



الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire

وزارة التعليم العالي والبحث العلمي
Ministère de L'Enseignement Supérieur et de la Recherche Scientifique

جامعة سعد دحلب البليدة-1-
Université Saad Dahlab Blida -1-



Mémoire de fin D'études

En vue de l'obtention du diplôme Master

Faculté de sciences
Département : Mathématiques

Spécialité : Modélisation stochastique et statistique

THEME

Modélisation et prévision par des modèles temporels,
application sur COVID19 en Algérie

Présenté par :
LAIB Mohammed
Devant le Jury :

Soutenu le : 26 / 09 /2020

Président : O.TAMI MAA Université de Blida 1

Promoteur : A .RASSOUL MCA ENSH de Blida

Examineur : R.FRIHI MAA Université de Blida 1

Promotion : 2019/2020

Dédicace

Je dédie ce modeste travail à

A mes parents qui peuvent être fiers et trouver ici le résultat de nombreuses années de sacrifice. Merci pour les nobles valeurs, l'éducation et le soutien continu qui sont venus de vous.

À mon cher frère et sœur Qui n'ont cessé d'être pour moi des exemples de persévérance, de courage et de Générosité. Dieu vous garde pour moi.

A mon encadreur RASOUL Abdelaziz, en espérant qu'il trouve dans ce travail le témoignage de ma profonde gratitude.

A tout mes enseignants de l'université de Saad Dahlab Blida1 sans exception.

A mes amis et camarades de classe.

REMERCIEMENTS

La chose la plus importante dans la vie d'un homme est d'être reconnaissant et d'être son plus grand atout

Il est dommage que cela ne soit utile à personne. Ce n'est pas mon habitude d'être ingrat. Il utilise

Je suis donc heureux de rendre mes premiers remerciements à Dieu Tout-Puissant, qu'il m'a donné la volonté et le courage de faire ce travail humble.

Je voudrais exprimer ma profonde gratitude au promoteur du RASOUL Abdelaziz, qui a été heureux de travailler avec lui sous sa supervision pour ses conseils et ses critiques constructives.

Je remercie sincèrement les membres du jury d'avoir accepté de faire partie du comité d'examinations.

Au final, je ne pourrai pas terminer cette partie sans exprimer ma gratitude mes parents, frères et sœurs qui m'ont toujours soutenu et encouragé pendant mes études.

Laib

ملخص

الهدف من هذا العمل هو دراسة تقنيات تحليل السلاسل الزمنية و نمذجتها من

اجل التنبؤ بالقيم المستقبلية

وقد قمنا في هذا العمل بدراسة نماذج أريما باستخدام تقنيات بوكس – جنكيس ومرشح كالمان من اجل التنبؤ بالقيم المستقبلية .

وقد طبقنا في نهاية العمل هذه التقنيات على معطيات حقيقية تتمثل في دراسة عدد الاصابات وعدد الوفيات بفيروس كورونا 19 في الجزائر باستعانة ببرنامج الاحصائي ار.

الكلمات المفتاحية : سلسلة زمنية ، بوكس _ جنكيز ، مرشح كالمان ، أرما ، أريما .

Résumé

L’objectif de ce travail est d’étudier les techniques d’analyse et de modélisation des séries chronologiques afin de prédire les valeurs futures.

Dans ce travail, nous avons étudié les modèles ARIMA en utilisant les techniques Box-Jenkins et le filtre de Kalman afin de prédire les valeurs futures.

A la fin des travaux, nous avons appliqué ces techniques à des données réelles, représentées par l’étude du nombre d’infections et du nombre de décès par Coronavirus 19 en Algérie, à l’aide du programme statistique R.

Mots clés : Série temporelle, Box-Jenkins, Filtre de Kalman , ARMA, ARIMA.

Summary

The objective of this work is to study the techniques of analysis and modeling of time series in order to predict future values.

In this work, we studied ARIMA models using Box techniques and the Kalman filter in order to predict future values.

At the end of the work, we applied these techniques to real data, represented by studying the number of infections and the number of deaths with Coronavirus 19 in Algeria, using the statistical program R.

Key words : Time Series, Box, Kalman Filter, ARMA, ARIMA

TABLE DES MATIÈRES

1 Généralités sur les séries chronologiques	2
1.1 Rappel sur les variables aléatoires	2
1.2 Généralités sur les séries chronologiques	3
1.2.1 Schémas de décomposition d'une série chronologique	4
1.2.2 Procédure de choix d'un schéma de décomposition	4
1.2.2.1 Plusieurs méthodes pour estimer la tendance	4
1.2.2.2 Méthode de calcul des coefficients saisonniers	5
1.2.2.3 désaisonnalisée la série	6
1.3 Définitions	6
1.3.1 Opérateur de retard	6
1.3.2 Opérateur différence	6
1.3.3 Opérateur différence saisonnière	7
1.4 Stationnarité	7
1.4.1 Fonction d'autocorrélation d'une série stationnaire	7
1.4.2 Fonction d'autocorrélation théorique.	8
1.4.3 Fonction d'autocorrélation empirique	8
1.4.4 La fonction d'autocorrélation partielle	9
1.4.5 Bruit blanc	11
1.4.6 Test de blancheur : le test du portemanteau.	12
1.4.7 Test de Durbin-Watson.	13
1.4.8 Série linéaire	13
1.4.9 Modèles autorégressifs, moyennes mobiles	14
1.4.9.1 Processus Auto Régressif d'ordre p , AR(p)	14
1.4.10 Processus Moyen mobile d'ordre q , MA(q)	15
1.4.10.1 Propriétés	16
1.4.11 Processus ARMA(p,q)	16
1.4.11.1 Fonction d'autocorrélation d'un AR	17

1.4.11.2	Fonction d'autocorrélation d'un MA	18
1.4.11.3	Inversibilité des modèles MA (q)	19
1.4.11.4	Processus ARMA(p,q) (Autoregressive-Moving Average)	21
1.4.11.5	ACF pour les modèles ARMA (p, q)	22
1.4.12	Processus autorégressif intégré à moyenne mobile ARIMA	23
1.4.13	Modèle saisonnière ARIMA (SARIMA)	24
1.5	Les processus aléatoires non stationnaires	25
1.5.1	Test de non-stationnarité : test Dickey-Fuller	27
1.5.2	Test de Dickey-Fuller augmenté	28
1.6	La méthodologie de Box et Jenkins	29
1.7	Critère d'information	30
1.8	Estimation des paramètres et validation du modèle	30
1.8.1	Estimation des paramètres	30
1.8.2	Vérification et validation	30
1.9	Prévision	32
2	Prévision par les séries chronologiques	33
2.1	Les modèles d'état	33
2.2	Filtre de Kalman	34
2.2.1	Equations du filtre	35
2.2.2	Le filtre d'information	37
2.3	Modèle d'état pour un processus ARMA	39
2.3.1	Modèle d'état pour un AR(p)	39
2.3.2	Modèle d'état pour un ARMA(p,q)	40
2.4	Identification "robuste" d'un processus ARMA	41
2.4.1	Cas d'un MA : l'algorithme des innovations	41
2.4.2	L'algorithme de Durbin-Levinson	42
2.4.3	Algorithme de Yule Walker	43
2.4.4	Cas général d'un ARMA(p, q)	43
2.4.5	Détails pour la prédiction	44
2.4.6	Identification paramétrique d'un processus ARMA	45
2.4.7	Validation d'un modèle : critères de contrôle sur le résidu	46
2.4.8	Méthodes de sélection	46
3	Application sur la prévision du COVID19	48
3.1	Analyse de COVID19 en algérie	49
3.1.1	La prédiction sur le nombre d'infections par COVID19	49
3.2	Test de tendance Mann-Kendall dans R	55
3.3	Prévision par la méthode de Box-Jenkins	57
3.3.1	Identification du modèle	57

3.3.1.1	Estimation des paramètres	57
3.3.2	Validation	58
3.3.3	Prévision	59
3.4	Prévision par filtre de Kalman	60
3.4.1	Représentation graphique de la prévision	62
3.5	Prévision du nombre de décès dus à le virus COVID19 en Algérie	63
3.5.1	Résultats de la Statistique descriptive	63
3.5.2	Analyse de la série temporelle	64
3.5.3	Application de la méthode de Box-Jenkins	68
3.5.3.1	Estimation des paramètres	68
3.5.3.2	Validation	70
3.5.3.3	Prévision	70
3.5.4	La prévision par filter de Kalman	71
3.5.4.1	Représentation graphique	74
3.5.5	Comparaison entre prévision par la méthode Box et filtre de Kalman	75

TABLE DES FIGURES

2.1	Filtre de Kalman	39
2.2	Schéma de prédiction par filtre de Kalman	47
3.1	Histogramme des observations des infectés	49
3.2	Représentation graphique de la série cas de d'infection par COVID19	50
3.3	Représentation graphique d'autocorrélation de la série cas de d'infection par COVID19	51
3.4	Représentation graphique d'autocorrélation partielle de la série obse	51
3.5	les composantes d'une série temporelle	52
3.6	Représentation graphique de la série obse différenciation	53
3.7	Autorrélation simple de la série obse différencier	54
3.8	Autorrélation partielle de la série différencier	55
3.9	Analyse de la tendance de la séries des infectés	56
3.10	Représentation graphique de la prévision par Box-Jenkins	60
3.11	Représentation graphique des observations	61
3.12	Représentation graphique des observations	61
3.13	Série chronologique ajustée par Filtre de Kalman	62
3.14	Représentation graphique de la prévision par filtre de Kalman	63
3.15	Histogramme de la série des décès en Algérie par COVID-19	64
3.16	Représentation graphique de la série des décès en Algérie par COVID-19	64
3.17	Autocorrélation simple ACF de la série des décès en Algérie par COVID-19	65
3.18	Autocorrélation partielle PACF de la série des décès en Algérie par COVID- 19	65
3.19	Décomposition de la série des décès en Algérie par COVID-19	66
3.20	Représentation graphique de la série différencier	67
3.21	Graphe de autorrélation simple de la série y différencier	68
3.22	Graphe de autorrélation partielle de la série y différencier	68

3.23	Repésentation graphique de la prévision des décès	71
3.24	Repésentation graphique de la série des décès	71
3.25	Série chronologique filtrée par Kalman	72
3.26	Série chronologique ajustée par Kalman	73
3.27	Représentation graphique de prévision des décès par filtre de Kalman . . .	75

LISTE DES TABLEAUX

3.1	Discription de la séries des cas d'infections en algérie	49
3.2	Résultats d'ajustements de la série des cas d'infections par COVID19 . . .	57
3.3	Résultats d'ajustements de la série des cas d'infections par COVID19 . . .	58
3.4	Estimation des coefficients du modèle d'ajustement	58
3.5	Résultats du prévision par la méthode de Box-Jenkins	59
3.6	Résultats du prévision pour dix jours avec les bornes de confiances	62
3.7	Résultats discriptive des décès à cause de COVID-19 en Algérie	63
3.8	Résultats d'ajustement de la série des décès par un modèle temporelle . .	69
3.9	Résultats de la prévision des décès par la méthode de Box-Jenkins	70
3.10	Résultats de la prévision des décès par la méthode de Filtre de Kalman . .	73
3.11	Bornes de confiances de la prévision des décès par la méthode de Filtre de Kalman	74
3.12	Résultats de la prévision des décès par la méthode de Filtre de Kalman . .	74
3.13	Comparaison de la prévision des décès par les deux méthodes avec le bilan réel	75
3.14	Comparaison de la prévision des décès par les deux méthodes avec le bilan réel	76

INTRODUCTION GÉNÉRALE

Connaître le futur, ou du moins avoir une idée du futur est l'un des soucis de l'Homme depuis toujours. De nos jours aussi, les raisons socioéconomiques poussent à anticiper l'avenir. Une question importante est de savoir sur quoi nous appuyer pour prédire l'avenir. Il est donc primordial d'arriver à prévoir le mieux possible le futur en s'appuyant sur le passé. D'une façon mathématique, on peut formuler le problème de la prévision en supposant avoir N observations (x_1, x_2, \dots, x_N) issues d'un processus (un ensemble des variables aléatoires) quantifiant une certaine activité, dans notre cas l'évolution du COVID19 et on souhaite connaître la valeur à une date future.

Dans le premier chapitre, nous présentons les définitions des séries chronologiques, la décomposition en principales composantes, et nous avons étudié et présenté les modèles des séries chronologiques, tels que : AR, MA, ARMA, ARIMA de toutes sortes, et à la fin de cet axe nous nous sommes familiarisés avec la méthode de prédiction utilisant la technologie BOX-JENKINS qui est centrée sur l'identification, l'estimation et la validation.

Dans le deuxième chapitre, nous avons traité de la prédiction de l'utilisation de filtre de KALMAN dont le principe est de corriger le chemin du modèle en combinant des observations et des informations fournies par le modèle pour réduire l'erreur entre l'état vrai et l'état filtré.

Dans le troisième nous avons appliqué ce que nous avons étudié dans le premier et le deuxième axe aux données qui représentent le nombre d'infections et de décès par COVID19 en Algérie, et à la fin de cet axe nous avons comparé entre la technologie BOX-JENKINS et la technique de KALMAN, et les résultats de la prédiction à l'aide de KALMAN étaient meilleurs que les résultats de la prédiction de BOX-JENKINS.

CHAPITRE 1

GÉNÉRALITÉS SUR LES SÉRIES CHRONOLOGIQUES

1.1 Rappel sur les variables aléatoires

Définition 1.1 L'espérance mathématique est la notion des probabilités qui correspond à la notion de moyenne en statistique descriptive. Formellement elle peut être définie comme somme des produits des valeurs d'une variable aléatoire par leur probabilité, ou une intégrale de produit de la variable par sa densité.

— Si X est une v.a. absolument continue de densité $f : \mathbb{R} \rightarrow \mathbb{R}$, la condition d'existence de l'espérance étant tout simplement la convergence absolue de cette intégrale généralisée, ce qui vu la positivité de f , se traduit par

$$\int_{-\infty}^{+\infty} |x| f(x) dx < +\infty,$$

donc

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

est l'espérance de X .

— si X est une v.a. discrète. $X(\omega)$ est un ensemble fini ou dénombrable. $\sum_{x \in X(\omega)} |x| P(X=x) < \infty$, dans ce cas la série est finie donc elle converge. Alors

$$E(X) = \sum_{x \in X(\omega)} x P(X=x)$$

est l'espérance de X .

