

Chapitre 1

Les Processus Aléatoires Stationnaires et les Processus ARMA

Le but de ce chapitre est d'introduire la notion de processus temporel et plus particulièrement la classe des processus *ARMA* qui sont particulièrement utiles pour décrire le comportement des séries temporelles univariées. Cette présentation suppose que l'on définisse au préalable un certain nombre de notions essentielles à l'analyse des séries temporelles, et en particulier la notion de stationnarité. L'étude des séries temporelles suppose que l'on fasse au préalable un certain nombre de rappels en probabilité et en statistiques.

1 Rappels de Probabilité et de Statistiques

Avant de définir la notion de série temporelle, il convient de faire un certain nombre de rappels succincts. Les rappels proposés dans le cadre de cette section portent avant tout sur les probabilités et les statistiques usuelles. Toutefois, la lecture de ces rappels doit nécessairement s'accompagner d'une étude plus systématique des fondements probabilistes des méthodes statistiques (*cf.* "Méthodes Statistiques", Philippe Tassi, *Economica* 1989).

On considère une *variable stochastique réelle* ou *variable aléatoire réelle* (*v.a.r.* en abrégé) *continue*, notée X , dont la loi de probabilité, pour une réalisation x , est définie par la fonction de densité $f_X(x)$, supposée *continue*, telle que $f_X(x) \geq 0$:

$$\forall (a, b) \in \mathbb{R}^2 \quad P(a \leq X \leq b) = \int_a^b f_X(x) dx \quad (1.1)$$

$$\int_0^1 f_X(x) dx = 1 \quad (1.2)$$

On note $F_X(\cdot)$ la *fonction de répartition* ou *fonction de distribution* cumulative associée à X , telle que :

$$F_X(a) \equiv P(X \leq a) \equiv P(X < a) = \int_{-\infty}^a f_X(x) dx \quad (1.3)$$

Nous allons à présent successivement proposer des rappels sur les notions de population des moments, de moments empiriques, distribution jointe, de distribution conditionnelle, d'indépendance, de convergence, ainsi que des rappels sur les propriétés d'un certain nombre de lois usuels.

1.1 La population des moments théoriques

Pour une variable aléatoire réelle (v.a.r.) X , l'espérance et la variance constituent deux moments particuliers, mais plus généralement il est possible, sous certaines hypothèses, de définir la population des moments et la population des moments centrés de la façon suivante :

Definition 1 Pour une variable aléatoire réelle continue X , de densité $f_X(\cdot)$, la population des moments associée à tout ordre $k \in \mathbb{N}$, est définie par :

$$E(X^k) = \int_{-\infty}^{\infty} x^k f_X(x) dx \quad (1.4)$$

La population des moments centrés associée à tout ordre $k \in \mathbb{N}$, notée μ_k , est définie par :

$$\mu_k = E[(X - \mu)^k] = \int_{-\infty}^{\infty} (x - \mu)^k f_X(x) dx \quad (1.5)$$

avec $\mu = E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$.

Ainsi l'espérance, notée $E(X)$, correspond au moment d'ordre un ($k = 1$) :

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \mu$$

De la même façon, la variance, notée $V(X)$, est définie par le moment centré d'ordre deux ($k = 2$) :

$$V(X) = \mu_2 = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx$$

En général, lorsque l'on cherche à établir la population des moments d'une transformée, linéaire ou non, de la v.a.r. X , on utilise la propriété suivante :

Propriété 1. On considère une v.a.r. transformée, telle que $Y = g(X)$. Soit $f_X(\cdot)$ la fonction de densité de la v.a.r. X . La population des moment et la population des moments centrés d'ordre k , de la transformée $g(X)$ sont alors définies par l'espérance respective des transformées $[g(X)]^k$ et $[g(X) - \mu^Y]^k$, avec $\mu^Y = E[g(x)]$, $\forall k \in \mathbb{N}$:

$$E\{[g(X)]^k\} = \int_{-\infty}^{\infty} [g(x)]^k f_X(x) dx \quad (1.6)$$

$$\mu_k^Y = E\{[g(X) - \mu^Y]^k\} = \int_{-\infty}^{\infty} [g(x) - \mu^Y]^k f_X(x) dx \quad (1.7)$$

En particulier, l'espérance de Y correspond à :

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx \quad (1.8)$$

Considérons l'exemple suivant. Soit Y la transformée linéaire de X , telle que $Y = a + bX$, $(a, b) \in \mathbb{R}^2$. Le moment d'ordre un de Y , est défini par :

$$\begin{aligned} E(a + bX) &= \int_{-\infty}^{\infty} (a + bx) f_X(x) dx \\ &= a \int_{-\infty}^{\infty} f_X(x) dx + b \int_{-\infty}^{\infty} x f_X(x) dx \\ &= a + bE(X) \end{aligned}$$

On retrouve ainsi le résultat $E(Y) = a + bE(X)$, selon lequel l'espérance est un opérateur linéaire. De la même façon, on peut calculer la variance de la variable transformée Y , notée $V(Y)$, qui correspond au moment centré d'ordre deux. On note μ , avec $\mu \in \mathbb{R}$, l'espérance de la variable X .

$$\begin{aligned} V(a + bX) &= \int_{-\infty}^{\infty} [(a + bx) - (a + b\mu)]^2 f_X(x) dx \\ &= b^2 \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx \end{aligned}$$

On obtient finalement la relation standard :

$$V(a + bX) = b^2 V(X) \quad (1.9)$$

Un autre résultat, découlant de la propriété 1, est particulièrement utile dans le cadre des modèles linéaires. Il s'agit du résultat suivant :

$$E(X^2) = V(X) + [E(X)]^2 \quad (1.10)$$

Pour démontrer ce résultat, il suffit de réécrire $E(X^2)$ sous la forme $E[(X - \mu + \mu)^2]$. Dès lors, il vient :

$$\begin{aligned} E(X^2) &= E[(X - \mu + \mu)^2] \\ &= E[(X - \mu)^2 + 2\mu(X - \mu) + \mu^2] \\ &= E[(X - \mu)^2] + 2\mu[E(X) - \mu] + \mu^2 \\ &= V(X) + [E(X)]^2 \end{aligned}$$

Dans l'application de certains tests, nous aurons par la suite besoin d'introduire les moments centrés d'ordre 3 et 4, qui correspondent respectivement à la *Skewness* et à la *Kurtosis* :

$$Skewness = \mu_3 = E[(X - \mu)^3] \quad (1.11)$$

$$\text{Kurtosis} = \mu_4 = E \left[(X - \mu)^4 \right] \quad (1.12)$$

La *Skewness* est une mesure de l'asymétrie de la distribution. Pour des fonctions de distributions symétriques, telles que $f_X(\mu - x) = f_X(\mu + x)$, la valeur de la Skewness est nulle, $S_u = 0$. En revanche, pour des fonctions de distributions asymétriques, la valeur de la Skewness est positive, si la partie "épaisse" de la distribution se situe dans la direction positive. La Kurtosis est une mesure de "l'épaisseur" des queues de distributions. En règle générale, on exprime ces deux mesures en contrôlant par une fonction puissance de la variance $V(X) = \sigma^2$. On définit ainsi deux nouvelles mesures : le *coefficient de Skewness* et le *degré d'excès de Kurtosis*.

$$\text{coefficient de Skewness} = \frac{\mu_3}{\sigma^3} \quad (1.13)$$

$$\text{degré d'excès de Kurtosis} = \frac{\mu_4}{\sigma^4} - 3 \quad (1.14)$$

Cette dernière mesure est fondée sur la distribution normale, qui est une distribution à queue "plate", et qui possède un degré d'excès de Kurtosis égal à 0.

1.2 Notions de Convergence

Considérons une séquence de T v.a.r. $\{X_1, X_2, \dots, X_i, \dots, X_T\}$, indicées par i . Supposons que l'on souhaite étudier le comportement de la moyenne empirique de ces v.a.r. lorsque T augmente. On cherche ainsi à déterminer *le comportement asymptotique* de la v.a.r. transformée, \bar{X}_T , telle que :

$$\bar{X}_T = \frac{1}{T} \sum_{i=1}^T X_i \quad (1.15)$$

Pour cela, il convient d'utiliser *la notion de convergences*.

1.2.1 Convergence en probabilité

La notion de convergence en probabilité est définie de la façon suivante :

Definition 2 (*Convergence en Probabilité*) *Soit $\{X_T\}_{T=1}^{\infty}$ une séquence de variables aléatoires scalaires. Cette séquence converge en probabilité vers $c, \forall c \in \mathbb{C}$, si pour toute valeurs arbitraires $\varepsilon > 0$ et $\delta > 0$, il existe une valeur N , telle que $\forall T \geq N$:*

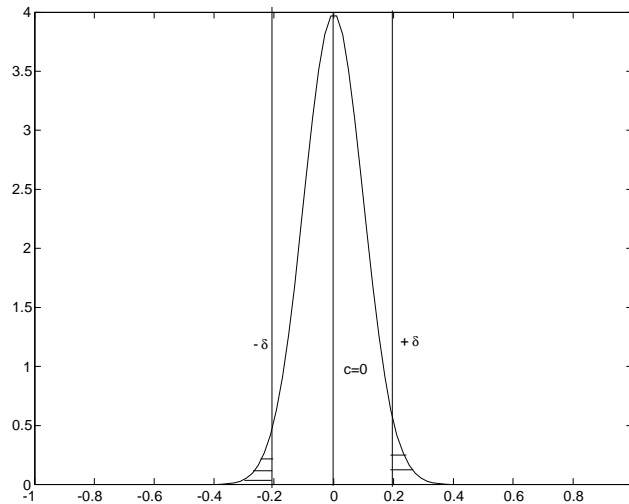
$$P[|X_T - c| > \delta] < \varepsilon \quad (1.16)$$

Alors, on note :

$$X_T \xrightarrow{p} c \iff \text{plim } X_T = c \quad (1.17)$$

Exprimée autrement, cette définition signifie que pour un échantillon de taille infinie, la probabilité que la réalisation de la variable X_T diffère de la valeur c de plus ou moins δ (δ étant aussi petit que l'on veut) est inférieure à toute valeur ε aussi petite soit-elle. En d'autres termes, les réalisations de la variable X_T sont concentrées au voisinage de la valeur c , comme l'illustre la figure (1.1). Sur cette figure, on a représenté la fonction de distribution (supposée normale), d'une variable X_T , pour un ordre $T \geq N$, avec $c = 0$. Si on se donne une valeur quelconque de δ (ici $\delta = 0.2$), et une valeur arbitraire $\varepsilon > 0$, alors la surface hachurée de la fonction de distribution, existant de part et d'autre des valeurs pivotales $-\delta$ et $+\delta$, doit être inférieure à la valeur de ε , si petite soit elle.

Figure 1.1: Illustration de la Notion de Convergence en Probabilité



Remarque Une suite de matrices de v.a.r. $\{X_T\}_{T=1}^{\infty}$, de dimension (m, n) , converge en probabilité vers une matrice C , de dimension (m, n) , si chaque élément de X_t converge en probabilité vers l'élément correspondant de C . De façon plus générale, si l'on considère deux séquences de v.a.r. $\{X_T\}_{T=1}^{\infty}$ et $\{Y_T\}_{T=1}^{\infty}$, de dimension (m, n) , alors :

$$X_T \xrightarrow{p} Y_T \quad (1.18)$$

si et seulement si, la différence entre les deux suites converge en probabilité vers zero :

$$X_T - Y_T \xrightarrow{p} 0 \quad (1.19)$$

Une des applications les plus courantes de la convergence en probabilité, consiste en *la loi faible des grands nombres*. Selon cette loi, la moyenne empirique est un estimateur convergent de

l'espérance de la population considérée. Ainsi, si l'on considère une séquence de v.a.r. $\{Y_t\}_{t=1}^{\infty}$ identiquement et indépendamment distribuées, telle que $E(Y_t) = \mu$ et $Var(Y_t) = \sigma^2, \forall t$. Selon la loi faible¹ des grands nombres :

$$\bar{Y}_T = \frac{1}{T} \sum_{t=1}^T Y_t \xrightarrow[T \rightarrow \infty]{p} E(Y_t) = \mu \quad (1.20)$$

La démonstration fait appel à la notion de convergence en moyenne quadratique (*m.s. pour mean square convergence*). Elle sera présentée dans la section suivante.

Enfin, il convient de rappeler deux propriétés qui nous seront utiles dans la caractérisation des distributions asymptotiques des estimateurs usuels.

Theorem 3 (*Théorème de Slutsky*) **Soit** $\{X_T\}_{T=1}^{\infty}$ **une suite de** $(n, 1)$ **vecteurs admettant une limite en probabilité définie par** c , **et soit** $g(\cdot)$ **une fonction continue en** c , **satisfaisant** $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$, **et ne dépendant pas de** T , **alors :**

$$g(X_T) \xrightarrow[T \rightarrow \infty]{p} g(c) \quad (1.21)$$

L'idée est la suivante. Si la fonction $g(\cdot)$ est continue, la quantité $g(X_T)$ se situera au voisinage de $g(c)$, dès lors que X_T se situe au voisinage de c . En choisissant une valeur de T suffisamment grande, la probabilité que la réalisation de X_T se situe au voisinage de c peut être définie aussi proche de l'unité que l'on le désire. Un exemple simple est le suivant. Considérons deux séquences de v.a.r. telles que $plim X_{1,T} = c_1$ et $plim X_{2,T} = c_2$, alors $plim (X_{1,T} + X_{2,T}) = c_1 + c_2$. La démonstration de ce résultat est immédiate dès lors que l'on montre que la fonction $g(X_{1,T}, X_{2,T}) = X_{1,T} + X_{2,T}$ est une fonction continue en (c_1, c_2) .

Propriété 1 *Une condition suffisante pour qu'une suite de v.a.r. $\{X_T\}_{T=1}^{\infty}$ converge en probabilité vers une constante réelle c est :*

$$\lim_{T \rightarrow \infty} E(X_T) = c \quad (1.22)$$

$$\lim_{T \rightarrow \infty} V(X_T) = 0 \quad (1.23)$$

L'intuition de cette propriété est simple. Si pour un ordre T suffisamment grand, la variable X_T admet c pour espérance et a une variance qui tend vers 0, alors la fonction de distribution de X_T sera infiniment concentrée autour de la valeur c .

¹La loi forte des grands nombres implique par opposition une convergence presque sûre.

1.2.2 Convergence en moyenne quadratique

Une forme de convergence plus restrictive que la convergence en probabilité est la convergence en moyenne quadratique (*m.s.* pour *mean square convergence*).

Definition 4 Une suite de suite de v.a.r. $\{X_T\}_{T=1}^{\infty}$ converge en moyenne quadratique vers c , si pour tout $\varepsilon > 0$, il existe une valeur N , telle $\forall T \geq N$:

$$E(X_T - c)^2 < \varepsilon \quad (1.24)$$

Alors, on note :

$$X_T \xrightarrow{m.s.} c \quad (1.25)$$

Naturellement, étant donné cette définition, la convergence en moyenne quadratique implique la convergence en probabilité, mais la réciproque n'est pas vraie :

$$X_T \xrightarrow{m.s.} c \implies X_T \xrightarrow{p} c$$

La notion de convergence en m.q. nous permet alors d'introduire l'inégalité de Chebyshev.

Proposition 5 (Inégalité de Chebyshev) Soit X une v.a.r. telle que la quantité $E(|X|^r)$ existe et soit finie pour $r > 0$. Pour tout $\delta > 0$, et toute valeur de c , on montre que :

$$P\{|X - c| > \delta\} \leq \frac{E(|X - c|^r)}{\delta^r} \quad (1.26)$$

Le résultat selon lequel la convergence en moyenne quadratique implique la convergence en probabilité peut être démontré à partir de l'inégalité de Chebyshev. Pour cela, il suffit de remarquer que si $X_T \xrightarrow{m.s.} c$, alors il existe un couple de valeurs positives (δ, ε) et une valeur N , tel que $E(X_T - c)^2 < \delta^2 \varepsilon$, pour tout $T \geq N$. Il s'ensuit que :

$$\frac{E(X - c)^2}{\delta^2} = \frac{E(|X - c|^2)}{\delta^2} < \varepsilon \quad \forall T \geq N$$

L'inégalité de Chebyshev implique alors que :

$$P\{|X - c| > \delta\} < \varepsilon \quad \forall T \geq N$$

Donc, on montre ainsi que $X_T \xrightarrow{p} c$.

Une autre application possible de la notion de convergence en m.q. consiste à démontrer la loi faible des grands nombres. Selon cette loi, si l'on considère une séquence de v.a.r.

$\{Y_t\}_{t=1}^{\infty}$ identiquement et indépendamment distribuées, telle que $E(Y_t) = \mu$ et $Var(Y_t) = \sigma^2$, $\forall t$, alors :

$$\bar{Y}_T = \frac{1}{T} \sum_{t=1}^T Y_t \xrightarrow[T \rightarrow \infty]{p} E(Y_t) = \mu \quad (1.27)$$

Pour démontrer ce résultat, étudions le comportement asymptotique de la quantité $E(\bar{Y}_T - \mu)^2$.

$$E(\bar{Y}_T - \mu)^2 = \frac{1}{T^2} Var\left(\sum_{t=1}^T Y_t\right) = \frac{1}{T^2} \sum_{t=1}^T Var(Y_t) = \frac{\sigma^2}{T}$$

Ainsi, on montre que $\lim_{T \rightarrow \infty} E(\bar{Y}_T - \mu)^2 = 0$, ce qui implique que $\bar{Y}_T \xrightarrow{m.s.} \mu$, et donc par conséquent que $\bar{Y}_T \xrightarrow{p} \mu$.

1.2.3 Convergence en loi

Le troisième type de convergence que nous utiliserons cette année est la *convergence en loi* ou *convergence en distribution*.

