

Séries Chronologiques

Université d'Angers – M2 Data Science

Support de cours

(F. Proïa)

Séries Chronologiques

Université d'Angers – M2 Data Science

Support de cours

(F. Proïa)

Table des matières

0	Introduction	2
0.1	Loi d'un processus stochastique (discret)	3
0.2	Bruits blancs et processus gaussiens	4
0.3	Tendance et saisonnalité	5
1	Estimation et élimination des tendances	7
1.1	Filtrage linéaire	7
1.2	Estimation de la tendance	8
1.3	Estimation de la saisonnalité	9
2	Processus stationnaires	11
2.1	Stationnarité stricte	11
2.2	Stationnarité au second ordre	12
2.3	Tester la stationnarité	14
2.4	Présentation des ARMA	16
2.5	Autocovariances et autocorrélations	16
2.6	Innovations	20
2.7	Autocorrélations partielles	22
3	Processus ARMA : théorie et pratique	26
3.1	Opérateur autorégressif	26
3.2	Autocovariances et autocorrélations	30
3.3	Portée pratique des racines anticipatives	31
3.4	Caractérisation	32
3.5	Écriture canonique	34
3.6	Équations de Yule-Walker	37
3.7	Estimation et propriétés des estimateurs	38
3.8	Prédictions ARMA	39
4	Quelques généralisations des ARMA	43
4.1	Processus ARIMA	43
4.2	Processus SARIMA	45
4.3	Prédictions ARIMA et SARIMA	47
4.4	Sélection et validation	49
4.5	Processus ARCH et GARCH	50
5	Introduction à la statistique spatiale	52
5.1	Variogramme	52
5.2	Contraintes du krigeage	55
5.3	Exemple du krigeage simple	55

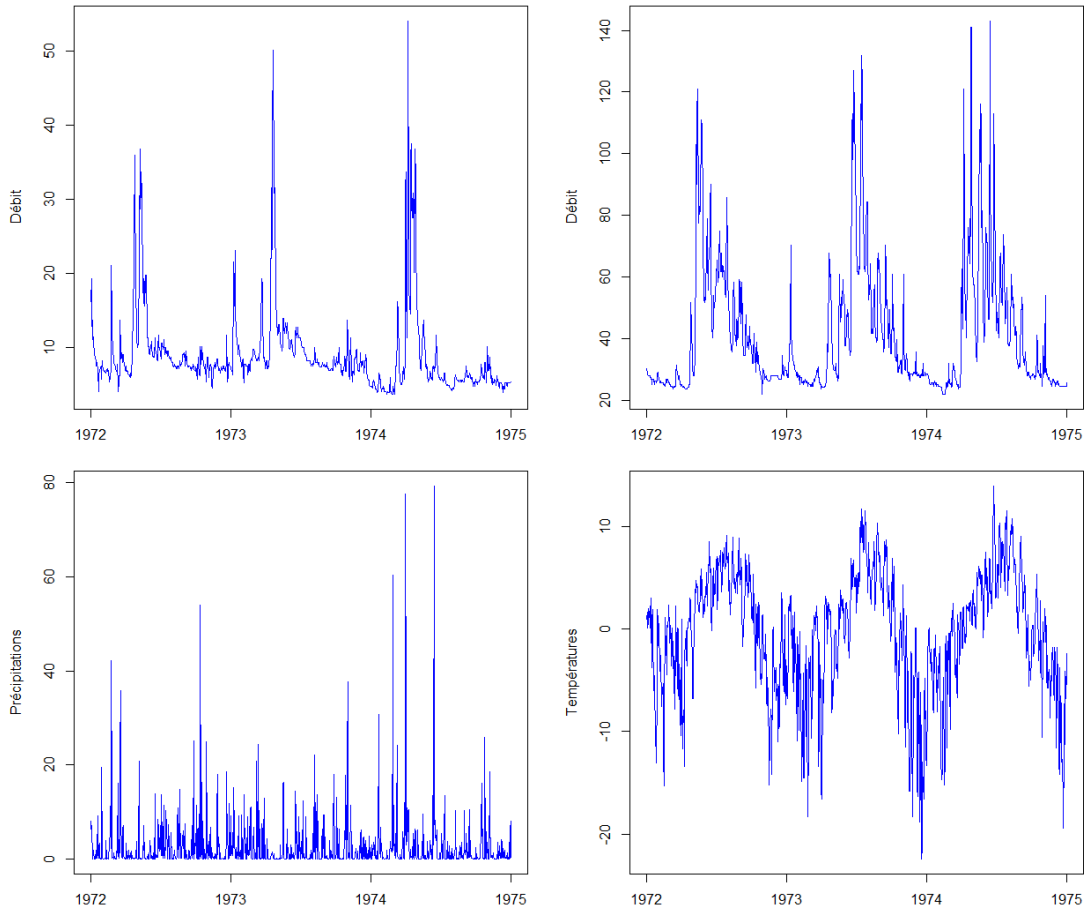
0 Introduction

Définition 0.1 On appelle *processus stochastique* toute famille $(X_t)_{t \in T}$ de v.a.r. définies sur un même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$, indexées par un espace T et à valeurs dans un même espace des états (E, \mathcal{E}) . Il s'agit d'une *série chronologique* dès que T est un espace des temps.

On dispose d'un échantillon de n observations récoltées séquentiellement dans le temps, et l'on suppose qu'il s'agit de la réalisation d'un vecteur aléatoire $\underline{X} = (X_1, \dots, X_n)$ extrait du processus $(X_t)_{t \in T}$. Cette modélisation est *chronologique* si l'indice est vu comme l'évolution du temps, et l'on va supposer l'existence de dépendances au sein de ce vecteur. Le processus est *discret* dès que T est dénombrable, sinon il est *continu*. Dans le cadre de ce cours, on considérera principalement des espaces des temps $T = \mathbb{Z}$ ou $T = \mathbb{N}$, des modélisation standards où l'échantillonnage est régulier, mais d'autres cas se rencontrent fréquemment (pas nécessairement temporels). En général :

- $T = \mathbb{R}$, le processus est à temps continu (ex : finance).
- $T = \mathbb{N}$ ou $T = \mathbb{R}^+$, le processus a un point de départ bien identifié (ex : marche aléatoire, mouvement brownien).
- $T = \mathbb{Z}^2$, le processus est quadrillé (ex : imagerie, statistique spatiale) ou $T = \mathcal{S}$, une sphère (ex : mesures à la surface de la Terre).
- $T = \{t_0, t_1, t_2, \dots\}$, l'échantillonnage est irrégulier et peut dépendre d'évènements extérieurs.

De manière formelle, le processus dépend de deux paramètres (ω, t) , respectivement l'aléa $\omega \in \Omega$ et le temps $t \in T$. À t fixé, l'application $\omega \mapsto X_t(\omega)$ est une variable aléatoire sur Ω . À ω fixé, l'application $t \mapsto X_t(\omega)$ est une fonction à valeurs dans E appelée la *trajectoire* du processus.

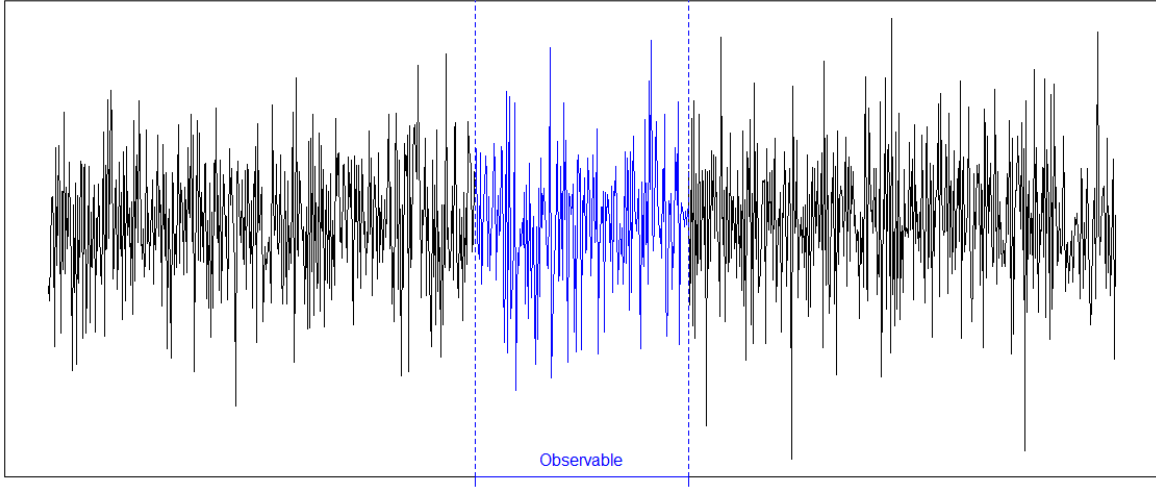


Ces quatre figures sont issues du jeu de données `ice.river` du package `tseries`. Un processus aléatoire défini sur \mathbb{Z} possède un *passé asymptotique* ainsi qu'un *futur asymptotique*, il évolue selon la chronologie

$$X_{t \rightarrow -\infty}, \dots, X_{-2}, X_{-1}, X_0, X_1, X_2, \dots, X_{t \rightarrow +\infty}$$

où $X_{t \rightarrow -\infty}$ et $X_{t \rightarrow +\infty}$ désignent de manière abusive son comportement asymptotique. S'il est défini sur \mathbb{N} , alors X_0 est sa valeur initiale et sa chronologie ne possède qu'un futur asymptotique. Sauf cas particuliers, le processus $(X_t)_{t \in T}$ n'est pas formé de v.a.r. indépendantes et identiquement distribuées. Au contraire, il existe généralement une corrélation forte entre deux valeurs proches dans le temps. Étant donné un processus $(X_t)_{t \in T}$ observable sur $T_n \subset T$, on cherche à :

- Identifier le type d'évolution du processus (corrélations, ...).
- Modéliser de façon pertinente cette évolution et estimer ses paramètres (ex : ARMA, ARIMA, ...).
- Être capable de prédire les valeurs futures du processus en fonction des observations passées.



0.1 Loi d'un processus stochastique (discret)

La loi d'une suite i.i.d. est très facile à caractériser. Dans le cas d'un processus aléatoire plus général, le problème devient extrêmement compliqué voire parfois insoluble, car la présence de dépendances implique de connaître la loi jointe de toutes les variables du processus. Pour formaliser cela, remarquons qu'on peut voir un processus $(X_t)_{t \in T}$ comme une application X de Ω dans l'ensemble E^T des trajectoires (des applications de T dans E), d'où la nécessité de commencer par munir E^T d'une tribu. Lorsque T est fini de cardinal n , $E^T \equiv E^n$ et on peut choisir la tribu produit $\mathcal{F} = \mathcal{E}^{\otimes n}$ mais dans le cas contraire, il est commun d'utiliser la plus petite tribu \mathcal{F} rendant mesurables toutes les applications coordonnées, soit

$$\forall t \in T, \quad \pi_t((x_s)_{s \in T}) = x_t.$$

Une telle tribu est engendrée par les *cylindres*

$$\mathcal{C} = \left\{ x \in E^T : x \in \prod_{t \in T} A_t \text{ où les } A_t \text{ vérifient } (*) \right\}$$

avec comme condition $(*)$: $A_t \in \mathcal{E}$ pour tout $t \in T$ et $A_t = E$ sauf pour un nombre *fini* d'entre eux (cette précision est importante pour la suite). Un cylindre $C = \{x \in E^T : x(t_1) \in A_{t_1}, \dots, x(t_k) \in A_{t_k}\} \in \mathcal{C}$ a pour image réciproque $X^{-1}(C) = \bigcap_i X_{t_i}^{-1}(A_{t_i})$ et c'est un élément de la tribu \mathcal{A} , en tant qu'intersection dénombrable d'éléments de \mathcal{A} . L'application $X : (\Omega, \mathcal{A}) \mapsto (E^T, \mathcal{F})$ est donc bien une variable aléatoire dont la loi \mathbb{P}_X est à construire sur (E^T, \mathcal{F}) .

Exemple. Si $T = \mathbb{N}$, les cylindres sont de la forme $B_0 \times \cdots \times B_n \times E \times E \times \cdots$ avec $B_i \in \mathcal{E}$, et la loi du processus jusqu'à l'instant n est définie par

$$\mathbb{P}_X^{(n)}(B_0 \times \cdots \times B_n) = \mathbb{P}(X_0 \in B_0, \dots, X_n \in B_n).$$

On remarque de plus qu'une *condition de compatibilité*

$$\forall n \in \mathbb{N}, \quad \mathbb{P}_X^{(n)}(B_0 \times \cdots \times B_n) = \mathbb{P}_X^{(n+1)}(B_0 \times \cdots \times B_n \times E)$$

est ici automatiquement satisfaite.

À l'aide d'un outillage sophistiqué (voir par exemple le *théorème d'extension de Kolmogorov*), on peut établir que les probabilités de ces cylindres déterminent entièrement \mathbb{P}_X .

Définition 0.2 Les lois fini-dimensionnelles, caractérisant la loi du processus $(X_t)_{t \in T}$, désignent l'ensemble des lois de $(X_t)_{t \in T_f}$ pour toutes les parties finies T_f de T , c'est-à-dire

$$\{\mathcal{L}(X_{t_1}, \dots, X_{t_k}) : (t_1, \dots, t_k) \in T^k, k \in \mathbb{N}^*\}.$$

Plus concrètement, si la loi de toute variable réelle X_t est bien caractérisée par sa fonction de répartition $\mathbb{P}_{X_t}(\cdot | -\infty, x] = \mathbb{P}(X_t \leq x)$, la loi du processus $(X_t)_{t \in T}$ dans son ensemble nécessite toutes les fonctions de répartition jointes (les *répartitions finies*)

$$\mathbb{P}_{(X_{t_1}, \dots, X_{t_k})}(\cdot | -\infty, x_1] \times \cdots \times \cdot | -\infty, x_k]) = \mathbb{P}(X_{t_1} \leq x_1, \dots, X_{t_k} \leq x_k)$$

et ce pour tous les vecteurs $(t_1, \dots, t_k) \in T^k$ de toutes les tailles $k \in \mathbb{N}^*$! À l'exception de cas notables, tels que les bruits blancs ou les processus gaussiens, on ne saura pas expliciter cette loi. Elle est pourtant déterminante pour faire des prévisions sur le comportement futur du processus ou estimer ses paramètres par maximum de vraisemblance. La notion plus faible de *stationnarité*, que l'on verra par la suite, se concentre sur les moments d'ordre 1 et 2 : sans connaître la loi mais seulement l'espérance et les covariances, on arrive malgré tout à quelques compromis intéressants.

0.2 Bruits blancs et processus gaussiens

Les fluctuations autour des tendances dans les modèles chronologiques sont bâties sur la notion fondamentale de bruit blanc.

Définition 0.3 On appelle *bruit blanc fort* tout processus $(X_t)_{t \in T}$ de v.a.r. indépendantes et identiquement distribuées.

Si $T = \{1, \dots, n\}$, il s'agit d'un n -échantillon au sens usuel du terme. Il est suffisant de connaître la loi de la variable parente X pour caractériser les répartitions finies d'un tel processus puisque dans ce cas,

$$\mathbb{P}(X_{t_1} \leq x_1, \dots, X_{t_k} \leq x_k) = \prod_{i=1}^k \mathbb{P}(X \leq x_i).$$

Définition 0.4 On appelle *bruit blanc (faible)* tout processus $(X_t)_{t \in T}$ de v.a.r. non corrélées, de même espérance m et de même variance σ^2 finie. On a

$$\forall t \in T, \quad \mathbb{E}[X_t] = m \quad \text{et} \quad \forall (t, s) \in T^2, \quad \text{Cov}(X_t, X_s) = \begin{cases} \sigma^2 & \text{si } t = s \\ 0 & \text{si } t \neq s. \end{cases}$$

Il est clair que le choix $m = 0$ est le plus adapté à la modélisation (si $m \neq 0$, on peut toujours considérer m comme une tendance déterministe et le bruit comme centré autour). Par la suite, le terme *bruit blanc* sera implicitement associé à $m = 0$.

Remarque. Un bruit blanc fort est donc également un bruit blanc faible s'il est de variance finie.

Définition 0.5 Un processus $(X_t)_{t \in T}$ est gaussien si pour tout $k \in \mathbb{N}^*$ et pour tout vecteur d'indices $T_k = (t_1, \dots, t_k) \in T^k$, on a

$$(X_{t_1}, \dots, X_{t_k}) \sim \mathcal{N}_k(\mu_{T_k}, \Sigma_{T_k}).$$

En particulier, les combinaisons linéaires $Z = a_1 X_{t_1} + \dots + a_k X_{t_k}$ sont également gaussiennes (si $\exists a_i \neq 0$... sinon il faut convenir qu'une loi gaussienne dégénérée $\mathcal{N}(0, 0)$ est assimilable à une masse de Dirac δ_0). Les répartitions finies du processus ne dépendent alors que d'une fonction d'espérance $t \mapsto \mathbb{E}[X_t]$ et d'une fonction de covariance $(t, s) \mapsto \text{Cov}(X_t, X_s)$.

Remarque. Un bruit blanc gaussien est nécessairement un bruit blanc fort.

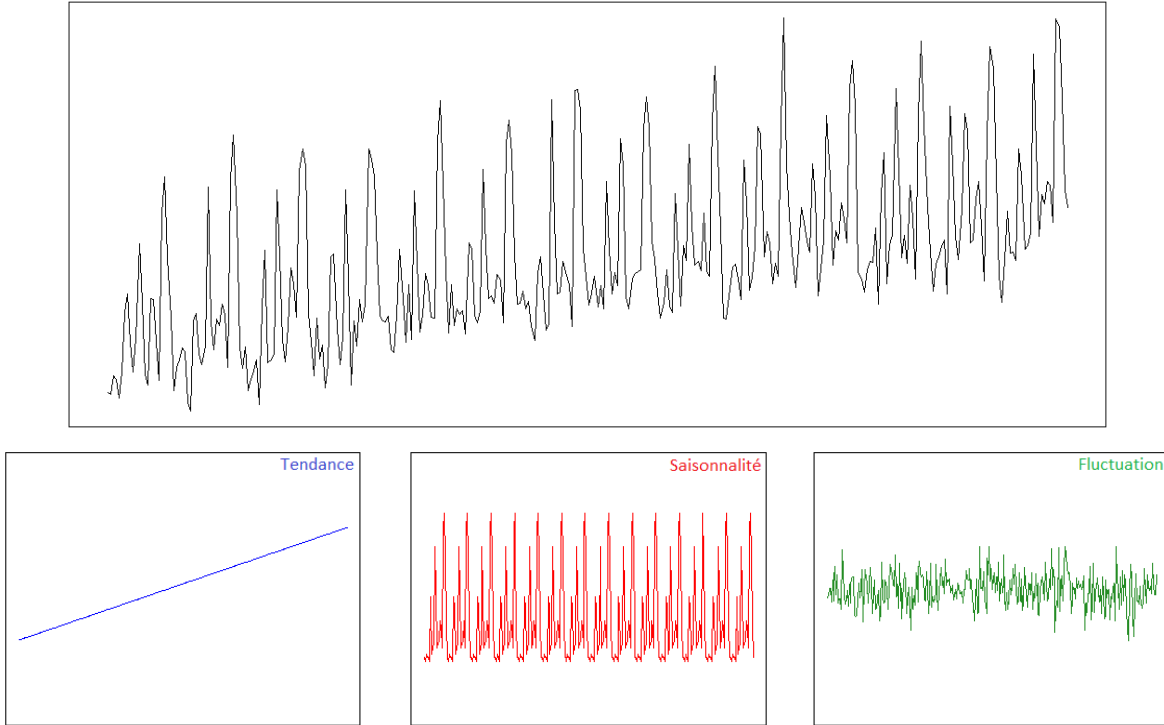
0.3 Tendances et saisonnalité

En général, les séries observées présentent des structures *macroscopiques* visuellement identifiables. En première analyse, on peut chercher à éliminer ces tendances déterministes avant de modéliser les fluctuations résiduelles. Par exemple,

$$X_t = m_t + S_t + Z_t$$

est une *modélisation additive* fréquemment rencontrée, dans laquelle m_t est une tendance (généralement polynomiale : $m_t = a_0$, $m_t = a_0 + a_1 t$, ...), S_t est une composante τ -périodique (qui peut prendre la forme de sinusôides ou simplement d'un motif répété) et Z_t est un terme de fluctuation. Dans cette décomposition :

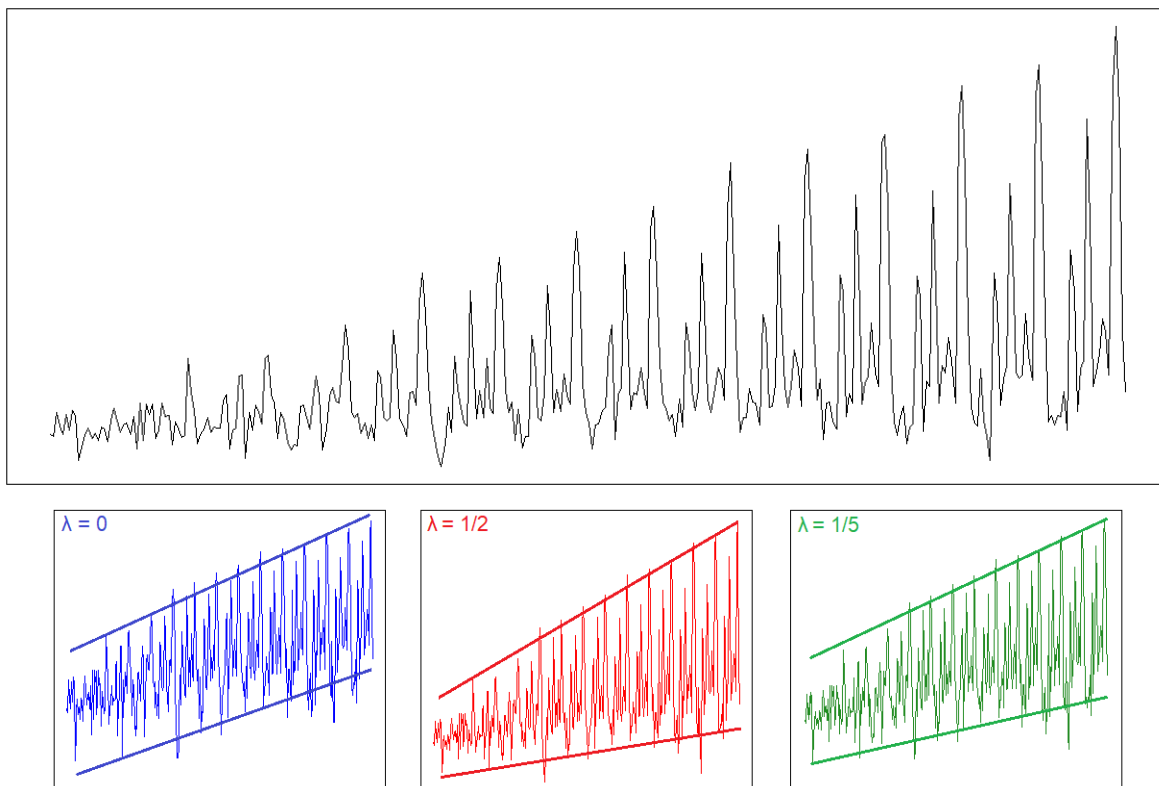
- Les processus (m_t) et (S_t) sont déterministes, ce sont des contributions certaines du temps.
- Le processus (Z_t) est stochastique, il contient les phénomènes de variabilité, de dépendance, etc.



On rencontre également des phénomènes s'amplifiant avec le temps, auxquels cas une *modélisation multiplicative* de la forme $X_t = m_t (1 + S_t) + Z_t$ ou $X_t = m_t S_t Z_t$ peut être pertinente. Une possibilité est de chercher à se ramener à des données adaptées à la modélisation additive par une *transformation de Box-Cox*,

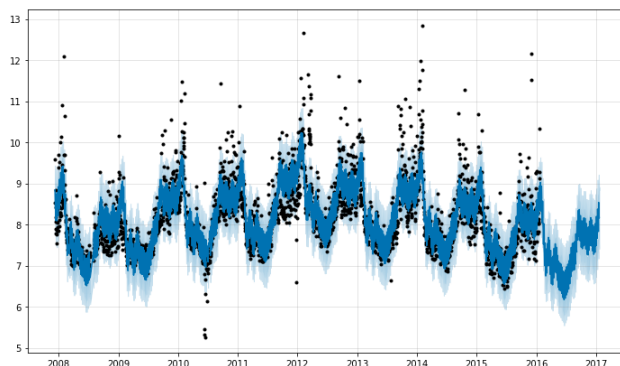
$$Y_t = \begin{cases} \frac{X_t^\lambda - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \ln X_t = \lim_{\lambda \rightarrow 0^+} \frac{X_t^\lambda - 1}{\lambda} & \text{si } \lambda = 0 \end{cases}$$

en prenant garde au signe de X_t , quitte à translater le processus au besoin. Dans l'exemple ci-dessous, la transformation logarithmique paraît la plus adéquate, ou dans une moindre mesure $\lambda = 1/5$. Au contraire, $\lambda = 1/2$ ne semble pas atténuer suffisamment l'amplification de la saisonnalité. On pourra caractériser (grossièrement) l'évolution des pics haut et bas de la saisonnalité par des droites et étudier visuellement leur parallélisme. Une règle générale à respecter : la modélisation chronologique d'une série passe nécessairement par une *représentation graphique* préalable, afin d'identifier les éventuelles transformations de données, tendances et/ou saisonnalités.