Proposition 1.2 Soient X et Y deux v.a. intégrables, a et b deux réels. Alors les v.a. $X+Y$ et $aX+bY$ sont intégrables et

$$E(X+Y) = E(X) + E(Y), \quad \text{et} \quad E(aX+bY) = aE(X) + bE(Y).$$

Définition 1.3 La covariance entre les deux variables aléatoires X et Y est :

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

Définition 1.4 La corrélation entre les deux variables aléatoires X et Y est :

$$\text{cor}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y}.$$

Définition 1.5 la méthode des moindres carrés permet alors de minimiser l'impact des erreurs expérimentales en « ajoutant de l'information » dans le processus de mesure

Définition 1.6 Un processus stochastique est une suite de variables aléatoires réelles qui sont indexées par le temps : $X_t, t \in \mathbb{Z}$ Ici t appartient à un espace discret, ce qui définit un processus en temps discret. Un processus stochastique est donc une famille de variables aléatoires X dont on va observer des valeurs réelles issues de l'espace S des échantillons selon une certaine loi de probabilité. Pour chaque point s de l'espace des échantillons S , la fonction qui associe $X_t(s)$ est appelée la trajectoire du processus.

1.2 Généralités sur les séries chronologiques

Définition 1.7 On appelle série chronologique (série temporelle) une série statistique à deux variables (t, Y_t) avec $t \in T$, $T = \{t_1, t_2, t_3, \dots, t_n\}$ où la première composante du couple t est le temps et la deuxième composante est une variable aléatoire numérique Y_t prenant ses valeurs aux instants t . Suivant la nature du problème étudié la chronique peut être journalière (cours d'une action en bourse), mensuelle (consommation mensuelle de gaz), trimestrielle (nombre trimestriel de chômeurs), annuelle (chiffre annuel des bénéfices des exportations) etc.

L'étude des séries chronologiques sert à faire de la prévision à court, moyen et long terme. Il existe des méthodes prévisionnelles quantitatives et qualitatives.

Définition 1.8 La tendance ou trend m_t capte l'orientation à long terme de la série.

Définition 1.9 La composante saisonnière s_t capte un comportement qui se répète avec une certaine périodicité (toutes les 12 périodes pour des données mensuelles, toutes les 7 périodes pour des données quotidiennes..).

Définition 1.10 Composante irrégulière est une composante d'erreur, u_t . Idéalement, elle est de faible variabilité par rapport aux autres composantes.

A ces trois composantes, on ajoute parfois un cycle.

Définition 1.11 On appelle cycle un comportement qui se répète assez régulièrement mais avec une périodicité inconnue et changeante.

1.2.1 Schémas de décomposition d'une série chronologique

La technique de décomposition d'une série chronologique, repose sur un modèle qui l'autorise, ce modèle porte le nom de schéma de décomposition. Il en existe essentiellement deux grands types :

- **Schéma additif** : Dans un modèle additif, on suppose que les 3 composantes : tendance, variations saisonnières et variations accidentelles sont indépendantes les unes des autres. On considère que la série Y_t s'écrit comme la somme de ces 3 composantes :

$$Y_t = m_t + s_t + u_t . \quad (1.1)$$

- **Schéma multiplicatif** : On suppose que les variations saisonnières et les variations accidentelles dépendent de la tendance et on considère que Y_t s'écrit de la manière suivante :

$$Y_t = m_t \times s_t \times u_t . \quad (1.2)$$

1.2.2 Procédure de choix d'un schéma de décomposition

- **La méthode de la bande** : La procédure de la bande consiste à partir de l'examen visuel du graphique de l'évolution de la série brute à relier, par une ligne brisée, toutes les valeurs « hautes » et toutes les valeurs « basses » de la chronique. Si les deux lignes sont parallèles, la décomposition de la chronique peut se faire selon un schéma additif; dans le cas contraire le schéma multiplicatif semble plus adapté.
- **Le test de Buys-Ballot** : Nous calculons, pour chacune des années, la moyenne et l'écart type, puis nous estimons par MCO (Moindre Carrés Ordinaires) les paramètres α_1 et α_2 de l'équation si $\sigma_i = \alpha_1 y_i + \alpha_2 + \varepsilon_i$. Dans le cas, où le paramètre α_1 n'est pas significativement différent de 0 (test de Student) alors nous acceptons l'hypothèse d'un schéma additif; dans le cas contraire, nous retenons un schéma multiplicatif.

1.2.2.1 Plusieurs méthodes pour estimer la tendance

- Définir la tendance par la régression linéaire (ou régression plus complexe) de $y(t)$ en fonction du temps.
- **Avantage** : simple et rapide à calculer. Dûs aux effets saisonniers, l'ajustement par une droite (ou autre) n'est pas toujours adéquat ($R^2 \ll 1$).
- Utiliser une méthode plus locale appelée méthode des moyennes mobiles que l'on combine le plus souvent à une régression linéaire (des moyennes mobiles en fonction du temps).

- L'estimation de la tendance ne dépend pas du modèle (additif, multiplicatif, . . .).

Méthode moyennes mobiles :

est utilisé pour éliminer les fluctuations à court terme et identifier la tendance à long terme dans la série originale (pour créer une série lisse afin de réduire les fluctuations aléatoires dans la série originale et estimer la composante tendance-cycle).

La moyenne mobile d'ordre k à l'instant t correspond à la moyenne pondérée de $k+1$, si k est pair, ou k , si k est impair, observations centrées autour de l'instant k . Cette série temporelle est notée $M_k(t)$

- si k est impair :

$$M_k(t) = \frac{1}{k}(Y_{t-1} + Y_t + Y_{t+1}) \quad (1.3)$$

- si k est pair :

$$M_k(t) = \frac{1}{k}\left(\frac{Y_{t-2}}{2} + Y_{t-1} + Y_t + Y_{t+1} + \frac{Y_{t+2}}{2}\right) \quad (1.4)$$

1.2.2.2 Méthode de calcul des coefficients saisonniers

La série Y_t est observée sur n année par période « p ». $p = 12$ mois ($j = 1, 2, \dots, 12$) ou 4 trimestres ($j = 1, 2, 3$ ou 4).

Les variations saisonnières S_j

— **cas : Modèle additif.**

- On calcule les données sans tendance $Y_t - m_t$.
- On calcule la moyenne des données sans tendance du mois j sur les n années, ceci pour chacun des p période. D'où

$$S_j = \frac{1}{n} \sum_{i=1}^n (Y_{ij} - m_{ij}). \quad \forall j = 1, 2, \dots, p. \quad \forall i = 1, 2, \dots, n.$$

- On calcule la moyenne des coefficients saisonniers (S_j)

$$\bar{S} = \frac{1}{p} \sum_{j=1}^p S_j.$$

- Si $\bar{S} \neq 0$ on corrige les coefficient saisonniers (CS_j) : $CS_j = S_j - \bar{S}$

cas : Modèle multiplicatif

- On calcule les données sans tendance $\frac{Y_t}{m_t}$.
- On calcule la moyenne des données sans tendance du mois j sur les n années, ceci pour chacun des p période

$$S_j = \frac{1}{n} \sum_{i=1}^n \frac{Y_{ij}}{m_{ij}}. \quad \forall j = 1, 2, \dots, p. \quad \forall i = 1, 2, \dots, n.$$

— On calcule la moyenne des coefficients saisonniers (S_j)

$$\bar{S} = \frac{1}{p} \sum_{j=1}^p S_j$$

— Si $\bar{S} \neq 1$ on corrige les coefficient saisonniers (CS_j)

$$CS_j = \frac{S_j}{\bar{S}}$$

1.2.2.3 désaisonnalisée la série

Modèle multiplicatif : $Y^* = \frac{Y_{ij}}{CS_{ij}}$

Modèle additif : $Y^* = Y_{ij} - CS_{ij}$.

1.3 Définitions

Définition 1.12 *La colinéarité exacte survient quand une combinaison linéaire des variables explicatives est égale à une autre variable explicative.*

Exemple 1.13 $X_1 = \alpha X_2 + \beta X_3 \quad \forall \alpha, \beta \in \mathbb{R}$.

1.3.1 Opérateur de retard

La manipulation pratique ou théorique des séries temporelles se trouve considérablement simplifiée par l'usage de l'opérateur retard (Lag operator). On donne ici ses propriétés élémentaires.

Opérateur retard. On note indifféremment B (backwards) ou L (lag), l'opérateur qui fait passer de X_t à X_{t-1} : $BX_t = X_{t-1}$, on a :

$$B^2 X_t = B(BX_t) = BX_{t-1} = X_{t-2}.$$

1.3.2 Opérateur différence

La différence première est :

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

parfois le premier opérateur de différence, utilisé pour supprimer la tendance de la série.

La différence seconde est :

$$\begin{aligned}\Delta^2 X_t &= \Delta(\Delta X_t) = \Delta((1 - B)X_t) = (1 - B)^2 X_t \\ &= (1 - 2B + B^2)X_t = X_t - 2X_{t-1} + X_{t-2}.\end{aligned}$$

1.3.3 Opérateur différence saisonnière

Etant donné une série mensuelle, il peut être important d'en examiner les accroissements d'une année sur l'autre (janvier sur janvier..).

L'opérateur différence saisonnière $\Delta_{12} = 1 - B^{12}$ est utile dans ce cas.

$$\Delta_{12} X_t = (1 - B^{12})X_t = X_t - X_{t-12}.$$

1.4 Stationnarité

Définition 1.14 Une série temporelle $\{Y_t\}$, ou processus stochastique, est dite strictement stationnaire si la distribution conjointe de $(Y_{t_1}, \dots, Y_{t_k})$ est identique à celle de $(Y_{t_1+t}, \dots, Y_{t_k+t})$, quels que soient k le nombre d'instants considérés, (t_1, \dots, t_k) les instants choisis et t , le décalage; c'est-à-dire que, quels que soient le nombre de dates et les dates choisis, quand on décale ces dates d'une même quantité, la distribution ne change pas. En somme, la stationnarité stricte dit que la distribution conjointe de tout sous-vecteur de $\{Y_t\}$, quels que soient sa longueur et les instants choisis, est invariante quand on translate ces instants d'une même quantité. Cette condition est difficile à vérifier et on utilise une version plus faible de stationnarité, la stationnarité faible ou du second ordre, souvent suffisante.

$\{Y_t\}$ est dite faiblement stationnaire si :

- $E(Y_t) = \mu$, constante indépendante de t .
- $cov(Y_t, Y_{t-l})$ ne dépend que de l entier et dans ce cas elle est notée :

$$\gamma_l = cov(Y_t, Y_{t-l}).$$

1.4.1 Fonction d'autocorrélation d'une série stationnaire

Définition 1.15 Soit $\{Y_t\}$ une série à valeurs réelles, stationnaire. La covariance $\gamma_l = cov(Y_t, Y_{t-l})$ est appelée autocovariance d'ordre (ou de décalage) l . La fonction : $l \rightarrow \gamma_l$, $l = \dots, -1, 0, 1, 2, \dots$ est la fonction d'autocovariance de $\{Y_t\}$.

Cette fonction vérifie notamment :

- $\gamma_0 = var(Y_t) \geq 0$.

- $|\gamma_l| \leq \gamma_0 \quad \forall l.$
- $\gamma_l = \gamma_{l-1} \quad \forall l.$

Cette fonction étant paire, on ne la représente que pour $l = 0, 1, 2, \dots$. On a également :

La fonction d'auto covariance d'une série $\{Y_t\}$ faiblement stationnaire est de type positif 1.

Cette propriété exprime le fait que la variance d'une combinaison linéaire de n v.a. Y_{t1}, \dots, Y_{tn} est positive.

1.4.2 Fonction d'autocorrélation théorique.

Définition 1.16 *Le coefficient d'autocorrélation d'ordre l est :*

$$\rho_l = \frac{\text{cov}(Y_t, Y_{t-l})}{\sqrt{\text{var}(Y_t) \text{var}(Y_{t-l})}} = \frac{\text{cov}(Y_t, Y_{t-l})}{\text{var}(Y_t)} = \frac{\gamma_l}{\gamma_0}.$$

La dernière égalité tient car $\text{var}(Y_{t-l}) = \text{var}(Y_t) = \gamma_0$. Enfin, en notant que par la stationnarité $E(Y_t) = \mu$, indépendant de t , on a en terme d'espérance mathématique :

$$\rho_l = \frac{E[(Y_t - \mu)(Y_{t-l} - \mu)]}{E[(Y_t - \mu)^2]} = \frac{\gamma_l}{\gamma_0}.$$

La fonction : $l \rightarrow \rho_l; l = 0, 1, 2, \dots$, est la fonction d'autocorrélation (théorique) de la série $\{Y_t\}$.

Nous utiliserons l'abréviation anglaise, ACF, qui est aussi celle des sorties de R, de préférence à FAC. On appelle son graphique corrélogramme. On voit que :

$$-1 \leq \rho_l \leq 1, \rho_0 = 1.$$

1.4.3 Fonction d'autocorrélation empirique

Etant donné une série observée $Y_t, t = 1, \dots, T$, notons $\bar{Y} = \sum_{t=1}^T \frac{Y_t}{T}$. L'autocovariance empirique d'ordre l est

$$\hat{\gamma}_l = \frac{\sum_{t=l+1}^T (Y_t - \bar{Y})(Y_{t-l} - \bar{Y})}{T}, \quad 0 \leq l \leq T-1.$$

Le coefficient d'autocorrélation empirique d'ordre l est

$$\hat{\rho}_l = \frac{\sum_{t=l+1}^T (Y_t - \bar{Y})(Y_{t-l} - \bar{Y})}{\sum_{t=1}^T (Y_t - \bar{Y})^2}. \quad 0 \leq l \leq T-1.$$

La fonction : $l \rightarrow \hat{\rho}_l, l = 0, 1, 2, \dots$, est la fonction d'autocorrélation empirique.

Remarque 1.17 Observons que le dénominateur dans l'équation de coefficient d'auto-corrélation empirique est T alors que le nombre de termes au numérateur dépend du décalage. Il faut se garder de corriger l'estimation en adoptant un dénominateur dépendant du nombre de termes dans la somme. En effet, avec un tel choix, la fonction d'autocovariance empirique $l \rightarrow \hat{\gamma}_l$ ne serait plus de type positif.

