

# Modélisation de séries stationnaires

MAP-STA2 : Séries chronologiques

Y. Goude - yannig.goude@edf.fr

2023-2024

## Contents

<b>Rappels et généralités</b>	<b>1</b>
Notion de stationnarité, auto-covariance, auto-corrélation . . . . .	4
<b>Corrélation et auto-corrélation partielle</b>	<b>7</b>

Nous avons vu précédemment comment, si une série chronologique présente des composantes déterministes (tendance ou saisonnalité(s)), les estimer de manière à pouvoir effectuer une prévision. Nous nous intéressons ici à la série corrigée de sa tendance et de sa ou ses saisonnalité(s) et à la notion de stationnarité.

## Rappels et généralités

Ce chapitre fait appel à différentes notions de probabilité et statistiques que nous rappelons ici.

**processus stochastique** un processus stochastique est une famille de variables aléatoires  $(Y_t)_{t \in I}$  définies sur  $(\Omega, A, P)$ . Les applications  $t \rightarrow Y_t(\omega)$ ,  $\omega \in \Omega$  sont appelés trajectoires du processus.

**processus Gaussien** Le processus  $(Y_t)_{t \in \mathbb{Z}}$  est un processus gaussien si toutes ses lois marginales sont gaussiennes ie si  $\forall k$  et  $\forall j_1, \dots, j_k$ ,  $(Y_{j_1}, Y_{j_2}, \dots, Y_{j_k})$  est un vecteur gaussien.

**bruit blanc fort** soit  $\varepsilon_t$  un processus aléatoire. On dira qu'il s'agit d'un bruit blanc fort si les  $\varepsilon_t$  sont indépendants et identiquement distribués. Il est dit centré si  $E(\varepsilon_t) = 0$  et réduit si  $\text{var}(\varepsilon_t) = 1$

**bruit blanc faible** soit  $\varepsilon_t$  un processus aléatoire. On dira qu'il s'agit d'un bruit blanc faible si les  $\varepsilon_t$  vérifient:

- $E(\varepsilon_t) = \mu$
- $E(\varepsilon_t^2) = \sigma^2$
- $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$  pour  $i \neq j$

**espace  $\mathbf{L}^2$**  L'ensemble des variables aléatoires  $X$  telles que  $E(X^2) < \infty$  peut être muni d'une structure d'espace vectoriel normé noté  $\mathbf{L}^2$ . On prend comme produit scalaire l'application  $(X, Y) \rightarrow E(XY)$  et comme norme  $\|X\|_{\mathbf{L}^2} = \sqrt{E(X^2)}$

**covariance** la covariance entre deux variables aléatoires  $X$  et  $Y$  est définie ainsi:

$$\text{cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$$

**produit scalaire** soit  $X$  et  $Y$  deux variables aléatoires,  $(X, Y) \rightarrow E(XY)$  est un produit scalaire vérifiant les propriétés suivantes:

- symétrie:  $E(XY) = E(YX)$
- bilinéarité:  $E((aX + bY)Z) = aE(XZ) + bE(YZ)$
- positivité:  $E(X^2) \geq 0$ , caractère défini  $E(X^2) = 0 \rightarrow P(X = 0) = 1$

**orthogonalité** soit  $X$  et  $Y$  deux variables aléatoires, elles sont orthogonales lorsque  $E(XY) = 0$  (si les variables sont centrées cela revient à  $\text{cov}(X,Y)=0$ )

**distance** on définit la distance entre  $X$  et  $Y$  deux variables aléatoires par  $d(X, Y) = \sqrt{E[(X - Y)^2]}$

**espace engendré** l'espace engendré par les variables aléatoires  $X_1, X_2, \dots, X_k$ , noté  $M(X_1, \dots, X_k)$  est l'ensemble des combinaisons linéaires de ces variables:

$$M(X_1, \dots, X_k) = \{\lambda_1 X_1 + \dots + \lambda_k X_k, \lambda \in \mathbf{R}^k\} = \{X\lambda, \lambda \in \mathbf{R}^k\}$$

en notant  $X = (X_1, \dots, X_k)$ ,  $\lambda = (\lambda_1, \dots, \lambda_k)$

**projection**  $P_{M(X_1, \dots, X_k)}(Y) = X\alpha$  la projection linéaire d'une variable  $Y$  sur  $M(X_1, \dots, X_k)$  est telle que

$$E[(Y - X\alpha)^2] \leq E[(Y - X\beta)^2], \forall \beta$$

c'est le vecteur de  $M(X_1, \dots, X_k)$  le plus proche de  $Y$  au sens de la distance  $d$  définie ci-dessus.

