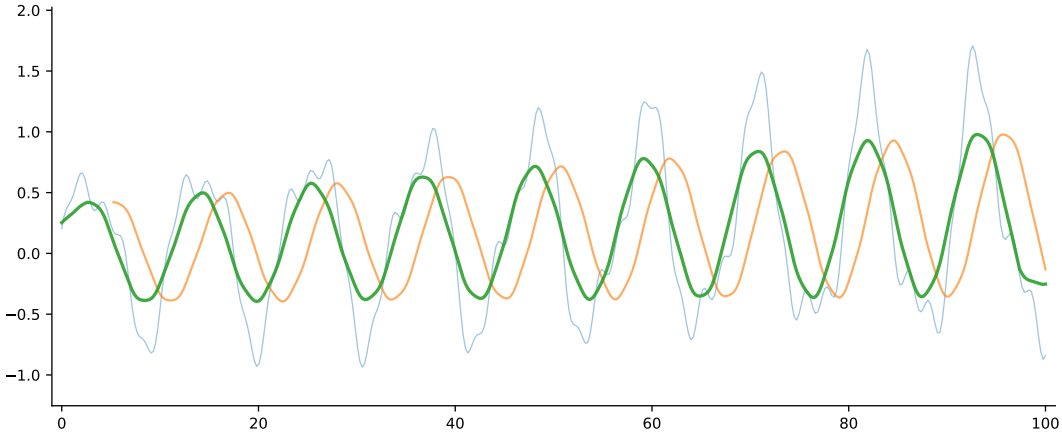
Séries temporelles

Notes de cours

Master 1 Actuariat



Solesne Bourguin

Université Gustave Eiffel

Version de travail
2026–2027

Avant-propos

Ces notes sont une version de travail pour un cours de Master 1 consacré aux séries temporelles. Elles ont deux ambitions simultanées. D'une part, elles doivent donner des bases mathématiques solides : processus stochastiques, stationnarité, autocovariance, filtrage, modèles ARMA, prédiction et estimation. D'autre part, elles doivent rester utilisables dans un cours appliqué, avec des données réelles, des graphiques et du code `R`. Le point de vue adopté est donc le suivant : les outils mathématiques sont introduits au moment où ils permettent de clarifier une difficulté concrète rencontrée sur une série.

Chaque chapitre contient des définitions, résultats, exemples, commentaires pratiques, sources des données, et exercices associés. Les références principales sont Brockwell–Davis pour l'ossature théorique classique, Shumway–Stoffer pour l'approche moderne avec exemples `R`, Aragon pour l'esprit pratique, et les notes de Chafaï–Lévy-Leduc pour le format mathématique en M1.

Références principales

[1] P. J. Brockwell et R. A. Davis, *Introduction to Time Series and Forecasting*, 3^{e} édition, Springer Texts in Statistics, Springer, 2016.

[2] R. H. Shumway et D. S. Stoffer, *Time Series Analysis and Its Applications : With R Examples*, 4^{e} édition, Springer Texts in Statistics, Springer, 2017.

[3] Y. Aragon, *Séries temporelles avec R : méthodes et cas*, Springer Paris, 2011.

[4] D. Chafaï, *Introduction aux séries temporelles*, notes de cours de Master 1 MIDO, Université Paris-Dauphine.

Plan prévisionnel du cours

Chapitre 1. Introduction, tendance, saisonnalité, stationnarité. Processus stochastique, chronogramme, modèle additif, résidus, opérateur retard, opérateurs différence, stationnarité faible et forte, autocovariance, bruit blanc, premiers exemples.

Chapitre 2. Outils `R` et analyse exploratoire. Objets `ts`, dates, sous-séries, graphiques, lag plots, month plots, valeurs manquantes, transformations.

Chapitre 3. Filtrage linéaire et séries linéaires. Filtres, convolution, stabilité, bruit blanc filtré, moyenne mobile infinie.

Chapitre 4. Modèles ARMA stationnaires. MA, AR, ARMA, causalité, inversibilité, autocovariance, ACF, PACF, identification.

Chapitre 5. Séries non stationnaires et ARIMA/SARIMA. Différenciation, racine unitaire, tendance déterministe, saisonnalité, modèles intégrés.

Chapitre 6. Estimation, diagnostic et prévision. Vraisemblance gaussienne, AIC/BIC, résidus, tests de blancheur, intervalles de prévision.

Chapitre 7. Lissage exponentiel et méthodes de prévision. Lissage simple, Holt, Holt–Winters, comparaison avec ARIMA.

Chapitre 8. Ouvertures. Analyse spectrale élémentaire et hétéroscédasticité conditionnelle.

Table des matières

Plan prévisionnel du cours **ii**

1 Introduction : séries observées et premiers modèles **1**
1.1 Séries observées et première lecture graphique . . . 1
1.1.1 Données temporelles et ordre des observations . . . 1
1.1.2 Deux premiers chronogrammes . . . 2
1.1.3 Objectifs d'une analyse de série temporelle . . . 2
1.1.4 Une série mensuelle : AirPassengers . . . 3
1.2 Décomposition d'une série observée . . . 4
1.2.1 Pourquoi séparer plusieurs composantes ? . . . 4
1.2.2 Tendance, saisonnalité et bruit théorique . . . 4
1.2.3 Valeur ajustée et résidu empirique . . . 5
1.2.4 Transformation logarithmique : passer des variations relatives aux variations additives . . . 5
1.2.5 Décomposition empirique de AirPassengers . . . 6
1.2.6 Estimation paramétrique d'une tendance . . . 8
1.2.7 Deux modélisations simples de la saisonnalité . . . 9
1.3 Graphiques pour repérer saisonnalité et dépendance . . . 9
1.3.1 Chronogramme, month plot, lag plot : trois questions différentes . . . 9
1.3.2 Month plot . . . 9
1.3.3 Lag plot . . . 10
1.4 Transformations temporelles : retard et différences . . . 11
1.4.1 Opérateur retard . . . 11
1.4.2 Différence première et différence saisonnière . . . 12
1.4.3 Effet sur les tendances polynomiales . . . 12
1.4.4 Effet sur une saisonnalité . . . 14
1.5 Modélisation probabiliste et stationnarité . . . 14
1.5.1 Processus stochastique et réalisation . . . 14
1.5.2 Moyenne, autocovariance et autocorrélation . . . 15
1.5.3 Stationnarité forte et stationnarité faible . . . 16
1.5.4 Trois exemples fondamentaux . . . 17
1.5.5 Matrices de Toeplitz et autocovariances possibles . . . 19
1.6 Synthèse du chapitre . . . 20

2 Outils `R` pour séries temporelles **21**
2.1 Pourquoi un chapitre sur les objets `R` ? . . . 21
2.2 L'objet `ts` . . . 21
2.2.1 Valeurs, début, fin et fréquence . . . 21
2.2.2 Construire soi-même un objet `ts` . . . 22
2.3 Temps numérique et cycle saisonnier . . . 23
2.3.1 Les fonctions `time()` et `cycle()` . . . 23
2.4 Extraire une sous-série avec `window()` . . . 23

2.5 Retards et différences dans R . . . 24
2.5.1 Retards : attention à la convention de lag() . . . 24
2.5.2 Différences ordinaires et saisonnières . . . 24
2.6 Valeurs manquantes . . . 25
2.6.1 Repérer les valeurs manquantes . . . 25
2.6.2 Faut-il remplacer les valeurs manquantes ? . . . 25
2.7 Synthèse du chapitre . . . 26

3 Filtrage linéaire et séries linéaires 28
3.1 Pourquoi introduire les filtres ? . . . 28
3.2 Filtres linéaires finis . . . 29
3.2.1 Définition et premiers exemples . . . 29
3.2.2 Support, causalité et filtres centrés . . . 30
3.2.3 Linéarité . . . 31
3.3 Effet d'un filtre sur des composantes simples . . . 31
3.3.1 Constantes et somme des coefficients . . . 31
3.3.2 Tendances polynomiales et différences . . . 32
3.3.3 Saisonnalité et différence saisonnière . . . 32
3.4 Des filtres finis aux filtres infinis . . . 33
3.5 Composition des filtres et convolution . . . 35
3.6 Filtrage d'un processus stationnaire et bruit blanc filtré . . . 36
3.6.1 Cas des filtres finis . . . 36
3.6.2 Bruit blanc filtré . . . 36
3.7 Séries linéaires causales . . . 37
3.7.1 Cas particulier : poids géométriques . . . 39
3.8 Filtres dans R . . . 40
3.9 Synthèse du chapitre . . . 40

4 Modèles MA, AR et ARMA stationnaires 42
4.1 Pourquoi introduire les modèles ARMA ? . . . 42
4.2 Processus moyenne mobile MA . . . 43
4.2.1 Définition et intuition . . . 43
4.2.2 Stationnarité et autocovariance d'un MA(q) . . . 44
4.2.3 Le cas MA(1) et le problème d'inversibilité . . . 44
4.3 Processus autorégressifs AR . . . 45
4.3.1 Définition et comparaison avec les MA . . . 45
4.3.2 Le cas AR(1) . . . 46
4.3.3 AR(p), racines et développement rationnel . . . 47
4.3.4 Calcul des coefficients par identification . . . 48
4.3.5 Équations de Yule–Walker . . . 48
4.4 Des AR et MA aux ARMA : les filtres rationnels . . . 49
4.4.1 Pourquoi combiner AR et MA ? . . . 49
4.4.2 Causalité, inversibilité et fractions rationnelles . . . 50
4.5 ACF, PACF et identification . . . 51
4.5.1 ACF théorique et ACF empirique . . . 51
4.5.2 Pourquoi l'ACF ne suffit pas toujours . . . 51
4.5.3 Autocorrélation partielle : définition et intuition . . . 52
4.5.4 Méthode pratique d'identification . . . 53
4.6 Simulation et premières commandes R . . . 53
4.7 Synthèse du chapitre . . . 54

5 Séries non stationnaires et modèles ARIMA/SARIMA 56
5.1 Pourquoi quitter les modèles ARMA stationnaires ? . . . 56
5.2 Deux formes de non-stationnarité . . . 57
5.2.1 Tendance déterministe . . . 57
5.2.2 Tendance stochastique et racine unitaire . . . 58
5.2.3 La marche aléatoire . . . 59
5.3 Différenciation et séries intégrées . . . 59
5.3.1 Différence première : passer des niveaux aux variations . . . 59
5.3.2 Séries intégrées . . . 61
5.3.3 Sur-différencier une série . . . 61
5.4 Modèles ARIMA . . . 62
5.4.1 Pourquoi introduire les ARIMA ? . . . 62
5.4.2 Définition . . . 62
5.5 Différenciation saisonnière et modèles SARIMA . . . 63
5.5.1 Pourquoi une différence saisonnière ? . . . 63
5.5.2 Effet sur une saisonnalité déterministe . . . 64
5.5.3 Définition d'un SARIMA . . . 65
5.6 Comment choisir les différences ? . . . 65
5.7 Premières commandes `R` . . . 66
5.8 Synthèse du chapitre . . . 66

6 Estimation, diagnostic et prévision 67
6.1 Pourquoi un nouveau chapitre ? . . . 67
6.2 Estimer les paramètres d'un modèle . . . 67
6.2.1 Du modèle théorique au modèle ajusté . . . 67
6.2.2 Un cas complètement explicite : AR(1) par moindres carrés . . . 68
6.2.3 Vraisemblance gaussienne . . . 69
6.3 Comparer plusieurs modèles : AIC et BIC . . . 69
6.3.1 Pourquoi un critère de comparaison ? . . . 69
6.4 Diagnostic : que doivent vérifier les résidus ? . . . 70
6.4.1 Le rôle du diagnostic . . . 70
6.4.2 Tests portemanteau : Box–Pierce et Ljung–Box . . . 70
6.5 Prévision . . . 71
6.5.1 Prévision ponctuelle et erreur de prévision . . . 71
6.5.2 Cas explicite : prévision d'un AR(1) . . . 71
6.6 Cas pratique : AirPassengers . . . 72
6.7 Validation hors-échantillon . . . 72
6.8 Synthèse du chapitre . . . 74

7 Lissage exponentiel et méthodes de prévision récursives 75
7.1 Pourquoi introduire le lissage exponentiel ? . . . 75
7.2 Du lissage par moyenne mobile au lissage exponentiel . . . 76
7.3 Lissage exponentiel simple . . . 77
7.4 Choix du paramètre de lissage . . . 78
7.5 Méthode de Holt : niveau et tendance . . . 78
7.6 Méthode de Holt–Winters : niveau, tendance et saisonnalité . . . 79
7.6.1 Version additive . . . 79
7.6.2 Version multiplicative . . . 80
7.7 Cas pratique : AirPassengers . . . 80
7.8 Diagnostic et limites . . . 80
7.9 Synthèse du chapitre . . . 81

8 Ouvertures : simulation, analyse spectrale et volatilité conditionnelle 82
8.1 Pourquoi terminer par ces ouvertures ? 82
8.2 Simuler pour comprendre un modèle 82
8.3 Période, fréquence et cycles 83
8.4 Transformée de Fourier discrète et périodogramme 84
8.5 Densité spectrale d'un processus stationnaire 85
8.6 Cas pratique : le cycle des taches solaires 86
8.7 Variabilité changeante et rendements financiers 87
8.8 Modèles ARCH et GARCH 88
8.9 Synthèse du chapitre 89

Chapitre 1

Introduction : séries observées et premiers modèles

Mots-clés. Série temporelle ; chronogramme ; tendance ; saisonnalité ; modèle additif ; valeur ajustée ; résidu empirique ; transformation logarithmique ; month plot ; lag plot ; opérateur retard ; différence ; processus stochastique ; réalisation ; stationnarité ; autocovariance ; autocorrélation ; bruit blanc.

1.1 Séries observées et première lecture graphique

1.1.1 Données temporelles et ordre des observations

Une *série temporelle* est une suite d'observations d'un phénomène, ordonnées par le temps. Contrairement à un échantillon statistique usuel, l'ordre des observations ne peut pas être ignoré. Les valeurs successives peuvent être liées : la température de demain dépend souvent de celle d'aujourd'hui, le trafic du mois prochain dépend de la saison, et le nombre de sinistres d'une année dépend parfois de l'évolution réglementaire ou économique.

Dans ce cours, les observations seront le plus souvent réelles et régulièrement espacées dans le temps. On les notera

$$x_1, x_2, \dots, x_n.$$

La lettre minuscule x_t désigne donc une *valeur observée*. Plus tard, la lettre majuscule X_t désignera une *variable aléatoire* censée modéliser l'observation faite au temps t. Cette distinction est importante : une série de données est finie ; un modèle probabiliste est souvent une suite infinie de variables aléatoires.

Définition 1.1.1: Chronogramme

Le *chronogramme* d'une série observée $x_1, \dots, x_n$ est le graphe des points (t, x_t), où l'abscisse représente le temps et l'ordonnée la valeur observée. Lorsque les dates sont régulièrement espacées, on relie souvent les points consécutifs pour rendre la dynamique visuellement lisible.

Le chronogramme n'est pas seulement une illustration. C'est le premier outil de diagnostic. Il permet de repérer des phénomènes qui ne sont pas interchangeables : une tendance globale, une périodicité, une rupture, une valeur atypique, ou encore une variance qui change au cours du temps. Ces éléments doivent être distingués avant tout calcul, car ils appellent des traitements différents.

1.1.2 Deux premiers chronogrammes

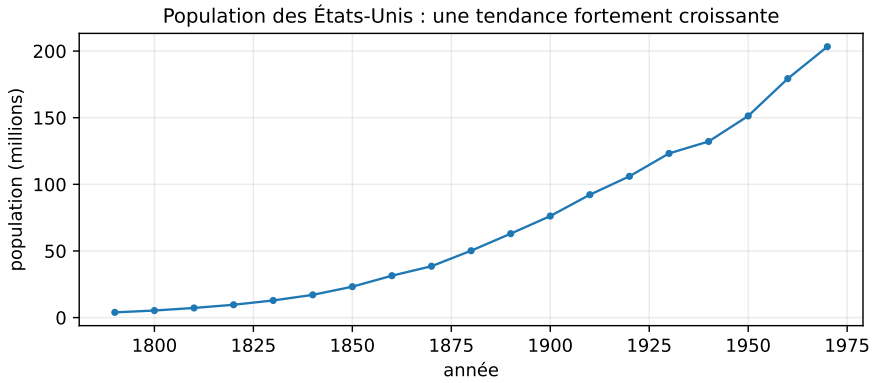


Figure 1.1 – Population des États-Unis, en millions d'habitants. Données : série `uspop`, fournie avec le package `datasets` de `R`. Le message visuel principal est une tendance croissante très marquée : le niveau de la série dépend fortement du temps.

Comment lire ce graphique ?

Sur ce graphique, il ne faut pas chercher d'abord un modèle compliqué. La première information est simple : la série augmente avec le temps. Une modélisation raisonnable devra donc contenir une composante de tendance, ou transformer la série pour éliminer cette tendance.

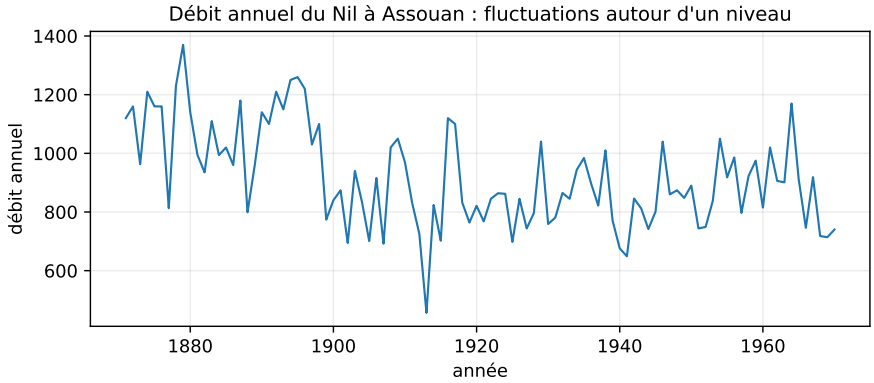


Figure 1.2 – Débit annuel du Nil à Assouan. Données : série `Nile`, fournie avec le package `datasets` de `R`. Contrairement à la population américaine, la série fluctue autour d'un niveau, avec une possible rupture autour de la fin du XIXe siècle.

Comment lire ce graphique ?

La série du Nil illustre une autre situation : il n'y a pas de tendance lisse aussi évidente que pour `uspop`. On observe plutôt des fluctuations et peut-être une rupture. Un même outil ne répond donc pas à toutes les séries : il faut d'abord regarder la série avant de choisir un modèle.

1.1.3 Objectifs d'une analyse de série temporelle

L'étude d'une série temporelle poursuit en général plusieurs objectifs :

1) *décrire* le phénomène : tendance, saisonnalité, rupture, valeurs atypiques ;
2) *modéliser* la dépendance temporelle par un processus aléatoire ;
3) *prévoir* des valeurs futures à partir de la partie observée ;
4) *expliquer* la série à l'aide de variables extérieures : température, calendrier, prix, réglementation, etc.

Ces objectifs ne conduisent pas toujours au même traitement. Pour une prévision à court terme, on peut accepter un modèle très empirique s'il prévoit bien. Pour comprendre un phénomène, on demandera au modèle d'être interprétable. Dans tous les cas, la première étape reste l'exploration graphique.

Il faut aussi garder en tête qu'une série temporelle contient moins d'information qu'il n'y paraît. On ne dispose pas de plusieurs répétitions indépendantes de la même expérience : on observe une seule histoire. Toute la difficulté consiste donc à exploiter l'ordre temporel sans confondre une régularité visible sur une période limitée avec une loi générale du phénomène.

1.1.4 Une série mensuelle : AirPassengers

Exemple 1.1.2: Nombre mensuel de passagers aériens internationaux

La série `AirPassengers`, fournie avec le package `datasets` de `R`, donne le nombre mensuel de passagers aériens internationaux de 1949 à 1960, en milliers. Elle est souvent utilisée comme exemple pédagogique, notamment pour illustrer tendance, saisonnalité et transformation logarithmique. Dans l'aide de `R`, elle est associée aux exemples classiques de Box–Jenkins.

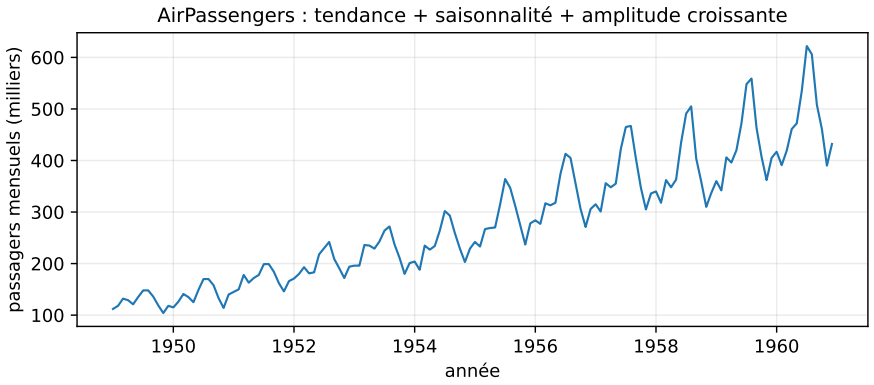


Figure 1.3 – Série `AirPassengers`. On voit simultanément une tendance croissante, une saisonnalité annuelle et une amplitude saisonnière qui augmente avec le niveau de la série. Ces trois éléments justifient l'utilisation d'une transformation et d'une décomposition.

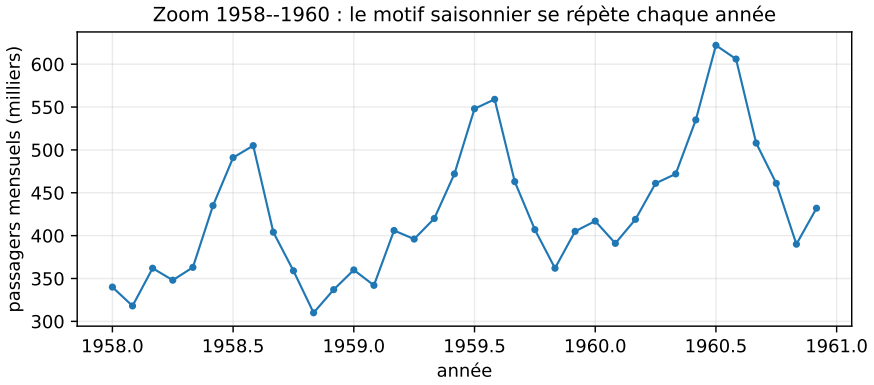


Figure 1.4 – Zoom sur `AirPassengers`, 1958–1960. Le motif se répète d'une année à l'autre : les creux et les pics apparaissent à des périodes comparables de l'année. C'est ce que l'on appellera une saisonnalité.

La figure de `AirPassengers` montre déjà trois difficultés classiques :

1) la série ne fluctue pas autour d'une constante ;
2) les mois comparables d'années différentes se ressemblent ;
3) les oscillations saisonnières deviennent plus grandes lorsque le niveau de la série augmente.

Ces trois observations conduisent à trois outils de base : une tendance, une saisonnalité, et parfois une transformation d'échelle. La suite du chapitre consiste à donner un sens précis à ces trois mots. L'objectif n'est pas encore de prévoir la série, mais de comprendre ce que l'on voit et de préparer un modèle plus simple pour la partie restante.

1.2 Décomposition d'une série observée

1.2.1 Pourquoi séparer plusieurs composantes ?

Une série temporelle brute mélange souvent plusieurs phénomènes. Dans `AirPassengers`, l'augmentation du trafic aérien sur plusieurs années n'a pas la même nature que les variations régulières entre les mois d'hiver et les mois d'été. Ces effets structurés n'ont pas non plus la même nature que les petites fluctuations irrégulières d'un mois à l'autre.

L'idée d'une décomposition est donc de remplacer une série compliquée par plusieurs objets plus simples :

$$\text{série observée} = \text{niveau lent} + \text{effet saisonnier} + \text{reste}.$$

Ce n'est pas seulement une opération graphique. C'est une étape de modélisation : on décide ce que l'on considère comme structurel et ce que l'on laisse dans le reste.

1.2.2 Tendance, saisonnalité et bruit théorique

Définition 1.2.1: Tendance et saisonnalité

On appelle *tendance* une fonction $t \mapsto m_t$ qui varie lentement avec le temps. On appelle *saisonnalité de période d* une fonction $t \mapsto s_t$ telle que

$$s_{t+d} = s_t$$

pour tout t. Dans le cas mensuel, une saisonnalité annuelle correspond à $d = 12$. Dans le cas trimestriel, une saisonnalité annuelle correspond à $d = 4$.

La tendance n'est pas nécessairement une droite : c'est le mouvement de fond que l'on verrait en regardant la série de loin. La saisonnalité est au contraire une répétition à une échelle plus courte : mois de l'année, jour de la semaine, trimestre, etc. Ces deux composantes sont utiles parce qu'elles séparent deux formes très différentes de régularité temporelle.

Définition 1.2.2: Modèle additif tendance-saisonnalité-bruit

Une *décomposition additive* d'un processus $X = (X_t)$ est une écriture de la forme

$$X_t = m_t + s_t + Z_t, \tag{1.1}$$

où :

i) m_t est la tendance ;

ii) s_t est la composante saisonnière ;

iii) Z_t est le *bruit théorique*, c'est-à-dire la partie que le modèle ne décrit pas par la tendance et la saisonnalité.

Cette écriture ne prétend pas que le monde se décompose réellement en trois morceaux parfaitement séparés. Elle fournit plutôt une grille de lecture. On décide d'appeler tendance ce qui évolue lentement, saisonnalité ce qui se répète, et bruit ce qui n'est pas expliqué par ces deux composantes. La qualité de cette décision se jugera ensuite sur les résidus.

1.2.3 Valeur ajustée et résidu empirique

Le modèle (1.1) contient un bruit théorique Z_t, mais ce bruit n'est pas observé directement. En pratique, on calcule des approximations $\widehat{m}_t$ et $\widehat{s}_t$, puis on regarde ce qui reste. Ce reste est le résidu.

Définition 1.2.3: Valeur ajustée et résidu empirique

Soit $x_1, \dots, x_n$ une série observée. Supposons qu'une méthode d'ajustement fournisse, pour chaque temps t, une *valeur ajustée* $\widehat{x}_t$, c'est-à-dire la valeur que le modèle attribue à l'observation x_t. Le *résidu empirique* au temps t est

$$\widehat{z}_t = x_t - \widehat{x}_t.$$

Dans une décomposition additive, la valeur ajustée est typiquement

$$\widehat{x}_t = \widehat{m}_t + \widehat{s}_t,$$

et le résidu vaut donc

$$\widehat{z}_t = x_t - \widehat{m}_t - \widehat{s}_t.$$

Point de vigilance

Il faut distinguer trois objets : le bruit théorique Z_t, la valeur ajustée $\widehat{x}_t$, et le résidu empirique $\widehat{z}_t$. Le bruit appartient au modèle probabiliste. La valeur ajustée et le résidu sont calculés à partir des données. Si l'on change de modèle ou de méthode d'ajustement, les résidus changent.

Les résidus sont importants parce qu'ils montrent ce que le modèle n'a pas expliqué. Si l'on retire une tendance et une saisonnalité raisonnables, mais que les résidus présentent encore une tendance, une saisonnalité ou une forte dépendance temporelle, alors le modèle est incomplet. Dans les modèles classiques de ce cours, on espère souvent que les résidus ressemblent à une série stationnaire, voire à un bruit blanc.

Concrètement, les résidus jouent le rôle d'un contrôle qualité. Un ajustement peut sembler satisfaisant sur la série observée parce qu'il suit correctement les grands mouvements, mais échouer à expliquer la dynamique fine. Examiner les résidus revient à demander : "ai-je vraiment retiré toute la structure visible, ou reste-t-il quelque chose à modéliser ?"

1.2.4 Transformation logarithmique : passer des variations relatives aux variations additives

Avant de décomposer `AirPassengers`, il est naturel de changer d'échelle. Sur la série brute, l'écart entre un mois creux et un mois haut est plus grand à la fin de la période qu'au début. Une partie de cette augmentation vient simplement du fait que le niveau global de la série est plus élevé.

Supposons qu'une série ait une variation saisonnière d'environ 10% autour de son niveau. Autour du niveau 100, une variation de 10% correspond à une amplitude d'environ 10. Autour du niveau 500, elle correspond à une amplitude d'environ 50. L'amplitude absolue augmente, alors que l'amplitude relative est la même.

Le logarithme transforme précisément les variations multiplicatives, ou relatives, en variations additives. Si l'on écrit localement

$$x_t = m_t(1 + u_t),$$

où m_t est le niveau moyen et u_t une fluctuation relative, alors

$$\log(x_t) = \log(m_t) + \log(1 + u_t).$$

Lorsque u_t est petit, $\log(1 + u_t) \simeq u_t$. Ainsi, une oscillation de 10% a à peu près la même taille sur l'échelle logarithmique, que le niveau soit 100 ou 500.

Idée à retenir

La question à se poser avant de prendre un logarithme est donc : les fluctuations sont-elles plutôt de taille absolue constante, ou plutôt de taille relative constante ? Dans le second cas, l'échelle logarithmique est souvent plus naturelle.

Il ne faut pas voir le logarithme comme une recette automatique. Si les écarts autour du niveau moyen sont à peu près constants en valeur absolue, le logarithme peut être inutile, voire rendre l'interprétation moins directe. Dans `AirPassengers`, il est naturel parce que les fluctuations saisonnières augmentent avec le niveau moyen de la série.

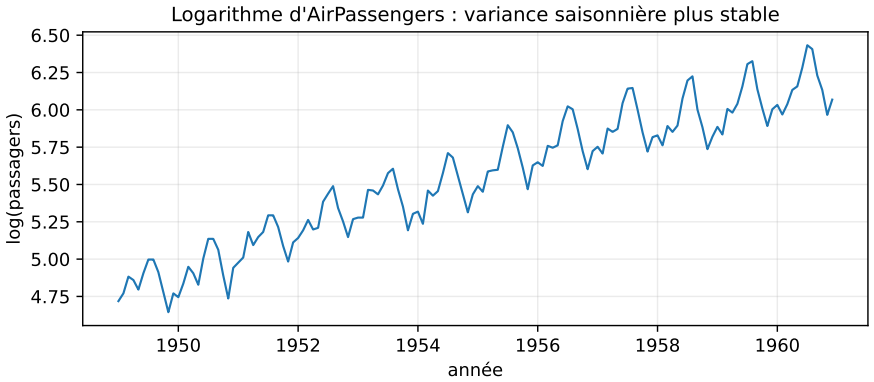


Figure 1.5 – Transformation logarithmique de `AirPassengers`. Le graphique représente $y_t = \log(x_t)$, où x_t est le nombre mensuel de passagers. Sur cette nouvelle échelle, les oscillations saisonnières du début et de la fin de période deviennent plus comparables, car le logarithme transforme les variations relatives en variations approximativement additives.

1.2.5 Décomposition empirique de AirPassengers

On applique maintenant une décomposition à $y_t = \log(x_t)$, où x_t est la série `AirPassengers`. On utilise ici la fonction `stl()` de `R`. Pour ce premier chapitre, on l'utilise comme un outil exploratoire : elle sépare automatiquement une série en une tendance, une composante saisonnière et un reste. Le détail de la méthode n'est pas l'objet du cours à ce stade ; ce qui compte est de comprendre ce que signifient les trois sorties produites. À partir de la série observée y_t, elle fournit trois séries estimées, notées

$$\widehat{m}_t, \qquad \widehat{s}_t, \qquad \widehat{z}_t,$$

qui vérifient l'identité de reconstruction

$$y_t = \widehat{m}_t + \widehat{s}_t + \widehat{z}_t.$$

Le terme $\widehat{m}_t$ est interprété comme la tendance estimée, $\widehat{s}_t$ comme la saisonnalité estimée, et $\widehat{z}_t$ comme le résidu empirique. Cette décomposition est donc une manière automatique de construire les valeurs ajustées

$$\widehat{y}_t = \widehat{m}_t + \widehat{s}_t$$

et les résidus $\widehat{z}_t = y_t - \widehat{y}_t$.

Dans les graphiques ci-dessous, la saisonnalité estimée est contrainte à être strictement périodique de période douze : l'effet de janvier est le même chaque année, l'effet de février aussi, etc. Ce choix est adapté à une première lecture de `AirPassengers`, où le motif annuel semble très régulier.

Une bonne façon de lire cette décomposition est de partir de la question suivante : que reste-t-il si l'on enlève ce que l'on sait déjà nommer ? La tendance décrit le mouvement de fond, la saisonnalité décrit le motif annuel, et le résidu est la partie que ces deux descriptions ne suffisent pas à expliquer.

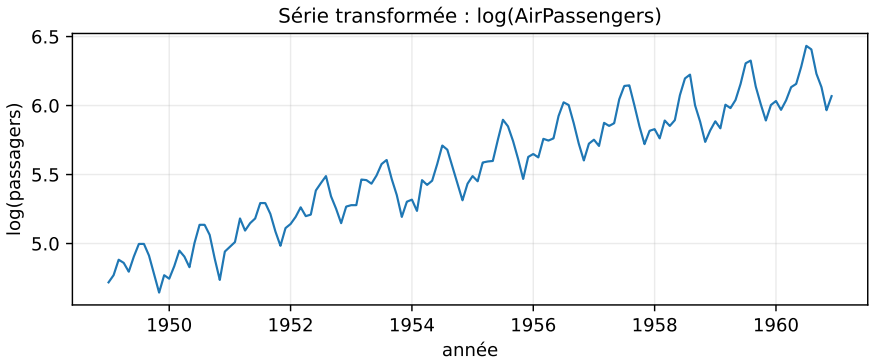


Figure 1.6 – Série à décomposer. Ce graphique représente $y_t = \log(\texttt{AirPassengers}_t)$. C'est cette série transformée, et non la série brute, que l'on écrit sous la forme $y_t = \widehat{m}_t + \widehat{s}_t + \widehat{z}_t$.