Ces modèles sont simplistes mais leur facilité de mise en œuvre leur donne un statut de référence : toujours avoir une base solide avant de se lancer dans des modèles complexes !

Remarque. Un modèle additif plus sophistiqué s'est popularisé récemment dans un contexte professionnel, intitulé – en toute humilité – *Facebook Prophet*. Ce dernier tient compte de tendances non-linéaires, de périodicités annuelle, mensuelle et journalière, et incorpore même un effet 'dimanches et jours fériés'. Un exercice de TP y est dédié, l'image ci-dessous provient du GitHub de Facebook.



1 Estimation et élimination des tendances

Dans cette section, on fait l'hypothèse que la série observée est une réalisation d'un modèle additif de la forme $X_t = m_t + S_t + Z_t$ sur un espace des temps T discret. Mais un tel modèle n'est pas *identifiable* (en posant $m'_t = m_t + \delta$ et $S'_t = S_t - \delta$, il est clair que $X_t = m_t + S_t + Z_t = m'_t + S'_t + Z_t$). On peut alors supposer que la saisonnalité est centrée sur sa période, c'est-à-dire que

$$\forall t \in T, \quad S_t = S_{t+\tau} \quad \text{avec} \quad \bar{S}_\tau = \frac{1}{\tau} \sum_{k=1}^{\tau} S_k = 0.$$

Cette hypothèse n'est pas restrictive car si la saisonnalité n'est pas centrée sur sa période, le décentrage sera absorbé par m_t (choisir $\delta = \bar{S}_\tau$). Il existe la commande **decompose** dans le package **stats** qui fait tout le travail d'estimation et d'élimination (ce sera vu en TP), mais notre objectif est ici de comprendre le principe du *filtrage* utilisé dans le cadre d'un tel modèle additif, le plus simple qui soit.

1.1 Filtrage linéaire

Définition 1.1 On appelle *opérateur retard* et l'on note généralement B l'opérateur qui à une donnée chronologique associe sa valeur précédente dans le temps.

La notation B provient de l'anglais *backward* ou *backshift*, mais il est fréquent de rencontrer aussi la notation L (comme *lag*). Par extension, $B^0 = I$ et $B^k = B \circ B \circ \dots \circ B$ à k reprises. Sur $T = \{\dots, t_{-1}, t_0, t_1, t_2, \dots\}$,

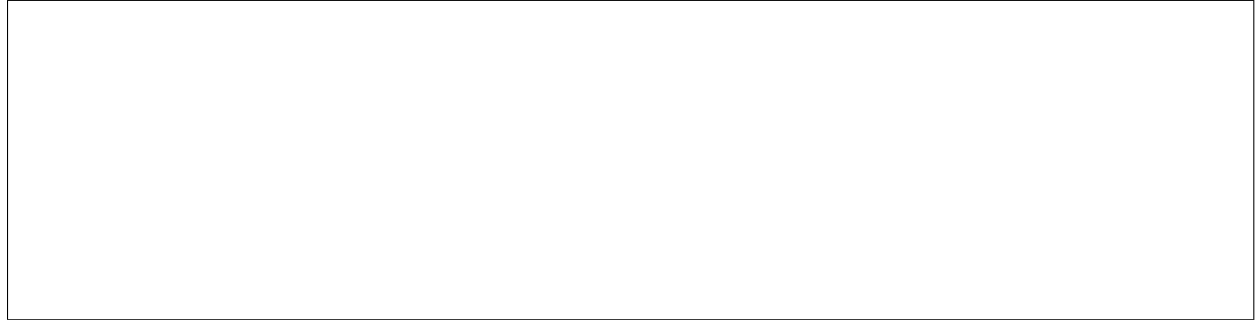
$$B^0 X_{t_i} = I X_{t_i} = X_{t_i} \quad \text{et} \quad B^k X_{t_i} = X_{t_{i-k}} \quad \text{pour } k \geq 1.$$

On peut utiliser les notations B^{-1} et $B^{-k} = B^{-1} \circ B^{-1} \circ \dots \circ B^{-1}$ pour travailler sur le futur d'un processus, l'opérateur B^{-1} étant l'*inverse* de l'opérateur B dans le sens où

$$B \circ B^{-1} = B^{-1} \circ B = I.$$

Sur $T = \mathbb{Z}$ ou \mathbb{N} , on a simplement $t_i = i$ et $BX_t = X_{t-1}$.

Exemples. Quelques exemples :



Définition 1.2 Soit $(a_k)_{k \in \mathbb{Z}}$ une suite de $\ell^1(\mathbb{Z})$. L'opérateur défini par

$$F(B) = \sum_{k \in \mathbb{Z}} a_k B^k$$

est un *filtre linéaire*. Un tel filtre est *causal* si $a_k = 0$ dès que $k < 0$.

Les filtres linéaires se combinent facilement par les règles polynomiales usuelles, par exemple

$$\begin{aligned} (a_0 I + a_1 B) \circ (b_1 B + b_2 B^2) &= (a_0(b_1 B + b_2 B^2) + a_1(b_1 B^2 + b_2 B^3)) \\ &= a_0 b_1 B + a_0 b_2 B^2 + a_1 b_1 B^2 + a_1 b_2 B^3 = (a_0 I + a_1 B)(b_1 B + b_2 B^2). \end{aligned}$$

Ceci vient du fait que $B^k \circ B^{k'} = B^{k'} \circ B^k = B^{k+k'}$ et de la linéarité de $B : B(X_t + \lambda Y_t) = BX_t + \lambda BY_t$.

Définition 1.3 Pour tout $m \in \mathbb{N}$, le filtre

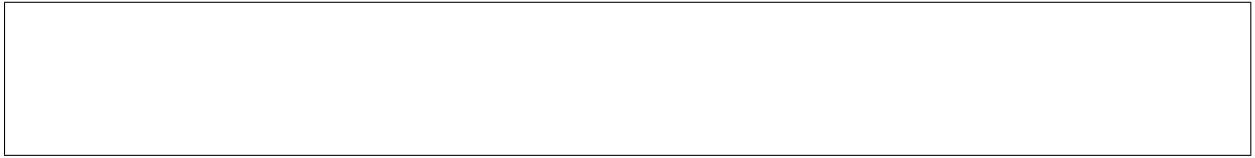
$$M_m(B) = \frac{1}{2m+1} \sum_{k=-m}^m B^{-k}$$

est une moyenne mobile arithmétique d'ordre $2m+1$. Pour tout $m \in \mathbb{N}^*$, le filtre

$$M_m^*(B) = \frac{1}{2m} \sum_{k=-m+1}^{m-1} B^{-k} + \frac{B^m + B^{-m}}{4m}$$

est une moyenne mobile arithmétique modifiée d'ordre $2m+1$.

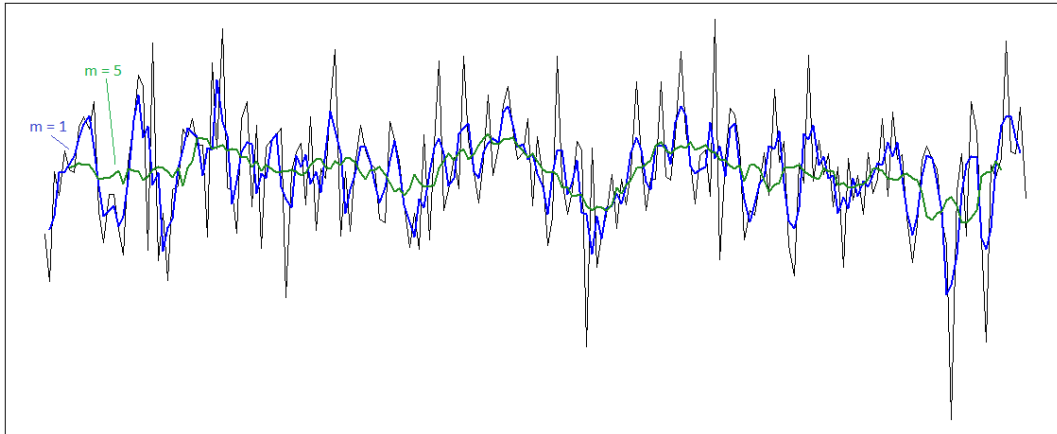
Exemples. Quelques exemples :



Le lissage par moyennisation de la série fait généralement diminuer sa variance (à condition de choisir des coefficients a_k raisonnables) mais, en contrepartie, on crée des corrélations artificielles. Par exemple,

$$\text{Cov}(\varepsilon_t, \varepsilon_{t-1}) = 0 \quad \text{mais} \quad \text{Cov}(M_m(B)\varepsilon_t, M_m(B)\varepsilon_{t-1}) = \frac{2m}{(2m+1)^2} \sigma^2 \neq 0.$$

Cette remarque n'a qu'une portée théorique car, dans notre contexte, ce n'est pas gênant : ces moyennes mobiles sont là pour supprimer la saisonnalité et lisser le signal en vue d'en extraire une estimation de la tendance de long terme. Par contre si la série est de taille finie (typiquement, sur des vraies données), on fausse également les m premières et dernières observations (**decompose** les remplace par des NA).

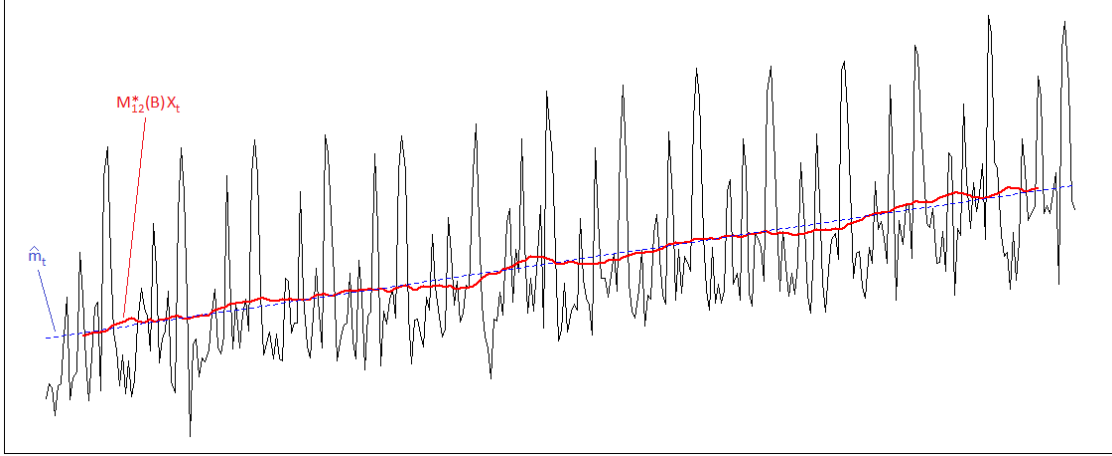


1.2 Estimation de la tendance

Proposition 1.1 On a les effets suivants :

- $M_m(B)$ annule les tendances τ -périodiques lorsque $\tau = 2m+1$ est impair.
- $M_m^*(B)$ annule les tendances τ -périodiques lorsque $\tau = 2m$ est pair.

Preuve. Il suffit d'exploiter la périodicité du signal et la contrainte de centrage sur chaque période (à faire pour s'entraîner...). Sans la contrainte, elles sont ramenées à des constantes.



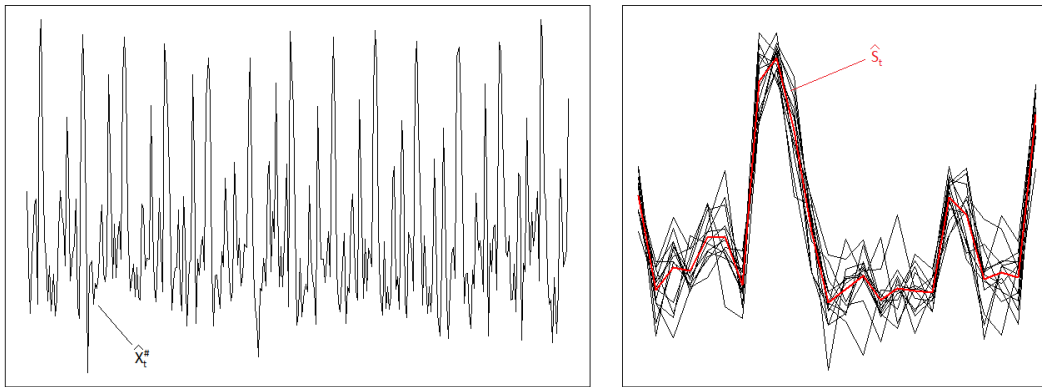
Si la saisonnalité est éliminée correctement, il est alors possible d'estimer la tendance. En général, on associe une forme paramétrique à m_t ($m_t = a_0$, $m_t = a_0 + a_1 t$, ...) puis on estime ses paramètres. L'exemple ci-dessus est simulé avec $\tau = 24$, on y trouve la série filtrée par $M_{12}^*(B)$ pour éliminer la saisonnalité ainsi qu'une estimation linéaire de la tendance. La méthode est très efficace à condition que le filtre ne perturbe 'pas trop' la tendance m_t (sinon ce qu'on gagne d'un côté, on le perd de l'autre). C'est le cas par exemple des moyennes mobiles arithmétiques pour les tendances linéaires $m_t = a_0 + a_1 t$, et c'est une situation courante (on suppose généralement que la tendance m_t varie lentement, de sorte qu'une approximation linéaire est souvent pertinente). On peut se contenter du signal filtré sans ajouter une modélisation, mais alors le modèle n'est pas prédictif (ci-dessus je peux *prédire* avec le bleu mais je peux seulement *décrire* avec le rouge).

1.3 Estimation de la saisonnalité

Une méthode (parmi d'autres) pour estimer la saisonnalité est de l'estimer sur une période, puis de la prolonger par périodicité. On va supposer pour simplifier que la série est observée sur un intervalle de temps $n = N\tau$ de sorte qu'un nombre exact de périodes est disponible. La série considérée est soit la série originale (si elle n'avait pas de tendance, auquel cas $\hat{m}_t = 0$), soit la série privée de son estimation de la tendance. On la note dans ce qui suit

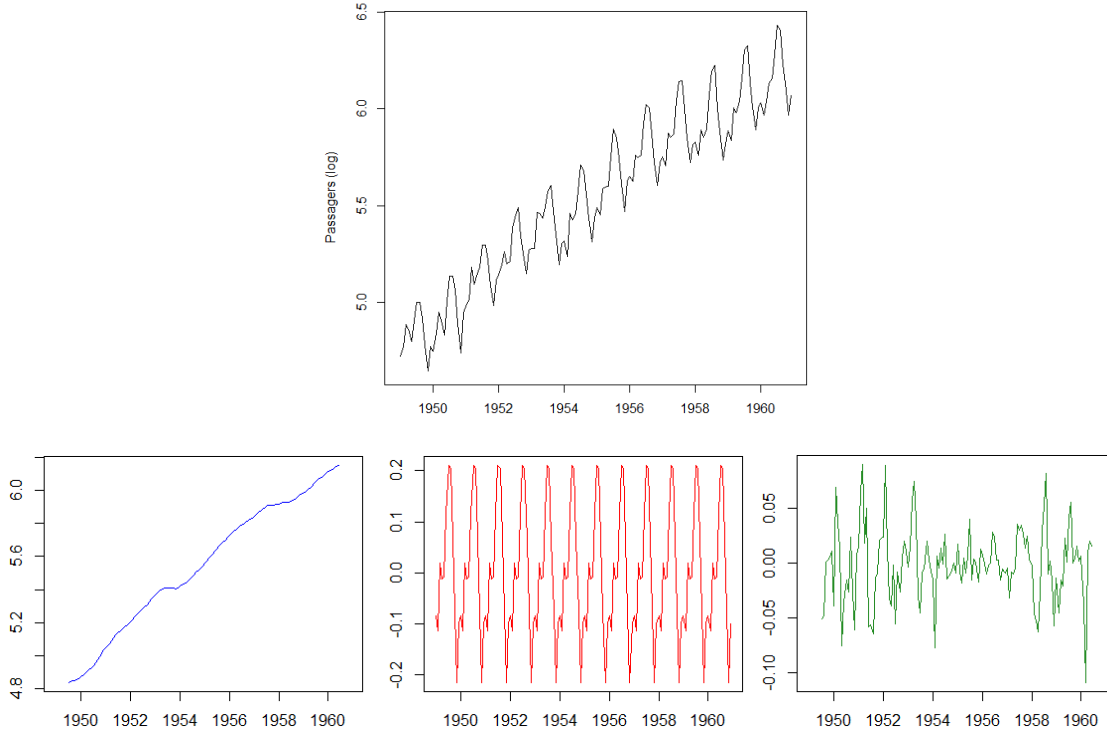
$$\hat{X}_t^\# = X_t - \hat{m}_t.$$

L'estimation de la saisonnalité consiste alors à superposer toutes les périodes disponibles, à les moyenniser pour obtenir le motif saisonnier, puis à recentrer le motif pour qu'il vérifie la contrainte $\hat{S}_1 + \dots + \hat{S}_\tau = 0$. Enfin, on le prolonge par périodicité.



On peut aussi effectuer des régressions sur des sinusôides (approche Fourier).

Que ce soit pour l'estimation de la tendance ou de la saisonnalité, on a supposé que la période τ était connue. Comment s'en assurer? Le plus souvent, c'est visuellement évident ou c'est inhérent aux données manipulées (annuelles, hebdomadaires, horaires, etc.). En cas de doute, on peut tracer le *périodogramme* de la série, notion étroitement liée à sa transformée de Fourier (`periodogram` du package `TSA`) qui mettra en évidence les phénomènes périodiques de la série. On représente ci-dessous la décomposition additive de la transformation logarithmique du jeu de données `AirPassengers`, disponible dans le package `datasets` directement obtenue par la commande `decompose`.



On peut éventuellement affiner la décomposition en réestimant une tendance sur le résidu $X_t - \hat{S}_t$ pour ensuite réestimer la saisonnalité sur $X_t - \hat{m}_t$, et ainsi de suite jusqu'à ce que les estimateurs ne varient plus significativement. Le processus résiduel

$$X_t - \hat{m}_t - \hat{S}_t$$

est censé contenir les phénomènes de nature stochastique (aléa, corrélation, etc.). Nous allons étudier quelques modélisations possibles de ces *fluctuations*.

Remarque. On vient de proposer quelques méthodes d'estimation des tendances mais on peut parfois simplement vouloir les éliminer, auquel cas les opérateurs de différenciation sont bien adaptés. On pourra vérifier facilement que :

- L'opérateur $(I - B)^d$ ramène les tendances polynomiales d'ordre d à des tendances constantes alors que $(I - B)^{d+1}$ les élimine totalement.
- L'opérateur $I - B^\tau$ élimine totalement les tendances τ -périodiques et ramène les tendances polynomiales d'ordre d à des tendances polynomiales d'ordre $d - 1$.

Exemple. Le filtre $(I - B)(I - B^4)$ ramène le processus $X_t = 1 + 2t + \cos(\frac{\pi t}{2}) + Z_t$ à une fluctuation (centrée).



2 Processus stationnaires

On commence par s'intéresser à la notion de *stationnarité* qui, dans sa version faible, est une façon simple (incomplète mais efficace, comme le montre la mise en pratique) de caractériser le comportement d'un processus sans passer par ses répartitions finies. Désormais, on considérera uniquement des séries définies sur $T = \mathbb{Z}$ ou sur $T = \mathbb{N} \cup \mathbb{I}$ où \mathbb{I} contient les instants initiaux. Avant toute chose, quelques rappels sur l'espace $L^2(\Omega, \mathcal{A}, \mathbb{P})$ s'imposent.

Définition 2.1 On note $L^2 = L^2(\Omega, \mathcal{A}, \mathbb{P})$ l'espace des v.a.r. X définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ telles que

$$\mathbb{E}[X^2] < +\infty.$$

L'espace L^2 est un espace de Hilbert (e.v. complet) pour le produit scalaire défini par

$$\langle X, Y \rangle = \mathbb{E}[XY].$$

La norme issue de ce produit scalaire est donnée par

$$\|X\| = \sqrt{\langle X, X \rangle} = \sqrt{\mathbb{E}[X^2]}.$$

Les outils fondamentaux pour étudier la stationnarité faible, que l'on introduira afin d'assouplir la stationnarité stricte, sont les moments d'ordre 1 et 2.

Définition 2.2 Soient $X, Y \in L^2$. La covariance de X et Y est définie par

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

En particulier,

$$\mathbb{V}(X) = \text{Cov}(X, X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

est la variance de X . Enfin, la corrélation entre X et Y est donnée par

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\mathbb{V}(X)}\sqrt{\mathbb{V}(Y)}}.$$

Remarque. On dira que X et Y ne sont pas corrélées lorsque $\text{Cov}(X, Y) = \text{Corr}(X, Y) = 0$. De plus, pour des variables centrées, la non corrélation équivaut à l'orthogonalité. En effet, on a alors $\langle X, Y \rangle = 0$, ce qui signifie que X et Y sont *orthogonales* dans L^2 .

2.1 Stationnarité stricte

Définition 2.3 Une série chronologique $(X_t)_{t \in \mathbb{Z}}$ est dite strictement stationnaire si, pour tout $n \geq 1$ et tout vecteur $(t_1, \dots, t_n) \in \mathbb{Z}^n$, les vecteurs

$$(X_{t_1}, \dots, X_{t_n}) \quad \text{et} \quad (X_{t_1+k}, \dots, X_{t_n+k})$$

ont la même loi, et ce pour tout décalage temporel $k \in \mathbb{Z}$.

La stationnarité forte caractérise la capacité d'un processus à se décorrérer totalement de l'indice temporel : sa loi est *invariante par translation*. La loi d'un processus strictement stationnaire ne dépend donc pas du temps. Cette propriété peut être complexe à établir et reposer sur des hypothèses fortes.

Exemple. Le bruit blanc fort est strictement stationnaire.

2.2 Stationnarité au second ordre

Définition 2.4 On appelle processus du second ordre tout processus stochastique $(X_t)_{t \in \mathbb{Z}}$ tel que

$$\forall t \in \mathbb{Z}, \quad X_t \in L^2.$$

Sa moyenne m et sa covariance Γ sont définies par

$$\forall t \in \mathbb{Z}, \quad m(t) = \mathbb{E}[X_t] \quad \text{et} \quad \forall (t, s) \in \mathbb{Z}^2, \quad \Gamma(t, s) = \text{Cov}(X_t, X_s).$$

Remarque. La moyenne et la covariance d'un processus stochastique sont donc des *fonctions du temps*. En particulier, $m : \mathbb{Z} \rightarrow \mathbb{R}$ alors que $\Gamma : \mathbb{Z}^2 \rightarrow \mathbb{R}$. On remarque que la fonction de covariance est symétrique.

Définition 2.5 Une série chronologique $(X_t)_{t \in \mathbb{Z}}$ du second ordre est dite stationnaire au second ordre si sa moyenne et sa covariance sont invariantes par translation dans le temps. Autrement dit :

$$\rightarrow \mathbb{E}[X_t] = m(t) = m \text{ pour tout } t \in \mathbb{Z}.$$

$$\rightarrow \text{Cov}(X_t, X_s) = \Gamma(t, s) = \Gamma(t + k, s + k) \text{ pour tous } (t, s) \in \mathbb{Z}^2 \text{ et tout décalage temporel } k \in \mathbb{Z}.$$

On en déduit en particulier qu'un processus stationnaire est de variance constante (prendre $t = s$). On a maintenant

$$m(t) = m \quad \text{et} \quad \Gamma(t, s) = \Gamma(t - s, 0) = \Gamma(h, 0) = \gamma(h)$$

en posant $h = t - s$, ce qui montre que la fonction de covariance d'un processus stationnaire au second ordre n'est fonction que d'une seule variable (le décalage temporel entre les deux instants de mesure).

