned}$$

- $\widehat{\beta}_{MCO}$, est BLUE (Best Linear Unbiased Estimator) : c'est-à-dire que c'est l'estimateur à variance minimale dans la classe des estimateurs linéaires centrés (Théorème de Gauss-Markov).
- $\widehat{\sigma}^2_{MCO} = \frac{SS}{T-k}$ estimateur centré de σ^2 car $E(SS) = \sigma^2 (T-k) \Rightarrow E\left(\widehat{\sigma}^2_{MCO}\right) = \sigma^2$
  En effet la somme des carrés des erreurs estimées SS s'écrit :

$$\begin{aligned} SS &= \widehat{\varepsilon}'\widehat{\varepsilon} = \left(y - X\widehat{\beta}_{MCO}\right)' \left(y - X\widehat{\beta}_{MCO}\right) \\ &= y'y - \widehat{\beta}'_{MCO}X'y - y'X\widehat{\beta}_{MCO} + \widehat{\beta}'_{MCO}X'X\widehat{\beta}_{MCO} \\ &= y'y - 2\widehat{\beta}'_{MCO}X'y + \widehat{\beta}'_{MCO}X'X\left(X'X\right)^{-1}X'y \\ &= y'y - 2\widehat{\beta}'_{MCO}X'y + \widehat{\beta}'_{MCO}X'y \\ &= y'y - \widehat{\beta}'_{MCO}X'y \end{aligned}$$

  Mais aussi :

$$SS = y'M_X y = \varepsilon' M_X \varepsilon \tag{7}$$

  où $M_X = I - X(X'X)^{-1} X'$ est une matrice idempotente et symétrique de rang $T-k$. SS est donc une forme quadratique idempotente en y ou en ε.
  En effet :

$$\begin{aligned} \widehat{\varepsilon} &= y - X\widehat{\beta}_{MCO} = y - X\left(X'X\right)^{-1}X'y = (I - X\left(X'X\right)^{-1}X')y = M_X y \\ &= \underset{car M_X X = 0}{M_X (X\beta + \varepsilon)} = M_X\varepsilon \end{aligned}$$

  Montrons que $\widehat{\sigma}^2_{MCO}$ est un estimateur centré de σ^2, calculons l'espérance mathématique de SS :

$$E(SS) = E\left(tr\varepsilon' M_X \varepsilon\right) = E\left(\mathrm{tr}\, M_X\varepsilon\varepsilon'\right) = \mathrm{tr}\, M_X E\left(\varepsilon\varepsilon'\right) = \sigma^2\, \mathrm{tr}\, M_X = \sigma^2 (T-k)$$

  Ce qui implique que $\widehat{\sigma}^2_{MCO} = \frac{SS}{T-k}$ est un estimateur centré de σ^2 car $E\left(\widehat{\sigma}^2_{MCO}\right) = \sigma^2$.
  La matrice des variances-covariances estimée de $\widehat{\beta}_{MCO}$ s'écrit alors :

$$\widehat{V}(\widehat{\beta}_{MCO}) = \widehat{\sigma}^2_{MCO}\left(X'X\right)^{-1}$$

Remarque 1 En échantillon de petite taille, sous l'hypothèse supplémentaire de normalité des erreurs, on montre que :

- $\widehat{\beta}_{MCO} \sim N(\beta, \sigma^2(X'X)^{-1})$.
- $SS/\sigma^2 \sim \chi^2_{n-k}$, indépendant de $\widehat{\beta}_{MCO}$.
- $\widehat{\beta}_{MCO}$ est l'estimateur du maximum de vraisemblance. Il atteint la borne inférieure de l'inégalité de Cramer-Rao. Il est donc efficient parmi tous les estimateurs centrés (linéaires ou non) sous l'hypothèse de normalité des erreurs.
- $\widehat{\sigma}^2_{MCO}$ est proportionnel à l'estimateur du maximum de vraisemblance : $\widehat{\sigma}^2_{MV} = \frac{SS}{n}$ qui est biaisé, et $\widehat{\sigma}^2_{MCO} = \frac{n}{n-k}\widehat{\sigma}^2_{MV}$. $\widehat{\sigma}^2_{MCO}$ n'atteint pas la limite inférieure de l'inégalité de Cramer-Rao. On montre cependant que $\widehat{\sigma}^2_{MCO}$ est efficient parmi tous les estimateurs centrés quadratiques : il est BQUE (Best Quadratic Unbiased Estimator) sous l'hypothèse de normalité des erreurs.