**propriété de la projection linéaire** la projection linéaire est orthogonale au plan  $M(X_1, \dots, X_k)$ .

Ainsi:

$$\forall j \in (1, \dots, k) : E[(Y - P_{M(X_1, \dots, X_k)}(Y))X_j] = 0$$

et donc  $\forall j \in (1, \dots, k)$ :

$$\alpha_1 E(X_1 X_j) + \dots + \alpha_k E(X_k X_j) = E(Y X_j)$$

en notant:

$$\Sigma_X = \begin{pmatrix} E(X_1^2) & \dots & E(X_1 X_k) \\ \vdots & \ddots & \vdots \\ E(X_1 X_k) & \dots & E(X_k^2) \end{pmatrix}$$

la matrice de covariance de  $X$ , et si cette matrice est inversible (si les  $X_j$  ne sont pas linéairement dépendants) on a

$$\alpha = \Sigma_X^{-1} \begin{pmatrix} E(Y X_1) \\ \vdots \\ E(Y X_k) \end{pmatrix}$$

## Pythagore

$$\|Y\|_{\mathbf{L}^2}^2 = \|P_{M(X_1, \dots, X_k)}\|_{\mathbf{L}^2}^2 + \|Y - P_{M(X_1, \dots, X_k)}\|_{\mathbf{L}^2}^2$$

**extension de la loi des grands nombres** lorsque l'on s'intéresse à un processus stochastique  $(Y_t)_{t \in \mathbf{Z}}$  on aimerait disposer d'un résultat de type loi des grands nombres sur une trajectoire du processus. Par exemple on aimerait que la moyenne empirique  $\frac{1}{T} \sum_{t=1}^T Y_t$  converge p.s. vers  $E(Y_0)$  lorsque  $T \rightarrow \infty$ . Cela n'est pas toujours vrai, notamment quand la dépendance temporelle entre les observations est trop "forte". Nous admettrons le théorème suivant:

**théorème** soient  $H : (\mathbf{R}^d)^{\mathbf{Z}} \rightarrow \mathbf{R}$  une fonction mesurable,  $(\varepsilon_i)_{i \in \mathbf{Z}}$  une suite i.i.d. de v.a. à valeur dans  $\mathbf{R}^d$ . Posons  $Y_t = H((\varepsilon_i)_{i \in \mathbf{Z}})$ , alors la série  $(Y_t)_{t \in \mathbf{Z}}$  est stationnaire forte (voir plus loin) et, si  $Y_0$  est intégrable:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T Y_t = E(Y_0), p.s.$$

exemple: voir plus loin le cas d'une moyenne mobile infinie.

## Notion de stationnarité, auto-covariance, auto-corrélation

Pour pouvoir espérer prévoir le futur d'une série chronologique  $(Y_t)$ , il est nécessaire que cette série présente une certaine reproductibilité. Cela permet que l'inférence effectuée sur certains paramètres de loi ou de modèle (corrélation, régression linéaire...) soit pérenne dans le temps.

**définition** soit un processus aléatoire  $(Y_t)_{t \in \mathbf{Z}}$ , il est dit stationnaire au sens fort (ou strictement) si pour toute fonction  $f$  mesurable  $f(Y_1, Y_2, \dots, Y_t)$  et  $f(Y_{1+h}, Y_{2+h}, \dots, Y_{t+h})$  ont la même loi.

Cette notion de stationnarité forte est très difficile à vérifier en pratique. On lui préfère généralement la notion de stationnarité faible qui porte sur les moments d'ordre 1 et 2 du processus.