Exemple. Le bruit blanc (faible) est stationnaire au second ordre, mais ce n'est pas nécessairement le cas du bruit blanc fort (encore faut-il que ses moments d'ordre 2 existent).

Proposition 2.1 Une série chronologique $(X_t)_{t \in \mathbb{Z}}$ strictement stationnaire est stationnaire au second ordre dès que sa fonction de covariance existe. Cependant, la réciproque est fausse.

Preuve.

Proposition 2.2 *Un processus gaussien $(X_t)_{t \in \mathbb{Z}}$ stationnaire au second ordre est également strictement stationnaire.*

Preuve. Sa moyenne et sa covariance déterminent entièrement ses répartitions finies (reprendre la Déf. 0.5 avec $T = \mathbb{Z}$ et exprimer μ_{T_k} et Σ_{T_k} en fonction de m et de Γ).

Remarque. On parle de série *stationnaire au second ordre*, ou de série *faiblement stationnaire*, *stationnaire dans L^2* ou même simplement de série *stationnaire*. De même, on parle de série *strictement stationnaire* ou de série *fortement stationnaire*.

Exemple. Soit le processus défini par

$$\forall t \in \mathbb{Z}, \quad X_t = \begin{cases} A + \varepsilon_t & \text{si } t \text{ est pair} \\ B + \varepsilon_t & \text{si } t \text{ est impair} \end{cases}$$

où $A \sim \mathcal{N}(\mu_A, \tau_A^2)$, $B \sim \mathcal{N}(\mu_B, \tau_B^2)$, $\text{Cov}(A, B) = \delta$ et (ε_t) est un bruit blanc de variance σ^2 indépendant de A et de B . Déterminer les conditions sur les paramètres pour que (X_t) soit stationnaire, puis étudier ses corrélations dans le temps.



Exemples. Étudier la stationnarité et les corrélations dans le temps des processus définis par

$$\forall t \in \mathbb{Z}, \quad X_t = A \cos(\theta t) \quad \text{et} \quad Y_t = A \cos(\theta t) + B \sin(\theta t)$$

où $\theta \in \mathbb{R}$, $\mathbb{E}[A] = \mathbb{E}[B] = 0$, $\mathbb{E}[A^2] = \mathbb{E}[B^2] = 1$ et $A \perp B$. On pourra traiter à part le cas où $\theta = 0$.



Remarque. Les définitions de stationnarité proposées sont valables sur \mathbb{Z} . Sur $\mathbb{N} \cup \mathbb{I}$, quelques précautions s'imposent mais l'essentiel est préservé. Les conditions doivent porter sur les vecteurs $(t_1, \dots, t_n) \in (\mathbb{N} \cup \mathbb{I})^n$ et les décalages temporels sur $k \geq 0$. Parfois des valeurs initiales arbitraires empêchent la stationnarité au sens de ces définitions rigides mais permettent de l'obtenir lorsque la taille de la série grandit, elles ont une influence décroissante avec le temps. On parle alors de *stationnarité asymptotique*.

Exemple. À partir d'une valeur initiale $X_0 \in L^2$ et d'un bruit $(\varepsilon_n)_{n \geq 1}$ indépendant de X_0 et de variance σ^2 finie, on génère le processus $(X_n)_{n \geq 0}$ par la relation

$$\forall n \geq 1, \quad X_n = \frac{1}{2} X_{n-1} + \varepsilon_n.$$

Montrer que le processus est asymptotiquement stationnaire.

2.3 Tester la stationnarité

Les résultats établis dans ce cours traitent en majorité de processus pour lesquels on fait l'hypothèse de stationnarité. Pour les mettre en pratique, encore faut-il que la trajectoire observable (X_1, \dots, X_n) puisse être considérée comme issue d'un processus stationnaire. On présente deux tests classiques :

→ Le *test ADF* (pour “Dickey-Fuller Augmenté”) permet de tester

$$\mathcal{H}_0 : \text{“la trajectoire est issue d'un processus non stationnaire”} \quad \text{contre} \quad \mathcal{H}_1 = \bar{\mathcal{H}}_0.$$

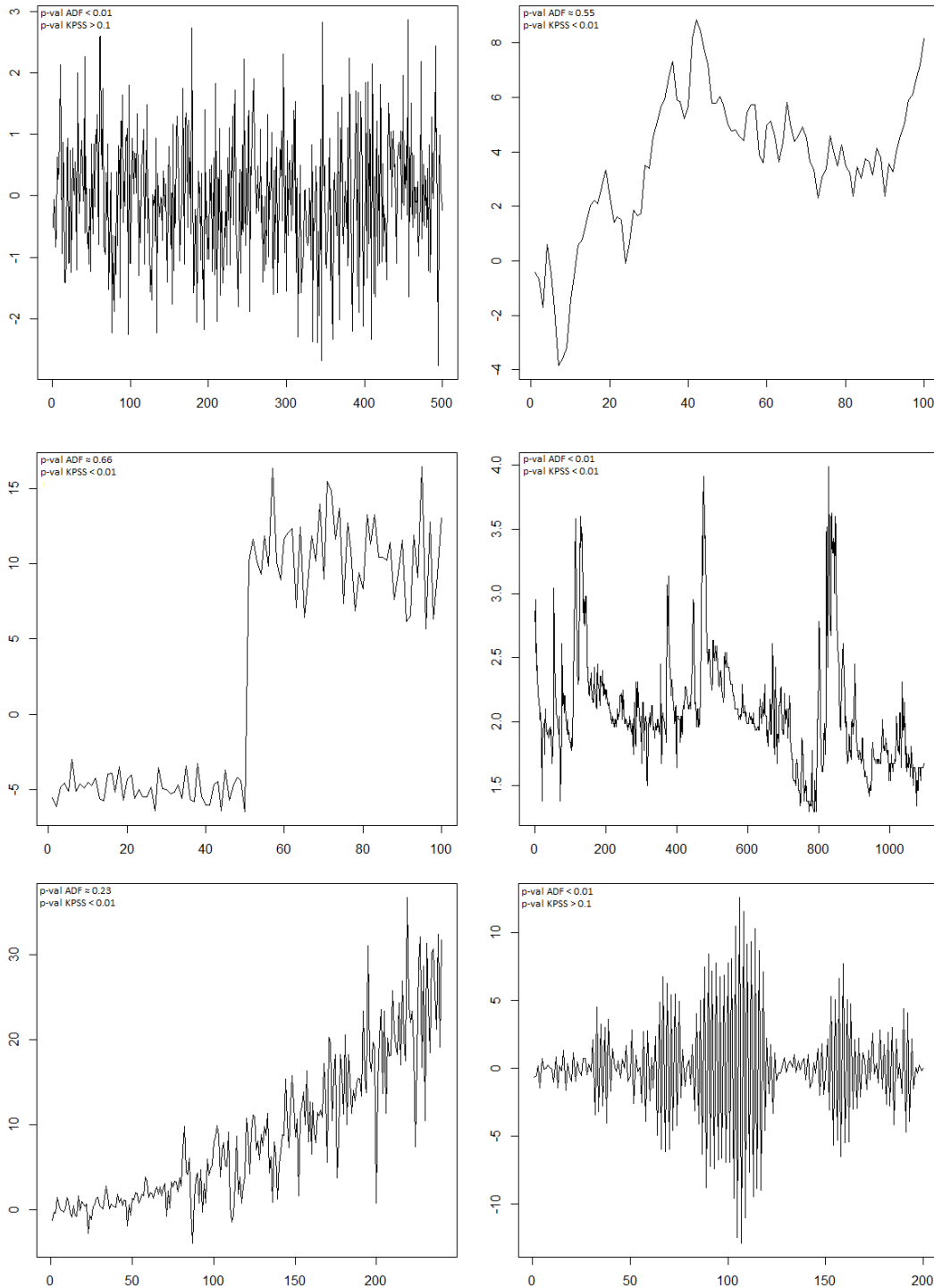
→ Le *test KPSS* (pour “Kwiatkowski-Phillips-Schmidt-Shin”) permet de tester

$$\mathcal{H}_0 : \text{“la trajectoire est issue d'un processus stationnaire”} \quad \text{contre} \quad \mathcal{H}_1 = \bar{\mathcal{H}}_0.$$

Ces tests (commandes `adf.test` et `kpss.test` du package `aTSA`) ont des hypothèses nulles complémentaires, il est donc intéressant de les combiner. Attention : selon les versions utilisées, il est possible de spécifier des tests *autour de tendances* (ce qui signifie qu'une tendance constante ou linéaire est éliminée avant de faire le test sur les résidus), ce qui peut être source d'erreurs et de conclusions hâtives si l'on n'y prend pas garde. On ne peut pas décider à l'œil nu qu'une trajectoire est stationnaire, par contre on peut avoir

l'intuition qu'une trajectoire *n'est pas* stationnaire. L'étape visuelle est indispensable car ces tests sont loin d'être infaillibles, ils reposent sur des structures génératrices particulières potentiellement inadaptées à la série étudiée et doivent être utilisés et interprétés correctement.

Exemples. Parmi ces trajectoires, lesquelles pourraient et lesquelles ne pourraient pas être considérées comme stationnaires ?



2.4 Présentation des ARMA

La modélisation chronologique la plus utilisée dans un cadre stationnaire est probablement l'écriture ARMA. Dans ce qui suit, on considère les opérateurs

$$\Phi(B) = I - \phi_1 B - \dots - \phi_p B^p \quad \text{et} \quad \Theta(B) = I + \theta_1 B + \dots + \theta_q B^q$$

avec $p, q \in \mathbb{N}$. Pour obtenir la *minimalité* de l'écriture, on supposera que $\phi_p \neq 0$ et que $\theta_q \neq 0$ (quitte à changer p en $p - 1$ et q en $q - 1$). On considère aussi les polynômes $\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ et $\Theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$ qui leur sont associés dans \mathbb{C} .

Définition 2.6 *On dit qu'une série chronologique $(X_t)_{t \in \mathbb{Z}}$ est un processus autorégressif moyenne mobile d'ordres (p, q) si elle est définie par*

$$\forall t \in \mathbb{Z}, \quad \Phi(B)(X_t - m) = \Theta(B)\varepsilon_t$$

où $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc de variance $\sigma^2 > 0$ et $m \in \mathbb{R}$. On note $(X_t) \sim \text{ARMA}(p, q)$. Si $p = 0$, il s'agit d'un processus moyenne mobile d'ordre q défini par

$$\forall t \in \mathbb{Z}, \quad X_t - m = \Theta(B)\varepsilon_t$$

que l'on note $(X_t) \sim \text{MA}(q)$. Si au contraire $q = 0$, il s'agit d'un processus autorégressif d'ordre p défini par

$$\forall t \in \mathbb{Z}, \quad \Phi(B)(X_t - m) = \varepsilon_t$$

que l'on note $(X_t) \sim \text{AR}(p)$. La notation ARMA vient de l'anglais "AutoRegressive Moving Average".

Exemples. Quelques exemples :



On démontrera dans la Sec. 3.2 (Prop. 3.2) la proposition suivante.

Proposition 2.3 *Si le polynôme autorégressif Φ est tel que $\Phi(z) \neq 0$ pour tout $|z| = 1$, alors le processus $(X_t)_{t \in \mathbb{Z}}$ engendré par la relation ARMA(p, q) est stationnaire.*

En particulier on voit qu'un MA(q) est toujours stationnaire : $\Phi(z) = 1$, le polynôme n'a aucune racine.

2.5 Autocovariances et autocorrélations

Définition 2.7 *Soit $(X_t)_{t \in \mathbb{Z}}$ une série chronologique stationnaire (de variance non nulle). Sa fonction d'autocovariance γ est définie par*

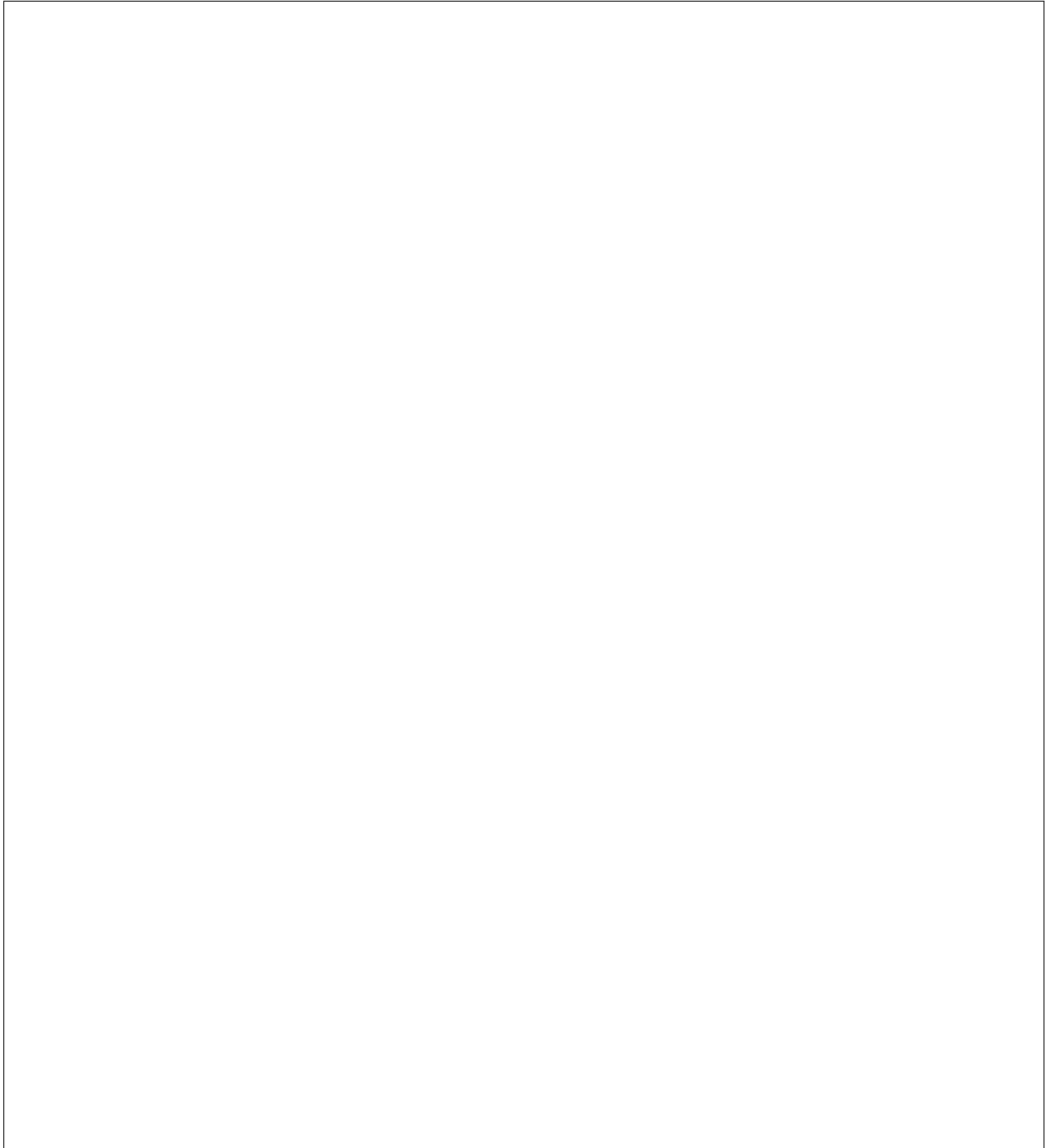
$$\forall h \in \mathbb{Z}, \quad \gamma(h) = \Gamma(h, 0) = \text{Cov}(X_h, X_0).$$

Sa fonction d'autocorrélation ρ est définie par

$$\forall h \in \mathbb{Z}, \quad \rho(h) = \text{Corr}(X_h, X_0) = \frac{\gamma(h)}{\gamma(0)}.$$

On l'appelle généralement *ACF*.

Exemples. Construire les fonctions d'autocovariance et d'autocorrélation du bruit blanc de variance $\sigma^2 > 0$. Faire de même avec les processus stationnaires des exemples précédents.



Proposition 2.4 *La fonction d'autocovariance γ d'un processus stationnaire satisfait*

$$\forall h \in \mathbb{Z}, \quad |\gamma(h)| \leq \gamma(0) \quad \text{et} \quad \gamma(h) = \gamma(-h).$$

Elle est de plus semi-définie positive dans le sens où

$$\forall n \geq 1, \forall (c_1, \dots, c_n) \in \mathbb{R}^n, \forall (t_1, \dots, t_n) \in \mathbb{Z}^n, \quad \sum_{i=1}^n \sum_{j=1}^n c_i \gamma(t_i - t_j) c_j \geq 0.$$

Cela implique en particulier que la matrice des autocovariances d'ordre n définie par

$$\Gamma_n = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{pmatrix}$$

est semi-définie positive.

Preuve.



On peut également montrer le résultat suivant, souvent bien pratique comme on le verra par la suite pour calculer les premiers termes d'une PACF.

Proposition 2.5 *Si $\gamma(0) > 0$ et que $\gamma(h) \rightarrow 0$ lorsque $|h| \rightarrow +\infty$, alors la matrice Γ_n définie ci-dessus est définie positive pour tout $n \geq 1$.*

En résumé, un processus stationnaire possède deux caractéristiques importantes : son espérance constante et sa fonction d'autocovariance. Sur les trajectoires observables (X_1, \dots, X_n) de tels processus, on les estime généralement par

$$\hat{m} = \frac{1}{n} \sum_{t=1}^n X_t \quad \text{et} \quad \hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (X_t - \hat{m})(X_{t+|h|} - \hat{m})$$

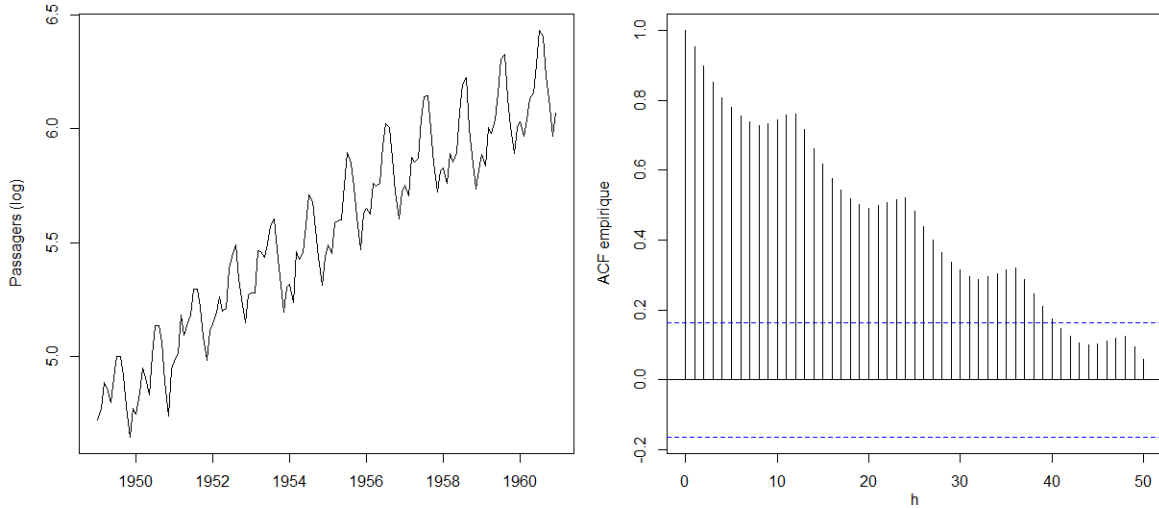
en posant $\hat{m} = 0$ si le processus est centré ($m = 0$). En fonction des données disponibles, le décalage est contraint par $0 \leq |h| \leq n - 1$. On en déduit aussi

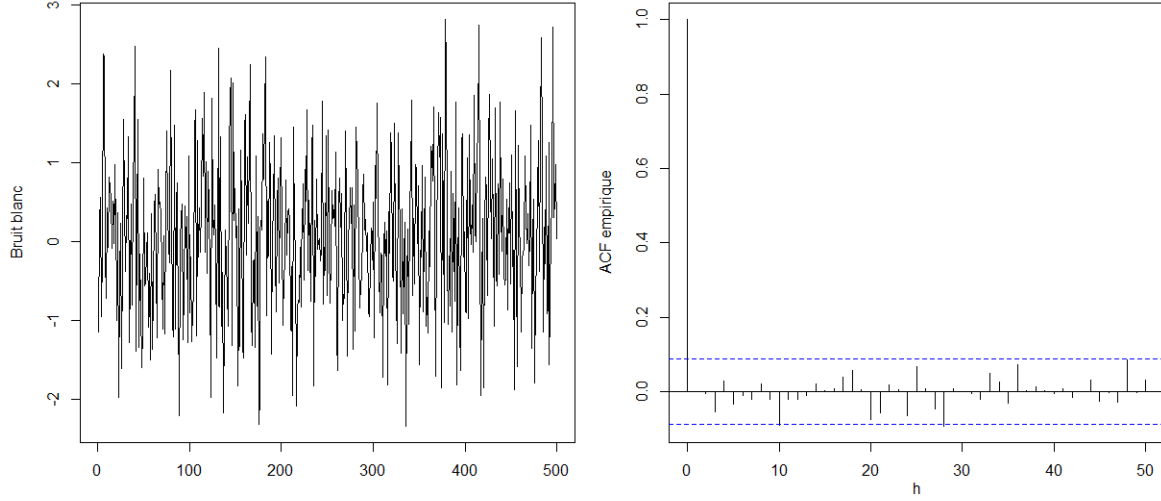
$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}.$$

L'autocorrélation empirique $\hat{\rho}$ (commande `acf`) contient donc beaucoup d'information quant au modèle de dépendance à proposer pour le processus, et à son éventuelle périodicité. De manière très intéressante, on peut montrer que, sous quelques hypothèses supplémentaires,

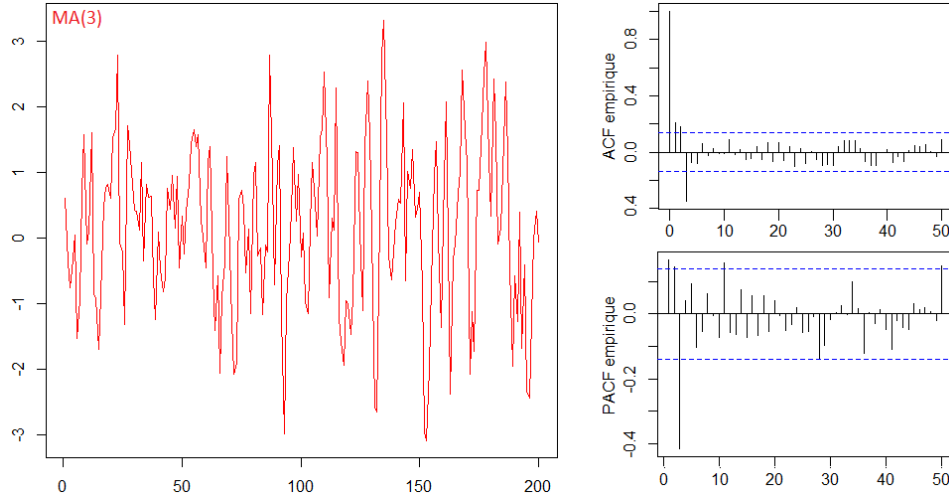
$$\forall \ell \geq 1, \quad \hat{\rho}(\ell) \xrightarrow{\text{p.s.}} \rho(\ell) \quad \text{et} \quad \sqrt{n} \begin{pmatrix} \hat{\rho}(1) - \rho(1) \\ \vdots \\ \hat{\rho}(\ell) - \rho(\ell) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, R_\ell).$$

On a donc accès à des intervalles de confiance et à des tests de $\mathcal{H}_0 : \rho(1) = \dots = \rho(\ell) = 0$ ou simplement de $\mathcal{H}_0 : \rho(h) = 0$ contre $\mathcal{H}_1 = \bar{\mathcal{H}}_0$. Le graphique issu de la commande `acf` représente l'intervalle de confiance sous \mathcal{H}_0 par des pointillés. Ci-dessous un exemple clairement non-stationnaire (les pointillés n'ont donc pas d'interprétation cohérente ici, c'est le phénomène périodique qui ressort), et un bruit blanc donc stationnaire (où cette fois les pointillés sont importants).





L'ACF est de plus un outil de caractérisation des MA comme on l'observe sur le schéma ci-dessous, et cela se justifie théoriquement par la proposition qui suivra.



Proposition 2.6 Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire et centré dont la fonction d'autocorrélation est telle que $\rho(h) = 0$ pour $|h| > q$ et $\rho(q) \neq 0$. Alors, il existe un bruit blanc tel que $(X_t) \sim MA(q)$.