1.4.4 La fonction d'autocorrélation partielle

Définition 1.18 L'autocorrélation partielle mesure la corrélation entre deux variables aléatoires Y_t et Y_{t+h} à différents décalages h après en supprimant la dépendance linéaire de Y_{t+1} à Y_{t+h-1} . PACF représente donc la séquence des corrélations conditionnelles :

$$\phi_{h,h} = \text{corr}(Y_t, Y_{t+h} \mid Y_{t+1}, \dots, Y_{t+h-1}), h = 1, 2, \dots$$

- La fonction d'autocorrélation (ACF) entre deux variables Y_t et Y_{t+h} à différents décalages h ne s'ajustent pas pour l'influence des décalages intermédiaires : l'ACF représente donc la séquence des corrélations inconditionnelles.

$$\begin{aligned} \phi_{h,h} &= \text{corr}(Y_t, Y_{t+h} \mid Y_{t+1}, \dots, Y_{t+h-1}) \\ &= \frac{\text{cov}[(Y_t \mid Y_{t+1}, \dots, Y_{t+h-1}), (Y_{t+h} \mid Y_{t+1}, \dots, Y_{t+h-1})]}{\sqrt{\text{var}(Y_t \mid Y_{t+1}, \dots, Y_{t+h-1})} \sqrt{\text{var}(Y_{t+h} \mid Y_{t+1}, \dots, Y_{t+h-1})}} \\ &= \frac{\text{cov}[(Y_t - \bar{Y}_t), (Y_{t+h} - \bar{Y}_{t+h})]}{\sqrt{\text{var}(Y_t - \bar{Y}_t)} \sqrt{\text{var}(Y_{t+h} - \bar{Y}_{t+h})}} \end{aligned}$$

où

$$\begin{aligned} \hat{Y}_t &= \alpha_1 Y_{t+1} + \alpha_2 Y_{t+2} + \dots + \alpha_{h-1} Y_{t+h-1}; \\ \hat{Y}_{t+h} &= \beta_1 Y_{t+1} + \beta_2 Y_{t+2} + \dots + \beta_{h-1} Y_{t+h-1}, \end{aligned}$$

et α_i, β_i , ($1 \leq i \leq h-1$) sont la régression linéaire quadratique moyenne coefficients obtenus en minimisant les $E(Y_t - \hat{Y}_t)^2$ et $E(Y_{t+h} - \hat{Y}_{t+h})^2$ respectivement

Équations de Yule-Walker et PACF pour le processus AR (p)

Les équations de Yule-Walker peuvent être utilisées pour dériver les coefficients d'autocorrélation aux décalages $1, 2, \dots, h$ comme suit :

- Ajuster le modèle de régression, où la variable dépendante Y_t à partir d'un processus stationnaire moyen nul est régressé sur le h variables retardées $Y_{t-1}, Y_{t-2}, \dots, Y_{t-h}$. c'est à dire

$$Y_t = \phi_{h,1} Y_{t-1} + \phi_{h,2} Y_{t-2} + \dots + \phi_{h,h} Y_{t-h} + \varepsilon_t.$$

où ϕ_{hh} désigne le h-ième paramètre de régression et t est un terme d'erreur de moyenne 0 et non corrélé à Y_{t-h} , pour $h \neq 0$

- Multipliez cette équation par Y_{t-1} ; prendre les attentes et diviser les résultats par la variance de Y_t . Faites la même opération avec $Y_{t-2}, Y_{t-3}, \dots, Y_{t-h}$ successivement pour obtenir l'ensemble suivant des équations h-Yule-Walker.
- Les équations de Yule-Walker sont une technique qui peut être utilisée pour estimer les paramètres d'autorégression du modèle $AR(h)$,

$$Y_t = \sum_{i=1}^h \phi_i Y_{t-i} + \varepsilon_t$$

à partir des données.

$$\begin{aligned} \rho_1 &= \phi_{h,1} + \phi_{h,2}\rho_1 + \phi_{h,3}\rho_2 + \dots + \phi_{h,h}\rho_{h-1} \\ \rho_2 &= \phi_{h,1}\rho_1 + \phi_{h,2} + \phi_{h,3}\rho_1 + \dots + \phi_{h,h}\rho_{h-2} \\ &\cdot \\ \rho_h &= \phi_{h,1}\rho_{h-1} + \phi_{h,2}\rho_{h-2} + \phi_{h,3}\rho_{h-3} + \dots + \phi_{h,h} \end{aligned}$$

Qui peut être représenté sous forme de matrice comme : $AX = b$;

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \cdot \\ \cdot \\ \cdot \\ \rho_h \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{h-1} \\ \rho_1 & 1 & & & \rho_{h-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{h-3} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{h-1} & \rho_{h-2} & \dots & & 1 \end{pmatrix} \begin{pmatrix} \phi_{h,1} \\ \phi_{h,2} \\ \phi_{h,3} \\ \cdot \\ \cdot \\ \cdot \\ \phi_{h,h} \end{pmatrix}$$

$$\rho_h^* = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_1 \\ \rho_1 & 1 & & & \rho_2 \\ \rho_2 & \rho_1 & 1 & \dots & \rho_3 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{h-1} & \rho_{h-2} & \dots & & \rho_h \end{pmatrix}$$

où

$$\hat{\phi}_{hh} = \frac{|\rho_h^*|}{|\rho_h|}$$

$|\rho_h^*|$ = le déterminant de la matrice (ρ_h) dans laquelle on remplace la dernière co-

lonne par le vecteur (ρ_1, \dots, ρ_h) .

Remarque 1.19 — La fonction d'autocorrélation partielle $\phi_{h;h}$ est une fonction de la autocorrélations $\rho_1, \rho_2, \dots, \rho_h$. Donc

$$-1 \leq \phi_{hh} \leq 1, \forall h > 0.$$

— Si (ε_t) est un processus de bruit blanc, alors l'autocorrélation partielle fonction $\phi_{hh} = 0$ pour tout $h \neq 0$, alors que

$$\phi_{00} = \rho_0 = 1.$$

— Si le processus sous-jacent est $AR(p)$,

$$\phi_{h;h} = 0, \forall h > p,$$

donc le graphique du PACF doit montrer un seuil après le décalage p .

— Remplacement de ρ_h (autocorrélations de population) par $\hat{\rho}_h$ (échantillon autocorrélations) $\forall h$ donnera l'échantillon PQCF $\hat{\phi}_{h;h}$ (voir Méthode de récursivité Levinson-Durbin)

Méthode récursive de Levinson-Durbin

En pratique, l'échantillon PACF est obtenu par Levinson-Durbin méthode de récursivité commençant par $\hat{\phi}_{1;1} = \hat{\rho}_1$, comme suit :

$$\hat{\phi}_{h+1;h+1} = \frac{\hat{\rho}_{h+1} - \sum_{j=1}^h \hat{\phi}_{h;j} \hat{\rho}_{h+1-j}}{1 - \sum_{j=1}^h \hat{\phi}_{h;j} \hat{\rho}_j};$$

et

$$\hat{\phi}_{h+1,j} = \hat{\phi}_{h,j} - \hat{\phi}_{h+1,h+1} \hat{\phi}_{h,h+1}, \quad \text{pour } j = 1, 2, \dots, h.$$

1.4.5 Bruit blanc

Définition 1.20 Un bruit blanc $\{Z_t\}$ est une suite de v.a. non corrélées (mais pas nécessairement indépendantes) de moyenne nulle et de variance constante σ_z^2 .

C'est donc une série faiblement stationnaire. On note $Z_t \sim BB(0, \sigma_z^2)$.

Définition 1.21 Un bruit blanc gaussien $\{Z_t\}$ est une suite de v.a. i.i.d. $N(0, \sigma_z^2)$, on note : $Z_t \sim BBN(0, \sigma_z^2)$.

Un bruit blanc gaussien est une série strictement stationnaire. Examinons ce que deviennent les coefficients d'autocorrélations empiriques quand ils sont calculés sur une série dont tous les coefficients d'autocorrélations théoriques sont nuls.

Proposition 1.22 Si $y_t, t = 1, \dots, T$ est une observation d'une suite de v.a. i.i.d. de moment d'ordre 2 fini, $E(y_t^2) < \infty$, alors les ρ_l sont approximativement indépendants et normalement distribués de moyenne 0 et de variance $1/T$.

1.4.6 Test de blancheur : le test du portemanteau.

Soit la série observée $Y_t, t = 1, \dots, T$, considérons la statistique : $Q(h) = T \sum_{j=1}^h \hat{\rho}_j^2$

où h est un décalage choisi par l'utilisateur et ρ_j l'estimation du coefficient d'autocorrélation d'ordre j de la série Y_t . $Q(h)$ est appelée statistique de BoxPierce. Elle permet de tester :

$$H_0^h : \rho_1 = \rho_2 = \dots = \rho_h = 0$$

H_1^h : au moins un des $\rho_1, \rho_2, \dots, \rho_h$ est non nul.

$Q(h)$ est la distance du χ^2 du vecteur $(\rho_1, \rho_2, \dots, \rho_h)$ au vecteur $(0, 0, \dots, 0)$ et on rejette l'hypothèse H_0^h pour les grandes valeurs de $Q(h)$.

En effet, sous l'hypothèse que $\{Y_t\}$ est une suite de v.a. i.i.d. et vu la propriété Le coefficient d'autocorrélation, $Q(h)$ n'est autre que

$$Q(h) = T \sum_{j=1}^h \hat{\rho}_j^2 = \sum_{j=1}^h \left(\frac{\hat{\rho}_j - 0}{1/\sqrt{T}} \right)^2$$

c'est-à-dire la somme des carrés de h variables approximativement $N(0, 1)$. Or, sachant que le carré d'une variable $N(0, 1)$ suit une loi χ_1^2 et que la somme de deux v.a. indépendantes et distribuées suivant des lois $\chi_{n_1}^2$ et $\chi_{n_2}^2$ suit une loi $\chi_{n_1+n_2}^2$, la loi de $Q(h)$ est bien approximativement χ_h^2 , sous l'hypothèse nulle. Notons qu'on doit choisir h , le nombre de coefficients dont on teste la nullité

Remarques (Variantes du test de blancheur)

♣ Pour des petits échantillons on utilise la statistique de Ljung-Box :

$$Q^*(h) = T(T+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{T-k}$$

Elle a une distribution de probabilité mieux approchée par un χ^2 que la statistique de Box-Pierce.

♣ Quand le test est appliqué non sur des v.a. indépendantes, mais sur les résidus d'un ajustement estimant m paramètres, la loi approchée sous l'hypothèse nulle est un χ^2 à $h - m$ degrés de liberté.

1.4.7 Test de Durbin-Watson.

Le test de Durbin-Watson est un test d'absence d'autocorrélation d'ordre 1 sur le résidu d'une régression linéaire. Il s'intéresse à la situation

$$y_t = x_t' \beta + u_t, t = 1, \dots, T, \quad u_t = \rho u_{t-1} + z_t$$

où x_t est un vecteur de $p + 1$ variables explicatives (dont la constante), $z_t \sim BB$.

Il teste $H_0 : \rho = 0$. La statistique de Durbin-Watson est

$$DW = \frac{\sum_{t=2}^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^T \hat{u}_t^2},$$

où u_t est le résidu de l'ajustement par moindres carrés ordinaires de y_t sur x_t . En développant numérateur et dénominateur, on voit que

$$DW \simeq 2(1 - \hat{\rho})$$

où

$$\hat{\rho} = \frac{\sum_{t=2}^T \hat{u}_{t-1} \hat{u}_t}{\sum_{t=1}^T \hat{u}_t^2}.$$

Les valeurs de DW proches de 0 indiquent une autocorrélation proche de 1.

Pour le test de : $H_0 : \rho = 0$ contre $H_1 : \rho > 0$, la région critique correspond à de faibles valeurs de DW (DW sensiblement inférieur à 2) et pour $H_1 : \rho < 0$, la région critique correspond à de fortes valeurs de DW

Remarque 1.23 *Pratiquement une statistique DW 2 peut être le signe d'une mauvaise spécification du modèle (par exemple, ajustement d'une tendance linéaire alors que la tendance réelle est quadratique).*

1.4.8 Série linéaire

Définition 1.24 *Une série $\{Y_t\}$ est dite linéaire si elle peut s'écrire :*

$$Y_t = \mu + \sum_{i=-\infty}^{+\infty} \psi_i z_{t-i},$$

où

$$z_t \sim BB(0, \sigma^2)$$

$\psi_0 = 1$, et la suite $\{\psi_i\}$ est absolument sommable, c'est-à-dire $\sum_i |\psi_i| < \infty$.

Définition 1.25 *Une série $\{Y_t\}$ est dite linéaire et causale si elle est linéaire avec $\psi_i = 0$,*

$i < 0$:

$$Y_t = \mu + \sum_{i=0}^{\infty} \psi_i z_{t-i}.$$

On admettra qu'une série linéaire est stationnaire. L'étude des séries non causales conduit à des résultats non intuitifs difficilement utilisables, aussi nous ne considérons parmi les séries linéaires que des séries causales. L'écriture

$$Y_t = \mu + \sum_{i=0}^{\infty} \psi_i z_{t-i},$$

comme somme de v.a. non corrélées permet d'obtenir facilement :

$$E(y_t) = \mu, \text{var}(y_t) = \sigma_z^2 \left(1 + \sum_{i=1}^{\infty} \psi_i^2\right)$$

$$\gamma_l = \text{cov}(y_t, y_{t-l}) = \sigma_z^2 \sum_{i=1}^{\infty} \psi_i \psi_{i-l}$$

1.4.9 Modèles autorégressifs, moyennes mobiles

1.4.9.1 Processus Auto Régressif d'ordre p, AR(p)

Dans le processus autorégressive d'ordre p , l'observation présente Y_t est générée par une moyenne pondérée des observations passées jusqu'à la p -ième période sous la forme suivante :

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

Où ε_t est un bruit blanc de moyenne nulle et de variance σ_ε^2 , avec $\phi_p \neq 0$

— En introduisant l'opérateur retard B : $\phi_p(B)Y_t = c + \varepsilon_t$. dans ce cas

$$E(Y_t) = \mu = \frac{c}{1 - \phi_1 - \phi_2 - \dots - \phi_p}$$

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (1.5)$$

polynome de degré p .