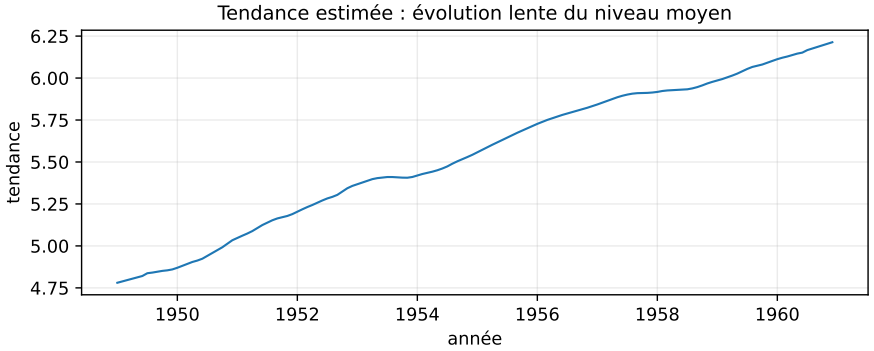


Figure 1.7 – Tendance estimée $\widehat{m}_t$. Elle représente l'évolution lente du niveau moyen de $\log(\texttt{AirPassengers})$, sans reproduire les oscillations mensuelles.

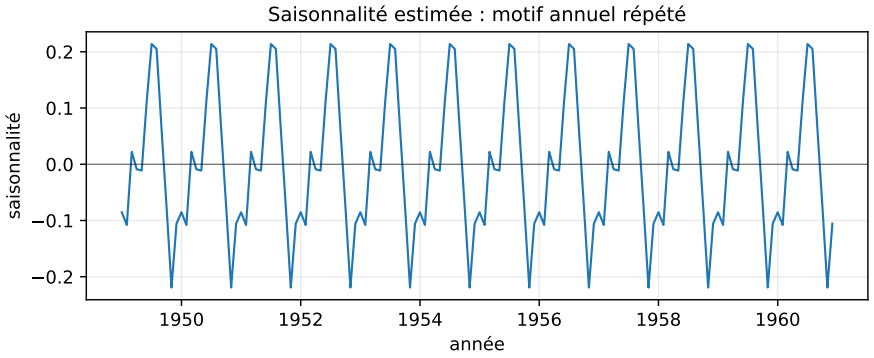


Figure 1.8 – Composante saisonnière estimée $\widehat{s}_t$. Le même motif de douze mois est répété chaque année. Les valeurs positives correspondent aux mois au-dessus de la tendance ; les valeurs négatives correspondent aux mois au-dessous.

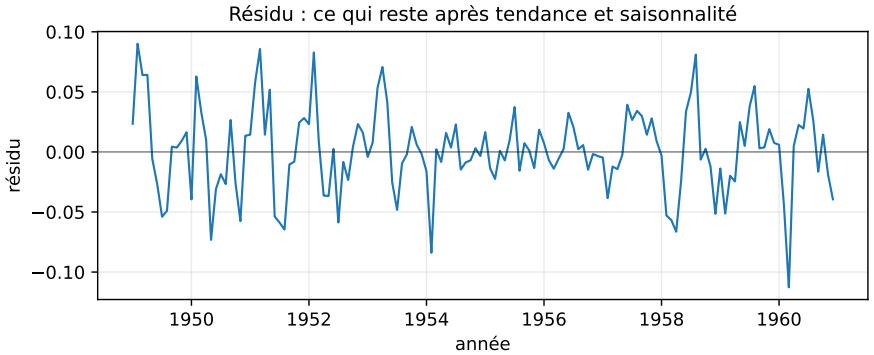


Figure 1.9 – Résidu de la décomposition $\widehat{z}_t = y_t - \widehat{m}_t - \widehat{s}_t$. Il s'agit de la partie restante après retrait de la tendance et de la saisonnalité estimées.

Comment lire ce graphique ?

Une décomposition se lit toujours dans l'ordre suivant : d'abord la série observée, ensuite la tendance, ensuite la saisonnalité, enfin les résidus. Le dernier graphique n'est pas un détail technique : il répond à la question suivante. Après avoir retiré ce que l'on comprend facilement, reste-t-il encore une dépendance temporelle ?

1.2.6 Estimation paramétrique d'une tendance

La décomposition précédente construit automatiquement une tendance et une saisonnalité estimées. On présente maintenant une méthode plus explicite : on choisit à l'avance une famille de fonctions pour la tendance, puis on estime ses paramètres. L'intérêt est double. D'abord, le modèle devient interprétable : une droite représente une croissance approximativement linéaire, un polynôme de degré deux une courbure, etc. Ensuite, les résidus ont une définition très claire : ce sont les écarts entre la série et la tendance ajustée. Supposons, pour commencer, qu'il n'y ait pas de saisonnalité et que l'on observe

$$x_t = m_t + z_t, \qquad t = 1, \dots, n.$$

Un modèle simple consiste à supposer que la tendance est polynomiale :

$$m_t = a_0 + a_1 t + \cdots + a_p t^p.$$

On estime $a_0, \dots, a_p$ par moindres carrés. Notons

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \qquad a = \begin{pmatrix} a_0 \\ \vdots \\ a_p \end{pmatrix}, \qquad A = \begin{pmatrix} 1 & t_1 & \cdots & t_1^p \\ \vdots & \vdots & & \vdots \\ 1 & t_n & \cdots & t_n^p \end{pmatrix}.$$

Alors le critère des moindres carrés est

$$\|X - Aa\|_2^2 .$$

Si A est de rang plein, l'estimateur est

$$\widehat{a} = \left(A^\top A\right)^{-1} A^\top X.$$

Les valeurs ajustées et les résidus associés sont

$$\widehat{X} = A\widehat{a}, \qquad \widehat{z} = X - A\widehat{a}.$$

Démonstration. La fonction à minimiser est $\varphi(a) = \|X - Aa\|_2^2$. Pour $h \in \mathbb{R}^{p+1}$,

$$\begin{aligned} \varphi(a+h) - \varphi(a) &= \|X - Aa - Ah\|_2^2 - \|X - Aa\|_2^2 \\ &= -2\left\langle A^\top (X - Aa), h \right\rangle + \|Ah\|_2^2 . \end{aligned}$$

Donc $\nabla\varphi(a) = 2A^\top Aa - 2A^\top X$. Si $A^\top A$ est inversible, l'unique point critique, donc l'unique minimiseur, est $\widehat{a} = \left(A^\top A\right)^{-1} A^\top X$. □

Point de vigilance

Une tendance ajustée par moindres carrés n'est pas une vérité. C'est une approximation choisie dans une famille de fonctions. Les résidus indiquent ce que cette approximation n'explique pas. Si les résidus gardent une forte structure, le modèle de tendance est probablement insuffisant.

1.2.7 Deux modélisations simples de la saisonnalité

Pour une série mensuelle, une manière très concrète de modéliser la saisonnalité est d'attribuer un effet propre à chaque mois. On écrit par exemple

$$x_t = m_t + \sum_{j=1}^{12} \alpha_j \mathbf{1}_{\{t \text{ est dans le mois } j\}} + z_t.$$

On impose souvent une contrainte comme $\alpha_1 + \cdots + \alpha_{12} = 0$ pour séparer le niveau moyen de l'effet saisonnier.

Une autre méthode consiste à utiliser des fonctions trigonométriques. Pour une saisonnalité de période d, les fonctions

$$\cos\left(\frac{2\pi kt}{d}\right), \qquad \sin\left(\frac{2\pi kt}{d}\right)$$

sont naturellement de période d. On peut donc écrire, pour un entier K,

$$s_t = \sum_{k=1}^{K} \left(a_k \cos\left(\frac{2\pi kt}{d}\right) + b_k \sin\left(\frac{2\pi kt}{d}\right)\right).$$

Avant de choisir une écriture, il faut se demander ce que l'on veut faire de la saisonnalité. Si l'objectif est d'interpréter les mois un par un, les indicatrices sont naturelles. Si l'objectif est de décrire un cycle régulier avec peu de paramètres, les fonctions trigonométriques sont souvent plus efficaces.

Remarque 1.2.4: Deux lectures de la saisonnalité

Les indicatrices mensuelles sont très interprétables : elles permettent de dire que tel mois est en moyenne au-dessus ou au-dessous du niveau moyen. Les fonctions trigonométriques sont plus compactes et annoncent l'analyse fréquentielle, mais elles sont parfois moins immédiates pour une première analyse appliquée.

1.3 Graphiques pour repérer saisonnalité et dépendance

1.3.1 Chronogramme, month plot, lag plot : trois questions différentes

Le chronogramme est indispensable, mais il ne suffit pas toujours. Deux graphiques sont particulièrement utiles au début d'une analyse : le *month plot*, qui aide à lire la saisonnalité, et le *lag plot*, qui aide à lire la dépendance temporelle. Ces graphiques ne remplacent pas le chronogramme : ils répondent à des questions différentes.

Idée à retenir

Le chronogramme répond à la question : "comment la série évolue-t-elle au cours du temps ?" Le month plot répond à la question : "à mois fixé, observe-t-on un comportement systématique ?" Le lag plot répond à la question : "une observation est-elle liée aux observations précédentes ?"

1.3.2 Month plot

Définition 1.3.1: Month plot

Pour une série mensuelle, le *month plot* représente les observations en regroupant les valeurs qui correspondent au même mois de l'année. Il sert à visualiser une saisonnalité annuelle et à comparer les mois entre eux. Concrètement, on ne trace plus la série dans son ordre chronologique complet : on compare les groupes "janvier", "février", ..., "décembre".

Le month plot est particulièrement utile quand le chronogramme donne l'impression d'un motif annuel, mais que ce motif est difficile à comparer d'une année à l'autre. Il réorganise les données pour faire apparaître la question saisonnière : janvier ressemble-t-il aux autres janvier ? août est-il systématiquement plus haut que février ?

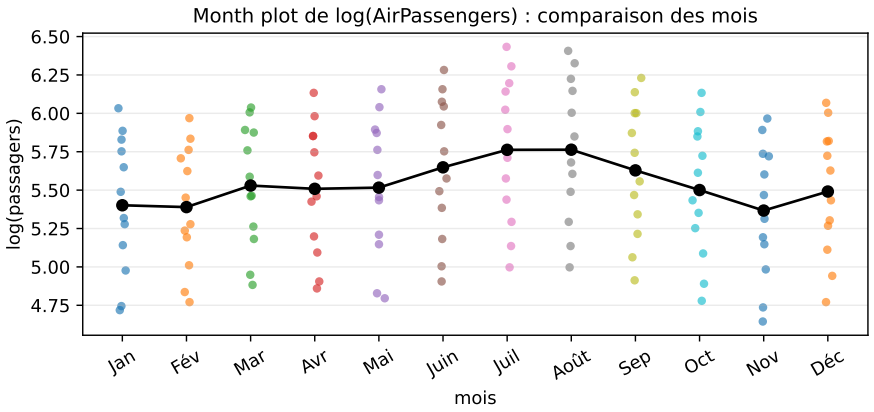


Figure 1.10 – Month plot de log(`AirPassengers`). Chaque point correspond à un mois observé une année donnée ; la ligne relie les valeurs d'un même mois au cours des années. On voit que juillet et août sont systématiquement plus élevés que les mois d'hiver.

1.3.3 Lag plot

Définition 1.3.2: Lag plot

Pour un entier $h \geq 1$, le *lag plot de retard* h est le nuage de points

$$(x_t, x_{t+h}), \qquad t = 1, \ldots, n-h.$$

Chaque point compare donc une observation à une observation future située h dates plus tard. Si le nuage est allongé autour d'une droite croissante, les valeurs séparées de h dates ont tendance à être simultanément grandes ou simultanément petites. Le lag plot sert ainsi à visualiser la dépendance entre la série et la même série décalée de h dates.

Le lag plot ne mesure pas encore l'autocorrélation au sens probabiliste : il donne une image empirique de la relation entre x_t et x_{t+h}. Un nuage très dispersé suggère une faible dépendance linéaire au retard h. Un nuage allongé suggère au contraire que les valeurs séparées de h dates portent une information l'une sur l'autre.

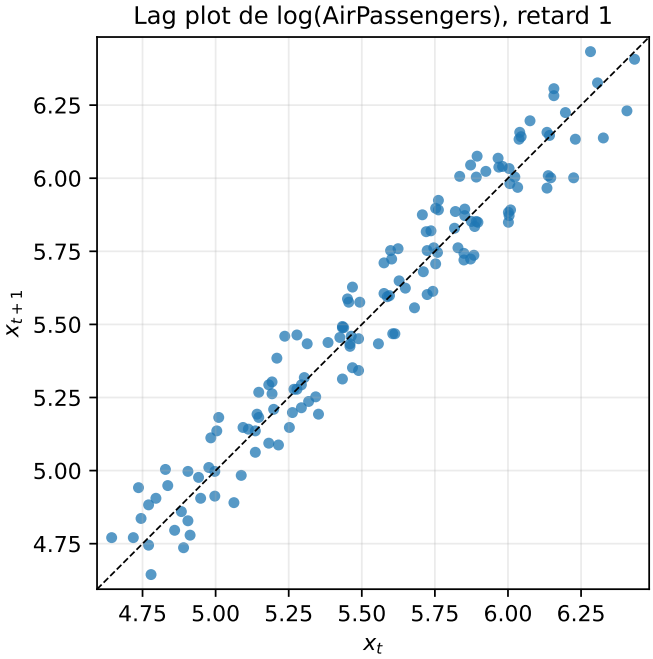


Figure 1.11 – Lag plot de log(`AirPassengers`) au retard 1. Chaque point compare un mois au mois suivant. Le nuage croissant indique que deux mois consécutifs ont tendance à avoir des niveaux proches.

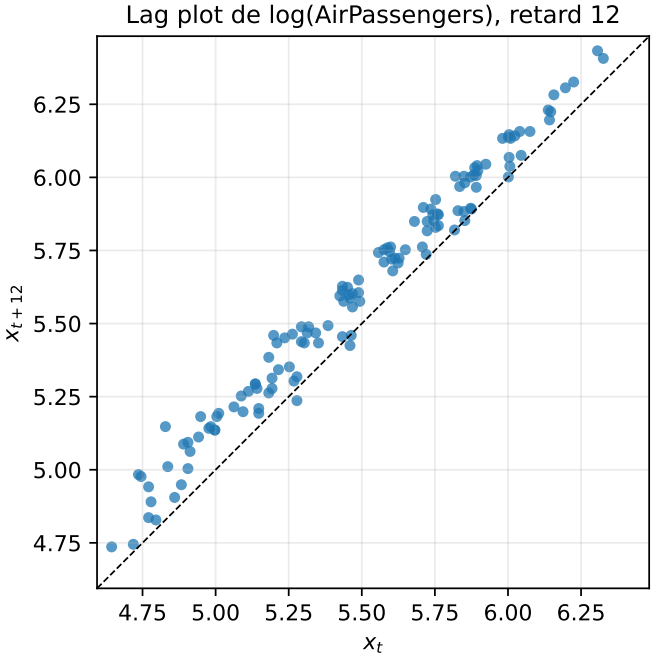


Figure 1.12 – Lag plot de log(`AirPassengers`) au retard 12. Chaque point compare un mois avec le même mois de l'année suivante. L'alignement marqué révèle la forte persistance annuelle de la série.

1.4 Transformations temporelles : retard et différences

1.4.1 Opérateur retard

Les notations deviennent rapidement lourdes si l'on écrit toujours les décalages temporels à la main. L'opérateur retard permet de rendre les transformations lisibles. Il faut le comprendre comme un déplacement de toute la série : on prend la courbe entière et on la décale d'un cran vers le passé.

Définition 1.4.1: Opérateur retard

Soit $X = (X_t)_{t\in\mathbb{Z}}$ une série, déterministe ou aléatoire. L'*opérateur retard*, noté B, est l'opérateur qui transforme X en la série BX définie par

$$(BX)_t = X_{t-1}, \qquad t \in \mathbb{Z}.$$

Par récurrence, pour tout entier $h \geq 1$,

$$\left(B^h X\right)_t = X_{t-h}, \qquad t \in \mathbb{Z}.$$

Dans les calculs, on utilisera souvent l'abréviation usuelle

$$BX_t = X_{t-1}, \qquad B^h X_t = X_{t-h},$$

qui signifie rigoureusement $(BX)_t = X_{t-1}$ et $\left(B^h X\right)_t = X_{t-h}$. Autrement dit, B agit sur la série indexée par le temps, et non sur une variable aléatoire isolée.

1.4.2 Différence première et différence saisonnière

Définition 1.4.2: Opérateurs différence

L'*opérateur différence première* Δ transforme une série X en la série ΔX définie par

$$(\Delta X)_t = X_t - X_{t-1} = ((1-B)\,X)_t\,.$$

Pour un entier $d \geq 1$, l'*opérateur différence saisonnière de période d* transforme X en la série $\Delta_d X$ définie par

$$(\Delta_d X)_t = X_t - X_{t-d} = \left(\left(1-B^d\right)X\right)_t\,.$$

On écrira souvent, par abus de notation,

$$\Delta X_t = (1-B)\,X_t, \qquad \Delta_d X_t = \left(1-B^d\right)X_t.$$

L'idée pédagogique est la suivante : différencier une série, c'est remplacer son niveau par son accroissement. Si une série augmente régulièrement, ses accroissements sont plus stables que ses niveaux. Pour les séries saisonnières, la différence saisonnière compare une valeur à la valeur observée une période plus tôt.

Cette opération est très fréquente car beaucoup de séries non stationnaires deviennent plus régulières une fois exprimées en variations. Par exemple, le niveau d'une population peut croître pendant longtemps, alors que son accroissement annuel peut fluctuer autour d'un ordre de grandeur plus stable.

1.4.3 Effet sur les tendances polynomiales

Proposition 1.4.3: Effet de la différence sur une tendance polynomiale

Si p est un polynôme de degré r, alors Δp est un polynôme de degré au plus $r-1$. En particulier, $\Delta^{r+1} p = 0$.

Démonstration. Il suffit de le vérifier sur les monômes. Pour $p(t) = t^r$,

$$\Delta p(t) = t^r - (t-1)^r\,.$$

Par la formule du binôme, les termes en t^r s'annulent et il reste un polynôme de degré au plus $r-1$. En itérant, on obtient $\Delta^{r+1} p = 0$. □

Ce résultat explique pourquoi la différenciation est si présente en séries temporelles. Elle n'est pas seulement une manipulation numérique : elle agit algébriquement sur les tendances polynomiales. Une tendance linéaire devient constante ; une tendance quadratique devient linéaire ; et ainsi de suite.

Exemple 1.4.4: Différenciation d'une tendance linéaire

Si $X_t = a + bt + Z_t$, alors $\Delta X_t = b + Z_t - Z_{t-1}$. La constante a a disparu et la tendance linéaire est devenue constante.

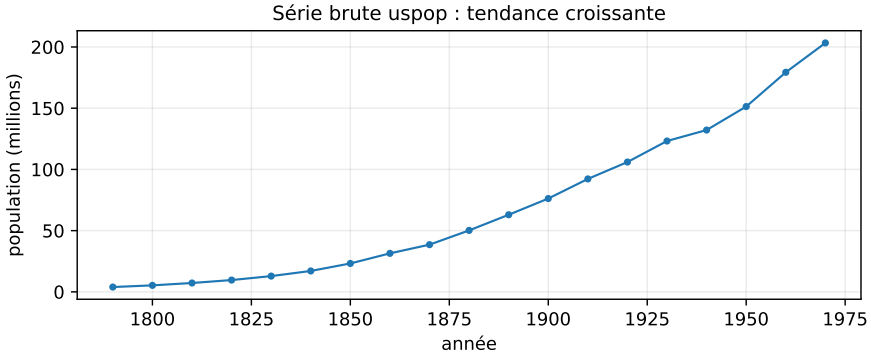


Figure 1.13 – Série brute `uspop`. La tendance est très forte : la série ne fluctue pas autour d'un niveau constant.

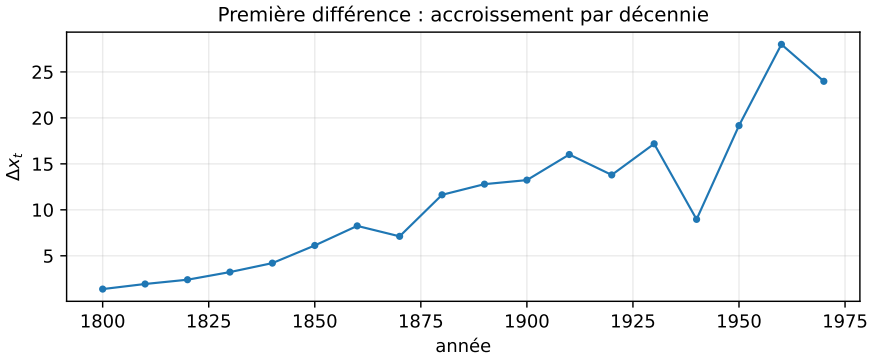


Figure 1.14 – Première différence de `uspop`. On ne représente plus la population, mais l'accroissement de population entre deux recensements successifs.

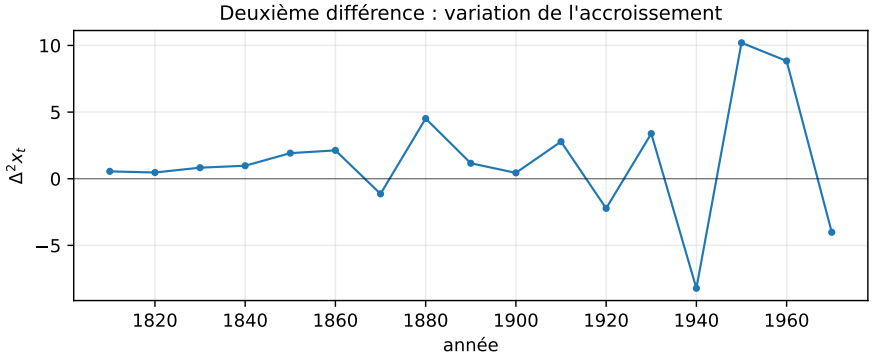


Figure 1.15 – Deuxième différence de `uspop`. On représente la variation de l'accroissement. La tendance initiale est beaucoup moins dominante.

1.4.4 Effet sur une saisonnalité

Exemple 1.4.5: Différenciation saisonnière

Si $s_{t+d} = s_t$, alors

$$\Delta_d (m_t + s_t + Z_t) = m_t - m_{t-d} + Z_t - Z_{t-d}.$$

La composante saisonnière s_t disparaît. Si m_t varie lentement, la nouvelle tendance $m_t - m_{t-d}$ est souvent plus simple que m_t.

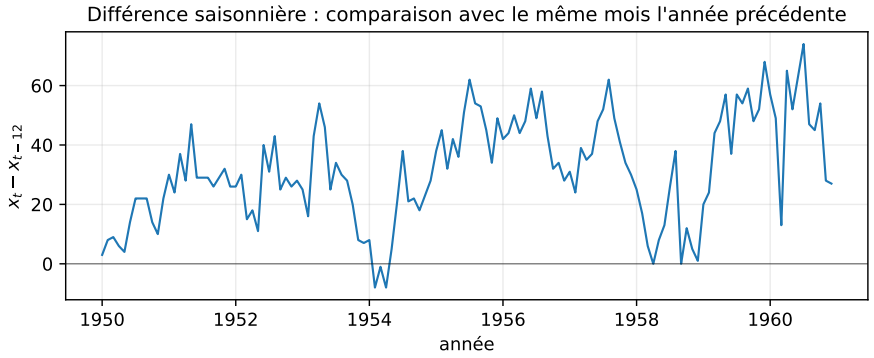


Figure 1.16 – Différence saisonnière de `AirPassengers`. Chaque valeur est comparée au même mois de l'année précédente : $x_t - x_{t-12}$. La saisonnalité annuelle visible dans la série brute est fortement réduite.

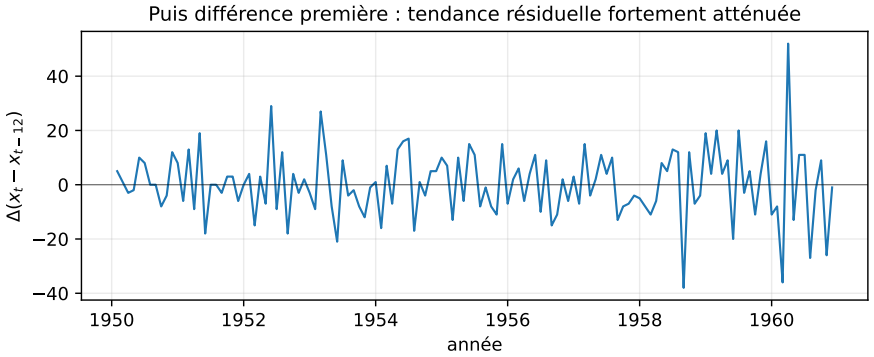


Figure 1.17 – Différence première après différence saisonnière. On applique $\Delta\Delta_{12}$, ce qui atténue à la fois la saisonnalité annuelle et la tendance résiduelle.

Point de vigilance

Différencier n'est pas une opération neutre. Elle peut rendre une série plus proche d'une série stationnaire, mais elle change aussi l'objet étudié : on passe des niveaux aux accroissements. Toute interprétation doit en tenir compte.

1.5 Modélisation probabiliste et stationnarité

1.5.1 Processus stochastique et réalisation

Les données observées $x_1, \ldots, x_n$ sont finies. Le modèle mathématique est souvent une suite infinie de variables aléatoires. Ce passage de la série observée au processus modélisant est fondamental.

L'idée est la même qu'en statistique classique, mais avec une difficulté supplémentaire. Dans un échantillon i.i.d., on observe plusieurs répétitions comparables. En séries temporelles, on observe souvent une seule trajectoire. Le modèle probabiliste sert donc à dire quelles régularités temporelles on accepte de transporter du passé vers le futur.

Définition 1.5.1: Processus stochastique

Un *processus stochastique* est une famille de variables aléatoires

$$(X_t)_{t\in T}$$

définies sur un même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Dans ce cours, sauf mention contraire, $T \subset \mathbb{Z}$, typiquement $T = \mathbb{Z}$, $T = \mathbb{N}$, ou $T = \{1, \dots, n\}$.

Définition 1.5.2: Réalisation ou trajectoire

Pour $\omega \in \Omega$ fixé, l'application

$$t \longmapsto X_t(\omega)$$

est appelée une *réalisation* ou *trajectoire* du processus. Les observations $x_1, \dots, x_n$ sont interprétées comme une partie d'une trajectoire.

Point de vigilance

On observe une seule trajectoire. La stationnarité, l'indépendance ou la gaussiannité sont des propriétés du processus, pas simplement de la courbe tracée. Une trajectoire régulière peut venir d'un processus stationnaire, et une trajectoire très irrégulière peut venir d'un processus non stationnaire.

1.5.2 Moyenne, autocovariance et autocorrélation

On commence par les moments d'ordre un et deux. Ils ne décrivent pas toute la loi du processus, mais ils sont la base de la plupart des modèles linéaires. La moyenne décrit le niveau moyen à chaque date ; l'autocovariance décrit comment deux dates communiquent linéairement entre elles.

Définition 1.5.3: Processus du second ordre, moyenne et autocovariance

Un processus $X = (X_t)_{t\in\mathbb{Z}}$ est dit *du second ordre* si

$$\mathbb{E}\left(|X_t|^2\right) < \infty$$

pour tout $t \in \mathbb{Z}$. On définit alors sa fonction moyenne par

$$\mu_X(t) = \mathbb{E}(X_t)$$

et sa fonction d'autocovariance par

$$\gamma_X(s,t) = \mathrm{Cov}(X_s, X_t) = \mathbb{E}\left[(X_s - \mu_X(s))(X_t - \mu_X(t))\right].$$

Pour un processus du second ordre, on a aussi

$$\gamma_X(s,t) = \mathbb{E}(X_s X_t) - \mathbb{E}(X_s)\,\mathbb{E}(X_t).$$

La covariance mesure une dépendance linéaire. Elle est symétrique : $\gamma_X(s,t) = \gamma_X(t,s)$. La variance au temps t est $\gamma_X(t,t) = \mathrm{Var}(X_t)$.

Pour une série temporelle, le point nouveau est que l'on ne s'intéresse pas seulement à la variance d'une observation, mais aussi à la covariance entre une observation et son passé. C'est cette information qui permettra plus tard de construire des prédicteurs.

Définition 1.5.4: Autocorrélation générale

Si X est du second ordre et si $\gamma_X(s,s) > 0$, $\gamma_X(t,t) > 0$, on définit l'autocorrélation entre les temps s et t par

$$\rho_X(s,t) = \frac{\gamma_X(s,t)}{\sqrt{\gamma_X(s,s)\gamma_X(t,t)}}.$$

On a toujours $-1 \leq \rho_X(s,t) \leq 1$.

Remarque 1.5.5: Pourquoi l'autocorrélation est-elle utile ?

La covariance dépend de l'unité de mesure. L'autocorrélation, elle, est sans unité et appartient à $[-1,1]$. Elle permet donc de comparer la force de dépendance entre différents retards et entre différentes séries.

1.5.3 Stationnarité forte et stationnarité faible

La stationnarité formalise l'idée que la structure probabiliste ne change pas au cours du temps. Elle permet d'apprendre à partir du passé pour prévoir le futur. Il ne faut pas la confondre avec l'idée qu'une trajectoire serait plate ou sans oscillations : un processus stationnaire peut produire des trajectoires très irrégulières, ou même périodiques dans certains exemples.

Définition 1.5.6: Stationnarité forte

Un processus $X = (X_t)_{t\in\mathbb{Z}}$ est *fortement stationnaire* si, pour tout $n \geq 1$, tous $t_1, \ldots, t_n \in \mathbb{Z}$ et tout $h \in \mathbb{Z}$, les vecteurs aléatoires

$$(X_{t_1}, \ldots, X_{t_n}) \quad \text{et} \quad (X_{t_1+h}, \ldots, X_{t_n+h})$$

ont la même loi.

La stationnarité forte est très naturelle, mais souvent trop exigeante et difficile à vérifier à partir d'une seule trajectoire. En pratique, dans les modèles linéaires, on utilise surtout la stationnarité faible.

L'idée de la stationnarité faible est de ne garder que ce qui est nécessaire pour travailler avec les moindres carrés linéaires : une moyenne stable et une structure de covariance stable par translation. C'est moins ambitieux qu'une invariance complète de la loi, mais beaucoup plus maniable.

Définition 1.5.7: Stationnarité faible

Un processus du second ordre $X = (X_t)_{t\in\mathbb{Z}}$ est *faiblement stationnaire*, ou simplement *stationnaire*, si

i) sa moyenne est constante : il existe $\mu \in \mathbb{R}$ tel que $\mathbb{E}(X_t) = \mu$ pour tout $t \in \mathbb{Z}$;

ii) sa covariance est invariante par translation : pour tous $s, t, h \in \mathbb{Z}$,

$$\gamma_X(s,t) = \gamma_X(s+h, t+h).$$

Si X est stationnaire, alors $\gamma_X(s,t)$ dépend seulement de l'écart temporel. On pose alors

$$\gamma_X(h) = \mathrm{Cov}(X_t, X_{t+h}), \qquad h \in \mathbb{Z}.$$

Cette quantité ne dépend pas de t. On appelle h le *retard* ou *lag*. C'est un point crucial : si la série est stationnaire, la covariance entre aujourd'hui et demain est la même que celle entre n'importe quel jour et le lendemain. On peut donc estimer une dépendance temporelle en moyennant le long d'une seule trajectoire.

Définition 1.5.8: Autocovariance et autocorrélation stationnaires

Si X est stationnaire, on définit sa fonction d'autocovariance par

$$\gamma_X(h) = \mathrm{Cov}\,(X_t, X_{t+h})\,, \qquad h \in \mathbb{Z},$$

et sa fonction d'autocorrélation par

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}$$

lorsque $\gamma_X(0) > 0$.

Théorème 1.5.9: Stationnarité forte et stationnarité faible

Un processus fortement stationnaire du second ordre est faiblement stationnaire. La réciproque est fausse en général. Pour un processus gaussien, la stationnarité faible implique la stationnarité forte.

Ce théorème explique pourquoi les deux notions coexistent. La stationnarité forte est conceptuellement plus propre, car elle parle de toute la loi. La stationnarité faible est plus calculable, car elle ne parle que de moyennes et de covariances. Dans le cas gaussien, il n'y a plus de différence : moyenne et covariance déterminent toute la loi.

Démonstration. Si X est fortement stationnaire et du second ordre, alors (X_t) et (X_{t+h}) ont même loi, donc même espérance. De même, (X_s, X_t) et (X_{s+h}, X_{t+h}) ont même loi, donc même covariance. Cela donne la stationnarité faible.