2 Convergence stochastique

2.1 Estimateur convergent en probabilité (consistent estimator)

Définition 1 Soit $\hat{\theta}_n$ un estimateur de θ fondé sur un échantillon de taille n, alors la suite d'estimateurs $\hat{\theta}_n$ est convergente si elle converge en probabilité vers θ, c'est-à-dire si, quels que soient les nombres δ et ε arbitrairement petits, il est possible de déterminer une taille d'échantillon n_0 telle que :

$$\forall\delta > 0, \forall\varepsilon > 0, \exists n_0 \text{ tel que } \forall n > n_0 \Longrightarrow \Pr\left\{\left|\hat{\theta}_n - \theta\right| < \varepsilon\right\} > 1 - \delta \tag{8}$$

Ceci signifie qu'en augmentant la taille de l'échantillon, on peut rendre l'estimateur $\hat{\theta}_n$ aussi proche de la vraie valeur de θ qu'on le souhaite avec une probabilité égale à 1. Nous pouvons aussi écrire :

$$\forall\varepsilon > 0, \lim_{n\longrightarrow\infty} \Pr\left[\left|\hat{\theta}_n - \theta\right| < \varepsilon\right] = 1 \tag{9}$$

On dit alors que $\hat{\theta}_n$ converge en probabilité vers θ et on note $\hat{\theta}_n \xrightarrow{p} \theta$ ou $\operatorname{plim}\hat{\theta}_n = \theta$. On dit aussi que $\hat{\theta}_n$ est un estimateur convergent en probabilité de θ.

Propriété 1
Supposons que $\operatorname{plim} S_n = \gamma$ *et* $\operatorname{plim} T_n = \theta$*, avec* γ *et* θ *des constantes finies.*

- $\operatorname{plim}\gamma = \gamma$ *: le plim d'une constante est égale à cette constante.*
- $\operatorname{plim} S_n \pm T_n = \operatorname{plim} S_n \pm \operatorname{plim} T_n = \gamma \pm \theta$
- $\operatorname{plim} S_n T_n = \operatorname{plim} S_n . \operatorname{plim} T_n = \gamma\theta$
- $\operatorname{plim} \dfrac{S_n}{T_n} = \dfrac{\operatorname{plim} S_n}{\operatorname{plim} T_n} = \dfrac{\gamma}{\theta}$ *avec* $\theta \neq 0$
- *Théorème de continuité (Slutsky) :*
  Pour toute fonction continue g *qui n'est pas fonction de* n *on a :* $\operatorname{plim} g(S_n) = g(\operatorname{plim} S_n) = g(\gamma)$.

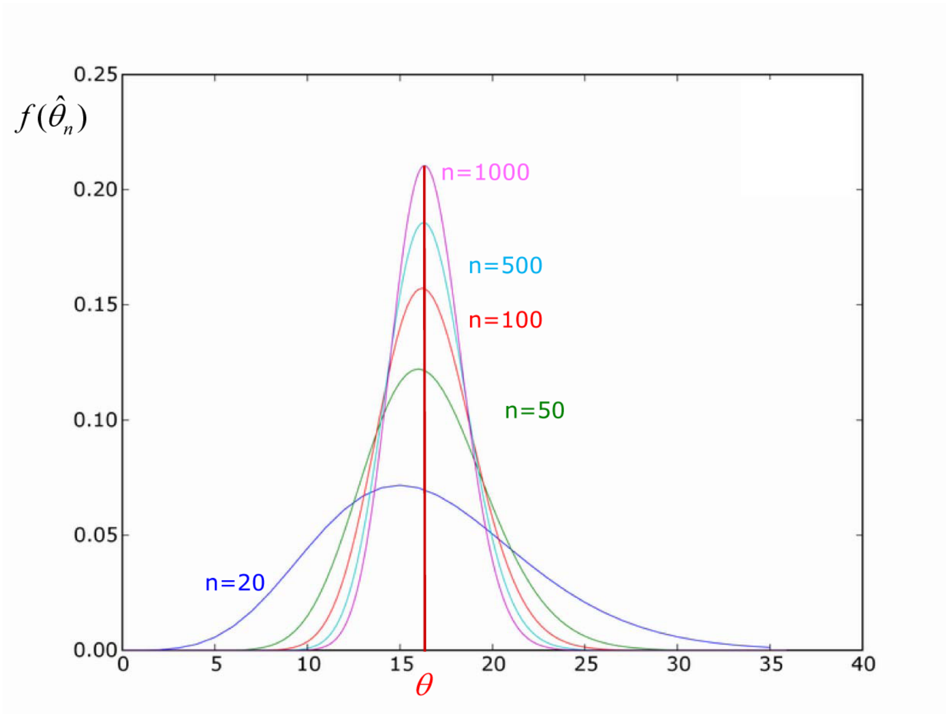

FIGURE 1 – Convergence en moyenne quadratique et convergence en probabilité

2.2 Estimateur convergent en moyenne quadratique

Définition 2 Soit $\hat{\theta}_n$ un estimateur de θ fondé sur un échantillon de taille n, $\hat{\theta}_n$ est un estimateur convergent en moyenne quadratique de θ si les deux premiers moments de$\hat{\theta}_n$ existent et sont finis : $E(\hat{\theta}_n) < \infty$ et $E(\hat{\theta}_n^2) < \infty$ et si $\lim\limits_{n \longrightarrow \infty} E\left[(\hat{\theta}_n - \theta)^2\right] = 0$.