En pratique, lorsque l'ACF empirique devient non significative à partir d'un rang $q+1$, il fait sens de proposer une modélisation $MA(q)$.

2.6 Innovations

Dans une modélisation chronologique, on souhaite expliquer X_t en fonction de son passé. Par exemple, si l'on se borne à un passé de n observations, on peut chercher dans L^2 la variable $\hat{X}_t = f(X_{t-1}, \dots, X_{t-n})$ qui minimise $\|X_t - Y\|^2 = \mathbb{E}[(X_t - Y)^2]$ par rapport à Y , soit

$$\hat{X}_t = \mathbb{E}[X_t | X_{t-1}, \dots, X_{t-n}].$$

Il s'agit de la meilleure approximation quadratique $\sigma(X_{t-1}, \dots, X_{t-n})$ -mesurable de Y et c'est une projection orthogonale sur le sous-espace de L^2 correspondant. Mais devant la difficulté voire l'impossibilité calculatoire

(on ne connaît pas nécessairement la loi du processus), on préfère chercher une expression de la forme

$$\hat{X}_t = \phi_0 + \sum_{k=1}^n \phi_k X_{t-k} = P_{\mathcal{H}_{t-1}^{t-n}} X_t$$

où

$$\forall n \geq 1, \quad \mathcal{H}_{t-1}^{t-n} = \text{Vect}\{1, X_{t-1}, \dots, X_{t-n}\}.$$

La suite $(\phi_k)_{0 \leq k \leq n}$ contient les coefficients de la projection orthogonale de X_t sur l'espace vectoriel engendré par ses n valeurs passées les plus proches et un terme de centrage (on se restreint donc aux combinaisons linéaires). Les coefficients sont déterminés grâce au système d'équations

$$\langle \hat{X}_t, 1 \rangle = \langle X_t, 1 \rangle \quad \text{et} \quad \forall j \in \{1, \dots, n\}, \quad \langle \hat{X}_t, X_{t-j} \rangle = \langle X_t, X_{t-j} \rangle.$$

Cette projection orthogonale offre de nombreux avantages, entre autres :

- Contrairement à $\mathbb{E}[X_t | X_{t-1}, \dots, X_{t-n}]$, on la calcule facilement.
- Elle ne fait apparaître que les moments d'ordre 1 et 2 du processus.
- Lorsque la série est gaussienne, $P_{\mathcal{H}_{t-1}^{t-n}} X_t = \mathbb{E}[X_t | X_{t-1}, \dots, X_{t-n}]$.

Remarque. Si le processus est centré, la projection se fait sur $\mathcal{H}_{t-1}^{t-n} = \text{Vect}\{X_{t-1}, \dots, X_{t-n}\}$.

Définition 2.8 Soit $(X_t)_{t \in \mathbb{Z}}$ un processus centré. On appelle passé linéaire de X_t l'espace

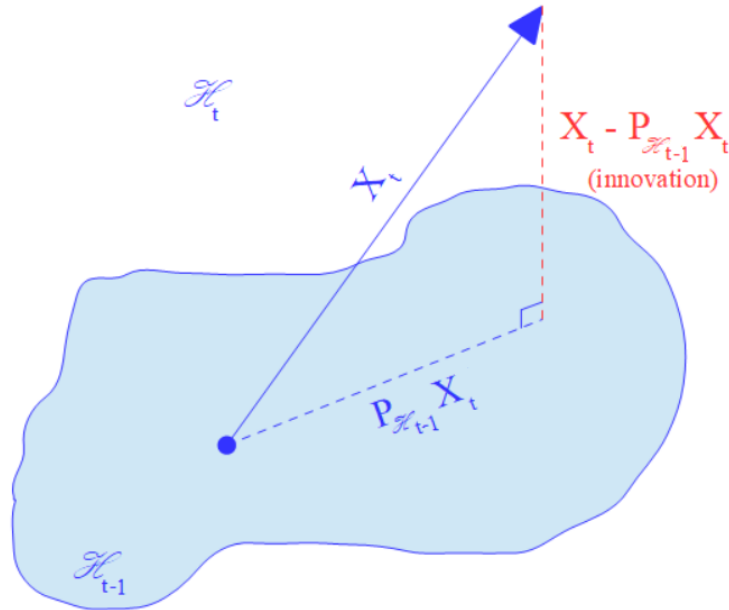
$$\mathcal{H}_t = \overline{\mathcal{H}_t^{-\infty}}$$

engendré par l'adhérence de toutes les combinaisons linéaires des valeurs passées du processus à l'instant t . Son enveloppe linéaire et son passé lointain sont alors définis par

$$\mathcal{H}_{+\infty} = \bigcup_{t \in \mathbb{Z}} \mathcal{H}_t = \overline{\mathcal{H}_{+\infty}^{-\infty}} \quad \text{et} \quad \mathcal{H}_{-\infty} = \bigcap_{t \in \mathbb{Z}} \mathcal{H}_t = \overline{\mathcal{H}_{-\infty}^{-\infty}}.$$

Si le processus n'est pas centré, on ajoute la composante constante aux espaces \mathcal{H}_t (mais, quitte à les recentrer par translation, on peut facilement se restreindre à l'étude des processus centrés). La suite $(\mathcal{H}_t)_{t \in \mathbb{Z}}$ est croissante dans le sens où $\mathcal{H}_{-\infty} \subset \dots \subset \mathcal{H}_{t-1} \subset \mathcal{H}_t \subset \mathcal{H}_{t+1} \subset \dots \subset \mathcal{H}_{+\infty}$. Sans entrer dans les détails, on note simplement que

$$P_{\mathcal{H}_{t-1}} X_t = \lim_{n \rightarrow +\infty} P_{\mathcal{H}_{t-1}^{t-n}} X_t \quad \text{dans } L^2.$$



Définition 2.9 On appelle *innovations* du processus $(X_t)_{t \in \mathbb{Z}}$ le processus défini par

$$\forall t \in \mathbb{Z}, \quad \varepsilon_t = X_t - P_{\mathcal{H}_{t-1}} X_t.$$

On peut interpréter l'*innovation* à l'instant t comme l'erreur commise lorsqu'on explique X_t par la meilleure combinaison linéaire de son passé, au sens de la distance dans L^2 . Clairement, $P_{\mathcal{H}_{t-1}} X_t \in \mathcal{H}_{t-1}$ et donc $\varepsilon_t = (I - P_{\mathcal{H}_{t-1}}) X_t \in \mathcal{H}_{t-1}^\perp$, d'où $\mathcal{H}_t = \mathcal{H}_{t-1} \oplus \text{Vect}\{\varepsilon_t\}$. L'innovation à l'instant t est l'aléa imprévisible permettant de passer de \mathcal{H}_{t-1} à \mathcal{H}_t .

Proposition 2.7 Si le processus $(X_t)_{t \in \mathbb{Z}}$ est stationnaire, alors $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc.

Preuve.

Remarque. Dans le cadre de la proposition précédente, si $\sigma^2 = \mathbb{E}[(X_t - P_{\mathcal{H}_{t-1}} X_t)^2]$ est la variance des innovations, alors $\sigma^2 = 0$ signifie que (p.s.) $X_t = P_{\mathcal{H}_{t-1}} X_t$ et donc que $\mathcal{H}_t \subset \mathcal{H}_{t-1}$ pour tout $t \in \mathbb{Z}$. Comme par définition $\mathcal{H}_{t-1} \subset \mathcal{H}_t$, on a $\mathcal{H}_{t-1} = \mathcal{H}_t$. Par suite, $\mathcal{H}_{-\infty} = \mathcal{H}_{+\infty}$ et le processus est déterministe dans le sens où toute son information était déjà là au départ.

2.7 Autocorrélations partielles

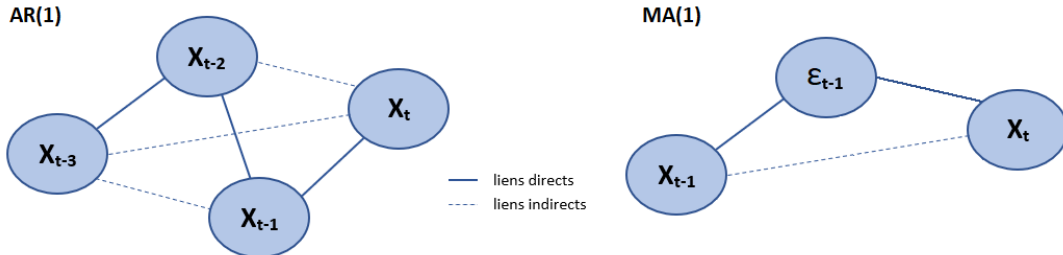
À partir de la projection orthogonale dans L^2 , on peut construire un autre type de corrélation. On rappelle que, dans le cas d'un processus centré, on a choisi de noter $\mathcal{H}_t^s = \text{Vect}\{X_t, \dots, X_s\}$.

Définition 2.10 Soit $(X_t)_{t \in \mathbb{Z}}$ une série chronologique stationnaire et centrée. Sa fonction d'autocorrélation partielle α est définie par

$$\alpha(1) = \rho(1) \quad \text{et} \quad \forall h \geq 2, \quad \alpha(h) = \text{Corr}(X_h - P_{\mathcal{H}_{h-1}^1} X_h, X_0 - P_{\mathcal{H}_{h-1}^1} X_0).$$

On l'appelle généralement *PACF*.

La PACF n'est pas aussi évidente à interpréter que l'ACF, on peut la voir comme la *corrélation pure* entre deux valeurs distantes de h dans le temps (c'est-à-dire entre lesquelles on a supprimé l'influence linéaire des valeurs intermédiaires). Dans ce schéma, explicatif à défaut d'être rigoureux, on représente des *liens indirects* (corrélations) et des *liens directs* (corrélations partielles). On y voit l'utilité de la PACF pour détecter les structures autorégressives ($\dots \rightarrow X_{t-2} \rightarrow X_{t-1} \rightarrow X_t \rightarrow \dots$). En effet, dans un AR(1) il existe une corrélation entre X_t et X_{t-2} ($X_t = \phi X_{t-1} + \varepsilon_t = \phi^2 X_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t$) mais on devine que X_{t-2} n'est pas directement liée à X_t , la jonction étant faite par X_{t-1} . Dans un MA(1) au contraire, X_t et X_{t-1} sont corrélées mais pas directement, c'est ε_{t-1} qui fait le lien ($\dots \leftarrow \varepsilon_{t-2} \rightarrow X_{t-1} \leftarrow \varepsilon_{t-1} \rightarrow X_t \leftarrow \dots$).



Elle est très difficile à construire selon sa définition, le résultat ci-dessous en propose une autre interprétation donnant lieu à une règle de calcul.

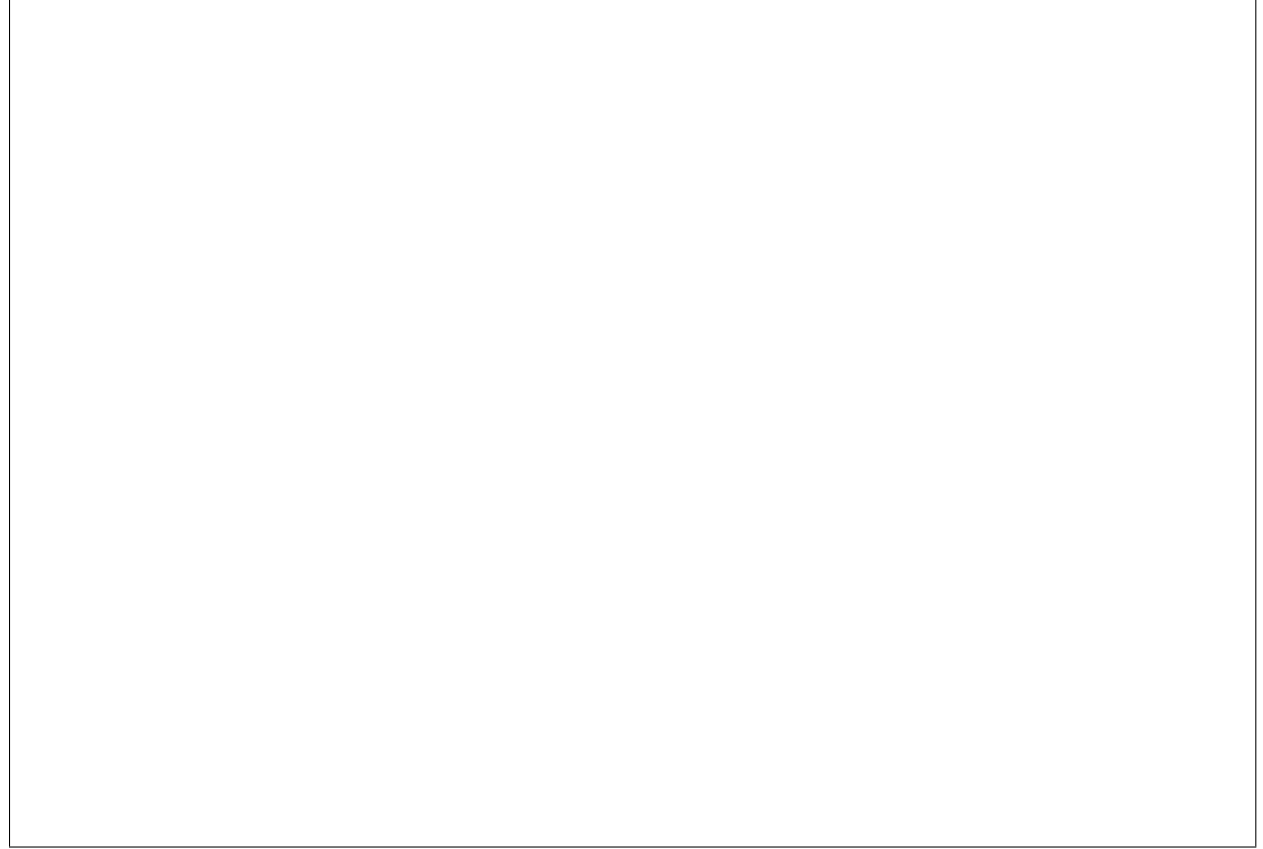
Proposition 2.8 *Soit $(X_t)_{t \in \mathbb{Z}}$ une série chronologique stationnaire et centrée. Alors,*

$$\forall h \geq 1, \quad \alpha(h) = \phi_h$$

où ϕ_1, \dots, ϕ_h désignent les coefficients de la projection orthogonale de X_h sur \mathcal{H}_{h-1}^0 ,

$$P_{\mathcal{H}_{h-1}^0} X_h = \phi_1 X_{h-1} + \dots + \phi_h X_0.$$

Preuve.



Par projection, on sait que $(I - P_{\mathcal{H}_{h-1}^0}) X_h \perp \mathcal{H}_{h-1}^0$ ce qui donne, axe par axe,

$$\begin{aligned} \forall k \in \{0, \dots, h-1\}, \quad \langle X_h - P_{\mathcal{H}_{h-1}^0} X_h, X_k \rangle &= \mathbb{E}[X_h X_k] - \sum_{j=1}^h \phi_j \mathbb{E}[X_{h-j} X_k] \\ &= \gamma(h-k) - \sum_{j=1}^h \phi_j \gamma(h-j-k) = 0. \end{aligned}$$

Ainsi on peut calculer le vecteur (ϕ_1, \dots, ϕ_h) en résolvant un système linéaire à h inconnues qui dépend de l'autocovariance γ . Explicitement,

$$\underbrace{\begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(h-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(h-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(h-1) & \gamma(h-2) & \cdots & \gamma(0) \end{pmatrix}}_{\Gamma_h} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_h \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(h) \end{pmatrix}.$$

Si $\gamma(0) > 0$ et $\gamma(h) \rightarrow 0$ lorsque $|h| \rightarrow +\infty$, on a vu que Γ_h était inversible (Prop. 2.5). On peut alors appliquer la méthode pour calculer $\alpha(h) = \phi_h$. Par exemple en utilisant la règle de Cramer,

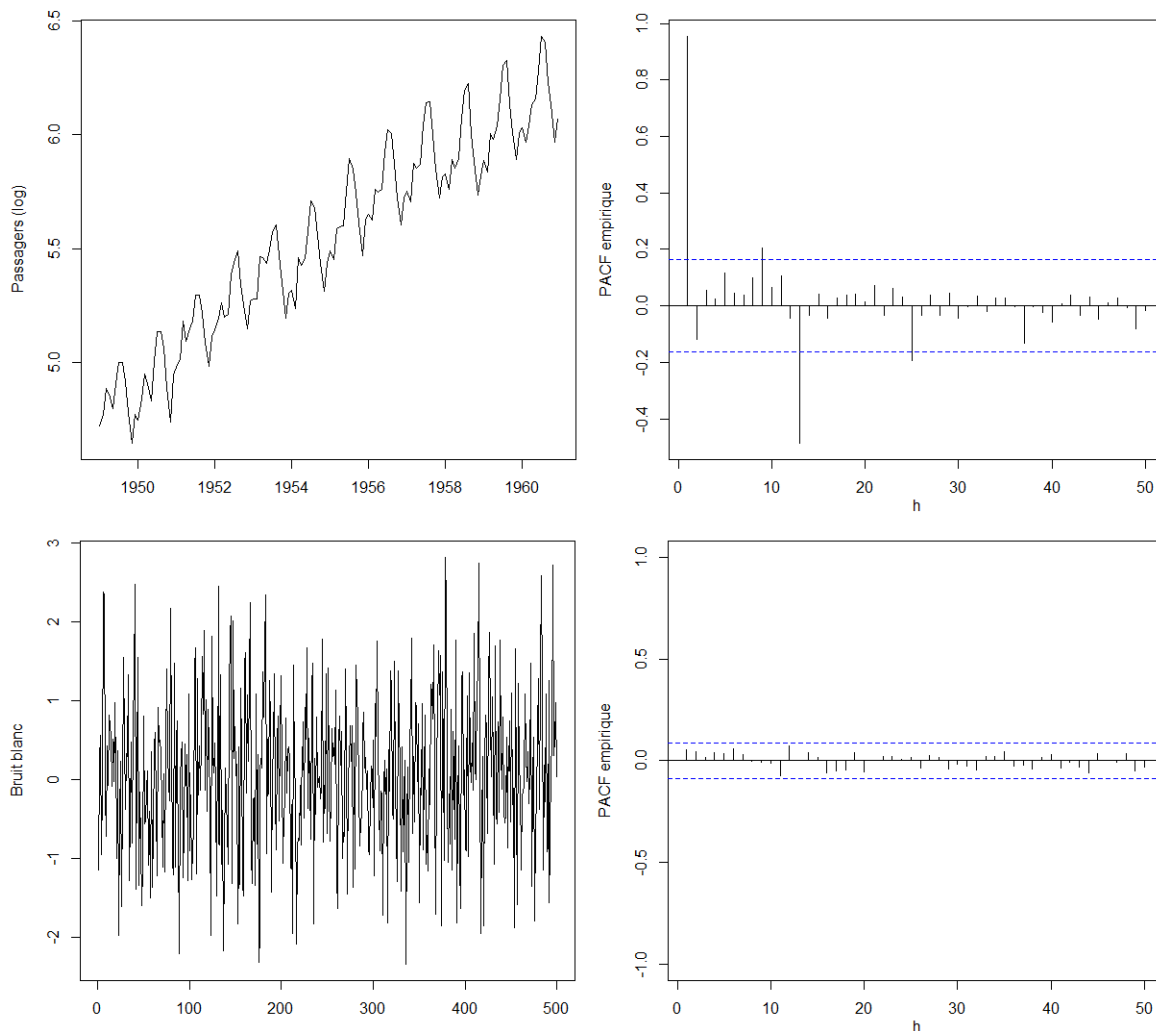
$$\alpha(h) = \frac{\begin{vmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(h-2) & \gamma(1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(h-3) & \gamma(2) \\ \vdots & \vdots & & \vdots & \vdots \\ \gamma(h-1) & \gamma(h-2) & \cdots & \gamma(1) & \gamma(h) \end{vmatrix}}{|\Gamma_h|}.$$

L'algorithme de *Durbin-Levinson* permet aussi d'effectuer itérativement le calcul de la PACF à partir de l'ACF.

Exemple. Construire la fonction d'autocorrélation partielle du bruit blanc de variance $\sigma^2 > 0$.



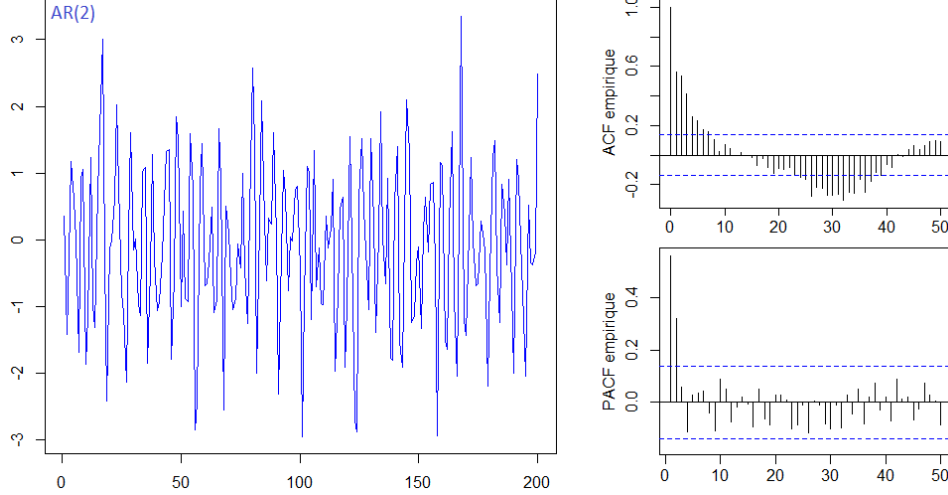
Sur les trajectoires observables (X_1, \dots, X_n) de tels processus, on peut estimer α en récupérant dans $\hat{\alpha}(h)$, pour $1 \leq h \leq n-1$, le dernier coefficient de la régression linéaire de X_n sur X_{n-1}, \dots, X_{n-h} (commande `pacf`). On peut aussi appliquer l'algorithme de Durbin-Levinson en y injectant $\hat{\rho}$ en lieu et place de ρ .



Ce sont aussi des estimateurs consistants et asymptotiquement normaux,

$$\forall \ell \geq 1, \quad \hat{\alpha}(\ell) \xrightarrow{\text{p.s.}} \alpha(\ell) \quad \text{et} \quad \sqrt{n} \begin{pmatrix} \hat{\alpha}(1) - \alpha(1) \\ \vdots \\ \hat{\alpha}(\ell) - \alpha(\ell) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, A_\ell).$$

Ils donnent donc accès à des tests de $\mathcal{H}_0 : \alpha(1) = \dots = \alpha(\ell) = 0$ ou simplement de $\mathcal{H}_0 : \alpha(h) = 0$ contre $\mathcal{H}_1 = \mathcal{H}_0^c$. Le graphique issu de la commande `pacf` représente l'intervalle de confiance sous \mathcal{H}_0 par des pointillés. Comme on l'a schématiquement évoqué avec les liens directs et indirects, la PACF est un outil de caractérisation des AR. On l'observe clairement sur le schéma ci-dessous, et cela se justifie théoriquement par la proposition qui suivra.



Proposition 2.9 Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stationnaire et centré dont la fonction d'autocorrélation partielle est telle que $\alpha(h) = 0$ pour $h > p$ et $\alpha(p) \neq 0$. Alors, il existe un bruit blanc tel que $(X_t) \sim \text{AR}(p)$.

En pratique, lorsque la PACF empirique devient non significative à partir d'un rang $p + 1$, il fait sens de proposer une modélisation $\text{AR}(p)$.

3 Processus ARMA : théorie et pratique

L'objectif est désormais d'approfondir la théorie des processus ARMA et de survoler leur mise en pratique. On peut montrer (ce résultat s'appelle la *décomposition de Wold*) que tous les processus stationnaires peuvent se mettre sous la forme linéaire

$$X_t - m = \sum_{k=0}^{+\infty} \psi_k \varepsilon_{t-k}$$

où $(\varepsilon_t)_{t \in \mathbb{Z}}$ est le bruit blanc d'innovation de $(X_t)_{t \in \mathbb{Z}}$ et $\mathbb{E}[X_t] = m$. Les processus ARMA couvrent une sous-partie de ces représentations : celles dont les coefficients (ψ_k) sont issus du développement en série entière d'une fraction rationnelle, c'est-à-dire pour lesquels il existe des polynômes tels que $\Theta(z)/\Phi(z) = \sum_k \psi_k z^k$. Les coefficients de Θ et Φ , en nombre fini et souvent réduit, sont suffisants pour générer la suite infinie (ψ_k) .