La réciproque est fausse car les deux premiers moments ne déterminent pas toujours la loi. En revanche, pour un vecteur gaussien, la loi est déterminée par le vecteur moyenne et la matrice de covariance. Donc, pour un processus gaussien, l'invariance des moyennes et covariances suffit à donner l'invariance des lois finies-dimensionnelles. □

1.5.4 Trois exemples fondamentaux

Définition 1.5.10: Bruit blanc

Un processus $Z = (Z_t)_{t\in\mathbb{Z}}$ est un *bruit blanc faible* de moyenne 0 et variance σ^2, noté $BB\left(0, \sigma^2\right)$, si

$$\mathbb{E}\,(Z_t) = 0, \qquad \mathrm{Cov}\,(Z_t, Z_{t+h}) = \sigma^2 \mathbf{1}_{\{h=0\}}.$$

S'il est de plus constitué de variables indépendantes et identiquement distribuées, on parlera de bruit blanc fort ou i.i.d., selon le niveau de précision souhaité.

Le bruit blanc est le modèle de référence pour une série sans mémoire linéaire. Il ne signifie pas nécessairement que les variables sont indépendantes, sauf si on le précise. Dans un diagnostic de modèle, dire que les résidus ressemblent à un bruit blanc signifie que le modèle a capturé l'essentiel de la dépendance linéaire visible.

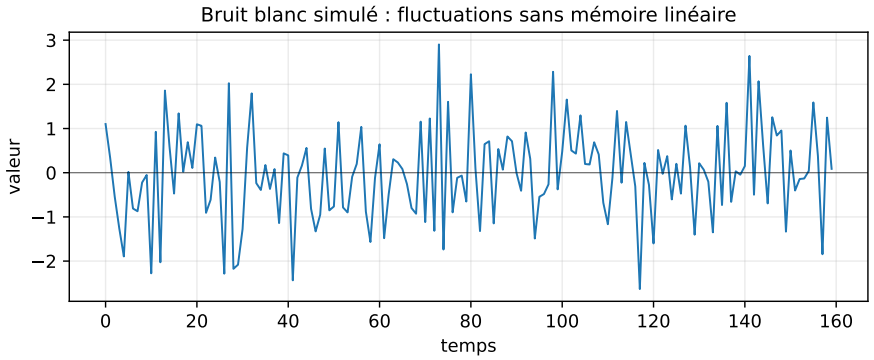


Figure 1.18 – Bruit blanc gaussien simulé. Les fluctuations sont centrées et ne montrent pas de mémoire linéaire visible : connaître une valeur ne permet pas de prévoir linéairement la suivante.

Exemple 1.5.11: Marche aléatoire

Soit Z_t un bruit blanc de variance σ^2. La marche aléatoire est définie par

$$X_0 = 0, \qquad X_t = X_{t-1} + Z_t, \quad t \geq 1.$$

Alors $X_t = Z_1 + \cdots + Z_t$, si bien que $\mathbb{E}(X_t) = 0$ et $\mathrm{Var}(X_t) = t\sigma^2$. La variance dépend de t, donc la marche aléatoire n'est pas stationnaire.
L'intuition est simple : les chocs ne s'effacent jamais. Chaque nouvelle innovation s'ajoute aux précédentes et reste dans le niveau de la série. C'est pourquoi deux trajectoires de marche aléatoire peuvent s'éloigner durablement l'une de l'autre, même si les chocs ont moyenne nulle.

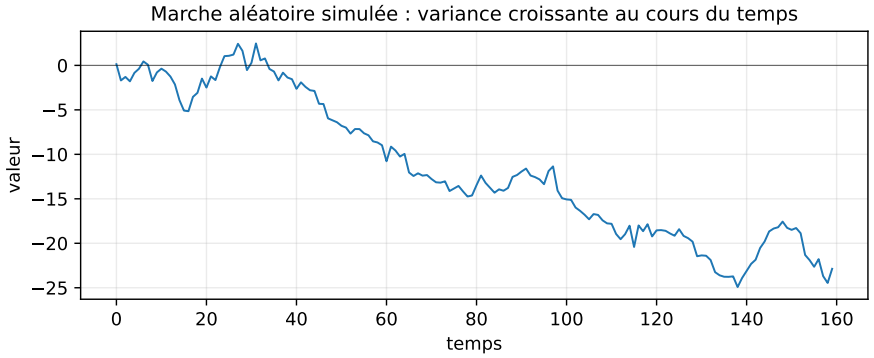


Figure 1.19 – Marche aléatoire simulée. La série semble dériver durablement parce que les chocs s'accumulent. Mathématiquement, sa variance vaut $t\sigma^2$, donc elle n'est pas stationnaire.

Exemple 1.5.12: Processus harmonique

Soient A et C deux variables aléatoires centrées, non corrélées, de même variance σ^2. Pour $\theta \in \mathbb{R}$, on pose

$$X_t = A\cos(\theta t) + C\sin(\theta t).$$

Alors $\mathbb{E}(X_t) = 0$ et

$$\begin{aligned}\mathrm{Cov}(X_t, X_{t+h}) &= \sigma^2\cos(\theta t)\cos(\theta(t+h)) + \sigma^2\sin(\theta t)\sin(\theta(t+h)) \\ &= \sigma^2\cos(\theta h).\end{aligned}$$

La covariance ne dépend donc que de h : le processus est stationnaire.
Cet exemple est volontairement troublant. Les trajectoires ressemblent à des sinusoïdes, donc à une saisonnalité déterministe. Pourtant le processus est stationnaire au sens faible. La raison est que l'aléa porte seulement sur les amplitudes A et C, et que la covariance reste inchangée par translation temporelle.

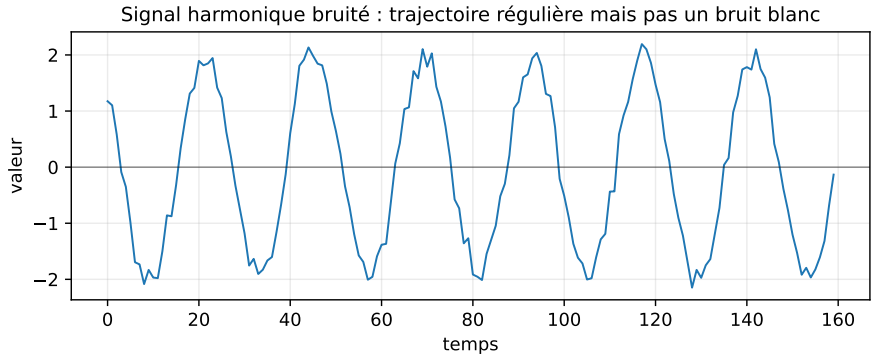


Figure 1.20 – Signal harmonique bruité. Le graphique ressemble à une saisonnalité régulière. Cet exemple rappelle qu'une apparence régulière ne suffit pas à conclure à la non-stationnarité : tout dépend de la source d'aléa et du modèle.

1.5.5 Matrices de Toeplitz et autocovariances possibles

La stationnarité impose une structure très particulière aux matrices de covariance finies. Cette structure est importante parce que la prédiction linéaire et l'estimation reposent souvent sur des matrices de covariance. Dans une série stationnaire, ces matrices ne sont pas quelconques : elles sont entièrement déterminées par les covariances aux différents retards.

Définition 1.5.13: Matrice de Toeplitz

Une matrice $M = (m_{j,k})_{1\le j,k\le n}$ est une *matrice de Toeplitz* si ses coefficients sont constants le long des diagonales, c'est-à-dire si $m_{j,k}$ ne dépend que de $j-k$.

Soit X un processus stationnaire. Pour $n \ge 1$, la matrice de covariance de $(X_1, \dots, X_n)$ est

$$\Gamma_n = (\gamma_X(j-k))_{1\le j,k\le n} = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{pmatrix}.$$

Les coefficients sont constants sur les diagonales : c'est une matrice de Toeplitz. Cela traduit exactement l'idée que seule la distance temporelle compte. Deux observations séparées de trois dates ont la même covariance, qu'elles soient au début ou à la fin de la série.

Théorème 1.5.14: Caractérisation des autocovariances stationnaires

Une fonction $\gamma : \mathbb{Z} \to \mathbb{R}$ est la fonction d'autocovariance d'un processus stationnaire si et seulement si elle est paire et de type positif, c'est-à-dire :

i) $\gamma(h) = \gamma(-h)$ pour tout $h \in \mathbb{Z}$;

ii) pour tout $n \geq 1$, tous $t_1, \ldots, t_n \in \mathbb{Z}$ et tous $a_1, \ldots, a_n \in \mathbb{R}$,

$$\sum_{j=1}^{n} \sum_{k=1}^{n} a_j \gamma(t_j - t_k) a_k \geq 0.$$

En particulier, pour toute autocovariance stationnaire,

$$\gamma(0) \geq 0, \qquad |\gamma(h)| \leq \gamma(0), \qquad \gamma(h) = \gamma(-h).$$

Ce résultat dit en particulier qu'on ne peut pas choisir arbitrairement une fonction $h \mapsto \gamma(h)$. Les nombres $\gamma(h)$ doivent pouvoir former des matrices de covariance, donc des matrices positives. Cette contrainte est invisible si l'on regarde les retards un par un, mais elle devient essentielle dès que l'on considère plusieurs dates simultanément.

Preuve de la partie nécessaire. Si γ est l'autocovariance d'un processus stationnaire X, alors

$$\gamma(-h) = \mathrm{Cov}\,(X_t, X_{t-h}) = \mathrm{Cov}\,(X_{t-h}, X_t) = \gamma(h).$$

Pour la positivité, posons $Y = \sum_{j=1}^{n} a_j \left(X_{t_j} - \mathbb{E}\left(X_{t_j}\right)\right)$. Alors

$$0 \leq \mathbb{E}\left(Y^2\right) = \sum_{j=1}^{n} \sum_{k=1}^{n} a_j \gamma(t_j - t_k) a_k.$$

Enfin, $\gamma(0) \geq 0$ est une variance, et $|\gamma(h)| \leq \gamma(0)$ découle de l'inégalité de Cauchy–Schwarz. □

1.6 Synthèse du chapitre

Le chapitre a introduit le vocabulaire minimal de tout le cours. À retenir :

1) Une série observée est une trajectoire finie ; un modèle est un processus.
2) Le chronogramme est indispensable, mais insuffisant : month plots et lag plots aident à voir saisonnalité et dépendance.
3) Le modèle additif $X_t = m_t + s_t + Z_t$ sert à séparer structure déterministe et bruit.
4) Un résidu empirique est ce qui reste après ajustement d'une partie structurée ; il sert à diagnostiquer ce que le modèle n'explique pas.
5) Les opérateurs B, Δ et Δ_d permettent d'écrire simplement les transformations qui éliminent tendance ou saisonnalité.
6) La stationnarité faible signifie : moyenne constante et covariance ne dépendant que du retard.
7) Le bruit blanc est le modèle de référence d'absence de corrélation temporelle.

Point de vigilance

La morale pratique est la suivante : on ne cherche pas à rendre la série brute stationnaire à tout prix ; on cherche à comprendre quelle partie est déterministe ou structurelle, puis à modéliser la partie résiduelle par un processus stationnaire simple.

Chapitre 2

Outils R pour séries temporelles

Mots-clés. Objet `ts` ; fréquence d'observation ; date de début ; date de fin ; temps numérique ; cycle ; sous-série ; `window()` ; valeurs manquantes ; retard en `R` ; différence ; différence saisonnière ; reproductibilité.

2.1 Pourquoi un chapitre sur les objets R ?

Le chapitre précédent a introduit les notions statistiques de base : chronogramme, tendance, saisonnalité, résidu, retard, différence et stationnarité. Avant d'aller plus loin dans les modèles, il faut savoir manipuler correctement les séries dans `R`. Ce point est plus important qu'il n'y paraît. Une série temporelle n'est pas seulement un vecteur de nombres : elle contient aussi une information temporelle. Si cette information est mal codée, les graphiques, les sous-séries, les différences saisonnières et les prévisions risquent d'être mal interprétés.

Dans ce chapitre, on ne cherche pas encore à choisir un modèle. L'objectif est plus élémentaire : comprendre comment `R` représente une série régulière, comment extraire une période, comment construire des transformations simples, et comment repérer les pièges les plus fréquents.

Idée à retenir

Un objet `ts` contient deux choses : des valeurs et une horloge. Les valeurs disent ce qui a été observé ; l'horloge dit quand elles ont été observées. En séries temporelles, perdre l'horloge revient souvent à perdre une partie du sens des données.

2.2 L'objet `ts`

2.2.1 Valeurs, début, fin et fréquence

L'objet de base de `R` pour les séries temporelles régulières est l'objet `ts`. Il est adapté lorsque les observations sont également espacées : données annuelles, trimestrielles, mensuelles, hebdomadaires régulières, etc.

Définition 2.2.1: Série régulière dans `R`

Un objet `ts` représente une série temporelle régulière. Il contient :

i) un vecteur ou une matrice de valeurs observées ;

ii) une date de début ;

iii) une date de fin ;

iv) une fréquence d'observation, c'est-à-dire le nombre d'observations par unité de temps.

Le mot *fréquence* peut prêter à confusion. Dans la commande `ts()`, `frequency = 12` signifie qu'il y a douze observations par unité de temps, par exemple douze mois par année. Ce n'est pas

encore la fréquence au sens de l'analyse spectrale. C'est une convention de calendrier.

Les valeurs les plus courantes sont :

Type de données	Fréquence R	Exemple de date
annuelles	`frequency = 1`	1980
trimestrielles	`frequency = 4`	`c(1980,3)` pour le 3e trimestre
mensuelles	`frequency = 12`	`c(1980,7)` pour juillet 1980

Exemple 2.2.2: La série `co2`

La série `co2`, fournie avec le package `datasets` de R, contient des concentrations atmosphériques mensuelles de dioxyde de carbone à Mauna Loa, en parties par million. Elle est déjà stockée sous forme d'objet `ts`.

```
data(co2)
start(co2)
end(co2)
frequency(co2)
tsp(co2)
```

La commande `start()` donne la date de début, `end()` la date de fin, `frequency()` le nombre d'observations par année, et `tsp()` résume les trois informations dans un même vecteur.

La figure suivante montre le chronogramme de cette série.

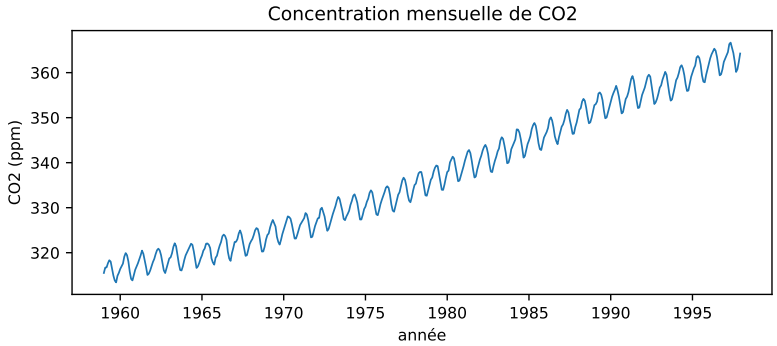


Figure 2.1 – Concentration mensuelle de CO2 à Mauna Loa. Données : série `co2`, fournie avec le package `datasets` de R. Le graphique montre à la fois une tendance croissante et une oscillation annuelle.

Comment lire ce graphique ?

La série `co2` est utile pédagogiquement parce que deux phénomènes sont visibles immédiatement : le niveau moyen augmente au cours du temps, et un motif annuel se répète. Ces deux phénomènes ne doivent pas être confondus : la tendance décrit l'évolution lente du niveau, tandis que la saisonnalité décrit le motif qui revient à l'intérieur de chaque année.

2.2.2 Construire soi-même un objet `ts`

Lorsqu'une série est fournie sous forme d'un simple vecteur, il faut ajouter l'information temporelle. Par exemple, si `x` contient des observations mensuelles commençant en janvier 2020, on écrit :

```
y <- ts(x, start = c(2020, 1), frequency = 12)
```

Si `x` contient des observations trimestrielles commençant au deuxième trimestre 2015, on écrit :

```
y <- ts(x, start = c(2015, 2), frequency = 4)
```

La difficulté n'est pas technique : la difficulté est de ne pas mentir à `R`. Une série mensuelle codée comme une série annuelle perdra sa saisonnalité mensuelle. Une série trimestrielle codée avec `frequency = 12` sera interprétée comme mensuelle. Le graphique pourra alors sembler correct, mais les calculs utilisant le calendrier seront faux.

2.3 Temps numérique et cycle saisonnier

2.3.1 Les fonctions `time()` et `cycle()`

Deux fonctions sont particulièrement utiles pour comprendre l'horloge attachée à un objet `ts`.

— `time(x)` renvoie les dates numériques associées aux observations.

— `cycle(x)` renvoie la position de chaque observation dans le cycle saisonnier.

Pour une série mensuelle, `cycle(x)` vaut 1 pour janvier, 2 pour février, etc. Pour une série trimestrielle, il vaut 1, 2, 3 ou 4.

Exemple 2.3.1: Interpréter `time()` et `cycle()` sur `co2`

Pour la série `co2`, on peut regarder les premières dates et les premiers mois :

```
head(time(co2))
head(cycle(co2))
```

La fonction `time()` renvoie des dates numériques, par exemple 1959, $1959 + 1/12$, $1959 + 2/12$, etc. La fonction `cycle()` indique simplement le mois de l'année. Ce sont deux informations différentes : l'une situe l'observation dans le temps absolu, l'autre situe l'observation dans l'année.

Point de vigilance

Ne pas confondre `time(x)` et `cycle(x)`. La première fonction répond à la question « à quelle date sommes-nous ? ». La seconde répond à la question « où sommes-nous dans le cycle annuel ou trimestriel ? ».

2.4 Extraire une sous-série avec `window()`

En pratique, on ne regarde pas toujours toute la série. On peut vouloir zoomer sur quelques années, séparer une période d'apprentissage et une période de test, ou étudier une période avant et après une rupture. La fonction `window()` sert à extraire une sous-série en conservant l'information temporelle.

Définition 2.4.1: Sous-série temporelle

Soit $x = (x_t)$ une série observée. Une sous-série est une restriction de la série à une période donnée. Dans `R`, si `x` est un objet `ts`, la commande `window(x, start = ..., end = ...)` extrait cette restriction en conservant sa structure temporelle.

Exemple 2.4.2: Zoom sur `co2`

Pour extraire les années 1990 à 1992, on écrit :

```
co2.9092 <- window(co2, start = c(1990, 1), end = c(1992, 12))
plot.ts(co2.9092)
```

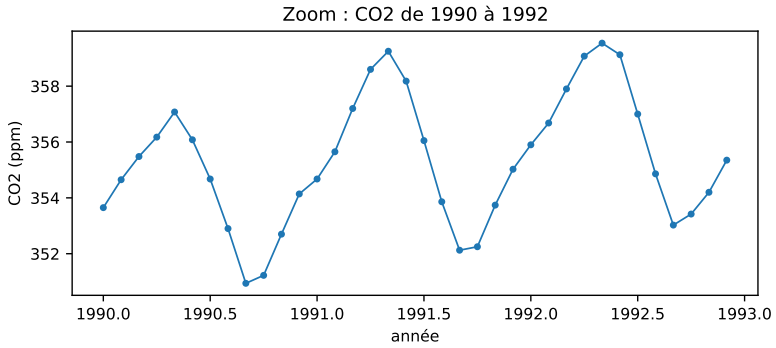


Figure 2.2 – Zoom sur la série `co2` entre 1990 et 1992. Le zoom rend le cycle annuel beaucoup plus visible que le chronogramme complet.

Ce zoom illustre un point important : un graphique global peut masquer une structure locale. Sur la série complète, la tendance croissante domine fortement. Sur trois années seulement, le motif annuel devient beaucoup plus facile à lire. Le choix de l'échelle temporelle est donc déjà un choix d'analyse.

2.5 Retards et différences dans R

2.5.1 Retards : attention à la convention de `lag()`

Le chapitre 1 a défini l'opérateur retard par $(BX)_t = X_{t-1}$. Dans `R`, la fonction `lag()` appliquée à un objet `ts` modifie surtout l'index temporel de la série. Cela peut surprendre, car le vecteur de valeurs affiché n'est pas toujours celui que l'on attend intuitivement.

Pour construire un nuage de points (x_{t-1}, x_t), il est souvent plus clair, au début du cours, d'utiliser directement les vecteurs :

```
x <- as.numeric(co2)
n <- length(x)
plot(x[-n], x[-1])
```

On évite ainsi les ambiguïtés de calendrier. Plus tard, quand la convention sera maîtrisée, on pourra utiliser des fonctions plus compactes.

Point de vigilance

La notation mathématique $BX_t = X_{t-1}$ et la fonction `lag()` de `R` ne doivent pas être identifiées trop vite. La première est une convention algébrique. La seconde manipule un objet `ts` avec son index temporel. Pour les premiers lag plots, la construction explicite `x[-n]` et `x[-1]` est plus sûre.

2.5.2 Différences ordinaires et saisonnières

La fonction `diff()` calcule des différences. Si `x` est une série, alors `diff(x)` correspond à la série des différences successives :

$$x_t - x_{t-1}.$$

On peut aussi utiliser un retard plus grand. Pour une série mensuelle, `diff(x, lag = 12)` calcule la différence entre une observation et celle du même mois l'année précédente :

$$x_t - x_{t-12}.$$

Exemple 2.5.1: Différence saisonnière de `co2`

La série `co2` est mensuelle. La commande suivante calcule donc la variation par rapport au même mois de l'année précédente :

```
dco2 <- diff(co2, lag = 12)
plot.ts(dco2)
```

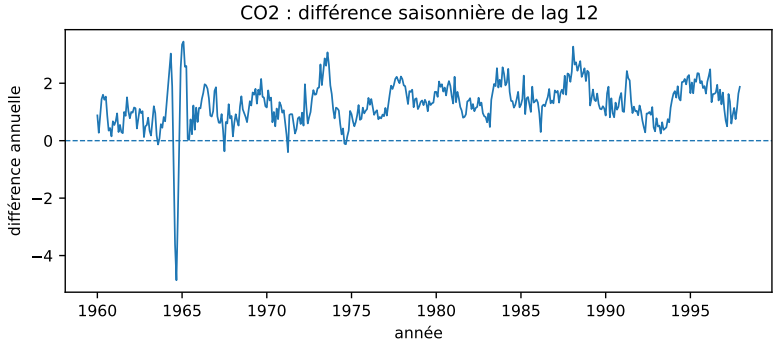


Figure 2.3 – Différence saisonnière de `co2` : chaque valeur est comparée au même mois de l'année précédente. L'oscillation annuelle est fortement réduite, mais une structure de long terme peut rester présente.

Il ne faut pas interpréter `diff()` comme une recette automatique. Une différence change la question posée. La série brute demande : « quel est le niveau du CO2 ? ». La différence annuelle demande : « combien a-t-on gagné par rapport au même mois de l'année précédente ? ». Ce n'est pas la même grandeur.

2.6 Valeurs manquantes

2.6.1 Repérer les valeurs manquantes

Les séries réelles contiennent souvent des observations manquantes : capteur en panne, changement de définition, donnée non publiée, erreur d'importation. Dans `R`, une valeur manquante est notée `NA`. On peut les repérer avec `is.na()`.

```
any(is.na(co2))
sum(is.na(co2))
```

Quand une série contient des valeurs manquantes, certains calculs ou graphiques peuvent être modifiés. Il faut donc toujours savoir si des `NA` sont présents avant d'ajuster un modèle.

2.6.2 Faut-il remplacer les valeurs manquantes ?

Remplacer une valeur manquante n'est pas un geste neutre. Cela revient à ajouter une information qui n'a pas été observée. Il peut être raisonnable d'interpoler une ou deux valeurs manquantes isolées dans une série très régulière ; il serait beaucoup plus discutable de remplacer une longue période manquante dans une série instable.

Exemple 2.6.1: Une lacune artificielle dans co2

On crée volontairement une lacune dans la série pour voir comment elle apparaît graphiquement.

```
co2.miss <- co2
t <- time(co2.miss)
ind <- t >= 1991 + 3/12 & t <= 1991 + 8/12
co2.miss[ind] <- NA
plot.ts(window(co2.miss, start = c(1990,1), end = c(1992,12)))
```

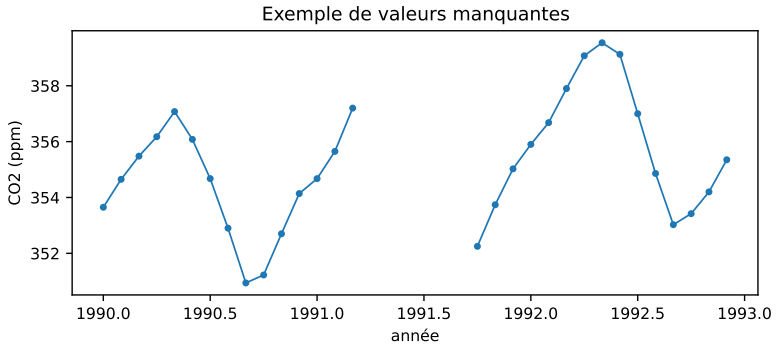


Figure 2.4 – Exemple de valeurs manquantes introduites artificiellement dans co2. Le trou dans la courbe rappelle qu'une valeur manquante est une absence d'information, pas une valeur nulle.

Une méthode simple consiste à interpoler linéairement entre les valeurs observées voisines. Cela peut être utile pour un graphique ou un exercice, mais cela ne doit pas être confondu avec une observation réelle.

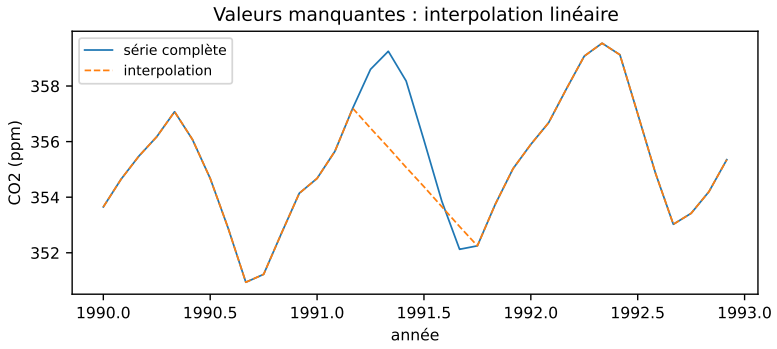


Figure 2.5 – Interpolation linéaire sur une courte période manquante. L'interpolation donne une courbe plausible visuellement, mais les points interpolés restent des valeurs reconstruites.

2.7 Synthèse du chapitre

Ce chapitre a introduit les manipulations élémentaires de R nécessaires pour la suite du cours. À retenir :

1) Un objet `ts` contient des valeurs et une structure temporelle.
2) La fréquence dans R est le nombre d'observations par unité de temps.
3) Les fonctions `start()`, `end()`, `frequency()`, `time()` et `cycle()` permettent de contrôler l'horloge de la série.
4) La fonction `window()` extrait une sous-série sans perdre l'information temporelle.
5) La fonction `diff()` calcule des différences ordinaires ou saisonnières.
6) Les valeurs manquantes doivent être repérées et traitées explicitement ; les remplacer n'est jamais automatique.

Point de vigilance

La règle pratique est simple : avant de modéliser une série dans `R`, il faut vérifier son calendrier. Une erreur de fréquence ou de date de début peut rendre faux un graphique saisonnier, une différence saisonnière, ou une séparation apprentissage/test.

Chapitre 3

Filtrage linéaire et séries linéaires

Mots-clés. Filtre linéaire ; moyenne mobile ; filtre causal ; filtre centré ; différence ; différence saisonnière ; filtre infini ; convergence dans L^2 ; convolution ; bruit blanc filtré ; série linéaire.

3.1 Pourquoi introduire les filtres ?

Dans les deux premiers chapitres, on a appris à regarder une série temporelle, à lire son calendrier, et à effectuer des transformations simples dans `R`. On a déjà rencontré la différence ordinaire $1 - B$, la différence saisonnière $1 - B^s$, et les moyennes locales utilisées pour lisser une série. Le but de ce chapitre est de comprendre ces opérations comme des cas particuliers d'un même objet : un *filtre linéaire*.

L'idée de départ est concrète. À partir d'une série $X = (X_t)_{t\in\mathbb{Z}}$, on construit une nouvelle série $Y = (Y_t)_{t\in\mathbb{Z}}$ en mélangeant plusieurs valeurs de X. Par exemple,

$$Y_t = \frac{X_{t-1} + X_t + X_{t+1}}{3}$$

est une moyenne locale de trois observations. La valeur au temps t est remplacée par une valeur plus stable, calculée à partir de ses voisines. On dit que l'on a appliqué un filtre.

Idée à retenir

Un filtre n'est pas seulement une commande `R`. C'est une opération mathématique sur une série. Il faut donc toujours se demander : quelles valeurs de la série initiale le filtre utilise-t-il ? avec quels poids ? et dans quel but ? Un même filtre peut être très utile pour décrire une série déjà observée, mais inutilisable pour prévoir si sa définition utilise des valeurs futures.

Il y a trois raisons de formaliser cette notion.

Premièrement, les filtres donnent un langage commun pour plusieurs opérations déjà vues. Les différences, les moyennes mobiles, les lissages élémentaires et certaines transformations saisonnières sont des filtres.

Deuxièmement, les filtres clarifient l'interprétation. Une moyenne mobile centrée lisse une série mais utilise des valeurs futures ; une moyenne mobile passée est moins symétrique, mais elle est utilisable en temps réel.

Troisièmement, les filtres sont à la base des modèles de dépendance temporelle. Une série dépendante peut souvent être vue comme un bruit blanc transformé par un filtre : le bruit initial ne contient pas de mémoire, mais le filtre peut créer de la dépendance entre observations successives.

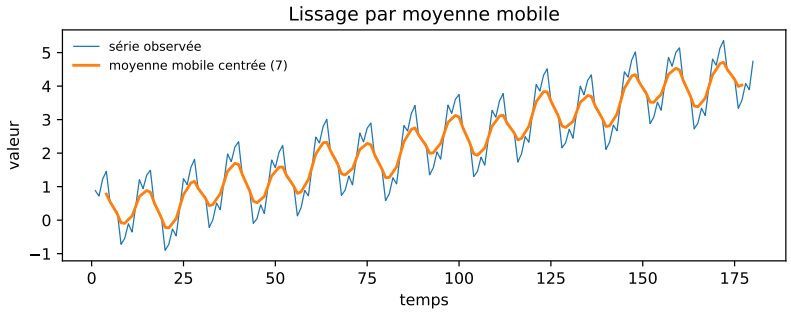


Figure 3.1 – Une série simulée et sa moyenne mobile centrée sur 7 points. Le filtre atténue les fluctuations rapides : il ne retrouve pas une “vraie” série cachée, il propose une nouvelle lecture de la série observée.

3.2 Filtres linéaires finis

3.2.1 Définition et premiers exemples

On commence par les filtres qui ne combinent qu'un nombre fini d'observations. Ils suffisent déjà pour formaliser les moyennes mobiles usuelles et les différences.

Définition 3.2.1: Filtre linéaire fini

Soit $a = (a_j)_{j\in\mathbb{Z}}$ une suite réelle nulle sauf pour un nombre fini d'indices. Le *filtre linéaire fini* de coefficients a transforme une série $X = (X_t)_{t\in\mathbb{Z}}$ en la série $Y = a(B)X$ définie par

$$Y_t = (a(B)X)_t = \sum_{j\in\mathbb{Z}} a_j X_{t-j}, \qquad t \in \mathbb{Z}.$$

La somme est bien définie sans condition supplémentaire, car seuls un nombre fini de coefficients a_j sont non nuls.

La notation $a(B)$ vient de l'opérateur retard. Formellement,

$$a(B) = \sum_{j\in\mathbb{Z}} a_j B^j,$$

où $(B^j X)_t = X_{t-j}$. Cette notation permet d'écrire un filtre comme un polynôme en B.

Remarque 3.2.2: Convention sur les indices

Dans la formule $Y_t = \sum_j a_j X_{t-j}$, un coefficient a_j avec $j > 0$ utilise une valeur passée X_{t-j}. Un coefficient a_j avec $j < 0$ utilise une valeur future $X_{t-j} = X_{t+|j|}$. Cette convention est celle qui rend la notation compatible avec l'opérateur retard B.

Exemple 3.2.3: Filtres élémentaires

1) **Identité.** Si $a_0 = 1$ et tous les autres coefficients sont nuls, alors $Y_t = X_t$.
2) **Retard d'un pas.** Si $a_1 = 1$, alors $Y_t = X_{t-1}$.
3) **Différence ordinaire.** Si $a_0 = 1$ et $a_1 = -1$, alors

$$Y_t = X_t - X_{t-1} = ((1 - B)X)_t.$$

4) **Différence saisonnière de période s.** Si $a_0 = 1$ et $a_s = -1$, alors

$$Y_t = X_t - X_{t-s} = ((1 - B^s)X)_t.$$

5) **Moyenne mobile passée sur trois valeurs.** Si $a_0 = a_1 = a_2 = 1/3$, alors

$$Y_t = \frac{X_t + X_{t-1} + X_{t-2}}{3}.$$

6) **Moyenne mobile centrée sur trois valeurs.** Si $a_{-1} = a_0 = a_1 = 1/3$, alors

$$Y_t = \frac{X_{t+1} + X_t + X_{t-1}}{3}.$$

Idée à retenir

Une moyenne mobile et une différence sont toutes les deux des filtres linéaires, mais elles ne cherchent pas à faire la même chose. Une moyenne mobile atténue les fluctuations locales. Une différence remplace un niveau par une variation. Les coefficients du filtre contiennent donc déjà une intention statistique.

3.2.2 Support, causalité et filtres centrés

Le *support* d'un filtre est l'ensemble des indices j tels que $a_j \neq 0$. Il indique quelles dates de la série initiale sont utilisées pour calculer la valeur filtrée.