Theorem 6 (Théorème de Paul Levy) *Soit $\{X_T\}_{T=1}^{\infty}$ une suite de v.a.r. et soit $F_{X_T}(x)$ la fonction de distribution cumulative de X_T . Si X_T converge en loi vers une v.a.r. X admettant $F_X(x)$ pour fonction caractéristique, alors :*

$$\lim_{T \rightarrow \infty} F_{X_T}(x) = F_X(x) \quad \forall x \in \mathbb{R} \quad (1.28)$$

On note alors :

$$X_T \xrightarrow[T \rightarrow \infty]{loi} X \quad \text{ou} \quad X_T \xrightarrow[T \rightarrow \infty]{\mathcal{L}} X \quad (1.29)$$

Un certain nombre de propriétés nous seront particulièrement utiles par la suite :

Propriété 1 *La convergence en probabilité implique la convergence en loi :*

$$X_T - X \xrightarrow[T \rightarrow \infty]{p} 0 \implies X_T \xrightarrow[T \rightarrow \infty]{\mathcal{L}} X \quad (1.30)$$

Propriété 2 *La convergence en loi vers une constante réelle implique la convergence en probabilité :*

$$\forall c \in \mathbb{R} \quad X_T \xrightarrow[T \rightarrow \infty]{\mathcal{L}} c \implies X_T \xrightarrow[T \rightarrow \infty]{p} c \quad (1.31)$$

Propriétés 3 *Soient deux suites de v.a.r. $\{X_T\}_{T=1}^{\infty}$ et $\{Y_T\}_{T=1}^{\infty}$ telle que $X_T \xrightarrow{\mathcal{L}} X$ et $Y_T \xrightarrow{p} c$, alors :*

- (i) $X_T + Y_T \xrightarrow{\mathcal{L}} X + c$
- (ii) $X_T Y_T \xrightarrow{\mathcal{L}} c X$
- (iii) $\frac{X_T}{Y_T} \xrightarrow{\mathcal{L}} \frac{X}{c}$ avec $c \neq 0$

Propriété 4 Soient X_T et X des vecteurs aléatoires de \mathbb{R}^p , tels que $X_T \xrightarrow[T \rightarrow \infty]{\mathcal{L}} X$, et soit $g(\cdot)$ une fonction continue définie de \mathbb{R}^p and \mathbb{R}^n , alors :

$$g(X_T) \xrightarrow[T \rightarrow \infty]{\mathcal{L}} g(X) \quad (1.32)$$

Naturellement, une des illustrations les plus connues de la notion de convergence en distribution réside dans *le théorème central limite (TCL en abrégé)*. Nous avons vu que selon la loi faible des grands nombres, la moyenne empirique \bar{Y}_T d'une suite $\{Y_t\}_{t=1}^{\infty}$ de v.a.r. identiquement et indépendamment distribuées (*i.i.d.* par la suite), telle que $E(Y_t) = \mu$ et $Var(Y_t) = \sigma^2$, $\forall t$, convergeait en probabilité vers la constante μ représentant le moment théorique d'ordre un de la v.a.r. Y_t . Ainsi, la v.a.r. transformée \bar{Y}_T converge en loi vers μ .

$$Var(\bar{Y}_T) = E(\bar{Y}_T - \mu)^2 = \frac{1}{T^2} Var\left(\sum_{t=1}^T Y_t\right) = \frac{1}{T^2} \sum_{t=1}^T Var(Y_t) = \frac{\sigma^2}{T}$$

Dès lors, on vérifie que $\lim_{T \rightarrow \infty} Var(\bar{Y}_T) = 0$. Autrement dit, la fonction de densité associée à la v.a.r. \bar{Y}_T est dégénérée en une masse ponctuelle pour des échantillons de grande taille. Toutefois, si l'on désire faire de l'inférence statistique sur la v.a.r. \bar{Y}_T , il convient de déterminer une distribution non dégénérée de cette variable. C'est précisément l'objet du *TCL*.

L'intuition est la suivante. On cherche une transformée de \bar{Y}_T possédant une distribution non dégénérée lorsque T tend vers l'infini, en particulier une distribution pour laquelle $Var(\bar{Y}_T) > 0$. Considérons la transformée $\sqrt{T}(\bar{Y}_T - \mu)$. Par définition, on a $E[\sqrt{T}(\bar{Y}_T - \mu)] = 0$. Mais ce qui nous intéresse plus particulièrement, c'est le comportement de la variance de cette transformée, or :

$$Var[\sqrt{T}(\bar{Y}_T - \mu)] = \frac{1}{T} Var(\bar{Y}_T - \mu) = \frac{1}{T} Var(\bar{Y}_T) = \sigma^2$$

Dès lors, on a bien défini *une transformée dont la variance asymptotique est non nulle*, et dont on peut ainsi espérer qu'elle possède une distribution asymptotique non dégénérée quand T tend vers l'infini. En effet, selon le *TCL*, la suite définie par $\sqrt{T}(\bar{Y}_T - \mu)$ converge vers une distribution normale d'espérance nulle et de variance finie σ^2 .

Theorem 7 (Théorème Central Limite) Soit $\{Y_t\}_{t=1}^{\infty}$ une suite de v.a.r. *i.i.d.* telle que $E(Y_t) = \mu$ et $Var(Y_t) = \sigma^2$, $\forall t$, alors :

$$\sqrt{T}(\bar{Y}_T - \mu) \xrightarrow[T \rightarrow \infty]{\mathcal{L}} N(0, \sigma^2) \quad (1.33)$$

avec $\bar{Y}_T = (1/T) \sum_{t=1}^T Y_t$.

Une des hypothèses importantes du *TCL* tient à l'indépendance des v.a.r. Y_t . Nous verrons par la suite comment étendre ce théorème au cas où les v.a.r. Y_t ne sont pas indépendantes.

1.3 Moments Empiriques

On considère *une réalisation* d'une suite de v.a.r. (Y_1, Y_2, \dots, Y_T) , issues d'un processus commun stationnaire², satisfaisant les hypothèses suivantes :

Hypothèses (H1). Les v.a.r. $\{Y_t\}_{t=1}^T$ satisfont les hypothèses suivantes :

- (i) $E(Y_t) = \mu, \forall t \in [1, T]$
- (ii) $E(Y_t - \mu)(Y_{t-j} - \mu) = \gamma_j, \forall t \in [1, T]$
- (iii) $\sum_{j=0}^{\infty} |\gamma_j| < \infty$

On cherche alors à proposer un ensemble d'estimateurs convergents³ des moments théoriques du processus des v.a.r. Y_t , c'est à dire des paramètres μ et γ_j , si l'on s'en tient aux moments d'ordre un et deux. Pour cela on construit des v.a.r. transformée, que l'on qualifie de *moments empiriques*, dont on s'assure qu'elles convergent en probabilité vers les moments théoriques. A partir de la réalisation d'un échantillon, (y_1, y_2, \dots, y_T) , on est alors en mesure de proposer une réalisation de ces estimateurs.

De façon générale, les moments empiriques (non centrés) sont définies de la façon suivante :

Definition 8 *Le moment empirique d'ordre r , avec $r \in N^*$, est définie par la v.a.r.*

$$\bar{Y}_T^{(r)} = \sum_{t=1}^T Y_t^r \quad (1.34)$$

La moyenne empirique, obtenue pour $r = 1$, notée \bar{Y}_T , est définie par :

$$\bar{Y}_T = \sum_{t=1}^T Y_t \quad (1.35)$$

Si les v.a.r. Y_t sont indépendantes, la simple application du TCL nous permet de montrer que la v.a.r. \bar{Y}_T converge en loi, et donc en probabilité, vers l'espérance théorique $E(Y_t) = \mu$. Toutefois, sous les hypothèses plus générales (H1), qui lèvent en particulier l'hypothèse d'indépendance des Y_t , il convient à nouveau de démontrer la convergence de l'estimateur \bar{Y}_T .

On cherche à montrer que sous les hypothèses (H_1) , $\text{plim} \bar{Y}_T = \mu$. Il existe plusieurs façons de démontrer ce résultat. Nous proposons ici d'utiliser la définition de la convergence en moyenne quadratique. On considère donc la v.a.r. \bar{Y}_T , telle $\bar{Y}_T = \sum_{t=1}^T Y_t$. Sous les hypothèses (H_1) ,

²Nous reviendrons par la suite sur la définition de ce terme.

³Un estimateur $\hat{\theta}_T$ d'un paramètre θ est dit convergent si :

$$\hat{\theta}_T \xrightarrow[T \rightarrow \infty]{p} \theta$$

on montre de façon évidente que $E(\bar{Y}_T) = \mu$. La moyenne empirique est donc un estimateur sans biais de l'espérance. Calculons maintenant la variance de \bar{Y}_T :

$$V(\bar{Y}_T) = E[(\bar{Y}_T - \mu)^2] = \frac{1}{T^2} E \left[\sum_{t=1}^T (Y_t - \mu) \right]^2$$

Sous les hypothèses (H_1) , on montre (cf. polycopié d'exercices), que :

$$V(\bar{Y}_T) = E(\bar{Y}_T - \mu)^2 = \frac{1}{T} \left[\gamma_0 + 2 \left(\frac{T-1}{T} \right) \gamma_1 + 2 \left(\frac{T-2}{T} \right) \gamma_2 + \dots + 2 \left(\frac{1}{T} \right) \gamma_{T-1} \right] \quad (1.36)$$

Reste à démontrer que la quantité $TE(\bar{Y}_T - \mu)^2$ est convergente, c.q.f.d. $TE(\bar{Y}_T - \mu)^2 < \infty$. Pour cela montrons que cette quantité peut être majorée par une quantité finie.

$$TE(\bar{Y}_T - \mu)^2 = \left| \gamma_0 + 2 \left(\frac{T-1}{T} \right) \gamma_1 + 2 \left(\frac{T-2}{T} \right) \gamma_2 + \dots + 2 \left(\frac{1}{T} \right) \gamma_{T-1} \right| \quad (1.37)$$

$$\leq |\gamma_0| + 2|\gamma_1| + 2|\gamma_2| + \dots + 2|\gamma_{T-1}| \quad (1.38)$$

C'est ici que le troisième point (iii) des hypothèses (H_1) prend toute son importance. En effet, si l'on suppose $\sum_{j=0}^{\infty} |\gamma_j| < \infty$, alors le membre de droite de l'inégalité (1.37) est convergent, $|\gamma_0| + 2|\gamma_1| + 2|\gamma_2| + \dots + 2|\gamma_{T-1}| < \infty$, quand T tend vers l'infini. Dès lors, on peut en conclure que sous l'hypothèse (H_1) , on a bien $TE(\bar{Y}_T - \mu)^2 < \infty$. Ainsi, on arrive finalement au résultat selon lequel $\lim_{T \rightarrow \infty} E(\bar{Y}_T - \mu)^2 = 0$.

En reprenant la définition de la convergence en *m.s.*, on sait que si :

$$\lim_{T \rightarrow \infty} E(\bar{Y}_T - \mu)^2 = 0 \iff \bar{Y}_T \xrightarrow[T \rightarrow \infty]{m.s.} \mu$$

Sachant que la convergence en moyenne quadratique implique la convergence en probabilité, on retrouve le résultat standard :

$$\bar{Y}_T \xrightarrow[T \rightarrow \infty]{p} \mu \quad (1.39)$$

Remarque La moyenne empirique est un estimateur convergent de l'espérance théorique. Toutefois, pour démontrer ce résultat sous des conditions générales, (notamment sans l'hypothèse d'indépendance), il nous a fallu introduire une condition supplémentaire, dite condition de sommabilité absolue, que l'on retrouvera souvent par la suite en analyse de séries temporelles. Cette condition implique que la somme en valeur absolue des autocovariances est elle-même convergente :

$$\sum_{j=0}^{\infty} |\gamma_j| < \infty \quad (1.40)$$

Ce résultat peut être généralisé sous la forme suivante :

Proposition 9 *Soit Y_T un processus dont les moments satisfont les hypothèses (H_1) et dont les autocovariances vérifient la condition de sommabilité absolue, alors la v.a.r. \bar{Y}_T vérifie les propriétés suivantes :*

$$(i) \bar{Y}_T \xrightarrow[T \rightarrow \infty]{m.s.} \mu$$

$$(ii) \lim_{T \rightarrow \infty} [TE(\bar{Y}_T - \mu)^2] = \sum_{j=-\infty}^{\infty} \gamma_j$$

1.4 Distributions Jointes, Distributions Marginales

On considère à présent un système de deux v.a.r. continues X et Y . Commençons par définir ce qu'est la *distribution jointe* de ces deux variables. Cette distribution nous permet d'établir la probabilité que les deux événements associés au fait que les réalisations de ces deux variables soient comprises dans deux intervalles, se réalisent conjointement.

Definition 10 *La densité jointe de deux v.a.r. continues X et Y , notée $f_{X,Y}(x,y) \geq 0$, satisfait les propriétés suivantes :*

$$\forall (a,b,c,d) \in \mathbb{R}^4 \quad P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f_{X,Y}(x,y) dy dx \quad (1.41)$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy dx = 1 \quad (1.42)$$

La fonction de distribution cumulative jointe, notée $F_{X,Y}(x,y)$ est alors définie par,
 $\forall (a,b) \in \mathbb{R}^2 :$

$$F_{X,Y}(a,b) = P(X \leq a, Y \leq b) = \int_{-\infty}^a \int_{-\infty}^b f_{X,Y}(x,y) dy dx \quad (1.43)$$

Partant de la distribution jointe, il est possible de définir la distribution marginale associée à l'une ou l'autre des deux variables X ou Y . Pour obtenir la distribution marginale, il convient d'intégrer la distribution jointe par rapport à la variable qui n'est pas la variable d'intérêt.

Definition 11 *La distribution marginale associée à la variable X , notée $f_X(x)$, est définie par :*

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy \quad (1.44)$$

De la même façon :

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx \quad (1.45)$$

Naturellement, il s'ensuit que deux variables indépendantes, dont par définition les réalisations ne sont pas liées, ont toujours une fonction de distribution jointe qui est précisément égale au produit des deux distributions marginales.

Proposition 12 *Les variables X et Y sont dites indépendantes si et seulement si :*

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) \quad (1.46)$$

L'indépendance implique par conséquent que :

$$F_{X,Y}(x, y) = F_X(x) F_Y(y) \quad (1.47)$$

En effet, dans ce cas, la probabilité que la réalisation de la variable X soit comprise dans un intervalle $[a, b]$ et que la réalisation de la variable Y soit comprise dans un intervalle $[c, d]$, est égale au produit des probabilités marginales puisque les deux événements n'ont aucun lien entre eux. De la même façon :

$$P(X \leq a, Y \leq b) = P(X \leq a) \cdot P(Y \leq b)$$

1.5 Moments Théoriques et Covariance : le Cas d'une Distribution Bivariée

On considère un système de deux v.a.r. continues X et Y . Soit $f_{X,Y}(x, y)$ la fonction de distribution jointe, et soient $f_X(x)$ et $f_Y(y)$ les distributions marginales respectives. Il convient à présent d'étudier les moments théoriques associés à ces deux variables. *La population des moments des v.a.r. dans une distribution jointe est alors définie à partir des distributions marginales respectives.* De façon générale, pour toute fonction $g(X, Y)$, on a :

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dy dx \quad (1.48)$$

De cette propriété, on peut alors déduire l'expression des moments d'ordre un et deux en fonction des distributions marginales.

Definition 13 *Soit $f_X(x)$ la distribution marginale de la v.a.r. X , et soit $f_{X,Y}(x, y)$ la fonction de distribution jointe des v.a.r. X et Y , l'espérance de X est définie de la façon suivante :*

$$E(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X,Y}(x, y) dy dx = \int_{-\infty}^{\infty} x f_X(x) dx \quad (1.49)$$

En effet, les deux définitions sont identiques puisque $f_X(x) = \int_y f_{X,Y}(x,y) dy$.

De la même façon, on peut définir les moments centrés d'ordre deux associés aux variables X et Y . On introduit alors la notion de *covariance*. Soient μ_x et μ_y les espérances respectives des v.a.r. X et Y .

Definition 14 Soit $f_{X,Y}(x,y)$ la fonction de distribution jointe des v.a.r. X et Y , soient $\mu_x = E(X)$ et $\mu_y = E(Y)$, la covariance de X et de Y , ou moment centré d'ordre deux, est :

$$\begin{aligned} cov(X,Y) &= E[(X - \mu_x)(Y - \mu_y)] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f_{X,Y}(x,y) dy dx \end{aligned} \quad (1.50)$$

La corrélation entre X et Y est alors définie par :

$$corr(X,Y) = \frac{cov(X,Y)}{[V(X)V(Y)]^{\frac{1}{2}}} \quad (1.51)$$

De façon générale, on montre que :

$$cov(X,Y) = E[(X - \mu_x)(Y - \mu_y)] = E[XY] - \mu_x\mu_y$$

Proposition 15 Lorsque les variables X et Y sont indépendantes, la covariance et la corrélation sont nulles.

$$cov(X,Y) = 0 \iff corr(X,Y) = 0 \quad (1.52)$$

Toutefois, la réciproque n'est pas vraie, puisque la nullité de la covariance (ou de la corrélation) n'implique pas nécessairement l'indépendance.

En effet, sous l'hypothèse d'indépendance on sait que $f_{X,Y}(x,y) = f_x(x)f_y(y)$. Dès lors, on montre que :

$$\begin{aligned} cov(X,Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f_x(x) f_y(y) dy dx \\ &= \int_{-\infty}^{\infty} (x - \mu_x) \left[\int_{-\infty}^{\infty} (y - \mu_y) f_y(y) dy \right] f_x(x) dx \\ &= \int_{-\infty}^{\infty} (x - \mu_x) \left[\int_{-\infty}^{\infty} y f_y(y) dy - \mu_y \int_{-\infty}^{\infty} f_y(y) dy \right] f_x(x) dx \\ &= \int_{-\infty}^{\infty} (x - \mu_x) [\mu_y - \mu_y] f_x(x) dx \\ &= 0 \end{aligned}$$

La réciproque n'est cependant pas valable, puisqu'il existe des cas où la nullité de la covariance n'implique pas l'indépendance. Considérons ainsi l'exemple suivant. Soient deux variables indépendantes Z et Y , d'espérance nulle. On considère la variable $X = ZY$, alors :

$$\text{cov}(X, Y) = E[(X - \mu_x)(Y - \mu_y)] = E(XY) = E[(ZY)Y] = E(Z)E(Y^2) = 0$$

Dans ce cas, les v.a.r. X et Y ne sont pas corrélées, pourtant elles ne sont pas indépendantes puisque les réalisations de $X = ZY$ dépendent de celles de Y .

Enfin, en séries temporelles, il est souvent nécessaire de calculer les moments théoriques d'une somme de deux v.a.r. ayant une distribution jointe $f_{X,Y}(x, y)$. Considérons la variable Z définie par $Z = aX + bY$, avec $(a, b) \in \mathbb{R}^2$. Calculons les moments théoriques associées à la v.a.r. Z .

$$\begin{aligned} E(Z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (ax + by) f_{X,Y}(x, y) dy dx \\ &= a \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X,Y}(x, y) dy dx + b \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{X,Y}(x, y) dy dx \\ &= a \int_{-\infty}^{\infty} x f_x(x) dx + b \int_{-\infty}^{\infty} y f_y(y) dy \end{aligned}$$

Ainsi, on obtient le résultat suivant :

$$E(aX + bY) = aE(X) + bE(Y) \quad (1.53)$$

De la même façon, calculons la variance de la v.a.r. Z :

$$\begin{aligned} V(Z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [(ax + by) - (a\mu_x + b\mu_y)]^2 f_{X,Y}(x, y) dy dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [(ax - a\mu_x)^2 + 2(ax - a\mu_x)(by - b\mu_y) + (by - b\mu_y)^2] f_{X,Y}(x, y) dy dx \\ &= a^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)^2 f_{X,Y}(x, y) dy dx + 2ab \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f_{X,Y}(x, y) dy dx \\ &\quad + b^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \mu_y)^2 f_{X,Y}(x, y) dy dx \end{aligned}$$

De façon générale, on obtient :

$$V(aX + bY) = a^2V(X) + 2ab \cdot \text{cov}(X, Y) + b^2V(Y) \quad (1.54)$$

Lorsque les v.a.r. X et Y sont indépendantes, alors cette expression se ramène à :

$$V(aX + bY) = a^2V(X) + b^2V(Y) \quad (1.55)$$

Ces deux derniers résultats peuvent être généralisés de la façon suivante.

