

Théorème de Wold - stationnarité optimalité des espérances conditionnelles - autocorrélations

G. Colletaz

30 septembre 2019

Résumé

On rappelle, sans démonstration, le théorème de Wold qui constitue le socle de la plupart des traitements de séries univariées que nous verrons dans ce cours. Par la suite nous discuterons des propriétés de prévisions construites au moyen de l'espérance conditionnelle d'un processus stationnaire d'ordre 2 sur la base de son écriture de Wold. On rappelle en particulier leur optimalité au sens du critère MSE dans la classe des prédicteurs linéaires. Enfin nous présentons les outils permettant de décrire les dépendances linéaires entre les observations d'un processus : fonction d'autocovariance, fonction d'autocorrélation, fonction d'autocorrélation partielle. Nous verrons dans la suite du cours que ces outils jouent un rôle important dans la méthodologie de Box-Jenkins, tout particulièrement lors de l'étape d'identification des processus

Table des matières

1	Théorème de Wold, stationnarité et ergodicité	1
1.1	Rappel du cadre d'analyse - Le théorème de Wold	1
1.2	Stationnarité et écriture de Wold	2
1.3	l'ergodicité	3
2	Les espérances conditionnelles comme prédicteurs des valeurs futures	4
2.1	Ce sont des estimateurs non biaisés des valeurs futures	4
2.2	Ce sont des estimateurs linéaires optimaux selon le critère MSE	4
3	Autocovariances, autocorrélations et autocorrélations partielles	5
3.1	La fonction d'autocovariance	5
3.2	La fonction d'autocorrélation	5
3.3	La fonction d'autocorrélation partielle	6
3.3.1	Son intérêt	6
3.3.2	Les équations de Yule-Walker	6

1 Théorème de Wold, stationnarité et ergodicité

1.1 Rappel du cadre d'analyse - Le théorème de Wold

Un processus x est stationnaire du second ordre, ou encore stationnaire au sens large, si ses deux premiers moments sont indépendants du temps, soit :

$$\forall t, E(x_t) = \mu_x, \text{ où } \mu_x \text{ est une constante,} \tag{1}$$

$$\forall t, k, cov(x_t, x_{t-k}) = \gamma_k \text{ où } \gamma_k \text{ est une constante} \tag{2}$$

Si x est un processus stationnaire du second ordre alors le théorème de Wold assure que :

$$x_t = \mu + \sum_{i=0}^{\infty} \psi_i u_{t-i} \tag{3}$$

où

1. les coefficients ψ_i sont des constantes avec $\psi_0 = 1$
2. u est un processus bruit blanc vérifiant

$$\forall t, E(u_t) = 0, \tag{4}$$

$$E(u_t u_s) = \begin{cases} \text{une constante } \sigma_u^2 & \text{si } t = s, \\ 0 & \text{si } t \neq s. \end{cases} \tag{5}$$

En d'autres termes, les u_t sont des aléatoires centrées, homoscédastiques et orthogonales entre elles. Si on pose, comme se sera le cas dans ce cours lorsqu'il sera nécessaire de préciser une distribution, que les aléatoires u sont des gaussiennes, alors l'orthogonalité équivaut à leur indépendance. Pour les autres distributions, l'orthogonalité signifie seulement l'absence de relation linéaire entre les aléatoires.

3. μ est un terme déterministe, *i.e.* ses valeurs futures peuvent être anticipées sans erreur de prévision. Des exemples de termes déterministes sont :
 - . $\mu = c$, où c est une constante,
 - . $\mu = a + bt$, un trend linéaire en t
 - . $\mu = a + bt + ct^2$, un trend quadratique en t

Ici, μ est obligatoirement une constante puisque si les aléatoires u sont centrées, alors via (3), $E(x_t) = \mu$ et l'on sait qu'avec la stationnarité, l'espérance de x est une constante.

L'équation (3) est ce que l'on appelle un processus linéaire causal : x_t est fonction linéaire des u contemporain et passés.

1.2 Stationnarité et écriture de Wold

Dans la section précédente, on dit que si x_t est stationnaire, alors il peut s'écrire selon l'équation (3). Nous nous intéressons ici à la proposition réciproque : si x_t est donné par une équation telle que (3), alors est-il stationnaire ?