— modèle $AR(p)$ est stationnaire ssi les racines en valeur absolue de $\phi_p(B) = 0 > 1$

Proposition 1.26 *Le processus autorégressif d'ordre p admet une représentation MA(∞) si elle stationnaire*

Représentation de AR(P) par MA(∞)

$$\phi_p(B)Y_t = \varepsilon_t, \quad \phi_p(B) = 1 - \sum_{i=1}^p \phi_i B^i, \quad \text{et } \varepsilon_t \sim BB(0, \sigma^2)$$

$AR(P)$ est un processus causal s'il peut être écrit dans un infini MA représentation $MA(\infty)$

$$\begin{aligned}\phi_p(B)^{-1}\phi_p(B)Y_t &= \phi_p(B)^{-1}\varepsilon_t \\ \Rightarrow Y_t &= \phi_p(B)^{-1}\varepsilon_t = \Psi_\infty(B)\varepsilon_t = \sum_{i=0}^{\infty} \Psi_i \varepsilon_{t-i}\end{aligned}$$

Où

$$\Psi_\infty(B) = \phi_p(B)^{-1}, \Psi_\infty(B) = 1 + \sum_{i=1}^{\infty} \psi_i B^i,$$

ψ_i satisfait $\sum_{i=0}^{\infty} |\psi_i| < \infty$, avec $\psi_0 = 1$.

les coefficients ψ_i peuvent être obtenus par égalisation des coefficients dans la relation $\phi_p(B)\Psi_\infty(B) = 1$.

Donc

$$\begin{aligned}\phi_p(B)\Psi_\infty(B) &= (1 - \phi_1 B - \dots - \phi_p B^p)(1 + \psi_1 B + \psi_2 B^2 + \dots) \\ &= 1 + (\psi_1 - \phi_1)B + (\psi_2 - \phi_1\psi_1 - \phi_2)B^2 + \dots \\ &\quad + (\psi_j - \phi_1\psi_{j-1} - \dots - \phi_p\psi_{j-p})B^j + \dots\end{aligned}$$

en assimilant des coefficients de différentes puissances B^i dans la relation $\phi_p(B)\Psi_\infty(B) = 1$, pour $i = 1, 2, \dots$, on a

$$\psi_i = \phi_1\psi_{i-1} + \phi_2\psi_{i-2} + \phi_3\psi_{i-3} + \dots + \phi_p\psi_{i-p},$$

où $\psi_0 = 1$, et $\psi_i = 0$ si $i < 0$

1.4.10 Processus Moyen mobile d'ordre q , $MA(q)$

Dans le processus moyenne mobile d'ordre q , chaque observation X_t est générée par une moyenne pondérée d'aléas jusqu'à la q -ième période.

$$X_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \dots + \theta_q\varepsilon_{t-q}$$

Où ε_t est un bruit blanc de moyenne nulle et de variance σ_ε^2 .

En introduisant l'opérateur moyenne mobile

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q.$$

on peut noter d'une façon équivalente : $X_t = \mu + \theta(B)\varepsilon_t$.

Un $MA(q)$ est toujours stationnaire quelles que soient les valeurs de θ ; il est de

moyenne μ .

On aimerait pouvoir exprimer ce processus en fonction de son passé (observé) et pas seulement en fonction du bruit passé non observé. C'est la question de l'inversibilité du processus. Examinons le cas d'un $MA(1)$ centré :

$$X_t = \varepsilon_t + \theta \varepsilon_{t-1} = (1 + \theta B) \varepsilon_t. \quad \varepsilon_t \sim BB(0, \sigma^2). \quad (1.6)$$

On voit que si $|\theta| < 1$, on peut développer $(1 + \theta B)^{-1}$ en série :

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

et écrire X_t , $AR(1)$, comme une autorégression infinie $AR(\infty)$:

$$X_t = \varepsilon_t + \theta X_{t-1} - \theta^2 X_{t-2} + \theta^3 X_{t-3} + \dots$$

on dit qu'il est inversible. Observons que la condition d'inversibilité d'un $MA(1)$ est parallèle à la condition de représentation causale d'un $AR(1)$.

Un $MA(q)$ est dit inversible si on peut le représenter comme une autorégression infinie.

1.4.10.1 Propriétés

Un $MA(q)$ est inversible si les racines de $1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q = 0$, sont en module, strictement supérieures à 1.

1.4.11 Processus ARMA(p,q)

Les processus ARMA sont des mélanges des processus AR et MA. Il sont nécessairement, en pratique, finis X_t obéit à un modèle ARMA(p, q) s'il est stationnaire et vérifie :

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}. \quad (1.7)$$

ε_t est un bruit blanc avec c constante arbitraire, $\phi_p \neq 0, \theta_q \neq 0$ et les polynômes $1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ et $1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$ n'ont pas de racines communes.

En utilisant l'opérateur retard, ce processus ARMA peut s'écrire comme suit :

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Y_t = c + (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) Y_t \varepsilon_t$$

Y_t obéissant est stationnaire si, comme dans le cas des autorégressifs, les racines du polynôme d'autorégression

$$1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p = 0 \quad (1.8)$$

sont en module strictement supérieures à 1.

Par un calcul identique à celui fait pour un $AR(p)$, on obtient que $\mu = E(Y_t)$ vérifie $(1 - \phi_1 - \phi_2 - \dots - \phi_p)c$ par la stationnarité, $1 - \phi_1 - \phi_2 - \dots - \phi_p$ et $\mu = c / (1 - \phi_1 - \phi_2 - \dots - \phi_p)$ ainsi peut encore s'écrire :

$$Y_t = \mu + \frac{1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q}{-\phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p} \varepsilon_t$$

Nombre de valeurs consécutives de la série chronologique à prédire. Si h est null, le nombre de valeurs consécutives à prédire est supposé être égal à la longueur de `timeseries.test`. Obligatoire lorsque `timeseries.test` est null.

On peut alors écrire une représentation $MA(\infty)$ de la série :

$$Y_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \quad \psi_0 = 1.$$

Par ailleurs, Y_t , $ARMA(p, q)$, est inversible si les racines de $\Theta(B)$ sont en module strictement supérieures à 1 et on peut écrire alors une représentation $AR(\infty)$ de la série :

$$Y_t = c + \sum_{i=1}^{\infty} \pi_i Y_{t-i} + \varepsilon_t.$$

- L'absence de racines communes dans (1.7) est une condition pour éviter la redondance des paramètres.
- Il arrive que certaines racines du polynôme autorégressif soient égales à 1. L'autorégressif est alors non stationnaire et on dit qu'il est intégré d'ordre d si 1 est d fois racine.

1.4.11.1 Fonction d'autocorrélation d'un AR

Partant de la représentation $MA(\infty)$ d'un $AR(1)$ on obtient :

$$var(Y_t) = \sigma^2 (1 + \phi^2 + \phi^4 + \dots) = \frac{\sigma^2}{1 - \phi^2}.$$

La fonction d'autocorrélation de l' $AR(1)$ est donc $\rho_k = \phi^k$, $k = 0, 1, 2, \dots$

Cette fonction décroît exponentiellement vers 0, en oscillant si $\phi < 0$.

la fonction d'autocovariance d'un $AR(p)$ Vérifier que :

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + \sigma_\varepsilon^2$$

$$\gamma_l = \phi_1 \gamma_{l-1} + \phi_2 \gamma_{l-2} + \dots + \phi_p \gamma_{l-p}, \quad l \geq 1.$$

donc la fonction d'autocorrélation d'un AR(p) :

$$\rho_l = \phi_1 \rho_{l-1} + \phi_2 \rho_{l-2} + \dots + \phi_p \rho_{l-p}, l \geq 1.$$

1.4.11.2 Fonction d'autocorrélation d'un MA

Commençons par calculer les moments d'ordre 2 d'un MA(1).

$$Y_t = \varepsilon_t + \theta \varepsilon_{t-1}$$

de moyenne $E(Y_t) = 0$.

La variance de Y_t définie par (1.6) est la variance d'une combinaison linéaire de variables non corrélées, donc :

$$\begin{aligned} \text{var}(Y_t) &= E(\varepsilon_t^2 + 2\theta \varepsilon_t \varepsilon_{t-1} + \theta^2 \varepsilon_{t-1}^2) = (1 + \theta^2) \sigma_\varepsilon^2, \\ \text{cov}(Y_t, Y_{t-1}) &= \text{cov}(\varepsilon_t + \theta \varepsilon_{t-1}, \varepsilon_{t-1} + \theta \varepsilon_{t-2}) = \theta \sigma_\varepsilon^2. \end{aligned}$$

on voit que $\text{cov}(Y_t, Y_{t-k}) = 0, k > 1$.

donc la fonction d'autocorrélation MA(1) :

$$\rho(k) = \begin{cases} 1 & \text{si } h = 0 \\ \frac{\theta}{1+\theta^2} & \text{si } h = 1 \\ 0 & \text{si } h > 1 \end{cases}$$

pour MA(q) : $Y_t = \sum_{i=0}^q \theta_i \varepsilon_{t-1}$, avec $\theta_0 = 1$, et $\varepsilon_{t-1} \sim BB(0, \sigma^2)$

$$\begin{aligned} E(Y_t) &= E\left(\sum_{i=0}^q \theta_i \varepsilon_{t-1}\right) = \sum_{i=0}^q E(\theta_i \varepsilon_{t-1}) = 0. \\ \text{var}(Y_t) &= \gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \delta^2 = \sigma^2 \sum_{i=0}^q \theta_i^2 \end{aligned}$$

$$\gamma_h = \begin{cases} \sigma^2 \sum_{i=0}^q \theta_i \theta_{i-h} & \text{si } h = 0, \pm 1, \pm 2, \dots, \pm q \\ 0 & \text{si } h > q \end{cases}$$

a ACF d'un MA(q) vérifie :

$$\gamma_h = \frac{\gamma_h}{\gamma_0} = \begin{cases} 0 & \text{si } h > q \\ \frac{\sum_{i=0}^q \theta_i \theta_{i-h}}{\sum_{i=0}^q \theta_i^2} & \text{sinon} \end{cases}$$

1.4.11.3 Inversibilité des modèles MA (q)

Définition 1.27 Le processus $MA(q)$ est inversible s'il peut être représenté forme AR infinie convergente $AR(\infty)$, $Y_t = \theta_q(B)\varepsilon_t$, avec

$$\theta_q(B) = 1 + \sum_{j=1}^q \theta_j B^j, \varepsilon_t \sim BB(0, \sigma^2)$$

$$\theta_q(B)^{-1} Y_t = \theta_q(B)^{-1} \theta_q(B) \varepsilon_t = \varepsilon_t \implies \pi_\infty(B) Y_t = \varepsilon_t$$

Parce que $\pi_\infty(B) = \theta_q(B)^{-1}$ es coefficients π_i peuvent être obtenus par égalisation des coefficients dans la relation $\theta_q(B)\pi_\infty(B) = 1$. Donc

$$\begin{aligned} \theta_q(B)\pi_\infty(B) &= (1 + \theta_1 B + \dots + \theta_q B^q)(1 - \pi_1 B - \pi_2 B^2 - \dots) \\ &= 1 - (\pi_1 - \theta_1)B - (\pi_2 + \theta_1 \pi_1 - \theta_2)B^2 - \dots \\ &\quad - (\pi_j + \theta_1 \pi_{j-1} + \dots + \theta_q \pi_{j-q})B^j - \dots \end{aligned}$$

en assimilant des coefficients de différentes puissances B^j dans la relation $\theta_q(B)\pi_\infty(B) = 1$ pour $j = 1, 2, \dots$

On a

$$\pi_j = -\theta_1 \pi_{j-1} - \theta_2 \pi_{j-2} - \dots - \theta_q \pi_{j-q},$$

avec $\pi_0 = -1$, et $\pi_j = 0$ pour $j < 0$, où

$$\pi_\infty(B) = \theta_q(B)^{-1} = 1 - \sum_{i=1}^{\infty} \pi_i B^i = - \sum_{i=0}^{\infty} \pi_i B^i$$

avec $\pi_0 = -1$, et π_i satisfaire ,

Notez que la condition de somme finie ($\sum_{i=0}^{\infty} |\pi_i| < +\infty$) assure que la série $AR(\infty)$ est convergente

Exemple 1.28 Un exemple de processus $MA(2)$ inversible est

$$Y_t = \varepsilon_t - 0.1 \varepsilon_{t-1} + 0.42 \varepsilon_{t-2}.$$

les racines de

$$(1 - 0.1B + 0.42B^2) = (1 - 0.7B)(1 + 0.6B) = 0$$

est

$$B_1 = \frac{1}{0.7} = 1.43 > 1$$

et

$$B_2 = \frac{-1}{0.6}, |B_2| = 1.67 > 1,$$

alors le processus est inversible .

Maintenant, le processus dans une représentation $AR(\infty)$ s'écrit comme suit :

$$Y_t = \sum_{i=1}^{\infty} \pi_j Y_{t-i} + \varepsilon_t,$$

avec

$$\begin{aligned} \pi_j &= -\theta_1 \pi_{j-1} - \theta_2 \pi_{j-2}, q = 2, \theta_1 = -0.1, \theta_2 = 0.42, \pi_0 = -1. \\ \pi_1 &= -\theta_1 \pi_0 - 0 = -(-0.1)(-1) = -0.1 \\ \pi_2 &= -\theta_1 \pi_1 - \theta_2 \pi_0 = -(-0.1)(-0.1) - (0.42)(-1) = 0.41. \\ \pi_3 &= -\theta_1 \pi_2 - \theta_2 \pi_1 - \theta_3 \pi_0 = -(-0.1)(0.41) - (0.42)(-0.1) = 0.083 \end{aligned}$$

Ainsi, le processus $MA(2)$ dans la représentation $AR(\infty)$ est

$$Y_t = \varepsilon_t - 0.1 Y_{t-1} + 0.41 Y_{t-2} + 0.083 Y_{t-3} + \dots$$

Propriété (Formule de Bartlett) :

Pour une série linéaire dont l'ACF vérifie : $\rho_k = 0, k > m$, on a : $\hat{\rho} \sim N(0, var(\hat{\rho}))$.

$$var(\hat{\rho}) \simeq \frac{1}{T} (1 + 2\rho_1^2 + \dots + 2\rho_m^2).$$

Ce résultat étend la Proposition (1.2). Il est précieux pour deviner (identifier) l'ordre de moyenne mobile convenable pour modéliser une série. En effet, en présence d'un corrélogramme empirique non significativement différent de 0 à partir d'un certain ordre $m + 1$, on essaiera d'ajuster à la série correspondante un modèle dont l'ACF est nulle à partir de l'ordre $m + 1$, un $MA(m)$. Mais comment savoir que l'ACF empirique à partir de l'ordre $m + 1$ est une estimation de 0? La formule de Bartlett permet de calculer des intervalles autour de 0 pour l'ACF d'un processus $MA(m)$, à partir du décalage $m + 1$: pour chaque retard $k > m$ on a en effet :

$$\hat{\rho} \in \left(-1.96 \sqrt{\frac{1}{T} (1 + 2\rho_1^2 + \dots + 2\rho_m^2)}, +1.96 \sqrt{\frac{1}{T} (1 + 2\rho_1^2 + \dots + 2\rho_m^2)} \right)$$

avec une probabilité d'environ 95%.