Proposition 16 *On considère une collection de n v.a.r. $\{X_1, X_2, \dots, X_n\}$, et un ensemble de paramètres, $a_i \in \mathbb{R}$, $\forall i = 1, \dots, n$, alors :*

$$E \left[\sum_{i=1}^n a_i X_i \right] = \sum_{i=1}^n a_i E(X_i) \quad (1.56)$$

$$V \left[\sum_{i=1}^n a_i X_i \right] = \sum_{i=1}^n a_i^2 V(X_i) + 2 \sum_{i=1}^n \sum_{j \neq i}^n a_i a_j \text{cov}(X_i, X_j) \quad (1.57)$$

Si les v.a.r. X_i sont indépendantes, cette expression se ramène à :

$$V \left[\sum_{i=1}^n a_i X_i \right] = \sum_{i=1}^n a_i^2 V(X_i) \quad (1.58)$$

1.6 Distributions Conditionnelles

Considérons un système de deux v.a.r. continues X et Y . Soit $f_{X,Y}(x,y)$ la fonction de distribution jointe, et soient $f_X(x)$ et $f_Y(y)$ les distributions marginales respectives.

Definition 17 *La densité conditionnelle de Y étant donné X est définie par :*

$$f_{Y/X}(y/x) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_X(x)} & \text{si } f_X(x) > 0 \\ 0 & \text{sinon} \end{cases} \quad (1.59)$$

On vérifie alors que la densité conditionnelle satisfait les propriétés standard d'une fonction de distribution, en particulier :

$$\begin{aligned} \int_{-\infty}^{\infty} f_{Y/X}(y/x) dy &= \int_{-\infty}^{\infty} \frac{f_{X,Y}(x,y)}{f_X(x)} dy \\ &= \frac{1}{f_X(x)} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy \\ &= \frac{f_X(x)}{f_X(x)} = 1 \end{aligned}$$

Une propriété importante est la suivante :

Propriété 1. *La densité jointe correspond au produit de la densité conditionnelle et de la densité marginale :*

$$f_{X,Y}(x,y) = f_{Y/X}(y/x) f_X(x) = f_{X/Y}(x/y) f_Y(y) \quad (1.60)$$

Propriété 2. *L'espérance conditionnelle de Y , sachant que la v.a.r. X prend une valeur particulière égale à x , est définie par :*

$$E(Y/X = x) = \int_{-\infty}^{\infty} y f_{Y/X}(y/x) dy \quad (1.61)$$

1.7 Lois de Probabilité Usuelles

Nous proposons à présent quelques rappels sur une sélection de lois de probabilité usuelles.

1.7.1 Loi gamma

La définition de la loi gamma est la suivante :

Definition 18 *La v.a.r. X suit une loi gamma de paramètre p et θ , notée $\gamma(p, \theta)$, si sa fonction de densité est :*

$$f_X(x) = \frac{\theta^p}{\Gamma(p)} e^{-\theta x} x^{p-1} \mathbf{1}_{\mathbb{R}^+(x)} \quad (1.62)$$

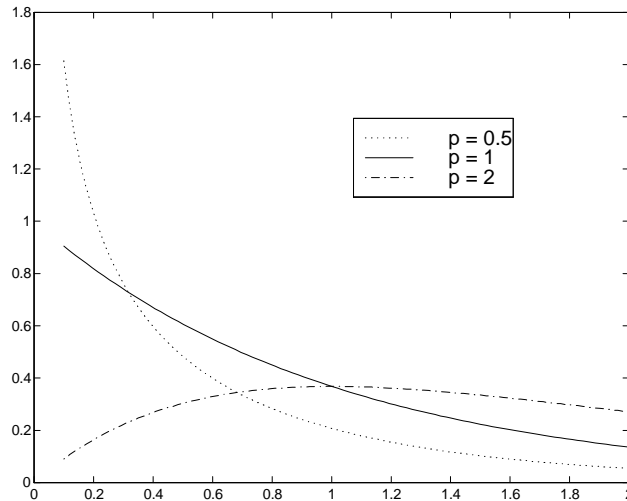
avec $p > 0$, $\theta > 0$ et $\Gamma(p) = \int_0^{+\infty} e^{-x} x^{p-1} dx$.

Lorsque le paramètre θ est égal à 1, on note la loi gamma $\gamma(p, 1)$ sous la forme $\gamma(p)$. La densité est alors définie par :

$$f_X(x) = \frac{1}{\Gamma(p)} e^{-x} x^{p-1} \mathbf{1}_{\mathbb{R}^+(x)} \quad (1.63)$$

Sur la figure (1.2) cette fonction est représentée pour différentes valeurs du paramètre p .

Figure 1.2: Fonction de Densité de la Loi $\gamma(p)$



On peut montrer que si X suit une loi $\gamma(p, \theta)$, alors :

$$E(X^r) = \frac{\Gamma(p+r)}{\theta^r \Gamma(p)}$$

On en déduit les moments d'ordre un et deux :

$$E(X) = \frac{p}{\theta} \quad V(X) = \frac{p}{\theta^2} \quad (1.64)$$

La fonction gamma $\Gamma(p)$ possède les propriétés suivantes :

Propriétés La fonction gamma $\Gamma(p)$ satisfait les propriétés suivantes :

- (i) $\Gamma(p) = \int_0^{+\infty} e^{-x} x^{p-1} dx, p > 0$
- (ii) $\Gamma(p) = (p-1) \Gamma(p-1)$
- (iii) $\Gamma(p) = (p-1)!$ avec $p \in \mathbb{N}^*$
- (iv) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$
- (v) $\Gamma(p+1) \sim \left(\frac{p}{e}\right)^p \sqrt{2\pi p} \left(1 + \frac{1}{12p}\right)$ avec $p \rightarrow \infty$

Enfin, pour une loi $\gamma(p)$, on montre que :

$$\frac{X-p}{\sqrt{p}} \xrightarrow[p \rightarrow \infty]{\mathcal{L}} N(0,1) \quad (1.65)$$

1.7.2 Loi Normale

Commençons par la définition de la loi normale uni-dimensionnelle.

Definition 19 *La v.a.r. X suit une loi normale d'espérance μ et de variance σ^2 , notée $N(\mu, \sigma^2)$, si sa densité est donnée par :*

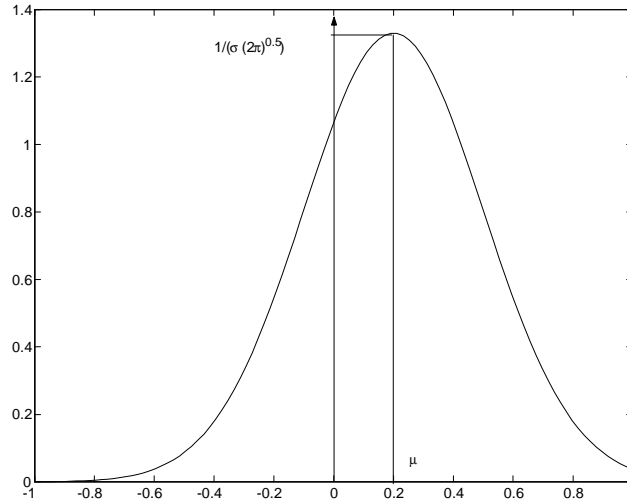
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \forall x \in \mathbb{R} \quad (1.66)$$

Les deux premiers moments sont les suivants : $E(X) = \mu, V(X) = \sigma^2$. La v.a.r. transformée $U = (X - \mu) / \sigma$ suit une loi normale d'espérance nulle et de variance égale à 1, notée $N(0,1)$, dite loi normale centrée réduite, de densité :

$$f_U(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \quad \forall u \in \mathbb{R}$$

La forme de la densité associée à une loi $N(0,1)$ est représentée sur la figure (1.3).

Propriété 1 *Soient X et Y deux v.a.r. indépendantes suivant respectivement les lois $N(\mu_1, \sigma_1)$ et $N(\mu_2, \sigma_2)$. La v.a.r. $X + Y$ suit une loi normale $N\left(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$.*

Figure 1.3: Densité d'une Loi Normale $N(0, 1)$


La v.a.r. définie par $(1/2)U^2$ suit une loi gamma $\gamma(\frac{1}{2}, 1)$. On peut en déduire les moments centrés de la v.a.r. X :

$$\mu_r = E[(X - \mu)^r] = \begin{cases} \sigma^r 2^{\frac{r}{2}} \frac{\Gamma(\frac{r+1}{2})}{\Gamma(\frac{1}{2})} & r \text{ pair} \\ 0 & r \text{ impair} \end{cases} \quad (1.67)$$

Ainsi, à partir de cette formule, on peut aisément retrouver la valeur de la Skewness (moment centré d'ordre 3, noté μ_3) et de la Kurtosis (moment centré d'ordre 4, noté μ_4). On retrouve en effet le résultat selon lequel la Skewness est nulle ($\mu_3 = 0$), ce qui traduit la symétrie de la fonction de distribution d'une loi normale. Quant à la Kurtosis, on montre que celle-ci prend une valeur égale à $3\sigma^4$, puisque en effet :

$$\mu_4 = 4\sigma^4 \frac{\Gamma(\frac{5}{2})}{\Gamma(\frac{1}{2})} = 4\sigma^4 \frac{(\frac{5}{2} - 1) \Gamma(\frac{5}{2} - 1)}{\Gamma(\frac{1}{2})} = 4 \left(\frac{3}{2}\right) \left(\frac{1}{2}\right) \sigma^4 = 3\sigma^4$$

On vérifie ainsi que pour une distribution normale (distribution à queue plate), le degré d'excès de Kurtosis, défini par $\mu_4/\sigma^4 - 3$ est nul.

Propriété 2 Si X suit une loi normale $N(\mu, \sigma^2)$, la Skewness associée est nulle ($\mu_3 = 0$) et la Kurtosis est égale à $\mu_4 = 3\sigma^4$. Par convention, le degré d'excès de Kurtosis, défini par $\mu_4/\sigma^4 - 3$, est nul.

On peut naturellement étendre la définition de la loi normale au cas multivarié.

Definition 20 Soit $\mu \in \mathbb{R}$ et une matrice (p, p) réelle symétrique définie positive, X est un vecteur normal de dimension p , si sa densité est donnée par :

$$f_X(x) = \frac{1}{(2\pi)^{\frac{p}{2}} \sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(x-\mu)' \Sigma^{-1}(x-\mu)} \quad (1.68)$$

Les propriétés mises en évidence pour la loi univariée peuvent alors être étendues au cas multivarié.

1.7.3 Loi du Khi-deux

La définition de la loi du Khi-deux est la suivante :

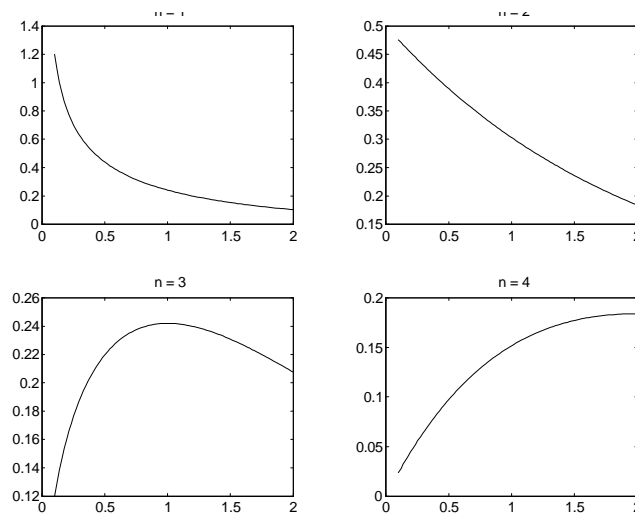
Definition 21 On considère une suite de n v.a.r. (X_1, X_2, \dots, X_n) indépendantes suivant toutes la loi normale $N(0, 1)$. La variable $U_n = \sum_{i=1}^n X_i^2$ suit une loi du khi-deux à n degrés de liberté, notée $\chi^2(n)$.

En utilisant les propriétés précédentes, on peut montrer que la v.a.r. transformée $(1/2)U_n$ suit une loi $\gamma(\frac{1}{2})$. On peut alors en déduire la fonction de densité de U_n .

$$f_{U_n}(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} e^{-\frac{x}{2}} x^{\frac{n}{2}-1} \mathbf{1}_{\mathbb{R}^+(x)}$$

Cette fonction de densité, pour différentes valeurs du paramètre n , est représentée sur la figure (1.4).

Figure 1.4: Fonction de Densité d'un $\chi^2(n)$



De la relation entre les lois du $\chi^2(n)$ et $\gamma(\frac{1}{2})$, on déduit que :

$$E[(U_n)^r] = 2^r \frac{\Gamma(\frac{n}{2} + r)}{\Gamma(\frac{n}{2})}$$

On en déduit les moments d'ordre un et deux :

$$E(U_n) = n \quad V(U_n) = 2n \quad (1.69)$$

Voici un résumé des propriétés de la loi du khi-deux :

Propriété 1. Soit un vecteur aléatoire X de dimension p , de loi $N(\mu, \Sigma)$, la v.a.r. $Y = (X - m)' \Sigma^{-1} (X - m)$ suit une loi du $\chi^2(p)$.

Propriété 2. Soient X et Y deux v.a.r. indépendantes suivant respectivement deux lois du χ^2 à n et m degrés de liberté, la v.a.r. $X + Y$ suit une loi du $\chi^2(n + m)$.

Propriété 3. On considère n variables aléatoires indépendantes X_i suivant respectivement des lois $N(m_i, \sigma)$. La variable $U_{n,\theta} = \sum_{i=1}^n X_i^2$ suit une loi du khi-deux décentrée à n degrés de liberté et de paramètre de décentrage $\theta = \sum_{i=1}^n m_i^2$. On note cette loi $\chi^2(n, \theta)$. On montre alors que :

$$E(U_{n,\theta}) = n + \theta \quad V(U_{n,\theta}) = 2(n + 2\theta) \quad (1.70)$$

1.7.4 Loi de Student

La définition de la loi de Student est la suivante :

Definition 22 Soient X et Y deux v.a.r. indépendantes suivant respectivement la loi $N(0, 1)$ et la loi $\chi^2(n)$. On appelle loi de Student à n degrés de liberté, la loi suivie par le rapport :

$$T_n = \frac{X}{\sqrt{Y/n}} \quad (1.71)$$

La densité de T_n est définie de la façon suivante :

$$f_{T_n}(x) = \frac{1}{\sqrt{n} B(\frac{1}{2}, \frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \quad (1.72)$$

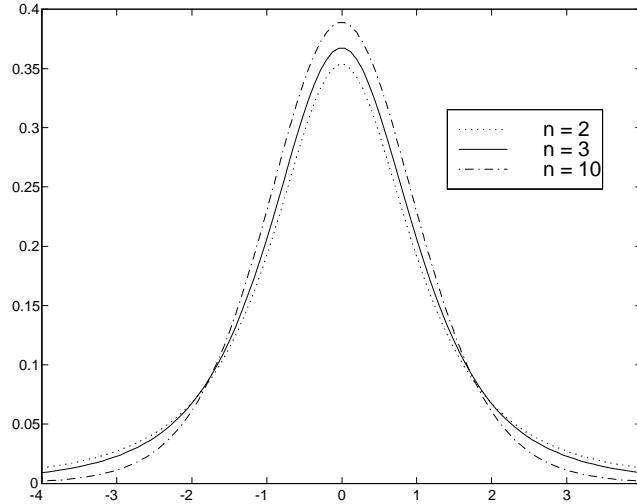
où $B(p, q)$ désigne la fonction beta :

$$B(p, q) = \frac{\Gamma(p) \Gamma(q)}{\Gamma(p+q)} \quad (1.73)$$

La fonction de densité de la v.a.r. T_n est représentée sur la figure (1.5) pour différentes valeurs de n .

Les deux premiers moments de la loi de Student sont les suivants :

$$E(T_n) = 0 \quad V(T_n) = \frac{n}{n-2} \quad n > 2 \quad (1.74)$$

Figure 1.5: Fonction de Densité de T_n


Enfin, on utilise souvent en économétrie l'approximation de la loi de Student par la loi normale centrée réduite.

Propriété 1. La v.a.r. T_n converge en loi vers $N(0, 1)$ lorsque n tend vers l'infini.

$$T_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, 1) \quad (1.75)$$

La démonstration de ce résultat se fait en deux étapes. La première étape consiste à démontrer que la v.a.r. transformée Y_n/n converge en probabilité vers la valeur certaine 1. En effet, on montre que :

$$E\left(\frac{Y_n}{n}\right) = 1 \quad \forall n$$

$$\lim_{n \rightarrow \infty} \left[V\left(\frac{Y_n}{n}\right) \right] = \lim_{n \rightarrow \infty} \left(\frac{2}{n}\right) = 0$$

Selon la propriété 1 de la convergence en probabilité, ces deux résultats impliquent que la v.a.r Y_n/n converge vers 1. Dès lors, la transformée $\sqrt{Y_n/n}$ converge elle aussi en probabilité vers 1. La seconde étape de la démonstration consiste alors à appliquer le troisième point (iii) de la propriété 3 de la convergence en loi. En effet, si l'on a :

$$plim_{n \rightarrow \infty} \left(\frac{Y_n}{n}\right)^{\frac{1}{2}} = 1 \quad X \xrightarrow{\mathcal{L}} N(0, 1)$$

alors, on en déduit que le rapport $X/\left(\sqrt{Y_n/n}\right)$ converge en loi vers une normale centrée réduite.

$$T_n = \frac{X}{\sqrt{Y/n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, 1) \quad (1.76)$$

Dans le cas où la variable X suit une loi normale non centrée, la variable T_n suit alors une loi de Student non centrée.

Definition 23 *Soient X et Y deux v.a.r. indépendantes suivant respectivement la loi $N(\delta, 1)$ et la loi $\chi^2(n)$. On appelle loi de Student supérieurement non centrée, à n degrés de liberté, de paramètre de décentrage δ , la loi suivie par le rapport :*

$$T_n(\delta) = \frac{X}{\sqrt{Y/n}} \quad (1.77)$$

Les deux premiers moments d'une loi de Student non centrée sont les suivants :

$$E[T_n(\delta)] = \sqrt{\frac{n}{2}} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})} \delta$$

$$V[T_n(\delta)] = \frac{n}{n-2} (1 + \delta^2) - \frac{\delta^2 n}{2} \frac{\Gamma^2(\frac{n-1}{2})}{\Gamma^2(\frac{n}{2})}$$

1.7.5 Loi de Fisher-Snedecor

La définition de la loi de Fisher-Snedecor est la suivante.