La réponse est négative : il faut introduire des contraintes sur les coefficients ψ_i . En effet, soit l'aléatoire centrée $y_t = x_t - \mu_X$, à l'évidence, si x_t est stationnaire, alors y_t l'est aussi et réciproquement. Maintenant, y_t et ses autocovariances sont donc donnés par les expressions suivantes :

$$y_t = u_t + \psi_1 u_{t-1} + \psi_2 u_{t-2} + \psi_3 u_{t-3} + \dots \text{ et,}$$

$$\gamma_0 = \text{var}(y_t) = E(y_t y_t) = \sigma_u^2 (\psi_0^2 + \psi_1^2 + \psi_2^2 + \psi_3^2 + \dots) = \sigma_u^2 \sum_{i=0}^{\infty} \psi_i^2$$

$$\gamma_1 = \text{cov}(y_t y_{t-1}) = E(y_t y_{t-1}) = \sigma_u^2 (\psi_0 \psi_1 + \psi_1 \psi_2 + \psi_2 \psi_3 + \dots) = \sigma_u^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+1}$$

$$\gamma_2 = \text{cov}(y_t y_{t-2}) = E(y_t y_{t-2}) = \sigma_u^2 (\psi_0 \psi_2 + \psi_1 \psi_3 + \psi_2 \psi_4 + \dots) = \sigma_u^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+2}$$

⋮

$$\gamma_k = \text{cov}(y_t y_{t-k}) = E(y_t y_{t-k}) = \sigma_u^2 (\psi_0 \psi_k + \psi_1 \psi_{k+1} + \psi_2 \psi_{k+2} + \dots) = \sigma_u^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}, \text{ où } k \text{ est un entier positif quelconque}$$

⋮

Les coefficients ψ_i étant des constantes toutes ces autocovariances sont obtenues par multiplications et additions de constantes. Si elles existent ce seront des constantes et les moments d'ordre 1 et 2 de y seront indépendants du temps : y sera stationnaire. Il faut seulement assurer l'existence de ces moments et cette condition d'existence devient en même temps une condition de stationnarité.

Pour cela, nous imposerons une condition suffisante : les coefficients ψ_i doivent être absolument sommables, *i.e.* :

$$\sum_{i=0}^{\infty} |\psi_i| < \infty \tag{6}$$

Sous cette condition on montre aisément que les sommes infinies précédentes convergent vers un réel $< \infty$:

$$\begin{aligned} \sum_{i=0}^{\infty} |\psi_i| < \infty &\Rightarrow \left(\sum_{i=0}^{\infty} |\psi_i| \right)^2 < \infty \text{ or} \\ \left(\sum_{i=0}^{\infty} |\psi_i| \right)^2 &= \sum_{i=0}^{\infty} |\psi_i|^2 + 2 \sum_{i=0}^{\infty} |\psi_i| |\psi_{i+1}| + 2 \sum_{i=0}^{\infty} |\psi_i| |\psi_{i+2}| + \dots + 2 \sum_{i=0}^{\infty} |\psi_i| |\psi_{i+k}| + \dots, \text{ et donc :} \\ \left(\sum_{i=0}^{\infty} |\psi_i| \right)^2 < \infty &\Rightarrow \sum_{i=0}^{\infty} |\psi_i|^2 < \infty \text{ et } \sum_{i=0}^{\infty} |\psi_i| |\psi_{i+1}| < \infty \text{ et } \sum_{i=0}^{\infty} |\psi_i| |\psi_{i+2}| < \infty \text{ et } \dots \text{ et } \sum_{i=0}^{\infty} |\psi_i| |\psi_{i+k}| < \infty \dots \end{aligned}$$

Enfin, comme σ_u^2 est elle même une constante,

$$\sum_{i=0}^{\infty} \psi_i^2 \leq \sum_{i=0}^{\infty} |\psi_i|^2 < \infty \Rightarrow \sigma_u^2 \sum_{i=0}^{\infty} \psi_i^2 = \gamma_0 < \infty \quad (7)$$

$$\sum_{i=0}^{\infty} \psi_i \psi_{i+1} \leq \sum_{i=0}^{\infty} |\psi_i| |\psi_{i+1}| < \infty \Rightarrow \sigma_u^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+1} = \gamma_1 < \infty \quad (8)$$

$$\sum_{i=0}^{\infty} \psi_i \psi_{i+2} \leq \sum_{i=0}^{\infty} |\psi_i| |\psi_{i+2}| < \infty \Rightarrow \sigma_u^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+2} = \gamma_2 < \infty \quad (9)$$

$$\vdots \quad (10)$$

$$\sum_{i=0}^{\infty} \psi_i \psi_{i+k} \leq \sum_{i=0}^{\infty} |\psi_i| |\psi_{i+k}| < \infty \Rightarrow \sigma_u^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k} = \gamma_k < \infty \quad (11)$$

$$\vdots \quad (12)$$

Ainsi, sous la condition énoncée, toutes les autocovariances de y existent et sont des constantes. Comme son espérance est nulle, y est bien stationnaire au sens large et il en va de même pour le processus non centré x .