Définition 3.2.4: Filtre causal et filtre centré

Un filtre fini $a(B)$ est dit *causal* si $a_j = 0$ pour tout $j < 0$. Dans ce cas, Y_t ne dépend que de $X_t, X_{t-1}, X_{t-2}, \ldots$, c'est-à-dire du présent et du passé.
Un filtre est dit *centré* lorsque ses poids sont répartis de manière symétrique autour de la date t. Par exemple,

$$Y_t = \frac{X_{t-1} + X_t + X_{t+1}}{3}$$

est une moyenne mobile centrée.

La distinction est importante. Pour décrire une série historique, un filtre centré est souvent naturel : il regarde autour de la date étudiée. Pour prévoir ou suivre une série en temps réel, il est impossible d'utiliser une valeur future. Il faut alors utiliser un filtre causal.

Point de vigilance

Une moyenne mobile centrée ne doit pas être utilisée comme outil de prévision immédiate : au temps t, elle utilise déjà X_{t+1}. Elle est parfaitement légitime pour comprendre une série historique, mais pas pour prévoir en temps réel.

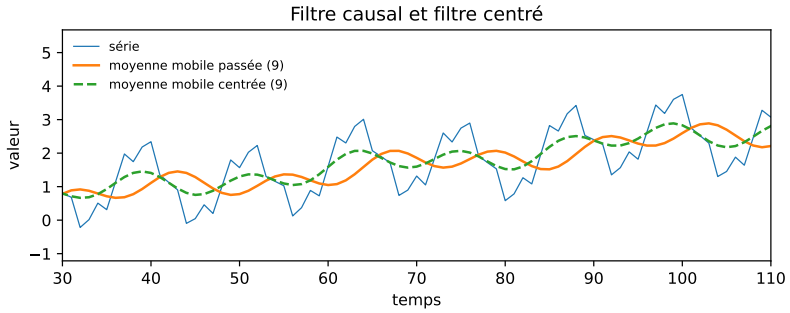


Figure 3.2 – Moyenne mobile passée et moyenne mobile centrée sur la même série. La moyenne mobile centrée suit mieux le milieu des oscillations, mais elle utilise des valeurs futures. La moyenne mobile passée est causale, mais elle réagit avec retard.

3.2.3 Linéarité

Le mot "linéaire" signifie que le filtre respecte les combinaisons linéaires.

Proposition 3.2.5: Linéarité d'un filtre fini

Soit $a(B)$ un filtre linéaire fini. Pour toutes séries X et W, et tous réels α, β, on a

$$a(B)(\alpha X + \beta W) = \alpha a(B)X + \beta a(B)W.$$

Démonstration. Pour tout t,

$$\begin{aligned}\big(a(B)(\alpha X + \beta W)\big)_t &= \sum_j a_j(\alpha X_{t-j} + \beta W_{t-j}) \\ &= \alpha \sum_j a_j X_{t-j} + \beta \sum_j a_j W_{t-j} \\ &= \big(\alpha a(B)X + \beta a(B)W\big)_t.\end{aligned}$$

□

Cette propriété explique pourquoi les filtres sont faciles à analyser : si une série est la somme d'une tendance, d'une composante saisonnière et d'un bruit, on peut étudier séparément l'effet du filtre sur chacune de ces composantes.

3.3 Effet d'un filtre sur des composantes simples

Avant de filtrer un processus aléatoire, il est utile de comprendre ce que fait un filtre sur des séries très simples. Ces calculs donnent des réflexes : un filtre conserve-t-il les constantes ? élimine-t-il une tendance ? atténue-t-il un motif périodique ?

3.3.1 Constantes et somme des coefficients

Proposition 3.3.1: Effet sur une constante

Soit $X_t = c$ pour tout t. Si $Y_t = \sum_j a_j X_{t-j}$, alors

$$Y_t = c \sum_j a_j.$$

En particulier, si $\sum_j a_j = 1$, le filtre conserve les constantes ; si $\sum_j a_j = 0$, le filtre annule les constantes.

Démonstration. Si $X_t = c$ pour tout t, alors $X_{t-j} = c$ pour tout j. Donc

$$Y_t = \sum_j a_j c = c \sum_j a_j.$$

□

Une moyenne mobile a des coefficients qui somment à 1 : si les données sont constantes, elle doit rester égale à cette constante. La différence ordinaire $1 - B$ a des coefficients 1 et -1, dont la somme est nulle : elle annule les constantes.

3.3.2 Tendances polynomiales et différences

La différence ordinaire $\Delta = 1 - B$ transforme une série de niveaux en série de variations. Son effet est particulièrement clair sur les polynômes.

Proposition 3.3.2: Différences et tendances polynomiales

Soit $p(t)$ un polynôme de degré $d \geq 1$. Alors $\Delta p(t) = p(t) - p(t-1)$ est un polynôme de degré $d - 1$. En particulier, $\Delta^r p(t) = 0$ dès que $r > d$.

Démonstration. Il suffit de vérifier le résultat sur les monômes. Si $p(t) = t^d$, alors

$$\Delta p(t) = t^d - (t-1)^d.$$

Par la formule du binôme,

$$(t-1)^d = t^d - dt^{d-1} + \text{termes de degré au plus } d-2.$$

Donc

$$t^d - (t-1)^d = dt^{d-1} + \text{termes de degré au plus } d-2,$$

ce qui est un polynôme de degré $d - 1$. Par linéarité, le résultat vaut pour tout polynôme. En appliquant plusieurs fois Δ, le degré baisse d'une unité à chaque étape jusqu'à disparaître. □

Exemple 3.3.3: Tendance linéaire

Si $X_t = a + bt$, alors

$$\Delta X_t = X_t - X_{t-1} = a + bt - (a + b(t-1)) = b.$$

La différence ordinaire transforme donc une tendance linéaire en constante. Une deuxième différence annule cette constante.

3.3.3 Saisonnalité et différence saisonnière

Si une composante saisonnière s_t a période s, alors $s_t = s_{t-s}$. La différence saisonnière annule donc cette composante :

$$(1 - B^s)s_t = s_t - s_{t-s} = 0.$$

Idée à retenir

La différence ordinaire compare une observation à l'observation immédiatement précédente. La différence saisonnière compare une observation à l'observation située au même endroit dans le cycle précédent : même mois de l'année précédente pour une série mensuelle, même trimestre de l'année précédente pour une série trimestrielle.

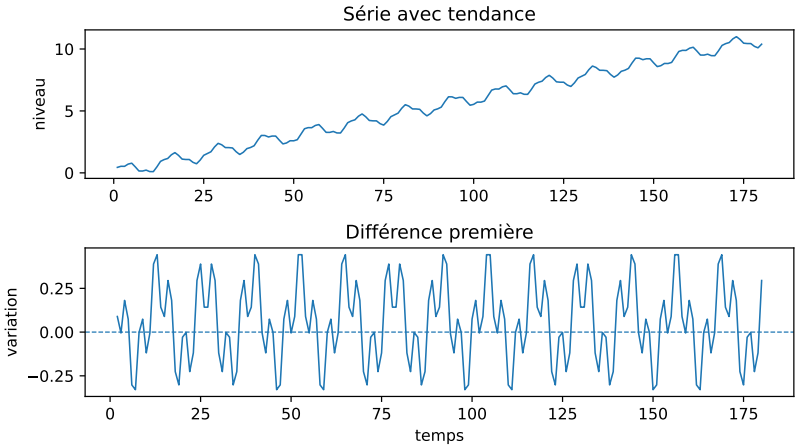


Figure 3.3 – Une tendance quadratique, sa première différence et sa deuxième différence. Les différences successives réduisent le degré de la tendance.

3.4 Des filtres finis aux filtres infinis

Les filtres précédents ne combinent qu'un nombre fini de valeurs. C'est suffisant pour lisser localement ou calculer des différences. Mais certains modèles naturels utilisent une infinité de valeurs passées, avec des poids de plus en plus petits. Par exemple,

$$Y_t = X_t + \frac{1}{2}X_{t-1} + \frac{1}{4}X_{t-2} + \frac{1}{8}X_{t-3} + \cdots .$$

Cette écriture est intuitive : les observations anciennes comptent encore, mais de moins en moins. Mathématiquement, il faut toutefois répondre à deux questions.

Première question : la somme infinie a-t-elle un sens ? Deuxième question : si X est stationnaire, la série filtrée Y reste-t-elle stationnaire ? Le théorème de filtrage répond à ces deux questions sous une condition simple sur les coefficients.

Définition 3.4.1: Suite absolument sommable

On note $\ell^1(\mathbb{Z})$ l'ensemble des suites réelles $a = (a_j)_{j\in\mathbb{Z}}$ telles que

$$\sum_{j\in\mathbb{Z}} |a_j| < \infty.$$

Une telle suite est dite *absolument sommable*.

Idée à retenir

La condition $a \in \ell^1(\mathbb{Z})$ signifie que l'influence totale des coefficients est contrôlée. Les poids peuvent être en nombre infini, mais leur masse absolue totale doit rester finie. C'est ce qui empêche les valeurs lointaines d'accumuler une influence incontrôlable.

Théorème 3.4.2: Théorème général de filtrage

Soit $X = (X_t)_{t\in\mathbb{Z}}$ un processus stationnaire du second ordre de moyenne μ et d'autocovariance γ_X. Soit $a = (a_j)_{j\in\mathbb{Z}} \in \ell^1(\mathbb{Z})$. Alors, pour tout $t \in \mathbb{Z}$, la série

$$Y_t = \sum_{j\in\mathbb{Z}} a_j X_{t-j}$$

converge absolument dans L^2, donc converge dans L^2. Le processus $Y = (Y_t)_{t\in\mathbb{Z}}$ ainsi défini est stationnaire du second ordre. Sa moyenne est

$$\mathbb{E}(Y_t) = \mu \sum_{j\in\mathbb{Z}} a_j,$$

et son autocovariance est donnée par

$$\gamma_Y(h) = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} a_j a_k \gamma_X(h+j-k), \qquad h \in \mathbb{Z}.$$

Démonstration. Comme X est stationnaire du second ordre, $\mathbb{E}(X_t^2)$ ne dépend pas de t. Posons $M = \|X_0\|_{L^2} < \infty$. Alors

$$\sum_{j\in\mathbb{Z}} \|a_j X_{t-j}\|_{L^2} = \sum_{j\in\mathbb{Z}} |a_j| \, \|X_{t-j}\|_{L^2} = M \sum_{j\in\mathbb{Z}} |a_j| < \infty.$$

La série converge donc absolument dans l'espace normé L^2, et donc dans L^2.

Pour la moyenne, on peut passer à la limite dans L^2, donc aussi dans L^1, et utiliser la linéarité de l'espérance :

$$\mathbb{E}(Y_t) = \sum_{j\in\mathbb{Z}} a_j \mathbb{E}(X_{t-j}) = \mu \sum_{j\in\mathbb{Z}} a_j.$$

Cette quantité ne dépend pas de t.

Pour la covariance, on commence par les sommes partielles

$$Y_t^{(n)} = \sum_{|j|\le n} a_j X_{t-j}.$$

Pour ces sommes finies,

$$\begin{aligned}\operatorname{Cov}\left(Y_t^{(n)}, Y_{t+h}^{(n)}\right) &= \sum_{|j|\le n} \sum_{|k|\le n} a_j a_k \operatorname{Cov}\left(X_{t-j}, X_{t+h-k}\right) \\ &= \sum_{|j|\le n} \sum_{|k|\le n} a_j a_k \gamma_X(h+j-k).\end{aligned}$$

En passant à la limite dans L^2, on obtient la formule annoncée. La double somme est absolument convergente car $|\gamma_X(r)| \le \gamma_X(0)$ pour tout r, donc

$$\sum_{j,k} |a_j a_k \gamma_X(h+j-k)| \le \gamma_X(0) \left(\sum_j |a_j|\right)^2 < \infty.$$

Enfin, la moyenne et la covariance de Y ne dépendent du temps qu'à travers le décalage h. Le processus Y est donc stationnaire du second ordre. □

Remarque 3.4.3: Pourquoi ce théorème arrive ici ?

Le théorème ne crée pas une nouvelle méthode. Il justifie rigoureusement une idée déjà rencontrée sur des exemples : construire une nouvelle série en mélangeant des valeurs de l'ancienne. Les filtres finis sont inclus dans le théorème, car une suite à support fini est automatiquement dans $\ell^1(\mathbb{Z})$. Le gain est que l'on peut maintenant autoriser une infinité de poids.

3.5 Composition des filtres et convolution

Une fois un filtre défini, on peut naturellement en appliquer un deuxième. Par exemple, on peut lisser une série puis la différencier, ou différencier puis lisser. La question est alors : quel filtre global a-t-on réellement appliqué ? La réponse est donnée par la convolution des coefficients.

Définition 3.5.1: Convolution de deux suites

Soient $a = (a_j)_{j\in\mathbb{Z}}$ et $c = (c_j)_{j\in\mathbb{Z}}$ deux suites absolument sommables. Leur *convolution* est la suite $a * c$ définie par

$$(a * c)_\ell = \sum_{j\in\mathbb{Z}} a_j c_{\ell-j}, \qquad \ell \in \mathbb{Z}.$$

Dans le cas des filtres finis, la définition est simplement une somme finie. La convolution dit comment les coefficients se combinent quand on applique deux filtres successivement.

Théorème 3.5.2: Composition des filtres et convolution

Soient $a, c \in \ell^1(\mathbb{Z})$. Alors $a * c \in \ell^1(\mathbb{Z})$, et pour tout processus stationnaire du second ordre X,

$$a(B)\big(c(B)X\big) = (a * c)(B)X.$$

Démonstration. On commence par montrer que la convolution est absolument sommable :

$$\begin{aligned}
\sum_{\ell\in\mathbb{Z}} |(a * c)_\ell| &\le \sum_{\ell\in\mathbb{Z}}\sum_{j\in\mathbb{Z}} |a_j|\,|c_{\ell-j}| \\
&= \sum_{j\in\mathbb{Z}} |a_j| \sum_{\ell\in\mathbb{Z}} |c_{\ell-j}| \\
&= \left(\sum_{j\in\mathbb{Z}} |a_j|\right)\left(\sum_{r\in\mathbb{Z}} |c_r|\right) < \infty.
\end{aligned}$$

On peut donc appliquer le théorème de filtrage. En utilisant la convergence absolue dans L^2, on peut réordonner les sommes :

$$\begin{aligned}
\big(a(B)(c(B)X)\big)_t &= \sum_{j\in\mathbb{Z}} a_j (c(B)X)_{t-j} \\
&= \sum_{j\in\mathbb{Z}} a_j \sum_{k\in\mathbb{Z}} c_k X_{t-j-k} \\
&= \sum_{\ell\in\mathbb{Z}} \left(\sum_{j\in\mathbb{Z}} a_j c_{\ell-j}\right) X_{t-\ell} \\
&= \big((a * c)(B)X\big)_t.
\end{aligned}$$

□

Exemple 3.5.3: Différence puis moyenne mobile

Considérons la moyenne mobile passée sur deux valeurs

$$M(B) = \frac{1}{2}(1 + B)$$

et la différence ordinaire

$$\Delta(B) = 1 - B.$$

Le filtre composé vaut

$$M(B)\Delta(B) = \frac{1}{2}(1 + B)(1 - B) = \frac{1}{2}(1 - B^2).$$

Donc lisser sur deux valeurs puis différencier revient ici à comparer X_t à X_{t-2}, avec un facteur $1/2$. Ce calcul explique pourquoi il ne faut pas voir deux commandes successives dans `R` comme deux opérations indépendantes : ensemble, elles forment un nouveau filtre.

3.6 Filtrage d'un processus stationnaire et bruit blanc filtré

3.6.1 Cas des filtres finis

Le théorème général de filtrage s'applique aux filtres finis. Dans ce cas, les formules sont plus simples à lire, car toutes les sommes sont finies.

Proposition 3.6.1: Filtrage fini d'un processus stationnaire

Soit X un processus stationnaire du second ordre de moyenne μ et d'autocovariance γ_X. Soit

$$Y_t = \sum_{j=-m}^{m} a_j X_{t-j}.$$

Alors Y est stationnaire du second ordre,

$$\mathbb{E}(Y_t) = \mu \sum_{j=-m}^{m} a_j,$$

et

$$\gamma_Y(h) = \sum_{j=-m}^{m} \sum_{k=-m}^{m} a_j a_k \gamma_X(h + j - k).$$

Ce résultat est important : filtrer une série stationnaire par un filtre linéaire à coefficients contrôlés ne détruit pas la stationnarité du second ordre. Cela ne veut pas dire que le filtre rend stationnaire une série qui ne l'était pas ; cela veut dire qu'il conserve cette propriété lorsqu'elle est déjà présente.

3.6.2 Bruit blanc filtré

Le cas le plus instructif est celui où la série initiale est un bruit blanc. Un bruit blanc n'a pas de corrélation temporelle. Après filtrage, la série obtenue peut en avoir.

Proposition 3.6.2: Autocovariance d'un bruit blanc filtré

Soit $Z = (Z_t)_{t\in\mathbb{Z}}$ un bruit blanc centré de variance σ^2, et soit

$$Y_t = \sum_j a_j Z_{t-j},$$

où le filtre (a_j) est fini. Alors Y est stationnaire, centré, et

$$\gamma_Y(h) = \sigma^2 \sum_j a_j a_{j+h},$$

où l'on pose $a_j = 0$ lorsque j est hors du support du filtre.

Démonstration. On applique le corollaire précédent avec $\gamma_Z(r) = \sigma^2 \mathbf{1}_{\{r=0\}}$. On obtient

$$\gamma_Y(h) = \sum_j \sum_k a_j a_k \sigma^2 \mathbf{1}_{\{h+j-k=0\}}.$$

La condition $h + j - k = 0$ équivaut à $k = j + h$. Donc

$$\gamma_Y(h) = \sigma^2 \sum_j a_j a_{j+h}.$$

□

Exemple 3.6.3: Le filtre $1 + \theta B$

Soit

$$Y_t = Z_t + \theta Z_{t-1},$$

où Z est un bruit blanc centré de variance σ^2. Les seuls coefficients non nuls sont $a_0 = 1$ et $a_1 = \theta$. On obtient

$$\gamma_Y(0) = (1+\theta^2)\sigma^2, \qquad \gamma_Y(1) = \theta\sigma^2, \qquad \gamma_Y(h) = 0 \quad \text{si } |h| > 1.$$

Donc l'autocorrélation vaut

$$\rho_Y(1) = \frac{\theta}{1+\theta^2}, \qquad \rho_Y(h) = 0 \quad \text{si } |h| > 1.$$

Ce calcul donne une première signature graphique d'un modèle : une dépendance au retard 1, puis plus rien au-delà. Cette idée sera centrale lorsque nous étudierons les modèles à moyenne mobile.

3.7 Séries linéaires causales

Le théorème général de filtrage impose une condition ℓ^1 sur les coefficients pour filtrer un processus stationnaire quelconque. Dans le cas particulier où l'on filtre un bruit blanc, on peut utiliser une condition plus faible, adaptée au second ordre : la sommabilité des carrés.

Définition 3.7.1: Suite de carré sommable

On note $\ell^2(\mathbb{N})$ l'ensemble des suites $\psi = (\psi_j)_{j\geq 0}$ telles que

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

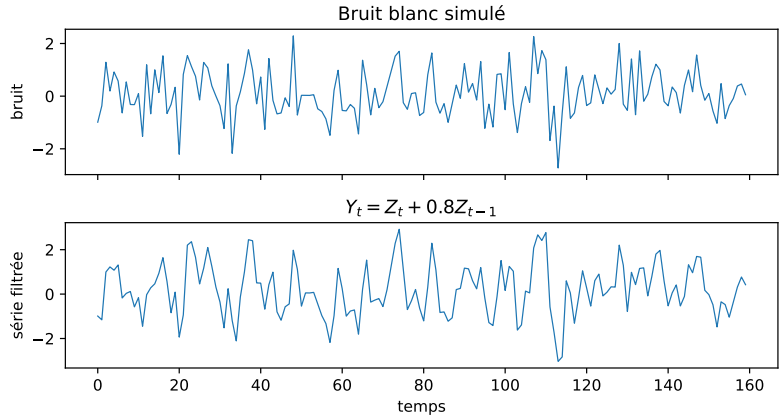


Figure 3.4 – Un bruit blanc simulé et la série filtrée $Y_t = Z_t + 0.8Z_{t-1}$. Deux valeurs consécutives de Y partagent un même choc, ce qui crée une dépendance temporelle visible.

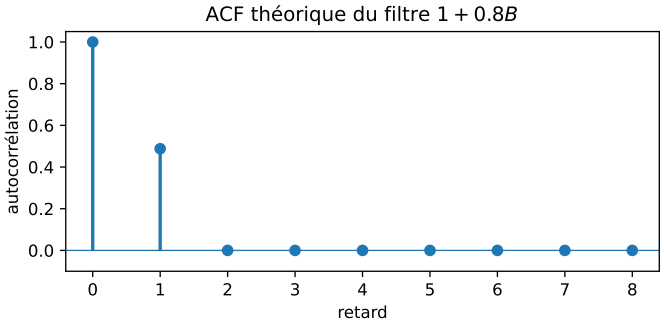


Figure 3.5 – Autocorrélation théorique du filtre $1 + 0.8B$ appliqué à un bruit blanc. La corrélation est non nulle au retard 1 et nulle ensuite.

Définition 3.7.2: Série linéaire causale

Soit $Z = (Z_t)_{t\in\mathbb{Z}}$ un bruit blanc centré de variance σ^2. Soit $(\psi_j)_{j\geq 0} \in \ell^2(\mathbb{N})$. La *série linéaire causale* associée à (ψ_j) est le processus défini dans L^2 par

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}.$$

La condition ℓ^2 signifie que l'énergie totale des coefficients est finie. Elle suffit ici parce que les variables Z_t sont orthogonales deux à deux : les variances des morceaux s'additionnent.

Théorème 3.7.3: Convergence, stationnarité et autocovariance

Sous les hypothèses de la définition précédente, la somme définissant X_t converge dans L^2. Le processus X est stationnaire du second ordre, centré, et pour $h \geq 0$,

$$\gamma_X(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}.$$

On complète par symétrie : $\gamma_X(-h) = \gamma_X(h)$.

Démonstration. Pour $n \geq 0$, posons

$$S_{n,t} = \sum_{j=0}^{n} \psi_j Z_{t-j}.$$

Si $m < n$, alors, par orthogonalité des variables d'un bruit blanc,

$$\mathbb{E}\left((S_{n,t} - S_{m,t})^2\right) = \mathbb{E}\left(\left(\sum_{j=m+1}^{n} \psi_j Z_{t-j}\right)^2\right) = \sigma^2 \sum_{j=m+1}^{n} \psi_j^2.$$

Comme $\sum_j \psi_j^2 < \infty$, le membre de droite tend vers 0 lorsque $m, n \to \infty$. Les sommes partielles sont donc de Cauchy dans L^2, et convergent dans L^2.

La moyenne de la limite est nulle. Pour $h \geq 0$, en passant à la limite dans les covariances des sommes partielles, seuls les termes vérifiant $t - j = t + h - k$, c'est-à-dire $k = j + h$, contribuent. On obtient

$$\mathrm{Cov}\,(X_t, X_{t+h}) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}.$$

Cette quantité dépend de h, mais pas de t. Le processus est donc stationnaire du second ordre. □

3.7.1 Cas particulier : poids géométriques

Un exemple fondamental est donné par

$$\psi_j = \phi^j, \qquad j \geq 0,$$

où $|\phi| < 1$. Alors $\sum_j \psi_j^2 < \infty$, car

$$\sum_{j=0}^{\infty} \phi^{2j} = \frac{1}{1 - \phi^2}.$$

Le processus

$$X_t = Z_t + \phi Z_{t-1} + \phi^2 Z_{t-2} + \cdots$$

a une mémoire longue mais décroissante : les chocs anciens comptent encore, mais leur poids diminue géométriquement.

Proposition 3.7.4: Autocovariance du filtre géométrique

Si $X_t = \sum_{j \geq 0} \phi^j Z_{t-j}$, avec $|\phi| < 1$, alors

$$\gamma_X(h) = \frac{\sigma^2 \phi^{|h|}}{1 - \phi^2}, \qquad h \in \mathbb{Z}.$$

Démonstration. Pour $h \geq 0$, le théorème précédent donne

$$\gamma_X(h) = \sigma^2 \sum_{j=0}^{\infty} \phi^j \phi^{j+h} = \sigma^2 \phi^h \sum_{j=0}^{\infty} \phi^{2j} = \frac{\sigma^2 \phi^h}{1 - \phi^2}.$$

La formule pour $h < 0$ s'obtient par symétrie de l'autocovariance. □

Ce cas servira de pont vers les modèles autorégressifs. Si

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j},$$

alors, au sens des séries convergentes dans L^2, on vérifie que

$$X_t - \phi X_{t-1} = Z_t.$$

On obtient ainsi une équation où le présent dépend de la valeur passée et d'un nouveau choc. Cette idée sera reprise lorsque nous introduirons les processus autorégressifs.

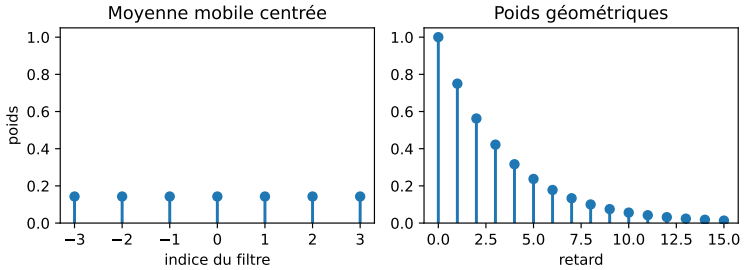


Figure 3.6 – Poids d'une moyenne mobile finie et poids géométriques. Un filtre fini oublie totalement les valeurs au-delà d'un certain retard ; un filtre géométrique n'oublie jamais totalement, mais l'influence des anciens chocs décroît.

3.8 Filtres dans R

Dans R, on utilise surtout deux fonctions de base.

La fonction `stats::filter()` applique un filtre linéaire. On écrit `stats::filter()` plutôt que `filter()` pour éviter une ambiguïté avec d'autres packages qui définissent aussi une fonction `filter()`.

```
y <- stats::filter(x, filter = rep(1/5, 5), sides = 1)
```

Cette commande calcule une moyenne mobile passée sur 5 valeurs. L'argument `sides = 1` indique que le filtre utilise le présent et le passé.

```
y <- stats::filter(x, filter = rep(1/5, 5), sides = 2)
```

Cette commande calcule une moyenne mobile centrée. L'argument `sides = 2` indique que le filtre utilise des valeurs avant et après la date courante. Aux bords de la série, certaines valeurs ne peuvent pas être calculées : R les note `NA`.

La fonction `diff()` calcule des différences :

```
dx <- diff(x)
d12x <- diff(x, lag = 12)
```

La première ligne calcule $x_t - x_{t-1}$. La seconde calcule $x_t - x_{t-12}$, ce qui correspond à une différence saisonnière pour une série mensuelle.

Point de vigilance

Après un filtrage, il faut toujours regarder les bords de la série. Une moyenne mobile centrée produit des `NA` au début et à la fin ; une différence supprime les premières observations nécessaires au calcul. Ces valeurs manquantes ne sont pas des erreurs : elles viennent de la définition même du filtre.

3.9 Synthèse du chapitre

Ce chapitre a introduit les filtres linéaires comme opérations fondamentales sur les séries temporelles. À retenir :

1) Un filtre linéaire construit une nouvelle série en combinant linéairement plusieurs valeurs de la série initiale.
2) Les coefficients du filtre disent quelles dates sont utilisées et avec quels poids.
3) Un filtre causal n'utilise que le présent et le passé ; un filtre centré peut utiliser des valeurs futures.
4) La somme des coefficients indique l'effet du filtre sur les constantes.
5) Les différences $1 - B$ et $1 - B^s$ sont des filtres particuliers.

6) Le théorème général de filtrage explique quand un filtre infini est bien défini et conserve la stationnarité du second ordre.
7) La composition de deux filtres correspond à la convolution de leurs coefficients.
8) Un bruit blanc filtré peut devenir une série dépendante dans le temps.
9) Une série linéaire est une combinaison, éventuellement infinie, de chocs présents et passés.

Le chapitre suivant introduira les modèles MA, AR et ARMA en utilisant ce langage : une partie de la dépendance temporelle sera décrite par des filtres appliqués à des bruits blancs, et une autre partie par des équations reliant X_t à ses valeurs passées.

Chapitre 4

Modèles MA, AR et ARMA stationnaires

Mots-clés. Modèle MA ; modèle AR ; modèle ARMA ; innovation ; causalité ; inversibilité ; polynôme autorégressif ; polynôme moyenne mobile ; filtre rationnel ; fraction rationnelle ; racines hors du disque unité ; fonction d'autocorrélation ; autocorrélation partielle ; identification.

4.1 Pourquoi introduire les modèles ARMA ?

Les chapitres précédents ont progressivement séparé deux niveaux de travail. D'abord, on observe une série réelle et l'on cherche à retirer ce qui est manifestement non stationnaire : tendance, saisonnalité, changement d'échelle, valeurs manquantes. Ensuite, une fois cette première étape faite, il reste souvent une série qui fluctue autour d'un niveau à peu près constant, mais dont les observations ne sont pas indépendantes. C'est cette dépendance temporelle résiduelle que les modèles MA, AR et ARMA cherchent à représenter.

Il faut donc garder en tête le rôle de ce chapitre : on ne modélise pas ici une série qui monte fortement, ou une série mensuelle avec une saisonnalité annuelle évidente. On modélise la partie que l'on espère stationnaire après les transformations appropriées. Par exemple, dans une série économique, on pourra d'abord prendre un logarithme puis une différence ; dans une série saisonnière, on pourra d'abord retirer une composante saisonnière. Les modèles ARMA arrivent après cette préparation.

Le chapitre 3 a introduit les filtres linéaires. Un filtre transforme une série en combinant des valeurs présentes, passées ou futures. Les modèles de ce chapitre reprennent exactement cette idée, mais avec une interprétation probabiliste : la série observée est construite à partir d'une suite de chocs aléatoires non corrélés, appelés innovations.

Trois mécanismes doivent être distingués.

1) Un modèle *moyenne mobile* MA considère que la série est une combinaison finie de chocs récents. Un choc a donc une durée de vie limitée.
2) Un modèle *autorégressif* AR considère que la valeur présente dépend directement des valeurs passées de la série. Un choc peut alors se propager par récurrence.
3) Un modèle ARMA combine les deux : une partie récursive AR et une partie moyenne mobile MA.

La différence entre ces trois mécanismes est essentielle. Un MA est un filtre fini appliqué à un bruit blanc. Un AR est défini par une équation de récurrence que l'on doit résoudre. Un ARMA est une équation de récurrence dont le second membre est lui-même un filtre fini du bruit. C'est pourquoi les polynômes en l'opérateur retard et les fractions rationnelles vont apparaître naturellement.

Point de vigilance

Dans tout le chapitre, sauf mention contraire, $Z = (Z_t)_{t\in\mathbb{Z}}$ désigne un bruit blanc centré de variance σ^2, c'est-à-dire

$$\mathbb{E}(Z_t) = 0, \qquad \mathrm{Cov}(Z_t, Z_{t+h}) = \sigma^2 \mathbf{1}_{\{h=0\}}.$$

Le mot *innovation* désigne l'idée suivante : au temps t, Z_t représente la partie nouvelle qui n'est pas expliquée par le passé du modèle.

Remarque 4.1.1: Deux acronymes utilisés dans ce chapitre

On utilisera parfois deux acronymes classiques, très fréquents dans les sorties de logiciels.

— **ACF** signifie *autocorrelation function*, c'est-à-dire *fonction d'autocorrélation*. Dans ce poly, il s'agit de la fonction déjà introduite au chapitre 1 :

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

Elle mesure la dépendance linéaire entre deux observations séparées de h dates.

— **PACF** signifie *partial autocorrelation function*, c'est-à-dire *fonction d'autocorrélation partielle*. Elle sera définie précisément plus loin ; son rôle est de mesurer la dépendance directe au retard h, après avoir retiré l'effet des retards intermédiaires.

Ces deux objets ne sont donc pas des outils mystérieux : ils servent à lire, dans les corrélations, la forme de mémoire produite par un modèle MA, AR ou ARMA.

4.2 Processus moyenne mobile MA

4.2.1 Définition et intuition

On commence par le modèle le plus directement lié au chapitre précédent. Un MA est simplement un filtre fini appliqué à un bruit blanc.

Définition 4.2.1: Processus MA(q)

Soit $q \in \mathbb{N}$. Un processus $X = (X_t)_{t\in\mathbb{Z}}$ est un processus moyenne mobile d'ordre q, ou MA(q), s'il s'écrit

$$X_t = \mu + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \qquad t \in \mathbb{Z},$$

où $\mu \in \mathbb{R}$, $\theta_1, \ldots, \theta_q \in \mathbb{R}$, et Z est un bruit blanc. En posant $\theta_0 = 1$, on peut aussi écrire

$$X_t = \mu + \sum_{j=0}^{q} \theta_j Z_{t-j}.$$

En notation opérateur retard,

$$X_t = \mu + \Theta(B) Z_t, \qquad \Theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q.$$

Le nom « moyenne mobile » ne signifie pas que les coefficients soient positifs ou qu'ils somment à 1. Il signifie que, à chaque date, on combine un petit nombre de chocs récents, puis que cette même combinaison est déplacée dans le temps. Le modèle a une mémoire finie : le choc Z_t intervient dans $X_t, X_{t+1}, \ldots, X_{t+q}$, puis n'intervient plus.