Definition 24 *Soient X et Y deux v.a.r. indépendantes suivant respectivement la loi $\chi^2(n)$ et la loi $\chi^2(m)$. On appelle loi de Fisher-Snedecor, à n et m degrés de liberté, la loi suivie par le rapport :*

$$F_{n,m} = \frac{X/n}{Y/n} \quad (1.78)$$

La fonction de densité de $F_{n,m}$ est donnée par :

$$f_F(x) = \frac{1}{B(\frac{n}{2}, \frac{m}{2})} n^{\frac{n}{2}} m^{\frac{m}{2}} \frac{x^{\frac{n}{2}-1}}{(m+nx)^{\frac{n+m}{2}}} \mathbf{1}_{\mathbb{R}^+}(x) \quad (1.79)$$

Cette fonction est représentée sur la figure (1.6) pour différentes valeurs de n et de m .

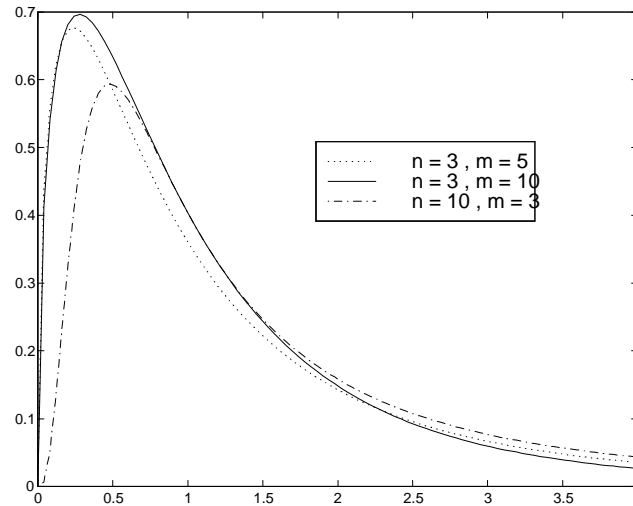
Les deux premiers moments de la variable $F_{n,m}$ sont définis par :

$$E(F_{n,m}) = \frac{m}{m-2} \quad m > 2 \quad (1.80)$$

$$V(F_{n,m}) = \frac{2m^2(n+m-2)}{n(m-4)(m-2)^2} \quad m > 4 \quad (1.81)$$

Etant donnée la définition de $F_{n,m}$ on peut faire plusieurs remarques. Tout d'abord, la variable T_n^2 , carré d'une variable de Student à n degrés de liberté, suit une loi Fisher-Snedecor à 1 et n degrés de liberté. De la même façon, la v.a.r. $Y = \log(F_{2,2})$ suit une loi logistique. Enfin, on utilise parfois la propriété suivante :

Figure 1.6: Fonction de Densité de $F_{n,m}$



Propriété 1. Soit $f_\alpha(n, m)$ le fractile d'ordre α de $F_{n,m}$, défini par $P\{F_{n,m} < f_\alpha(n, m)\} = \alpha$, alors :

$$f_{1-\alpha}(m, n) = \frac{1}{f_\alpha(n, m)} \quad (1.82)$$

Par exemple, soit $F_{5,8}$, dont on cherche le fractile d'ordre 0.05. On a $f_{0.05}(5, 8) = 1/f_{0.95}(8, 5) = 1/4.82 = 0.207$.

2 Séries Temporelles

Essayons à présent de comprendre ce qu'est une *série temporelle* ou de façon équivalente un *processus temporel*. Pour cela, considérons un échantillon de réalisations de T variables aléatoires indicées par le temps, notées Y_t , $t = 1, \dots, T$.

$$\{y_1, y_2, y_3, \dots, y_T\} \quad (2.83)$$

Pour simplifier supposons que cette collection de réalisations de T variables aléatoires soit *indépendamment et identiquement distribuée* (par la suite nous noterons *i.i.d*) selon une même loi normale.

$$Y_t \sim N(\mu, \sigma^2) \quad \forall t = 1, \dots, T \quad (2.84)$$

L'échantillon observé (2.83) constitue ainsi une réalisation particulière sur T unités de temps d'un même *processus stochastique* qui génère les données de la variable Y_t . Ce processus générateur (ou *DGP* pour *Data Generating Process*) est ainsi qualifié de *série stochastique temporelle*.

De façon plus générale, il est possible d'obtenir dans l'absolu une collection de réalisations de ce processus sur une infinité de périodes. On note généralement cette collection de réalisations sous la forme :

$$\{y_t\}_{t=-\infty}^{\infty} = \{\dots, y_{-1}, y_0, y_1, y_2, \dots, y_T, y_{T+1}, \dots\} \quad (2.85)$$

Si l'on connaît le processus générateur (qui dans l'exemple précédent se limite à la loi normale $N(0, \sigma^2)$), il est possible d'obtenir I collections de réalisations $\{y_t\}_{t=-\infty}^{\infty}$, que l'on notera $\{y_t^i\}_{t=-\infty}^{\infty}$, $\forall i = 1, \dots, I$. A partir de ces collections de réalisations, on peut alors se ramener à un échantillon de I observations de la variable aléatoire Y_t et cela quelque soit l'indice de temps retenu. Cette variable aléatoire possède une fonction de densité, notée $f_{Y_t}(y_t)$, qui correspond à la *densité inconditionnelle* de Y_t . Dans notre exemple, cette densité inconditionnelle est donnée par la relation, $\forall t \in \mathbb{Z}$:

$$f_{Y_t}(y_t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-y_t^2}{2\sigma^2}\right)$$

Partant de la fonction de densité de Y_t , il est possible de déterminer l'ensemble des *moments* de cette variable aléatoire. En analyse de séries temporelles, on se limite souvent à la définition des moments d'ordre un (*l'espérance*) et d'ordre deux (*la fonction d'autocovariance*).

L'espérance théorique du processus Y_t est déterminée par la formule standard, $\forall t \in \mathbb{Z}$:

$$E(Y_t) = \int_{-\infty}^{\infty} y_t f(y_t) dy_t \quad (2.86)$$

Cette intégrale peut être approximée par la *limite en probabilité*, notée *plim*, de la somme infinie suivante :

$$E(Y_t) = \text{plim}_{I \rightarrow \infty} \sum_{i=1}^I y_t^i \quad (2.87)$$

ou de façon équivalente :

$$\frac{1}{I} \sum_{i=1}^I y_t^i \xrightarrow{I \rightarrow \infty} E(Y_t) \quad (2.88)$$

On utilise ici la notion de *convergence en probabilité* (ou de *limite en probabilité*) puisque rappelons-le, les termes $\{y_t^i\}_{i=1}^I$ sont des réalisations d'une variable aléatoire, donc la somme de ces termes est elle-même une variable aléatoire. Pour ce qui nous intéresse, lorsque les variables $Y_t, \forall t \in \mathbb{Z}$, sont *indépendamment et identiquement distribuées (i.i.d)*, on peut alors déterminer l'espérance $E(Y_t)$ non seulement à partir des réalisations de cette variable, notée $\{y_t^i\}_{i=1}^I$, mais aussi tout simplement à partir des réalisations des variables Y_1, Y_2, \dots , notées $\{y_t\}_{t=-\infty}^{\infty}$. En effet, sous l'hypothèse *i.i.d.* ces différentes variables ont la même distribution et en particulier le même moment d'ordre un. Dès lors, on obtient la relation usuelle :

$$\frac{1}{T} \sum_{t=1}^T y_t \xrightarrow{T \rightarrow \infty} E(Y_t) \quad \forall t \in \mathbb{Z} \quad (2.89)$$

Dans notre exemple, sous l'hypothèse de normalité on obtient évidemment que $E(Y_t) = \mu$, mais de façon plus générale lorsque l'on lève l'hypothèse *i.i.d.*, l'espérance peut dépendre du temps. On a alors :

$$E(Y_t) = \mu_t \quad \forall t \in \mathbb{Z}$$

Nous verrons *dans ce chapitre que le fait que le moment d'ordre un soit ou non fonction du temps (indiqué en t) caractérise en partie la notion de stationnarité qui est essentielle en analyse de série temporelle.*

En ce qui concerne les moments d'ordre deux de la variable Y_t , ceux-ci ne se limitent pas à la simple variance du processus. En effet, il est nécessaire d'étudier les éventuelles corrélations (ou covariance) entre la variable Y_t et les variables $Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots$, c'est à qui traduisent le lien entre les valeurs courantes et passées de la série temporelle étudiée. Cette information (variance et covariance avec les valeurs passées et futures) est résumée par la *fonction d'autocovariance*⁴, notée généralement $\gamma_t(h)$. Reprenons à nouveau nos I collections de réalisations $\{y_t\}_{t=-\infty}^{\infty}$, que l'on notera $\{y_t^i\}_{t=-\infty}^{\infty}, \forall i = 1, \dots, I$. On définit à présent la *distribution jointe* des variables $Y_j, j = t, \dots, t-h$ par la fonction $f_{Y_t, Y_{t-1}, \dots, Y_{t-h}}(y_t, y_{t-1}, \dots, y_{t-h})$. La *fonction d'autocovariance* est définie par $\forall t \in \mathbb{Z}, \forall h \in \mathbb{Z}$:

$$\gamma_t(h) = E\{[Y_t - E(Y_t)][Y_{t-h} - E(Y_{t-h})]\} \quad (2.90)$$

⁴Ou de façon équivalente par la fonction d'autocorrélation qui n'est autre que la fonction d'autovariance normalisée par la variance du processus.

Ce qui peut se réécrire sous la forme :

$$\begin{aligned} \gamma_t(h) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [y_t - E(y_t)] [y_{t-h} - E(y_{t-h})] \\ &\quad \times f_{Y_t, Y_{t-1}, \dots, Y_{t-h}}(y_t, y_{t-1}, \dots, y_{t-h}) dy_t dy_{t-1} \dots dy_{t-h} \end{aligned}$$

La fonction $\gamma_t(h)$ définit ainsi la covariance entre la variable Y_t et la variable Y_{t-h} , notée $cov(Y_t, Y_{t-h}) = E\{[Y_t - E(Y_t)][Y_{t-h} - E(Y_{t-h})]\}$. Le fait que cette covariance soit définie par rapport à une valeur passée de Y_t implique que l'on qualifie cette fonction de *fonction d'autocovariance*. A l'ordre 0, on reconnaît bien entendu la variance de la variable Y_t :

$$\gamma_t(0) = E\{[Y_t - E(Y_t)]^2\} = var(Y_t) \quad (2.91)$$

Tout comme pour le moment d'ordre un, on peut réexprimer cette quantité comme une limite en probabilité :

$$\frac{1}{I} \sum_{i=1}^I [y_t^i - E(Y_t)] [y_{t-h}^i - E(Y_{t-h})] \xrightarrow{I \rightarrow \infty} \gamma_t(h) \quad (2.92)$$

Lorsque les variables $Y_t, \forall t \in \mathbb{Z}$ sont *i.i.d.*, on est en droit de supposer que l'ensemble des moments d'ordre deux sont identiques pour toutes les variables Y_t , la fonction d'autocovariance peut alors être définie par :

$$\frac{1}{T} \sum_{t=1}^T [y_t - E(Y_t)] [y_{t-h} - E(Y_{t-h})] \xrightarrow{T \rightarrow \infty} \gamma_t(h) = \gamma(h) \quad \forall t \in \mathbb{Z} \quad (2.93)$$

Mais de façon générale, la fonction d'autocorrélation peut dépendre de l'indice de temps t , c'est à dire de la variable Y_t considérée. C'est pourquoi on conserve la notation $\gamma_t(h)$. Nous verrons que le fait que cette fonction soit indépendante du temps, c'est à dire que $\gamma_t(h) = \gamma(h), \forall t \in \mathbb{Z}$, caractérise en partie la notion de stationnarité.

De façon pratique, l'économètre dispose d'une seule collection de réalisations de T variables aléatoires :

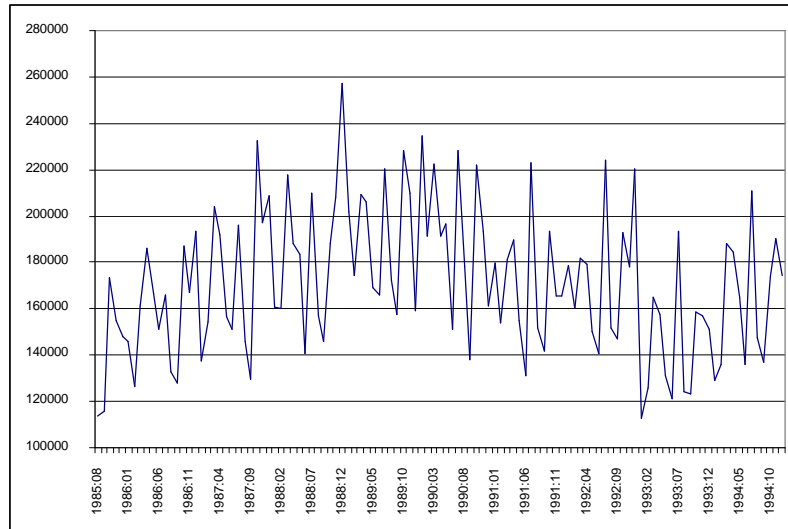
$$\{y_1, y_2, y_3, \dots, y_T\} \quad (2.94)$$

La première chose est de choisir, ou de se restreindre suivant la disponibilité des données, à la meilleure périodicité possible. Considérons deux exemples, l'un relevant de la microéconomie, l'autre d'une dimension plus macroéconomique. La première série, correspondant à la figure (2.7), est la série historique des immatriculations de voitures neuves en France d'août 1985 à novembre 1994⁵. Cette série est disponible en périodicité mensuelle, on dispose ainsi de 114 réalisations historiques.

Le second exemple, correspondant à la figure (2.8), correspond au PIB français exprimé en volume (sans effet prix) du second semestre de 1960 au second semestre de 1998. Cette

⁵Série tirée de Bourbonnais (2000).

Figure 2.7: Immatriculations de Voitures Neuves en France (1985:08 à 1994:10)



série est de périodicité semestrielle, on dispose ainsi de 79 réalisations historiques pour cette série. Dans ce cas, il est possible d'augmenter la périodicité et donc par là même le nombre de réalisations en utilisant des données trimestrielles. Mais ce choix n'est pas toujours possible (données budgétaires par exemple).

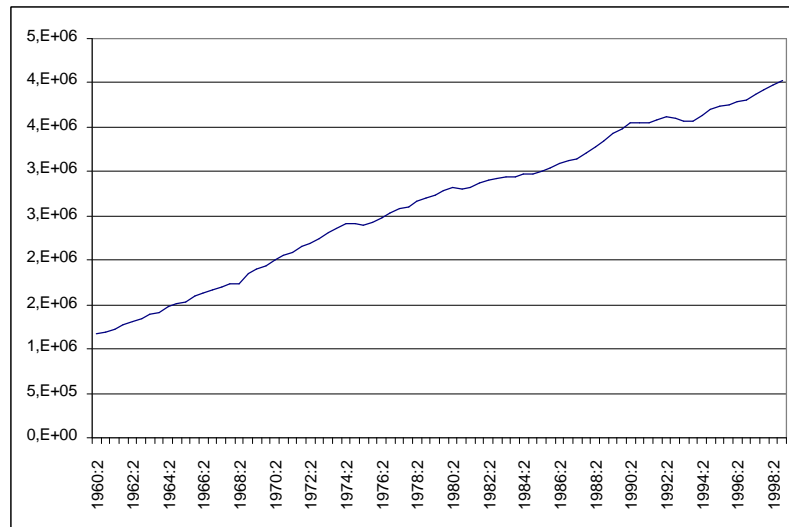
Ensuite en utilisant ces réalisations, on va chercher à identifier le meilleur processus stochastique générateur des données historiques (*DGP*). Partant de ce processus, il sera alors possible d'effectuer des prévisions, mais aussi de réaliser des simulations d'autres pseudo échantillons historiques à des fins statistiques ou à des fins d'interprétation économique.

Nous verrons dans le chapitre suivant, que lorsque l'on considère une réalisation historique d'une série temporelle, la première étape de la phase de modélisation consiste à déterminer si le processus générateur est stationnaire. La notion de stationnarité que nous retiendrons consiste tout simplement à déterminer si les moments des variables aléatoires Y_t , sont indépendants du temps. En particulier si l'on envisage une notion de stationnarité du second ordre, la question est de savoir si les grandeurs $E(Y_t)$ et $\gamma_t(h)$ sont constantes dans le temps. En effet, dans ce cas on est garanti qu'il existe un processus *stochastique indépendant du temps, stationnaire*, permettant de représenter les variables Y_1, Y_2, \dots, Y_t .

Reste, dans une seconde étape à déterminer la ou les formes possibles de ce processus stationnaire. La première représentation possible, valable pour tout processus stationnaire, est alors donnée par la décomposition de Wold qui permet d'exprimer le processus comme une somme pondérée de bruits blancs. Ce sera l'objet de notre deuxième section.

Mais certains processus stationnaires peuvent aussi être représentés par des processus in-

Figure 2.8: PIB en Volume de la France (1960:2 à 1998:2)



tégrant une partie autorégressive et une partie moyenne mobile dans les chocs : les processus *ARMA*. La définition de ces processus et de leurs propriétés fera l'objet de notre troisième section.

Nous allons donc à présent poser la définition de la stationnarité. Partant de là, nous pourrions, dans les chapitres suivants, proposer des tests de cette hypothèse de stationnarité à mener à partir d'un échantillon de réalisations historiques.

3 Les Processus Stationnaires

Nous commencerons par poser la définition d'un processus stationnaire au sens strict (ou *stationnarité forte*) pour ensuite étudier les propriétés de la *stationnarité du second ordre* (ou *stationnarité faible*). Partant de là nous étudierons des processus stationnaires particuliers que sont les *bruits blancs*.

3.1 Définition d'un processus stationnaire au sens strict : la stationnarité forte

Soit un processus temporel aléatoire $(x_t, t \in \mathbb{Z})$.

Definition 25 *Le processus x_t est dit strictement ou fortement stationnaire si \forall le n -uplet du temps $t_1 < t_2 < \dots < t_n$, tel que $t_i \in \mathbb{Z}$ et pour tout temps $h \in \mathbb{Z}$ avec $t_i + h \in \mathbb{Z}, \forall i, i = 1, \dots, n$, la suite $(x_{t_1+h}, \dots, x_{t_n+h})$ à la même loi de probabilité que la suite $(x_{t_1}, \dots, x_{t_n})$.*

Une autre façon équivalente de définir la stationnarité forte (ou stationnarité stricte) :

Definition 26 *Un processus est dit stationnaire au sens strict si, pour toutes valeurs j_1, j_2, \dots, j_n la distribution jointe de la suite $(x_t, x_{t+j_1}, \dots, x_{t+j_n})$ dépend uniquement des intervalles de temps j_1, j_2, \dots, j_n et est indépendante de la période t .*

3.2 La stationnarité d'ordre deux ou la stationnaire faible

Dans la pratique, on se limite généralement à requérir la stationnarité du second ordre (ou *stationnarité faible*) du processus étudié.