Nous avons vu que l'erreur quadratique moyenne (mean square error, MSE) s'écrit :

$$E(\hat{\theta}_n - \theta)^2 = V(\hat{\theta}_n) + B^2(\hat{\theta}_n)$$

par conséquent :

$$\lim_{n \longrightarrow \infty} E\left[(\hat{\theta}_n - \theta)^2\right] = 0 \qquad \Longleftrightarrow \qquad \lim_{n \longrightarrow \infty} V(\hat{\theta}_n) = 0 \text{ et } \lim_{n \longrightarrow \infty} B(\hat{\theta}_n) = 0$$

Un estimateur est convergent en moyenne quadratique s'il est asymptotiquement sans biais et que sa variance tend vers zéro lorsque la taille de l'échantillon tend vers l'infini et on note $\hat{\theta}_n \overset{m.s.}{\longrightarrow} \theta$ ou $\hat{\theta}_n \overset{m.q.}{\longrightarrow} \theta$.

Une condition suffisante pour que $\hat{\theta}_n$ soit un estimateur convergent en probabilité est qu'il soit convergent en moyenne quadratique, c'est-à-dire que son biais et sa variance tendent tous les deux vers zéro lorsque la taille de l'échantillon tend vers l'infini. Cette condition est simple à vérifier en pratique, mais ce n'est pas une condition nécessaire. Un estimateur peut être convergent en probabilité même si son biais ne tend pas vers zéro.

Théorème 2
La convergence en moyenne quadratique implique la convergence en probabilité, la réciproque est fausse :

$$\hat{\theta}_n \xrightarrow{m.s.} \theta \qquad \Longrightarrow \qquad \hat{\theta}_n \xrightarrow{p} \theta$$

Exemple 1 Soit une population caractérisée par une variable aléatoire X. Supposons que dans la population $E(X) = \mu$ et $V(X) = \sigma^2$.

Soit un échantillon aléatoire de taille n. Soit $\overline{X}_n$ la moyenne de l'échantillon : $\overline{X}_n = \sum_{i=1}^{n} X_i$

Nous avons vu que :

- $E(\overline{X}_n) = \mu$ estimateur centré, sans biais de μ : $B(\overline{X}_n) = 0$
- $V(\overline{X}_n) = \frac{\sigma^2}{n}$ et $\lim\limits_{n \longrightarrow \infty} V(\overline{X}_n) = 0$

Par conséquent $\overline{X}_n$ est un estimateur convergent en moyenne quadratique de μ, ce qui implique que $\overline{X}_n$ est un estimateur convergent de μ.

2.3 Convergence en loi ou en distribution

Définition 3 Soit $X_1, X_2, ..., X_n$ une suite de variable aléatoires de fonctions de répartition $F_1, F_2, ..., F_n$.

On dit que la suite $\{X_n\}$ converge en loi ou en distribution vers la variable aléatoire X de fonction de répartition F si :

$$\lim_{n \longrightarrow \infty} |F_n(x) - F(x)| = 0$$

en tout point de continuité de F.

On note $X_n \xrightarrow{d} X$ ou $X_n \xrightarrow{L} X$. F est la fonction de répartition limite. La distribution limite est souvent donnée en terme de fonction de densité de probabilité : par exemple $X_n \xrightarrow{d} N(0, 1)$: la distribution limite de X_n normal centrée réduite.

Théorème 3
La convergence en probabilité implique la convergence en loi.

$$X_n \xrightarrow{p} X \qquad \Longrightarrow \qquad X_n \xrightarrow{L} X$$

La réciproque est fausse sauf si X est égale à une constante c, on a alors :

$$X_n \xrightarrow{L} c \qquad \Longrightarrow \qquad X_n \xrightarrow{p} c$$

Théorème 4 (Théorème de Slutsky)
Soit la suite $\{X_n\}$ de variables aléatoires telle que $X_n \xrightarrow{L} X$, soit la suite $\{Y_n\}$ de variables aléatoires telle que $Y_n \xrightarrow{p} c$

1. $X_n + Y_n \xrightarrow{L} X + c$
2. $X_n . Y_n \xrightarrow{L} cX$
3. $X_n / Y_n \xrightarrow{L} X/c$
4. *Généralisation : $X_n \xrightarrow{L} X$ et g est une fonction continue qui ne dépend pas de la taille de l'échantillon n, alors : $g(X_n) \xrightarrow{L} g(X)$.*