4.2.2 Stationnarité et autocovariance d'un MA(q)

Proposition 4.2.2: Stationnarité d'un MA(q)

Tout processus MA(q) est stationnaire du second ordre. Sa moyenne vaut μ. Si $\theta_0 = 1$, alors son autocovariance est donnée, pour $h \geq 0$, par

$$\gamma_X(h) = \begin{cases} \sigma^2 \displaystyle\sum_{j=0}^{q-h} \theta_j \theta_{j+h}, & 0 \leq h \leq q, \\ 0, & h > q. \end{cases}$$

On complète par symétrie : $\gamma_X(-h) = \gamma_X(h)$.

Démonstration. La moyenne est

$$\mathbb{E}(X_t) = \mu + \sum_{j=0}^{q} \theta_j \mathbb{E}(Z_{t-j}) = \mu.$$

Elle ne dépend donc pas de t. Pour $h \geq 0$,

$$\begin{aligned} \operatorname{Cov}(X_t, X_{t+h}) &= \operatorname{Cov}\left(\sum_{j=0}^{q} \theta_j Z_{t-j}, \sum_{k=0}^{q} \theta_k Z_{t+h-k} \right) \\ &= \sum_{j=0}^{q} \sum_{k=0}^{q} \theta_j \theta_k \operatorname{Cov}(Z_{t-j}, Z_{t+h-k}). \end{aligned}$$

La covariance du bruit blanc est nulle sauf si $t - j = t + h - k$, c'est-à-dire si $k = j + h$. Lorsque $0 \leq h \leq q$, cela donne

$$\gamma_X(h) = \sigma^2 \sum_{j=0}^{q-h} \theta_j \theta_{j+h}.$$

Si $h > q$, les deux combinaisons ne partagent aucun choc commun, et la covariance est nulle. La covariance ne dépend que de l'écart h, donc X est stationnaire du second ordre. □

Idée à retenir

Pour un MA(q), deux observations séparées de plus de q périodes ne partagent aucun choc commun. Leur covariance est donc nulle. C'est la signature essentielle d'un MA : son ACF théorique est coupée après le retard q.

4.2.3 Le cas MA(1) et le problème d'inversibilité

Le cas MA(1) est le premier exemple à connaître parfaitement :

$$X_t = \mu + Z_t + \theta Z_{t-1}.$$

D'après la proposition précédente,

$$\gamma_X(0) = \sigma^2(1 + \theta^2), \qquad \gamma_X(1) = \sigma^2 \theta, \qquad \gamma_X(h) = 0 \quad \text{si } |h| \geq 2.$$

Ainsi,

$$\rho_X(1) = \frac{\theta}{1 + \theta^2}, \qquad \rho_X(h) = 0 \quad \text{si } |h| \geq 2.$$

Ce résultat illustre un point important : $\rho_X(1)$ ne détermine pas toujours θ de manière unique, car

$$\frac{\theta}{1 + \theta^2} = \frac{1/\theta}{1 + 1/\theta^2} \qquad (\theta \neq 0).$$

Les deux valeurs θ et $1/\theta$ donnent donc la même autocorrélation au retard 1. Pour choisir une écriture canonique, on impose une condition d'inversibilité.

Pour un MA(1) centré,

$$X_t = (1 + \theta B)Z_t.$$

Si $|\theta| < 1$, alors la série géométrique donne

$$(1 + \theta B)^{-1} = 1 - \theta B + \theta^2 B^2 - \theta^3 B^3 + \cdots,$$

et donc

$$Z_t = X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \theta^3 X_{t-3} + \cdots.$$

On dit que le MA(1) est inversible : l'innovation présente peut être reconstruite à partir du présent et du passé de la série observée.

Proposition 4.2.3: Condition d'inversibilité d'un MA(q)

Un processus MA(q) est inversible si les racines du polynôme

$$\Theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$$

sont toutes de module strictement supérieur à 1. Dans ce cas, il existe des coefficients $(\pi_j)_{j\geq 0}$ absolument sommables tels que

$$Z_t = \sum_{j=0}^{\infty} \pi_j (X_{t-j} - \mu).$$

Idée de la preuve. Si les racines de Θ sont hors du disque unité, alors $1/\Theta(z)$ admet un développement en série entière autour de 0 dont le rayon de convergence est strictement supérieur à 1. Les coefficients de ce développement sont absolument sommables, ce qui permet de remplacer z par l'opérateur retard B et d'obtenir une représentation convergente de Z_t en fonction de $X_t, X_{t-1}, \ldots$. □

4.3 Processus autorégressifs AR

4.3.1 Définition et comparaison avec les MA

Dans un MA, la série est explicitement donnée comme un filtre fini du bruit. Dans un AR, la série est définie implicitement par une équation reliant le présent au passé. Cela change beaucoup l'interprétation : le choc Z_t n'affecte pas seulement X_t, il peut se transmettre à X_{t+1}, puis à X_{t+2}, etc., à travers la récurrence.

Définition 4.3.1: Processus AR(p)

Soit $p \geq 1$. Un processus $X = (X_t)_{t\in\mathbb{Z}}$ est un processus autorégressif d'ordre p, ou AR(p), s'il vérifie

$$X_t - c = \phi_1(X_{t-1} - c) + \cdots + \phi_p(X_{t-p} - c) + Z_t,$$

où c est le niveau moyen autour duquel fluctue le processus. Dans le cas centré $c = 0$, on écrit

$$\Phi(B)X_t = Z_t, \qquad \Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p.$$

Le terme *autorégressif* signifie que l'on régresse le présent sur le passé de la même série. La nouveauté par rapport à une régression linéaire classique est que le passé est aléatoire et fait partie du même processus. L'erreur Z_t est l'innovation : ce qui reste au temps t après utilisation du passé inclus dans l'équation.

4.3.2 Le cas AR(1)

Le modèle AR(1) centré est

$$X_t = \phi X_{t-1} + Z_t.$$

Il contient déjà l'idée essentielle. Si $\phi > 0$, une valeur positive tend à être suivie par une valeur positive. Si $\phi < 0$, les signes tendent à alterner. Si $|\phi|$ est proche de 1, le retour vers zéro est lent.

Théorème 4.3.2: Solution stationnaire causale d'un AR(1)

Si $|\phi| < 1$, alors l'équation

$$X_t = \phi X_{t-1} + Z_t$$

admet une solution stationnaire causale donnée par

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}.$$

Cette série converge dans L^2. De plus,

$$\gamma_X(h) = \frac{\sigma^2 \phi^{|h|}}{1-\phi^2}, \qquad \rho_X(h) = \phi^{|h|}.$$

Démonstration. Comme $|\phi| < 1$, la suite $(\phi^j)_{j \geq 0}$ est de carré sommable. Le résultat sur les séries linéaires filtrant un bruit blanc assure donc que

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$$

converge dans L^2 et définit un processus stationnaire. De plus,

$$\begin{aligned} X_t - \phi X_{t-1} &= \sum_{j=0}^{\infty} \phi^j Z_{t-j} - \phi \sum_{j=0}^{\infty} \phi^j Z_{t-1-j} \\ &= Z_t + \sum_{j=1}^{\infty} \phi^j Z_{t-j} - \sum_{j=1}^{\infty} \phi^j Z_{t-j} = Z_t. \end{aligned}$$

Enfin, pour $h \geq 0$,

$$\begin{aligned} \gamma_X(h) &= \operatorname{Cov}\left(\sum_{j=0}^{\infty} \phi^j Z_{t-j}, \sum_{k=0}^{\infty} \phi^k Z_{t+h-k} \right) \\ &= \sigma^2 \sum_{j=0}^{\infty} \phi^j \phi^{j+h} = \sigma^2 \phi^h \sum_{j=0}^{\infty} \phi^{2j} = \frac{\sigma^2 \phi^h}{1-\phi^2}. \end{aligned}$$

La symétrie de l'autocovariance donne la formule pour $h < 0$, puis on divise par $\gamma_X(0) = \sigma^2/(1-\phi^2)$. □

Point de vigilance

Le cas $\phi = 1$ donne la marche aléatoire $X_t = X_{t-1} + Z_t$, dont la variance croît avec le temps. Ce n'est pas un modèle stationnaire. Le chapitre sur les modèles ARIMA reviendra sur ce cas, car la différenciation transforme une marche aléatoire en bruit blanc.

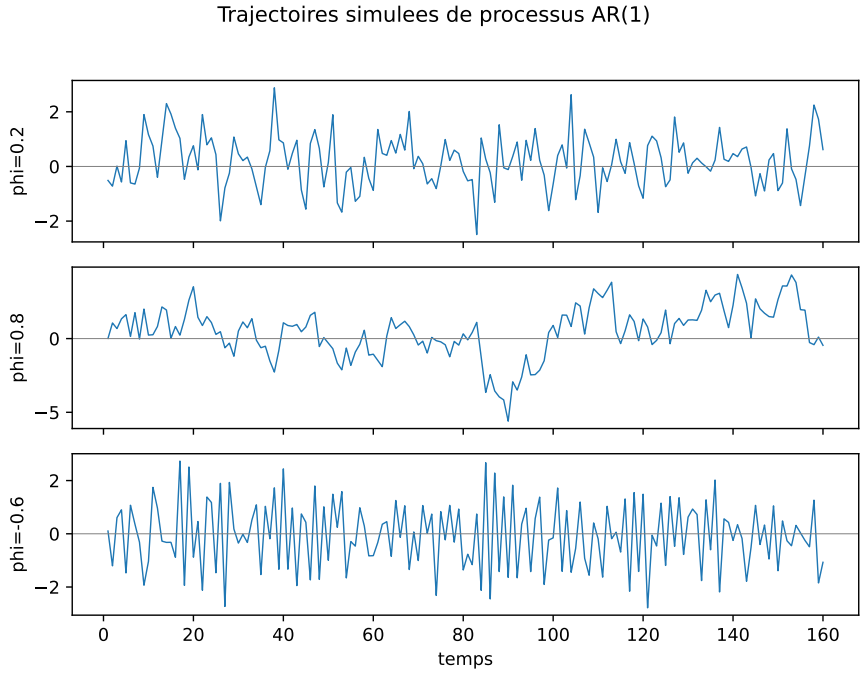


Figure 4.1 – Trajectoires simulées de trois AR(1). Quand $\phi = 0.8$, les valeurs restent longtemps du même côté de zéro. Quand $\phi = -0.6$, on observe une alternance plus marquée.

4.3.3 AR(p), racines et développement rationnel

Pour $p \geq 1$, l'équation centrée d'un AR(p) s'écrit

$$\Phi(B)X_t = Z_t, \qquad \Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p.$$

Pour résoudre cette équation, on cherche à inverser $\Phi(B)$:

$$X_t = \frac{1}{\Phi(B)} Z_t.$$

Cette notation n'est pas une simple décoration formelle. Elle signifie que l'on veut développer la fraction rationnelle $1/\Phi(z)$ en série entière autour de 0, puis remplacer z par B. Si

$$\frac{1}{\Phi(z)} = \sum_{j=0}^{\infty} \psi_j z^j,$$

alors la solution candidate est

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}.$$

La question de stationnarité devient alors une question de convergence des coefficients ψ_j.

Théorème 4.3.3: Condition de causalité d'un AR(p)

Si les racines du polynôme autorégressif

$$\Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$$

sont toutes de module strictement supérieur à 1, alors l'équation $\Phi(B)X_t = Z_t$ admet une solution stationnaire causale. Cette solution peut s'écrire

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \qquad \sum_{j=0}^{\infty} |\psi_j| < \infty.$$

Démonstration. Si toutes les racines de Φ sont de module strictement supérieur à 1, il existe $r > 1$ tel que $\Phi(z) \neq 0$ pour tout $|z| < r$. La fonction $1/\Phi(z)$ est donc analytique sur le disque $|z| < r$, et admet un développement

$$\frac{1}{\Phi(z)} = \sum_{j=0}^{\infty} \psi_j z^j$$

avec un rayon de convergence strictement supérieur à 1. En particulier, les coefficients décroissent au moins géométriquement le long d'un rayon r_0 tel que $1 < r_0 < r$, ce qui implique $\sum_j |\psi_j| < \infty$. Le théorème de filtrage assure alors que $\sum_j \psi_j Z_{t-j}$ définit un processus stationnaire. Enfin, l'identité

$$\Phi(z) \sum_{j=0}^{\infty} \psi_j z^j = 1$$

se traduit, par composition de filtres, en $\Phi(B)X_t = Z_t$. □

Idée à retenir

La condition sur les racines dit que les pôles de la fraction rationnelle $1/\Phi(z)$ doivent être à l'extérieur du disque unité. Si un pôle se rapproche du cercle unité, les coefficients ψ_j décroissent lentement : les chocs anciens persistent longtemps. Si un pôle entre dans le disque unité, le développement causal n'est plus stable.

4.3.4 Calcul des coefficients par identification

Dans la pratique, on n'a pas toujours besoin de factoriser complètement Φ pour calculer les premiers coefficients du filtre causal. On peut les obtenir en identifiant les coefficients dans

$$\Phi(z)\Psi(z) = 1, \qquad \Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j.$$

Par exemple, pour un AR(2),

$$\Phi(z) = 1 - \phi_1 z - \phi_2 z^2.$$

L'identité

$$(1 - \phi_1 z - \phi_2 z^2)(\psi_0 + \psi_1 z + \psi_2 z^2 + \cdots) = 1$$

donne

$$\psi_0 = 1, \qquad \psi_1 = \phi_1, \qquad \psi_j = \phi_1 \psi_{j-1} + \phi_2 \psi_{j-2} \quad (j \geq 2).$$

On retrouve donc une idée très naturelle : les coefficients de la représentation causale obéissent à la même récurrence linéaire que le modèle.

4.3.5 Équations de Yule–Walker

Les autocovariances d'un AR(p) satisfont aussi une récurrence. En multipliant l'équation

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + Z_t$$

par X_{t-h} et en prenant l'espérance, on obtient, pour $h \geq 1$,

$$\gamma(h) = \phi_1 \gamma(h-1) + \cdots + \phi_p \gamma(h-p),$$

car l'innovation Z_t est non corrélée au passé. Ces équations sont appelées *équations de Yule–Walker*.

Pour un AR(1), elles donnent immédiatement

$$\gamma(h) = \phi \gamma(h-1),$$

donc $\rho(h) = \phi^h$ pour $h \geq 0$, ce qui retrouve la formule précédente.

4.4 Des AR et MA aux ARMA : les filtres rationnels

4.4.1 Pourquoi combiner AR et MA ?

Les modèles MA et AR capturent deux formes simples de dépendance. Les MA donnent une mémoire finie : l'ACF s'annule après un certain retard. Les AR donnent une mémoire infinie mais structurée : les effets des chocs décroissent selon une récurrence. En pratique, beaucoup de séries stationnaires ont une dépendance qui n'est ni parfaitement coupée, ni purement autorégressive d'ordre très petit. Les ARMA servent alors de compromis : ils combinent une partie récursive et une partie moyenne mobile.

Le point de vue algébrique est particulièrement clair. Un MA s'écrit

$$X_t - \mu = \Theta(B)Z_t.$$

Un AR s'écrit

$$\Phi(B)(X_t - \mu) = Z_t.$$

Un ARMA remplace simplement le bruit blanc du membre de droite par un bruit filtré :

$$\Phi(B)(X_t - \mu) = \Theta(B)Z_t.$$

Ainsi, lorsque l'équation est causale,

$$X_t - \mu = \frac{\Theta(B)}{\Phi(B)}Z_t.$$

La fraction rationnelle $\Theta(z)/\Phi(z)$ est donc le filtre qui transforme les innovations Z_t en observations X_t.

Définition 4.4.1: Processus ARMA(p, q)

Un processus stationnaire $X = (X_t)_{t\in\mathbb{Z}}$ est un processus ARMA(p, q) s'il vérifie

$$\Phi(B)(X_t - \mu) = \Theta(B)Z_t,$$

où Z est un bruit blanc centré,

$$\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p, \qquad \Theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q.$$

On suppose en général que Φ et Θ n'ont pas de racine commune, afin d'éviter les écritures redondantes.

Remarque 4.4.2: Pourquoi exclure les racines communes ?

Si Φ et Θ ont un facteur commun, on peut parfois simplifier le modèle. Par exemple,

$$(1 - aB)X_t = (1 - aB)Z_t$$

se réduit formellement à $X_t = Z_t$, sous les conditions permettant la simplification. Écrire ce modèle comme un ARMA$(1, 1)$ masquerait donc le fait qu'il s'agit simplement d'un bruit blanc. La convention « pas de racine commune » évite ces représentations non minimales.

4.4.2 Causalité, inversibilité et fractions rationnelles

Théorème 4.4.3: Causalité et inversibilité d'un ARMA

Considérons un ARMA(p, q) sans racines communes entre Φ et Θ.

1) Si les racines de $\Phi(z)$ sont toutes de module strictement supérieur à 1, alors le processus est causal :
$$X_t - \mu = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \qquad \sum_{j=0}^{\infty} |\psi_j| < \infty.$$
Les coefficients ψ_j sont ceux du développement
$$\frac{\Theta(z)}{\Phi(z)} = \sum_{j=0}^{\infty} \psi_j z^j.$$

2) Si les racines de $\Theta(z)$ sont toutes de module strictement supérieur à 1, alors le processus est inversible :
$$Z_t = \sum_{j=0}^{\infty} \pi_j (X_{t-j} - \mu), \qquad \sum_{j=0}^{\infty} |\pi_j| < \infty.$$
Les coefficients π_j sont ceux du développement
$$\frac{\Phi(z)}{\Theta(z)} = \sum_{j=0}^{\infty} \pi_j z^j.$$

Démonstration. Pour la causalité, les racines de Φ hors du disque unité garantissent que la fraction rationnelle $\Theta(z)/\Phi(z)$ n'a pas de pôle dans un disque $|z| < r$ avec $r > 1$. Elle admet donc un développement en série entière dont les coefficients sont absolument sommables. Le théorème de filtrage donne la représentation causale.

Pour l'inversibilité, on applique le même raisonnement à la fraction $\Phi(z)/\Theta(z)$. Les racines de Θ hors du disque unité permettent d'exprimer Z_t comme un filtre causal stable appliqué au passé de X. □

Exemple 4.4.4: Un ARMA$(1, 1)$

Considérons
$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}, \qquad |\phi| < 1.$$
Alors
$$X_t = \frac{1 + \theta B}{1 - \phi B} Z_t.$$
Comme
$$\frac{1}{1 - \phi B} = 1 + \phi B + \phi^2 B^2 + \cdots,$$
on obtient
$$\begin{aligned} X_t &= (1 + \theta B)(1 + \phi B + \phi^2 B^2 + \cdots) Z_t \\ &= Z_t + (\phi + \theta) Z_{t-1} + \phi(\phi + \theta) Z_{t-2} + \phi^2(\phi + \theta) Z_{t-3} + \cdots. \end{aligned}$$
Ainsi, sauf cas particulier $\phi + \theta = 0$, un ARMA$(1, 1)$ a une mémoire infinie, mais géométriquement décroissante.

Idée à retenir

Le modèle ARMA$(1,1)$ montre très bien le rôle de la fraction rationnelle. Le numérateur $1+\theta B$ modifie les premiers effets du choc ; le dénominateur $1-\phi B$ propage ensuite ces effets dans le temps. C'est pourquoi un ARMA peut ressembler à un AR sur le long terme, tout en ayant une structure de court terme plus flexible.

4.5 ACF, PACF et identification

Cette section répond à une question pratique qui motive une grande partie des modèles ARMA : lorsque l'on observe une série que l'on suppose stationnaire, comment deviner un modèle plausible avant de lancer une estimation automatique ? On ne connaît pas les innovations Z_t, ni les paramètres du modèle. En revanche, on peut estimer à partir des données la dépendance entre X_t et ses valeurs passées. C'est précisément le rôle de l'ACF et de la PACF.

L'idée est la suivante. Un modèle MA, un modèle AR et un modèle ARMA ne laissent pas la même empreinte dans les autocorrélations. Un MA a une mémoire finie des chocs ; un AR propage les chocs par récurrence ; un ARMA mélange les deux. Les graphiques d'ACF et de PACF ne donnent donc pas une preuve, mais ils donnent une première orientation.

4.5.1 ACF théorique et ACF empirique

La *fonction d'autocorrélation* d'un processus stationnaire a déjà été définie par

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}, \qquad h \in \mathbb{Z}.$$

Elle est souvent notée ACF. Elle répond à la question : si je connais X_t, à quel point cela m'informe-t-il linéairement sur X_{t+h} ?

Dans un cours, on calcule souvent l'ACF *théorique*, c'est-à-dire celle du modèle. Sur une série observée, on trace une ACF *empirique*, calculée à partir des données. Ces deux objets ne doivent pas être confondus. L'ACF théorique d'un MA(1) est exactement nulle après le retard 1 ; l'ACF empirique d'une trajectoire simulée d'un MA(1) ne sera presque jamais exactement nulle, car elle est calculée sur un échantillon fini.

Idée à retenir

Il faut lire une ACF empirique comme une photographie bruitée de l'ACF théorique. On cherche une forme générale : coupure nette, décroissance lente, alternance de signes, pics saisonniers éventuels. Il ne faut pas surinterpréter une petite barre isolée.

Pour les modèles de base, les signatures théoriques sont les suivantes.

Modèle	ACF théorique	Interprétation
MA(q)	coupée après le retard q	mémoire finie des chocs
AR(p)	décroissance progressive	propagation récursive
ARMA(p,q)	décroissance progressive	mélange AR et MA

4.5.2 Pourquoi l'ACF ne suffit pas toujours

L'ACF mesure une dépendance totale au retard h. Cette dépendance peut être directe, mais elle peut aussi passer par les dates intermédiaires. Par exemple, dans un AR(1), X_t influence X_{t+1}, qui influence X_{t+2}, etc. Il est donc normal que X_t et X_{t+h} soient corrélés pour beaucoup de valeurs de h, même si le modèle ne dépend directement que du retard 1.

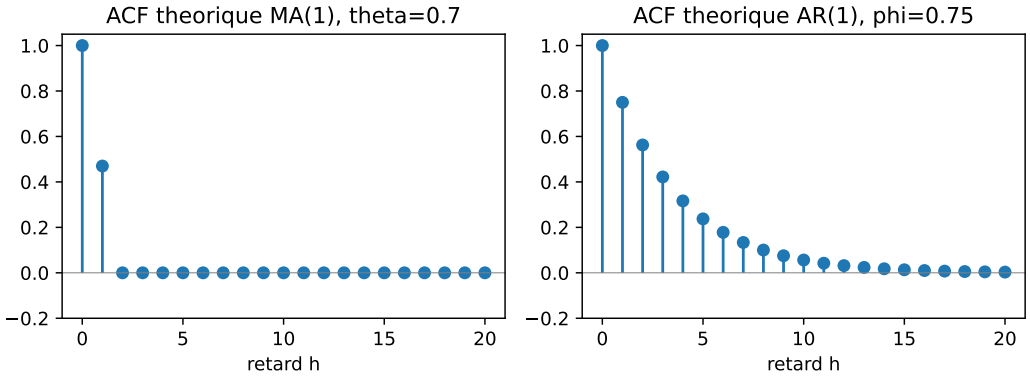


Figure 4.2 – ACF théoriques d'un MA(1) et d'un AR(1). Le MA(1) a une ACF nulle après le retard 1 ; l'AR(1) a une ACF géométrique qui décroît progressivement.

Cela explique pourquoi l'ACF d'un AR(1) ne se coupe pas après le retard 1. Une lecture naïve pourrait faire croire à une dépendance de très longue portée ; en réalité, cette dépendance vient de la propagation répétée d'une dépendance directe d'ordre 1.

On introduit alors l'autocorrélation partielle pour répondre à une question différente : après avoir tenu compte des retards intermédiaires, reste-t-il une dépendance directe entre X_t et X_{t-h} ?

4.5.3 Autocorrélation partielle : définition et intuition

Définition 4.5.1: Autocorrélation partielle

Soit X un processus stationnaire du second ordre. Pour $h \geq 1$, l'autocorrélation partielle au retard h, notée $\alpha(h)$, est la corrélation entre les deux résidus obtenus en projetant X_t et X_{t-h} sur l'espace linéaire engendré par les variables intermédiaires

$$X_{t-1}, X_{t-2}, \ldots, X_{t-h+1}.$$

Pour $h = 1$, il n'y a pas de variable intermédiaire, et l'on pose simplement

$$\alpha(1) = \rho(1).$$

La fonction $h \mapsto \alpha(h)$ est appelée PACF, pour *partial autocorrelation function*.

Cette définition peut sembler plus abstraite que celle de l'ACF, mais son interprétation est très simple. Pour regarder la dépendance au retard 4, on enlève d'abord ce qui peut déjà être expliqué par les retards 1, 2 et 3. Ce qui reste mesure la contribution propre du retard 4.

Exemple 4.5.2: Lecture intuitive dans le cas AR(1)

Dans un AR(1), on a

$$X_t = \phi X_{t-1} + Z_t.$$

Le retard 1 est directement présent dans l'équation. En revanche, X_{t-2} n'intervient dans X_t qu'à travers X_{t-1}. Une fois X_{t-1} pris en compte, le retard 2 n'apporte donc plus de dépendance directe. C'est pourquoi la PACF d'un AR(1) est coupée après le retard 1.

Théorème 4.5.3: Signature PACF d'un AR(p)

Pour un processus AR(p) causal, l'autocorrélation partielle est nulle à tous les retards strictement supérieurs à p. Autrement dit,

$$\alpha(h) = 0, \qquad h > p.$$

où α désigne la PACF du processus.

Idée de la preuve. L'équation AR(p) écrit X_t comme combinaison linéaire de $X_{t-1}, \dots, X_{t-p}$ plus l'innovation Z_t. Pour tester un retard $h > p$, les variables intermédiaires $X_{t-1}, \dots, X_{t-h+1}$ contiennent déjà $X_{t-1}, \dots, X_{t-p}$. Après projection sur ces variables intermédiaires, il ne reste donc dans X_t que l'innovation Z_t, qui est non corrélée au passé. La corrélation partielle au retard h est alors nulle. □

On obtient ainsi la règle de lecture classique :

Modèle	ACF	PACF
MA(q)	coupée après q	décroissance progressive
AR(p)	décroissance progressive	coupée après p
ARMA(p, q)	décroissance progressive	décroissance progressive

Point de vigilance

Ces règles sont des guides d'identification, pas des théorèmes de décision automatique. Sur une trajectoire finie, les ACF et PACF empiriques sont bruitées. De plus, plusieurs modèles peuvent donner des signatures proches. L'objectif est donc de proposer quelques modèles plausibles, pas de conclure définitivement à partir d'un seul graphique.

4.5.4 Méthode pratique d'identification

Une première démarche pratique est la suivante.

1) Tracer la série et vérifier qu'elle ressemble raisonnablement à une série stationnaire.
2) Tracer l'ACF empirique : cherche-t-on une coupure nette ou une décroissance progressive ?
3) Tracer la PACF empirique : observe-t-on une coupure nette après un petit nombre de retards ?
4) Proposer quelques modèles candidats, par exemple MA(1), AR(1), ARMA(1, 1), sans prétendre que le graphique suffit à choisir définitivement.

Dans ce chapitre, on s'arrête à cette lecture qualitative. Les chapitres suivants expliqueront comment estimer les paramètres et comparer plus systématiquement les modèles.

4.6 Simulation et premières commandes R

La simulation est utile pour relier les formules du chapitre aux graphiques que l'on observe réellement. Quand on simule un AR(1), un MA(1) ou un ARMA(1, 1), on connaît le modèle qui a produit la série. On peut donc vérifier si les chronogrammes, les ACF et les PACF empiriques ressemblent bien aux signatures théoriques attendues.

Dans R, la fonction de base pour simuler des modèles ARMA est `arima.sim()`.

```
x.ar <- arima.sim(model = list(ar = 0.8), n = 200)
x.ma <- arima.sim(model = list(ma = 0.7), n = 200)
x.arma <- arima.sim(model = list(ar = 0.6, ma = 0.5), n = 200)
```

La fonction `acf()` trace l'autocorrélation empirique ; `pacf()` trace l'autocorrélation partielle empirique.

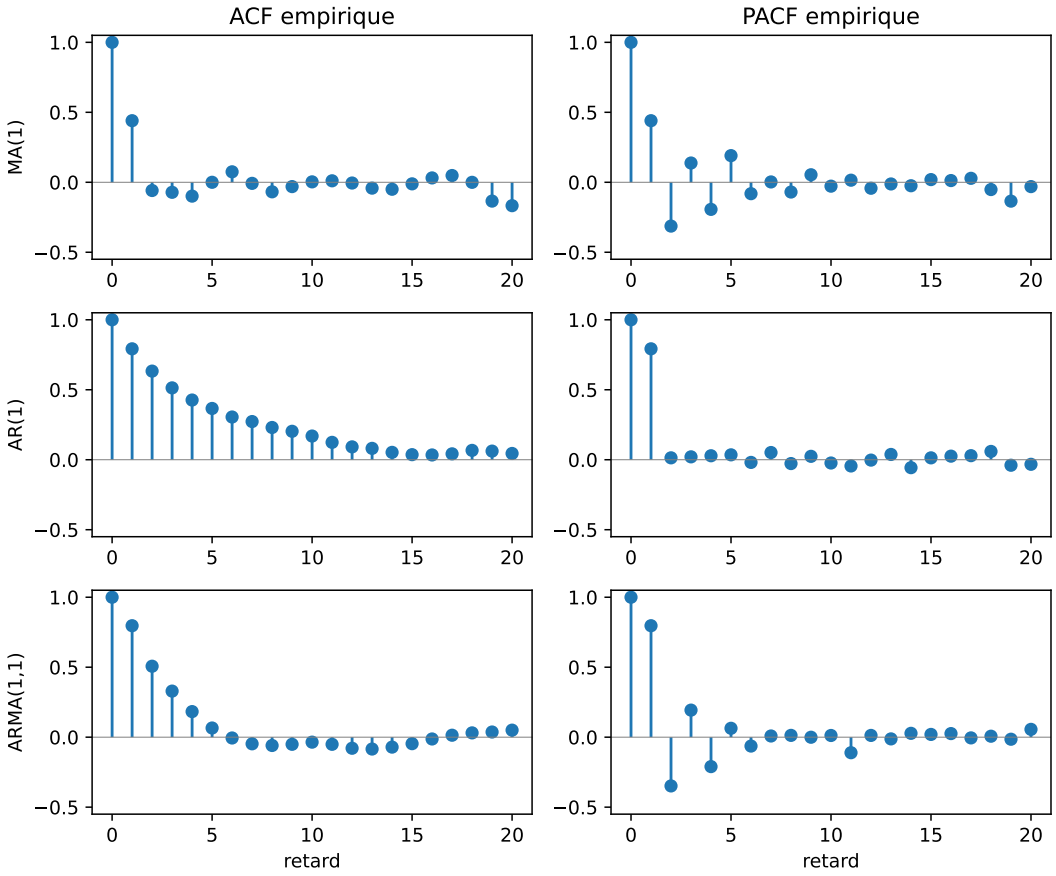


Figure 4.3 – ACF et PACF empiriques de trois séries simulées. Les règles théoriques sont visibles mais imparfaites : l'échantillon est fini, donc les barres ne sont jamais exactement nulles.

```
acf(x.ar)
pacf(x.ar)
```

Dans ce chapitre, on utilise ces graphiques pour comprendre les signatures des modèles. L'estimation systématique des paramètres et le choix de modèle par AIC/BIC seront traités plus tard.

4.7 Synthèse du chapitre

À retenir :

1) Un MA(q) est une combinaison finie de chocs récents ; il est toujours stationnaire du second ordre.
2) L'ACF d'un MA(q) est coupée après le retard q.
3) L'inversibilité d'un MA signifie que l'on peut reconstruire les innovations à partir du présent et du passé observés.
4) Un AR(p) relie X_t à ses propres valeurs passées et à une innovation présente.
5) Un AR(1) stationnaire causal existe lorsque $|\phi| < 1$, et son ACF vaut $\rho(h) = \phi^{|h|}$.
6) Pour un AR(p), la condition standard de causalité est : toutes les racines de $\Phi(z)$ sont de module strictement supérieur à 1.
7) Les fractions rationnelles $1/\Phi(z)$ et $\Theta(z)/\Phi(z)$ décrivent les filtres causaux associés aux AR et ARMA.
8) Un ARMA(p, q) combine un polynôme autorégressif Φ et un polynôme moyenne mobile Θ.

9) Les racines de Φ contrôlent la causalité ; les racines de Θ contrôlent l'inversibilité.
10) L'ACF est la fonction d'autocorrélation ; la PACF est la fonction d'autocorrélation partielle, qui retire l'effet des retards intermédiaires.
11) ACF et PACF servent à l'identification, mais seulement comme guides empiriques.

Le chapitre suivant traitera des séries non stationnaires. On y verra comment utiliser les différences pour transformer certaines séries en objets auxquels les modèles ARMA peuvent être appliqués.

Chapitre 5

Séries non stationnaires et modèles ARIMA/SARIMA

Mots-clés. Non-stationnarité ; tendance déterministe ; tendance stochastique ; racine unitaire ; marche aléatoire ; différenciation ; série intégrée ; ARIMA ; SARIMA ; différence saisonnière ; sur-différenciation ; identification.