Definition 27 *Un processus $(x_t, t \in \mathbb{Z})$ est dit stationnaire au second ordre, ou stationnaire au sens faible, ou stationnaire d'ordre deux si les trois conditions suivantes sont satisfaites :*

- $\forall t \in \mathbb{Z}, E(x_t^2) < \infty$
- $\forall t \in \mathbb{Z}, E(x_t) = m$, *indépendant de t*
- $\forall (t, h) \in \mathbb{Z}^2, cov(x_t, x_{t+h}) = E[(x_{t+h} - m)(x_t - m)] = \gamma(h)$, *indépendant de t*

La première condition $E(x_t^2) < \infty$ garantit tout simplement l'existence (ou la *convergence*) des moments d'ordre deux. La seconde condition $E(x_t) = m, \forall t \in \mathbb{Z}$ porte sur les moments d'ordre un et signifie tout simplement que les variables aléatoires x_t doivent avoir la même espérance

quelle que soit la date t . Autrement dit, *l'espérance du processus x_t doit être indépendante du temps*. Enfin, la troisième condition, $\gamma(h)$ indépendant de t , porte sur les moments d'ordre deux résumés par la fonction d'autocovariance. Cette condition implique que ces moments doivent être indépendants de la date considérée et ne doivent dépendre uniquement que de l'ordre des retards. Autrement dit *la fonction d'autocovariance du processus x_t doit être indépendante du temps*. Sous ces trois conditions, il existe alors une ou plusieurs représentations possibles de la série temporelle x_t , ces représentations ou processus ayant la même distribution dans le temps (même moments). Cette hypothèse d'invariance de la distribution dans le temps permet ainsi de se limiter à une certaine classe de processus.

En résumé, un processus est stationnaire au second ordre si l'ensemble de ses moments sont indépendants du temps. Par la suite le terme stationnaire fera référence à la stationnarité faible.

Quelques précisions doivent être apportées à cette définition. Tout d'abord, il convient de noter que la fonction $\gamma(h)$, qui désigne la fonction génératrice d'autocovariance du processus $(x_t, t \in \mathbb{Z})$, est symétrique

$$\gamma(h) = \gamma(-h)$$

En effet, de manière évidente on montre que :

$$\begin{aligned} \gamma(h) &= E \{ [Y_t - E(Y_t)] [Y_{t-h} - E(Y_{t-h})] \} \\ &= E \{ [Y_{t-h} - E(Y_{t-h})] [Y_t - E(Y_t)] \} \\ &= \gamma(-h) \end{aligned}$$

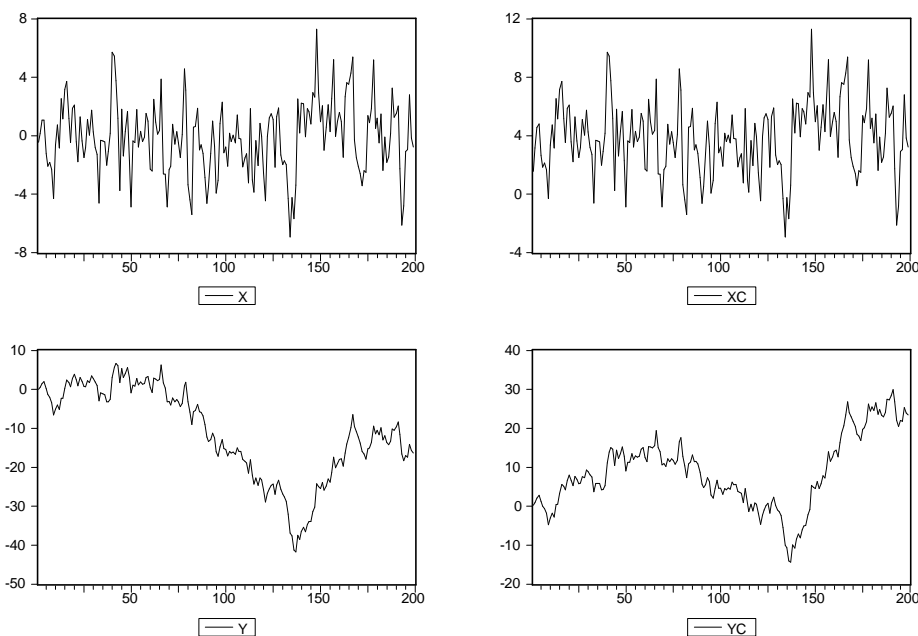
En second lieu, on note que la troisième condition de cette définition implique que en particulier que la variance du processus x_t soit constante et indépendante du temps :

$$v(x_t) = \gamma(0) = \sigma^2 \text{ (indépendant de } t) \quad (3.95)$$

Sur la figure (3.9) sont représentées quelques réalisations de processus stationnaires et non stationnaires sur un échantillon de 200 points. On constate que les variables y_c et y (quadrants sud-est et sud-ouest) sont a priori non stationnaires. La première série est décroissante sur les 150 premiers points ce qui viole la condition de stationnarité du moment d'ordre un, puisque l'espérance des variables $y_{c,t}$ semble ici décroissante avec t . La seconde série, semble quant à elle violer la condition sur les moments d'ordre deux puisque la variance ne semble pas identique sur l'ensemble de l'échantillon. La série semble plus volatile sur les 100 derniers points que sur la première partie de l'échantillon.

En revanche, les deux séries x_c et x (quadrants nord-est et nord-ouest) peuvent être des réalisations d'un processus stationnaire. La première série est centrée sur 0, tandis que la

Figure 3.9: Simulation de Processus Stationnaires et Non Stationnaires



seconde est centrée sur la valeur 3. La variance semble identique sur l'ensemble de l'échantillon et les autocovariances ne semblent pas être fonction du temps.

Mais attention, un tel examen graphique, s'il est toujours nécessaire ne constitue pas un test statistique de la stationnarité et peut parfois induire en erreur. Il est toujours nécessaire de tester l'hypothèse de stationnarité ou de non stationnarité par un test approprié (*cf.* chapitre suivant).

3.3 Le processus bruit blanc (white noise)

Parmi la classe des processus stationnaire, il existe des processus particuliers que sont les *processus bruit blanc* (ou white noise). Ces processus sont très souvent utilisés en analyse des séries temporels car *ils constituent en quelque sorte les "briques élémentaires" de l'ensemble des processus temporels*. En effet, nous verrons par la suite que tout processus stationnaire peut s'écrire comme une somme pondérée de bruits blancs (*théorème de Wold*).

Un bruit blanc est un processus stationnaire à accroissements indépendants. On parle aussi de processus *i.i.d.* (variables indépendantes et identiquement distribuées). Commençons donc par définir ce qu'est un processus à accroissements indépendants.

Definition 28 *Un processus $(x_t, t \in \mathbb{Z})$ est un processus accroissements indépendants si pour tout n -uplet du temps $t_1 < t_2 < \dots < t_n$, $t_i \in \mathbb{Z}$, les variables aléatoires réelles $x_{t_2} - x_{t_1}, \dots, x_{t_n} - x_{t_{n-1}}$ sont indépendantes.*

Le processus $(x_t, t \in \mathbb{Z})$ est un processus stationnaire à accroissements indépendants si la loi de probabilité des accroissements $(x_{t+h} - x_t)$ est indépendante de t , $\forall h \in \mathbb{Z}$. C'est donc une classe particulière des processus stationnaires.

Mais dans la pratique, on retient plus généralement la définition suivante du bruit blanc.

Definition 29 *Un processus $(x_t, t \in \mathbb{Z})$ est un bruit blanc s'il satisfait les deux conditions suivantes $\forall t \in \mathbb{Z}$:*

$$E(x_t) = 0$$

$$\gamma(h) = E[x_t x_{t-h}] = \begin{cases} \sigma^2 & \text{si } h = 0 \\ 0 & \forall h \neq 0 \end{cases}$$

La première condition signifie tout simplement que l'espérance du processus est indépendante du temps, et de plus qu'elle est nulle. La seconde condition implique bien entendu l'indépendance de la fonction d'autocovariance par rapport au temps (stationnarité). Mais elle implique en outre que les termes d'autocovariances (pour $h \neq 0$) sont tous nuls. Seule la variance est non nulle. Autrement dit, cela signifie que *les bruits blancs sont des processus stationnaires particuliers sans "mémoire"*. Le niveau de la série considéré aujourd'hui n'a aucune incidence sur son niveau de demain, tout comme le niveau d'hier n'a aucune incidence sur le niveau d'aujourd'hui.

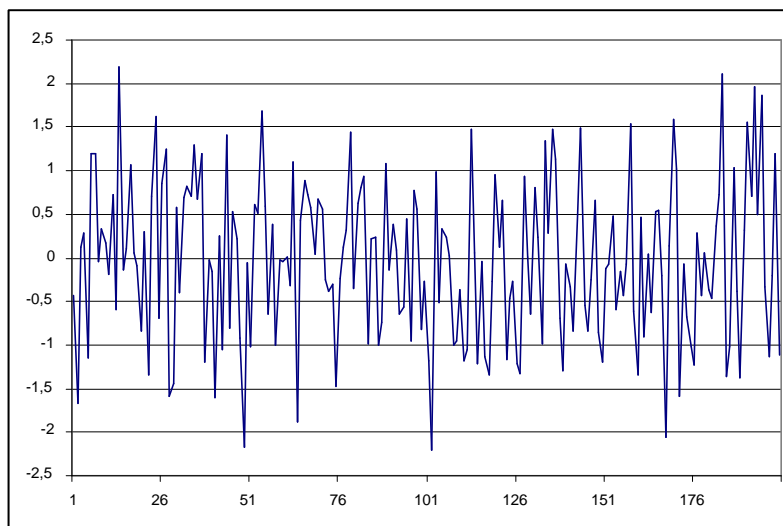
C'est pourquoi, le terme *bruit blanc* provient de l'analogie dans le domaine des fréquences entre la densité spectrale d'une variable *i.i.d.* (constante) et le spectre de la lumière blanche dans le spectre des couleurs.

En outre, on parle de *bruit blanc gaussien* lorsque la loi de probabilité du processus est elle-même gaussienne.

$$\varepsilon_t \text{ i.i.d. } \sim \mathcal{N}(0, \sigma^2) \quad (3.96)$$

Sur la figure (3.10) est représentée une simulation d'un bruit blanc gaussien de variance unitaire pour un échantillon de 200 réalisations. On constate qu'il n'existe aucune *réurrence* (aucune similitude dans l'évolution historique) pour ce type de processus. On dit qu'il est *dynamiquement indépendant*.

Figure 3.10: Simulation d'un Bruit Blanc Gaussien (0, 1)



4 Le Théorème de Wold

Le *théorème de Wold* (1938) est le théorème fondamental de l'analyse des séries temporelles stationnaires. Nous commencerons par donner l'énoncé de ce théorème, puis nous verrons différentes applications, et nous définirons ensuite l'opérateur retard.

4.1 Le théorème de la décomposition de Wold

L'énoncé du théorème de Wold est le suivant

Theorem 30 (Décomposition de Wold) *Tout processus stationnaire d'ordre deux $(x_t, t \in \mathbb{Z})$ peut être représenté sous la forme :*

$$x_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + \kappa_t \quad (4.97)$$

où les paramètres ψ_j satisfont $\psi_0 = 1$, $\psi_j \in \mathbb{R}, \forall j \in \mathbb{N}^*$, $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ et où ε_t est un bruit blanc i.i.d. $(0, \sigma_\varepsilon^2)$. On dit que la somme des chocs passés correspond à la composante linéaire stochastique de x_t . Le terme κ_t désigne la composante linéaire déterministe telle que $\text{cov}(\kappa_t, \varepsilon_{t-j}) = 0, \forall j \in \mathbb{Z}$.

La démonstration de ce théorème est donnée dans Wold (1938), voir Sargent, *Macroeconomic Theory*, Boston Academic Press, p 286-290 pour l'intuition de la démonstration. Ainsi, d'après le théorème de Wold, si l'on omet la composante déterministe κ_t , *tout processus stationnaire peut s'écrire comme une somme pondérée infinie de chocs passés*, ces chocs étant représentés par un bruit blanc de variance finie. L'implication forte de ce théorème est que, si l'on connaît les pondérations $\psi_j, \forall j \in \mathbb{N}$, et si l'on connaît la variance σ_ε^2 du bruit blanc, on est mesure de proposer une représentation de n'importe quel processus stationnaire. Cette représentation est aussi qualifiée de *représentation moyenne mobile infinie*.

Reste à comprendre ce que peut être cette composante linéaire déterministe κ_t . La condition $\text{cov}(\kappa_t, \varepsilon_{t-j}) = 0$, implique que ce terme est, par définition (déterministe), indépendant des chocs. Alors le cas le plus simple est celui d'un processus stationnaire $(x_t, t \in \mathbb{Z})$ d'espérance non nulle, tel que $E(x_t) = m \neq 0$. Puisque le bruit blanc est par définition un processus centré, une somme pondérée de ces chocs est elle-même centrée. Par conséquent, la représentation de Wold du processus $(x_t, t \in \mathbb{Z})$ suppose que l'on ajoute à cette somme pondérée des chocs passés, une composante déterministe qui n'est autre que l'espérance du processus : $\kappa_t = m$. On a donc :

$$x_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + m$$

et l'on vérifie que

$$E(x_t) = E\left[\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}\right] + m = \sum_{j=0}^{\infty} \psi_j E(\varepsilon_{t-j}) + m = m$$

4.1.1 Conditions sur les pondérations ψ_j

L'énoncé du théorème de Wold fait clairement apparaître *trois conditions sur les pondérations* ψ_j . La première condition $\psi_0 = 1$ impose tout d'abord que *le poids du choc présent ε_t dans la définition du processus soit unitaire*. Il s'agit ici tout simplement d'une condition de normalisation qui porte sur la détermination de la variance du bruit blanc. Considérons le processus suivant :

$$x_t = \sum_{j=0}^{\infty} \tilde{\psi}_j v_{t-j} = \frac{1}{2}v_t + \left(\frac{1}{2}\right)^2 v_{t-1} + \left(\frac{1}{2}\right)^3 v_{t-2} + \dots$$

avec $v_t \sim \mathcal{N}(0, \sigma_v^2)$ et $\sigma_v^2 = 1$. Il est possible de normaliser la variance du bruit blanc de sorte à ce que le poids du chocs présent soit unitaire. Pour cela on pose :

$$\varepsilon_t = \frac{1}{2}v_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$$

avec $\sigma_\varepsilon^2 = 1/4$. Le processus $(x_t, t \in \mathbb{Z})$ peut alors être réexprimé en fonction du bruit blanc $(\varepsilon_t, t \in \mathbb{Z})$ de la façon suivante :

$$x_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \varepsilon_t + \frac{1}{2} \varepsilon_{t-1} + \left(\frac{1}{2}\right)^2 \varepsilon_{t-2} + \left(\frac{1}{2}\right)^3 \varepsilon_{t-3} + \dots$$

On obtient ainsi la représentation de Wold du processus $(x_t, t \in \mathbb{Z})$, où le poids du choc contemporains ε_t est unitaire.

La seconde condition, $\psi_j \in \mathbb{R}, \forall j \in \mathbb{N}^*$, est triviale et signifie simplement que les pondérations des chocs strictement passés peuvent être éventuellement nulles pour certains retards.

En revanche, la troisième condition, $\sum_{j=0}^{\infty} \psi_j^2 < \infty$, est particulièrement importante.

Definition 31 *La condition $\sum_{j=0}^{\infty} \psi_j^2 < \infty$, dite condition de sommabilité des carrés (ou d'intégrabilité des carrés) assure l'existence des moments d'ordre deux du processus $(x_t, t \in \mathbb{Z})$. Sous cette condition, on dit alors que x_t converge en moyenne quadratique.*

Pour comprendre en quoi cette condition de sommabilité assure l'existence des moments théoriques d'ordre deux, il faut revenir à la définition de la fonction génératrice d'autocovariance du processus $(x_t, t \in \mathbb{Z})$ et utiliser la représentation de Wold (équation 4.97). Si l'on suppose pour simplifier que $E(x_t) = 0$, on obtient alors :

$$\gamma(h) = E(x_{t+h}x_t) = E\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{t+h-j} \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}\right) \quad \forall h \in \mathbb{Z}$$

Etant donné la définition d'un bruit blanc on a $E(\varepsilon_t \varepsilon_{t-j}) = 0, \forall j \neq 0, E(\varepsilon_t^2) = \sigma_\varepsilon^2$. Dès lors, puisque l'espérance est un opérateur linéaire, la fonction $\gamma(h)$ peut se réécrire sous la forme :

$$\gamma(h) = \begin{cases} \sum_{j=0}^{\infty} \psi_j^2 E(\varepsilon_{t-j}^2) & \text{si } h = 0 \\ \sum_{j=0}^{\infty} \psi_{j+h} \psi_j E(\varepsilon_{t-j}^2) & \forall h \neq 0 \end{cases}$$

D'ou finalement on tire que :

$$\gamma(h) = \begin{cases} \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j^2 & \text{si } h = 0 \\ \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_{j+h} \psi_j & \forall h \neq 0 \end{cases} \quad (4.98)$$

On peut montrer que la condition $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ implique $\sum_{j=0}^{\infty} \psi_{j+h} \psi_j < \infty, \forall h \neq 0$. Dès lors cette condition suffit à garantir la convergence, et donc l'existence, de l'ensemble des moments d'ordre deux du processus x_t .

4.1.2 Prévisions à partir de la décomposition de Wold

Comme nous l'avons dit, une des conséquences de ce théorème est que, si l'on connaît les pondérations $\psi_j, \forall j \in \mathbb{N}$, et si l'on connaît la variance σ_ε^2 du bruit blanc, on est mesure de proposer une représentation de n'importe quel processus stationnaire. Partant de cette représentation, on peut alors réaliser une prévision, notée \widehat{x}_{T+h} , pour la réalisation du processus x_t à la date $T+h$, à partir d'un échantillon de T observations de x_t .