Supposons en particulier que le processus étudié est un bruit blanc, alors $\rho_k, k > 0$ doit appartenir à l'intervalle $-1.96/\sqrt{T}, +1.96/\sqrt{T}$ à 95% environ. En superposant le graphique de l'ACF ρ_k et cet intervalle ou son approximation $-2/\sqrt{T}, +2/\sqrt{T}$, on peut voir si l'hypothèse de blancheur est raisonnable. On peut tracer ces intervalles pour une série supposée bruit blanc (cf. Proposition 1.2). On représente habituellement ces intervalles sur les graphiques d'ACF empirique. On note que les autocorrélations sont bien incluses dans l'intervalle à 95% autour de zéro. Il peut arriver qu'avec

une autre simulation, une autocorrélation sorte de l'intervalle.

1.4.11.4 Processus ARMA(p,q) (Autoregressive-Moving Average)

Les processus *ARMA* sont des mélanges des processus *AR* et *MA*. Ils sont nécessairement, en pratique, finis Y_t obéit à un modèle *ARMA*(p, q) s'il est stationnaire et vérifie :

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (*)$$

ε_t est un bruit blanc, $\phi_p \neq 0$, $\theta_q \neq 0$, et les polynômes $1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ et $1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$ n'ont pas de racines communes.

— En utilisant l'opérateur retard, ce processus ARMA peut s'écrire comme suit :

$$\begin{aligned} (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Y_t &= (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) \varepsilon_t \\ \phi_p(B) Y_t &= \theta_q(B) \varepsilon_t \end{aligned}$$

partie *AR*(p) partie *MA*(q), avec

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p = 1 - \sum_{i=1}^p \phi_i B^i$$

et

$$\theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q = 1 + \sum_{i=1}^q \theta_i B^i$$

— Si moyenne de Y_t est μ , remplacer Y_t par $Y_t - \mu$ obtenir :

$$\phi_p(B)(Y_t - \mu) = \theta_q(B) \varepsilon_t.$$

peut également être écrit comme :

$$Y_t = \alpha + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

où

$$\alpha = \mu(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p).$$

1.4.11.4.1 Conditions de modèle ARMA Le modèle *ARMA* est supposé stationnaire, inversible et identifiable, où :

- La condition de stationnaire c'est pareil pour le Processus *AR*(p), c'est-à-dire les racines en valeur absolue de $\phi_p(B)$ strictement supérieures à 1.
- La condition d'inversible c'est pareil pour le Processus *MA*(q), c'est-à-dire les racines en valeur absolue de $\theta_q(B)$ strictement supérieures à 1.

— La condition d'identifiable signifie que le modèle n'est pas redondant, c'est-à-dire $\phi_p(B) = 0$ et $\theta_q(B) = 0$ n'ont pas de racines communes.

Exemple 1.29 Considérer $ARMA(1,2)$:

$$Y_t = 0.2Y_{t-1} + \varepsilon_t - 1.1\varepsilon_{t-1} + 0.18\varepsilon_{t-2}$$

ce modèle peut s'écrire :

$$(1 - 0.2B)Y_t = (1 - 1.1B + 0.18B^2)\varepsilon_t$$

ou équivalent

$$(1 - 0.2B)Y_t = (1 - 0.2B)(1 - 0.9B)\varepsilon_t,$$

Il y a une racine commune donc le modèle est redondant annulation $(1 - 0.2B)$ des deux côtés pour obtenir

$$Y_t = (1 - 0.9B)\varepsilon_t.$$

Ainsi, le processus n'est pas vraiment un $ARMA(1,2)$, mais c'est un $MA(1) \equiv ARMA(0,1)$.

1.4.11.5 ACF pour les modèles $ARMA(p, q)$

Pour un modèle $ARMA$

$$Y_t = \sum_{j=1}^p \phi_j Y_{t-j} + \sum_{j=0}^q \theta_j \varepsilon_{t-j},$$

avec $\theta_0 = 1$,

$$\begin{aligned} \gamma_h &= cov(Y_{t+h}, Y_t) = E(Y_{t+h}Y_t) = E\left[\left(\sum_{j=1}^p \phi_j Y_{t+h-j} + \sum_{j=0}^q \theta_j \varepsilon_{t+h-j}\right)Y_t\right] \\ &= \sum_{j=1}^p \phi_j E[Y_{t+h-j}Y_t] + \sum_{j=0}^q \theta_j E\left[\varepsilon_{t+h-j} \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}\right] \\ &= \sum_{j=1}^p \phi_j \gamma_{h-j} + \sigma^2 \sum_{j=h}^q \theta_j \psi_{j-h}, \text{ pour } h \geq 0. \end{aligned}$$

Cela donne l'équation de différence homogène générale pour γ_h :

$$\gamma_h - \phi_1 \gamma_{h-1} - \phi_2 \gamma_{h-2} - \dots - \phi_p \gamma_{h-p} = 0, \text{ pour } h \geq \max(p, q + 1)$$

aux conditions initiales

$$\gamma_h - \sum_{j=1}^p \phi_j \gamma_{h-j} = \sigma^2 \sum_{j=h}^q \theta_j \psi_{j-h}, \text{ pour } 0 \leq h \leq \max(p, q + 1).$$

◇ Le processus $ARIMA(1, 1, 1)$ peut être écrit comme suit :

$$(1 - \phi B)(1 - B)Y_t = (1 + \theta B)\varepsilon_t$$

◇ Le processus $ARIMA(2, 1, 1)$ peut être écrit comme suit :

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B)Y_t = (1 + \theta B)\varepsilon_t$$

◇ Le processus $ARIMA(1, 2, 2)$ peut être écrit comme suit :

$$(1 - \phi_1 B)(1 - B)^2 Y_t = (1 + \theta_1 B + \theta_2 B^2)\varepsilon_t$$

Remarque 1.31 Les modèles $ARIMA$ sont appliqués dans certains cas où les données montrent preuve de non stationnarité, où une première étape de différenciation peut être appliquée une ou plusieurs fois pour éliminer le non stationnarité.

- $AR(p) \equiv ARIMA(p, 0, 0)$
- $MA(q) \equiv ARIMA(0, 0, q)$
- $ARI(p, d) \equiv ARIMA(p, d, 0)$,
- $IMA(d, q) \equiv ARIMA(0, d, q)$
- $ARMA(p, q) \equiv ARIMA(p, 0, q)$,
- $BB \equiv ARIMA(0, 0, 0)$, BB est un bruit blanc.

1.4.13 Modèle saisonnière $ARIMA$ ($SARIMA$)

Un processus non stationnaire possède souvent une composante saisonnière qui se répète après une période de temps régulière, où le plus petit la période indiquée par s est appelée période saisonnière.

Définition 1.32 Le modèle multiplicatif *Seasonal* $ARIMA$ ($SARIMA$) désigné par $ARIMA(p, d, q) \times (P, D, Q)_s$, où s est le nombre de saisons :

$$\phi_p(B)\Phi_p(B^s)(1 - B)^d(1 - B^s)^D Y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t,$$

- ◇ $\phi_p(B) = 1 - \sum_{i=1}^p \phi_i B^i$ polynôme en B de degré p ,
- ◇ $\theta_q(B) = 1 + \sum_{i=1}^q \theta_i B^i$ polynôme en B de degré q ,
- ◇ $\Phi_p(B^s) = 1 - \sum_{i=1}^P \phi_i B^{is}$ polynôme en B^s de degré P ,
- ◇ $\Theta_Q(B^s) = 1 + \sum_{i=1}^Q \theta_i B^{is}$ polynôme en B^s de degré Q . sans racines communes entre $\Phi_p(B^s)$ et $\Theta_Q(B^s)$, p, d et q sont l'ordre du modèle AR non saisonnier, du modèle MA et ordinaire différenciation respectivement, alors que P, D , et

Q sont de l'ordre de modèle autorégressif saisonnier (*SAR*), moyenne mobile saisonnière (*SMA*) et différenciation saisonnière respectivement

- L'idée est que les *SARIMA* sont des modèles $ARIMA(p, d, q)$ dont les résidus t sont $ARIMA(P, D, Q)$ dont les opérateurs sont définis sur les B^s et les puissances successives, où p, q et d sont les commande *AR* non saisonnière, commande *MA* non saisonnière et différenciation non saisonnière respectivement, tandis que P, Q et D sont l'ordre saisonnier *AR* (*SAR*), l'ordre saisonnier *MA* (*SMA*), et différenciation saisonnière au décalage s respectivement.
- La différenciation saisonnière

$$\Delta_s Y_t = (1 - B^s) Y_t = Y_t - Y_{t-s}$$

supprimera la saisonnalité de la même manière que la différenciation ordinaire $\Delta Y_t = Y_t - Y_{t-1}$ supprimera une tendance polynomiale.

- ◇ Modèle $SARIMA(0, 1, 0) \times (0, 1, 0)_5$ peut être écrit comme suit :

$$(1 - B)(1 - B^5) Y_t = \varepsilon_t$$

- ◇ Modèle $SARIMA(0, 1, 0) \times (0, 1, 1)_4$ peut être écrit comme suit :

$$(1 - B)(1 - B^4) Y_t = (1 - \theta B^4) \varepsilon_t.$$

1.5 Les processus aléatoires non stationnaires

Les processus stochastiques non stationnaires sont caractérisés par des propriétés stochastiques qui évoluent en fonction du temps. On distingue deux types de processus stochastiques non stationnaires : une non stationnarité de nature déterministe (TS) et une non stationnarité de nature stochastique (DS)

— Description des processus TS

Un processus (Y_t) présente une non stationnarité de type déterministe TS (Trend Stationary), s'il peut se décomposer en une somme de deux fonctions : $Y_t = f_t + \varepsilon_t$

Tel que : ε_t : est un bruit blanc. f_t : est une fonction polynomiale du temps.

polynomiale de degré 1, il s'écrit : $Y_t = \alpha_0 + \alpha_1 t + \varepsilon_t$, où $\alpha_0, \alpha_1 \in \mathbb{R}$.

Les caractéristiques de ce processus sont :

$$E(Y_t) = \alpha_0 + \alpha_1 t$$

$$var(Y_t) = \sigma^2, cov(Y_t, Y_{t-h}) = 0, \forall h \neq 0.$$

La non stationnarité de ce processus est dû au fait que son espérance dépend du temps.

La méthode pour stationnariser un processus TS est d'estimer les coefficients α_0 , α_1 par MCO (Moindre Carrés Ordinaires) et de retrancher de la valeur de Y_t En t la valeur estimée de sa moyenne $\hat{\alpha}_0 + \hat{\alpha}_1 t$.

— **Description d'un processus DS**

DS sans dérive

Soit le processus DS sans dérive (AR(1)) :

$$Y_t = Y_{t-1} + \varepsilon_t \iff (1 - B)Y_t = \varepsilon_t$$

La racine du polynôme caractéristique $(1 - B)$ est égale à 1. On dit que le processus Y_t a une racine unité, il est donc non stationnaire.

Ce processus DS sans dérive peut se réécrire sous la forme :

$$\triangleright Y_t = Y_{t-1} + \varepsilon_t \quad (1)$$

$$\triangleright Y_{t-1} = Y_{t-2} + \varepsilon_{t-1} \quad (2)$$

$$\triangleright Y_{t-2} = Y_{t-3} + \varepsilon_{t-2} \quad (3)$$

la somme de (1)+(2) équations : $Y_t = Y_{t-2} + \varepsilon_{t-1} + \varepsilon_t$

la somme de (1)+(2)+(3) équations : $Y_t = Y_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t$

Si le premier terme de la chronique est Y_0 , le modèle s'écrit alors : $Y_t = Y_0 + \sum_{i=0}^t \varepsilon_i$

Les caractéristiques de ce processus sont (en supposant Y_0 certain) : $E(Y_t) = Y_0$

$$var(Y_t) = t\sigma^2.$$

$$cov(Y_t, Y_{\check{t}}) = \sigma^2 \min(t, \check{t}) \text{ si } t \neq \check{t}$$

Un processus DS sans dérive est un processus stationnaire en moyenne et non stationnaire en variance.

DS avec dérive

Considérons un processus DS avec dérive :

$$Y_t = \mu + Y_{t-1} + \varepsilon_t = t\mu + Y_0 + \sum_{j=1}^t \varepsilon_j.$$

Un processus DS avec dérive est un processus non stationnaire en moyenne et en variance. Ces moments évoluent en fonction du temps t. Un processus DS est un processus que l'on peut stationnariser par l'application du filtre aux différences :

$$Y_t = Y_{t-1} + \varepsilon_t \implies \Delta Y_t = \varepsilon_t$$

1.5.1 Test de non-stationnarité : test Dickey-Fuller

Dickey et Fuller ont construit leur test à partir des modèles de base suivants :

♣ Modèle(1) : $Y_t = \phi Y_{t-1} + \varepsilon_t$ $AR(1)$

♣ Modèle(2) : $Y_t = c + \phi Y_{t-1} + \varepsilon_t$ $AR(1)$ avec constante.

♣ Modèle(3) : $Y_t = c + bt + \phi Y_{t-1} + \varepsilon_t$ $AR(1)$ avec constante et tendance.