1.3 l'ergodicité

Il s'agit d'une propriété mathématique qui, pour ce qui nous intéresse ici, nous permet d'appliquer la loi des grands nombres sur des observations corrélées entre elles. En pratique, cette hypothèse assure que les moments empiriques convergent vers les moments théoriques lorsque le nombre d'observation augmente, et donc que l'on peut construire des estimateurs ayant au moins asymptotiquement cette propriété de convergence au sens où pour une série X stationnaire et ergodique on vérifiera :

$$plim \bar{x} = E[X], \text{ avec } \bar{x} = \frac{1}{T} \sum_t x_t,$$

$$plim \frac{1}{T} \sum_t (x_t - \bar{x})^2 = E([X - E[X])^2] = var[X] = \gamma_0$$

$$plim \frac{1}{T} \sum_t (x_t - \bar{x})(x_{t-1} - \bar{x})^2 = E([X_t - E[X])([X_{t-1} - E[X]]) = cov[X_t, X_{t-1}] = \gamma_1$$

\vdots

$$plim \frac{1}{T} \sum_t (x_t - \bar{x})(x_{t-k} - \bar{x})^2 = E([X_t - E[X])([X_{t-k} - E[X]]) = cov[X_t, X_{t-k}] = \gamma_k$$

Pour qu'elle soit vérifiée, il faut que la corrélation entre deux observations distantes dans le temps tende à s'annuler avec l'augmentation de cette distance. Intuitivement, cela implique que l'information contenue dans ces observations n'est pas la même : chaque ajout d'observation apporte alors de l'information pertinente non entièrement contenue dans les observations précédentes sur les moments théoriques.

2 Les espérances conditionnelles comme prédicteurs des valeurs futures

Soit ${}_t x_{t+h}$ la prévision faite en t pour un horizon de h périodes. On pose :

$${}_t x_{t+h} = E(x_{t+h}|I_t) \quad (13)$$

i.e. cette prévision est égale à l'espérance de ce que l'on cherche à prévoir, espérance conditionnée par l'ensemble des informations disponibles à la date où on fait cette prévision, I_t . L'équation précédente s'écrira encore :

$${}_t x_{t+h} = E_t(x_{t+h}) \quad (14)$$

Dans ce qui suit, x est un processus stationnaire du second ordre.

2.1 Ce sont des estimateurs non biaisés des valeurs futures

On a :

$$\begin{aligned} x_{t+h} &= \mu_x + u_{t+h} + \psi_1 u_{t+h-1} + \psi_2 u_{t+h-2} + \dots + \psi_{h-1} u_{t+1} \\ &\quad + \psi_h u_t + \psi_{h+1} u_{t-1} + \psi_{h+2} u_{t-2} + \dots \end{aligned} \quad (15)$$

et donc,

$$\begin{aligned} {}_t x_{t+h} &\equiv E_t(x_{t+h}) \\ &= \mu_x + \psi_h u_t + \psi_{h+1} u_{t-1} + \psi_{h+2} u_{t-2} + \dots \end{aligned} \quad (16)$$

En conséquence, l'erreur associée à cette prévision est

$$\begin{aligned} {}_t e_{t+h} &\equiv x_{t+h} - E_t(x_{t+h}) \\ &= u_{t+h} + \psi_1 u_{t+h-1} + \psi_2 u_{t+h-2} + \dots + \psi_{h-1} u_{t+1} \end{aligned} \quad (17)$$

Très logiquement elle ne dépend que des u qui se réaliseront postérieurement à la date de construction de la prévisions, t . Avec ce mécanisme de construction de prévision, il vient :

$$E({}_t e_{t+h}) = 0 : \text{ en moyenne on ne se trompe pas, et,} \quad (18)$$

$$E_t({}_t e_{t+h}) = 0 : \text{ à la date } t, \text{ lorsque la prévision est construite, on espère ne pas se tromper en moyenne.} \quad (19)$$