Considérons, pour commencer, une prévision à période ($h = 1$). La meilleure prévision possible de la réalisation de x_{T+1} , notée \widehat{x}_{T+1} , connaissant les valeurs passées $x_T, x_{T-1}, x_{T-2}, \dots$ est donnée par l'espérance conditionnelle suivante $E(x_{T+1} | x_T, x_{T-1}, x_{T-2}, \dots)$. Or, d'après le théorème de Wold, la connaissance des valeurs passées de x_T est équivalente à la connaissance des valeurs passées des chocs $\varepsilon_T, \varepsilon_{T-1}, \varepsilon_{T-2}, \dots$. En effet, on sait que :

$$\begin{aligned} x_T &= \sum_{j=0}^{\infty} \psi_j \varepsilon_{T-j} = \varepsilon_T + \psi_1 \varepsilon_{T-1} + \psi_2 \varepsilon_{T-2} + \psi_3 \varepsilon_{T-3} + \dots \\ x_{T-1} &= \sum_{j=0}^{\infty} \psi_j \varepsilon_{T-j-1} = \varepsilon_{T-1} + \psi_1 \varepsilon_{T-2} + \psi_2 \varepsilon_{T-3} + \psi_3 \varepsilon_{T-4} + \dots \\ x_{T-p} &= \sum_{j=0}^{\infty} \psi_j \varepsilon_{T-j-p} = \varepsilon_{T-p} + \psi_1 \varepsilon_{T-p-1} + \psi_2 \varepsilon_{T-p-2} + \psi_3 \varepsilon_{T-p-3} + \dots \quad \forall p > 0 \end{aligned}$$

Par conséquent, l'espace de conditionnement de notre prévision peut être modifié de la façon suivante :

$$\widehat{x}_{T+1} = E(x_{T+1} | x_T, x_{T-1}, x_{T-2}, \dots) = E(x_{T+1} | \varepsilon_T, \varepsilon_{T-1}, \varepsilon_{T-2}, \dots)$$

Reste alors à considérer la représentation de Wold du processus x_{T+1} :

$$x_{T+1} = \sum_{j=0}^{\infty} \psi_j \varepsilon_{T-j+1} = \varepsilon_{T+1} + \psi_1 \varepsilon_T + \psi_2 \varepsilon_{T-1} + \psi_3 \varepsilon_{T-2} + \dots$$

A partir de cette formule, il est facile de déterminer l'espérance conditionnelle :

$$\begin{aligned} \widehat{x}_{T+1} &= E(x_{T+1} | \varepsilon_T, \varepsilon_{T-1}, \varepsilon_{T-2}, \dots) \\ &= E\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{T-j+1} | \varepsilon_T, \varepsilon_{T-1}, \varepsilon_{T-2}, \dots\right) \\ &= \sum_{j=0}^{\infty} \psi_j E(\varepsilon_{T-j+1} | \varepsilon_T, \varepsilon_{T-1}, \varepsilon_{T-2}, \dots) \end{aligned}$$

On obtient alors la forme générale de notre prévision :

$$\widehat{x}_{T+1} = \sum_{j=1}^{\infty} \psi_j \varepsilon_{T-j+1} = \psi_1 \varepsilon_T + \psi_2 \varepsilon_{T-1} + \psi_3 \varepsilon_{T-2} + \dots \quad (4.99)$$

et par conséquent de l'erreur de prévision :

$$x_{T+1} - \hat{x}_{T+1} = \varepsilon_{T+1} \quad (4.100)$$

Par définition cette erreur de prévision est d'espérance nulle et de variance égale à σ_ε^2 . Ainsi, pour un processus stationnaire l'erreur de prévision optimale à un horizon d'une période correspond à l'innovation fondamentale (au choc) associé à cette même période. De plus, cette erreur de prévision a de "bonnes" propriétés puisqu'elle est strictement indépendante de l'erreur de prévision que l'on a pu commettre aux dates précédentes puisque $E(\varepsilon_T, \varepsilon_{T-j}) = 0$. Ce n'est pas parce qu'hier l'économètre s'est trompé, qu'il se trompera nécessairement demain.

Bien entendu, on peut généraliser ces résultats pour des prévisions à des ordres h quelconques. On montre de la même façon les résultats suivants (à démontrer) :

Résultats Soit \hat{x}_{T+h} la meilleure prévision possible de la réalisation du processus x_t à la date $T+h$ compte tenu de l'information disponible à la date T .

$$\hat{x}_{T+h} = \sum_{j=1}^{\infty} \psi_j \varepsilon_{T+h-j} \quad (4.101)$$

L'erreur de prévision associée est alors déterminée par :

$$x_{T+h} - \hat{x}_{T+h} = \sum_{j=0}^{h-1} \psi_j \varepsilon_{T+h-j} \quad (4.102)$$

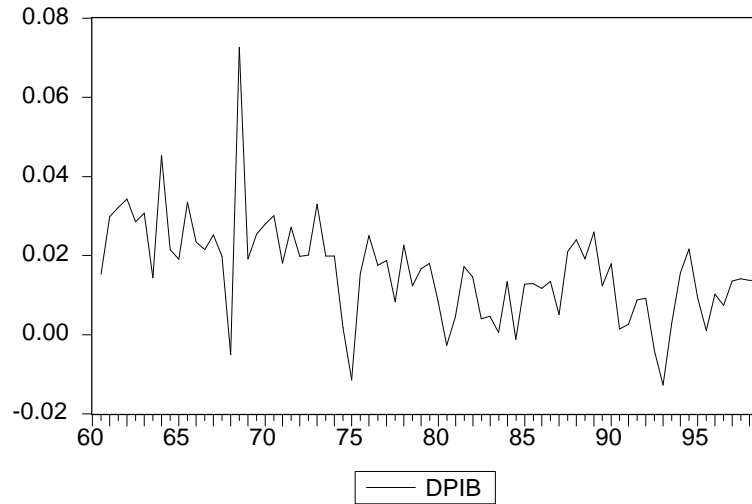
4.2 Applications

Le théorème de Wold n'est jamais directement applicable, en tant que tel, dans la pratique puisque il suppose que l'on exprime un processus stationnaire comme une somme pondérée *infinie* de chocs passés. Or, dans la pratique informatique il n'est pas possible de manier des objets mathématiques de dimension infinie.

Pour autant, on peut démontrer qu'il existe toujours pour tout processus stationnaire, un ordre fini, tel que ce processus puisse être approximé par une somme pondérée de chocs passés jusqu'à cet ordre (décomposition de Wold tronquée).

Concrètement, cela signifie que n'importe quelle série **stationnaire** peut être représentée de façon satisfaisante par une somme pondérée finie de chocs passés. Reste ensuite à estimer les coefficients de pondération (les ψ_j) et la variance du bruit blanc (σ_ε^2) et il est alors possible de proposer une prévision du processus étudié.

Figure 4.11: Taux de Croissance du PIB en Volume France (1960:2 à 1998:2)



Considérons un échantillon de 77 observations semestrielles du PIB français en volume de 1960:2 à 1998:2. Cette série est représentée sur la figure (4.11). Admettons que l'on ait montré que cette série était stationnaire.

Proposons une première représentation pour de ce processus sous la forme d'une somme pondérée finie de chocs passés de la forme :

$$x_t = m + \sum_{j=0}^q \psi_j \varepsilon_{t-j} = m + \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \dots + \psi_q \varepsilon_{t-q} \quad (4.103)$$

où ε_t *i.i.d.* $(0, \sigma_\varepsilon^2)$ et où $q > 0$. Considérons un ordre arbitraire q suffisamment grand, compte tenu du nombre de points disponibles et donc par conséquent du degré de liberté de notre estimation. On retiendra pour l'exemple $q = 20$. Nous n'expliquerons pas ici la façon dont nous avons obtenu les estimateurs des paramètres ψ_j et de la variance σ_ε^2 , puisque cela fera l'objet d'un chapitre ultérieur. Les résultats de l'estimation sont reportés sur la figure (4.12).

Dans ce tableau figurent les réalisations des estimateurs de la constante (c) et des 20 paramètres ψ_j , pour $j = 1, \dots, 20$. Ces estimateurs sont notés ici sous la forme $MA(j)$. On observe que la qualité d'ajustement est moyenne puisque le R^2 est de 0.50. Le premier coefficient n'est pas ici normalisé à l'unité, mais il serait possible de déterminer une nouvelle représentation où le poids du choc contemporains serait unitaire.

Ainsi, un simple modèle de ce type permet ainsi de proposer une première représentation possible du PIB français, si l'on admet que ce dernier peut être représenté par un processus stationnaire. Mais attention, *cette approximation que l'on qualifie de $MA(20)$, ne correspond pas à la véritable décomposition de Wold* puisque celle-ci est par définition infinie.

Figure 4.12: Estimation d'un Processus MA

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.016481	0.003098	5.319266	0.0000
MA(1)	0.405355	0.135034	3.009283	0.0039
MA(2)	0.160709	0.147399	1.080300	0.2803
MA(3)	-0.033793	0.147542	-0.229089	0.8197
MA(4)	0.412280	0.145879	2.826173	0.0065
MA(5)	-0.065285	0.165882	-0.395113	0.7236
MA(6)	0.299534	0.149528	2.003194	0.0500
MA(7)	0.191218	0.152549	1.252666	0.2195
MA(8)	-0.206489	0.152592	-1.352206	0.1814
MA(9)	0.452963	0.157323	2.879183	0.0066
MA(10)	-0.339223	0.167262	-2.028097	0.0473
MA(11)	0.114034	0.170505	0.668809	0.5064
MA(12)	-0.047615	0.162015	-0.293696	0.7689
MA(13)	0.344490	0.166484	2.069146	0.0432
MA(14)	0.177854	0.170351	1.044044	0.3010
MA(15)	0.504153	0.168807	2.986568	0.0042
MA(16)	0.051337	0.182075	0.501649	0.6179
MA(17)	-0.066000	0.174953	-0.377245	0.7074
MA(18)	0.134403	0.174677	0.769439	0.4449
MA(19)	0.061061	0.171203	0.356660	0.7227
MA(20)	0.164617	0.178551	0.921959	0.3605
R-squared	0.503179	Mean dependent var	0.016043	
Adjusted R-squared	0.332527	S.D. dependent var	0.012742	
S.E. of regression	0.010410	Akaike info criterion	-6.065036	
Sum squared resid	0.006069	Schwarz criterion	-5.425816	
Log likelihood	254.5039	F-statistic	2.693116	
Durbin-Watson stat	1.963246	Prob(F-statistic)	0.000923	

La figure (4.13) montre quelle peut être alors la qualité de l'ajustement, et donc par conséquent des prévisions que l'on peut en tirer.

4.3 Opérateur Retard et Décomposition de Wold

L'énoncé du théorème de Wold est souvent donné en introduisant *un polynôme défini en l'opérateur retard*. Plus généralement, les modèles de séries temporels sont souvent exprimés sous la forme de *polynôme retard*. Nous commencerons donc par définir ce qu'est l'opérateur retard, puis nous réexprimerons le théorème de Wold à l'aide d'un polynôme retard.

4.3.1 Définition de l'opérateur retard

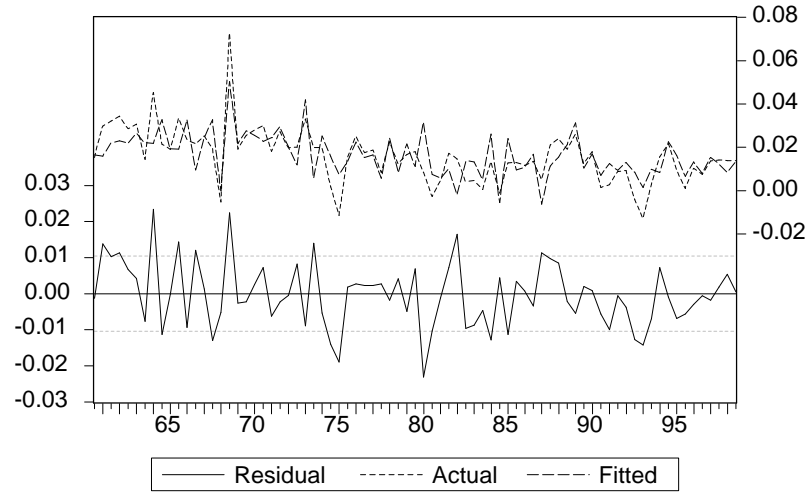
L'opérateur retard (noté L pour *lag* ou B suivant les ouvrages) est défini de la façon suivante :

Definition 32 *On considère un processus stochastique $(x_t, t \in \mathbb{Z})$, l'opérateur retard, noté L , est défini par la relation :*

$$Lx_t = x_{t-1} \quad \forall t \in \mathbb{Z} \quad (4.104)$$

Cet opérateur permet ainsi de définir une application qui à toute variable x_t associe la variable retardée (d'où son nom) x_{t-1} . Cet opérateur possède les propriétés suivantes. On considère un processus stochastique stationnaire $(x_t, t \in \mathbb{Z})$

Propriété 1. $L^j x_t = x_{t-j}, \forall j \in \mathbb{Z}$, en particulier on a $L^0 x_t = x_t$

Figure 4.13: Estimation du PIB français par un MA(q)


Propriété 2. Si $x_t = c, \forall t \in \mathbb{Z}$ avec $c \in \mathbb{R}$, $L^j x_t = L^j c = c, \forall j \in \mathbb{Z}$

Propriété 3. $L^i (L^j x_t) = L^{i+j} x_t = x_{t-i-j} \quad \forall (i, j) \in \mathbb{Z}^2$

La preuve est la suivante :

$$L^i (L^j x_t) = L^i (x_{t-j}) = x_{t-i-j}$$

Propriété 4. $L^{-i} x_t = x_{t+i} \quad \forall i \in \mathbb{Z}$.

Propriété 5. $(L^i + L^j) x_t = L^i x_t + L^j x_t = x_{t-i} + x_{t-j} \quad \forall (i, j) \in \mathbb{Z}^2$

Propriété 6. Si $|a| < 1$

$$(1 - aL)^{-1} x_t = \frac{x_t}{(1 - aL)} = \lim_{j \rightarrow \infty} (1 + aL + a^2 L^2 + \dots + a^j L^j) x_t$$

Cette dernière propriété est particulièrement utile pour inverser des polynômes d'ordre 1 définis en l'opérateur retard. La démonstration de cette propriété est la suivante. Démontrons que :

$$(1 - aL) \lim_{j \rightarrow \infty} (1 + aL + a^2 L^2 + \dots + a^j L^j) x_t = x_t$$

En développant le terme de gauche, il vient :

$$\lim_{j \rightarrow \infty} (1 - aL + aL - a^2 L^2 + a^2 L^2 + \dots - a^{j-1} L^{j-1} + a^{j-1} L^{j-1} + a^j L^j) x_t = \lim_{j \rightarrow \infty} (1 + a^j L^j) x_t$$

De là, on montre que sous l'hypothèse $|a| < 1$, on a bien :

$$\lim_{j \rightarrow \infty} (1 + a^j L^j) x_t = x_t$$

4.3.2 Théorème de Wold et polynômes retard

Partant de la définition de cet opérateur retard, on peut alors construire des *polynômes à coefficients réels en cet opérateur*. Par exemple, considérons un polynôme $\Psi(L)$ d'ordre q :

$$\Psi(L) = \psi_0 L^0 + \psi_1 L^1 + \psi_2 L^2 + \dots + \psi_q L^q$$

L'application de ce polynôme à une processus aléatoire $(\varepsilon_t, t \in \mathbb{Z})$ nous donne donc :

$$\begin{aligned} \Psi(L) \varepsilon_t &= \psi_0 L^0 \varepsilon_t + \psi_1 L^1 \varepsilon_t + \psi_2 L^2 \varepsilon_t + \dots + \psi_q L^q \varepsilon_t \\ &= \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots + \psi_q \varepsilon_{t-q}. \end{aligned}$$

Si $\psi_0 = 1$ et $q \rightarrow \infty$ on retrouve ici une *forme moyenne mobile infinie* identique à celle du théorème de Wold. Le nouvel énoncé de ce théorème (identique, bien entendu, au précédent) est le suivant :

Theorem 33 (Décomposition de Wold) *Tout processus stationnaire d'ordre deux $(x_t, t \in \mathbb{Z})$ peut être représenté sous la forme :*

$$x_t = \Psi(L) \varepsilon_t + \kappa_t \tag{4.105}$$

où le polynôme de degré infini $\Psi(L)$ est défini par $\Psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$, les paramètres ψ_j satisfaisant $\psi_0 = 1, \psi_j \in \mathbb{R}, \forall j \in \mathbb{N}^, \sum_{j=0}^{\infty} \psi_j^2 < \infty$ et où ε_t est un bruit blanc i.i.d. $(0, \sigma_\varepsilon^2)$ et où la composante linéaire déterministe κ_t vérifie $\text{cov}(\kappa_t, \varepsilon_{t-j}) = 0, \forall j \in \mathbb{Z}$.*

5 Les processus ARMA

Jusqu'à présent nous avons vu que tout processus stationnaire pouvait s'écrire sous la forme d'une somme pondérée infinie de chocs passés (théorème de Wold). Pour toute cette classe de processus la décomposition de Wold est donc une première représentation possible. Toutefois, *cette représentation n'est jamais la représentation optimale parmi toutes les représentations possibles d'un même processus.*

En effet, nous verrons dans le troisième chapitre, que lorsque que l'on cherche à modéliser une série temporelle, on applique toujours *un principe de parcimonie* qui veut que, à qualité d'ajustement égale, l'on adopte en priorité la représentation nécessitant l'estimation du minimum de paramètres. Or par définition, si l'on devait appliquer la décomposition de Wold, cela supposerait que l'on estime une infinité de paramètres (les ψ_j et le σ_ε^2). Donc dans la pratique, il convient de rechercher d'autres représentations possibles pour les processus temporels.

Parmi les représentations les plus utilisées figurent les représentations ARMA pour Autoregressive Moving Average. Cette représentation consiste en l'adjonction d'une composante autorégressive d'ordre fini (AR) et d'une composante moyenne mobile d'ordre fini (MA).

Nous allons donc commencer par définir la classe des processus AR, MA et ARMA. Nous étudierons ensuite sous quels conditions ces processus satisfont l'hypothèse de stationnarité. Enfin, nous introduirons des processus ARMA saisonniers : les SARMA.

5.1 Définition des processus ARMA

Définissons à présent la classe des processus AR, MA et ARMA.

5.1.1 Les processus MA

La définition générale d'un processus MA est la suivante :

Definition 34 *Le processus $(x_t, t \in \mathbb{Z})$ satisfait une représentation MA d'ordre q , notée $MA(q)$, si et seulement si :*

$$x_t = m + \Theta(L) \varepsilon_t \quad (5.106)$$

avec $E(x_t) = m$. Le polynôme $\Theta(L)$ étant défini par $\Theta(L) = \sum_{j=0}^q \theta_j L^j$ où $\forall j < q, \theta_j \in \mathbb{R}, \theta_0 = 1$ et $\theta_q \in \mathbb{R}^$, avec ε_t i.i.d. $(0, \sigma_\varepsilon^2)$.*

On reconnaît ici la forme moyenne mobile que nous avons identifié dans la décomposition de Wold, mais qui dans le cas présent est un *forme moyenne mobile d'ordre q fini*. Par opposition, la *décomposition de Wold, définie par l'équation (4.97), est aussi qualifiée de $MA(\infty)$*

Les différentes conditions sur les paramètres θ_j signifient que dans un $MA(q)$, tous les paramètres du polynôme $\Theta(L)$ peuvent être nuls à l'exception du paramètre correspondant au $q^{\text{ème}}$ retard. En effet, si $\theta_q = 0$, alors le processus $(x_t, t \in \mathbb{Z})$ satisfait une représentation $MA(q-1)$. De la même façon que pour la décomposition de Wold, le premier coefficient θ_0 associé au choc contemporain est, par convention, normalisé à l'unité.