Où ε_t est un bruit blanc de variance σ^2 et b, c des constantes réelles. Le principe du test consiste à tester l'hypothèse nulle de racine unitaire contre l'hypothèse alternative d'absence de racine unitaire

$$H_0 : \phi = 1 \text{ \& } H_1 : |\phi| < 1$$

Les modèles de base du test étant théoriques, l'application du test requiert l'estimation en pratique de modèles :

♣ Modèle(1)' : $\Delta Y_t = (\phi - 1) Y_{t-1} + \varepsilon_t$

♣ Modèle(2)' : $\Delta Y_t = c + (\phi - 1) Y_{t-1} + \varepsilon_t$

♣ Modèle(3)' : $\Delta Y_t = c + bt + (\phi - 1) Y_{t-1} + \varepsilon_t$

On calcule la t-statistique $t_{\hat{\phi}}$ qui est donnée par : $t_{\hat{\phi}} = \frac{\hat{\phi} - 1}{\hat{\sigma}^2}$

$t_{\hat{\phi}}$ sera comparée à la valeur critique tabulée notée t_{tab} et on applique la règle suivante :

- ▶ Si $t_{\hat{\phi}} < t_{tab}$ on rejette H_0
- ▶ Si $t_{\hat{\phi}} \geq t_{tab}$ on accepte H_0

En pratique, on n'effectue pas ce test sur les trois modèles mais on procède par une stratégie séquentielle en trois étapes suivantes :

Etape 1 : On estime le modèle (3)' et on teste la significativité de la tendance déterministe (test de Student sur le paramètre b)

— Si cette tendance estimée n'est pas significativement différente de zero (donc la t-statistique de la tendance est inférieure aux valeurs critiques de la tendance tabulée par Dickey-Fuller) alors on passe à l'étape 2.

— Si la tendance est différente de zero, on teste l'hypothèse nulle unitaire :

- ▶ Si on accepte H_0 , Y_t est non stationnaire de type DS
- ▶ Si on rejette H_0 , Y_t est non stationnaire de type TS

Etape 2

On aura à appliquer cette étape que si à l'étape 1 on a rejeté l'idée d'une tendance significative. On estime le modèle (2)' et on teste la significativité de la constante c .

— Si H_0 est acceptée Y_t est non stationnaire de type DS

— Si H_0 est rejetée Y_t est stationnaire.

Etape 3

Si l'étape 2 detecte une constante nulle, alors on estime le modèle(1) et on effectue le test de racine unitaire tel que :

- Si H_0 est acceptée Y_t est non stationnaire de type DS
- Si H_0 est rejetée Y_t est alors stationnaire.

1.5.2 Test de Dickey-Fuller augmenté

En passant au test de Dickey-Fuller, l'existence d'une racine unitaire conduit donc à la différenciation de la série en question. Une fois la série différenciée, on peut se demander si la nouvelle série obtenue après différenciation est stationnaire ou non. On applique donc à nouveau le test de Dickey-Fuller et ainsi de suite. Selon Dickey et Pantula, cette procédure dite séquentielle ascendante peut donner des résultats faux car les distributions statistiques diffèrent suivant qu'il existe une ou deux racines unitaires. Ainsi, ils ont proposé, en se référant aux tables de Dickey-Fuller, une nouvelle procédure dite séquentielle descendante. Cette procédure permet de tester en même temps l'existence de plusieurs racines unitaires. Supposons que l'on veuille tester l'existence de deux racines unitaires, alors le test se basera sur le modèle suivant :

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-1} + \phi_1 \phi_2 Y_{t-2} + \varepsilon_t$$

Et en pratique on estime le modèle suivant :

$$\Delta^2 Y_t = \phi_1 Y_{t-1} + \Delta \phi_2 Y_{t-1} + \varepsilon_t$$

Où l'on pose $\theta_1 = -(\phi_1 - 1)(\phi_2 - 1)$

et $\theta_2 = (\phi_1 \phi_2 - 1)$. Le test se fait en deux étapes suivantes :

Etape 1

On teste l'hypothèse nulle de deux racines unitaires contre l'alternative d'une seule racine unitaire

$$H_0 : \theta_1 = \theta_2 = 0 \text{ \& } H_1 : \theta_1 \neq 0$$

Sous l'hypothèse H_1 , on estime $\Delta^2 Y_t = \theta_2 \Delta Y_{t-1} + \varepsilon_t$ et on fait l'usage de la règle suivante :

- Si $t_{\theta_2} \leq t_{tab}$ on rejette H_0 et on passe à l'étape 2
- Si $t_{\theta_2} > t_{tab}$ on accepte H_0 , le processus contient donc deux racines unitaires.

Etape 2

Si à l'étape 1 on rejette l'hypothèse H_0 , alors on peut appliquer l'étape 2.

On teste donc l'hypothèse nulle de la présence d'une racine unitaire contre l'hypothèse alternative d'aucune racine unitaire dans

$\Delta Y_t = (\phi_1 - 1)Y_{t-1} + \varepsilon_t$. Ce test sur $(\phi_1 - 1)$ est l'équivalent de celui sur θ_1 dans le modèle $\Delta^2 Y_t$.

1.6 La méthodologie de Box et Jenkins

une méthodologie de modélisation d'une série chronologique univariée basée sur les modèles linéaires ARMA, ARIMA. Cette méthodologie possède trois étapes : identification, estimation et validation

— Identification du modèle

En premier lieu, on examine le graphe représentatif de la série temporelle, ceci peut donner une idée préliminaire sur le comportement de la série (stationnarité, tendance, saisonnalité,...). Si la série présente une tendance et/ou une saisonnalité, des transformations adéquates doivent être appliquées afin de stationnariser la série. L'idée générale de l'identification dans la méthodologie Box-Jenkins, consiste à comparer la structure des corrélations estimées que présente la série à travers le corrélogramme (diagramme représentatif des autocorrélations estimées) avec la structure de corrélation théorique exhibée par des modèles bien connus. Ainsi l'étude du corrélogramme est très utile pour la détermination des ordres p et q , puisque les fonctions d'autocorrélation simples et partielles peuvent indiquer la présence d'un modèle moyenne mobile ou auto-régressif respectivement. Plus précisément si la fonction d'autocorrélation simple décroît rapidement vers 0 et la fonction d'autocorrélation partielle présente un cut-off après p retard, on peut conclure que la série provient d'un processus AR d'ordre p (AR(p)). Si la fonction d'autocorrélation simple présente un cut-off après q retards et que la fonction d'autocorrélations partielles décroît rapidement vers 0, alors on peut conclure que la série est générée à partir d'un modèle moyenne mobile d'ordre q (MA(q)). On note que si les fonctions d'autocorrélations simples et partielles présentent une forme exponentielle ou sinusoidale, on constate qu'on est en présence d'un processus auto-régressif à moyenne mobile ARMA(p, q). Cette étape n'est pas aisée et demande beaucoup d'expertise, il existe cependant des méthodes d'identifications automatiques, basées sur le critère d'information.

	<i>ACF</i>	<i>PACF</i>
Bruit blanc	Tous les zéros	Tous les zéros
$AP(P)$	S'évanouit avec une décroissance exponentielle	coupure après le décalage p
$MA(q)$	coupure après le décalage q	S'évanouit avec une décroissance expon
$ARMA(p, q)$	Fin après le décalage $q-p$	Fin après le décalage $p-q$

1.7 Critère d'information

Il existe des critères d'informations qui sont utilisés, comme guide, dans le choix du modèle, ce qui nous permet d'éviter la sélection arbitraire des paramètres p et q du modèle. Parmi ces critères, il existe les critères d'information qui mesurent l'écart entre la vraie loi inconnue et celle du modèle proposé; les estimations de la qualité d'information qui ont été proposées sont :

1. Critère d'Akaike(1969) appelé aussi *AIC*, il est défini comme suit

$$AIC(p, q) = \log \sigma^2 + \frac{2(p + q)}{N}.$$

2. Critère Bayésien(1977) appelé aussi *BIC*, il est défini comme suit :

$$BIC(p, q) = \log \sigma^2 + 2(p + q) \frac{\log(N)}{N}.$$

1.8 Estimation des paramètres et validation du modèle

1.8.1 Estimation des paramètres

Après avoir terminé l'identification, il convient d'estimer les paramètres qui sont les coefficients des polynômes AR et MA et la variance des résidus ε_t . La méthode d'estimation la plus utilisée est celle du maximum de vraisemblance ou la méthode des moindres carrés. Le principe consiste à construire une fonction dite de fonction de vraisemblance et par la suite à maximiser son logarithme par rapport aux paramètres θ_i, θ_j , (avec $i = 1, \dots, p; j = 1, \dots, q$), permettant ainsi de trouver la valeur numérique la plus vraisemblable pour ces paramètres. L'étape d'estimation finie, l'étape suivante va nous permettre de valider le modèle estimé.

1.8.2 Vérification et validation

Au début de cette étape on dispose de plusieurs processus ARMA dont on a estimé les paramètres. Il faut maintenant valider ces modèles afin de les départager. Pour cela, on applique des tests sur les paramètres et sur les résidus. Si plusieurs modèles sont validés, l'étape de validation doit se poursuivre par une comparaison de qualité de ces derniers.

- a) Tests concernant les paramètres :** Après avoir estimé les paramètres d'un modèle, on peut se poser la question de savoir si ces paramètres sont significativement différents de zéro. Ces tests sont aussi appelés tests sur le modèle, car si le test de significativité des paramètres détecte des paramètres non significatifs,

cela entraîne automatiquement un changement dans l'ordre du modèle. Soit un modèle $ARMA(p; q)$.

1. On peut par exemple tester $\hat{p} = p - 1$ et $\hat{q} = q - 1$. Ici il est question de savoir si, on peut diminuer d'une unité l'ordre de la partie AR . Pour cela, on utilise le test de Student qui va tester la signification du coefficient ϕ_p . Soit $\hat{\phi}_p$ l'estimateur de ϕ_p et $\hat{V}(\hat{\phi}_q)$ sa variance estimée. En supposant que les estimateurs sont normalement distribués au risque de 0.05 on compare la valeur de la statistique t_c donnée par,

$$t_c = \frac{|\hat{\phi}|}{(V(\hat{\phi}_q))^2}$$

à la valeur critique 1.96. Si t_c est supérieure à 1.96, on rejette l'hypothèse $\phi_p = 0$. Dans le cas contraire on accepte l'hypothèse de nullité de ϕ_p .

2. On peut également tester $\hat{p} = p + 1$ et $\hat{q} = q$. Dans ce cas, il s'agit de connaître la possibilité d'augmenter l'ordre de la partie autoregressive. Donc il faut tester la signification du coefficient ϕ_{p+1} . Comme précédemment, on utilise la statistique de Student. On compare alors le rapport

$$t_c = \frac{|\phi_{p+1}|}{(\hat{V}(\hat{\phi}_{p+1}))^{\frac{1}{2}}}$$

à 1.96 (au risque de 0.05)

b) Tests concernant le bruit blanc : L'une des hypothèses qui doivent être vérifiées de manière rigoureuse pour la validité d'un modèle ajustée est celle de bruit blanc. En effet, si les résidus ne forment pas un bruit blanc on peut penser à une mauvaise stationnarisation des données ou même à un mauvais choix du modèle.

1. **Test Portmanteau (Box et Pierce)**
2. **Test de Durbin-Watson**
3. **Test de Ljung et Box(amélioration du test du Portemanteau) :** Ce test est d'une part comme une amélioration du test de Box et Pierce. D'autre part, si on ne connaît rien sur la structure du processus comme souvent c'est le cas, le test de Ljung-Box est plus général que le test de Durbin-Watson. Il est basé sur la statistique :

$$\hat{Q} = n(n+2) \sum_{h=1}^K \frac{\hat{\rho}_\varepsilon^2(h)}{n-h}$$

\hat{Q} suit une loi de χ^2 à $(K - p - q)$ degrés de liberté et le test se déroule comme celui de Box-Pierce.

On peut résumer la méthodologie de Box et Jeninks par le schéma ci-dessous.

1.9 Préviation

Soit $(Y_t)_{t \in T}$, un processus au second ordre réel et centré. On a que la fonction d'autocovariance est notée $\gamma(i, j) = E(Y_i Y_j)$. Soit Y_1, Y_2, \dots, Y_n un échantillon de $(Y_t)_{t \in T}$, on note $H_n = Y_1, Y_2, \dots, Y_n$ le sous espace fermé de $L^2(\Omega, A, P)$ engendré par $(Y_t)_{1 \leq j \leq n}$. On pose que $\hat{Y} = 0$ et $\hat{Y}_j = H_{n-1}(Y_j)$. On suppose que la matrice $K(i, j)$ est définie positive, on montre

$$H(n) = [(Y_1 - \hat{Y}_1), (Y_2 - \hat{Y}_2), \dots, (Y_n - \hat{Y}_n)]$$

Et on déduit que

$$\hat{Y}_{n+1} = \sum_{j=1}^n \theta_{n,j} (Y_{n+1-j} - \hat{Y}_{n+1-j})$$

pour $n \geq 1$.

Pour des prévisions d'ordre $h \geq 1$

$$\hat{Y}_{n+h} = \sum_{j=1}^{n+h-1} \theta_{n+h-1,j} (Y_{n+1-j} - \hat{Y}_{n+1-j})$$

Si on écrit (Y_t) sous forme d'une moyenne mobile infinie :

$$Y_t = \varepsilon_t + \sum_{j=1}^{\infty} (b_j \varepsilon_{t-j})$$

On a l'intervalle de confiance au seuil de $\alpha = 0.05$ d'où

$$Y_t \in \left[\hat{Y}_{n+h} - 1.96\sigma_\varepsilon \left(\sum_{j=0}^{h-1} (b_j^2) \right), \hat{Y}_{n+h} + 1.96\sigma_\varepsilon \left(\sum_{j=0}^{h-1} (b_j^2) \right) \right]$$

CHAPITRE 2

PRÉVISION PAR LES SÉRIES CHRONOLOGIQUES

Le filtre de Kalman est une approche statistique, d'assimilation de données, dont le principe est de corriger la trajectoire du modèle en combinant les observations avec l'information fournie par le modèle de façon à minimiser l'erreur entre l'état vrai et l'état filtré. Le filtre de Kalman a été utilisé dans un certain nombre d'études pour analyser la variabilité des paramètres dans le temps .

2.1 Les modèles d'état

Les lettres en gras représentent des vecteurs-colonne. Un modèle d'état est constitué de deux équations :

— une équation d'observation (ou de mesure) :

$$Y_t = G_t X_t + W_t, \quad t = 1, 2, \dots \quad (2.1)$$

— une équation d'état :

$$X_{t+1} = F_t X_t + V_t, \quad t = 1, 2, \dots \quad (2.2)$$

Dans ces équations :

- ♣ Y_t : est un vecteur de dimension ω
- ♣ X_t : est un vecteur de dimension ν
- ♣ G_t : est une séquence temporelle de matrices (ω, ν)
- ♣ F_t : est une séquence temporelle de matrices (ν, ν)
- ♣ W_t : est une séquence temporelle de vecteurs aléatoires indépendants de dimension ω (bruit de mesure), de moyenne nulle et de matrice de covariance $R_t(\omega, \omega)$

- ♣ V_t : est une séquence temporelle de vecteurs aléatoires indépendants de dimension v (bruit d'état), de moyenne nulle et de matrice de covariance $Q_t(v, v)$.