Pour un horizon d'une période on peut remarquer que ${}_t e_{t+1} = x_{t+1} - E_t(x_{t+1}) = u_{t+1}$: l'erreur de prévision se confond avec le processus causal lui-même et donc $x_{t+1} = E_t(x_{t+1}) + u_{t+1}$. Comme $E_t(x_{t+1})$ est par construction connu à la date t , cette dernière égalité signifie que la valeur de x en $t+1$ est décomposée en deux éléments, l'un qui ne dépend que d'informations passées, et l'autre, u_{t+1} , qui est imprévisible en t , au sens où $E_t[u_{t+1}] = E[u_{t+1}] = 0$. Ainsi u_{t+1} est l'innovation où l'effet de surprise qui affecte x en $t+1$ et u est d'ailleurs également souvent nommé processus d'innovation.

2.2 Ce sont des estimateurs linéaires optimaux selon le critère MSE

Pour le montrer on considère un autre prédicteur linéaire, fondé sur le même ensemble d'informations que l'espérance conditionnelle précédente. Soit :

$${}_t \hat{x}_{t+h} = \mu_x + a_h u_t + a_{h+1} u_{t-1} + a_{h+2} u_{t-2} + \dots \quad (20)$$

Ainsi ${}_t x_{t+h}$ le prédicteur construit sur l'espérance conditionnelle donné par (16), et cet autre prédicteur ${}_t \hat{x}_{t+h}$ partagent bien le même ensemble d'informations en t , à savoir les valeurs présente et passées du processus causal u et μ_x , l'espérance de x_t . Évidemment pour qu'ils diffèrent, on suppose que les coefficients a sont différents des ψ .

On regarde la MSE de ce second prédicteur :

$$MSE({}_t\hat{x}_{t+h}) \equiv E [(x_{t+h} - {}_t\hat{x}_{t+h})^2] \quad (21)$$

$$\begin{aligned} &= E[(\mu_x + u_{t+h} + \psi_1 u_{t+h-1} + \psi_2 u_{t+h-2} + \dots + \psi_{h-1} u_{t+1} \\ &+ \psi_h u_t + \psi_{h+1} u_{t-1} + \psi_{h+2} u_{t-2} + \dots \\ &- \mu_x - a_h t - a_{h+1} u_{t-1} - a_{h+2} u_{t-2} - \dots)^2] \end{aligned} \quad (22)$$

$$= E[\underbrace{(u_{t+h} + \psi_1 u_{t+h-1} + \psi_2 u_{t+h-2} + \dots + \psi_{h-1} u_{t+1})}_a] \quad (23)$$

$$+ \underbrace{(\psi_h - a_h)u_t + (\psi_{h+1} - a_{h+1})u_{t-1} + (\psi_{h+2} - a_{h+2})u_{t-2} + \dots}_b^2] \quad (24)$$

Le membre de droite de cette dernière égalité est la somme de deux termes. Le premier (a) est constitué de h termes qui chacun font référence à des u de temps différents, $u_{t+1}, u_{t+2}, \dots, u_{t+h}$, ils sont donc orthogonaux entre eux. Si on développe le carré, ce premier terme fera ainsi apparaître leur carré et leur produits croisés. Avec le passage à l'espérance, ces produits croisés vont donner des covariances nulles car u est un bruit blanc. Au final, le développement du carré et la prise de l'espérance font que ce premier terme est égal à :

$$\sigma_u^2(1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{h-1}^2)$$

Par ailleurs un raisonnement identique mené sur le terme (b) montre qu'il conduit après élévation au carré et prise de l'espérance à :

$$\sigma_u^2((\psi_h - a_h)^2 + (\psi_{h+1} - a_{h+1})^2 + (\psi_{h+2} - a_{h+2})^2 + \dots)$$

Enfin, comme (a) et (b) sont constitués d'ensembles de u disjoints (postérieurs à t pour (a), contemporain et antérieurs à t pour (b)), l'espérance du double produit (a)(b) sera nulle. La MSE de ${}_t\hat{x}_{t+h}$ est donc la somme des deux dernières expressions :

$$MSE({}_t\hat{x}_{t+h}) = \sigma_u^2 \sum_{i=0}^{h-1} \psi_i^2 + \sigma_u^2 \sum_{i=0}^{\infty} (\psi_{h+i} - a_{h+i})^2 \quad (25)$$