Pour un processus $MA(q)$, la constante m correspond à l'espérance du processus puisque les bruits blancs sont par définition centrés $E(\varepsilon_{t-j}) = 0, \forall j$:

$$E(x_t) = E\left[m + \sum_{j=0}^q \theta_j \varepsilon_{t-j}\right] = m + \sum_{j=0}^q \theta_j E(\varepsilon_{t-j}) = m$$

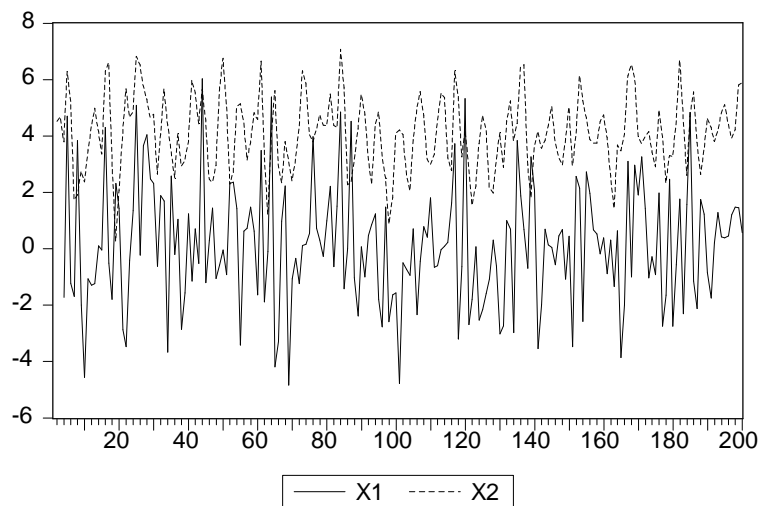
Voici deux exemples de processus $MA(q)$:

$x_{1,t} = \varepsilon_t + 2\varepsilon_{t-3}$: $x_{1,t}$ est un processus $MA(3)$ de moyenne nulle

$x_{2,t} = 4 + \varepsilon_t + 0.8\varepsilon_{t-1}$: $x_{2,t}$ est un processus $MA(1)$ de moyenne 4

Une simulation de taille $T = 200$ de ces deux processus pour $\varepsilon_t \sim \mathcal{N}(0, 1)$ est représentée sur la figure (5.14). On vérifie au passage que ces deux processus semblent stationnaires.

Figure 5.14: Simulation de Processus $MA(q)$



5.1.2 Les processus AR

La définition générale d'un processus AR est la suivante :

Definition 35 *Le processus stationnaire* $(x_t, t \in \mathbb{Z})$ *satisfait une représentation AR d'ordre* p , *notée* $AR(p)$, *si et seulement si :*

$$\Phi(L)x_t = c + \varepsilon_t \quad (5.107)$$

avec $c \in \mathbb{R}$, $\Phi(L) = \sum_{j=0}^p \phi_j L^j$ où $\forall j < p$, $\phi_j \in \mathbb{R}$, $\phi_0 = 1$ et $\phi_p \in \mathbb{R}^*$, avec ε_t i.i.d. $(0, \sigma_\varepsilon^2)$.

On parle ici de *représentation autorégressive*, dans le sens où la variable x_t est déterminée par les valeurs passées $x_{t-1}, x_{t-2}, \dots, x_{t-p}$. Les différentes conditions sur les paramètres ϕ_j signifient que dans un $AR(p)$, tous les paramètres du polynôme $\Phi(L)$ peuvent être nuls à l'exception du paramètre correspondant au $p^{\text{ème}}$ retard. En effet, si $\phi_p = 0$, alors le processus $(x_t, t \in \mathbb{Z})$ satisfait une représentation $AR(p-1)$. De la même façon que pour la décomposition de Wold, le premier coefficient θ_0 associé au choc contemporain est, par convention, normalisé à l'unité.

Contrairement au cas du MA , la constante c de l'équation (5.107) ne correspond pas à l'espérance du processus $(x_t, t \in \mathbb{Z})$. Cette espérance, dans le cas d'un processus stationnaire, est déterminée par la relation suivante si $\Phi(1) \neq 0$:

$$E(x_t) = \Phi(L)^{-1} c = \Phi(1)^{-1} c \quad (5.108)$$

En effet, reprenons l'écriture de l' $AR(p)$ donnée par l'équation (5.107) :

$$E(x_t) = c - \phi_1 E(x_{t-1}) - \phi_2 E(x_{t-2}) - \dots - \phi_p E(x_{t-p}) + E(\varepsilon_t)$$

Par définition $E(\varepsilon_t) = 0$. Si l'on suppose maintenant que $(x_t, t \in \mathbb{Z})$, son espérance, que l'on notera m , ne dépend pas du temps :

$$E(x_{t-j}) = m \quad \forall j \in \mathbb{Z}$$

Donc on a :

$$E(x_t) + \phi_1 E(x_{t-1}) + \phi_2 E(x_{t-2}) + \dots + \phi_p E(x_{t-p}) = m + \phi_1 m + \phi_2 m + \dots + \phi_p m = c$$

Ce qui peut se récrire sous la forme :

$$E(x_t) = m = \frac{c}{1 + \phi_1 + \phi_2 + \dots + \phi_p}$$

On retrouve bien alors la formule générale puisque

$$\Phi(1)^{-1} = \left[\sum_{j=0}^p \phi_j \right]^{-1} = \frac{1}{\phi_0 + \phi_1 + \phi_2 + \dots + \phi_p} \quad \text{avec } \phi_0 = 1$$

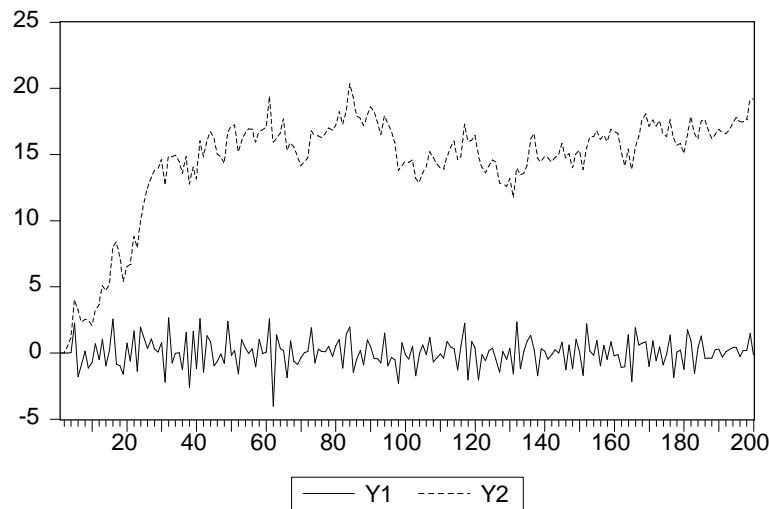
Voici deux exemples de processus $AR(p)$:

$$y_{1,t} = -0.5y_{1,t-1} + 0.2y_{1,t-3} + \varepsilon_t : y_{1,t} \text{ est un processus } AR(3) \text{ de moyenne nulle}$$

$y_{2,t} = 0.8 + 0.7y_{2,t-1} + 0.25y_{2,t-2} + \varepsilon_t$: $y_{2,t}$ est un processus $AR(2)$ d'espérance $0.8 / (1 - 0.7 - 0.25) = 16$.

Une simulation de taille $T = 200$ de ces deux processus pour $\varepsilon_t \sim \mathcal{N}(0, 1)$ est représentée sur la figure (5.15). Ces deux processus semblent stationnaires autour de leur moyenne respective (0 et 16), et cela si l'on excepte la dynamique sur les premières périodes liées aux valeurs initiales nulles dans les deux cas. *Mais nous verrons que contrairement aux processus $MA(q)$ ce résultat n'est pas général.* Il existe des processus AR non stationnaires.

Figure 5.15: Simulations de Processus $AR(p)$



5.1.3 Les processus ARMA

Naturellement, les processus $ARMA$ se définissent par l'adjonction d'une composante autorégressive AR et d'une composante moyenne mobile MA .

Definition 36 *Le processus stationnaire $(x_t, t \in \mathbb{Z})$ satisfait une représentation ARMA, d'ordre p et q , notée $ARMA(p, q)$, si et seulement si :*

$$\Phi(L)x_t = c + \Theta(L)\varepsilon_t \quad (5.109)$$

avec $c \in \mathbb{R}$, $\Theta(L) = \sum_{j=0}^q \theta_j L^j$, $\Phi(L) = \sum_{j=0}^p \phi_j L^j$ où $\forall j < q, \theta_j \in \mathbb{R}^2, \forall j < p, \phi_j \in \mathbb{R}^2$, $\theta_0 = \phi_0 = 1$ et $(\phi_p, \theta_q) \in \mathbb{R}^{2}$, avec ε_t i.i.d. $(0, \sigma_\varepsilon^2)$.*

Les différentes conditions sur les paramètres ϕ_j et θ_j signifient que dans un $ARMA(p, q)$, tous les paramètres du polynôme $\Phi(L)$ peuvent être nuls à l'exception du paramètre correspondant au $p^{\text{ème}}$ retard et que tous les paramètres du polynôme $\Theta(L)$ peuvent être nuls à

l'exception du paramètre correspondant au $q^{\text{ème}}$ retard . En effet, si $\phi_p = 0$, alors le processus $(x_t, t \in \mathbb{Z})$ satisfait une représentation $ARMA(p - 1, q)$. De la même façon, si $\theta_q = 0$, alors le processus $(x_t, t \in \mathbb{Z})$ satisfait une représentation $ARMA(p, q - 1)$.

Ainsi, on constate que les processus AR et MA ne sont que des cas particuliers des processus $ARMA$. Un $AR(p)$ correspond à un $ARMA(p, 0)$, de la même façon un $MA(q)$ correspond à un $ARMA(0, q)$.

Tout comme pour les AR , la constante c ne correspond pas dans ce cas à l'espérance du processus $(x_t, t \in \mathbb{Z})$. Cette dernière est donnée par la même relation que celle obtenue pour l' AR :

$$E(x_t) = \Phi(1)^{-1} c \tag{5.110}$$

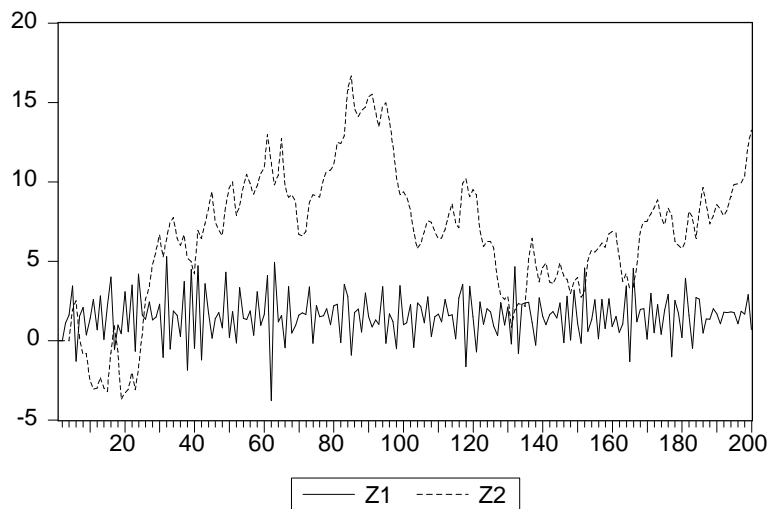
Voici deux exemples de processus $ARMA(p, q)$:

$z_{1,t} = 2 - 0.5z_{1,t-1} + 0.2z_{1,t-3} + \varepsilon_t - 0.5\varepsilon_{t-1}$: $z_{1,t}$ est un processus $ARMA(3, 1)$ de moyenne nulle $2 / (1 + 0.5 - 0.2) = 1.5385$

$z_{2,t} - 0.2z_{2,t-1} - 1.5z_{2,t-2} = \varepsilon_t + 0.5\varepsilon_{t-1} + 0.6\varepsilon_{t-4}$ est un $ARMA(2, 4)$ de moyenne nulle

Une simulation de taille $T = 200$ de ces deux processus pour $\varepsilon_t \sim \mathcal{N}(0, 1)$ est représentée sur la figure (5.16). Dans ce cas, on constate que le processus z_1 est centré autour de sa moyenne et semble-t-être un processus stationnaire. En revanche, le processus z_2 , de moyenne nulle, ne semble pas être un processus stationnaire. En effet, comme nous allons le voir il existe certaines conditions sur les processus $ARMA$ pour que ces derniers soient stationnaires. Ici, ces conditions ne sont pas satisfaites.

Figure 5.16: Simulations de Processus $ARMA(p, q)$



5.1.4 Les processus ARMA saisonniers

Il existe en outre des processus ARMA saisonniers ou processus SARMA

Definition 37 *Un processus stationnaire $(x_t, t \in \mathbb{Z})$ satisfait une représentation ARMA saisonnière (ou SARMA), notée $ARMA_{s,s'}(p, q)$, si*

$$\sum_{j=0}^p \phi_{js} x_{t-js} = c + \sum_{j=0}^q \theta_{js'} \varepsilon_{t-js'} \quad (5.111)$$

avec $c \in \mathbb{R}$, $\forall j < p$, $\phi_j \in \mathbb{R}$, $\forall j < q$, $\theta_j \in \mathbb{R}$, $\phi_0 = \theta_0 = 1$ et $(\phi_p, \theta_q) \in \mathbb{R}^*$, $\{\varepsilon_t\}$ i.i.d. $(0, \sigma_\varepsilon^2)$ et où s désigne la période de saisonnalité de la composante AR et s' désigne la période de saisonnalité de la composante MA.

Ainsi un $ARMA_{12,6}(2, 2)$ s'écrit :

$$x_t + \phi_{12} x_{t-12} + \phi_{24} x_{t-24} = c + \varepsilon_t + \theta_6 \varepsilon_{t-6} - \theta_{12} \varepsilon_{t-12}$$

Bien entendu, on peut construire des processus ARMA non saisonniers et saisonniers à la fois. Il en existe deux sortes. Les processus ARMA additifs (rarement employés) :

$$ARMA(p, q) + ARMA_{s,s'}(\tilde{p}, \tilde{q})$$

et les processus ARMA multiplicatifs

$$ARMA(p, q) \times ARMA_{s,s'}(\tilde{p}, \tilde{q})$$

Les processus ARMA multiplicatifs sont équivalents à des $ARMA(p + \tilde{p}s, q + \tilde{q}s)$.

5.2 La stationnarité et l'inversibilité des processus ARMA

Considérons à présent des processus ARMA. La question est alors de savoir sous quelles conditions sur les paramètres des polynômes $\Phi(L)$ et $\Theta(L)$ ces processus sont ils stationnaires. Nous allons en outre introduire la notion d'inversibilité qui consiste à déterminer s'il existe une représentation MA (respectivement AR) équivalente pour un AR (respectivement pour un MA).

5.2.1 Conditions de stationnarité et d'inversibilité

Commençons par énoncer un premier théorème concernant les processus AR.

Theorem 38 *Un processus $(x_t, t \in \mathbb{Z})$ satisfaisant une représentation AR(p) est toujours inversible. Il est stationnaire lorsque toutes les racines du polynôme $\Phi(L)$, notées $\lambda_j \in \mathbb{C}$, $\forall j \leq p$, sont de module strictement supérieur à l'unité.*

$$\Phi(\lambda_j) = \sum_{i=0}^p \phi_i \lambda_j^i = 0 \iff \prod_{j=1}^p \left(1 - \frac{1}{\lambda_j} L\right) = 0 \quad |\lambda_j| > 1, \forall j \quad (5.112)$$

Ce théorème indique tout d'abord que tout processus AR peut être "inversé" et représenté sous la forme d'un processus MA . Le second point, plus fondamental, porte sur la stationnarité du processus. Ce théorème indique que *lorsque les racines du polynôme autorégressif sont toutes supérieures à l'unité en module, le processus satisfait les 3 conditions de la stationnarité du second ordre.*

Nous ne fournirons pas de démonstration générale, mais nous considérerons un exemple simple d'un processus $AR(1)$ centré défini par :

$$\Phi(L)x_t = (1 - \phi L)x_t = \varepsilon_t \quad (5.113)$$

On sait que si le processus x_t est stationnaire, il peut être exprimé sous la forme d'un $MA(\infty)$, suivant le théorème de Wold. D'après le théorème précédent la condition de stationnarité se ramène à montrer que la racine de $\Phi(L)$, notée λ , est strictement supérieure à l'unité en module :

$$\Phi(\lambda) = 1 - \phi\lambda = 0 \iff \lambda = \frac{1}{\phi}$$

Donc si $|\phi| < 1$, alors $|\lambda| > 1$, le processus x_t est stationnaire et peut s'écrire sous la forme $MA(\infty)$ suivante :

$$x_t = \Phi(L)^{-1} \varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

Les paramètres ψ_j sont alors définis par inversion du polynôme $\Phi(L)$. D'après la 6^{ème} propriété de l'opérateur retard, on a $\forall j \in \mathbb{N}$:

$$\psi_j = \phi^j$$

puisque

$$\Phi(L)^{-1} = \frac{1}{1 - \phi L} = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j} \quad (5.114)$$

Reste à démontrer que sous l'hypothèse $|\phi| < 1$, le processus $(x_t, t \in \mathbb{Z})$ qui peut s'écrire sous la forme $\forall t \in \mathbb{Z}$:

$$x_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$$

satisfait les trois conditions de la définition de la stationnarité du second ordre. La première condition, dite condition de sommabilité, est immédiate :

$$\sum_{j=0}^{\infty} \psi_j^2 = \sum_{j=0}^{\infty} \phi^{2j} < \infty$$

Sous l'hypothèse $|\phi| < 1$ cette condition est satisfaite. La seconde condition porte sur le moment d'ordre un :

$$E(x_t) = E \left[\sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j} \right] = \sum_{j=0}^{\infty} \phi^j E(\varepsilon_{t-j}) = 0 \text{ indépendant de } t$$

Enfin, la troisième condition porte sur les moments d'ordre deux et sur la fonction d'autocovariance :

$$\begin{aligned} \gamma(h) &= E(x_{t+h}x_t) = E\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{t+h-j} \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}\right) \quad \forall h \in \mathbb{Z} \\ \Leftrightarrow \gamma(h) &= \begin{cases} \sum_{j=0}^{\infty} \psi_j^2 E(\varepsilon_{t-j}^2) & \text{si } h = 0 \\ \sum_{j=0}^{\infty} \psi_{j+h} \psi_j E(\varepsilon_{t-j}^2) & \forall h \neq 0 \end{cases} \end{aligned}$$

D'ou finalement on tire que :

$$\gamma(h) = \begin{cases} \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \phi^{2j} & \text{si } h = 0 \\ \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \phi^{2j+h} & \forall h \neq 0 \end{cases}$$

La fonction d'autocovariance est donc bien indépendante du temps. Donc sous l'hypothèse $|\phi| < 1$, le processus $(x_t, t \in \mathbb{Z})$ est stationnaire.