5.1 Pourquoi quitter les modèles ARMA stationnaires ?

Le chapitre précédent a donné une première famille de modèles pour les séries stationnaires : MA, AR et ARMA. Ces modèles sont puissants, mais ils reposent sur une hypothèse structurante : la série que l'on modélise doit fluctuer autour d'un comportement stable. Sa moyenne ne doit pas dériver indéfiniment, sa variance ne doit pas changer brutalement avec le temps, et sa dépendance temporelle doit être comparable d'une période à l'autre.

Or beaucoup de séries observées ne satisfont pas cette description. Une série de ventes peut croître pendant plusieurs années ; une série trimestrielle de consommation d'énergie peut présenter un motif saisonnier très marqué ; un indice financier peut changer durablement de niveau après une crise. Dans ces cas, essayer d'ajuster directement un ARMA à la série brute revient souvent à demander à un modèle stationnaire de décrire un phénomène qui ne l'est manifestement pas.

L'idée de ce chapitre est de ne pas abandonner les modèles ARMA, mais de les utiliser au bon endroit. On ne modélise pas toujours le niveau brut de la série. On modélise parfois ses variations. Par exemple, si la série brute (X_t) n'est pas stationnaire mais si

$$\Delta X_t = X_t - X_{t-1}$$

est stationnaire, alors on peut ajuster un modèle ARMA à ΔX. La série X sera alors décrite par un modèle ARIMA. Si la non-stationnarité est saisonnière, on utilise la différence saisonnière

$$(1 - B^s)X_t = X_t - X_{t-s}.$$

Idée à retenir

Un modèle ARMA décrit une série déjà stationnaire. Un modèle ARIMA décrit une série qui devient stationnaire après différenciation. Un modèle SARIMA ajoute une différenciation saisonnière, par exemple d'une année à l'autre pour une série mensuelle ou d'un trimestre à l'autre pour une série trimestrielle.

Ce changement de point de vue est très concret. Dans une série économique, le niveau peut croître fortement, alors que les taux de croissance sont plus stables. Dans une série mensuelle, les mois de juillet peuvent être systématiquement plus hauts que les mois de février ; comparer chaque mois au même mois de l'année précédente peut alors supprimer une grande partie de la saisonnalité.

Exemple 5.1.1: Concentration atmosphérique de CO_2

La série `co2`, fournie avec `R` dans le package `datasets`, donne des mesures mensuelles de concentration atmosphérique de CO_2 à Mauna Loa. Elle montre deux phénomènes bien visibles : une tendance croissante de long terme et une oscillation saisonnière annuelle. Sur la série brute, un modèle ARMA stationnaire n'est pas naturel. En revanche, une différence saisonnière de période 12 compare chaque mois au même mois de l'année précédente et retire une grande partie du cycle annuel.

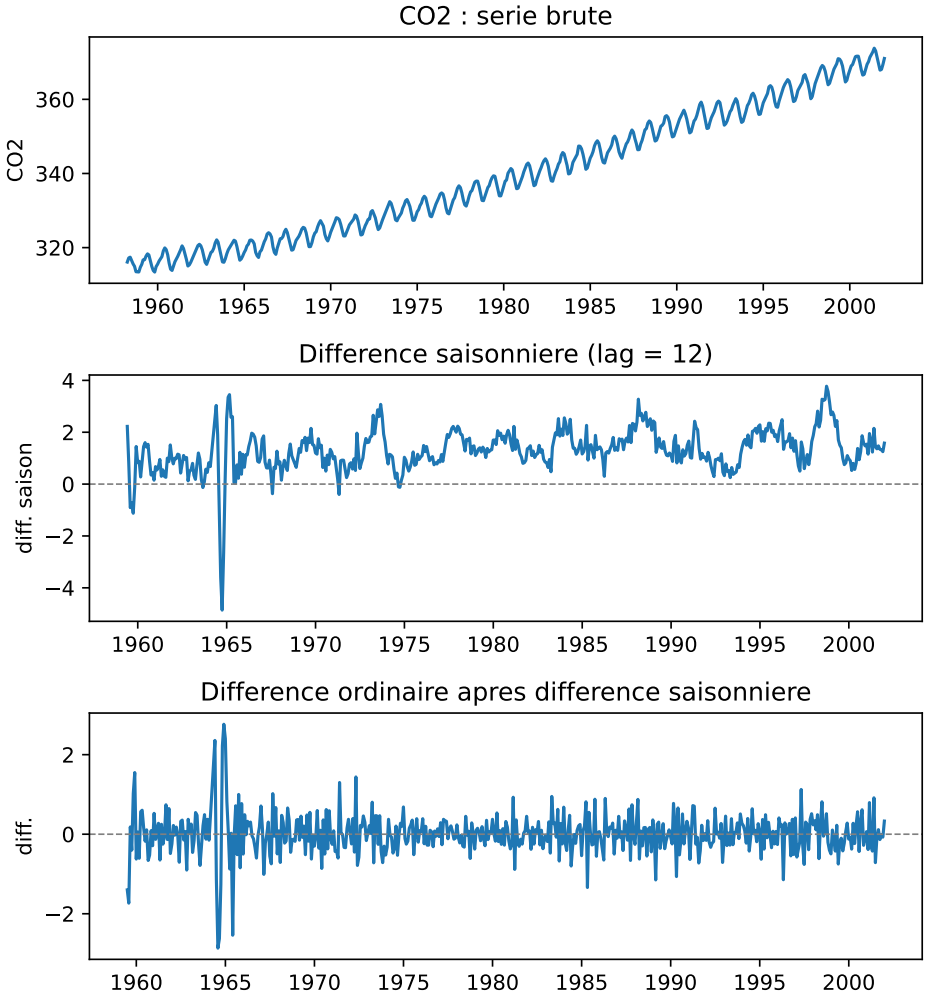


Figure 5.1 – Série de concentration de CO_2 à Mauna Loa, analogue à la série `co2` fournie avec `R`. La série brute combine tendance croissante et saisonnalité annuelle ; la différence saisonnière réduit fortement le motif annuel ; une différence ordinaire supplémentaire aide à stabiliser la tendance restante.

5.2 Deux formes de non-stationnarité

Avant de différencier une série, il faut comprendre quel type de non-stationnarité on essaie de traiter. Deux mécanismes peuvent produire des chronogrammes qui montent ou descendent dans le temps, mais ils n'ont pas du tout la même interprétation.

5.2.1 Tendance déterministe

Une première possibilité est une *tendance déterministe*. On écrit par exemple

$$X_t = a + bt + Z_t,$$

où Z est stationnaire. Ici, la partie non stationnaire est la droite $t \mapsto a + bt$. Le bruit Z_t fluctue autour de cette droite. Si l'on connaissait a et b, on pourrait retirer exactement la tendance et retrouver Z_t. En pratique, on estime a et b par régression, puis on étudie les résidus.

Ce modèle correspond à une situation où le système possède une trajectoire moyenne bien déterminée. Par exemple, une production industrielle pourrait augmenter approximativement linéairement sur une période courte, autour d'une tendance de croissance stable. Dans ce cas, la question naturelle est : quelle est la tendance moyenne, et les fluctuations autour de cette tendance sont-elles stationnaires ?

5.2.2 Tendance stochastique et racine unitaire

Une deuxième possibilité est une *tendance stochastique*. L'exemple fondamental est la marche aléatoire

$$X_t = X_{t-1} + Z_t.$$

Dans ce modèle, les chocs s'accumulent. Un choc positif aujourd'hui augmente le niveau de la série aujourd'hui, mais aussi demain, après-demain, etc. Il n'y a pas de droite déterministe autour de laquelle la série revient. Le niveau courant contient toute l'histoire des innovations passées.

On peut écrire la marche aléatoire sous la forme

$$(1 - B)X_t = Z_t.$$

Le polynôme autorégressif est donc $1 - B$. Il possède une racine en 1, appelée *racine unitaire*. Cette racine est le signe algébrique d'une non-stationnarité : le filtre inverse $(1 - B)^{-1}$ n'est pas un filtre causal stable au sens du chapitre 3.

Point de vigilance

Une tendance déterministe et une tendance stochastique peuvent se ressembler graphiquement sur un intervalle fini. Pourtant, elles ne prédisent pas le même futur. Avec une tendance déterministe, la série revient autour d'une courbe moyenne. Avec une marche aléatoire, les chocs déplacent durablement le niveau.

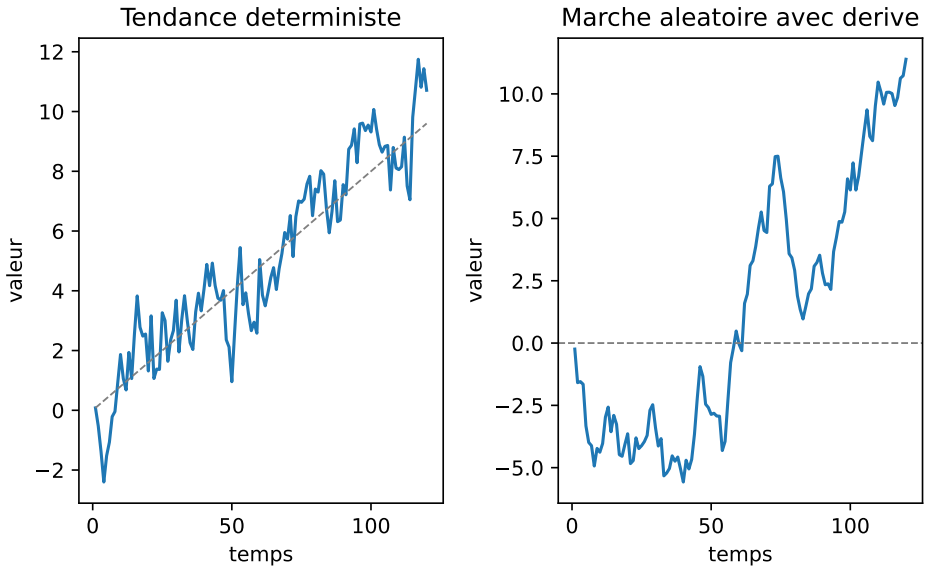


Figure 5.2 – Deux séries qui peuvent sembler croître, mais pour des raisons différentes. À gauche, une tendance déterministe linéaire plus bruit stationnaire. À droite, une marche aléatoire avec dérive : les chocs s'accumulent et modifient durablement le niveau.

5.2.3 La marche aléatoire

Définition 5.2.1: Marche aléatoire

Soit $(Z_t)_{t\geq 1}$ un bruit blanc centré de variance σ^2. La marche aléatoire issue de 0 est le processus défini par

$$X_0 = 0, \qquad X_t = X_{t-1} + Z_t \quad (t \geq 1).$$

On a donc

$$X_t = Z_1 + \cdots + Z_t.$$

Proposition 5.2.2: La marche aléatoire n'est pas stationnaire

La marche aléatoire $(X_t)_{t\geq 0}$ n'est pas stationnaire du second ordre. En revanche, sa différence première est le bruit blanc Z :

$$(\Delta X)_t = X_t - X_{t-1} = Z_t.$$

Démonstration. On a $\mathbb{E}(X_t) = 0$. En revanche,

$$\mathrm{Var}(X_t) = \mathrm{Var}(Z_1 + \cdots + Z_t) = t\sigma^2,$$

car les innovations sont non corrélées. La variance dépend de t, ce qui contredit la stationnarité du second ordre. Enfin, l'équation de définition donne directement $X_t - X_{t-1} = Z_t$. □

Cette proposition donne le prototype des modèles ARIMA : la série brute n'est pas stationnaire, mais une différence l'est. On ne force donc pas un ARMA sur X ; on modélise ΔX.

5.3 Différenciation et séries intégrées

5.3.1 Différence première : passer des niveaux aux variations

La différence première a déjà été introduite au chapitre 1 :

$$(\Delta X)_t = X_t - X_{t-1} = (1 - B)X_t.$$

Elle remplace un niveau par une variation. C'est un changement d'objet : on ne demande plus « quel est le niveau de la série ? », mais « de combien le niveau a-t-il changé depuis la période précédente ? ».

Pour une série économique positive, on applique souvent cette opération au logarithme :

$$\Delta \log(X_t) = \log(X_t) - \log(X_{t-1}).$$

Quand les variations relatives sont petites, cette quantité est proche du taux de croissance relatif entre $t-1$ et t. Ainsi, $\Delta \log(X_t)$ décrit une croissance relative plutôt qu'une variation absolue.

Exemple 5.3.1: Indice économique ou chiffre d'affaires

Si un chiffre d'affaires passe de 100 à 105, la variation absolue vaut 5. S'il passe de 1000 à 1050, la variation absolue vaut 50. Pourtant, dans les deux cas, la croissance relative est de 5%. Sur l'échelle logarithmique, ces deux variations sont presque de même taille. C'est pour cette raison que l'on modélise souvent les différences de logarithmes plutôt que les différences brutes.

Le point mathématique à retenir est que la différenciation supprime certaines tendances, mais transforme aussi la structure du bruit. Elle n'est donc pas une opération neutre.

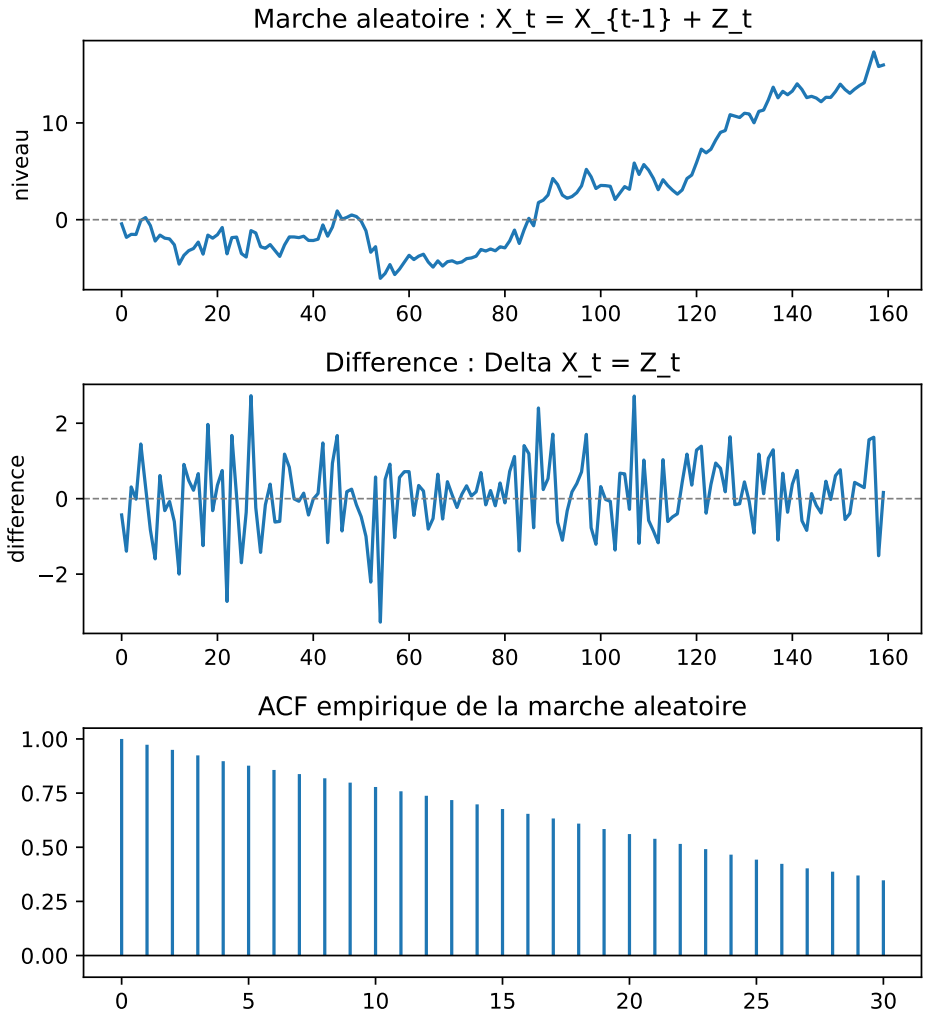


Figure 5.3 – Une marche aléatoire simulée, sa différence première, et l'ACF empirique de la marche aléatoire. La série brute a une mémoire apparente très longue ; après différenciation, on retrouve les innovations.

Proposition 5.3.2: Effet de la différence sur une tendance déterministe

Soit

$$X_t = a + bt + Z_t,$$

où Z est stationnaire. Alors

$$(\Delta X)_t = b + Z_t - Z_{t-1}.$$

En particulier, ΔX est stationnaire si Z est stationnaire.

Démonstration. Il suffit de calculer :

$$\begin{aligned}(\Delta X)_t &= X_t - X_{t-1} \\ &= (a + bt + Z_t) - (a + b(t-1) + Z_{t-1}) \\ &= b + Z_t - Z_{t-1}.\end{aligned}$$

Si Z est stationnaire, alors $Z_t - Z_{t-1}$ est obtenu par filtrage linéaire fini d'un processus stationnaire. Il est donc stationnaire. □

Point de vigilance

Différencier une série avec tendance déterministe fonctionne parfois, mais ce n'est pas toujours la meilleure modélisation. Si le vrai modèle est $X_t = a + bt + Z_t$, une régression sur le temps peut être plus interprétable qu'une différenciation. La différenciation est surtout naturelle quand les chocs semblent avoir un effet permanent, comme dans une marche aléatoire.

5.3.2 Séries intégrées

Définition 5.3.3: Série intégrée d'ordre d

Soit $d \in \mathbb{N}$. On dit qu'une série (X_t) est intégrée d'ordre d, et l'on écrit parfois $X \sim I(d)$, si $\Delta^d X$ est stationnaire, tandis que $\Delta^{d-1} X$ ne l'est pas lorsque $d \geq 1$.

Cette définition formalise l'idée suivante : il faut différencier d fois pour obtenir un objet stationnaire. Le cas $d = 0$ correspond simplement à une série déjà stationnaire.

Exemple 5.3.4: Quelques cas simples

1) Un bruit blanc est $I(0)$.
2) Une marche aléatoire est $I(1)$, car sa différence est un bruit blanc.
3) Si $X_t = a+bt+Z_t$, avec Z stationnaire, alors $\Delta X_t = b+Z_t-Z_{t-1}$ est stationnaire. La série devient stationnaire après une différence, mais la non-stationnarité est ici déterministe, ce qui demande une interprétation différente.

5.3.3 Sur-différencier une série

On pourrait être tenté de différencier jusqu'à ce que le graphique ait l'air très désorganisé. Ce n'est pas une bonne règle. Si l'on différencie une série déjà stationnaire, on peut introduire artificiellement une dépendance négative.

Proposition 5.3.5: Différence d'un bruit blanc

Si (Z_t) est un bruit blanc centré de variance σ^2, et si

$$Y_t = \Delta Z_t = Z_t - Z_{t-1},$$

alors Y est un MA(1) et

$$\rho_Y(1) = -\frac{1}{2}, \qquad \rho_Y(h) = 0 \quad \text{pour } |h| \geq 2.$$

Démonstration. On reconnaît

$$Y_t = Z_t - Z_{t-1},$$

c'est-à-dire un MA(1) avec $\theta = -1$. On peut aussi calculer directement :

$$\gamma_Y(0) = \mathrm{Var}(Z_t - Z_{t-1}) = 2\sigma^2,$$

et

$$\gamma_Y(1) = \mathrm{Cov}(Z_t - Z_{t-1}, Z_{t+1} - Z_t) = -\sigma^2.$$

Donc $\rho_Y(1) = -1/2$. Pour $|h| \geq 2$, les deux différences ne partagent aucune innovation commune, donc la covariance est nulle. □

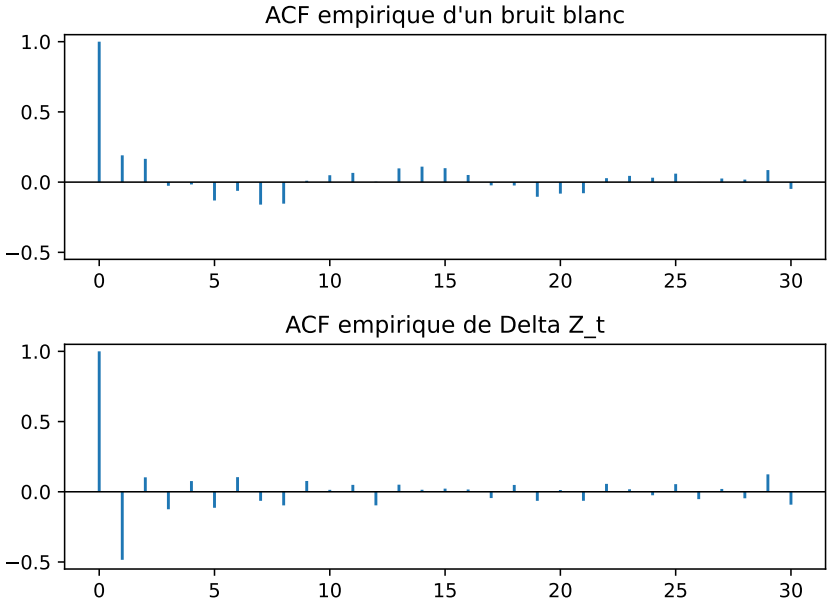


Figure 5.4 – Différencier un bruit blanc crée une corrélation négative au retard 1. La différenciation n'est pas une opération à appliquer automatiquement sans diagnostic.

5.4 Modèles ARIMA

5.4.1 Pourquoi introduire les ARIMA ?

Les modèles ARMA du chapitre 4 décrivent une dynamique stationnaire : les chocs ont un effet qui s'amortit, et la série reste autour d'un niveau stable. Les modèles ARIMA sont introduits pour traiter le cas où cette dynamique stationnaire n'apparaît qu'après différenciation.

La logique est donc en deux étapes :

1) transformer la série brute X en une série $Y = \Delta^d X$ qui semble stationnaire ;
2) modéliser Y par un ARMA(p, q).

Cette construction permet de garder toute la théorie ARMA pour la partie stationnaire, tout en autorisant le niveau brut à être non stationnaire. Elle est très utile lorsque l'on observe des niveaux qui dérivent, mais des variations qui ont un comportement plus stable.

5.4.2 Définition

Définition 5.4.1: Modèle ARIMA(p, d, q)

On dit que (X_t) suit un modèle ARIMA(p, d, q) si la série différenciée

$$Y_t = \Delta^d X_t$$

est un processus ARMA(p, q). Autrement dit,

$$\Phi(B)\Delta^d X_t = \Theta(B)Z_t,$$

où

$$\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p, \qquad \Theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q.$$

Le mot *integrated* dans ARIMA correspond au fait que la série brute X est obtenue en réintégrant, c'est-à-dire en cumulant, une série stationnaire. Par exemple, si $\Delta X_t = Z_t$, alors

$$X_t = X_0 + Z_1 + \cdots + Z_t.$$

La série brute accumule les innovations.

Exemple 5.4.2: Trois modèles ARIMA élémentaires

1) ARIMA$(0,1,0)$: $\Delta X_t = Z_t$. C'est la marche aléatoire.
2) ARIMA$(0,1,0)$ avec dérive : $\Delta X_t = c + Z_t$. Les accroissements ont une moyenne non nulle.
3) ARIMA$(1,1,0)$: $(1-\phi B)\Delta X_t = Z_t$. La série des variations suit un AR(1).

Proposition 5.4.3: Prévision d'une marche aléatoire avec dérive

Soit

$$X_t = X_{t-1} + c + Z_t,$$

où (Z_t) est un bruit blanc centré. Si $\mathcal{F}_n$ représente l'information disponible jusqu'au temps n, alors, pour $h \geq 1$,

$$\mathbb{E}(X_{n+h} \mid \mathcal{F}_n) = X_n + hc.$$

Si les innovations sont non corrélées de variance σ^2, alors l'erreur de prévision à horizon h a variance $h\sigma^2$.

Démonstration. En itérant l'équation,

$$X_{n+h} = X_n + hc + Z_{n+1} + \cdots + Z_{n+h}.$$

Les innovations futures sont centrées et non corrélées à l'information $\mathcal{F}_n$, d'où la formule de moyenne conditionnelle. La variance de la somme des h innovations futures vaut $h\sigma^2$. □

Idée à retenir

Pour une marche aléatoire, la meilleure prévision sans dérive est le dernier niveau observé. Mais l'incertitude augmente avec l'horizon : plus on prédit loin, plus on additionne d'innovations inconnues.

5.5 Différenciation saisonnière et modèles SARIMA

5.5.1 Pourquoi une différence saisonnière ?

Une série saisonnière ne dépend pas seulement de la période précédente : elle dépend souvent fortement de la même saison précédente. Dans une série mensuelle de trafic, décembre se compare plus naturellement au décembre précédent qu'au mois de novembre. Dans une série trimestrielle de consommation de gaz, le premier trimestre d'une année se compare au premier trimestre de l'année précédente.

La différence saisonnière répond à cette idée. Pour une série de période saisonnière s, on pose

$$(\Delta_s X)_t = (1 - B^s)X_t = X_t - X_{t-s}.$$

Pour une série mensuelle, $s = 12$; pour une série trimestrielle, $s = 4$.

Exemple 5.5.1: Trafic aérien mensuel

La série `AirPassengers`, fournie avec `R`, donne un exemple classique de série mensuelle avec croissance et saisonnalité. Le logarithme rend les oscillations plus comparables sur toute la période. La différence saisonnière du logarithme compare chaque mois au même mois de l'année précédente. Il reste ensuite une légère tendance, que l'on réduit par une différence ordinaire supplémentaire.

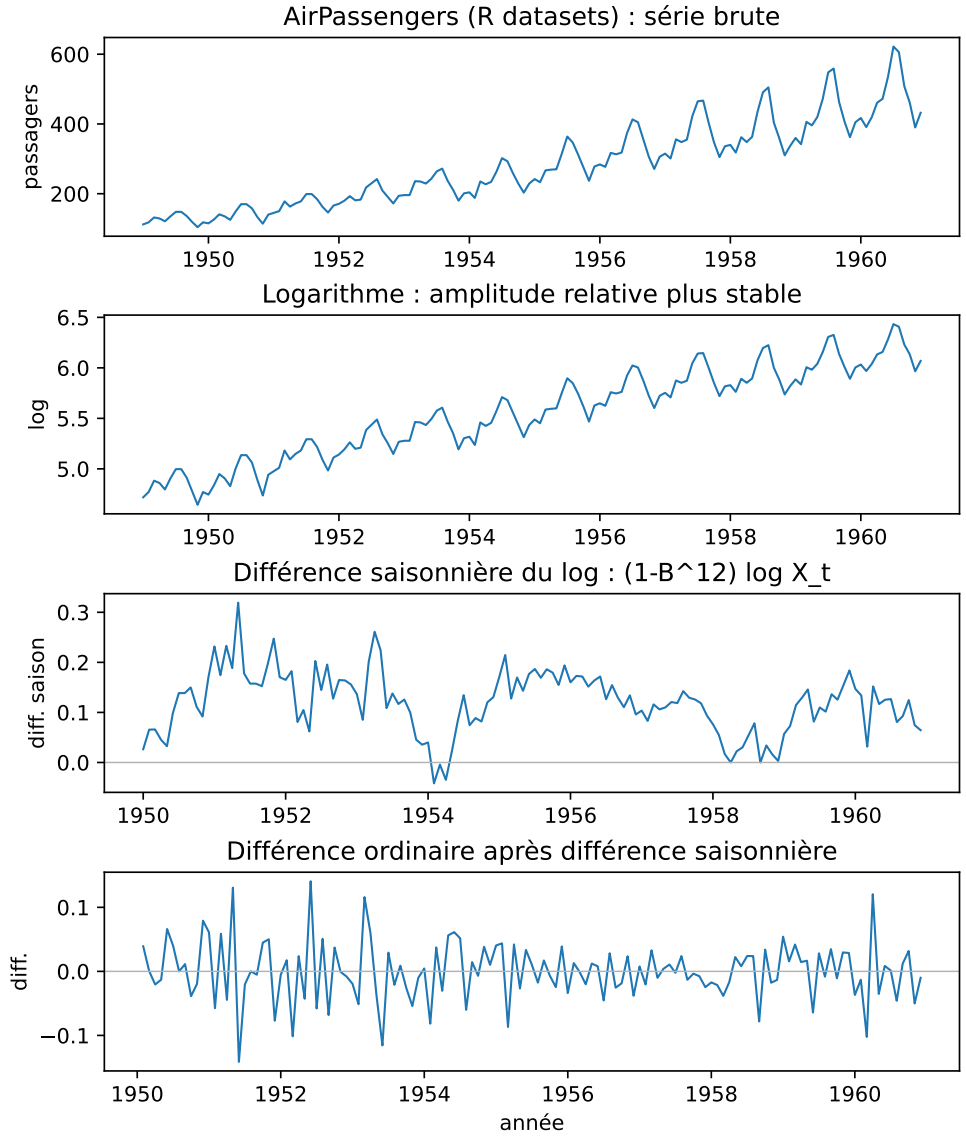


Figure 5.5 – Transformation de la série `AirPassengers`, fournie avec `R` dans le package `datasets`. Le logarithme stabilise l'amplitude relative ; la différence saisonnière retire une grande partie du motif annuel ; la différence ordinaire supplémentaire réduit la tendance restante.

5.5.2 Effet sur une saisonnalité déterministe

Proposition 5.5.2: Effet sur une saisonnalité déterministe

Supposons que

$$X_t = m_t + s_t + Z_t,$$

où $s_{t+s} = s_t$ pour tout t. Alors

$$(1 - B^s)X_t = (m_t - m_{t-s}) + (Z_t - Z_{t-s}).$$

La composante saisonnière déterministe est supprimée.

Démonstration. On calcule

$$\begin{aligned} X_t - X_{t-s} &= (m_t + s_t + Z_t) - (m_{t-s} + s_{t-s} + Z_{t-s}) \\ &= (m_t - m_{t-s}) + (s_t - s_{t-s}) + (Z_t - Z_{t-s}). \end{aligned}$$

Comme $s_{t-s} = s_t$, le terme saisonnier disparaît. □

Si $m_t = a + bt$, alors $m_t - m_{t-s} = bs$. La différence saisonnière transforme donc une tendance linéaire en constante, mais elle ne supprime pas nécessairement toute tendance plus complexe. C'est pourquoi une différence ordinaire peut encore être utile après une différence saisonnière.

5.5.3 Définition d'un SARIMA

Un modèle SARIMA est un modèle ARIMA avec une partie saisonnière. Il contient des polynômes ordinaires en B, mais aussi des polynômes saisonniers en B^s. Cette notation semble lourde, mais chaque facteur a un rôle précis : certaines transformations retirent la non-stationnarité, d'autres modélisent la dépendance restante.

Définition 5.5.3: Modèle SARIMA

On dit que (X_t) suit un modèle

$$\text{SARIMA}(p, d, q) \times (P, D, Q)_s$$

si

$$\Phi(B)\Phi_s(B^s)(1-B)^d(1-B^s)^D X_t = \Theta(B)\Theta_s(B^s)Z_t,$$

où s est la période saisonnière, d est le nombre de différences ordinaires, et D le nombre de différences saisonnières.

La lecture de la notation se fait de gauche à droite :

1) $(1-B)^d$ retire une tendance non saisonnière ;
2) $(1-B^s)^D$ retire une non-stationnarité saisonnière ;
3) $\Phi(B), \Theta(B)$ modélisent la dépendance ordinaire restante ;
4) $\Phi_s(B^s), \Theta_s(B^s)$ modélisent la dépendance saisonnière restante.

Exemple 5.5.4: Modèle $\text{SARIMA}(0,1,1) \times (0,1,1)_{12}$

Le modèle

$$(1-B)(1-B^{12})X_t = (1+\theta B)(1+\Theta B^{12})Z_t$$

est un modèle saisonnier classique pour des séries mensuelles de type trafic, ventes ou passagers. Le membre de gauche signifie que l'on applique une différence saisonnière, puis une différence ordinaire. Le membre de droite autorise une dépendance MA au retard 1 et au retard 12.

5.6 Comment choisir les différences ?

Le choix de d et D est une étape délicate. Il ne faut ni trop peu différencier, ni trop différencier. Trop peu différencier laisse une série non stationnaire ; trop différencier crée une dépendance artificielle et peut rendre le modèle inutilement compliqué.

Une démarche raisonnable est la suivante.

1) Tracer la série brute : cherche-t-on une tendance, une saisonnalité, une variance croissante ?
2) Si la série est positive avec amplitude croissante, regarder le logarithme.
3) Si une saisonnalité forte apparaît, essayer une différence saisonnière.
4) Si une tendance reste après la différence saisonnière, essayer ensuite une différence ordinaire.
5) Contrôler les ACF/PACF empiriques de la série transformée : une ACF qui décroît très lentement suggère souvent qu'il reste une non-stationnarité.

Point de vigilance

Les fonctions automatiques peuvent proposer des valeurs de d et D, mais elles ne remplacent pas le diagnostic. Dans un cours d'introduction, on veut surtout comprendre ce que font les transformations : passer du niveau aux variations, comparer une saison à la même saison précédente, puis modéliser une série raisonnablement stationnaire.

5.7 Premières commandes R

Les fonctions essentielles sont `diff()`, `log()`, `acf()` et, plus tard, `arima()`.

```
data(AirPassengers)
air <- AirPassengers
log.air <- log(air)
sair <- diff(log.air, lag = 12)
dsair <- diff(sair)

plot.ts(log.air)
plot.ts(sair)
plot.ts(dsair)
acf(dsair)
```

Dans ce code, `diff(log.air, lag = 12)` calcule

$$\log(X_t) - \log(X_{t-12}),$$

et `diff(sair)` applique ensuite une différence ordinaire à la série déjà différenciée saisonnièrement.