Si on cherche les valeurs des coefficients a qui conduisent au prédicteur optimal au sens du critère MSE, il faut donc les choisir de sorte à minimiser cette dernière expression. Celle-ci est constituée de deux termes, le premier est positif et ne dépend pas des coefficients a . Le second est positif ou nul, la nullité étant obtenue en posant $a_{h+i} = \psi_{h+i}$. C'est évidemment avec cette condition que l'on minimise la MSE. Dans ce cas, l'équation (20) qui définit alors le ${}_t\hat{x}_{t+h}$ optimal devient :

$${}_t\hat{x}_{t+h} = \mu_x + \psi_h u_t + \psi_{h+1} u_{t-1} + \psi_{h+2} u_{t-2} + \dots \quad (26)$$

On s'aperçoit alors que cette dernière équation est identique à (16) qui définissait le prédicteur fondé sur l'espérance conditionnelle, ${}_t x_{t+h}$ qui est donc bien optimal au sens du critère MSE.

3 Autocovariances, autocorrélations et autocorrélations partielles

Il s'agit de décrire les dépendances qui existent entre les réalisations de temps différents d'une variable aléatoire. L'intérêt est immédiatement perçu si on se pose un problème de prévision : si on connaît la structure de dépendance entre un état actuel et les états passés alors, pour un processus stationnaire, cette même structure doit pouvoir être utile pour anticiper l'état futur en fonctions des états présent et passés qui eux sont connus.

3.1 La fonction d'autocovariance

Vous savez que la covariance est une mesure de dépendance linéaire entre deux aléatoires. Ainsi, si on se limite aux relations linéaires et que l'on veuille mesurer ces dépendances entre l'état présent, x_t et les états passés, x_{t-1}, x_{t-2}, \dots on s'intéressera à la fonction d'autocovariance, dont les éléments sont $\gamma_k = cov(x_t, x_{t-k}), k = \pm 0, 1, 2, 3 \dots$. Cette fonction étant symétrique, il suffira de l'étudier sur une seule moitié, $\gamma_k = cov(x_t, x_{t-k}), k = 0, 1, 2, 3 \dots$. Notez que $\gamma_0 = var(x_t)$.

3.2 La fonction d'autocorrélation

Elle se déduit de la précédente, son avantage étant évidemment que ses coefficients sont bornés par 0 et 1, sachant que la valeur absolue d'un coefficient reflète l'intensité de la dépendance avec une orthogonalité parfaite lorsqu'il est nul. Ses coefficients se définissent par $\rho_k = \frac{\gamma_k}{\gamma_0}$. Comme la précédente elle est symétrique au sens ou $\forall k, \rho_k = \rho_{-k}$.

3.3 La fonction d'autocorrélation partielle

3.3.1 Son intérêt

Lorsque l'on veut décrire la dépendance linéaire de x au temps t à son passé, il paraît raisonnable de regarder les autocorrélations $\rho_j = \text{cor}(x_t, x_{t-j}), j = 1, 2, \dots$. Pour autant les informations fournies par cette fonction d'autocorrélation peuvent être ambiguës comme le montre cet exemple simple : supposons que ρ_2 soit non nul laissant ainsi penser que peut exister une relation linéaire entre x_t et x_{t-2} , mais supposons que dans le même temps ρ_1 soit également non nul. La question est alors de savoir s'il existe effectivement une dépendance directe de x_t à x_{t-2} , ou si cette dépendance résulte d'une combinaison de dépendances à l'ordre 1 : x_{t-2} impactant x_{t-1} puis x_{t-1} impactant x_t . Vous devez comprendre que la dépendance directe à l'ordre 2 existe vraiment si, même en tenant compte de la dépendance à l'ordre 1 on continue d'avoir une relation linéaire entre x_t à x_{t-2} . L'équation à considérer pour éclaircir les choses est évidemment :

$$x_t = \phi_{2,1}x_{t-1} + \phi_{2,2}x_{t-2} + u_t$$

ou u_t est un processus d'innovation pour x_t et donc orthogonal à x_{t-1} et x_{t-2} . Dans cette équation $\phi_{2,2}$ mesure bien la dépendance linéaire de x_t à x_{t-2} compte-tenu de celle existant entre x_t et x_{t-1} mesurée par $\phi_{2,1}$. Ce coefficient $\phi_{2,2}$ est la corrélation partielle à l'ordre 2 afférent au processus x . Vous devez d'ailleurs être en mesure de montrer que la corrélation ρ_2 est quant à elle égale au coefficient ϕ_2 dans l'équation