**définition** la fonction d'auto-covariance d'un processus  $Y_{t \in \mathbf{Z}}$

$$\text{cov}(Y_t, Y_{t+h}) = \gamma(h)$$

**définition** la fonction d'auto-corrélation d'un processus  $Y_{t \in \mathbf{Z}}$

$$\rho(h) = \gamma(h)/\gamma(0)$$

*remarque*  $\gamma(h)$  et  $\rho(h)$  sont des fonctions symétriques,  $\rho(0) = 1$ .

**définition** soit un processus aléatoire  $(Y_t)_{t \in \mathbf{Z}}$  tel que  $E(Y_t^2) < \infty$ , il est dit stationnaire au sens faible (ou d'ordre 2) si son espérance est constante et ses auto-covariances sont stables dans le temps ie:

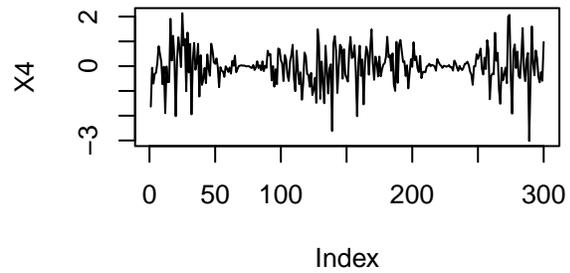
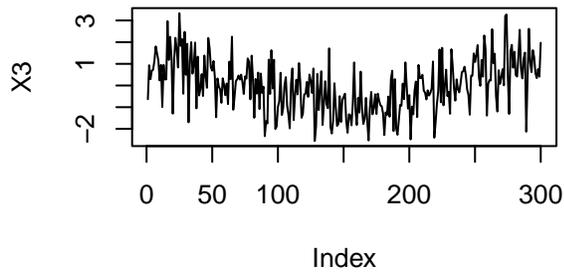
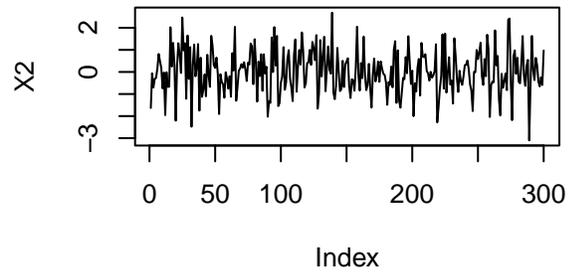
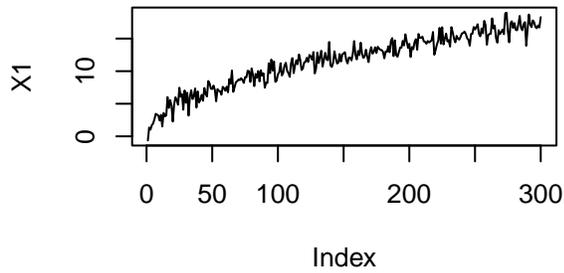
$$\forall t \quad E(Y_t) = \mu$$

$$\forall t, \forall h \quad \text{cov}(Y_t, Y_{t+h}) = \gamma(h)$$

On remarque que  $\text{var}(Y_t) = \gamma(0)$  et donc qu'un processus stationnaire faible à une variance constante dans le temps.

En pratique, pour apprécier la stationnarité d'un processus, on commence d'abord par vérifier que sa moyenne et sa variance sont constantes dans le temps.

*exercice* selon vous quel(s) processus ci-dessous est(sont) stationnaire(s)? Pourquoi?