5.8 Synthèse du chapitre

À retenir :

1) Les modèles ARMA s'appliquent à des séries stationnaires ; les modèles ARIMA permettent de traiter certaines séries non stationnaires par différenciation.
2) Une marche aléatoire n'est pas stationnaire, car sa variance croît avec le temps, mais sa différence est un bruit blanc.
3) Une tendance déterministe et une tendance stochastique peuvent se ressembler graphiquement, mais elles n'ont pas la même interprétation.
4) Un modèle ARIMA(p, d, q) signifie que $\Delta^d X_t$ suit un ARMA(p, q).
5) Une différence saisonnière $(1 - B^s)X_t = X_t - X_{t-s}$ compare une observation à celle de la même saison précédente.
6) Un modèle SARIMA combine différences ordinaires, différences saisonnières, dépendance ARMA ordinaire et dépendance ARMA saisonnière.
7) Trop différencier peut créer une dépendance artificielle, par exemple une autocorrélation négative au retard 1.

Le chapitre suivant sera consacré à l'estimation, au diagnostic et à la prévision. On y prendra une série transformée, on ajustera plusieurs modèles, puis on vérifiera si les résidus ressemblent à un bruit blanc.

Chapitre 6

Estimation, diagnostic et prévision

Mots-clés. Estimation ; vraisemblance ; moindres carrés ; AIC ; BIC ; résidus d'innovation ; diagnostic ; ACF empirique des résidus ; test de Ljung–Box ; prévision ponctuelle ; intervalle de prévision ; validation hors-échantillon.

6.1 Pourquoi un nouveau chapitre ?

Les chapitres précédents ont répondu à une première question : *quel type de modèle peut décrire la dynamique d'une série ?* Nous avons construit des modèles ARMA pour des séries stationnaires, puis des modèles ARIMA et SARIMA pour des séries que l'on rend stationnaires par différence. Il reste maintenant trois questions indispensables en pratique.

1) Une fois la forme du modèle choisie, comment estimer les coefficients inconnus ?
2) Comment vérifier que le modèle ajusté a réellement capturé la dépendance temporelle importante ?
3) Comment utiliser le modèle pour prévoir les observations futures, et comment quantifier l'incertitude de cette prévision ?

Ce chapitre est donc le passage du modèle écrit au tableau au modèle effectivement utilisé sur une série. Il faut garder à l'esprit une distinction centrale : le modèle théorique contient un bruit blanc inconnu (Z_t), alors que sur les données on ne dispose que de *résidus estimés*. Le diagnostic consiste précisément à vérifier si ces résidus estimés ressemblent raisonnablement à un bruit blanc.

Idée à retenir

Dans un modèle correctement ajusté, les résidus doivent être la partie que le modèle n'arrive pas à expliquer. Si les résidus contiennent encore une tendance, une saisonnalité ou une autocorrélation visible, alors le modèle n'a pas fini son travail.

6.2 Estimer les paramètres d'un modèle

6.2.1 Du modèle théorique au modèle ajusté

Considérons par exemple un ARMA$(1,1)$ stationnaire,

$$X_t = \phi X_{t-1} + Z_t + \theta Z_{t-1}.$$

La forme du modèle dit comment le présent dépend du passé, mais elle ne dit pas quelles sont les valeurs de ϕ, θ et de la variance σ^2 du bruit. Sur une série observée $x_1, \dots, x_n$, on doit donc remplacer les paramètres inconnus par des estimations.

Le vocabulaire est le suivant :

— le *modèle* est la famille de lois indexée par les paramètres ;

— l'*estimation* choisit une valeur des paramètres à partir des données ;

— le *modèle ajusté* est le modèle avec les paramètres estimés.

Cette distinction est importante pour les diagnostics. Les résidus ne sont pas les vrais Z_t. Ce sont des objets calculés à partir du modèle ajusté. Ils dépendent donc du modèle choisi et des paramètres estimés.

Définition 6.2.1: Modèle ajusté et résidus d'innovation

Soit un modèle temporel ajusté à des observations $x_1, \dots, x_n$. On appelle *valeur prédite à un pas* la quantité $\widehat{x}_{t|t-1}$ prédite pour x_t à partir du passé disponible $x_1, \dots, x_{t-1}$. Le *résidu d'innovation estimé* au temps t est

$$\widehat{z}_t = x_t - \widehat{x}_{t|t-1}.$$

L'expression $\widehat{x}_{t|t-1}$ signifie : prévision de x_t sachant l'information disponible jusqu'au temps $t-1$. L'indice rappelle donc que l'on ne triche pas : pour prédire x_t, on n'utilise pas encore x_t.

6.2.2 Un cas complètement explicite : AR(1) par moindres carrés

Avant de parler de vraisemblance, il est utile de traiter un cas simple où les calculs sont transparents. Supposons que

$$X_t = \phi X_{t-1} + Z_t,$$

avec (Z_t) centré. Si l'on observe $x_0, x_1, \dots, x_n$, une idée naturelle est de choisir ϕ pour rendre les écarts

$$x_t - \phi x_{t-1}$$

les plus petits possible.

Proposition 6.2.2: Estimateur des moindres carrés pour un AR(1)

Si $\sum_{t=1}^{n} x_{t-1}^2 > 0$, le minimum de

$$S(\phi) = \sum_{t=1}^{n} (x_t - \phi x_{t-1})^2$$

est atteint en

$$\widehat{\phi} = \frac{\sum_{t=1}^{n} x_t x_{t-1}}{\sum_{t=1}^{n} x_{t-1}^2}.$$

Démonstration. La fonction S est un polynôme quadratique en ϕ. On dérive :

$$\begin{aligned} S'(\phi) &= -2 \sum_{t=1}^{n} x_{t-1} (x_t - \phi x_{t-1}) \\ &= -2 \sum_{t=1}^{n} x_t x_{t-1} + 2\phi \sum_{t=1}^{n} x_{t-1}^2. \end{aligned}$$

L'unique point critique vérifie donc

$$\phi \sum_{t=1}^{n} x_{t-1}^2 = \sum_{t=1}^{n} x_t x_{t-1}.$$

Comme $S''(\phi) = 2 \sum_{t=1}^{n} x_{t-1}^2 > 0$, ce point critique est bien le minimum. □

Ce résultat ressemble à une régression linéaire sans constante : on explique x_t par x_{t-1}. Il donne une bonne intuition de l'estimation en séries temporelles : on choisit les paramètres pour que les innovations estimées soient petites et sans structure restante.

6.2.3 Vraisemblance gaussienne

Pour les modèles ARMA et ARIMA généraux, les calculs explicites deviennent vite plus lourds. Les logiciels utilisent alors des méthodes numériques, souvent fondées sur la vraisemblance. L'idée est simple : pour chaque valeur possible du paramètre, le modèle attribue une probabilité aux données observées. On choisit la valeur qui rend les données les plus plausibles.

Définition 6.2.3: Vraisemblance et log-vraisemblance

Soit θ un vecteur de paramètres et supposons que le modèle attribue aux observations $x_1, \dots, x_n$ une densité $L(\theta)$. La fonction L est appelée *vraisemblance*. La *log-vraisemblance* est

$$\ell(\theta) = \log L(\theta).$$

Un estimateur du maximum de vraisemblance est une valeur $\widehat{\theta}$ telle que

$$\ell\left(\widehat{\theta}\right) = \max_{\theta} \ell(\theta).$$

Dans un modèle gaussien, les innovations sont supposées normales. Maximiser la vraisemblance revient alors, de manière schématique, à chercher des paramètres qui rendent les innovations estimées petites et compatibles avec une loi normale centrée de variance constante. Cette description n'est pas une preuve générale, mais elle donne la bonne intuition.

Point de vigilance

La fonction `arima()` de `R` estime les paramètres numériquement. Les valeurs affichées ne sont donc pas obtenues par une formule simple en général. Il faut apprendre à lire la sortie : coefficients estimés, erreurs standards, log-vraisemblance, AIC et résidus.

6.3 Comparer plusieurs modèles : AIC et BIC

6.3.1 Pourquoi un critère de comparaison ?

Si l'on augmente le nombre de paramètres, on améliore presque toujours l'ajustement aux données déjà observées. Mais un modèle trop riche peut apprendre des fluctuations accidentelles au lieu d'apprendre la structure de la série. C'est le phénomène de sur-ajustement.

Les critères AIC et BIC pénalisent donc la complexité. Ils partent de la log-vraisemblance maximisée, puis ajoutent une pénalité qui augmente avec le nombre de paramètres.

Définition 6.3.1: Critères AIC et BIC

Pour un modèle ajusté par maximum de vraisemblance, on note $\widehat{\ell}$ la log-vraisemblance maximisée, k le nombre de paramètres estimés et n le nombre d'observations. On définit

$$\mathrm{AIC} = -2\widehat{\ell} + 2k, \qquad \mathrm{BIC} = -2\widehat{\ell} + k \log n.$$

À données identiques, on préfère le modèle dont le critère est le plus petit.

La différence entre les deux critères est la pénalité. Le BIC pénalise davantage les paramètres quand n est grand, car $\log n > 2$ dès que $n \geq 8$. Il a donc tendance à choisir des modèles plus simples que l'AIC.

Remarque 6.3.2: Ce qu'un critère ne dit pas

Un critère numérique ne remplace pas le diagnostic. Un modèle peut avoir un bon AIC et des résidus encore autocorrélés. Dans ce cas, le critère indique seulement que le modèle est meilleur que les modèles comparés, pas qu'il est satisfaisant en soi.

6.4 Diagnostic : que doivent vérifier les résidus ?

6.4.1 Le rôle du diagnostic

Après estimation, on ne doit pas conclure immédiatement que le modèle est bon. Il faut regarder ce qu'il reste. L'idée est la même que dans les premiers chapitres : un résidu est un reste après ajustement. Mais ici, le reste a une interprétation plus précise. Dans un ARMA ou un ARIMA bien ajusté, les résidus d'innovation doivent ressembler aux innovations du modèle, donc à un bruit blanc.

Concrètement, on vérifie trois points :

1) les résidus sont-ils centrés autour de zéro ?
2) leur variance semble-t-elle à peu près constante ?
3) leur ACF empirique montre-t-elle encore des autocorrélations significatives ?

Définition 6.4.1: ACF empirique des résidus

Soient $\widehat{z}_1, \ldots, \widehat{z}_n$ des résidus et $\overline{z}$ leur moyenne. L'autocorrélation empirique des résidus au retard h est

$$\widehat{\rho}_{\widehat{z}}(h) = \frac{\sum_{t=1}^{n-h} (\widehat{z}_t - \overline{z})(\widehat{z}_{t+h} - \overline{z})}{\sum_{t=1}^{n} (\widehat{z}_t - \overline{z})^2}.$$

Si les résidus sont proches d'un bruit blanc, les valeurs $\widehat{\rho}_{\widehat{z}}(h)$ doivent être petites pour $h \geq 1$. Il ne faut pas exiger qu'elles soient exactement nulles : on travaille avec un échantillon fini. La question est plutôt de savoir si elles sont compatibles avec des fluctuations aléatoires.

6.4.2 Tests portemanteau : Box–Pierce et Ljung–Box

Les tests portemanteau résument plusieurs autocorrélations résiduelles en une seule statistique. Ils testent globalement si les premières autocorrélations sont toutes nulles.

Définition 6.4.2: Statistique de Ljung–Box

Pour des résidus $\widehat{z}_1, \ldots, \widehat{z}_n$, la statistique de Ljung–Box au retard maximal m est

$$Q_{LB} = n(n+2) \sum_{h=1}^{m} \frac{\widehat{\rho}_{\widehat{z}}(h)^2}{n-h}.$$

Elle sert à tester l'hypothèse nulle : les autocorrélations des résidus jusqu'au retard m sont nulles.

Théorème 6.4.3: Approximation asymptotique du test de Ljung–Box

Sous des hypothèses de régularité et si le modèle ajusté est correct, la statistique de Ljung–Box est approximativement distribuée comme une loi du χ^2 à $m - k$ degrés de liberté, où k est le nombre de paramètres ARMA estimés dans la partie stationnaire du modèle.

Remarque 6.4.4: Lire une p-valeur

Dans `R`, la commande `Box.test(..., type = "Ljung-Box")` renvoie une p-valeur. Une petite p-valeur indique que les résidus ne ressemblent pas à un bruit blanc : il reste de l'autocorrélation. Une grande p-valeur ne prouve pas que le modèle est vrai ; elle dit seulement que ce test n'a pas détecté d'autocorrélation résiduelle significative.

6.5 Prévision

6.5.1 Prévision ponctuelle et erreur de prévision

Une prévision n'est pas seulement un nombre. Elle doit être accompagnée d'une incertitude. Un modèle de série temporelle donne typiquement une prévision ponctuelle et une variance d'erreur de prévision.

Définition 6.5.1: Erreur de prévision

Soit $\widehat{X}_{n+h|n}$ la prévision de X_{n+h} à partir de l'information disponible jusqu'au temps n. L'erreur de prévision à horizon h est

$$e_{n+h|n} = X_{n+h} - \widehat{X}_{n+h|n}.$$

Sa variance mesure l'incertitude associée à la prévision.

En général, l'incertitude augmente avec l'horizon h. C'est normal : prévoir demain est plus facile que prévoir dans deux ans. Pour un ARMA stationnaire, les prévisions finissent par revenir vers la moyenne. Pour un ARIMA non stationnaire, les prévisions portent sur un niveau obtenu en ré-accumulant des variations ; l'incertitude peut croître plus fortement.

6.5.2 Cas explicite : prévision d'un AR(1)

Le cas AR(1) permet de voir les formules sans boîte noire. Supposons

$$X_t = \phi X_{t-1} + Z_t, \qquad |\phi| < 1,$$

avec (Z_t) bruit blanc centré de variance σ^2, et supposons X_n observé.

Proposition 6.5.2: Prévision d'un AR(1)

Pour $h \geq 1$, la prévision à horizon h est

$$\widehat{X}_{n+h|n} = \phi^h X_n.$$

L'erreur de prévision vaut

$$e_{n+h|n} = \sum_{j=0}^{h-1} \phi^j Z_{n+h-j},$$

et sa variance est

$$\mathrm{Var}\left(e_{n+h|n}\right) = \sigma^2 \sum_{j=0}^{h-1} \phi^{2j} = \sigma^2 \frac{1-\phi^{2h}}{1-\phi^2}.$$

Démonstration. On itère l'équation du modèle :

$$X_{n+h} = \phi^h X_n + \phi^{h-1} Z_{n+1} + \phi^{h-2} Z_{n+2} + \cdots + Z_{n+h}.$$

À la date n, les innovations futures ont espérance conditionnelle nulle. La partie prévisible est donc $\phi^h X_n$. L'erreur de prévision est la somme des innovations futures pondérées. Comme elles sont non corrélées et de variance σ^2, la variance de la somme est la somme des variances. □

Si les innovations sont gaussiennes, un intervalle de prévision approximatif à 95% est

$$\widehat{X}_{n+h|n} \pm 1.96\sqrt{\mathrm{Var}\left(e_{n+h|n}\right)}.$$

Cette formule explique ce que fait un logiciel : il ne donne pas seulement une courbe future, mais une zone d'incertitude autour de cette courbe.

6.6 Cas pratique : AirPassengers

La série `AirPassengers`, déjà rencontrée, est un bon exemple pour comprendre le flux de travail complet : transformer, ajuster, diagnostiquer, prévoir. On travaille sur le logarithme, puis on applique une différence saisonnière et une différence ordinaire. Une classe naturelle de modèles est alors

$$\text{SARIMA}(p,1,q)\times(P,1,Q)_{12}$$

sur la série brute, ou de manière équivalente un modèle ARMA sur la série transformée $(1-B)(1-B^{12})\log X_t$.

```
data(AirPassengers)
y <- log(AirPassengers)
fit1 <- arima(y, order = c(0,1,1),
              seasonal = list(order = c(0,1,1), period = 12))
fit2 <- arima(y, order = c(1,1,0),
              seasonal = list(order = c(0,1,1), period = 12))
AIC(fit1, fit2)
```

La comparaison par AIC donne une première indication. Elle doit être suivie d'un diagnostic des résidus.

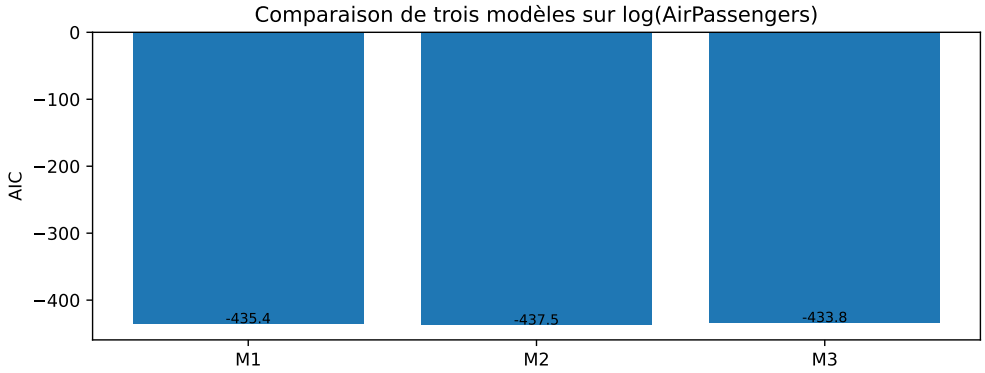


Figure 6.1 – Comparaison de trois modèles SARIMA sur log(`AirPassengers`). Le critère AIC compare l'ajustement pénalisé par le nombre de paramètres. Le plus petit AIC est préféré parmi les modèles comparés, mais il faut encore examiner les résidus.

```
res <- residuals(fit1)
plot.ts(res)
acf(res)
Box.test(res, lag = 20, type = "Ljung-Box", fitdf = 2)
```

Si le diagnostic est raisonnable, on peut produire une prévision. La commande `predict()` fournit des prévisions et des erreurs standards sur l'échelle utilisée pour ajuster le modèle. Si l'on a ajusté le modèle à $\log X_t$, il faut revenir à l'échelle initiale en exponentiant les prévisions. Cette étape demande de la prudence : exponentier une prévision logarithmique donne une prévision médiane approximative sur l'échelle initiale, mais pas exactement l'espérance.

```
prev <- predict(fit1, n.ahead = 24)
prev.moy <- exp(prev$pred)
prev.inf <- exp(prev$pred - 1.96 * prev$se)
prev.sup <- exp(prev$pred + 1.96 * prev$se)
```

6.7 Validation hors-échantillon

Une méthode très utile consiste à ne pas utiliser toutes les données pour estimer le modèle. On garde une fin de série comme période test. On ajuste le modèle sur la période d'apprentissage, puis

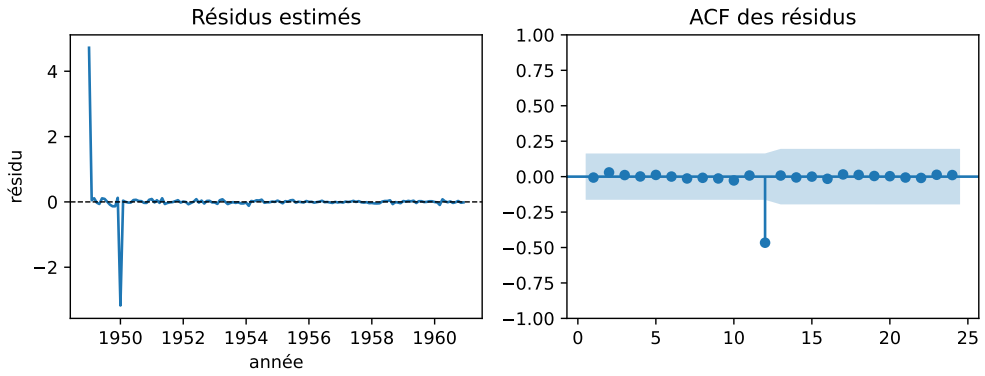


Figure 6.2 – Diagnostic typique après ajustement d'un modèle SARIMA à log(`AirPassengers`). À gauche, on cherche une structure visible dans les résidus ; à droite, on vérifie si l'ACF empirique des résidus contient encore des pics importants.

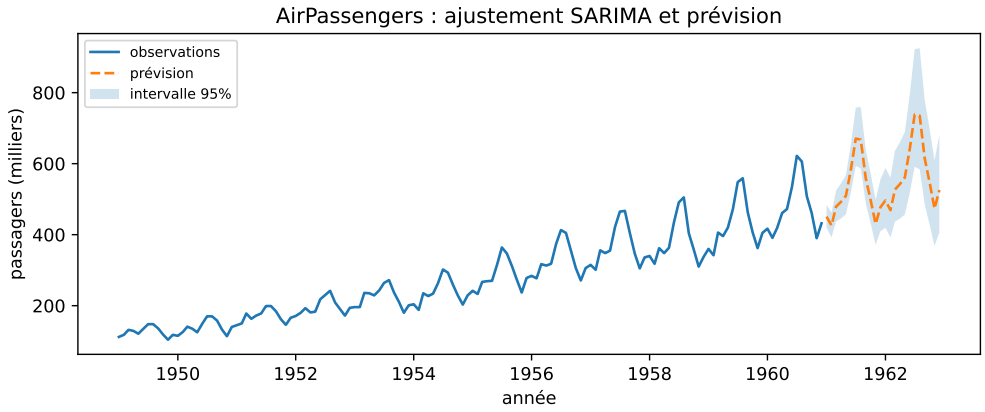


Figure 6.3 – Prévision à 24 mois pour `AirPassengers`, obtenue après ajustement d'un modèle SARIMA sur le logarithme de la série. L'éventail de prévision s'élargit avec l'horizon, ce qui traduit l'accumulation de l'incertitude.

on compare les prévisions aux observations réellement gardées de côté.

Définition 6.7.1 : Échantillon d'apprentissage et échantillon test

On appelle *échantillon d'apprentissage* la partie de la série utilisée pour estimer le modèle. On appelle *échantillon test* une partie ultérieure de la série, non utilisée pour l'estimation, qui sert à évaluer la qualité prédictive du modèle.

Cette validation est particulièrement importante lorsque l'objectif du modèle est la prévision. Un modèle peut très bien s'ajuster au passé et prévoir mal le futur. La séparation apprentissage/test force le modèle à être évalué sur des données qu'il n'a pas vues.

Définition 6.7.2 : Erreurs de prévision MAE et RMSE

Si $x_{n+1}, \dots, x_{n+m}$ sont les valeurs test et $\widehat{x}_{n+1|n}, \dots, \widehat{x}_{n+m|n}$ les prévisions correspondantes, on définit

$$\mathrm{MAE} = \frac{1}{m} \sum_{h=1}^{m} \left| x_{n+h} - \widehat{x}_{n+h|n} \right|$$

et

$$\mathrm{RMSE} = \sqrt{\frac{1}{m} \sum_{h=1}^{m} \left(x_{n+h} - \widehat{x}_{n+h|n} \right)^2}.$$

La MAE mesure une erreur moyenne en valeur absolue. La RMSE pénalise plus fortement les

grandes erreurs, car elle utilise des carrés. Il n'y a pas un critère universellement meilleur ; il faut choisir en fonction du contexte.

6.8 Synthèse du chapitre

À retenir :

1) Estimer un modèle, c'est choisir ses paramètres à partir des données.
2) Les résidus d'innovation estimés doivent ressembler à un bruit blanc si le modèle a capturé la dépendance temporelle principale.
3) AIC et BIC comparent des modèles ajustés en pénalisant le nombre de paramètres ; ils ne remplacent pas le diagnostic.
4) Le test de Ljung–Box cherche une autocorrélation globale dans les résidus.
5) Une prévision doit être accompagnée d'une incertitude, souvent sous forme d'intervalle de prévision.
6) La validation hors-échantillon est le meilleur moyen simple de savoir si un modèle prédit raisonnablement bien.

Le chapitre suivant présentera le lissage exponentiel. Cette famille de méthodes part d'une idée différente : au lieu d'écrire explicitement un modèle ARIMA, on met à jour récursivement un niveau, une tendance et éventuellement une saisonnalité pour produire des prévisions.

Chapitre 7

Lissage exponentiel et méthodes de prévision récursives

Mots-clés

Lissage exponentiel simple ; poids exponentiellement décroissants ; niveau ; tendance ; méthode de Holt ; saisonnalité additive ; saisonnalité multiplicative ; méthode de Holt–Winters ; paramètres de lissage ; erreur de prévision.

7.1 Pourquoi introduire le lissage exponentiel ?

Les chapitres précédents ont suivi une logique de modélisation : on cherche d'abord à rendre la série stationnaire, puis on ajuste un modèle ARMA ou ARIMA, on diagnostique les résidus, et enfin on prévoit. Cette démarche est fondamentale, car elle oblige à se demander quelle dépendance temporelle est présente dans les données.

Il existe cependant une autre famille de méthodes, très utilisée en pratique, dont le point de départ est différent. Au lieu de commencer par écrire un modèle ARIMA explicite, on met à jour récursivement quelques composantes simples de la série : un niveau, parfois une tendance, parfois une saisonnalité. La prévision est alors obtenue en prolongeant ces composantes.

Ces méthodes s'appellent les *méthodes de lissage exponentiel*. Leur idée centrale est simple : pour prévoir demain, les observations récentes doivent en général compter plus que les observations anciennes. Mais on ne veut pas oublier brutalement le passé ; on veut plutôt lui donner un poids qui décroît progressivement.

Idée à retenir

Un modèle ARIMA décrit la dépendance temporelle à l'aide d'équations sur les observations et les innovations. Le lissage exponentiel part d'une logique plus directement prédictive : il construit une estimation récursive du niveau, de la tendance et de la saisonnalité, puis extrapole ces quantités.

Cette différence ne signifie pas que les deux approches sont sans lien. Certaines méthodes de lissage exponentiel peuvent s'interpréter comme des modèles ARIMA particuliers. Mais, pédagogiquement, il est utile de les présenter séparément : les modèles ARIMA mettent l'accent sur la structure probabiliste, tandis que le lissage exponentiel met l'accent sur la mise à jour et la prévision.

7.2 Du lissage par moyenne mobile au lissage exponentiel

Au chapitre 3, nous avons vu que l'on pouvait lisser une série en prenant une moyenne mobile, par exemple

$$Y_t = \frac{1}{3}\left(X_{t-1} + X_t + X_{t+1}\right).$$

Cette opération réduit les fluctuations rapides, mais elle a deux limites importantes pour la prévision.

D'abord, une moyenne mobile centrée utilise le futur : pour calculer Y_t, elle utilise X_{t+1}. Elle est donc adaptée à l'analyse descriptive, mais pas directement à une prévision en temps réel. Ensuite, elle donne souvent le même poids à toutes les observations de la fenêtre, puis un poids nul aux observations plus anciennes. Cette coupure est assez artificielle.

Le lissage exponentiel remplace cette coupure par des poids qui décroissent régulièrement. Les observations très récentes ont un poids élevé ; les observations anciennes ne disparaissent pas immédiatement, mais leur poids devient de plus en plus faible.

Définition 7.2.1: Poids exponentiellement décroissants

Soit $\alpha \in (0,1)$. Les poids exponentiellement décroissants associés à α sont

$$w_j = \alpha(1-\alpha)^j, \qquad j \geq 0.$$

Ils vérifient

$$\sum_{j=0}^{+\infty} w_j = 1.$$

Démonstration. Il s'agit d'une série géométrique :

$$\sum_{j=0}^{+\infty} \alpha(1-\alpha)^j = \alpha \frac{1}{1-(1-\alpha)} = 1.$$

□

Le paramètre α règle la mémoire de la méthode. Si α est grand, le poids de l'observation la plus récente est grand et les poids anciens décroissent vite : la méthode réagit rapidement aux changements. Si α est petit, les poids décroissent lentement : la méthode est plus stable, mais réagit moins vite.

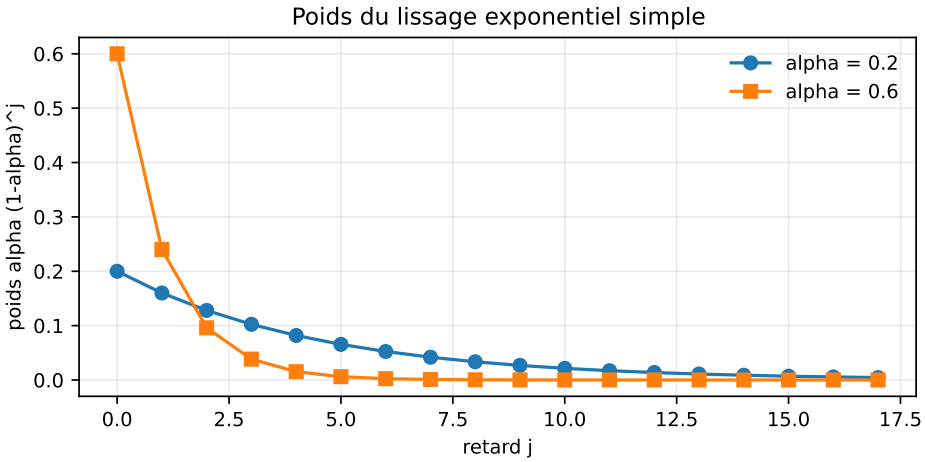


Figure 7.1 – Poids $\alpha(1-\alpha)^j$ donnés à l'observation située j périodes dans le passé. Pour $\alpha = 0.6$, les observations très récentes dominent fortement. Pour $\alpha = 0.2$, le passé plus ancien continue à compter.

7.3 Lissage exponentiel simple

Le lissage exponentiel simple est adapté aux séries qui fluctuent autour d'un niveau qui peut évoluer lentement, mais sans tendance persistante ni saisonnalité nette. L'objectif n'est pas de modéliser toutes les oscillations : on cherche une estimation du niveau courant de la série.

Définition 7.3.1: Lissage exponentiel simple

Soit (x_t) une série observée. Le lissage exponentiel simple avec paramètre $\alpha \in (0,1)$ définit récursivement un niveau ℓ_t par

$$\ell_t = \alpha x_t + (1-\alpha)\ell_{t-1}.$$

La prévision à tout horizon $h \geq 1$, faite à la date t, est

$$\widehat{x}_{t+h|t} = \ell_t.$$

La formule dit que le nouveau niveau est un compromis entre l'observation actuelle x_t et l'ancien niveau ℓ_{t-1}. Le paramètre α mesure la confiance accordée à la nouvelle observation.

Proposition 7.3.2: Forme en moyenne pondérée

Si le lissage exponentiel simple est initialisé dans le passé lointain, alors on peut écrire formellement

$$\ell_t = \sum_{j=0}^{+\infty} \alpha(1-\alpha)^j x_{t-j}.$$

Ainsi, ℓ_t est une moyenne pondérée des observations passées, avec des poids exponentiellement décroissants.

Démonstration. En remplaçant ℓ_{t-1} par sa définition, on obtient

$$\begin{aligned}\ell_t &= \alpha x_t + (1-\alpha)\ell_{t-1} \\ &= \alpha x_t + \alpha(1-\alpha)x_{t-1} + (1-\alpha)^2\ell_{t-2}.\end{aligned}$$

En répétant l'opération m fois,

$$\ell_t = \sum_{j=0}^{m} \alpha(1-\alpha)^j x_{t-j} + (1-\alpha)^{m+1}\ell_{t-m-1}.$$

Si le dernier terme devient négligeable lorsque m est grand, on obtient la formule annoncée. □

Exemple 7.3.3: Lissage simple du débit du Nil

La série `Nile`, fournie avec `R`, donne le débit annuel du Nil à Assouan. Elle ne présente pas de saisonnalité, car elle est annuelle. Le lissage exponentiel simple peut donc servir à construire un niveau lissé.

```
data(Nile)
fit02 <- HoltWinters(Nile, alpha = 0.2, beta = FALSE, gamma = FALSE)
fit08 <- HoltWinters(Nile, alpha = 0.8, beta = FALSE, gamma = FALSE)
```

Avec $\alpha = 0.2$, la courbe lissée est stable : elle réagit lentement. Avec $\alpha = 0.8$, elle suit beaucoup plus fortement les dernières observations. Le choix de α est donc un choix de compromis entre stabilité et réactivité.

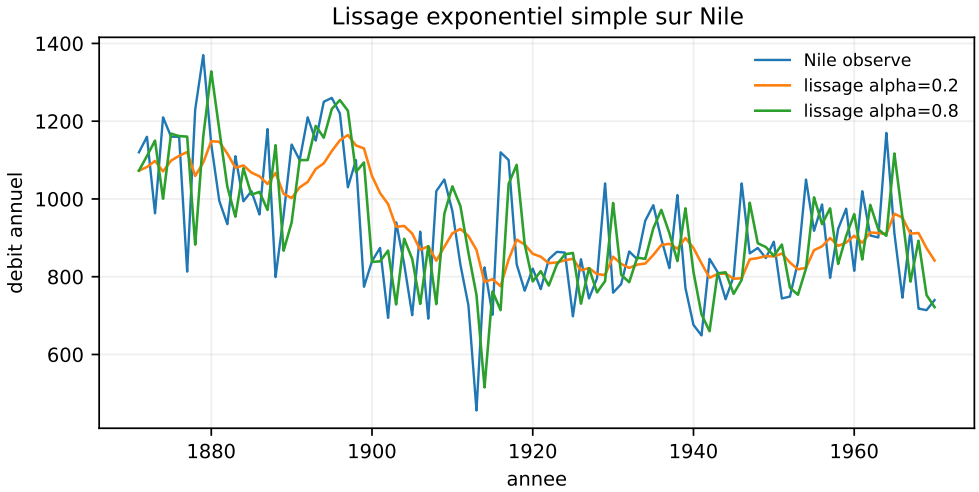


Figure 7.2 – Lissage exponentiel simple de la série `Nile`. Une valeur faible de α produit un lissage plus stable ; une valeur élevée produit un lissage plus réactif.