$$x_t = \phi_2x_{t-2} + u_t$$

Cet exemple se généralise à un ordre k quelconque : pour mesurer la dépendance directe entre x_t et x_{t-k} , il faut tenir compte des dépendances intermédiaires qui pourraient l'expliquer, *i.e.* de celles pouvant exister entre x_t et x_{t-1} , x_t et x_{t-2} , ..., x_t et x_{t-k-1} . Pour cela, on va considérer l'équation :

$$x_t = \phi_{k,1}x_{t-1} + \phi_{k,2}x_{t-2} + \dots + \phi_{k,k-1}x_{t-k-1} + \phi_{k,k}x_{t-k} + u_t,$$

la corrélation partielle d'ordre k sera alors donnée par $\phi_{k,k}$.

3.3.2 Les équations de Yule-Walker

Ces équations de Yule-Walker sont simplement des équations de passage des autocorrélations ρ_j vers les coefficients de l'écriture autorégressive du processus et donc en particulier, pour une longueur p donnée de cette écriture, vers son dernier coefficient, *i.e.* son autocorrélation partielle d'ordre p . On va tout d'abord préciser la fonction d'autocovariance d'un processus ayant une écriture autorégressive, pour en tirer sa fonction d'autocorrélation puis les coefficients ϕ_i de l'écriture autorégressive, ce qui explicitera les équations de Yule-Walker permettant alors le calcul des autocorrélations partielles.

On considère donc l'écriture autorégressive d'un processus X_t stationnaire d'ordre p quelconque avec son processus d'innovation u_t :

$$\begin{aligned} X_t &= c + \phi_1X_{t-1} + \phi_2X_{t-2} + \dots + \phi_pX_{t-p} + u_t, \text{ soit encore,} \\ &= c + \sum_{i=1}^{\infty} \phi_iX_{t-i} + u_t, \text{ avec } \phi_i = 0 \text{ si } i > p. \end{aligned}$$

Si on note μ_X l'espérance de X , alors en prenant l'espérance de X_t dans l'équation ci-dessus, il vient :

$$\begin{aligned} \mu_x &= c + \phi_1\mu_x + \phi_2\mu_x + \dots + \phi_p\mu_x + u_t, \text{ soit encore} \\ \mu_x &= \frac{c}{1 - \phi_1 - \phi_2 - \dots - \phi_p} \end{aligned}$$

Sur le processus centré, $x_t = X_t - \mu_x$ on a une écriture équivalente :

$$\begin{aligned} x_t &= \phi_1x_{t-1} + \phi_2x_{t-2} + \dots + \phi_px_{t-p} + u_t, \text{ soit encore,} \\ &= \sum_{i=1}^{\infty} \phi_ix_{t-i} + u_t, \text{ avec } \phi_i = 0 \text{ si } i > p. \end{aligned} \tag{27}$$

On calcule alors les autocovariances, avec toujours $\phi_i = 0$ si $i > p$:

$$\begin{aligned}\gamma_0 &= E[x_t x_t] = \sum_{i=1}^{\infty} \phi_i E[x_t x_{t-i}] + E[x_t u_t] = \sum_{i=1}^{\infty} \phi_i \gamma_i + E[x_t u_t], \\ \gamma_1 &= E[x_t x_{t-1}] = \sum_{i=1}^{\infty} \phi_i E[x_{t-1} x_{t-i}] + E[x_{t-1} u_t] = \sum_{i=1}^{\infty} \phi_i \gamma_{i-1} + E[x_{t-1} u_t], \\ &\vdots \\ \gamma_k &= E[x_t x_{t-k}] = \sum_{i=1}^{\infty} \phi_i E[x_{t-k} x_{t-i}] + E[x_{t-k} u_t] = \sum_{i=1}^{\infty} \phi_i \gamma_{i-k} + E[x_{t-k} u_t],\end{aligned}$$