3.1 Opérateur autorégressif

Ce qui suit est la clé pour comprendre les principales propriétés des ARMA. On considère l'opérateur $I - \lambda B$ avec $\lambda \in \mathbb{C}$ pour lequel on cherche un opérateur inverse prenant la forme

$$\Psi(B) = \sum_{k \in \mathbb{Z}} \psi_k B^k \quad \text{avec} \quad \sum_{k \in \mathbb{Z}} |\psi_k| < +\infty.$$

Le fait de chercher une forme linéaire pour opérateur inverse provient de l'analogie avec le développement en série entière des fractions rationnelles. En effet, on sait que

$$\frac{1}{1 - \lambda z} = \sum_{k=0}^{+\infty} (\lambda z)^k \quad \text{pour tout } |\lambda z| < 1$$

alors que

$$\frac{1}{1-\lambda z} = -\frac{1}{\lambda z} \frac{1}{1-\frac{1}{\lambda z}} = -\frac{1}{\lambda z} \sum_{k=0}^{+\infty} (\lambda z)^{-k} = -\sum_{k=1}^{+\infty} (\lambda z)^{-k} \quad \text{pour tout } |\lambda z| > 1.$$

On retrouve cette analogie dans le résultat ci-dessous.

Proposition 3.1 *Pour $|\lambda| < 1$, on a*

$$\Psi(B) = (I - \lambda B)^{-1} = \sum_{k=0}^{+\infty} \lambda^k B^k \quad \text{avec} \quad \sum_{k=0}^{+\infty} |\lambda^k| < +\infty.$$

Pour $|\lambda| > 1$, on a

$$\Psi(B) = (I - \lambda B)^{-1} = -\sum_{k=1}^{+\infty} \lambda^{-k} B^{-k} \quad \text{avec} \quad \sum_{k=1}^{+\infty} |\lambda^{-k}| < +\infty.$$

Pour $|\lambda| = 1$, l'opérateur $I - \lambda B$ n'est pas inversible.

Preuve.

Exemple. Comme application immédiate de l'opérateur $I - \lambda B$, on considère le processus (X_t) défini par

$$X_t = \lambda X_{t-1} + \varepsilon_t$$

où (ε_t) est un bruit blanc de variance $\sigma^2 > 0$ et $\lambda \in \mathbb{R}$. Étudier la stationnarité du processus selon les valeurs de λ , et donner le cas échéant sa fonction d'autocorrélation.



Soit maintenant l'opérateur autorégressif

$$M(B) = I + \phi_1 B + \dots + \phi_n B^n.$$

Le polynôme $M(z) = 1 + \phi_1 z + \dots + \phi_n z^n$ satisfait $M(0) = 1$, il possède donc n racines $\mu_1, \dots, \mu_n \in \mathbb{C}^*$ et l'opérateur admet la factorisation (on se rappelle que $B^k \circ B^{k'} = B^{k+k'}$)

$$M(B) = \prod_{i=1}^n (I - \mu_i^{-1} B).$$

De l'ensemble $R = \{\mu_1, \dots, \mu_n\}$ des racines de M , on fait trois sous-ensembles :

$$\rightarrow R_c = \{\mu_i, |\mu_i| > 1\}.$$

$$\rightarrow R_a = \{\mu_i, |\mu_i| < 1\}.$$

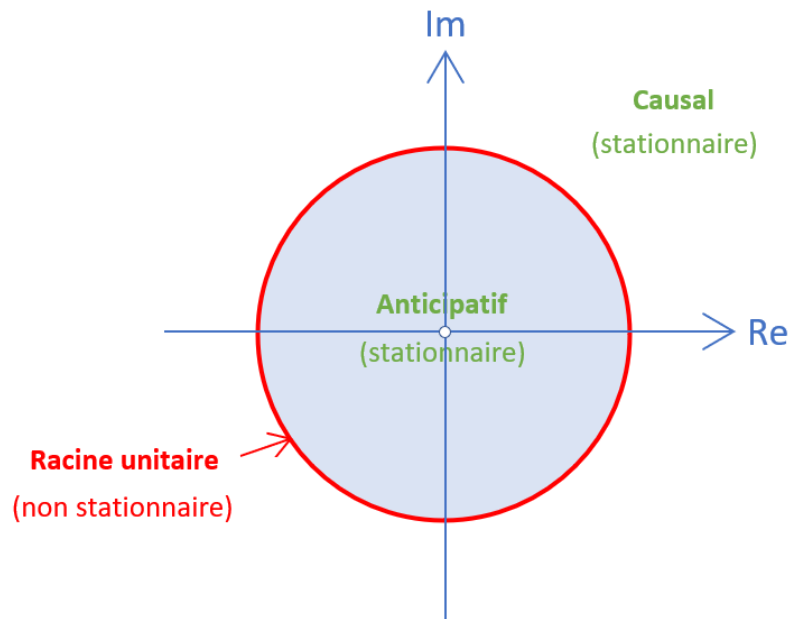
$$\rightarrow R_u = \{\mu_i, |\mu_i| = 1\}.$$

On a donc $R_c \cup R_a \cup R_u = R$, soit

$$M(B) = \prod_{R_c} (I - \mu_i^{-1} B) \prod_{R_a} (I - \mu_i^{-1} B) \prod_{R_u} (I - \mu_i^{-1} B)$$

avec la convention que $\prod_{\emptyset} = I$. Ainsi dès que $R_u \neq \emptyset$ (et donc qu'il existe une *racine unitaire*), la proposition précédente nous dit que $M(B)$ n'est pas inversible. Si au contraire $R_u = \emptyset$, alors $M(B)$ est inversible et l'on peut construire $M^{-1}(B)$ en inversant chacun des facteurs. Si de plus $R_a = \emptyset$, alors toutes les racines sont dans R_c , on a donc $|\mu_i| > 1$ pour tout i , d'où $|\mu_i^{-1}| < 1$: l'opérateur inverse est *causal* car il ne fait apparaître que le passé. Sinon, le futur apparaît et l'opérateur inverse est *anticipatif*. On fera le lien par la suite (voir la figure ci-dessous) entre la position des racines d'un polynôme autorégressif et la stationnarité du processus AR qui en découle.

Exemples. Étudier l'inversibilité des opérateurs $M(B) = I - \frac{B}{2}$, $N(B) = I - 2B$, $P(B) = M(B)N(B)$, $Q(B) = (I - B)M(B)$ et $R(B) = (I + B^2)N(B)$.



*Remarque.*¹ Lorsqu'on cherche à inverser $M(B) = (I - \lambda_1 B)(I - \lambda_2 B)$ avec $|\lambda_1| < 1$ et $|\lambda_2| < 1$ alors, comme dans ce cas l'opérateur inverse est causal, on peut écrire

$$M(B) \sum_{k=0}^{+\infty} \psi_k B^k = I.$$

Par identification,

$$\begin{cases} \psi_0 = 1 \\ \psi_1 - (\lambda_1 + \lambda_2) \psi_0 = 0 \\ \psi_k - (\lambda_1 + \lambda_2) \psi_{k-1} + \lambda_1 \lambda_2 \psi_{k-2} = 0 \end{cases} \quad \text{pour } k \geq 2.$$

1. Elle est importante celle-là... Chaque année je demande un tel opérateur à inverser et chaque année personne ne pense à utiliser cette formule toute simple...

Ainsi, on peut retrouver les coefficients (ψ_k) en résolvant une équation de récurrence linéaire d'ordre 2. Dans ce cas précis, on obtient facilement

$$\psi_k = \frac{\lambda_1^{k+1} - \lambda_2^{k+1}}{\lambda_1 - \lambda_2} \quad \text{ou} \quad \psi_k = (1+k) \lambda_0^k$$

selon que $\lambda_1 \neq \lambda_2$ ou que $\lambda_1 = \lambda_2 = \lambda_0$.

3.2 Autocovariances et autocorrélations

Rappelons (voir Déf. 2.6) qu'un processus ARMA(p, q) s'écrit

$$\forall t \in \mathbb{Z}, \quad \Phi(B) (X_t - m) = \Theta(B) \varepsilon_t$$

où $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un bruit blanc de variance $\sigma^2 > 0$, $m \in \mathbb{R}$, et où les opérateurs sont définis par

$$\Phi(B) = I - \phi_1 B - \dots - \phi_p B^p \quad \text{et} \quad \Theta(B) = I + \theta_1 B + \dots + \theta_q B^q$$

avec $\phi_p \neq 0$ et $\theta_q \neq 0$. Grâce au résultat suivant, on prouve et on enrichit la Prop. 2.3 énoncée dans la Sec. 2.4 lors de la présentation des ARMA.

Proposition 3.2 *Si le polynôme autorégressif Φ est tel que $\Phi(z) \neq 0$ pour tout $|z| = 1$, alors le processus $(X_t)_{t \in \mathbb{Z}}$ engendré par la relation ARMA(p, q) est stationnaire de moyenne m et de fonction d'autocorrélation donnée par*

$$\forall h \in \mathbb{Z}, \quad \rho(h) = \frac{\sum_{k \in \mathbb{Z}} \psi_{k+|h|} \psi_k}{\sum_{k \in \mathbb{Z}} \psi_k^2}$$

où $(\psi_k) \in \ell^1(\mathbb{Z})$ est la suite des coefficients de l'opérateur linéaire $\Phi^{-1}(B) \Theta(B)$.

Preuve.

Corollaire 3.1 *Si le polynôme autorégressif Φ est tel que $\Phi(z) \neq 0$ pour tout $|z| = 1$, alors le processus $(X_t)_{t \in \mathbb{Z}}$ engendré par la relation AR(p) est stationnaire de moyenne m et de fonction d'autocorrélation donnée par*

$$\forall h \in \mathbb{Z}, \quad \rho(h) = \frac{\sum_{k \in \mathbb{Z}} \xi_{k+|h|} \xi_k}{\sum_{k \in \mathbb{Z}} \xi_k^2}$$

où $(\xi_k) \in \ell^1(\mathbb{Z})$ est la suite des coefficients de l'opérateur linéaire $\Phi^{-1}(B)$.

Corollaire 3.2 *Le processus $(X_t)_{t \in \mathbb{Z}}$ vérifiant le modèle MA(q) est stationnaire de moyenne m et de fonction d'autocorrélation donnée par*

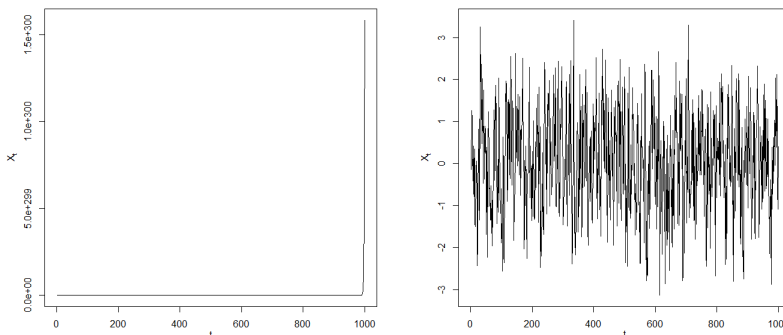
$$\forall h \in \mathbb{Z}, \quad \rho(h) = \begin{cases} \frac{\theta_{|h|} + \sum_{k=1}^{q-|h|} \theta_{k+|h|} \theta_k}{1 + \sum_{k=1}^q \theta_k^2} & \text{si } |h| \leq q \\ 0 & \text{si } |h| > q \end{cases}$$

avec les conventions $\Sigma_{\emptyset} = 0$ et $\theta_0 = 1$.

3.3 Portée pratique des racines anticipatives

Si l'on en croit la Prop. 3.2, le processus $X_t = 2X_{t-1} + \varepsilon_t$ serait donc stationnaire. Voici un bout de code R simulant un tel processus et le graphique obtenu (essayez!), et à côté le même en remplaçant 2 par 1/2.

```
> n = 1000; E = rnorm(n); X = rep(0, n)
> for (i in 2:n){ X[i] = 2*X[i-1] + E[i] }
```

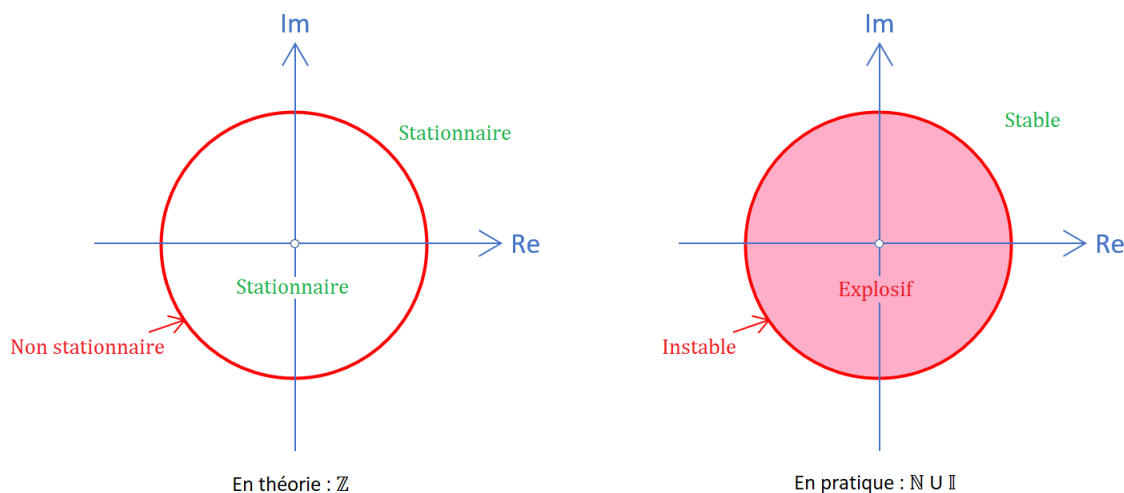


Ces deux processus sont-ils vraiment de même nature? En fait tout repose sur l'espace des temps. Le raisonnement développé suppose $T = \mathbb{Z}$ mais pour $T = \mathbb{N} \cup \mathbb{I}$ ($\mathbb{I} = \{0\}$ dans l'exemple), les conclusions sont différentes. Poser un AR(1) sur une série observée à partir d'un instant initial revient à dire que

$$\forall n \geq 1, \quad X_n = \lambda X_{n-1} + \varepsilon_n = \lambda^n X_0 + \sum_{t=0}^{n-1} \lambda^t \varepsilon_{n-t}.$$

Si l'on fait l'hypothèse que la valeur initiale $X_0 \in L^2$ est décorrélée du bruit $(\varepsilon_n)_{n \geq 1}$, comme si la série débutait réellement à l'instant 0 et était ensuite engendrée par la relation de récurrence et des sauts aléatoires décorrélés du passé, alors (on a déjà fait ce calcul dans un exemple du chapitre précédent)

$$\mathbb{E}[X_n] = \lambda^n \mathbb{E}[X_0] \quad \text{et} \quad \text{Cov}(X_n, X_m) = \begin{cases} \lambda^{n+m} \mathbb{V}(X_0) + \sigma^2 \lambda^{|n-m|} \frac{1-\lambda^{2(n \wedge m)}}{1-\lambda^2} & \text{si } |\lambda| \neq 1 \\ \mathbb{V}(X_0) + \sigma^2 \lambda^{|n-m|} (n \wedge m) & \text{si } |\lambda| = 1. \end{cases}$$



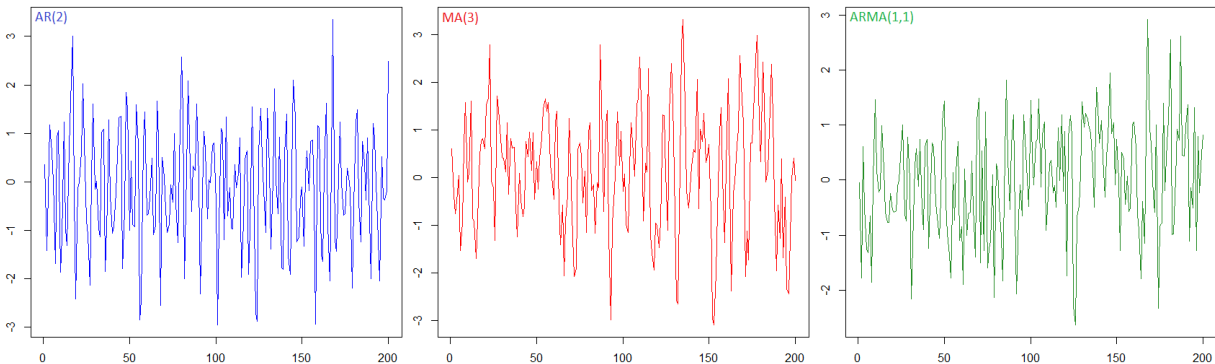
On voit (c'est un bon exercice...) que, pour que le processus soit stationnaire, il faut et il suffit que

$$\mathbb{E}[X_0] = 0 \quad \text{et que} \quad \mathbb{V}(X_0) = \frac{\sigma^2}{1 - \lambda^2}$$

et cela n'est possible que si $|\lambda| < 1$ (si $X_0 \in L^2$ on a seulement la stationnarité asymptotique). Ainsi, dans ce cas de figure, la stationnarité du processus est liée à la condition $|\lambda| < 1$ (cadre causal) alors que la condition $|\lambda| \geq 1$ (cadres unitaire et anticipatif) engendre un comportement explosif (au moins en variance). Ce raisonnement s'étend facilement aux ARMA et en pratique on travaille plutôt avec le vocabulaire du schéma ci-dessus (pour $1/\lambda \dots$ n'oublions pas qu'on travaille avec la racine de $1 - \lambda z$). On retiendra donc que l'étude sur $\mathbb{N} \cup \mathbb{I}$ n'est un cas particulier de l'étude sur \mathbb{Z} que dans le cadre causal ! Sinon les facilités théoriques n'ont aucune portée pratique. Et en statistique, la portée pratique c'est plutôt important...

3.4 Caractérisation

Il n'est pas possible de faire la différence visuellement entre les AR, les MA et les ARMA, comme le montrent les simulations ci-dessous. Les critères d'identification pour les AR et les MA sont très utiles à cet égard.



On a vu (Sec. 2.5 et 2.7) que l'ACF et la PACF possèdent des propriétés remarquables lorsque $(X_t)_{t \in \mathbb{Z}}$ est un processus $MA(q)$ ou $AR(p)$, et donc que l'étude de leurs corrélations permet de guider le choix d'une modélisation. Nous avons maintenant les outils pour détailler un peu mieux ces résultats :

- Rappel de la Prop. 2.6. Si un processus stationnaire et centré est tel que sa fonction d'autocorrélation vérifie $\rho(h) = 0$ pour $|h| > q$ et $\rho(q) \neq 0$, alors il se comporte comme un $MA(q)$.
- Rappel de la Prop. 2.9. Si un processus stationnaire et centré est tel que sa fonction d'autocorrélation partielle vérifie $\alpha(h) = 0$ pour $h > p$ et $\alpha(p) \neq 0$, alors il se comporte comme un $AR(p)$.

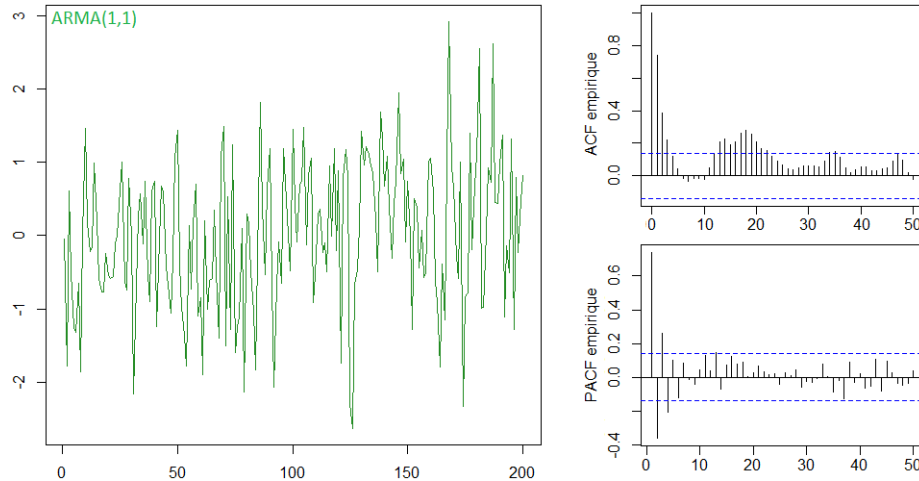
Preuve (partielle).



Hélas il n'existe pas de comportement remarquable pour les $\text{ARMA}(p, q)$ qui ne sont ni AR ni MA purs. On peut montrer par un calcul fastidieux que

$$\forall h \geq \max(0, q + 1 - p), \quad \gamma(h) = \sum_{i=1}^s \sum_{j=0}^{m_i-1} \beta_{ij} h^j \mu_i^{-h}$$

où β_{ij} sont des constantes identifiables, s est le nombre de racines distinctes μ_i de Φ et m_i est la multiplicité de μ_i . Lorsque $\Phi(z) \neq 0$ pour $|z| \leq 1$, on voit que $|\gamma(h)|$ décroît très rapidement vers 0, géométriquement vite. On peut donc justifier une tentative de modélisation ARMA lorsque la série est considérée comme stationnaire et que ses ACF et PACF empiriques montrent une décroissance rapide (en module).



Concernant les ordres p et q , en général on teste une batterie de modèles afin d'optimiser un certain critère. On peut également se fier à la significativité des estimateurs, à la *méthode du coin*, à la *blancheur* des résidus, etc. C'est finalement un choix très empirique.

3.5 Écriture canonique

Face à toutes ces subtilités (racines causales, unitaires, anticipatives, dans Φ ou dans Θ), il existe une écriture d'un intérêt particulier pour l'interprétation : si l'on peut trouver des coefficients $(\psi_k) \in \ell^1(\mathbb{N})$ tels que

$$\forall t \in \mathbb{Z}, \quad X_t = m + \sum_{k=0}^{+\infty} \psi_k \varepsilon_{t-k}$$

où le bruit blanc $(\varepsilon_t)_{t \in \mathbb{Z}}$ est le processus des innovations de $(X_t)_{t \in \mathbb{Z}}$, alors on parvient à décomposer le processus en une composante déterministe (sa moyenne m) et la somme de toutes ses innovations passées.

Définition 3.1 *Le processus $(X_t)_{t \in \mathbb{Z}}$ vérifiant l'équation $ARMA(p, q)$ est mis sous forme canonique si ses paramètres vérifient $\phi_p \neq 0$ et $\theta_q \neq 0$, si de plus Φ et Θ n'ont pas de racine commune, et si $\Phi(z) \neq 0$ et $\Theta(z) \neq 0$ pour tout $|z| \leq 1$.*

Exemples. Donner quelques exemples d'écritures canoniques et non canoniques.

Proposition 3.3 *Si le processus $(X_t)_{t \in \mathbb{Z}}$ vérifiant l'équation $ARMA(p, q)$ est sous forme canonique, alors la décomposition ci-dessus peut être obtenue.*

Preuve.



Un peu de vocabulaire :

- Lorsqu'un processus admet la décomposition (1), on dit qu'il est *causal*. On reconnaît par ailleurs une écriture $MA(\infty)$. La notion de *causalité* traduit le fait que l'on peut exprimer un processus en fonction du *présent* et du *passé* d'un autre.
- Lorsqu'un processus admet décomposition (2), on dit qu'il est *inversible*. On reconnaît par ailleurs une écriture $AR(\infty)$.
- Un processus ARMA sous forme canonique est donc à la fois causal et inversible.