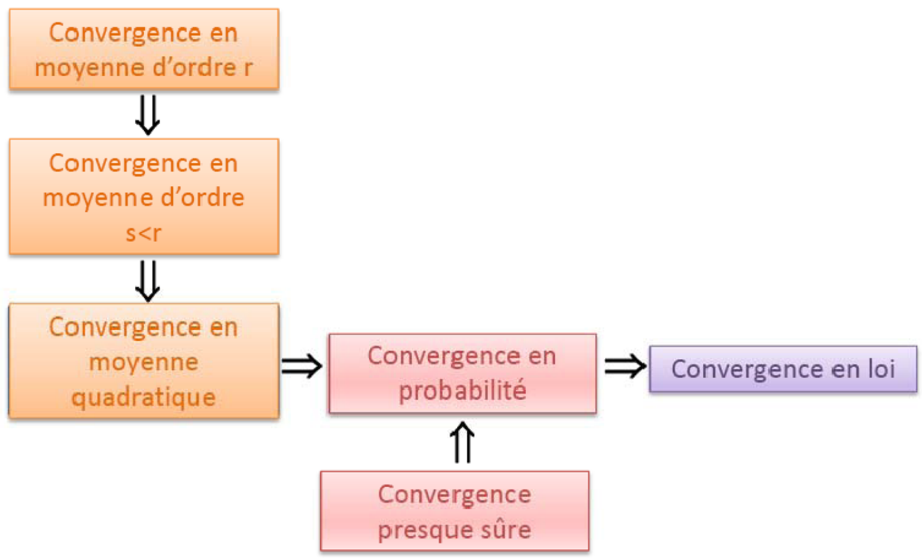

FIGURE 2 – Relation entre les différents modes de convergence

2.4 Loi faible des grands nombres

Théorème 5 (Théorème de Khintchine)
Soit $Y_1, Y_2, ..., Y_n$ une suite de variables aléatoires identiquement, indépendamment distribuées telles que $\mathrm{E}(Y_i)$ existe pour tout i. Alors :

$$\mathrm{E}(Y_i) = \mu < \infty \Longrightarrow \bar{Y}_n = \frac{1}{n}\sum Y_i \xrightarrow{p} \mu$$

Théorème 6 (Convergence des moments d'un échantillon aléatoire)
Soit une population caractérisée par une variable aléatoire X. On distingue les moments simples d'ordre r : $\mathrm{E}(X^r) = m_r$ et les moments centrés d'ordre r : $\mathrm{E}[(X-\mu)^r] = \mu_r$ de la population, où r

Soit un échantillon aléatoire $X_1, X_2, ..., X_n$, c'est-à-dire une suite de variables aléatoires identiquement, indépendamment distribuées telles que $\mathrm{E}(Y_i)$ existe pour tout i. Les moments simples et centrés d'ordre r de l'échantillon sont respectivement définis de la manière suivante :

$$\frac{1}{n}\sum X_i^r \qquad \text{et} \qquad \frac{1}{n}\sum (X_i - \bar{X})^r$$

Alors on montre que les moments simples et centrés d'ordre r de l'échantillon convergent en probabilité vers les moments simples et centrés d'ordre r de la population :

$$\operatorname{plim} \frac{1}{n}\sum X_i^r = \mathrm{E}(X_i^r) = m_r \tag{10}$$

$$\operatorname{plim} \frac{1}{n}\sum (X_i - \bar{X})^r = \mathrm{E}[(X-\mu)^r] = \mu_r \tag{11}$$

3 Théorèmes Central-Limit

Théorème 7 (Théorème de Lindberg-Lévy)
Soit $\{Y_n\}$ une suite de variables aléatoires, indépendantes et identiquement distribuées telle que $\mathrm{E}(Y_i) = \mu$ et $\mathrm{V}(Y_i) = \sigma^2$ finie pour tout $i = 1, ..., n$. Calculons la l'espérance mathématique et la variance de la somme de ces n variables aléatoires i.i.d.(μ, σ^2) :

$$\mathrm{E}(\sum_i Y_i) = \sum_i E(Y_i) = n\mu$$

$$\mathrm{V}(\sum_i Y_i) = \sum_i V(Y_i) = n\sigma^2$$

Définissons la somme standardisée de ces n variables aléatoires :

$$Z_n = \frac{\sum_{i=1}^n Y_i - n\mu}{\sqrt{n}\sigma} = \frac{\sum_{i=1}^n Y_i - \mu)}{\sqrt{n}\sigma} = \frac{1/n\sum_{i=1}^n Y_i - \mu)}{\sigma/\sqrt{n}} = \frac{\bar{Y}_n - \mu}{\sigma/\sqrt{n}} = \sqrt{n}\,\frac{(\bar{Y}_n - \mu)}{\sigma}$$

Alors Z_n converge en loi vers la loi normale centrée réduite :

$$Z_n \xrightarrow{L} N(0,1)$$

Théorème 8 (Théorème de Liapounov)
Spanos A. (Statistical Foundations of Econometric Modelling, 1986, p. 174)
Soit $\{Y_n\}$ une suite de variables aléatoires, indépendantes et identiquement distribuées telle que $\mathrm{E}(Y_i) = \mu_i$, $\mathrm{V}(Y_i) = \sigma_i^2 < \infty$ et $\mathrm{E}(|Y_i|^{2+\delta}) < \infty, \delta > 0$ pour tout $i = 1, ..., n$. Définissons $c_n = (\sum_{i=1}^n \sigma_i^2)^{1/2}$. Si, de plus :

$$\lim_{n\longrightarrow\infty} \frac{1}{c_n^{2+\delta}} \sum_{i=1}^{\infty} \mathrm{E}(|Y_i - \mu_i|^{2+\delta}) = 0$$