Les derniers termes des membres de droite de ces équations, à l'exception de la première d'entre elles, sont nuls. En effet, u_t est orthogonal à tous les u qui lui sont antérieurs et donc à toute combinaison linéaire de ces u passés, et donc à x_{t-k} si $k \geq 1$. Maintenant, pour $k = 0$:

$$\begin{aligned}E[x_t u_t] &= E[(\phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + u_t) u_t] \\ &= E[(\phi_1 x_{t-1} u_t + \phi_2 x_{t-2} u_t + \dots + \phi_p x_{t-p} u_t + u_t^2)] \\ &= \phi_1 E[x_{t-1} u_t] + \phi_2 E[x_{t-2} u_t] + \dots + \phi_p E[x_{t-p} u_t] + E[u_t^2] \\ &= 0 + 0 + \dots + 0 + \sigma_u^2, \text{ la nullité étant obtenue pour les raisons qui viennent d'être énoncées} \\ &= \sigma_u^2\end{aligned}$$

Au final on a donc deux types d'équations, l'une pour γ_0 , l'autre pour toutes les autocovariances d'ordre supérieur :

$$\begin{aligned}\gamma_0 &= E[x_t x_t] = \sum_{i=1}^{\infty} \phi_i \gamma_i + \sigma_u^2, \\ \gamma_k &= E[x_t x_{t-k}] = \sum_{i=1}^{\infty} \phi_i \gamma_{k-i} \text{ pour } k \geq 1.\end{aligned}$$

Ci-dessus on a utilisé le fait que la fonction d'autocovariance est symétrique : $cov(x_{t-k}, x_{t-i}) = \gamma_{i-k} = cov(x_{t-i}, x_{t-k}) = \gamma_{k-i}$. On rappelle également que dans ces écritures $\phi_j = 0$ si $j > p$. En termes de corrélations, si on divise la seconde équation par γ_0 , il vient :

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \sum_{i=1}^{\infty} \phi_i \rho_{k-i} \text{ pour } k \geq 1, \quad (28)$$

avec naturellement $\rho_0 = 1$ et toujours $\phi_i = 0$ si $i > p$. On peut noter que lorsque le processus x a une écriture autorégressive alors x_t , sa fonction d'autocovariance et sa fonction d'autocorrélation obéissent à la même équation de récurrence. Par exemple, pour une autorégression de longueur p :

$$\begin{aligned}x_t &= \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + u_t \\ \gamma_k &= \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p} \\ \rho_k &= \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p}\end{aligned}$$

Supposons maintenant que l'on connaisse les autocorrélations et que l'on veuille en déduire les valeurs des coefficients d'une écriture autorégressive de longueur p , *i.e.* les $\phi_1, \phi_2, \dots, \phi_p$. Il nous faut alors considérer p équations du type de la dernière ci-dessus, c'est un tel ensemble que décrit le système matriciel suivant :

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}$$

D'où l'on déduit le résultat recherché :

$$\begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\ \vdots & & & & \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix}$$

Ce dernier système constitue les équations de Yule-Walker.

Enfin, si l'on veut maintenant retrouver, à partir des autocorrélations ρ_i la fonction d'autocorrélation partielle, soit les coefficients $\phi_{p,p}$ égaux à ϕ_p pour $p = 1, 2, 3, \dots$, il suffit de résoudre ce système pour les différentes valeurs de p . Ainsi :

— pour $p = 1$, le système précédent se réduit à

$$[\phi_{11}] = [1]^{-1} [\rho_1] \Rightarrow \text{la corrélation partielle au rang 1 est égale à la corrélation de même rang.}$$

L'intuition de ce dernier résultat est simplement que la corrélation partielle d'ordre p mesure la dépendance linéaire entre x_t et x_{t-p} en tenant compte des dépendances intermédiaires entre x_t et chacun des $x_{t-1}, \dots, x_{t-p-1}$ contrairement à la corrélation ρ_p qui n'en tient pas compte. Dans ces conditions, s'il n'y a pas de x intermédiaire, les deux mesures vont coïncider et c'est évidemment le cas entre les instants t et $t-1$: sans surprise on aura toujours $\rho_1 = \phi_{11}$.

— pour $p = 2$ il vient

$$\begin{bmatrix} \phi_{21} \\ \phi_{22} \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix}$$

Une fois résolu ce système, la seconde équation va donner la valeur de l'autocorrélation partielle d'ordre 2, qui évidemment n'a plus de raison d'être égale à ρ_2 . Il en ira évidemment de même pour les autocorrélations partielles d'ordre supérieur $\phi_{33}, \phi_{44}, \dots$: généralement $\rho_p \neq \phi_{pp}$ pour $p > 1$.