Remarque. Puisqu'on cherche en quelque sorte une écriture *minimale* pour le modèle, les conditions $\phi_p \neq 0$ et $\theta_q \neq 0$ sont faciles à comprendre. À propos des éventuelles racines communes entre Φ et Θ : s'il existe $\lambda_0 \in \mathbb{C}^*$ tel que $\Phi(\lambda_0) = \Theta(\lambda_0) = 0$, alors on peut écrire

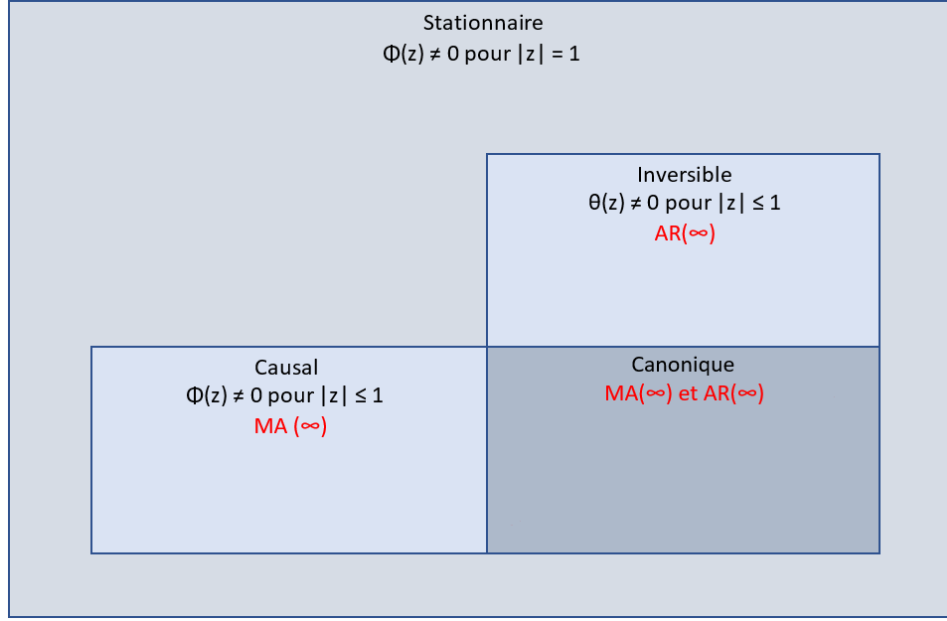
$$\Phi(B) = (I - \lambda_0^{-1}B) \Phi^*(B) \quad \text{et} \quad \Theta(B) = (I - \lambda_0^{-1}B) \Theta^*(B).$$

Si $|\lambda_0| \neq 1$, on a $(X_t) \sim \text{ARMA}(p, q)$ et

$$\begin{aligned} (I - \lambda_0^{-1}B) \Phi^*(B) X_t = (I - \lambda_0^{-1}B) \Theta^*(B) \varepsilon_t &\iff \Phi^*(B) X_t = (I - \lambda_0^{-1}B)^{-1} (I - \lambda_0^{-1}B) \Theta^*(B) \varepsilon_t \\ &\iff \Phi^*(B) X_t = \Theta^*(B) \varepsilon_t. \end{aligned}$$

On a donc en fait $(X_t) \sim \text{ARMA}(p-1, q-1)$ dès que, à leur tour, Φ^* et Θ^* n'ont pas de racine commune. Sinon, on applique le même raisonnement jusqu'à obtenir une écriture irréductible.

De manière très intéressante, on peut montrer que, si un processus ARMA stationnaire n'est pas sous forme canonique, on peut souvent s'y ramener par une transformation de l'écriture du modèle. Cela justifie en particulier qu'on peut se limiter aux formes canoniques dans les études théoriques. Résumons par un schéma (Φ et Θ n'ont pas de racine commune, avec $\phi_p \neq 0$ et $\theta_q \neq 0$).



3.6 Équations de Yule-Walker

Soit $(X_t)_{t \in \mathbb{Z}}$ un processus $AR(p)$ centré et causal pour simplifier (on peut avoir des résultats aussi dans le cadre des ARMA). On a

$$\forall i \in \mathbb{N}, \quad \mathbb{E}[X_t X_{t-i}] = \sum_{k=1}^p \phi_k \mathbb{E}[X_{t-k} X_{t-i}] + \mathbb{E}[\varepsilon_t X_{t-i}].$$

Si l'on choisit $i \geq 1$, il vient

$$\gamma(i) = \sum_{k=1}^p \phi_k \gamma(i-k) \quad \text{et donc} \quad \gamma(i) - \phi_1 \gamma(i-1) - \dots - \phi_p \gamma(i-p) = 0$$

car en effet par la causalité (et le fait qu'on ait un bruit blanc),

$$\mathbb{E}[\varepsilon_t X_{t-i}] = \sum_{k \geq 0} \psi_k \mathbb{E}[\varepsilon_t \varepsilon_{t-i-k}] = 0.$$

En faisant varier i dans $\{1, \dots, p\}$, on construit le système de p équations linéaires à p inconnues donné par

$$\begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \cdots & \gamma(0) \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(p) \end{pmatrix}.$$

Ces équations permettent de mettre en relation les coefficients autorégressifs et les autocorrélations. Avec $i = 0$, on a aussi

$$\mathbb{E}[X_t^2] = \sum_{k=1}^p \phi_k \mathbb{E}[X_{t-k} X_t] + \mathbb{E}[\varepsilon_t X_t]$$

et donc une $(p+1)$ -ème équation associée au $(p+1)$ -ème paramètre,

$$\gamma(0) = \phi_1 \gamma(1) + \dots + \phi_p \gamma(p) + \sigma^2.$$

Ces équations sur les moments d'ordre 2 sont classiques dans les modèles chronologiques, elles peuvent permettre d'obtenir très rapidement des estimateurs (car on se rappelle qu'on sait très bien estimer la fonction γ , voir la Sec. 2.5).

3.7 Estimation et propriétés des estimateurs

Supposons maintenant que l'on dispose d'une trajectoire (X_1, \dots, X_n) sur laquelle on veut poser un modèle ARMA de paramètres (ϕ, θ, σ^2) . Un estimateur explicite n'existe pas dans le cas général, c'est pourquoi on fait 'comme si' le bruit était gaussien, donc que $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. Sous toutes les hypothèses du modèle, le vecteur $\underline{X} = (X_1, \dots, X_n)$ est lui-même gaussien, de covariance Γ_n et de log-vraisemblance

$$\ell\ell_{\phi, \theta, \sigma^2}(\underline{X}) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln \det(\Gamma_n) - \frac{1}{2} \underline{X}^t \Gamma_n^{-1} \underline{X}.$$

Il est alors possible de construire un algorithme d'optimisation pour maximiser $\ell\ell_{\phi, \theta, \sigma^2}(\underline{X})$ par rapport à ses paramètres (ϕ, θ, σ^2) , cachés dans Γ_n , qui converge vers un EMV. L'estimateur obtenu possède toutes les propriétés utiles aux tests de significativité (consistant et asymptotiquement normal, passons les détails pour simplifier...). En résumé, il existe une covariance, qui s'estime bien et qui du coup donne des possibilités de tests. Un exemple :

```
> arima(X, order = c(3, 0, 3))
```

Coefficients:

	ar1	ar2	ar3	ma1	ma2	ma3	intercept
	0.4581	0.5174	-0.7140	0.0550	-0.7562	0.1887	0.0024
s.e.	0.1050	0.0879	0.7867	0.0019	0.0986	0.1309	0.0226

sigma^2 estimated as 0.1109: log likelihood = -33.37, aic = 82.74

Ici le ratio estimation/écart-type indique que à 95% (borne 1.96), le troisième coefficient de la partie AR et le troisième de la partie MA n'ont que très peu d'importance, et il en va de même de la constante du modèle (et ça tombe bien, j'ai simulé un ARMA(2, 2) centré). Dans le cas des AR purs la procédure peut être simplifiée. En utilisant les équations de Yule-Walker, il suffit de remplacer γ par $\hat{\gamma}$ dans les équations, car on a vu que ces estimateurs étaient consistants sous nos hypothèses (Sec. 2.5),

$$\begin{pmatrix} \tilde{\phi}_1 \\ \tilde{\phi}_2 \\ \vdots \\ \tilde{\phi}_p \end{pmatrix} = \begin{pmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \cdots & \hat{\gamma}(p-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \cdots & \hat{\gamma}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}(p-1) & \hat{\gamma}(p-2) & \cdots & \hat{\gamma}(0) \end{pmatrix}^{-1} \begin{pmatrix} \hat{\gamma}(1) \\ \hat{\gamma}(2) \\ \vdots \\ \hat{\gamma}(p) \end{pmatrix}$$

et

$$\tilde{\sigma}^2 = \hat{\gamma}(0) - \tilde{\phi}_1 \hat{\gamma}(1) - \dots - \tilde{\phi}_p \hat{\gamma}(p).$$

La matrice limite Γ_p^{-1} étant définie positive (voir Prop. 2.5), on sait que, pour n suffisamment grand, le système admettra une solution unique. Dans l'exemple ci-dessous, on a simulé un ARMA(2,1) à bruit gaussien ($\phi_1 = 0.6$, $\phi_2 = 0.2$, $\theta_1 = 0.7$, $\sigma^2 = 1$) de taille $n = 200$.

```
> Arima(X, order = c(2,0,1), include.mean = FALSE)
```

Series: X

ARIMA(2,0,1) with zero mean

Coefficients:

```

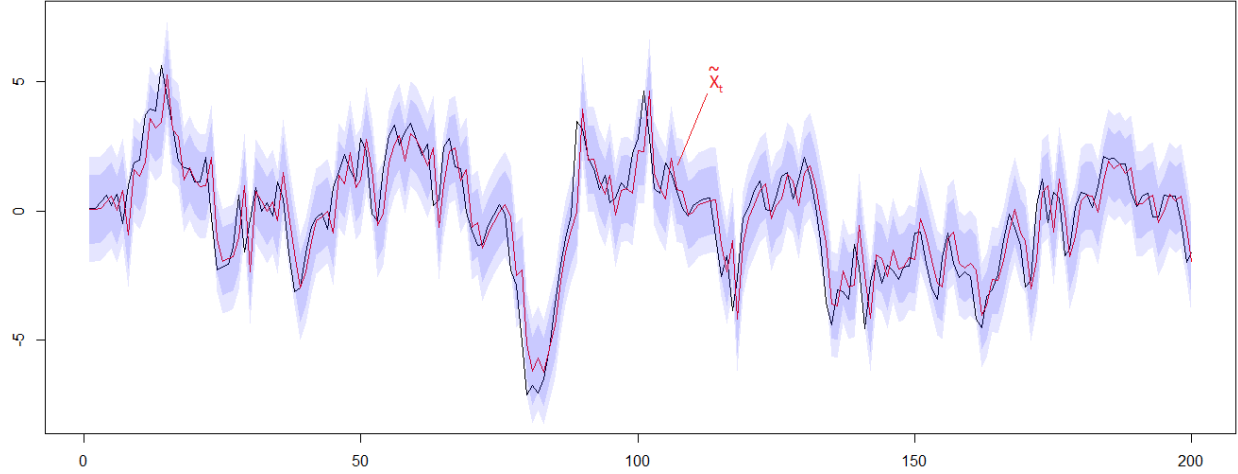
          ar1    ar2    ma1
0.7151  0.079  0.6325
s.e.  0.1323  0.127  0.1060

```

```

sigma^2 estimated as 0.8675:  log likelihood=-269.14
AIC=546.27   AICc=546.48   BIC=559.46

```



On modélise la trajectoire, sans estimation de m , par la procédure **Arima** du package **forecast**. À partir des estimateurs, on reconstruit la trajectoire à *un pas en avant* par la commande **fitted** (on verra dans la section suivante comment obtenir cette reconstruction et les résidus qui lui sont associés). On y ajoute un intervalle de confiance $\pm \hat{\sigma} u_{1-\frac{\alpha}{2}}$ pour $\alpha = 20\%$ et pour $\alpha = 5\%$.

3.8 Prédictions ARMA

On considère un processus $(X_t)_{t \in \mathbb{Z}}$ supposé stationnaire et centré. Le principe est le suivant : sachant $X_t, X_{t-1}, X_{t-2}, \dots$ et tout le passé du processus à l'instant t , on souhaite prévoir X_{t+h} pour un *horizon de prédiction* $h \in \mathbb{N}^*$ par un *prédicteur* de la forme

$$\hat{X}_{t+h} = f(X_t, X_{t-1}, X_{t-2}, \dots).$$

Comme on l'a vu en introduisant la notion d'innovations, la minimisation de $\|X_{t+h} - Y\|^2$ conduit à une espérance conditionnelle qu'on ne sait pas toujours calculer. Selon la même logique, on cherche des prédicteurs dans la classe des fonctions f linéaires. On a alors

$$\hat{X}_{t+h} = P_{\mathcal{H}_t} X_{t+h} \quad \text{où l'on rappelle que} \quad \mathcal{H}_t = \overline{\text{Vect}\{(X_s)_{s \leq t}\}}.$$

Supposons maintenant que $(X_t)_{t \in \mathbb{Z}}$ soit un $\text{ARMA}(p, q)$ centré et mis sous forme canonique. Le résultat suivant nous permet de déterminer le prédicteur \hat{X}_{t+h} et la variance de l'erreur de prédiction.

Proposition 3.4 *Soit le prédicteur $\hat{X}_{t+h} = P_{\mathcal{H}_t} X_{t+h}$ pour un horizon $h \geq 1$. Alors,*

$$\hat{X}_{t+h} = - \sum_{k=1}^{+\infty} \pi_k \hat{X}_{t+h-k} = \sum_{k=h}^{+\infty} \psi_k \varepsilon_{t+h-k}$$

où $(\pi_k)_{k \in \mathbb{N}}$ et $(\psi_k)_{k \in \mathbb{N}}$ sont les coefficients respectifs de l'écriture $\text{AR}(\infty)$ et $\text{MA}(\infty)$ du processus. De plus,

$$\mathbb{E}[(X_{t+h} - \hat{X}_{t+h})^2] = \sigma^2 \sum_{k=0}^{h-1} \psi_k^2.$$

Preuve.

Remarque. Lorsque $h = 1$, on a

$$\hat{X}_{t+1} = - \sum_{k=1}^{+\infty} \pi_k \hat{X}_{t+1-k} = -\pi_1 X_t - \pi_2 X_{t-1} - \pi_3 X_{t-2} - \dots$$

car $X_t, X_{t-1}, X_{t-2}, \dots \in \mathcal{H}_t$. Lorsque $h > 1$, on travaille récursivement avec

$$\hat{X}_{t+2} = -\pi_1 \hat{X}_{t+1} - \pi_2 X_t - \pi_3 X_{t-1} - \dots$$

et ainsi de suite.

Si le processus est gaussien, on en tire l'intervalle de prédiction pour X_{t+h} (sachant \mathcal{H}_t) donné par

$$\text{IP}_{1-\alpha}(X_{t+h}) = \left[\hat{X}_{t+h} \pm \sigma \sqrt{\sum_{k=0}^{h-1} \psi_k^2} u_{1-\alpha/2} \right].$$

On voit que la largeur de l'intervalle grandit avec h , en toute logique : plus l'horizon de prédiction est lointain et moins la prédiction est précise.

Exemples. Donner la valeur explicite du prédicteur à horizon h et de l'erreur quadratique moyenne de prédiction dans les modèles MA(q) et AR(2) pour simplifier (mais avec l'AR(p) c'est le même principe).



La théorie précédente est inadaptée pour effectuer concrètement une prédiction. En effet, la trajectoire (X_1, \dots, X_n) observée possède un instant initial, et tous les paramètres de la modélisation ainsi que ses innovations sont inconnus. On va donc en pratique estimer les opérateurs

$$\hat{\Phi}(B) = I - \hat{\phi}_1 B - \dots - \hat{\phi}_p B^p \quad \text{et} \quad \hat{\Theta}(B) = I + \hat{\theta}_1 B + \dots + \hat{\theta}_q B^q$$

pour estimer à leur tour les $(\hat{\pi}_k)$ en tant que coefficients de $\hat{\Theta}^{-1}(B)\hat{\Phi}(B)$. Le processus est quant à lui tronqué au plus lointain passé observable,

$$\hat{X}_{n+h} = - \sum_{k=1}^{+\infty} \pi_k \hat{X}_{n+h-k} \approx - \sum_{k=1}^{n+h-1} \hat{\pi}_k \tilde{X}_{n+h-k} = \tilde{X}_{n+h}.$$

En itérant les prédictions, il vient

$$\tilde{X}_{n+h} = \begin{cases} -\hat{\pi}_1 X_n - \hat{\pi}_2 X_{n-1} - \dots - \hat{\pi}_n X_1 & \text{si } h = 1 \\ -\hat{\pi}_1 \tilde{X}_{n+h-1} - \dots - \hat{\pi}_{h-1} \tilde{X}_{n+1} - \hat{\pi}_h X_n - \dots - \hat{\pi}_{n+h-1} X_1 & \text{si } h > 1. \end{cases}$$

La suite $(\hat{\psi}_k)$ des coefficients de $\hat{\Phi}^{-1}(B)\hat{\Theta}(B)$ et l'estimateur $\hat{\sigma}^2$ peuvent servir à construire les intervalles de prédiction. Cette stratégie se simplifie en exploitant la relation de récurrence de l'ARMA et la prédiction à un pas en avant :

→ On fixe des valeurs initiales $\tilde{X}_1, \dots, \tilde{X}_c$ où $c \geq \max(p, q)$, selon une stratégie propre à l'utilisateur (l'algorithme d'initialisation peut varier avec les packages, les hypothèses retenues, ...).

→ Pour tout $c+1 \leq t \leq n$, sachant $(X_1, \dots, X_{t-1}, \tilde{X}_1, \dots, \tilde{X}_{t-1})$, on prédit

$$\tilde{X}_t = \hat{\phi}_1 X_{t-1} + \dots + \hat{\phi}_p X_{t-p} + \underbrace{\hat{\theta}_1 (X_{t-1} - \tilde{X}_{t-1})}_{\tilde{\varepsilon}_{t-1}} + \dots + \underbrace{\hat{\theta}_q (X_{t-q} - \tilde{X}_{t-q})}_{\tilde{\varepsilon}_{t-q}}$$

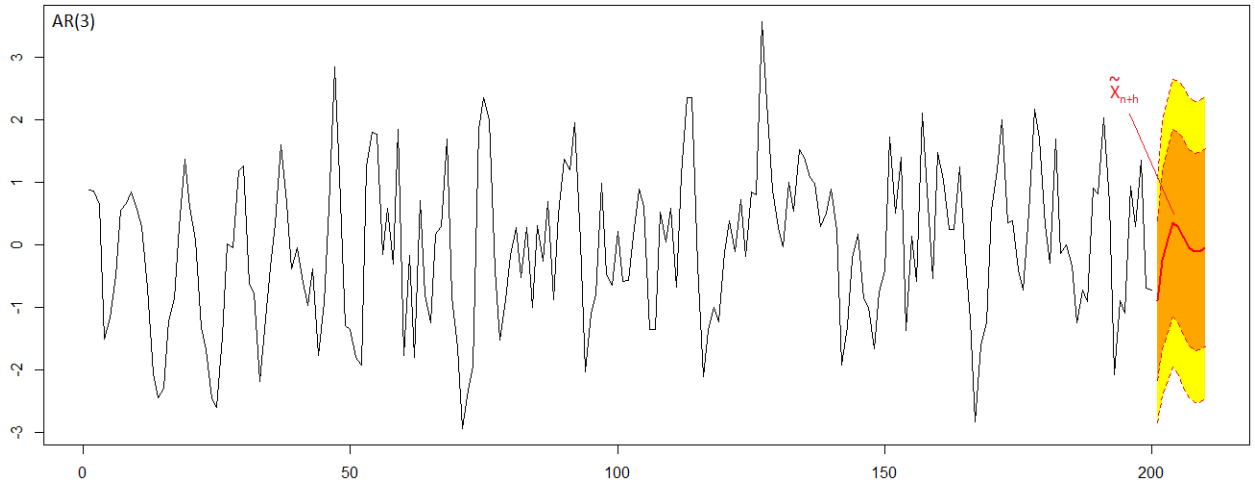
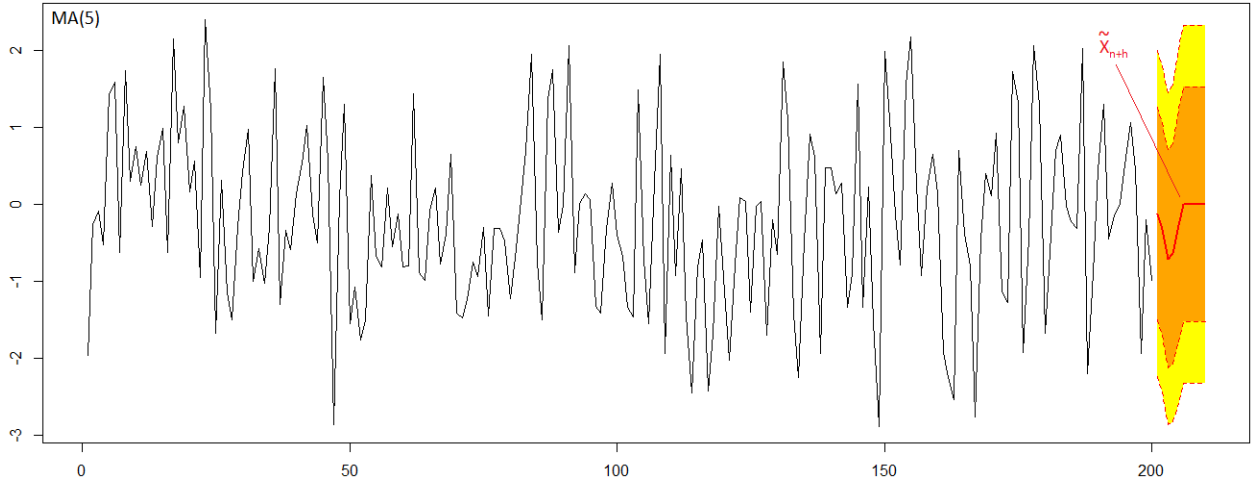
et l'on pose $\tilde{\varepsilon}_t = X_t - \tilde{X}_t$. La suite $(\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_n)$ forme les *résidus* de la modélisation, ou encore les erreurs commises par la prédiction à un pas en avant.

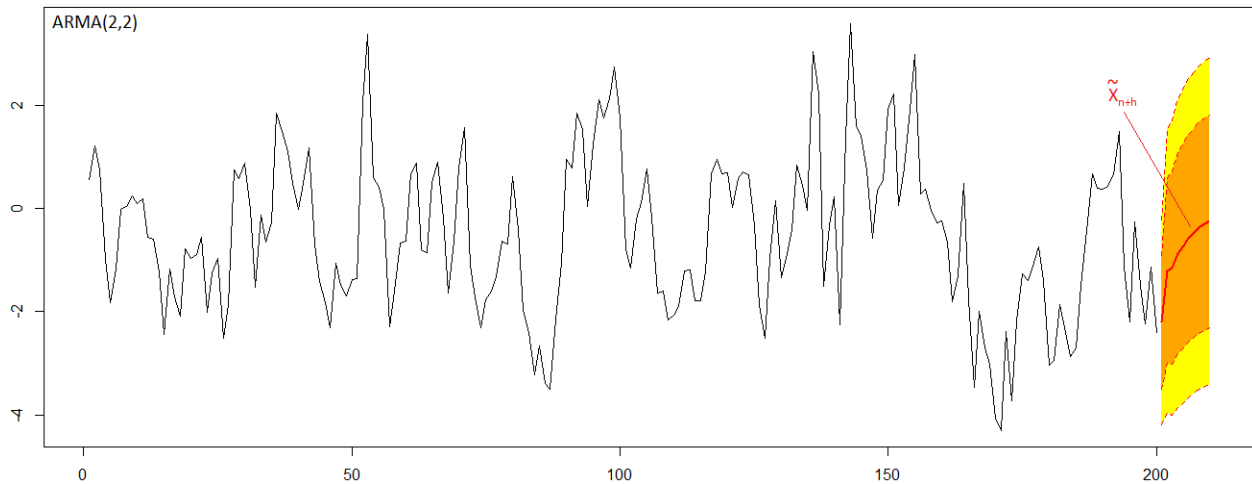
→ La prédiction à horizon $h = 1$ qui en découle naturellement est donnée par

$$\tilde{X}_{n+1} = \hat{\phi}_1 X_n + \dots + \hat{\phi}_p X_{n-p+1} + \hat{\theta}_1 (X_n - \tilde{X}_n) + \dots + \hat{\theta}_q (X_{n-q+1} - \tilde{X}_{n-q+1}).$$

→ Par manque d'observations, on fixe $X_{n+h} = \tilde{X}_{n+h}$.

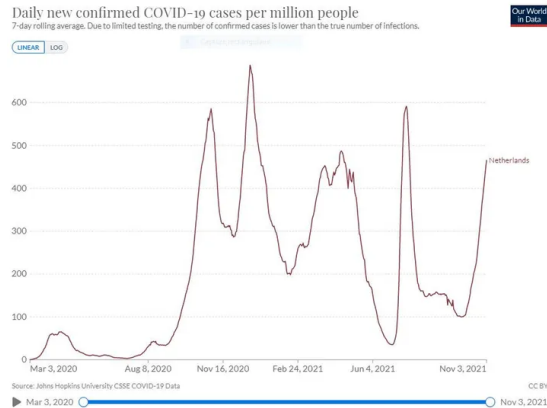
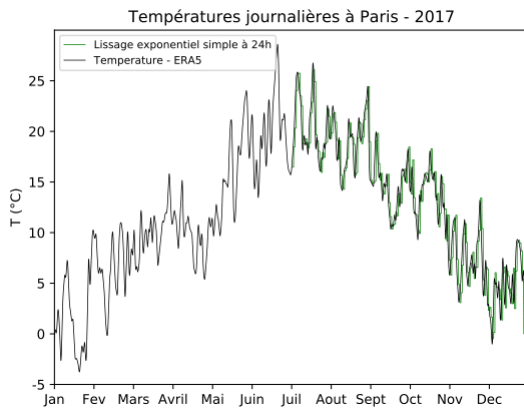
On donne quelques simulations avec des prédictions à horizon $h = 10$ (commande **forecast**). On y ajoute les intervalles de prédiction pour $\alpha = 20\%$ et pour $\alpha = 5\%$.