Alors la somme standardisée Z_n converge en loi vers la loi normale centrée réduite :

$$Z_n \xrightarrow{L} N(0,1)$$

Théorème 9 (Théorème de Lindberg-Lévy / multidimensionnel)
Soit $\{Y_n\}$ une suite de vecteurs aléatoires de dimension $m \times 1$, indépendants et identiquement distribués d'espérance mathématique $\mathrm{E}(Y_i) = \mu$ et de matrice de variance-covariance Ω, définie positive.
Alors la somme standardisée Z_n converge en loi vers la loi normale centrée réduite :

$$Z_n \xrightarrow{L} N(0, \Omega)$$

où $Z_n = \sqrt{n}(\bar{Y}_n - \mu)$ et $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$ sont des vecteurs aléatoires de dimension $m \times 1$.

Théorème 10 (Théorème Central-Limit généralisé de Malinvaud / unidimensionel)
Malinvaud E. (Méthodes statistiques de l'économétrie, 1974, p. 269-271)
Soit $\{Y_n\}$ une suite de variables aléatoires, indépendantes et identiquement distribuées

telle que $\mathrm{E}(Y_i) = 0$, $\mathrm{V}(Y_i) = \sigma^2 < \infty$.
Définissons la variable standardisée :

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} v_i Y_i$$

où $v_i, i-1, ..., n$ *sont des scalaires uniformément bornés, tels que* $\lim_{n \longrightarrow \infty} \frac{\sigma^2}{n} \sum_{i=1}^{n} v_i^2 = q$ *avec* $0 < q < \infty$.
Alors le vecteur standardisé Z_n *converge en loi vers la loi normale centrée réduite :*

$$Z_n \xrightarrow{L} N(0,1)$$

Théorème 11 (Théorème Central-Limit généralisé de Malinvaud / multidimensionel)
Malinvaud E. (Méthodes statistiques de l'économétrie, 1974, p. 269-271)
Soit $\{Y_n\}$ *une suite de vecteurs aléatoires de dimension* $m \times 1$*, indépendants et identiquement distribués d'espérance mathématique* $\mathrm{E}(Y_i) = 0$ *et de matrice de variance-covariance* Ω*, définie positive.*
Définissons le vecteur aléatoire standardisé Z_n *de dimension* $k \times 1$ *:*

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_i Y_i$$

où $A_i, i = 1, ..., n$ *est une matrice de dimension* $k \times m$ *composé d'éléments certains uniformément bornés tels que :*

$$\lim_{n \longrightarrow \infty} \frac{\sigma^2}{n} \sum_{i=1}^{n} A_i \Omega A_i' = Q$$

où Q *est matrice de dimension* $k \times k$*, définie positive. Alors le vecteur standardisé* Z_n *converge en loi vers la loi normale centrée réduite :*

$$Z_n \xrightarrow{L} N(0,Q)$$

Théorème 12 (Théorème sur le produit de deux suites)
Soit $\{Y_n\}$ *une suite de vecteurs aléatoires de dimension* $m \times 1$*, soit* $\{W_n\}$ *une suite de matrice d'éléments certains ou aléatoires de dimension* $k \times m$.
Supposons que :

$$Y_n \xrightarrow{L} Y \text{ et } W_n \xrightarrow{p} W$$

où W *est une matrice de constantes finies de rang* $k \leq m$*. Alors la suite* $W_n Y_n$ *converge en loi vers le vecteur aléatoire* WY.

Exemple 2 Supposons que $Y_n \xrightarrow{L} Y \sim N(\mu, \Omega)$ et $W_n \xrightarrow{p} W$ alors $W_n Y_n \xrightarrow{L} N(W\mu, W\Omega W')$.

4 Propriétés asymptotiques des estimateurs des MCO

Propriété 13 (Convergence en probabilité)
Sous les hypothèses H_1, H_2 avec $\lim_{n\to\infty}\frac{1}{n}X'X = Q$, matrice définie positive (ou H_2' avec $\text{plim}_{n\to\infty}\frac{1}{n}X'X = Q$, matrice définie positive) et H_{3a} on montre que l'estimateur des MCO est convergent en probabilité, de même que l'estimateur de la variance résiduelle $\hat{\sigma}^2$:

$$\text{plim}\,\hat{\beta}_{MCO} = \beta$$
$$\text{plim}\,\hat{\sigma^2}_{MCO} = \sigma^2$$

On montre également que ces estimateurs sont convergents en moyenne quadratique.

Démonstration.