Voilà quelques exemples de processus stationnaires:

- un bruit blanc  $\varepsilon_t$  vérifiant  $E(\varepsilon_t) = \mu$  et  $\text{var}(\varepsilon_t) = \sigma^2$

*preuve* on a par définition  $\text{cov}(\varepsilon_t, \varepsilon_{t+h}) = 0$

- Le processus gaussien  $(Y_t)_{t \in \mathbf{Z}}$  tel que  $E(Y_t) = \mu$  et  $\text{cov}(Y_t, Y_{t+h}) = \alpha^{|h|}$  ( $|\alpha| < 1$ ) est faiblement stationnaire. Tout processus gaussien stationnaire faible est stationnaire fort.
- le processus moyenne mobile  $X_t = \varepsilon_t + a_1\varepsilon_{t-1} + a_2\varepsilon_{t-2} + \dots + a_q\varepsilon_{t-q}$

*preuve*

$$\begin{aligned} \gamma(0) &= \sigma^2(1 + a_1^2 + \dots + a_q^2) \\ \gamma(1) &= \sigma^2(a_1 + a_1a_2 + \dots + a_{q-1}a_q) \\ &\dots \\ \gamma(q) &= \sigma^2(a_q) \\ \gamma(q+h) &= 0 \end{aligned}$$

- processus autorégressif d'ordre 1:

$$Y_t = aY_{t-1} + \varepsilon_t$$

en supposant que  $|a| < 1$  on a bien  $E(Y_t) = 0$  et

$$\text{var}(Y_t) = \sigma^2(1 + a + a^2 + \dots) = \frac{\sigma^2}{1 - a^2}$$

pour tout  $h > 0$ :

$$\gamma(h) = \sigma^2(a^h + a^{h+2} + \dots) = \frac{\sigma^2 a^h}{1 - a^2}$$

comme de plus  $\gamma(h) = \gamma(-h)$ ,

$$\gamma(h) = \frac{\sigma^2 a^{|h|}}{1 - a^2}$$

on remarque que pour ce processus  $\rho(h) = a^{|h|}$ , donc l'autocorrélation tend vers 0 à une vitesse exponentielle.

- soit  $(X_t)_{t \in \mathbf{Z}}$  un processus stationnaire d'espérance  $\mu$  et de fonction d'autocovariance  $\gamma(h)$ ,  $(a_i)_{i \in \mathbf{Z}}$  une suite de nombres réels absolument sommable  $\sum_{i \in \mathbf{Z}} |a_i| < \infty$ , alors  $Y_t = \sum_{i \in \mathbf{Z}} a_i X_{t-i}$  est un processus stationnaire.

tout d'abord, on remarque que

$$\sum_{i \in \mathbf{Z}} \|a_i X_{t-i}\| = \sum_{i \in \mathbf{Z}} |a_i| \|X_{t-i}\| = \sqrt{\mathbf{Var}(X) + E(X)^2} \sum_{i \in \mathbf{Z}} |a_i| < \infty$$

$\sum_{i \in \mathbf{Z}} a_i X_{t-i}$  est donc bien convergente au sens de  $\mathbf{L}^2$  et  $Y_t$  est de carré intégrable.

on vérifie également que  $E(Y_t)$  ne dépend pas de  $t$ :

$$E\left(\sum_{i \in \mathbf{Z}} a_i X_{t-i}\right) = \sum_{i \in \mathbf{Z}} a_i E(X_{t-i}) = \mu \sum_{i \in \mathbf{Z}} a_i$$

ainsi que sa fonction d'autocovariance:

$$\text{cov}(Y_t, Y_{t+h}) = \text{cov}\left(\sum_{i \in \mathbf{Z}} a_i X_{t-i}, \sum_{j \in \mathbf{Z}} a_j X_{t+h-j}\right)$$

$$\text{cov}(Y_t, Y_{t+h}) = \sum_{i \in \mathbf{Z}} \sum_{j \in \mathbf{Z}} a_i a_j \text{cov}(X_{t-i}, X_{t+h-j})$$

$$\text{cov}(Y_t, Y_{t+h}) = \sum_{i \in \mathbf{Z}} \sum_{j \in \mathbf{Z}} a_i a_j \gamma(h + i - j)$$

En pratique, on ne connaît pas explicitement les fonctions d'auto-covariance et d'auto-corrélation. Il est donc nécessaire de les estimer en se basant sur des observations.