4 Quelques généralisations des ARMA

Les ARMA sont donc dévolus aux séries stationnaires et ne permettent pas de tenir compte des tendances éventuelles. Or, les séries réelles (économiques, financières, météorologiques, industrielles, etc.) ne sont que rarement stationnaires et possèdent généralement des tendances. Pour s'en convaincre, ces deux séries (issues d'une simple recherche sur le net) montrent des phénomènes réels, on voit bien l'incohérence d'y associer un modèle stationnaire.



Donc les ARMA seront associés aux fluctuations, et l'hypothèse de stationnarité est souvent crédible en ce qui concerne les incréments de ces séries (consécutifs ou saisonniers). Les modèles ARIMA et SARIMA généralisent l'écriture ARMA en y intégrant les éventuelles racines unitaires dans le polynôme autorégressif ainsi qu'une exploration périodique du passé. Le principe est le suivant :

- Tant que la série n'est pas stationnaire, on la différencie (simplement : $X_t - X_{t-1}$ pour les tendances linéaires, saisonnièrement : $X_t - X_{t-s}$ pour les tendances s -périodiques, ...).
- Une fois qu'on a obtenu la stationnarité, on pose un ARMA sur les différences obtenues.

On survolera aussi les modèles hétéroscédastiques ARCH et GARCH, couramment utilisés en finance.

4.1 Processus ARIMA

Définition 4.1 On dit que la série chronologique $(X_t)_{t \in \mathbb{Z}}$ est un processus autorégressif moyenne mobile intégré d'ordres (p, d, q) si la série $Y_t = (I - B)^d X_t$ admet une écriture $ARMA(p, q)$ causale. Son ordre

d'intégration est alors $d \in \mathbb{N}$. On note $(X_t) \sim \text{ARIMA}(p, d, q)$. La notation I vient de l'anglais "*Integrated*".

De manière condensée, on a

$$\Phi^*(B) X_t = (I - B)^d \Phi(B) X_t = \Theta(B) \varepsilon_t$$

où $\Phi(z) \neq 0$ pour $|z| \leq 1$. Cette écriture amène quelques remarques :

- Un tel processus n'est pas stationnaire pour $d \geq 1$, il existe des racines unitaires dans Φ^* .
- Par ailleurs, le filtre $I - B$ permet également d'éliminer les tendances linéaires pour les ramener à des constantes. En d'autres termes,

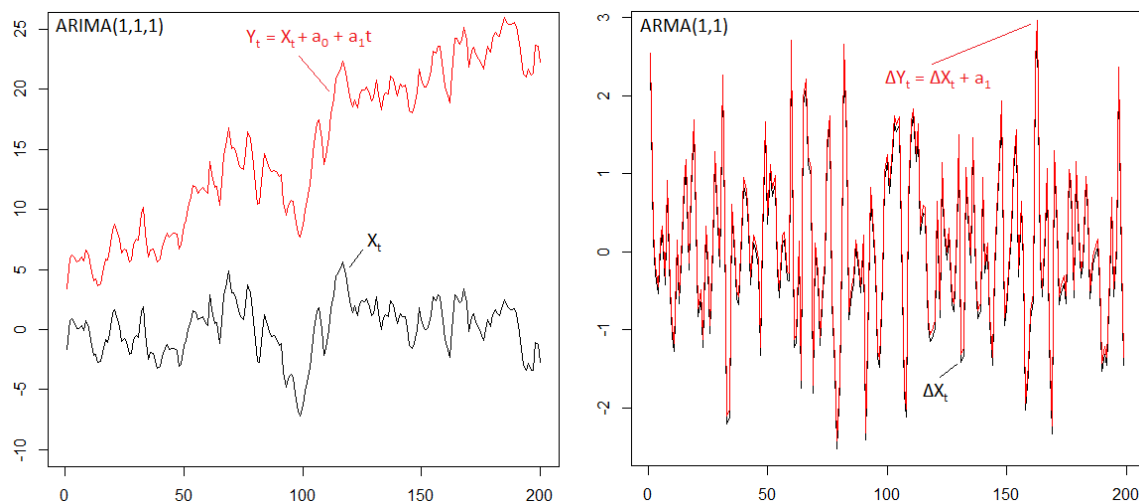
$$Y_t = X_t + a_0 + a_1 t$$

vérifie une équation $\text{ARIMA}(p, d, q)$ dès que X_t la vérifie aussi. Cela signifie qu'il faut être prudent lors de l'application du filtre $I - B$, on risque de passer à côté d'une tendance linéaire en l'éliminant involontairement. Les packages usuels permettent d'inclure une tendance constante ou linéaire en plus du modèle ARIMA, au besoin, c'est-à-dire que, pour $d = 1$, on aura accès à une estimation de la pente a_1 (le *drift*) dans un modèle de la forme

$$(I - B) \Phi(B) \underbrace{(Y_t - a_0 - a_1 t)}_{X_t} = \Theta(B) \varepsilon_t.$$

C'est aussi vrai pour $(I - B)^d$ évidemment mais on ne rencontre pas ce cas (ou presque).

Lorsqu'une série n'est pas stationnaire, on peut donc la *différencier* par l'opérateur $I - B$ et tester la stationnarité de ses incréments. Le cas échéant, on choisit un modèle ARIMA avec $d = 1$. Si les incréments ne sont pas stationnaires, on essaie $d = 2$ mais en pratique on évite d'aller au-delà.



```
# ARIMA(1,1,1) sur la série X sans tendance
> Arima(X, order = c(1,1,1), include.mean = FALSE)
```

```
Series: X
ARIMA(1,1,1)
```

```
Coefficients:
```

```
      ar1      ma1
      0.5886 0.6916
s.e.   0.0614 0.0506
```

```
sigma^2 estimated as 1.142: log likelihood=-295.44
AIC=596.88  AICc=597.01  BIC=606.76
```

```
# ARIMA(1,1,1) sur la série Y avec tendance linéaire
> Arima(Y, order = c(1,1,1), include.drift = TRUE)
```

```
Series: Y
ARIMA(1,1,1) with drift
```

```
Coefficients:
          ar1      ma1      drift
      0.5749  0.6938  0.2118
s.e.  0.0624  0.0504  0.3196
```

```
sigma^2 estimated as 1.139:  log likelihood=-294.68
AIC=597.36  AICc=597.57  BIC=610.53
```

Remarque. Attention à bien saisir l'utilité des arguments `include.mean` et `include.drift` ainsi que l'estimation du paramètre `drift` lorsqu'il est demandé.

Exemples. Donner l'écriture explicite de X_t dans les modèles :

→ $(X_t) \sim \text{ARIMA}(1, 1, 0)$.

→ $(X_t) \sim \text{ARIMA}(2, 1, 1)$.

→ $(X_t) \sim \text{ARIMA}(0, 2, 2)$.

4.2 Processus SARIMA

Ce qui suit pourra paraître très compliqué mais c'est surtout une façade, car les techniques utilisées sont essentiellement les mêmes que pour l'ARIMA. En outre c'est fondamental, car très utilisé dans les entreprises manipulant des données chronologiques, et c'est populaire car englobant beaucoup de phénomènes (tendance et saisonnalité dans un même modèle). Considérons, pour $P, Q \in \mathbb{N}$, les nouveaux opérateurs

$$\Phi_s(B) = I - \alpha_1 B - \dots - \alpha_P B^P \quad \text{et} \quad \Theta_s(B) = I + \beta_1 B + \dots + \beta_Q B^Q.$$

Définition 4.2 On dit que la série chronologique $(X_t)_{t \in \mathbb{Z}}$ est un processus autorégressif moyenne mobile intégré et saisonnier d'ordres $(p, d, q) \times (P, D, Q)_s$ si la série $Y_t = (I - B^s)^D (I - B)^d X_t$ admet une écriture ARMA causale caractérisée par

$$\Phi_s(B^s) \Phi(B) Y_t = \Theta_s(B^s) \Theta(B) \varepsilon_t.$$

Ses ordres d'intégration sont alors $d \in \mathbb{N}$ et $D \in \mathbb{N}$, et sa période est $s \in \mathbb{N} \setminus \{0, 1\}$. On note $(X_t) \sim \text{SARIMA}(p, d, q) \times (P, D, Q)_s$. La notation S vient de l'anglais "Seasonal".

De manière condensée, on a

$$\Phi^*(B) X_t = (I - B^s)^D (I - B)^d \Phi_s(B^s) \Phi(B) X_t = \Theta_s(B^s) \Theta(B) \varepsilon_t = \Theta^*(B) \varepsilon_t$$

où $\Phi_s(z) \Phi(z) \neq 0$ pour $|z| \leq 1$. Par analogie avec l'ARIMA, on remarque que :

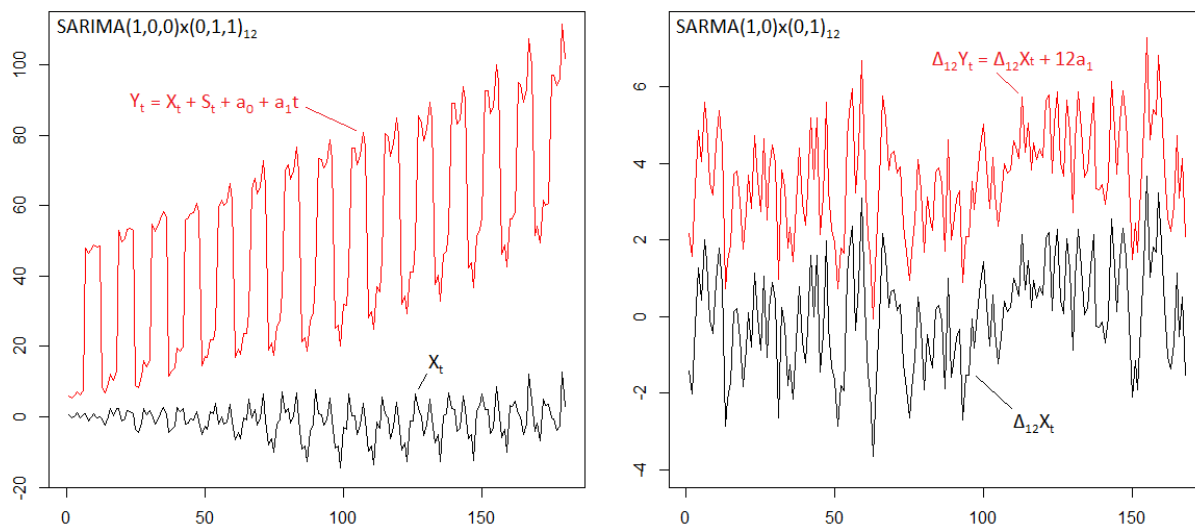
- Un tel processus n'est pas stationnaire pour $d + D \geq 1$, il existe des racines unitaires dans Φ^* .
- Le filtre $I - B^s$ agit sur les tendances linéaires pour les ramener à des constantes et il en va de même pour toute tendance s -périodique (S_t). En d'autres termes,

$$Y_t = X_t + S_t + a_0 + a_1 t$$

vérifie une équation SARIMA(p, d, q) \times (P, D, Q) $_s$ dès que X_t la vérifie aussi. On voit cela en remarquant que $(I - B^s)S_t = S_t - S_{t-s} = 0$ et que $(I - B^s)(a_0 + a_1 t) = a_1 s$. C'est vrai pour des ordres bien plus grands avec $(I - B^s)^D (I - B)^d$ mais ça ne se rencontre pas.

- En pratique, on a typiquement $d, D \leq 1$, $p, q \leq 5$, $P, Q \leq 1$. Les modèles plus gros doivent absolument être évités.
- La différenciation $I - B$ est peu coûteuse : on perd la première valeur de la série. La différenciation saisonnière $I - B^s$ est bien plus coûteuse lorsque s est grand, car on perd les s premières valeurs. Il faudra donc disposer de suffisamment de périodes pour que la modélisation soit pertinente.

Lorsqu'une série est périodique, on peut donc la *différencier* par l'opérateur $I - B^s$, quitte à combiner avec $I - B$ au besoin, et vérifier la stationnarité des incréments. Le cas échéant, on choisit un modèle SARIMA avec $D = 1$ et $d = 0/1$. On prendra garde à ne pas oublier les tendances potentiellement éliminées par les différenciations. En pratique on peut tester $(d, D) = (0, 1)$, $(1, 0)$ ou $(1, 1)$ mais on évite d'aller au-delà.



```
# SARIMA(1,0,0)x(1,1,1)12 sur la série X sans tendance
```

```
> Arima(X, order = c(1,0,0), seasonal = list(order = c(0,1,1), period = 12),
  include.mean = FALSE)
```

```
Series: X
```

```
ARIMA(1,0,0)(0,1,1)[12]
```

```
Coefficients:
```

```
      ar1      sma1
    0.4327  0.7246
s.e.  0.0704  0.0693
```

```
sigma^2 estimated as 0.9898: log likelihood=-241.09
```

```
AIC=488.18  AICc=488.32  BIC=497.55
```

```
# SARIMA(1,0,0)x(1,1,1)12 sur la série Y avec tendance linéaire
> Arima(Y, order = c(1,0,0), seasonal = list(order = c(0,1,1), period = 12),
      include.drift = TRUE)
```

```
Series: Y
ARIMA(1,0,0)(0,1,1)[12] with drift
```

```
Coefficients:
      ar1      sma1      drift
      0.4320  0.7245  0.3037
s.e.   0.0705  0.0693  0.0187
```

```
sigma^2 estimated as 0.9957: log likelihood=-241.07
AIC=490.14   AICc=490.38   BIC=502.63
```

Exemples. Donner l'écriture explicite de X_t dans les modèles :

$\rightarrow (X_t) \sim \text{SARIMA}(1,0,0) \times (0,1,1)_{12}.$

$\rightarrow (X_t) \sim \text{SARIMA}(0,1,1) \times (1,0,1)_7.$

$\rightarrow (X_t) \sim \text{SARIMA}(0,1,0) \times (1,1,0)_{24}.$

4.3 Prédictions ARIMA et SARIMA

Il est facile de généraliser les prédictions ARMA aux ARIMA et SARIMA, il suffit pour cela d'inverser localement les opérateurs de différenciation. Par exemple pour un $\text{ARIMA}(p,1,q)$, le processus défini par $Y_t = (I - B) X_t$ est un $\text{ARMA}(p,q)$ causal sur lequel on sait faire une prédiction \hat{Y}_{t+1} . Alors,

$$\hat{X}_{t+1} = X_t + \hat{Y}_{t+1}.$$

À titre d'exemple, on applique un modèle $\text{SARIMA}(1,0,1) \times (0,1,1)_{12}$ avec une tendance linéaire au logarithme de la série `AirPassengers`, disponible dans le package `datasets`.

```
> SARIMA = Arima(X, order = c(1,0,1), seasonal = list(order = c(0,1,1), period = 12),
      include.drift = TRUE)
```

```
Series: X
ARIMA(1,0,1)(0,1,1)[12] with drift
```

```
Coefficients:
      ar1      ma1      sma1      drift
```



```

      0.9017 -0.3324 -0.5561 0.0098
s.e. 0.0469 0.1039 0.0754 0.0008

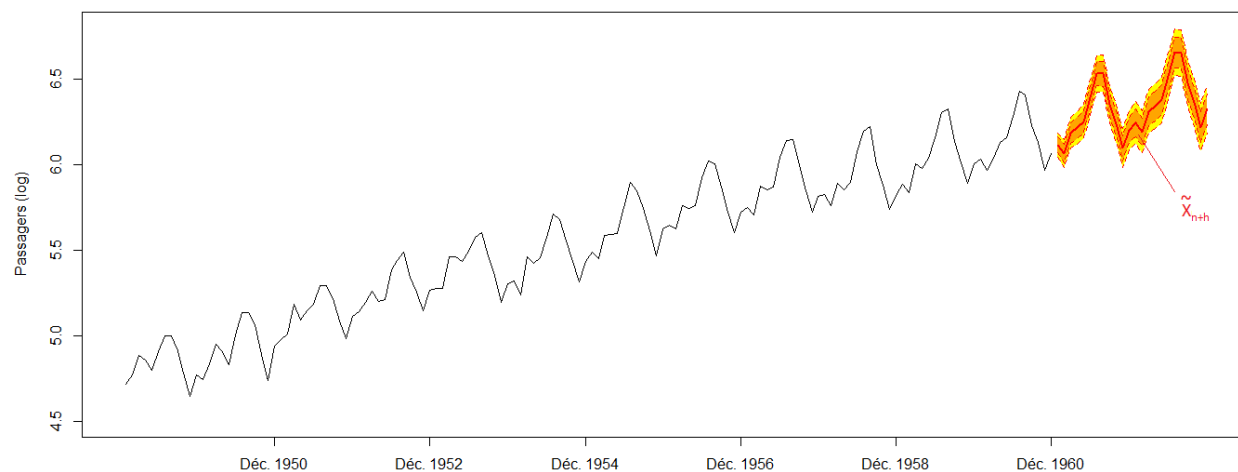
```

```

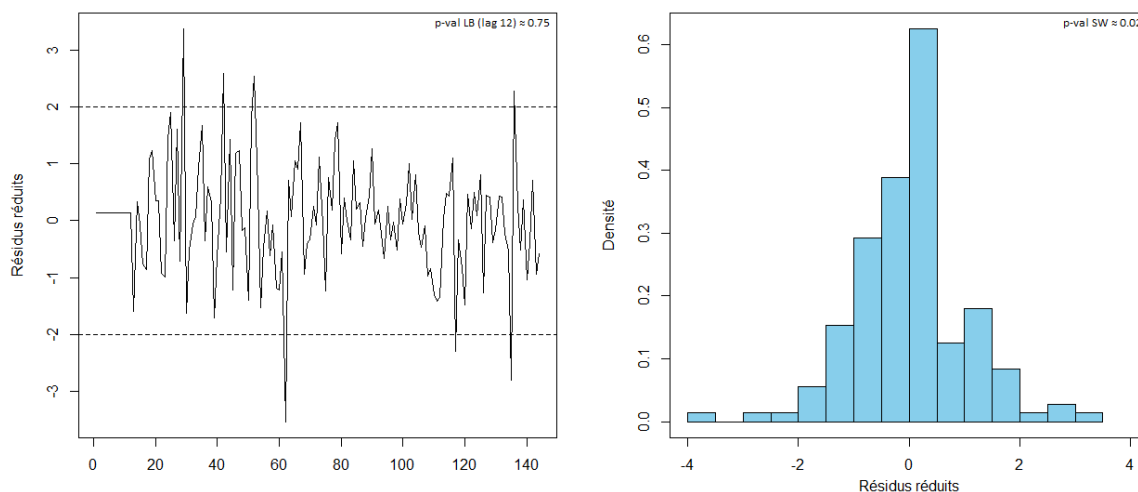
sigma^2 estimated as 0.001325: log likelihood=249.59
AIC=-489.19 AICc=-488.71 BIC=-474.77

```

```
> Pred = forecast(SARIMA, h = 12, level = c(0.95, 0.80))
```



On a représenté les données observées et la prédiction de deux périodes supplémentaires (donc pour $h = 24$) avec les intervalles de prédiction à 80 % et 95 %, cela pour la série logarithmique (ci-dessus) et la série originale (ci-dessous). On étudie également les résidus de modélisation de la série logarithmique.



Remarque. Soient (Z_t) la série originale et (X_t) sa transformation logarithmique. Si \hat{X}_{n+1} est le prédictor de X_{n+1} , alors aussi bizarre que ça puisse paraître, il n'est pas toujours judicieux de poser

$$\hat{Z}_{n+1} = e^{\hat{X}_{n+1}}.$$

De manière informelle, considérer que (X_t) se comporte comme un SARIMA revient à écrire $X_t = C_{t-1} + \varepsilon_t$ où C_{t-1} contient une information accessible à l'instant $t - 1$ et (ε_t) est un bruit additif. Cela engendre une

relation multiplicative de la forme $Z_t = e^{C_{t-1}} e^{\varepsilon_t}$. Or sous l'hypothèse $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, on a $e^\varepsilon \sim \mathcal{LN}(0, \sigma^2)$ et la loi log-normale n'étant pas centrée, on corrige généralement notre prédiction par le facteur

$$\mathbb{E}[e^\varepsilon] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} e^u e^{-\frac{u^2}{2\sigma^2}} du = e^{\frac{\sigma^2}{2}}$$

que l'on estime naturellement en injectant $\hat{\sigma}^2$ en lieu et place de σ^2 . Finalement, le prédicteur de Z_{n+1} est donné par

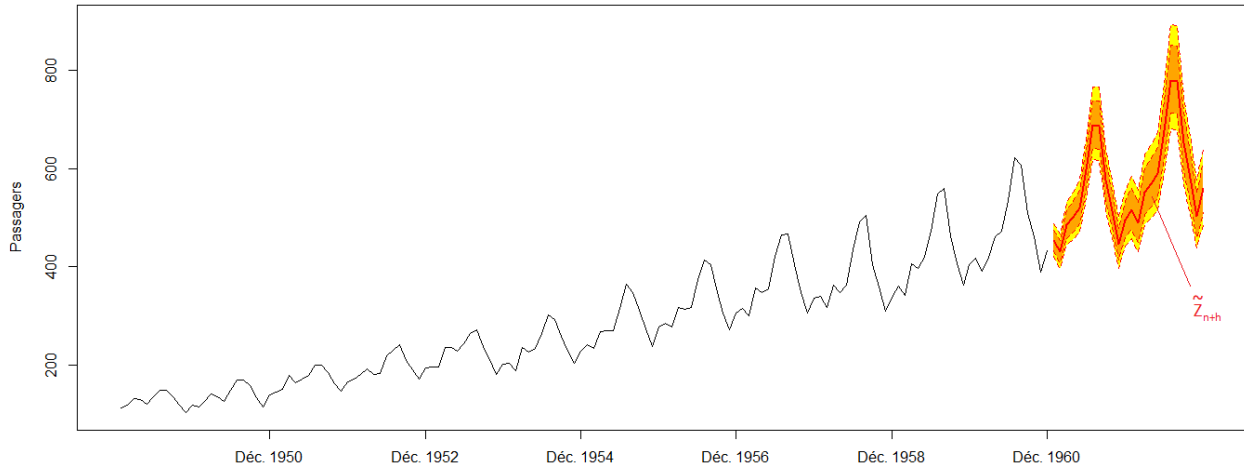
$$\hat{Z}_{n+1} = e^{\hat{X}_{n+1} + \frac{\hat{\sigma}^2}{2}}.$$

Si l'hypothèse de normalité de ε n'est pas crédible sur le jeu de données, on peut également estimer le facteur de correction par

$$\hat{\mathbb{E}}[e^\varepsilon] = \frac{1}{n} \sum_{k=1}^n e^{\tilde{\varepsilon}_k}$$

à l'aide des résidus de la modélisation. Cependant les intervalles de prédictions ne sont pas impactés par ce coefficient, puisque

$$\mathbb{P}(X_{n+1} \in [b_{\inf}, b_{\sup}]) = \mathbb{P}(Z_{n+1} \in [e^{b_{\inf}}, e^{b_{\sup}}]).$$



Le *pouvoir prédictif* d'un modèle est un critère en général très apprécié afin de sélectionner une modélisation. L'idée est de retirer une certaine quantité de mesures en fin de série, d'estimer le modèle sur la série tronquée, puis de reprédire les valeurs supprimées. C'est une forme de validation croisée... mais pas question ici de retirer des données aléatoirement comme dans la 'vraie' validation croisée. C'est chronologique !