7.4 Choix du paramètre de lissage

Dans un exercice, on peut fixer α pour comprendre son effet. Dans une analyse réelle, on le choisit généralement à partir des données. L'idée la plus simple consiste à choisir α pour rendre les erreurs de prévision à un pas aussi petites que possible.

Définition 7.4.1: Erreurs à un pas et somme des carrés

Dans le lissage exponentiel simple, la prévision de x_t faite à la date $t-1$ est ℓ_{t-1}. L'erreur à un pas est donc

$$e_t = x_t - \ell_{t-1}.$$

On peut choisir α en minimisant

$$\mathrm{SSE}(\alpha) = \sum_t e_t^2.$$

Ce critère n'est pas une vérité absolue, mais il formalise une idée naturelle : un bon paramètre de lissage doit produire de bonnes prévisions à court terme sur le passé observé. Les logiciels comme `HoltWinters()` estiment les paramètres de lissage en optimisant un critère de ce type.

7.5 Méthode de Holt : niveau et tendance

Le lissage exponentiel simple prévoit toujours un futur plat : toutes les prévisions sont égales au dernier niveau ℓ_t. C'est raisonnable si la série fluctue autour d'un niveau stable, mais c'est insuffisant si la série présente une tendance persistante.

La méthode de Holt ajoute une deuxième composante : une estimation de la pente locale. On met donc à jour deux quantités : le niveau ℓ_t et la tendance b_t. La prévision prolonge ensuite la droite locale.

Définition 7.5.1: Méthode de Holt

La méthode de Holt, ou lissage exponentiel double, dépend de deux paramètres $\alpha, \beta \in (0,1)$. Elle est définie par

$$\ell_t = \alpha x_t + (1-\alpha)(\ell_{t-1} + b_{t-1}),$$
$$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta) b_{t-1}.$$

La prévision à horizon $h \geq 1$ est

$$\widehat{x}_{t+h|t} = \ell_t + h b_t.$$

La première équation met à jour le niveau en comparant l'observation actuelle avec ce que l'on attendait à partir du niveau et de la tendance précédents. La deuxième met à jour la tendance : la nouvelle pente est un compromis entre la variation récente du niveau et l'ancienne pente.

Remarque 7.5.2: Pourquoi ce modèle n'est pas seulement une droite

Une régression linéaire ajuste une seule droite sur toute la période. La méthode de Holt ajuste une tendance locale qui peut évoluer au cours du temps. Elle est donc plus souple pour des séries dont la pente change lentement, sans pour autant introduire un modèle ARIMA complet.

7.6 Méthode de Holt–Winters : niveau, tendance et saisonnalité

La méthode de Holt capture une tendance, mais pas une saisonnalité. Or beaucoup de séries actuarielles, économiques ou commerciales présentent un motif qui revient chaque mois, trimestre ou année. On a alors besoin de mettre à jour trois composantes : niveau, tendance et saisonnalité.

Il faut distinguer deux cas. Si l'amplitude saisonnière est à peu près constante en valeur absolue, on utilise une saisonnalité additive. Si l'amplitude saisonnière augmente avec le niveau de la série, on utilise plutôt une saisonnalité multiplicative, ou bien on travaille sur le logarithme de la série.

7.6.1 Version additive

Définition 7.6.1: Holt–Winters additif

Soit m la période saisonnière. La méthode de Holt–Winters additive met à jour un niveau ℓ_t, une tendance b_t et une saisonnalité s_t par

$$\ell_t = \alpha(x_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}),$$
$$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta) b_{t-1},$$
$$s_t = \gamma(x_t - \ell_t) + (1-\gamma) s_{t-m}.$$

La prévision à horizon h est

$$\widehat{x}_{t+h|t} = \ell_t + h b_t + s_{t-m+1+((h-1) \bmod m)}.$$

Dans la version additive, la série est pensée sous la forme

$$x_t \simeq \text{niveau}_t + \text{tendance}_t + \text{saisonnalité}_t.$$

Le même mois ou trimestre ajoute à peu près la même quantité, quel que soit le niveau général de la série.

7.6.2 Version multiplicative

Définition 7.6.2: Holt–Winters multiplicatif

La méthode de Holt–Winters multiplicative remplace les corrections additives de saisonnalité par des corrections multiplicatives. Une écriture usuelle est

$$\begin{aligned}
\ell_t &= \alpha \frac{x_t}{s_{t-m}} + (1-\alpha)(\ell_{t-1} + b_{t-1}),\\
b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta) b_{t-1},\\
s_t &= \gamma \frac{x_t}{\ell_t} + (1-\gamma) s_{t-m}.
\end{aligned}$$

La prévision à horizon h est

$$\widehat{x}_{t+h|t} = (\ell_t + h b_t)\, s_{t-m+1+((h-1) \bmod m)}.$$

Dans la version multiplicative, un mois donné multiplie le niveau par un facteur. C'est adapté lorsqu'un pic saisonnier représente par exemple environ 20% de plus que le niveau moyen, et non une quantité fixe.

Point de vigilance

La méthode multiplicative demande des observations positives. Si la série peut prendre des valeurs nulles ou négatives, on ne peut pas diviser par une composante saisonnière multiplicative de manière simple.

7.7 Cas pratique : AirPassengers

La série `AirPassengers` est un exemple classique : elle possède une tendance croissante et une saisonnalité dont l'amplitude augmente avec le niveau. C'est précisément une situation où la version multiplicative ou le logarithme sont naturels.

```
data(AirPassengers)
fit.add <- HoltWinters(AirPassengers, seasonal = "additive")
fit.mul <- HoltWinters(AirPassengers, seasonal = "multiplicative")
prev.add <- predict(fit.add, n.ahead = 24)
prev.mul <- predict(fit.mul, n.ahead = 24)
```

La méthode additive prolonge une saisonnalité en amplitude absolue. La méthode multiplicative prolonge une saisonnalité proportionnelle au niveau. Sur `AirPassengers`, la deuxième est souvent plus cohérente visuellement, car les pics saisonniers deviennent plus grands lorsque le nombre de passagers augmente.

7.8 Diagnostic et limites

Le lissage exponentiel est simple et efficace, mais il ne dispense pas de contrôler les erreurs. Après ajustement, on regarde les résidus : s'ils contiennent encore une structure évidente, la méthode choisie n'a pas capturé toute l'information temporelle utile.

```
res <- residuals(fit.mul)
plot.ts(res)
acf(res)
```

On peut également utiliser une validation hors-échantillon, comme au chapitre 6. On ajuste la méthode sur une partie de la série et on compare les prévisions aux observations gardées de côté.

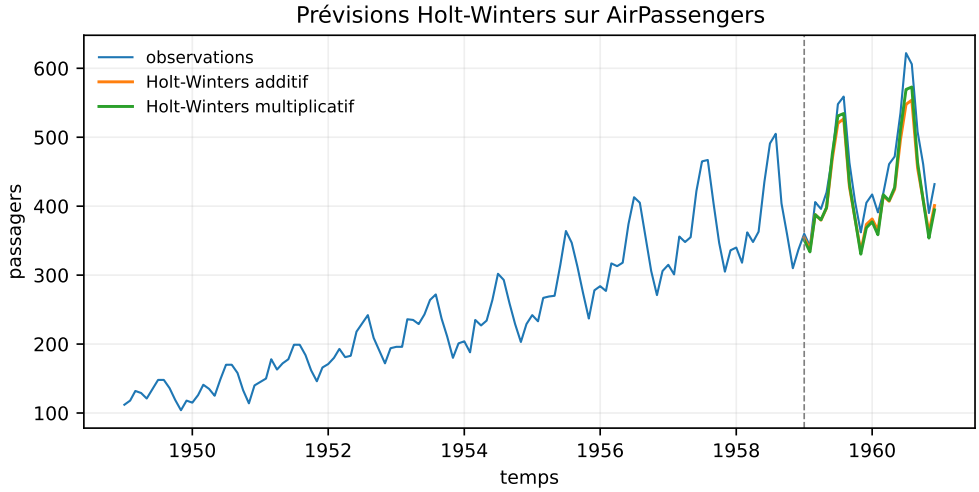


Figure 7.3 – Prévisions Holt–Winters sur `AirPassengers`. La série présente une saisonnalité dont l'amplitude augmente avec le niveau, ce qui rend la version multiplicative particulièrement naturelle.

Idée à retenir

Le lissage exponentiel est très utile quand l'objectif principal est la prévision et que la série s'explique bien par niveau, tendance et saisonnalité. Il est moins adapté si l'on veut interpréter finement une structure de dépendance ARMA ou tester un modèle probabiliste précis.

7.9 Synthèse du chapitre

À retenir :

1) Le lissage exponentiel donne plus de poids aux observations récentes qu'aux observations anciennes.
2) Le lissage exponentiel simple estime un niveau et produit une prévision plate.
3) La méthode de Holt ajoute une tendance locale et produit des prévisions linéaires.
4) La méthode de Holt–Winters ajoute une composante saisonnière.
5) La saisonnalité additive convient lorsque l'amplitude saisonnière est à peu près constante ; la saisonnalité multiplicative convient lorsque l'amplitude est proportionnelle au niveau.
6) Les paramètres de lissage règlent le compromis entre stabilité et réactivité.
7) Comme pour ARIMA, il faut contrôler les résidus et évaluer la qualité prédictive.

Le chapitre suivant reviendra sur la simulation. Simuler des séries permet de comprendre concrètement les modèles : on peut générer un processus dont on connaît le mécanisme, puis vérifier si les outils graphiques et statistiques permettent de retrouver ce mécanisme.

Chapitre 8

Ouvertures : simulation, analyse spectrale et volatilité conditionnelle

Mots-clés

Simulation ; fréquence ; période ; composante cyclique ; transformée de Fourier discrète ; périodogramme ; densité spectrale ; volatilité ; rendement ; hétéroscédasticité conditionnelle ; ARCH ; GARCH.

8.1 Pourquoi terminer par ces ouvertures ?

Les chapitres précédents ont construit une chaîne complète pour traiter une série temporelle : observer la série, la transformer si nécessaire, la rendre approximativement stationnaire, choisir un modèle ARMA ou ARIMA, estimer ses paramètres, contrôler les résidus, puis prévoir. Cette démarche est déjà très utile en pratique.

Il reste cependant deux phénomènes importants que les outils précédents ne décrivent pas toujours bien.

Le premier est la présence de *cycles*. Une saisonnalité de période 12 est facile à voir sur un chronogramme mensuel, mais certains cycles sont moins réguliers, plus bruités, ou mélangés à d'autres fluctuations. Dans ce cas, il peut être plus naturel de regarder la série non plus seulement dans le temps, mais aussi selon les *fréquences* qui la composent. C'est l'idée de l'analyse spectrale.

Le second est la présence d'une variabilité qui change au cours du temps. Dans une série financière, par exemple, les rendements peuvent avoir une moyenne proche de zéro et une autocorrélation faible, mais alterner entre périodes calmes et périodes très agitées. Un modèle ARMA standard capture une dépendance dans la moyenne, mais pas directement cette dépendance dans la variance. Cela mène aux modèles ARCH et GARCH.

Idée à retenir

Ce chapitre est une ouverture. L'analyse spectrale répond à la question : "quelles périodes ou fréquences sont présentes dans la série ?" Les modèles ARCH/GARCH répondent à la question : "la variabilité future dépend-elle de la variabilité récente ?"

Nous commençons par la simulation, car elle permet de comprendre ces deux idées sur des exemples dont le mécanisme est connu.

8.2 Simuler pour comprendre un modèle

Simuler une série temporelle consiste à fabriquer une trajectoire à partir d'un modèle choisi. Cela peut paraître moins concret que l'analyse d'une série réelle, mais c'est un outil pédagogique très puissant : quand on simule un AR(1), une série saisonnière bruitée ou un processus à volatilité

variable, on connaît la vérité. On peut donc vérifier si les graphiques et diagnostics utilisés dans le cours retrouvent effectivement la structure que l'on a mise dans le modèle.

Exemple 8.2.1: Une oscillation bruitée

Supposons que l'on observe

$$X_t = 2\cos\left(\frac{2\pi t}{12}\right) + Z_t,$$

où (Z_t) est un bruit blanc. La partie déterministe possède une période 12 : elle revient au même niveau tous les 12 pas de temps. Le bruit rend cette périodicité moins évidente sur le chronogramme, mais l'information périodique reste présente.

La figure suivante illustre l'effet de la période. Une oscillation de période 12 varie plus lentement qu'une oscillation de période 6.

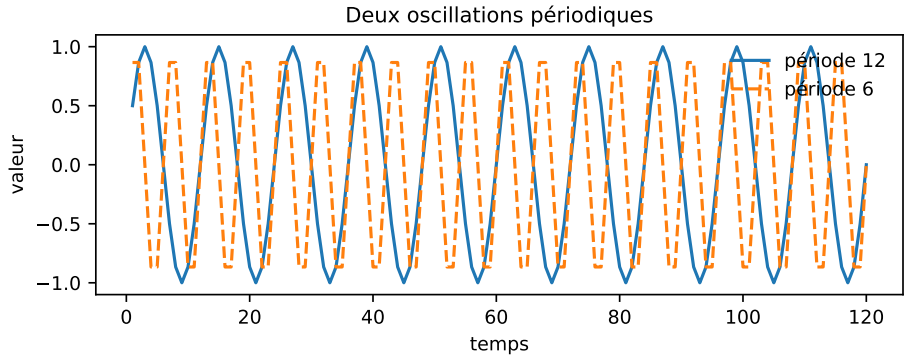


Figure 8.1 – Deux oscillations périodiques. La période est le nombre de pas de temps nécessaires pour retrouver le même état du cycle.

8.3 Période, fréquence et cycles

Avant de parler de périodogramme, il faut fixer le vocabulaire. Dans les chapitres précédents, nous avons surtout parlé de saisonnalité en termes de période : période 4 pour une série trimestrielle, période 12 pour une série mensuelle. L'analyse spectrale utilise plutôt le langage des fréquences.

Définition 8.3.1: Période et fréquence

Une fonction déterministe (s_t) est périodique de période entière $p \geq 1$ si

$$s_{t+p} = s_t \qquad \text{pour tout } t.$$

La fréquence associée est l'inverse de la période :

$$f = \frac{1}{p}.$$

Elle mesure le nombre de cycles par unité de temps.

Cette définition est volontairement simple. Si une série est mensuelle et possède une saisonnalité annuelle, sa période est 12 mois et sa fréquence est $1/12$ cycle par mois. Si une série annuelle possède un cycle d'environ 11 ans, sa fréquence est environ $1/11$ cycle par an.

Proposition 8.3.2: Sinusoïde et période

Soit

$$s_t = A\cos(\lambda t) + B\sin(\lambda t),$$

où $A, B \in \mathbb{R}$. Si $\lambda = 2\pi/p$, avec $p \in \mathbb{N}^*$, alors (s_t) est périodique de période p.

Démonstration. Pour tout t, on a

$$\cos(\lambda(t+p)) = \cos(\lambda t + \lambda p) = \cos(\lambda t + 2\pi) = \cos(\lambda t),$$

et de même pour le sinus. Donc $s_{t+p} = s_t$. □

Idée à retenir

La période dit au bout de combien de temps un motif revient. La fréquence dit combien de motifs complets on observe par unité de temps. Ce sont deux manières de dire la même chose, mais l'analyse spectrale préfère le langage des fréquences.

8.4 Transformée de Fourier discrète et périodogramme

Le chronogramme décrit la série dans le domaine temporel : il montre comment x_t évolue avec t. Le périodogramme décrit la même série dans le domaine fréquentiel : il cherche à repérer quelles fréquences expliquent une part importante des variations observées.

L'idée est analogue à celle d'un accord musical. Un son complexe peut être décomposé en plusieurs notes. Une série temporelle peut, elle aussi, contenir plusieurs oscillations de fréquences différentes, auxquelles s'ajoute du bruit.

Définition 8.4.1: Transformée de Fourier discrète

Soit $x_1, \dots, x_n$ une série réelle finie. Pour $k = 0, \dots, n-1$, on définit le coefficient de Fourier discret

$$d_k = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t \exp\left(-2\pi i \frac{kt}{n}\right).$$

La fréquence associée est k/n.

Le coefficient d_k mesure la présence d'une oscillation de fréquence k/n dans la série observée. Sa taille ne s'interprète pas directement sur l'échelle de la série ; on regarde plutôt son module au carré.

Définition 8.4.2: Périodogramme

Le périodogramme de $x_1, \dots, x_n$ est la quantité

$$I_n\left(\frac{k}{n}\right) = \frac{1}{n}\left|\sum_{t=1}^{n} x_t \exp\left(-2\pi i \frac{kt}{n}\right)\right|^2, \qquad k = 0, \dots, n-1.$$

En pratique, on centre souvent la série avant de le calculer, afin d'éviter que le niveau moyen domine le graphique.

Théorème 8.4.3: Orthogonalité de Fourier et conservation de l'énergie

Avec la normalisation de la définition 8.4.1, on a l'identité

$$\sum_{t=1}^{n} |x_t|^2 = \sum_{k=0}^{n-1} |d_k|^2.$$

Démonstration. On utilise l'orthogonalité des exponentielles complexes : pour tous $s, t \in \{1, \dots, n\}$,

$$\frac{1}{n} \sum_{k=0}^{n-1} \exp\left(2\pi i \frac{k(s-t)}{n}\right) = \begin{cases} 1, & s = t, \\ 0, & s \neq t. \end{cases}$$

Alors

$$\begin{aligned} \sum_{k=0}^{n-1} |d_k|^2 &= \sum_{k=0}^{n-1} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t e^{-2\pi i k t/n}\right) \left(\frac{1}{\sqrt{n}} \sum_{s=1}^{n} x_s e^{2\pi i k s/n}\right) \\ &= \sum_{t=1}^{n} \sum_{s=1}^{n} x_t x_s \left(\frac{1}{n} \sum_{k=0}^{n-1} e^{2\pi i k(s-t)/n}\right) = \sum_{t=1}^{n} x_t^2. \end{aligned}$$

□

Ce résultat explique pourquoi le périodogramme est une notion naturelle : il répartit l'énergie de la série entre différentes fréquences. Un pic important indique qu'une fréquence contribue fortement aux variations observées.

Point de vigilance

Un pic dans un périodogramme n'est pas automatiquement une preuve mathématique d'un cycle exact. Sur une série courte ou très bruitée, il faut rester prudent. Le périodogramme est d'abord un outil exploratoire.

8.5 Densité spectrale d'un processus stationnaire

Le périodogramme est calculé sur une série finie observée. Pour un processus stationnaire théorique, l'objet correspondant est la densité spectrale. Elle contient la même information que l'autocovariance, mais exprimée selon les fréquences plutôt que selon les retards.

Dans les chapitres précédents, l'autocovariance $\gamma(h)$ répondait à la question : "comment X_t dépend-il de X_{t+h} ?" La densité spectrale répond à une autre question : "quelles fréquences dominent la variabilité du processus ?"

Définition 8.5.1: Densité spectrale

Soit $(X_t)_{t\in\mathbb{Z}}$ un processus stationnaire du second ordre, de fonction d'autocovariance γ. On suppose que

$$\sum_{h\in\mathbb{Z}} |\gamma(h)| < +\infty.$$

La densité spectrale de X est la fonction

$$f(\lambda) = \frac{1}{2\pi} \sum_{h\in\mathbb{Z}} \gamma(h) e^{-ih\lambda}, \qquad \lambda \in [-\pi, \pi].$$

La condition de sommabilité absolue de γ garantit que la série de Fourier converge bien. Elle est vérifiée pour beaucoup de modèles ARMA causaux, dont l'autocovariance décroît assez vite.

Théorème 8.5.2: Formule d'inversion

Sous l'hypothèse de la définition précédente, on a, pour tout $h \in \mathbb{Z}$,

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda)\, d\lambda.$$

Ainsi, la densité spectrale et l'autocovariance contiennent la même information de second ordre.

Démonstration. En remplaçant f par sa définition, on obtient formellement

$$\int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda)\, d\lambda = \frac{1}{2\pi} \sum_{k\in\mathbb{Z}} \gamma(k) \int_{-\pi}^{\pi} e^{i(h-k)\lambda}\, d\lambda.$$

La sommabilité absolue permet d'échanger somme et intégrale. Or

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i(h-k)\lambda}\, d\lambda = \begin{cases} 1, & h = k, \\ 0, & h \neq k. \end{cases}$$

Il reste donc seulement le terme $k = h$, c'est-à-dire $\gamma(h)$. □

Exemple 8.5.3: Bruit blanc

Si (Z_t) est un bruit blanc centré de variance σ^2, alors

$$\gamma(h) = \sigma^2 \mathbf{1}_{h=0}.$$

Sa densité spectrale vaut donc

$$f(\lambda) = \frac{\sigma^2}{2\pi}.$$

Elle est constante : aucune fréquence ne domine. C'est l'origine du mot "blanc", par analogie avec la lumière blanche qui contient toutes les fréquences visibles.

Exemple 8.5.4: Densité spectrale d'un AR(1)

Soit (X_t) le processus AR(1) causal

$$X_t = \phi X_{t-1} + Z_t, \qquad |\phi| < 1,$$

où (Z_t) est un bruit blanc de variance σ^2. Comme

$$X_t = \sum_{j=0}^{+\infty} \phi^j Z_{t-j},$$

le filtre associé a pour fonction de transfert

$$\frac{1}{1 - \phi e^{-i\lambda}}.$$

La densité spectrale est

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{|1 - \phi e^{-i\lambda}|^2} = \frac{\sigma^2}{2\pi} \frac{1}{1 + \phi^2 - 2\phi\cos\lambda}.$$

Si $\phi > 0$, le spectre est plus élevé près de $\lambda = 0$, ce qui correspond à des variations lentes. Si $\phi < 0$, il est plus élevé près de $\lambda = \pi$, ce qui correspond à une alternance plus rapide.

8.6 Cas pratique : le cycle des taches solaires

La série `sunspot.year`, fournie avec `R`, contient le nombre annuel moyen de taches solaires. Elle est un exemple classique de série présentant un comportement cyclique. Le cycle n'est pas parfaitement régulier, mais le périodogramme permet d'identifier une période dominante proche de 11 ans.

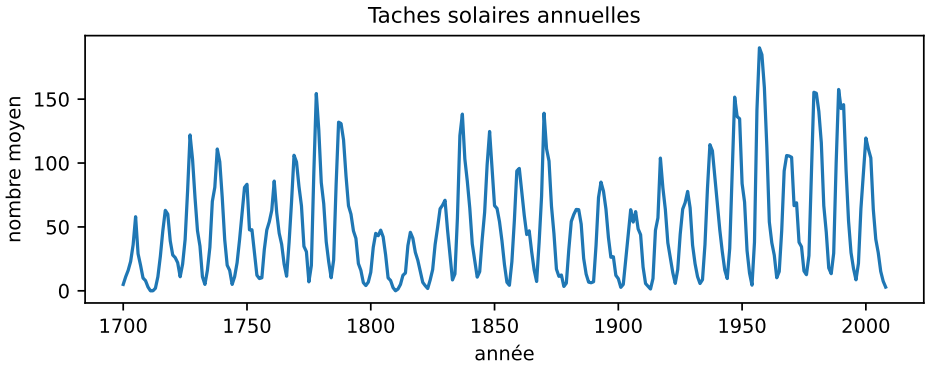


Figure 8.2 – Nombre annuel moyen de taches solaires. La série est irrégulière, mais des cycles d'environ une dizaine d'années sont visibles.

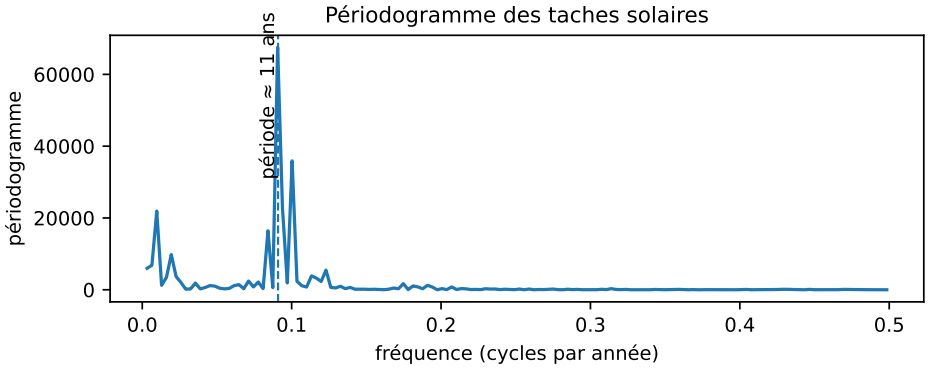


Figure 8.3 – Périodogramme de `sunspot.year`. Le pic proche de la fréquence 1/11 correspond à un cycle d'environ 11 ans.

Idée à retenir

Le chronogramme permet de voir qu'il existe des oscillations. Le périodogramme permet de dire plus précisément quelles périodes dominent. Ici, il ne remplace pas l'analyse visuelle : il la complète.

8.7 Variabilité changeante et rendements financiers

Les modèles ARMA et ARIMA décrivent surtout la dépendance de la moyenne conditionnelle. Après ajustement, on espère que les résidus ressemblent à un bruit blanc : pas de structure visible, pas d'autocorrélation importante, variance à peu près stable.

Mais dans de nombreuses séries financières, on observe autre chose. Les rendements eux-mêmes peuvent être faiblement autocorrélés, alors que leurs carrés ou leurs valeurs absolues sont fortement dépendants. Cela signifie que le signe des variations est difficile à prévoir, mais que l'intensité des variations peut être persistante.

Définition 8.7.1: Rendement logarithmique

Si (P_t) est une série de prix strictement positive, le rendement logarithmique est

$$R_t = \log(P_t) - \log(P_{t-1}).$$

On utilise souvent $100R_t$ pour exprimer ce rendement en pourcentage approximatif.

Idée à retenir

Une série de prix est souvent non stationnaire. Les rendements sont beaucoup plus proches d'une série stationnaire. Mais stationnarité ne signifie pas forcément variance constante conditionnellement au passé : les périodes agitées peuvent se regrouper.

La figure suivante montre une trajectoire simulée avec volatilité conditionnelle. Les valeurs positives et négatives alternent de manière difficilement prévisible, mais les grandes amplitudes apparaissent par paquets.

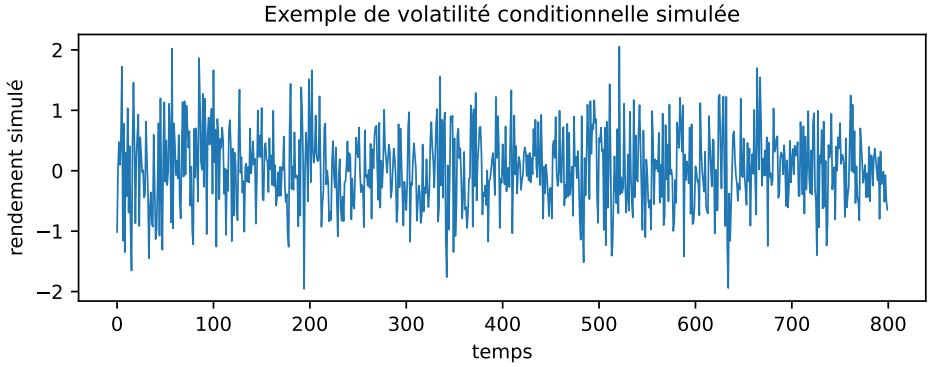


Figure 8.4 – Exemple simulé de rendements avec volatilité conditionnelle. Les périodes calmes et les périodes agitées tendent à se regrouper.

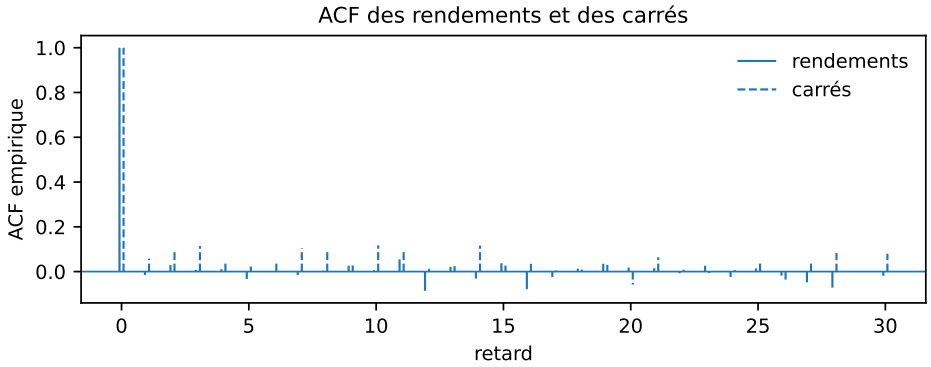


Figure 8.5 – Dans cet exemple, l'ACF des rendements est faible, tandis que l'ACF des carrés est positive sur plusieurs retards. C'est un signal typique d'hétéroscédasticité conditionnelle.

8.8 Modèles ARCH et GARCH

Les modèles ARCH et GARCH introduisent une idée différente de celle des ARMA. Dans un ARMA, le passé sert à prévoir la moyenne de X_t. Dans un modèle ARCH/GARCH, le passé sert surtout à prévoir la variance conditionnelle de X_t.

Définition 8.8.1: Hétéroscédasticité conditionnelle

On dit qu'une série centrée (X_t) présente une hétéroscédasticité conditionnelle lorsque

$$\mathbb{E}(X_t \mid \mathcal{F}_{t-1}) = 0,$$

mais

$$\mathrm{Var}(X_t \mid \mathcal{F}_{t-1})$$

dépend du passé $\mathcal{F}_{t-1}$.

Autrement dit, le modèle ne prévoit pas directement si la prochaine variation sera positive ou négative, mais il peut prévoir si elle risque d'être petite ou grande.

Définition 8.8.2: Modèle ARCH(1)

Soit (Z_t) un bruit blanc très fort, centré, réduit. Un processus (X_t) suit un modèle ARCH(1) si

$$X_t = \sigma_t Z_t, \qquad \sigma_t^2 = \omega + \alpha X_{t-1}^2,$$

où $\omega > 0$ et $\alpha \geq 0$.

Si X_{t-1}^2 est grand, alors σ_t^2 est grand : une grande variation récente augmente la variance conditionnelle de la prochaine variation. C'est le mécanisme de regroupement de volatilité.

Proposition 8.8.3: Moments conditionnels d'un ARCH(1)

Dans le modèle ARCH(1), on a

$$\mathbb{E}(X_t \mid \mathcal{F}_{t-1}) = 0, \qquad \mathrm{Var}(X_t \mid \mathcal{F}_{t-1}) = \sigma_t^2.$$

Si le processus est stationnaire du second ordre et si $\alpha < 1$, alors

$$\mathrm{Var}(X_t) = \frac{\omega}{1-\alpha}.$$

Démonstration. Comme Z_t est centré, réduit et indépendant du passé,

$$\mathbb{E}(X_t \mid \mathcal{F}_{t-1}) = \sigma_t \mathbb{E}(Z_t \mid \mathcal{F}_{t-1}) = 0,$$

et

$$\mathrm{Var}(X_t \mid \mathcal{F}_{t-1}) = \mathbb{E}(X_t^2 \mid \mathcal{F}_{t-1}) = \sigma_t^2 \mathbb{E}(Z_t^2) = \sigma_t^2.$$

Si $v = \mathbb{E}(X_t^2)$ ne dépend pas de t, alors

$$v = \mathbb{E}(\sigma_t^2) = \omega + \alpha \mathbb{E}(X_{t-1}^2) = \omega + \alpha v.$$

Donc $v(1-\alpha) = \omega$, ce qui donne la formule annoncée. □

Le modèle ARCH(1) est déjà très instructif, mais il est souvent trop court en mémoire. Dans les données financières, une période de forte volatilité peut durer longtemps. Le modèle GARCH ajoute donc une dépendance de la variance conditionnelle à sa propre valeur passée.

Définition 8.8.4: Modèle GARCH(1,1)

Un processus (X_t) suit un modèle GARCH(1,1) si

$$X_t = \sigma_t Z_t, \qquad \sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2,$$

où $\omega > 0$, $\alpha \geq 0$ et $\beta \geq 0$.

Remarque 8.8.5: Persistance de la volatilité

Dans un GARCH(1,1), la quantité $\alpha + \beta$ mesure la persistance de la volatilité. Si $\alpha + \beta$ est proche de 1, un choc de volatilité disparaît lentement. Si $\alpha + \beta$ est nettement plus petit que 1, la variance conditionnelle revient plus vite vers son niveau moyen.

8.9 Synthèse du chapitre

À retenir :

1) L'analyse spectrale étudie une série selon ses fréquences plutôt que seulement selon le temps.
2) Le périodogramme est un outil exploratoire pour repérer des périodes dominantes.
3) La densité spectrale est l'équivalent théorique du périodogramme pour un processus stationnaire.
4) Un bruit blanc a un spectre plat : aucune fréquence ne domine.
5) Les séries financières peuvent avoir des rendements peu autocorrélés mais des carrés autocorrélés.
6) Les modèles ARCH/GARCH servent à modéliser une variance conditionnelle qui dépend du passé.

Ce chapitre clôt le parcours principal du cours. Les outils précédents ne donnent pas une recette automatique, mais une méthode de travail : observer, transformer, modéliser, diagnostiquer, prévoir, puis vérifier si le modèle choisi répond vraiment à la question posée par les données.