**définition** soit une série d'observations  $(y_t)_{t \in (1, \dots, n)}$ , notons  $\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t$ , alors la fonction d'auto-covariance empirique vaut, pour tout  $h \in (0, \dots, n-1)$

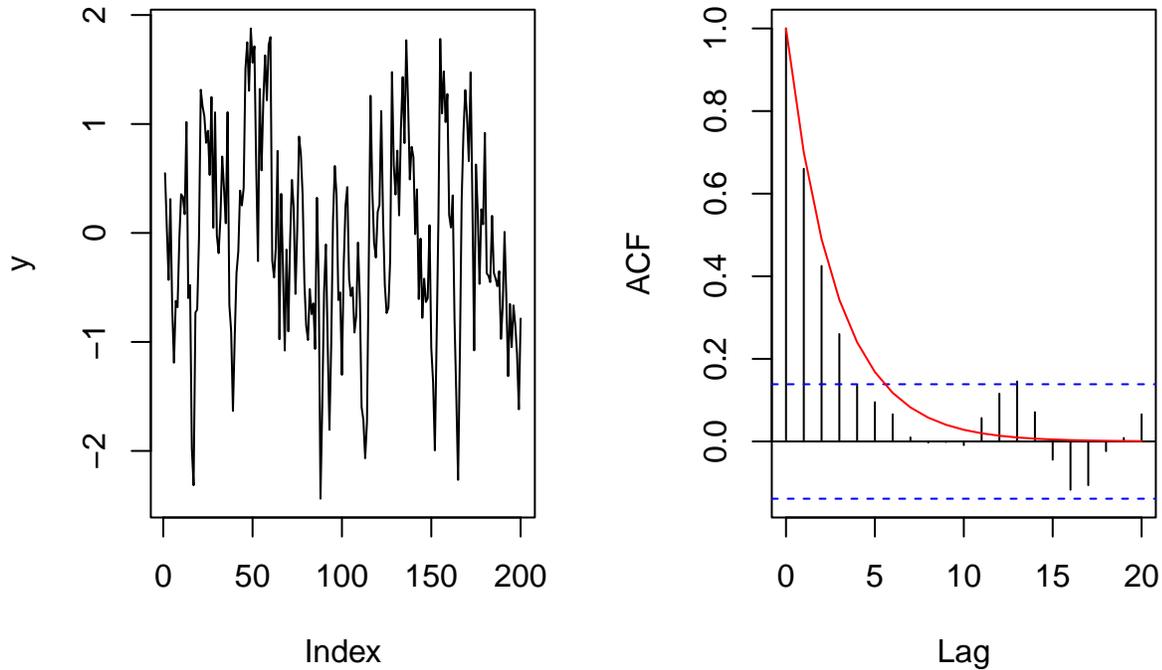
$$\hat{\gamma}(h) = \frac{1}{n-h} \sum_{t=h+1}^n (y_t - \bar{y})(y_{t-h} - \bar{y})$$

**définition** soit une série d'observations  $(y_t)_{t \in (1, \dots, n)}$ , notons  $\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t$ , alors la fonction d'auto-corrélation empirique vaut, pour tout  $h \in (0, \dots, n-1)$

$$\hat{\rho}(h) = \frac{\frac{1}{n-h} \sum_{k=h+1}^n (y_t - \bar{y})(y_{t-h} - \bar{y})}{\frac{1}{n} \sum_{t=1}^n (y_t - \bar{y})^2}$$

Le graphique représentant la fonction d'auto-corrélation empirique est appelé l'auto-corrélogramme.

*exemple* voilà un exemple de série et son auto-corrélogramme, à votre avis de quel type de série s'agit-il?



## Corrélation et auto-corrélation partielle

Lorsque l'on s'intéresse à caractériser les dépendances d'au moins 3 variables aléatoires, il est nécessaire d'introduire la notion de corrélation partielle. En effet, si l'on considère les variables  $X_1, \dots, X_k$ ,  $X_1$  peut être corrélée à  $X_3$  parce que  $X_1$  et  $X_3$  sont toutes deux corrélées à  $X_2$ . Voilà quelques exemples frappants:

**définition** soit les variables aléatoires  $X_1, \dots, X_k$ , le coefficient de corrélation partielle entre  $X_1$  et  $X_k$  abstraction faite de  $X_2, \dots, X_{k-1}$  est définie par:

$$r_{X_2, \dots, X_{k-1}}(X_1, X_k) = \rho(X_1 - P_{M(X_2, \dots, X_{k-1})}, X_k - P_{M(X_2, \dots, X_{k-1})})$$

*exercice* soit  $X, Z_1, Z_2$  3 variables aléatoires indépendantes centrées de variance 1, et les variables  $X_1$  et  $X_2$  définies ainsi:

$$X_1 = X + Z_1 \text{ et } X_2 = X + Z_2$$

Calculer  $\rho(X_1, X_2)$  puis  $r_X(X_1, X_2)$ . Commenter.

**définition** soit un processus aléatoire  $(Y_t)_{t \in \mathbf{Z}}$  stationnaire faible centré. La fonction d'auto-corrélation partielle est définie de la manière suivante:

$$\begin{aligned} r(1) &= \rho(1) \\ r(h) &= r_{Y_2, \dots, Y_h}(Y_1, Y_{h+1}), \forall h \geq 2 \\ r(h) &= r(-h) \end{aligned}$$

on remarque que  $r(h) = r_{Y_2, \dots, Y_h}(Y_1, Y_{h+1}) = r_{Y_k, \dots, Y_{k+h}}(Y_k, Y_{k+1+h})$  car le processus est stationnaire.

Pour obtenir l'auto-corrélation partielle d'un processus, voyons les propriétés suivantes.

**propriété** considérons un processus stationnaire faible  $(Y_t)_{t \in \mathbf{Z}}$ . Supposons le centré et notons  $\gamma$  sa fonction d'autocovariance.