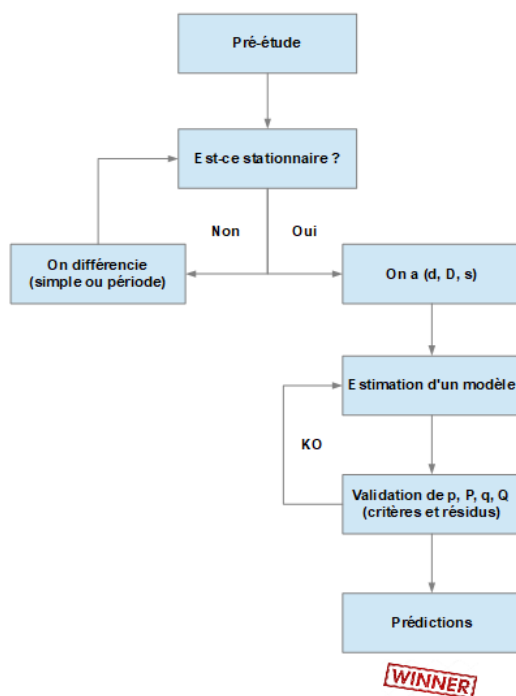
4.4 Sélection et validation

Une fois le modèle AR/MA/ARMA/ARIMA/SARIMA mis en place sur le jeu de données, il reste à se demander *a posteriori* si l'on a fait un bon choix et si les hypothèses retenues sont approximativement vérifiées. Quelques étapes conseillées :

- A-t-on bien pensé à afficher la série à l'écran pour s'en faire une première idée, avant de se lancer dans son traitement ?
- La série ne gagnerait-elle pas à être prétraitée ? (passage au log, ruptures de niveau, valeurs aberrantes et/ou manquantes, ...).
- La série était-elle bien stationnaire ? Sinon, l'a-t-on suffisamment différenciée pour la rendre stationnaire et éliminer son éventuelle périodicité ? (ADF, KPSS, ACF/PACF, aspect visuel, ...).
- N'a-t-on pas oublié une tendance en cours de traitement, éliminée par la différenciation ?

- Les ACF et PACF empiriques se comportent-elles comme la théorie des ARMA le prédit ? L'ACF ou la PACF ne s'annulent-elles pas à partir d'un certain rang ?
- Les rangs p et q retenus sont-ils pertinents ? (significativité des estimateurs, ...).
- Si le modèle est saisonnier, a-t-on correctement choisi s , P et Q ? (Pics saisonniers dans l'ACF et la PACF, significativité des estimateurs, périodogramme, ...).
- Ne peut-on pas améliorer les résultats avec des modèles voisins ? A-t-on fait varier les paramètres dans leur voisinage et optimisé un certain critère ? (AIC, BIC, MSE, MAPE, procédure `auto.arima`, pouvoir prédictif, ...).
- Les résidus forment-ils un bruit blanc ? (Box-Pierce et Ljung-Box avec `Box.test`, Durbin-Watson avec `dwtest`, Breusch-Godfrey avec `bgtest`, ACF/PACF, ...). S'il reste des corrélations, n'est-il pas judicieux d'augmenter p et/ou q ? Sont-ils homoscedastiques (visuellement, Breusch-Pagan avec `bptest`, ...) ?
- Les résidus sont-ils approximativement gaussiens ? (Shapiro-Wilk, Kolmogorov-Smirnov, histogramme, majoritairement dans l'intervalle $[-2, 2]$ une fois réduits, ...).

La stratégie consistant à différencier une série jusqu'à la rendre stationnaire, puis à étudier les ACF/PACF pour ajuster un modèle ARMA sur la série différenciée jusqu'à obtenir un bruit blanc résiduel est aussi appelée *méthode de Box-Jenkins*.



4.5 Processus ARCH et GARCH

En finance, il est fréquent que l'on considère la *volatilité* (liée à l'écart-type du bruit) comme évoluant au cours du temps. Cela permet d'englober des périodes agitées et des périodes calmes dans un même modèle. Les chroniques financières sont connues pour contenir des *clusters de volatilité* ("large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes" (B. Mandelbrot)), ce qui signifie que les ARMA ne sont pas adaptés pour de telles séries et l'on voudrait plutôt modéliser un phénomène d'*hétéroscédasticité conditionnelle*. Un processus autorégressif causal (centré) satisfait

$$\mathbb{E}[X_t] = 0 \quad \text{et} \quad \mathbb{E}[X_t | \mathcal{H}_{t-1}] = \sum_{i=1}^p \phi_i X_{t-i}$$

puisque, par causalité, $\varepsilon_t \perp \mathcal{H}_{t-1}$. Il adapte donc sa valeur moyenne au passé récent. En revanche, le fait que

$$\mathbb{V}(X_t) = \gamma(0) \quad \text{et} \quad \mathbb{V}(X_t | \mathcal{H}_{t-1}) = \sigma^2$$

revient à dire qu'il ne peut pas modéliser des variations locales de variance : quel que soit le passé observé (agité, calme, ...), la variance est inchangée. Essayons de construire une structure autorégressive remédiant à cela et considérons les deux nouveaux opérateurs

$$\alpha(B) = \alpha_1 B + \dots + \alpha_q B^q \quad \text{et} \quad \beta(B) = \beta_1 B + \dots + \beta_p B^p$$

avec $\alpha_i \geq 0$ pour $i \in \{1, \dots, p\}$ et $\beta_j \geq 0$ pour $j \in \{1, \dots, q\}$.

Définition 4.3 On dit que la série chronologique $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un processus autorégressif conditionnellement hétéroscédastique d'ordre q si elle est définie par

$$\forall t \in \mathbb{Z}, \quad \begin{cases} \varepsilon_t &= \sigma_t \eta_t \\ \sigma_t^2 &= \omega + \alpha(B) \varepsilon_t^2 \end{cases}$$

où $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc fort de variance 1 et $\omega > 0$. On note $(\varepsilon_t) \sim \text{ARCH}(q)$. La notation ARCH vient de l'anglais "AutoRegressive Conditional Heteroscedastic".

Définition 4.4 On dit que la série chronologique $(\varepsilon_t)_{t \in \mathbb{Z}}$ est un processus autorégressif conditionnellement hétéroscédastique généralisé d'ordres (p, q) si elle est définie par

$$\forall t \in \mathbb{Z}, \quad \begin{cases} \varepsilon_t &= \sigma_t \eta_t \\ \sigma_t^2 &= \omega + \alpha(B) \varepsilon_t^2 + \beta(B) \sigma_t^2 \end{cases}$$

où $(\eta_t)_{t \in \mathbb{Z}}$ est un bruit blanc fort de variance 1 et $\omega > 0$. On note $(\varepsilon_t) \sim \text{GARCH}(p, q)$. La notation G vient de l'anglais "Generalized".

Les autorégressions portent donc sur le processus mais également sur sa volatilité. Mais... la proposition suivante montre qu'un processus (G)ARCH se comporte comme un bruit blanc, sous certaines hypothèses.

Proposition 4.1 Si les polynômes autorégressifs α et β sont tels que $\alpha(1) + \beta(1) < 1$, alors le processus $(\varepsilon_t)_{t \in \mathbb{Z}}$ engendré par la relation $\text{GARCH}(p, q)$ est un bruit blanc centré et de variance

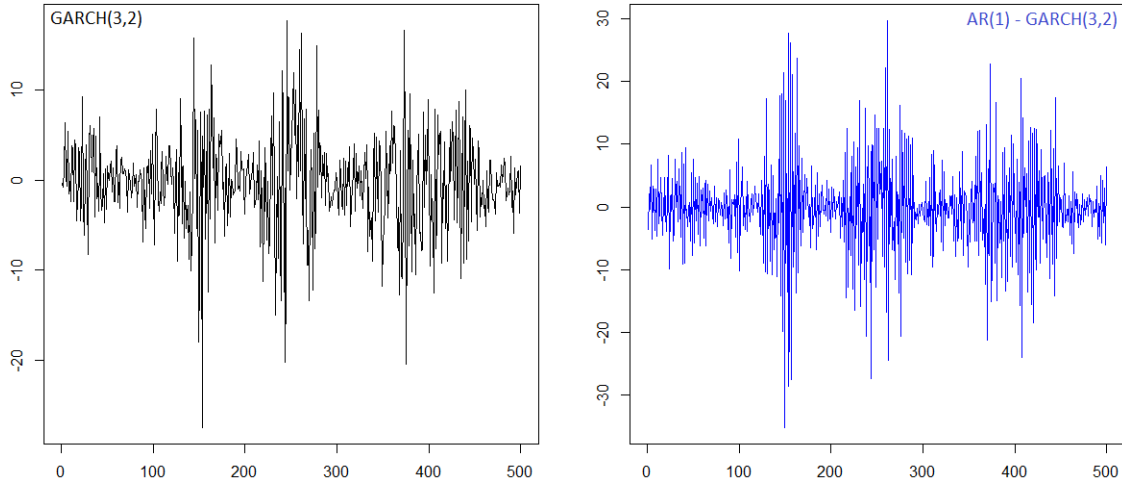
$$\gamma(0) = \frac{\omega}{1 - \alpha(1) - \beta(1)}.$$

Preuve (partielle).

Par contre, la relation

$$\mathbb{E}[\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots] = \sigma_t^2 \neq \sigma^2$$

montre que la variance conditionnelle du processus est influencée par le passé récent. Cela motive l'utilisation des modèles GARCH très répandus en finance et en économétrie, mais il n'empêche que tout ça reste un bruit blanc, on a donc perdu la corrélation... En définitive, on combine : un modèle ARMA avec bruit GARCH vérifie nos contraintes. Les paramètres sont généralement estimés par maximum de vraisemblance. Les packages `tseries`, `TSA` ou encore `fGarch` contiennent des outils pour manipuler les GARCH. On voit bien sur les simulations ci-dessous les changements de variance.



En conclusion générale pour le domaine chronologique, on a développé dans ce cours assez largement la théorie ARMA parce que ce choix est naturel et très utilisé, mais il existe d'autres procédures, parfois plus simples et plus rapides en temps de calcul, parfois également trop simples. Comme souvent en statistique il n'y a pas de règle générale, on gagnera à modéliser une série par un SARIMA pendant qu'une autre sera plus réceptive à une approche additive. Les lissages exponentiels, le *Facebook Prophet* qu'on a évoqué ou encore l'approche non-paramétrique par des noyaux qui mesurent la ressemblance entre des blocs passés et le bloc présent pour prédire le bloc futur forment autant d'alternatives aux SARIMA. On en verra quelques applications en TP.

5 Introduction à la statistique spatiale

Lorsque le processus est indexé par un ensemble S qui ne représente plus le temps mais l'espace, on parle de *statistique spatiale*. Les problèmes rencontrés sont alors essentiellement bidimensionnels (en géostatistique, climatologie, épidémiologie, etc.). On va aborder les modèles les plus simples de *krigeage*, consistant à reconstruire par interpolation spatiale un phénomène bidimensionnel à partir d'un échantillonnage (par exemple, les cartes météorologiques).

5.1 Variogramme

Posons

$$\forall s \in S, \quad Z_s = \mu(s) + \delta_s$$

où $\mu : S \rightarrow \mathbb{R}$ est une fonction déterministe donnant la valeur moyenne du processus, et $(\delta_s)_{s \in S}$ est une collection de v.a.r. centrées de fonction de covariance $\Gamma : S \times S \rightarrow \mathbb{R}$. Le processus $(Z_s)_{s \in S}$ est donc spatial. Généralement, la moyenne prend l'une des formes suivantes :

- $\mu(s) = m$: m connue (krigeage simple).
- $\mu(s) = m$: m inconnue (krigeage ordinaire).
- $\mu(s) = \sum_k \alpha_k f_k(s)$: combinaison linéaire de fonctions de base f_k (krigeage universel).

Il faut également spécifier la structure de dépendance dans le bruit parce que c'est justement le principe de supposer que le phénomène évolue de manière corrélée. Par analogie avec les séries chronologiques, la collection $(\delta_s)_{s \in S}$ doit satisfaire une hypothèse de stationnarité, au sens spatial.

Définition 5.1 La collection $(\delta_s)_{s \in S}$ est stationnaire au second ordre si elle satisfait

$$\forall s \in S, \quad \mathbb{E}[\delta_s] = 0 \quad \text{et} \quad \forall (s, s+h) \in S^2, \quad \text{Cov}(\delta_s, \delta_{s+h}) = c(h).$$

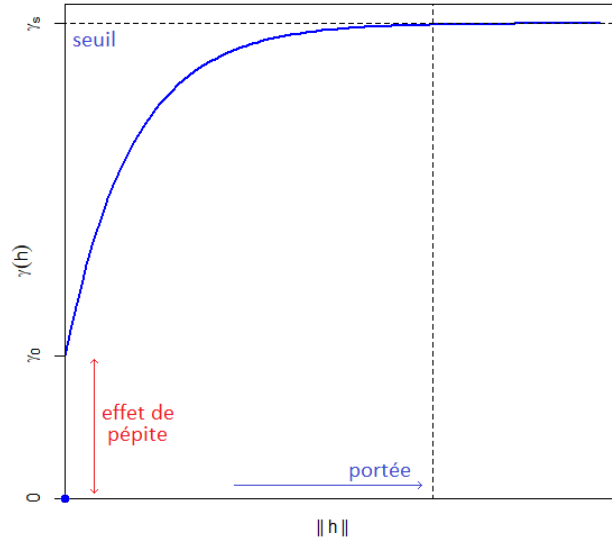
La fonction c est le covariogramme. La collection est intrinsèquement stationnaire si elle satisfait

$$\forall s \in S, \quad \mathbb{E}[\delta_s] = 0 \quad \text{et} \quad \forall (s, s+h) \in S^2, \quad \mathbb{V}(\delta_{s+h} - \delta_s) = 2\gamma(h).$$

La fonction γ est le semi-variogramme.

Exemple. Choisir le symbole qui convient : au second ordre (\Leftarrow / \Leftrightarrow / \Rightarrow) intrinsèquement.

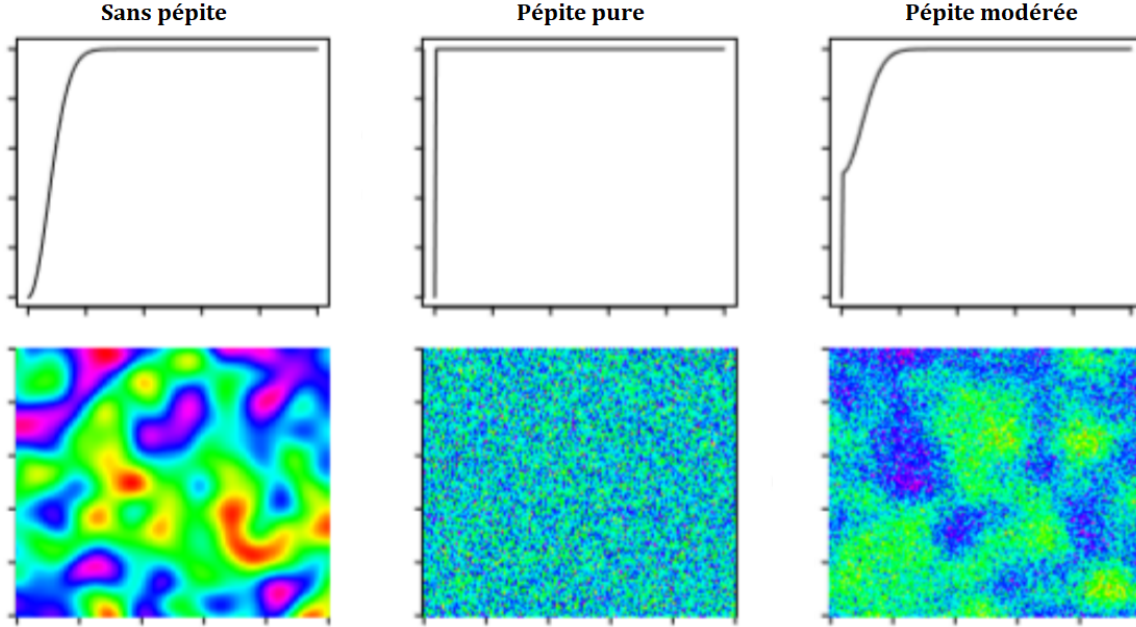
Lorsque ces quantités ne dépendent de h qu'à travers $\|h\|$, on parle d'*isotropie* : seule la distance au sens géométrique du terme intervient, et non la direction. On trouve dans la littérature toute une gamme de semi-variogrammes, à choisir selon les applications. Un exemple typique dont on va décrire le vocabulaire par la suite :



Parmi ceux-là, les modèles à *effet de pépité* sont tels que

$$\lim_{\|h\| \rightarrow 0^+} \gamma(h) = \gamma_0 > \gamma(0) = 0,$$

c'est-à-dire qu'il existe une variabilité à petite échelle. Le graphique ² ci-dessous montre bien l'influence de l'effet de pépité sur le résultat obtenu : aucune pépité conduit à une variabilité inexistante à petite échelle (et donc à un effet de lissage très marqué) alors qu'une pépité pure conduit à une variabilité maximale à petite échelle (et donc à un résultat inexploitable, de granulosité maximale). En générale elle est petite mais non nulle.



Les modèles à *seuil* vérifient

$$\lim_{\|h\| \rightarrow +\infty} \gamma(h) = \gamma_s.$$

Si le seuil γ_s est atteint, la *portée* marque la frontière à partir de laquelle l'influence des points n'évolue plus significativement. Si l'on dispose de suffisamment de données observées dans $S_{\text{obs}} \subset S$, on peut également estimer les paramètres de $\gamma(h)$ ou même la fonction $\gamma(h)$ elle-même en travaillant sur les couples de points distants de h . En krigeage simple ou ordinaire, on voit en effet que

$$\gamma(h) = \frac{\mathbb{V}(\delta_{s+h} - \delta_s)}{2} = \frac{\mathbb{V}(Z_{s+h} - Z_s)}{2} = \frac{\mathbb{E}[(Z_{s+h} - Z_s)^2]}{2}.$$

Une estimation naturelle est donc

$$\hat{\gamma}(h) = \frac{1}{2|V_h|} \sum_{(s_1, s_2) \in V_h} (Z_{s_1} - Z_{s_2})^2$$

où $V_h = \{(s_1, s_2) \in S_{\text{obs}}^2 : \|s_1 - s_2\| \approx \|h\|\}$. On utilise une approximation pour assouplir la condition (à moins d'un quadrillage parfait...). Certains packages séparent les écarts (par exemple en 10 classes) et mettent chaque couple de points dans la case correspondante pour estimer $\gamma(h)$ directement ou ses paramètres, il y a de nombreuses façons de faire mais toutes sont très critiquables en raison du manque d'observations... Sous l'hypothèse de stationnarité au second ordre, $\gamma(h)$ détermine $c(h)$ et donc toute la structure de corrélation du bruit, elle est indispensable aux besoins de l'étude.

2. http://math.univ-lille1.fr/~calgaro/TER_2019/wa_files/langlois_pudlicki.pdf

5.2 Contraintes du krigage

Ayant observé $(Z_{s_1}, \dots, Z_{s_n})$ pour $\{s_1, \dots, s_n\} = S_{\text{obs}} \subset S$, le krigage consiste à reconstruire la valeur du processus en tout point $s_0 \in S$ selon les quatre contraintes suivantes :

- 1) (Linéarité) La prédiction est une combinaison linéaire des observations.
- 2) (Autorisation) La variance de l'erreur de prédiction existe.
- 3) (Non biais) La prédiction est sans biais.
- 4) (Optimalité) La variance de l'erreur de prédiction est minimale.

Il s'agit donc d'une prédiction additive (combinaison linéaire) dans laquelle on retrouve les fondements de la notion inférentielle d'*efficacité* (sans biais et de variance minimale).

5.3 Exemple du krigage simple

On va se contenter pour conclure le module de décrire la mise en œuvre de ces contraintes dans le krigage simple, donc pour $Z_s = m + \delta_s$ avec une moyenne m connue. Dans ce qui suit, on note

$$\mathbb{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad Z = \begin{pmatrix} Z_{s_1} \\ \vdots \\ Z_{s_n} \end{pmatrix}, \quad \Sigma = \mathbb{V}(Z) = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nn} \end{pmatrix} \quad \text{et} \quad c_0 = \text{Cov}(Z, Z_{s_0}) = \begin{pmatrix} c_{01} \\ \vdots \\ c_{0n} \end{pmatrix}$$

avec $c_{k\ell} = \text{Cov}(Z_{s_k}, Z_{s_\ell}) = \text{Cov}(\delta_{s_k}, \delta_{s_\ell})$. On se place sous l'hypothèse de stationnarité au second ordre (des résultats similaires peuvent être établis sous l'hypothèse de stationnarité intrinsèque). On commence par recentrer le processus, puisque la moyenne est connue, et l'on pose

$$\forall s_0 \in S, \quad \hat{Z}_{s_0} - m = \sum_{i=1}^n \lambda_i (Z_{s_i} - m).$$

C'est une application de la première contrainte : on veut un prédicteur au site s_0 qui soit combinaison linéaire des observations s_1, \dots, s_n . Résoudre les équations de krigage revient donc ici à trouver la valeur des coefficients $\lambda_1, \dots, \lambda_n$.

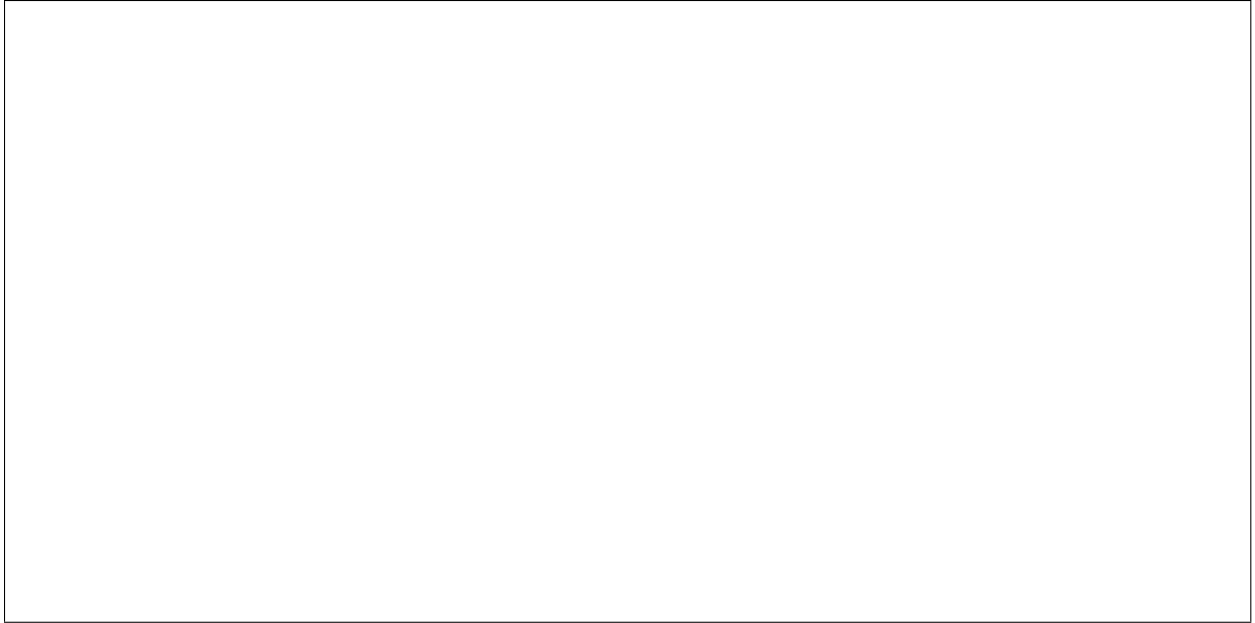
Proposition 5.1 *Les quatre contraintes du krigage conduisent à*

$$\hat{Z}_{s_0} = m + c_0^t \Sigma^{-1} (Z - m \mathbb{1})$$

avec une erreur de prédiction de variance

$$\mathbb{V}(\hat{Z}_{s_0} - Z_{s_0}) = c_{00} - c_0^t \Sigma^{-1} c_0.$$

Preuve.



Ce cas est le plus simple, mais les méthodes de résolution sont similaires pour les autres krigeages. En particulier pour le krigeage ordinaire, m étant inconnue, elle fait aussi partie des valeurs à déterminer. Et la solution dépend de la structure de corrélation, d'où la nécessité de bien l'estimer.

Remarque. On verra en TD une particularité de ce type de krigeage : l'erreur de prédiction est exactement nulle aux sites sur lesquels s'appuient l'estimation. Cela peut compliquer la validation du modèle si la quantité de mesures n'est pas suffisante (car alors il faut en laisser de côté).

Des packages comme **gstat**, **geoR** ou encore **sp** proposent des procédures de krigeage et des représentations graphiques adaptées. L'exemple ci-dessous est construit sur le jeu de données **meuse** : il s'agit de la log-concentration en zinc mesurée dans le sol (à gauche) et interpolée par un krigeage ordinaire.