- On calcule la limite en probabilité de $\hat{\beta}_{MCO}$:

$$\text{plim}\,\hat{\beta}_{MCO} = \beta + \text{plim}\left[\left(\frac{1}{n}X'X\right)^{-1}\cdot\frac{1}{n}X'\varepsilon\right] = \beta + \left(\frac{1}{n}X'X\right)^{-1}\cdot\text{plim}\left[\frac{1}{n}X'\varepsilon\right] \quad (12)$$

or on sait que par hypothèse $\text{plim}\left(\frac{1}{n}X'X\right)^{-1} = Q^{-1}$ et $\text{plim}\left(\frac{1}{n}X'\varepsilon\right) = 0$. Par conséquent $\text{plim}\,\hat{\beta}_{MCO} = \beta$. L'estimateur des MCO de β est convergent en probabilité.

- On calcule la limite en probabilité de σ^2 :

$$\hat{\sigma}^2_{MCO} = \frac{SS}{n-k} = \frac{1}{n-k}\varepsilon' M_X \varepsilon = \frac{n}{n-k}\cdot\frac{1}{n}\varepsilon'(I_n - X(X'X)^{-1}X')\varepsilon \quad (13)$$
$$\hat{\sigma}^2_{MCO} = \frac{n}{n-k}\cdot\frac{1}{n}\varepsilon'\varepsilon - \frac{n}{n-k}\cdot\frac{1}{n}\varepsilon' X(X'X)^{-1}X'\varepsilon \quad (14)$$

$$\text{plim}\,\hat{\sigma}^2_{MCO} = \underbrace{\text{plim}\left(\frac{n}{n-k}\right)}_{=1}\cdot\,\text{plim}\left(\frac{1}{n}\varepsilon'\varepsilon\right) \quad (15)$$
$$-\underbrace{\text{plim}\left(\frac{n}{n-k}\right)}_{=1}\cdot\underbrace{\text{plim}\left(\frac{1}{n}\varepsilon' X\right)}_{=0}\cdot\underbrace{\text{plim}\left(\frac{1}{n}X'X\right)^{-1}}_{Q^{-1}}\cdot\underbrace{\text{plim}\left(\frac{1}{n}X'\varepsilon\right)}_{=0} \quad (16)$$
$$\text{plim}\,\hat{\sigma}^2_{MCO} = \text{plim}\left(\frac{1}{n}\varepsilon'\varepsilon\right) = \text{plim}\left(\frac{1}{n}\sum\varepsilon_i^2\right) \quad (17)$$

On applique la loi faible des grands nombres (théorème de Khintchine) à la suite de variables aléatoires $\varepsilon_1^2, \varepsilon_2^2, ..., \varepsilon_n^2$ identiquement, indépendamment distribuée d'espérance mathématique $\text{E}(\varepsilon_i^2) = \sigma^2$ finie :

$$\text{E}(\varepsilon_i^2) = \sigma^2 < \infty \Longrightarrow \frac{1}{n}\sum\varepsilon_i^2 \xrightarrow{p} \sigma^2$$

Par conséquent, on obtient $\text{plim}\,\hat{\sigma}^2_{MCO} = \sigma^2$. L'estimateur des MCO de σ^2 est convergent en probabilité.

Propriété 14 (Distribution asymptotique)
Sous les hypothèses H_1, H_2 avec $\lim_{n\to\infty}\frac{1}{n}X'X = Q$, matrice définie positive (ou H_2' avec $\text{plim}_{n\to\infty}\frac{1}{n}X'X = Q$, matrice définie positive) et H_{3b}, on montre que l'estimateur des MCO est asymptotiquement normalement distribué, de même que l'estimateur de la variance résiduelle $\hat{\sigma}^2$:

$$\sqrt{n}(\hat{\beta}_{MCO}-\beta) \xrightarrow{d} N(0,\sigma^2 Q^{-1}) \qquad V_{AS}(\hat{\beta}_{MCO}) = \frac{\sigma^2}{n}Q^{-1}$$
$$\sqrt{n}(\hat{\sigma}^2_{MCO}-\sigma^2) \xrightarrow{d} N(0,\mu_4-\sigma^4) \qquad V_{AS}(\hat{\sigma^2}_{MCO}) = \frac{\mu_4-\sigma^4}{n}$$

où $\mu_4 = \mathrm{E}(\varepsilon_i^4)$, le moment d'ordre 4 du terme d'erreur. On dira que $\hat{\beta}_{MCO}$ est asymptotiquement normal de variance asymptotique $\frac{\sigma^2}{n}Q^{-1}$ et que $\hat{\sigma}^2_{MCO}$ est asymptotiquement normal de variance asymptotique $\frac{\mu_4-\sigma^4}{n}$.

Sous l'hypothèse supplémentaire de Normalité H_{3d}, on a $\mu^4 = 3\sigma^4$ et $\hat{\beta}_{MCO}$ et $\hat{\sigma}^2_{MCO}$ ont des variances asymptotiques qui atteignent la borne inférieure de l'inégalité de Cramer-Rao. Ils sont donc asymptotiquement efficients.