Considérons trois exemples de processus AR définis par les relations suivantes :

$$x_{1,t} = 2x_{1,t-1} - x_{1,t-2} + \varepsilon_t$$

$$x_{2,t} = 2x_{2,t-1} - 3x_{2,t-2} + 2x_{2,t-3} + \varepsilon_t$$

$$x_{3,t} = 0.1x_{3,t-1} - 0.5x_{3,t-2} + \varepsilon_t$$

Les trois polynômes autorégressifs associés sont les suivants :

$$\Phi_i(L) x_{i,t} = \varepsilon_t \quad i = 1, 2, 3$$

avec

$$\Phi_1(L) = 1 - 2L + L^2 = 0 \Leftrightarrow \lambda_1^1 = \lambda_1^2 = 1$$

$$\Phi_2(L) = 1 - 2L + 3L^2 - 2L^3 \Leftrightarrow \lambda_1^2 = 1 \text{ et } \lambda_j^2 = 0.25 \pm 0.25i\sqrt{7} \quad j = 2, 3$$

$$\Phi_3(L) = 1 - 0.1L + 0.5L^2 \Leftrightarrow \lambda_j^3 = 0.1 \pm 1.4107i \quad j = 1, 2$$

Par conséquent, le processus $(x_{1,t}, t \in \mathbb{Z})$ est non stationnaire puisque les deux racines sont unitaires. Le processus $(x_{2,t}, t \in \mathbb{Z})$ est aussi non stationnaire puisqu'il a une racine unitaire et deux racine de module strictement inférieures à un. Seul le processus $(x_{3,t}, t \in \mathbb{Z})$ est stationnaire puisque ses deux racines sont de module strictement supérieures à l'unité.

Un second théorème peut être énoncé pour les processus MA.

Theorem 39 *Un processus $(x_t, t \in \mathbb{Z})$ satisfaisant une représentation $MA(q)$ est toujours stationnaire. Il est inversible lorsque toutes les racines du polynôme $\Theta(L)$, notées $\beta_j \in \mathbb{C}, \forall j \leq q$, sont de module strictement supérieure à l'unité.*

$$\Theta(\beta_j) = \sum_{i=0}^q \theta_i \beta_j^i = 0 \iff \prod_{j=1}^q \left(1 - \frac{1}{\beta_j} L\right) = 0 \quad |\beta_j| > 1, \forall j$$

Ce théorème nous indique tout d'abord que tout processus MA , qui n'est autre qu'une somme pondérée de bruits blancs, est toujours stationnaires. Ce résultat se comprend aisément compte tenu des propriétés des bruits blancs. La seconde partie nous indique qu'un processus x_t défini comme un MA peut être exprimé sous la forme d'un AR si le polynôme associé $\Theta(L)$ est inversible, c'est à dire si ses racines sont toutes supérieures à 1 en module.

Considère l'exemple d'un processus centré $(x_t, t \in \mathbb{Z})$ suivant un $MA(1)$ défini par :

$$x_t = \Theta(L) \varepsilon_t = (1 - \theta L) \varepsilon_t \quad (5.115)$$

Par inversion du polynôme $\Theta(L)$, on peut exprimer x_t sous la forme d'un $AR(\infty)$:

$$\Theta(L)^{-1} x_t = \varepsilon_t$$

Déterminons l'inverse de $\Theta(L)$. De la même façon que précédemment, on montre que si $|\theta| < 1$, $\Theta(L)$ est inversible et son inverse est :

$$(1 - \theta L)^{-1} = \sum_{j=0}^{\infty} \theta^j L^j$$

Dès lors, le processus x_t peut être défini comme un $AR(\infty)$ dont les paramètres sont donnés par la relation suivante :

$$\Theta(L)^{-1} x_t = \sum_{j=0}^{\infty} \theta^j x_{t-j} = \varepsilon_t \quad (5.116)$$

Bien entendu, il nous reste maintenant à établir les mêmes résultats en ce qui concerne les processus $ARMA$.

Corollary 40 *Les conditions de stationnarité d'un processus ARMA sont déterminées par les racines du polynôme associé à sa composante AR. Les conditions d'inversibilité sont déterminées par les racines du polynôme associé à sa composante MA.*

Ainsi, un processus centré $(x_t, t \in \mathbb{Z})$ défini par un $ARMA(p, q)$ est stationnaire dès lors que les racines du polynômes associé à la composante AR sont strictement supérieures à l'unité en module.

Reprenons les deux exemples étudiés précédemment :

$$z_{1,t} = 2 - 0.5z_{1,t-1} + 0.2z_{1,t-3} + \varepsilon_t - 0.5\varepsilon_{t-1}$$

$$z_{2,t} - 0.2z_{2,t-1} - 1.5z_{2,t-2} = \varepsilon_t + 0.5\varepsilon_{t-1} + 0.6\varepsilon_{t-4}$$

La représentation de ces deux processus est la suivante :

$$\Phi_i(L) z_{i,t} = \Theta_i(L) \varepsilon_t \quad i = 1, 2$$

avec

$$\Phi_1(L) = 1 + 0.5L - 0.2L^3 = 0 \iff \lambda_1^1 = 2.18 \text{ et } \lambda_j^1 = -1.09 \pm 1.04i \quad j = 2, 3$$

$$\Phi_2(L) = 1 - 0.2L - 1.5L^2 = 0 \iff \lambda_1^2 = -0.88 \text{ et } \lambda_2^2 = 0.75$$

On vérifie au passage la conclusion que nous avons faite précédemment, à savoir que le processus z_1 est stationnaire, tandis que le processus z_2 ne l'est pas.

5.2.2 Recherche des conditions de stationnarité et d'inversibilité des processus ARMA

Ainsi, la recherche des conditions de stationnarité (ou des conditions d'inversibilité suivant les cas) revient à déterminer la position relative des racines d'un polynôme défini en l'opérateur retard par rapport au cercle unité du plan complexe.

On considère un polynôme $\Phi(L)$ d'ordre p à coefficients réels, défini en l'opérateur retard L , avec $\phi_0 = 1$. Soit $(\lambda_i)_i$ l'ensemble des racines de ce polynôme avec $(\lambda_i) \in \mathbb{C}^p$

$$\Phi(L) = \sum_{j=0}^p \phi_j L^j = \prod_{j=1}^p \left(1 - \frac{1}{\lambda_j} L\right) = 0 \quad (5.117)$$

Dans le cas $p = 2$, les racines du polynôme s'obtiennent directement à partir de l'équation, $\forall z \in \mathbb{C}$:

$$1 + \phi_1 z + \phi_2 z^2 = \left(1 - \frac{z}{\lambda_1}\right) \left(1 - \frac{z}{\lambda_2}\right) = 0$$

Suivant la valeur de la quantité $\Delta = \phi_1^2 - 4\phi_2$, on montre que :

$$\begin{aligned} \Delta > 0 \quad \lambda_i &= 2\phi_2 \left[-\phi_1 \pm \sqrt{\phi_1^2 - 4\phi_2} \right]^{-1} \quad i = 1, 2 \\ \Delta < 0 \quad \lambda_i &= 2\phi_2 \left[-\phi_1 \pm i\sqrt{-\phi_1^2 + 4\phi_2} \right]^{-1} \quad i = 1, 2 \end{aligned}$$

Dans le cas général, les racines du polynôme à coefficients réels d'ordre p , noté $\phi(z)$, correspondent aux valeurs propres de la matrice F associée à l'application linéaire

$$Z = FZ \quad (5.118)$$

avec

$$\underset{(p,1)}{Z} = \begin{pmatrix} z \\ z^2 \\ \dots \\ z^p \end{pmatrix} \quad \underset{(p,p)}{F} = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_p \\ 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & 1 & 0 \end{pmatrix}$$

5.2.3 Inversion d'un polynôme d'ordre p

En analyse des séries temporelles, il est souvent utile *d'inverser* des processus. Par exemple, partant d'un processus *AR* stationnaire, on peut par inversion du polynôme autorégressif, déterminer la forme *MA*(∞) associée à la décomposition de Wold. On obtient ainsi *des représentations équivalentes d'un même processus*. Pour cela, il est nécessaire d'inverser des polynômes définis en l'opérateur retard. Nous avons déjà vu comment réaliser cette opération pour des polynômes de degré un. Nous allons à présent généraliser cette méthode au cas de polynôme de degré supérieur ou égal à un.

Le problème est donc le suivant. Soit $\Phi(z)$ un polynôme inversible d'ordre p à coefficients réels avec $\phi_0 = 1$. Il s'agit de déterminer $\tilde{\Phi}(z)$, le polynôme inverse de $\Phi(z)$. Par définition, $\forall z \in \mathbb{C}$

$$\Phi(z) \tilde{\Phi}(z) = \Phi(z) \Phi(z)^{-1} = \sum_{j=0}^p \phi_j z^j \sum_{j=0}^{\infty} \tilde{\phi}_j z^j = 1$$

Plusieurs solutions existent pour déterminer $\tilde{\Phi}(z)$. Parmi celles-ci, nous n'en retiendrons que deux.

1. Méthode d'identification :

On part de la relation $\Phi(z) \tilde{\Phi}(z) = 1$ et l'on cherche à déterminer les paramètres $\tilde{\phi}_j$, tels que $\sum_{j=0}^n \tilde{\phi}_j z^j$, par identification les coefficients de même degré. De façon générale, les n équations du système d'identification sont définies par :

$$\begin{aligned} \sum_{k=0}^i \tilde{\phi}_{i-k} \phi_k &= 0 \quad \forall i \in [1, p] \\ \sum_{k=0}^p \tilde{\phi}_{i-k} \phi_k &= 0 \quad \forall i > p \end{aligned} \tag{5.119}$$

Considérons l'exemple suivant. On cherche à inverser le polynôme *AR*(2) défini par

$$\Phi(z) x_t = (1 + \phi_1 z + \phi_2 z^2) x_t = \varepsilon_t$$

avec $\phi_1 = 0.6$ et $\phi_2 = -0.3$. Le polynôme est inversible puisque les racines sont de module strictement supérieur à 1 : $\lambda_1 = -1.23$ et $\lambda_2 = 3.23$. Soit $\tilde{\Phi}(z)$ le polynôme inverse de $\Phi(z)$, que l'on suppose de degré infini. On part de la relation d'identification :

$$\Phi(z) \tilde{\Phi}(z) = 1$$

En développant on obtient :

$$\begin{aligned}
 & (1 + \phi_1 z + \phi_2 z^2) \left(\tilde{\phi}_0 + \tilde{\phi}_1 z + \tilde{\phi}_2 z^2 + \tilde{\phi}_3 z^3 + \dots + \tilde{\phi}_p z^p + \dots \right) = 1 \\
 \Leftrightarrow & \quad \tilde{\phi}_0 + \tilde{\phi}_1 z + \tilde{\phi}_2 z^2 + \tilde{\phi}_3 z^3 + \dots + \tilde{\phi}_p z^p + \dots \\
 & \quad \tilde{\phi}_0 \tilde{\phi}_1 z + \tilde{\phi}_1^2 z^2 + \tilde{\phi}_2 \tilde{\phi}_1 z^3 + \tilde{\phi}_3 \tilde{\phi}_1 z^4 + \dots + \tilde{\phi}_p \tilde{\phi}_1 z^{p+1} + \dots \\
 & \quad \tilde{\phi}_0 \tilde{\phi}_2 z^2 + \tilde{\phi}_1 \tilde{\phi}_2 z^3 + \tilde{\phi}_2^2 z^4 + \tilde{\phi}_3 \tilde{\phi}_2 z^5 + \dots + \tilde{\phi}_p \tilde{\phi}_2 z^{p+2} + \dots = 1
 \end{aligned}$$

Par identification des termes de même degré à droite et à gauche du signe égal, on obtient alors le système suivant :

$$\begin{cases} \tilde{\phi}_0 = 1 \\ \tilde{\phi}_1 + \phi_1 = 0 \\ \tilde{\phi}_2 + \phi_1 \phi_1 + \phi_2 = 0 \\ \tilde{\phi}_3 + \phi_2 \phi_1 + \phi_1 \phi_2 = 0 \\ \dots \\ \tilde{\phi}_n + \phi_{n-1} \phi_1 + \phi_{n-2} \phi_2 = 0 \quad \forall n > 2 \end{cases}$$

La résolution de ce système fournit alors une suite de récurrence qui définit les coefficients de la représentation $MA(\infty)$ du processus x_t :

$$x_t = \tilde{\Phi}(z) \varepsilon_t = \sum_{j=0}^{\infty} \tilde{\phi}_j \varepsilon_{t-j} \quad (5.120)$$

où les paramètres $\tilde{\phi}_j$ sont définis par la relation suivante :

$$\tilde{\phi}_0 = 1 \quad (5.121)$$

$$\tilde{\phi}_1 = -0.6 \quad (5.122)$$

$$\tilde{\phi}_n = -0.6 \tilde{\phi}_{n-1} + 0.3 \tilde{\phi}_{n-2} \quad \forall n \geq 2 \quad (5.123)$$

2. Méthode :

L'autre méthode d'inversion des polynômes est la suivante. On considère \tilde{p} racines λ_i réelles distinctes du polynôme $\Phi(z)$ d'ordre p , $(\lambda_i)_i \in \mathbb{R}^{\tilde{p}}$, avec $\tilde{p} \leq p$. On supposera ici pour simplifier que $\tilde{p} = p$, mais cette méthode peut être étendue au cas général $\tilde{p} \leq p$. Une autre façon d'obtenir les paramètres $\tilde{\phi}_i$ du polynôme inverse $\tilde{\Phi}(z) = \Phi(z)^{-1}$ consiste à déterminer les paramètres a_j tels que :

$$\tilde{\Phi}(z) = \frac{1}{\prod_{j=1}^p (1 - \tilde{\lambda}_j z)} = \sum_{j=1}^p \left(\frac{a_j}{1 - \tilde{\lambda}_j z} \right)$$

où les $\tilde{\lambda}_j$ sont définis comme l'inverse des racines de $\Phi(z)$:

$$\tilde{\lambda}_j = \frac{1}{\lambda_j} \quad \forall j \in [1, \tilde{p}] \quad (5.124)$$

De façon générale, les paramètres a_j sont définis par la relation suivante.

$$a_j = \frac{\tilde{\lambda}_j^{p-1}}{\prod_{\substack{k=1 \\ k \neq j}}^p (\tilde{\lambda}_j - \tilde{\lambda}_k)} \quad \forall j \leq p \quad (5.125)$$

Or on montre que :

$$\sum_{j=1}^p \left(\frac{a_j}{1 - \tilde{\lambda}_j z} \right) = \sum_{j=1}^p a_j \sum_{k=0}^{\infty} \tilde{\lambda}_j^k z^k = \sum_{k=0}^{\infty} \left[\sum_{j=1}^p a_j \tilde{\lambda}_j^k \right] z^k \quad (5.126)$$

Dès lors, par identification, on obtient :

$$\tilde{\phi}_i = \left[\sum_{j=1}^p a_j \tilde{\lambda}_j^i \right] \quad \forall i \geq 0 \quad (5.127)$$

Considère à nouveau l'exemple du polynôme $AR(2)$ défini par $\Phi(z) = 1 + \phi_1 z + \phi_2 z^2$, avec $\phi_1 = 0.6$ et $\phi_2 = -0.3$. Les deux racines réelles sont $\lambda_1 = -1.23$ et $\lambda_2 = 3.23$. On cherche tout d'abord à déterminer les paramètres a_i tels que $\forall z \in \mathbb{C}$

$$\frac{1}{(1 - \tilde{\lambda}_1 z)(1 - \tilde{\lambda}_2 z)} = \frac{a_1}{(1 - \tilde{\lambda}_1 z)} + \frac{a_2}{(1 - \tilde{\lambda}_2 z)}$$

avec

$$\begin{aligned} \tilde{\lambda}_1 &= \frac{1}{\lambda_1} = -\frac{1}{1.23} \\ \tilde{\lambda}_2 &= \frac{1}{\lambda_2} = \frac{1}{3.23} \end{aligned}$$

En développant, on obtient l'égalité suivante, $\forall z \neq \lambda_i, i = 1, 2$:

$$\begin{aligned} a_1 (1 - \tilde{\lambda}_2 z) + a_2 (1 - \tilde{\lambda}_1 z) &= 1 \\ \iff (a_1 + a_2) - (a_1 \tilde{\lambda}_2 + a_2 \tilde{\lambda}_1) z &= 1 \end{aligned}$$

Par identification des termes de même degré, on obtient le système :

$$\begin{cases} a_1 + a_2 = 1 \\ a_1 \tilde{\lambda}_2 + a_2 \tilde{\lambda}_1 = 0 \end{cases}$$

D'où l'on tire finalement que :

$$a_1 = \frac{\tilde{\lambda}_1}{\tilde{\lambda}_1 - \tilde{\lambda}_2} \quad a_2 = \frac{\tilde{\lambda}_2}{\tilde{\lambda}_2 - \tilde{\lambda}_1}$$

On peut alors construire le polynôme inverse.

$$\begin{aligned} \tilde{\Phi}(z) &= \sum_{k=0}^{\infty} \left[\sum_{j=1}^p a_j \tilde{\lambda}_j^k \right] z^k = \sum_{k=0}^{\infty} \left[a_1 \tilde{\lambda}_1^k + a_2 \tilde{\lambda}_2^k \right] z^k \\ &= \sum_{k=0}^{\infty} \left[\frac{\tilde{\lambda}_1^{k+1}}{\tilde{\lambda}_1 - \tilde{\lambda}_2} + \frac{\tilde{\lambda}_2^{k+1}}{\tilde{\lambda}_2 - \tilde{\lambda}_1} \right] z^k \\ &= \sum_{k=0}^{\infty} \left[\frac{\tilde{\lambda}_2^{k+1} - \tilde{\lambda}_1^{k+1}}{\tilde{\lambda}_2 - \tilde{\lambda}_1} \right] z^k \end{aligned}$$

On obtient donc finalement :

$$\tilde{\Phi}(z) = \sum_{j=0}^{\infty} \tilde{\phi}_j z^j = \sum_{j=0}^{\infty} \left(\frac{\tilde{\lambda}_2^{j+1} - \tilde{\lambda}_1^{j+1}}{\tilde{\lambda}_2 - \tilde{\lambda}_1} \right) z^j \quad (5.128)$$

On peut démontrer que les paramètres $\tilde{\phi}_j$ ainsi définis satisfont l'équation de récurrence définie en (5.123).