Les séquences W_t et V_t sont indépendantes entre elles, et décorréées avec X_1 .

Dans de nombreux cas pratiques, les matrices F_t , G_t , Q_t et R_t sont indépendantes du temps, et sont alors notées respectivement F , G , Q et R .

Il existe des formes plus générales de ces équations :

- les séquences W_t et V_t peuvent être corrélées,
- un terme de contrôle $H_t u_t$ peut venir s'ajouter dans le terme de droite de l'équation d'état, pour imposer un changement de X_{t+1} .

Proposition 2.1 Soient (Y_1, \dots, Y_n) des vecteurs aléatoires de R_k , U un vecteur aléatoire de R_d et V un vecteur aléatoire de R_q .

- i) Si A est une matrice $v \times d$, alors :

$$P(AU|Y_1, \dots, Y_n) = AP(U|Y_1, \dots, Y_n).$$

- ii) En notant $M = E(UY_1^T)(E(Y_1Y_1^T))^{-1}$, où S^{-1} désigne une matrice telle que

$$SS^{-1}S = S$$

(c'est une inverse généralisée), on a :

$$P(U|Y_1) = MU,$$

Notamment si V est décorréé de Y_1 , alors :

$$P(U|Y_1) = 0.$$

- iii) Si V est décorréé de Y_1 :

$$P(U|Y_1, V) = P(U|Y_1) + P(U|V).$$

2.2 Filtre de Kalman

Les formules récursives de Kalman ont pour objectif général de trouver des estimateurs linéaires optimaux, du vecteur d'état X_k en fonction des observations Y_1, Y_2, \dots et d'un vecteur aléatoire Y_0 orthogonal à V_t et W_t pour tout $t \geq 1$ (on prend en général pour Y_0 le vecteur $(1, 1, \dots, 1)^t$). On notera $P_t(X_k)$ le meilleur prédicteur linéaire de X_k en fonction de Y_0, Y_1, \dots, Y_t . Les définitions de prédiction, filtrage, lissage sont les suivantes :

- ◆ Prédiction : estimer (au sens défini ci-dessus) X_t en fonction de Y_0, Y_1, \dots, Y_{t-1}
- ◆ Filtrage : estimer X_t en fonction de Y_0, Y_1, \dots, Y_t
- ◆ Lissage : estimer X_t en fonction de Y_0, Y_1, \dots, Y_t avec $n > t$

Soit $P_t(X_k)$ le meilleur prédicteur linéaire de X_k en fonction de Y_0, Y_1, \dots, Y_t

Le filtre de Kalman a donc pour objectif de calculer à chaque instant t la valeur de $P_t(X_k)$, ainsi que la matrice de covariance de l'erreur commise sur cet estimateur.

Pour cela, le filtre opère à chaque itération t en plusieurs étapes :

- **Prédiction de l'état** : calcul de $P_{t-1}(X_t)$ et covariance associée $C_{t|t-1}$ de l'erreur de prédiction
- **Prédiction de la mesure** : calcul de $P_{t-1}(X_t)$
- **Calcul du terme d'innovation**, égal à l'écart entre la mesure Y_t et sa valeur prédite, et de sa covariance S_t
- **Calcul du gain de filtrage** K_t
- **Filtrage** : calcul de $P_t(X_t)$ comme combinaison linéaire de la prédiction et de l'innovation, et covariance associée $C_{t|t}$ de l'erreur d'estimation.

2.2.1 Equations du filtre

- **Initialisation** : Celle-ci consiste à donner une estimée de X à l'instant 0, soit $P_0(X_0)$ ainsi que la matrice de covariance $C_{0|0}$. Ceci se fait à partir des informations a priori dont on dispose sur l'état à l'instant 0; en absence d'information a priori, $C_{0|0}$ sera pris très grand, ce qui a pour effet d'accorder un poids négligeable à la valeur initiale.
- **Formules récursives** $t = 1, 2, \dots$

1. Prédiction de l'état

$$\begin{aligned} P_{t-1}(X_t) &= F_{t-1}P_{t-1}(X_{t-1}) \\ C_{t|t-1} &= F_{t-1}C_{t-1|t-1}F_{t-1}^T + Q_{t-1} \end{aligned}$$

2. Prédiction de la mesure

$$P_{t-1}(Y_t) = G_t P_{t-1}(X_t)$$

3. Calcul de l'innovation à l'instant t et de sa covariance S_t

$$\begin{aligned} v_t &= Y_t - P_{t-1}(Y_t) \\ S_t &= G_t C_{t|t-1} G_t^T + R_t \end{aligned}$$

4. Calcul du gain K_t

$$K_t = C_{t|t-1} G_t^T S_t^{-1}$$

5. Estimation de l'état courant, comme combinaison linéaire de la valeur prédite et de l'innovation

$$\begin{aligned} P_t(X_t) &= P_{t-1}(X_t) + K_t v_t \\ C_{t|t} &= C_{t|t-1} - K_t S_t K_t^T \end{aligned}$$

Note 1

La dernière équation : $C_{t|t} = C_{t|t-1} - K_t S_t K_t^T$ fait apparaître explicitement la diminution de la covariance de l'erreur d'estimation apportée par une nouvelle mesure. On peut l'écrire de façon équivalente :

$$C_{t|t} = (I - K_t G_t) C_{t|t-1} (I - K_t G_t)^T + K_t R_t K_t^T$$

issue de l'équation :

$$P_t(X_t) = (I - K_t G_t) P_{t-1}(X_t) + K_t Y_t$$

Note 2 :

Une mesure Y_t de mauvaise qualité ($R_t \rightarrow \infty$) conduit à $P_t(X_t) = P_{t-1}(X_t)$ (la mesure Y_t n'est pas prise en compte) ; une mesure Y_t sans erreur ($R_t = 0$) conduit à $G_t P_t(X_t) = Y_t$. Ci-dessous, on donne la démonstration des formules récursives du filtre de Kalman.

Démonstration : Rappelons que V_{t-1} est décorrélé de X_{t-1} , et W_t est décorrélé de X_t .

$$P_{t-1}(X_t) = P_{t-1}(F_{t-1} X_t + V_t) = F_{t-1} P_{t-1}(X_t) + P_{t-1}(V_t) = F_{t-1} P_{t-1}(X_t).$$

Et de même, $P_{t-1}(Y_t) = G_t P_{t-1}(X_t)$. Pour estimer X_t à l'instant t , i.e. en ayant observé Y_1, \dots, Y_t on décompose la projection à l'aide du point (iii) de la Proposition 2.1 :

$$P(X_t | Y_0, \dots, Y_t) = P(X_t | Y_0, \dots, Y_t) = P(X_t | Y_0, \dots, Y_t) + P(X_t | v_t).$$

C'est à dire

$$P_t(X_t) = P_{t-1}(X_t) + P(X_t | v_t).$$

En utilisant le point (ii) de la Proposition 2.1 :

$$P(X_t | v_t) = E(X_t v_t^T) S_t^{-1} v_t,$$

où $S_t = E(v_t v_t^T)$ est la matrice de covariance des innovations à l'instant t . Posons $K_t = E(X_t v_t^T) S_t^{-1}$.

Il reste à évaluer récursivement K_t . En écrivant $v_t = G_t(X_t - P_{t-1}(X_t)) + W_t$, et en notant que W_t est décorrélé de X_t ,

$$E(X_t v_t^T) = E(X_t(X_t - P_{t-1}(X_t))^T)G_t^T = C_{t|t-1}G_t^T,$$

où

$$C_{t|t-1} = E[(X_t - P_{t-1}(X_t))(X_t - P_{t-1}(X_t))^T].$$

On a :

$$\begin{aligned} C_{t|t-1} &= E[X_t X_t^T] - E[P_{t-1}(X_t)P_{t-1}(X_t)^T] \\ &= F_{t-1}E[X_{t-1}X_{t-1}^T]F_{t-1}^T + Q_{t-1} - F_{t-1}E[[P_{t-1}(X_{t-1})P_{t-1}(X_{t-1})^T]F_{t-1}^T] \\ &= C_{t|t-1} + Q_{t-1}, \end{aligned}$$

où

$$C_{t|t} = E[(X_t - P_t(X_t))(X_t - P_t(X_t))^T].$$

De même,

$$\begin{aligned} S_t &= E(v_t v_t^T) = G_t E[X_t X_t^T] G_t^T + R_t - G_t E[P_{t-1}(X_{t-1})P_{t-1}(X_{t-1})^T] G_t^T \\ &= G_t C_{t|t-1} G_t^T + R_t. \end{aligned}$$

Enfin,

$$\begin{aligned} C_{t|t} &= C_{t|t-1} - E[(P_t(X_t) - P_{t-1}(X_t))(P_t(X_t) - P_{t-1}(X_t))^T] \\ &= C_{t|t-1} - E[(K_t v_t)(K_t v_t)^T] \\ &= C_{t|t-1} - K_t S_t K_t^T. \end{aligned}$$

2.2.2 Le filtre d'information

Les formules précédentes mènent en parallèle le calcul de l'estimateur et celui de sa covariance d'erreur.

Dans certains cas, seul le calcul de covariance est intéressant, en particulier si l'on s'intéresse seulement à la performance potentielle du filtre. Il n'est alors pas utile de calculer la covariance de l'innovation et le gain du filtre, seules les valeurs successives de C nous intéressent.

Soit I la matrice inverse de C , pour toutes les valeurs des indices; I est appelée matrice d'information, elle est homogène à l'inverse d'une covariance.

La formule simple suivante donne l'évolution de I :

$$I_{t|t} = I_{t|t-1} + G_t^T R_{t-1} G_t$$

avec :

$$I_{t|t-1} = (F_{t-1} I_{t-1|t-1}^{-1} F_{t-1}^T + Q_{t-1})^{-1}$$

qui se réduit à :

$$I_{t|t-1} = (F_{t-1}^T)^{-1} I_{t-1|t-1} F_{t-1}^{-1}$$

dans le cas où il n'y a pas de bruit d'état. Dans ce cas, on obtient une formule globale pour

passer de $I_{t-1|t-1}$ à $I_{t|t}$:

$$I_{t|t} = (F_{t-1}^T)^{-1} I_{t-1|t-1} F_{t-1}^{-1} + G_t^T R_{t-1} G_t$$

Cette formule se simplifie encore si l'état est constant, soit si la matrice de transition F se réduit à l'identité :

$$I_{t|t} = I_{t-1|t-1} + G_t^T R_{t-1} G_t$$

Le terme $G_t^T R_{t-1} G_t$ est l'information apportée par une nouvelle mesure Y_t à la connaissance de X_t , elle est d'autant plus grande que l'erreur de mesure est plus petite.

Cette formule, qui peut être vérifiée à partir du lemme d'inversion matricielle, exprime le caractère additif de l'information.

Le cas d'une initialisation sans connaissance a priori se traduit en prenant $I_{0|0} = (0)_{v,v}$.

La Figure 4.1 illustre le fonctionnement du filtre de Kalman : soit une série égale à la somme d'une tendance linéaire et d'un bruit blanc . Cette série est représentée sur la figure de gauche. La figure de droite représente l'évolution au cours du temps de $P_t(Y_t) = G_t P_t(X_t)$ qui est dans ce cas une quantité scalaire (courbe continue en bleu), avec un intervalle de confiance à $\pm 1\sigma$ calculé à partir de sa variance $G_t C_{t|t} G_t^T$ (courbes pointillées en rouge).

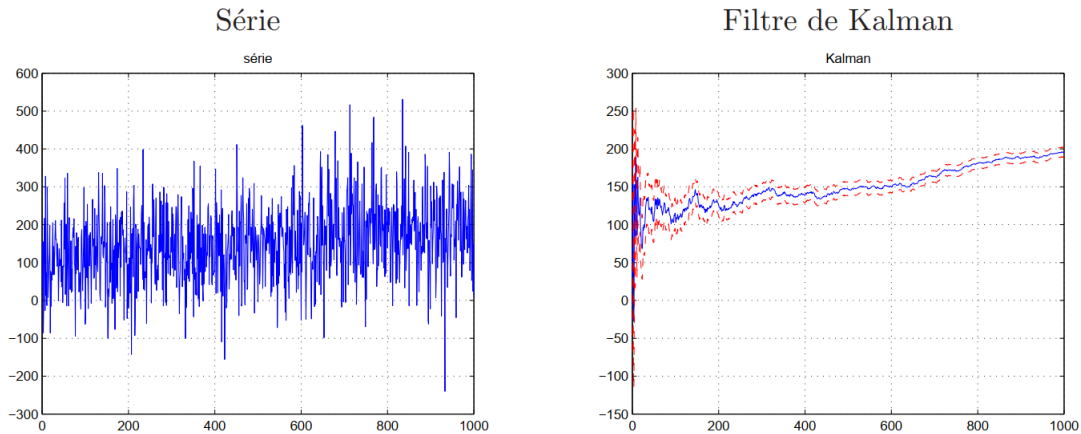


FIGURE 2.1 – Filtre de Kalman

2.3 Modèle d'état pour un processus ARMA

Les processus ARMA peuvent être représentés par des modèles d'état. On donne ici successivement la représentation par modèle d'état :

- ◇ d'un processus AR causal,
- ◇ d'un processus ARMA causal

2.3.1 Modèle d'état pour un AR(p)

Pour la cohérence avec les notations sur les modèles d'état, le processus AR(p) considéré est noté Y .