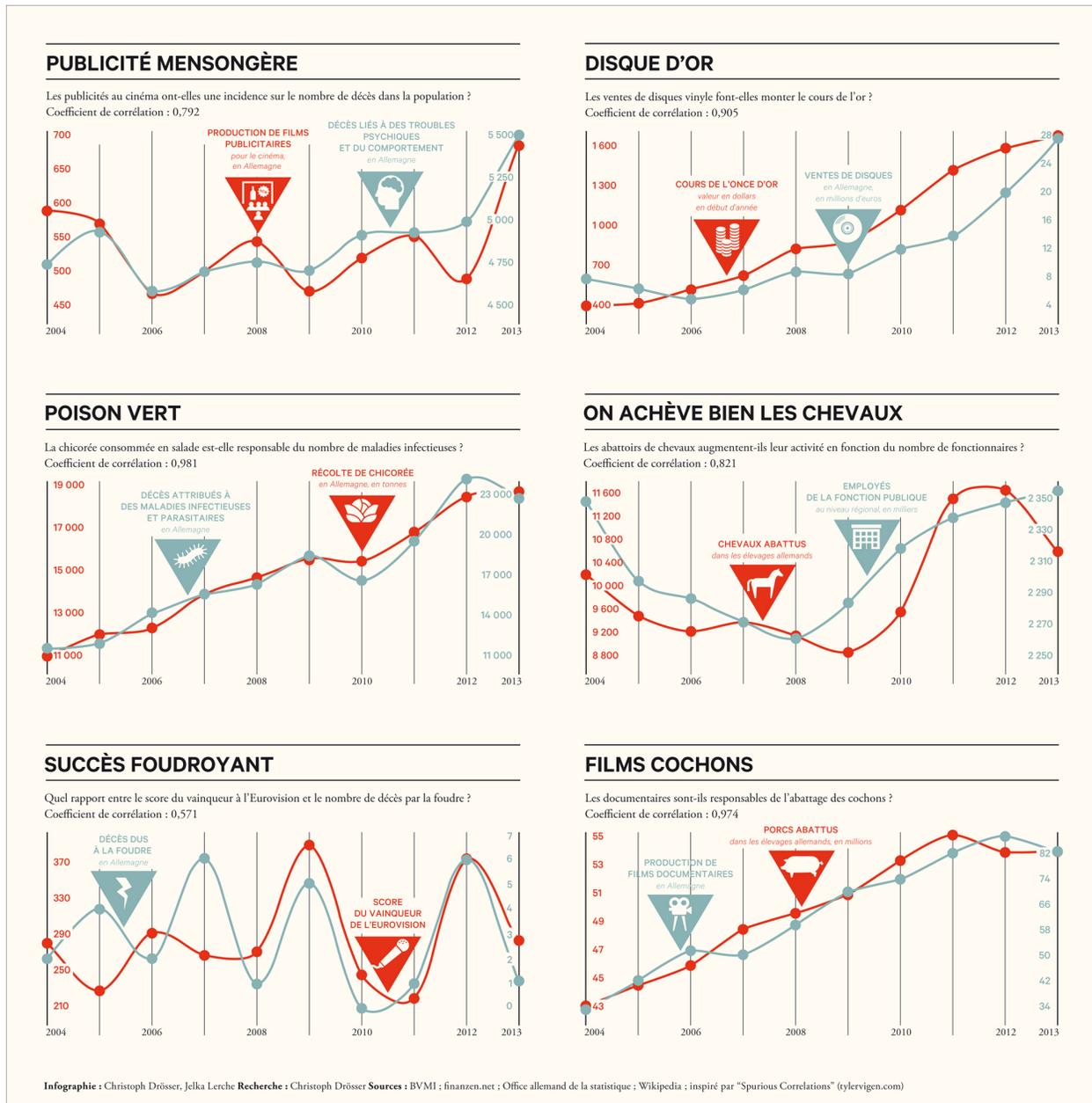


Figure 1: exemple de corrélations fallacieuses, source: <http://www.courrierinternational.com/grand-format/statistiques-les-correlations-de-labsurde>

La projection de  $Y_t$  sur son passé (espace vectoriel engendré par  $(Y_{t-1}, \dots, Y_{t-k})$ ) s'écrit:

$$P_{M_{t-1}^k}(Y_t) = \sum_{j=1}^k b_{k,j} Y_{t-j}$$

ou les coefficients s'obtiennent ainsi:

$$E(Y_t Y_{t-i}) = \sum_{j=1}^k b_{k,j} E(Y_{t-j} Y_{t-i})$$

$$\gamma(i) = \sum_{j=1}^k b_{k,j} \gamma(i-j)$$

en notant  $\Gamma_k$  la matrice  $[\gamma(i-j)]_{1 \leq i, j \leq k}$  on a:

$$\begin{pmatrix} b_{k,1} \\ \vdots \\ b_{k,k} \end{pmatrix} = \Gamma_k^{-1} \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(k) \end{pmatrix}$$

**propriété** soit un processus stationnaire faible  $(Y_t)_{t \in \mathbf{Z}}$  et  $P_{M_{t-1}^k}(Y_t) = b_{k,1} Y_{t-1} + \dots + b_{k,k} Y_{t-k}$  sa projection sur son passé, alors on a son auto-corrélation partielle d'ordre  $k$  vaut  $r(k) = b_{k,k}$ .

*preuve*

Pour  $k = 1$  c'est évident.

Pour  $k \geq 2$ , calculons

$$\text{cov}(Y_t - P_{M_{t-1}^{k-1}}(Y_t), Y_{t-k} - P_{M_{t-1}^{k-1}}(Y_{t-k}))$$

remarquons que

$$P_{M_{t-1}^k}(Y_t) = \sum_{j=1}^{k-1} b_{k,j} Y_{t-j} + b_{k,k} (Y_{t-k} - P_{M_{t-1}^{k-1}}(Y_{t-k})) + b_{k,k} P_{M_{t-1}^{k-1}}(Y_{t-k})$$

comme  $Y_{t-k} - P_{M_{t-1}^{k-1}}(Y_{t-k})$  est orthogonal à  $M_{t-1}^{k-1}(Y_t)$ , en projetant sur  $M_{t-1}^{k-1}(Y_t)$  on a:

$$P_{M_{t-1}^{k-1}}(Y_t) = \sum_{j=1}^{k-1} b_{k,j} Y_{t-j} + b_{k,k} P_{M_{t-1}^{k-1}}(Y_{t-k})$$

et donc:

$$\begin{aligned} \text{cov}(Y_t - P_{M_{t-1}^{k-1}}(Y_t), Y_{t-k} - P_{M_{t-1}^{k-1}}(Y_{t-k})) &= \text{cov}(Y_t - P_{M_{t-1}^k}(Y_t) + b_{k,k} Y_{t-k} - b_{k,k} P_{M_{t-1}^{k-1}}(Y_{t-k}), \dots) \\ &= \text{cov}(Y_t - P_{M_{t-1}^k}(Y_t), Y_{t-k} - P_{M_{t-1}^{k-1}}(Y_{t-k})) + b_{k,k} \text{var}(Y_{t-k} - P_{M_{t-1}^{k-1}}(Y_{t-k})) \\ &= b_{k,k} \text{var}(Y_{t-k} - P_{M_{t-1}^{k-1}}(Y_{t-k})) \end{aligned}$$

car  $Y_t - P_{M_{t-1}^k}(Y_t)$  est orthogonal à  $M_{t-1}^k(Y_t)$ . De plus,  $\text{var}(Y_{t-k} - P_{M_{t-1}^{k-1}}(Y_{t-k})) = \text{var}(Y_t - P_{M_{t-1}^{k-1}}(Y_t))$  et on a bien:

$$r(k) = b_{k,k}$$

On a donc que l'auto-corrélation partielle d'ordre  $k$  s'obtient comme le dernier coefficient du vecteur:

$$\begin{pmatrix} b_{k,1} \\ \vdots \\ b_{k,k} \end{pmatrix} = \Gamma_k^{-1} \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(k) \end{pmatrix}$$

En pratique, pour estimer la fonction des auto-corrélations partielles, on utilise l'algorithme de Durbin qui permet d'obtenir les auto-corrélations partielles de manière récursive (notons que, plus généralement, cet algorithme permet d'obtenir efficacement les coefficients d'une régression linéaire si on ajoute une variable à un modèle existant).

**algorithme de Durbin** l'algorithme de Durbin permet d'obtenir les auto-corrélation partielles  $r(k) = b_{k,k}$  d'un processus processus stationnaire faible  $(Y_t)_{t \in \mathbf{Z}}$  via la formule récursive suivante:

$$b_{k,j} = b_{k-1,j} - b_{k,k} b_{k-1,k-j}, \quad \forall j = 1, \dots, k-1$$

$$b_{k,k} = \frac{\gamma(k) - \sum_{j=1}^{k-1} \gamma(k-j) b_{k-1,j}}{\gamma(0) - \sum_{j=1}^{k-1} \gamma(j) b_{k-1,j}}$$

*preuve*

On a

$$P_{M_{t-1}^{k-1}}(Y_t) = \sum_{j=1}^{k-1} b_{k,j} Y_{t-j} + b_{k,k} P_{M_{t-1}^{k-1}}(Y_{t-k})$$

De plus  $P_{M_{t-1}^{k-1}}(Y_t) = \sum_{j=1}^{k-1} b_{k-1,j} Y_{t-j}$  et

$$P_{M_{t-1}^{k-1}}(Y_{t-k}) = \sum_{j=1}^{k-1} b_{k-1,j} Y_{t-(k-j)} = \sum_{j=1}^{k-1} b_{k-1,k-j} Y_{t-j}$$

ce qui donne les équations,  $\forall j \in (1, \dots, k-1)$

$$b_{k,j} = b_{k-1,j} - b_{k,k} b_{k-1,k-j}$$

or rappelons nous que:

$$\begin{pmatrix} b_{k,1} \\ \vdots \\ b_{k,k} \end{pmatrix} = \Gamma_k^{-1} \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(k) \end{pmatrix}$$

et donc:

$$\sum_{j=1}^k b_{k,j} \gamma(k-j) = \gamma(k)$$

ce qui permet d'avoir finalement:

$$b_{k,k} = \frac{\gamma(k) - \sum_{j=1}^{k-1} \gamma(k-j)b_{k-1,j}}{\gamma(0) - \sum_{j=1}^{k-1} \gamma(j)b_{k-1,j}}$$