Démonstration.

- On calcule la distribution asymptotique de $\hat{\beta}_{MCO}$:

$$\hat{\beta}_{MCO} = \beta + (X'X)^{-1}X'\varepsilon \tag{18}$$

La distribution asymptotique s'obtient en calculant la distribution limite de l'estimateur centré multiplié par le facteur de normalisation qui est ici égal à $\sqrt{n}$:

$$\sqrt{n}(\hat{\beta}_{MCO}-\beta) = \left(\frac{1}{n}X'X\right)^{-1}\cdot\frac{1}{\sqrt{n}}X'\varepsilon \tag{19}$$

Cette expression correspond au produit de deux suites : on sait que plim $\left(\frac{1}{n}X'X\right)^{-1} = Q^{-1}$ par hypothèse. Il nous faut donc calculer la distribution limite de $\frac{1}{\sqrt{n}}X'\varepsilon$ en utilisant le théorème Central-Limit généralisé de Malinvaud :

$$\frac{1}{\sqrt{n}}X'\varepsilon = \frac{1}{\sqrt{n}}(x_1\varepsilon_1+...+x_i\varepsilon_i+...+x_n\varepsilon_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i\varepsilon_i \tag{20}$$

où x_i est la $i^{\text{ème}}$ colonne de la matrice X et donc x_i' est la $i^{\text{ème}}$ ligne de cette matrice.
1 $\varepsilon_i \sim i.i.d.(0,\sigma^2)$ par hypothèse
2 les x_i sont fixes, non aléatoires, uniformément bornés tel que :

$$\frac{1}{n}\sum x_i V(\varepsilon)x_i' = \frac{\sigma^2}{n}\sum x_i x_i' = \sigma^2\frac{1}{n}X'X \tag{21}$$

en passant aux limites, on obtient :

$$\lim_{n\longrightarrow\infty}\sigma^2\frac{1}{n}X'X = \sigma^2\lim_{n\longrightarrow\infty}\frac{1}{n}X'X = \sigma^2 Q \text{ définie positive} \tag{22}$$

par conséquent les deux conditions du théorème Central-Limit généralisé de Malinvaud sont satisfaites et on a :

$$\frac{1}{\sqrt{n}}X'\varepsilon \xrightarrow{d} N(0,\sigma^2 Q) \tag{23}$$

On applique maintenant le théorème sur le produit de deux suites :

$$\sqrt{n}(\hat{\beta}_{MCO} - \beta) \xrightarrow{d} N(0, Q^{-1}\sigma^2 Q Q^{-1}) \tag{24}$$

$$\sqrt{n}(\hat{\beta}_{MCO} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1}) \tag{25}$$

d'où nous déduisons que $\hat{\beta}_{MCO}$ est asymptotiquement normalement distribué d'espérance mathématique β et de matrice de variance-covariance $\frac{\sigma^2}{n}Q^{-1}$. En effet, nous pouvons donc écrire la vraie matrice de variance-covariance asymptotique :

$$V_{as}(\hat{\beta}_{MCO}) = \frac{\sigma^2}{n}Q^{-1} \tag{26}$$

qui sera estimée par :

$$\hat{V}_{as}(\hat{\beta}_{MCO}) = \frac{\hat{\sigma}^2}{n}\left(\frac{1}{n}X'X\right)^{-1} = \hat{\sigma}^2(X'X)^{-1} \tag{27}$$

avec $\hat{\sigma}^2 = SS/(n-k)$.
Notons que la variance asymptotique de $\hat{\beta}_{MCO}$ atteint la borne inférieure de l'inégalité de Cramer-Rao (cf. Rappels Econométrie, p.24). Cet estimateur est donc asymptotiquement efficient, on dit parfois qu'il est BAN (Best Asymptotically Normal).

- On calcule la distribution asymptotique de σ^2_{MCO}, exprimons l'estimateur centré multiplié par le facteur de normalisation $\sqrt{n}$:

$$\begin{aligned}
\sqrt{n}(\hat{\sigma}^2_{MCO} - \sigma^2) &= \frac{\sqrt{n}}{(n-k)}\varepsilon' M_X \varepsilon - \sqrt{n}\sigma^2 \qquad (28)\\
&= \frac{\sqrt{n}}{(n-k)}\varepsilon'\varepsilon - \frac{\sqrt{n}}{(n-k)}\varepsilon' X(X'X)^{-1}X'\varepsilon - \sqrt{n}\sigma^2\\
&= \frac{n}{(n-k)}\frac{1}{\sqrt{n}}\varepsilon'\varepsilon - \sqrt{n}\sigma^2 - \frac{n}{(n-k)}\frac{1}{\sqrt{n}}\varepsilon' X\left(\frac{1}{n}X'X\right)^{-1}\frac{1}{n}X'\varepsilon\\
&= \frac{n}{(n-k)}\frac{1}{\sqrt{n}}\varepsilon'\varepsilon - \frac{(n-k)\sqrt{n}}{(n-k)}\sqrt{n}\sigma^2 - \frac{n}{(n-k)}\frac{1}{\sqrt{n}}\varepsilon' X\left(\frac{1}{n}X'X\right)^{-1}\frac{1}{n}X'\varepsilon\\
&= \frac{n}{(n-k)}\frac{1}{\sqrt{n}}\varepsilon'\varepsilon - \frac{n\sqrt{n}}{(n-k)}\sigma^2 + \frac{k\sqrt{n}}{(n-k)}\sigma^2 - \frac{n}{(n-k)}\frac{1}{\sqrt{n}}\varepsilon' X\left(\frac{1}{n}X'X\right)^{-1}\frac{1}{n}X'\varepsilon\\
&= \frac{n}{(n-k)}\frac{1}{\sqrt{n}}(\varepsilon'\varepsilon - n\sigma^2) + \frac{k\sqrt{n}}{(n-k)}\sigma^2 - \frac{n}{(n-k)}\frac{1}{\sqrt{n}}\varepsilon' X\left(\frac{1}{n}X'X\right)^{-1}\frac{1}{n}X'\varepsilon\\
&= \frac{n}{(n-k)}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\varepsilon_i^2 - \sigma^2) + \frac{k\sqrt{n}}{(n-k)}\sigma^2 - \frac{n}{(n-k)}\frac{1}{\sqrt{n}}\varepsilon' X\left(\frac{1}{n}X'X\right)^{-1}\frac{1}{n}X'\varepsilon
\end{aligned}$$

On note que :

$$\lim_{n\longrightarrow\infty}\frac{n}{n-k} = 1 \qquad \lim_{n\longrightarrow\infty}\frac{k\sqrt{n}}{(n-k)} = 0$$

$$\text{plim}\left(\frac{1}{n}X'\varepsilon\right) = 0 \qquad \text{plim}\left(\frac{1}{n}X'X\right)^{-1} = Q^{-1}$$

La distribution limite de $\sqrt{n}(\hat{\sigma}^2_{MCO} - \sigma^2)$ est donc celle de $\frac{1}{\sqrt{n}} \sum v_i$ avec $v_i = \varepsilon_i^2 - \sigma^2$. Or :

$$\begin{aligned}
\mathrm{E}(v_i) &= \mathrm{E}(\varepsilon_i^2 - \sigma^2) = 0 \\
\mathrm{V}(v_i) &= \mathrm{E}(v_i^2) = \mathrm{E}(\varepsilon_i^4 - 2\varepsilon_i^2\sigma^2 + \sigma^4) \\
\mathrm{V}(v_i) &= \mathrm{E}(\varepsilon_i^4) - 2\sigma^2\,\mathrm{E}(\varepsilon_i^2) + \sigma^4 \\
\mathrm{V}(v_i) &= \mathrm{E}(\varepsilon_i^4) - 2\sigma^4 + \sigma^4 \\
\mathrm{V}(v_i) &= \mathrm{E}(\varepsilon_i^4) - \sigma^4
\end{aligned}$$

Il faut que le moment d'ordre 4 existe : si $\mathrm{E}(\varepsilon_i^4) = \mu_4 < \infty$ alors $v_i \sim i.i.d.(0, \mu_4 - \sigma^4)$. On applique alors le théorème Central-Limit de Lindeberg-Levy :

$$\frac{1}{\sqrt{n}} \sum v_i = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\varepsilon_i^2 - \sigma^2) \xrightarrow{d} N(0, \mu_4 - \sigma^4) \tag{29}$$

et donc :

$$\sqrt{n}(\hat{\sigma}^2_{MCO} - \sigma^2) \xrightarrow{d} N(0, \mu_4 - \sigma^4) \tag{30}$$

d'où nous déduisons que $\hat{\sigma}^2_{MCO}$ est asymptotiquement normalement distribué d'espérance mathématique σ^2 et de variance $\frac{\mu_4 - \sigma^4}{n}$. En effet, nous pouvons donc écrire la vraie variance asymptotique :

$$V_{as}(\hat{\sigma}^2_{MCO}) = \frac{\mu_4 - \sigma^4}{n} \tag{31}$$

Si de plus, on suppose la normalité des erreurs, on a $\mu^4 = 3\sigma^4$ et :

$$\sqrt{n}(\hat{\sigma}^2_{MCO} - \sigma^2) \xrightarrow{d} N(0, 2\sigma^4) \tag{32}$$

et

$$V_{as}(\hat{\sigma}^2_{MCO}) = \frac{2\sigma^4}{n} \tag{33}$$

Notons que la variance asymptotique de $\hat{\sigma}^2_{MCO}$ atteint la borne inférieure de l'inégalité de Cramer-Rao (cf. Rappels Econométrie, p.24). Cet estimateur est donc asymptotiquement efficient, on dit parfois qu'il est BAN (Best Asymptotically Normal).

5 Méthode Delta : test asymptotique d'un ensemble de restrictions non linéaires