Si Y est un processus $AR(p)$, on a l'équation suivante :

$$Y_{t+1} = \phi_1 Y_t + \dots + \phi_p Y_{t-p+1} + Z_{t+1} \quad t \in \mathbb{Z}$$

Soit le vecteur d'état X de taille p défini par :

$$X_t = (Y_{t-p+1}, Y_{t-p+2}, \dots, Y_t)^T \quad t \in \mathbb{Z}$$

alors on peut vérifier que Y_t est un processus $AR(p)$ causal si l'on prend pour matrices (constantes) G et F :

$$G = [0 \ 0 \dots 1] \text{ de dimension } p$$

$$F = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ & & \dots & & \\ 0 & 0 & 0 & \dots & 0 \\ \phi_p & \phi_{p-1} & \dots & \phi_1 & \end{bmatrix}$$

de dimension (p, p)

2.3.2 Modèle d'état pour un ARMA(p,q)

Le processus $ARMA(p, q)$ considéré est encore noté Y . Y satisfait l'équation suivante :

$$\Phi(B)Y_t = \Theta(B)Z_t \quad t \in \mathbb{Z}$$

On définit le processus U_t qui est un $AR(p)$ causal par $\Phi(B)Y_t = Z_t$, soit :

$$Y_t = \Theta(B)U_t$$

Alors si l'on pose $r = \max(p, q + 1)$, on a :

$$Y_t = [\theta_{r-1}\theta_{r-2}\dots\theta_0]X_t$$

où X_t est le vecteur de dimension r :

$$X_t = (U_{t-r+1}, U_{t-r+2}, \dots, U_t)^T \quad t \in \mathbb{Z}$$

On peut alors étendre les équations précédentes en prenant : $G = [\theta_{r-1}\theta_{r-2}\dots\theta_0]$ de dimension r

$$F = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ & & \dots & & \\ 0 & 0 & 0 & \dots & 0 \\ \phi_r & \phi_{r-1} & \dots & \phi_1 & \end{bmatrix}$$

de dimension (r, r) et si l'on pose $V_t = (0, \dots, Z_{t+1})^T$ et $W_t = 0$.

On voit que :

- G contient des zéros si $r - 1 > q$, c'est-à-dire si $p > q + 1$,
- la dernière ligne de la matrice F contient des zéros si $r > p$, c'est-à-dire si $q + 1 > p$

Cette représentation d'un processus $ARMA$ par modèle d'état conduit à prendre

un vecteur d'état de dimension $r = \max(p, q + 1)$

Notons qu'il est possible d'avoir une représentation plus concise (de taille $\max(p, q)$), auquel cas W est non nul.

Cette représentation sera en particulier utilisée pour la prédiction, qui est une phase-clé de l'identification des processus.

Nota : il est nécessaire pour cela d'appliquer un filtrage de Kalman aux observations Y_t , $t = 1, \dots, n$ (voir initialisation et formules récursives). X_t étant lui-même stationnaire de moyenne nulle, défini pour $t \in \mathbb{Z}$, mais observé à des instants t tels que $t = 1, \dots, n$ ($Y_t = GX_t$), la partie initialisation doit être traitée de la façon suivante : prendre $P_0(X_0) = (0)_2$ (vecteur nul de taille 2) et $C_{0|0} = \text{Cov}(X_t) = E(X_t X_t^T)$ qui est indépendant de t , car X_t est stationnaire. Comme $X_t = (U_{t-r+1}, \dots, U_t)^T$ où U_t est le processus $AR(p)$ causal défini par $\Phi(B)U_t = Z_t$, on obtient

$$C_{0|0}(i, j) = \gamma U(|i - j|) \text{ pour } i, j = 1 \dots r.$$

2.4 Identification "robuste" d'un processus ARMA

On suppose ici que l'hypothèse sur l'ordre (p, q) a déjà été faite présente un schéma général pour l'estimation robuste d'un processus ARMA causal, appelé méthode de Box-Jenkins. Dans la suite, on décrit en détail l'estimation des coefficients de l'ARMA lorsque p et q sont imposés.

2.4.1 Cas d'un MA : l'algorithme des innovations

Cet algorithme permet d'estimer les coefficients θ d'un processus MA. Il peut être utilisé de plusieurs façons :

- cas d'un processus MA d'ordre connu : calcul des coefficients θ ,
- cas d'un processus ARMA, qu'on cherche dans un premier temps à identifier à un MA, comme d'écrit au début de ce paragraphe. Dans ce cas, les coefficients trouvés par l'algorithme des innovations sont les valeurs de ψ , puis on utilisera l'algorithme suivant pour en déduire ϕ et θ .

Détermination des valeurs de θ (ou ψ)

Initialisation :

$$\hat{v}_0 = \gamma_e(0), \hat{\theta}_{1,1} = \frac{1}{\hat{v}_0} \gamma_e(1), \hat{v}_1 = \gamma_e(0) - \hat{\theta}_{1,1}^2 \hat{v}_0$$

Formules récursives pour $m \geq 2$:

$$\begin{aligned}\hat{\theta}_{m,m} &= \frac{1}{\hat{v}_0} \gamma_e(m) \\ \hat{\theta}_{m,m-k} &= \frac{1}{\hat{v}_k} \left[\gamma_e(m-k) \sum_{j=0}^{k-1} \hat{\theta}_{m,m-j} \hat{\theta}_{k,k-j} \hat{v}_j \right] \quad \forall k \in \{1, \dots, m-1\} \\ \hat{v}_m &= \gamma_e(0) - \sum_{j=0}^{m-1} \hat{\theta}_{m,m-j}^2 \hat{v}_j.\end{aligned}$$

Les estimées de θ pour l'itération m sont données par le vecteur : $\hat{\theta}_m = (\hat{\theta}_{m,1}, \dots, \hat{\theta}_{m,m})$ et l'estimée de σ_Z^2 pour l'itération m est donnée par : $\sigma_Z^2 = \hat{v}_m$

Cet algorithme doit fonctionner jusqu'à une valeur de m suffisante pour que les coefficients estimés se "stabilisent".

2.4.2 L'algorithme de Durbin-Levinson

Les entrées et sorties sont décrites ci-dessous selon l'utilisation de l'algorithme.

$$\begin{aligned}\phi_{1,1} &= \frac{\gamma(1)}{\gamma(0)} \\ v_1 &= \gamma(0)(1 - \phi_{1,1}^2)\end{aligned}$$

Formules récursives pour $m \geq 2$

$$\begin{aligned}\phi_{m,m} &= \frac{1}{v_{m-1}} \left[\gamma(m) - \sum_{j=1}^{m-1} \phi_{m-1,j} \gamma(m-j) \right] \\ \phi_{m,k} &= \phi_{m-1,k} - \phi_{m,m} \phi_{m-1,m-k} \quad k = 1, \dots, m-1 \\ v_m &= v_{m-1} (1 - \phi_{m,m}^2)\end{aligned}$$

Dans le processus d'identification, cet algorithme peut être utilisé à deux niveaux :

- Calcul de la fonction de corrélation partielle empirique : l'entrée est la fonction de covariance empirique $\gamma_e(\cdot)$, la sortie est la séquence des $\phi_{m,m}$,
- Estimation du processus s'il est censé être un AR : c'est le même algorithme, on utilise en plus la sortie v_m comme estimateur de la variance de Z . Dans ce cas, on trouve les coefficients de l' $AR(p)$ dans la ligne $n - p$ (par ordre croissant d'indice). Les lignes suivantes (à partir de $m = p + 1$) contiennent les mêmes coefficients pour les p premiers, puis des zéros jusqu'à m ; la valeur de v_m reste constante à partir de $m = p$.

On verra par la suite que, si l'on utilise une méthode alternative au filtre de Kalman pour la prédiction à un pas, l'algorithme de Durbin Levinson est également utilisé à ce

niveau.

2.4.3 Algorithme de Yule Walker

Cet algorithme permet de calculer les mêmes coefficients $\phi_{m,k}$ que l'algorithme de Durbin Levinson, mais de façon non récursive à chaque itération m .

Soient (à chaque itération m) :

- Φ_m le vecteur (à déterminer) composé des coefficients $\phi_{m,k}$ pour $k = 1, \dots, m$,
- Γ_m la matrice (m, m) définie par $\Gamma_m = [\gamma_e(i - j)]_{i,j=1}^m$
- γ_m le vecteur constitué des valeurs de la fonction de covariance de 1 à m : $\gamma_m = (\gamma_e(1), \dots, \gamma_e(m))^T$

On a alors les relations suivantes :

$$\Gamma_m \Phi_m = \gamma_m, \quad (2.3)$$

$$v_m = \gamma_e(0) - (\Phi_m)^T \gamma_m, \quad (2.4)$$

On voit que la détermination des coefficients $\phi_{m,k}$ nécessite d'inverser une matrice de dimension m . L'algorithme de Yule Walker est donc utilisé essentiellement pour déterminer les coefficients d'un processus auto-régressif d'ordre p connu; le vecteur Φ_p contient précisément les coefficients cherchés.

2.4.4 Cas général d'un ARMA(p, q)

Un processus ARMA causal peut être écrit sous la forme d'un MA d'ordre infini :

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

avec $\psi_0 = 1$

$$\psi_j = \theta_j + \sum_{i=0}^{\min(j,p)} \phi_i \psi_{j-i}$$

où par convention $\psi_0 = 1$. Les ψ_j sont en nombre infini, mais sont liés aux coefficients ϕ et θ , dont le nombre est $p + q$, donc la connaissance des $p + q$ premières valeurs de ψ (ψ_1 à ψ_{p+q}) entraîne en général celle de toutes les autres (ψ_{p+q+1} à ψ_{∞}) et surtout suffit pour calculer les coefficients ϕ et θ .

On suppose qu'on a déjà déterminé l'ordre M d'un processus MA d'ordre fini qui modélise correctement le processus ARMA, mais cet ordre sera confirmé par l'algorithme des innovations. On a généralement $M \geq p + q$.

Pour la détermination des coefficients $\hat{\psi}_j$ et $\hat{\sigma}^2$, on pourra utiliser l'algorithme des 'innovations' présenté plus loin en calculant les $\theta_{m,m-k}$ suffisamment loin pour confirmer la valeur de l'ordre M qui satisfait les critères suivants : pour m suffisamment grand, les valeurs de $\hat{\theta}_{m,j}$ sont stables pour tout $j \leq M$ et petites pour $j > M$ et les valeurs de v_m sont stables. On pose alors $\hat{\psi}_j =$ valeur stabilisée de $\hat{\theta}_{m,j}$ pour $j = 1, \dots, M$ et $\hat{\sigma}^2 =$ valeur stabilisée de \hat{v}_m .

Il s'agit ensuite d'estimer ϕ et θ ($p + q$ valeurs) à partir des $\hat{\psi}_j$ pour $j = 1 \dots p + q$ (voir ci-dessous). On peut montrer que $\hat{\Phi}$ (le vecteur estimé des coefficients ϕ) satisfait l'équation matricielle :

$$\hat{\Psi}\hat{\Phi} = \hat{\psi}, \quad (2.5)$$

où $\hat{\Psi}$ est la matrice (p, p) dont les éléments sont :

$$\begin{aligned} \hat{\Psi}(i, j) &= \hat{\psi}_{q+i-j} \quad \text{si } q+i-j \geq 0 \\ \hat{\Psi}(i, j) &= 0 \quad \text{si } q+i-j < 0 \end{aligned}$$

et $\hat{\psi}$ est le vecteur colonne (p) : $\hat{\psi} = (\hat{\psi}_{q+1}, \dots, \hat{\psi}_{q+p})^T$

Les valeurs de $\hat{\theta}$ sont ensuite calculées au moyen des formules :

$$\hat{\theta}_j = \hat{\psi}_j - \sum_{i=1}^{\min(j,p)} \hat{\phi}_i \hat{\psi}_{j-i}, \quad j = 1, \dots, q.$$

2.4.5 Détails pour la prédiction

Dans le filtre de Kalman, le résidu est égal à l'innovation divisée par son écart-type :

$$r = \frac{v_t}{\sqrt{s_t}}$$

S est un scalaire positif, puisque la série est univariée.

Dans le schéma de la Figure 2.2, on utilise un filtrage de Kalman pour la prédiction. où la prédiction se fait en calculant la fonction de covariance estimée $\gamma(\cdot)$ en appliquant l'équation aux différences aux coefficients $\hat{\phi}$, $\hat{\theta}$, σ_Z^2 et $\hat{\psi}$, et en faisant la prédiction par l'algorithme de Durbin Levinson ou celui de Yule-Walker ou encore l'algorithme des innovations. Avec cette alternative, on peut donc utiliser pour la prédiction :

- Soit l'algorithme de Durbin Levinson (ou son équivalent non récursif l'algorithme de Yule Walker), auquel cas la fonction d'entrée est $\hat{\gamma}(\cdot)$, la sortie est l'ensemble des coefficients $\phi_{m,k}$ et la variance de l'erreur de prédiction est donnée par v_m . La formule qui donne la valeur de X prédite à l'instant $n + 1$ à partir des valeurs de Y_1 à Y_n est :

$$\hat{Y}_{n+1} = \sum_{j=1}^n \phi_{n,j} Y_{n+1-j}.$$

Proposition 2.2 *La vraisemblance L d'un processus ARMA gaussien, de moyenne nulle, de paramètres $(\phi, \theta, \sigma_Z^2)$, la vraisemblance L est donnée par la formule suivante :*

$$L(\phi, \theta, \sigma_Z^2) = (2\pi\sigma_Z^2)^{-n/2} (r_0 \dots r_{n-1})^{-1/2} \exp\left(\frac{-1}{2\sigma_Z^2} \sum_j \frac{(Y_j - \hat{Y}_j)^2}{r_{j-1}}\right)$$

où

$$\hat{Y}_{j+1} = \hat{Y}_{j+1}(\phi, \theta, \sigma_Z^2)$$

est la prédiction de Y_{j+1} sachant Y_1, \dots, Y_j , dans le modèle ARMA donné par les paramètres $(\phi, \theta, \sigma_Z^2)$, et les r_j sont les variances des erreurs de prédiction, divisées par la variance de Z :

$$r_j = r_j(\phi, \theta, \sigma_Z^2) = \frac{1}{2} v_j(\phi, \theta, \sigma_Z^2)$$

où

$$v_j(\phi, \theta, \sigma_Z^2) = E(Y_{j+1} - \hat{Y}_{j+1})^2.$$

La maximisation de L par rapport aux r paramètres conduit aux équations suivantes :

$$\hat{\sigma}_Z^2 = \frac{1}{n} s(\hat{\phi}, \hat{\theta}) \tag{2.6}$$

$$s(\hat{\phi}, \hat{\theta}) = \sum_j \frac{(Y_j - \hat{Y}_j)^2}{r_{j-1}} \tag{2.7}$$

où $\hat{\phi}$ et $\hat{\theta}$ sont les valeurs de ϕ et θ qui minimisent :

$$l(\phi, \theta) = \log\left(\frac{1}{n} s(\phi, \theta)\right) + \frac{1}{n} \sum_{j=1}^n \log(r_{j-1}) \tag{2.8}$$

où “log” désigne le logarithme népérien. La fonction $l(\phi, \theta)$ est appelée vraisemblance réduite.

2.4.6 Identification paramétrique d'un processus ARMA

La méthode exposée précédemment (identification “robuste”) est basée entièrement sur la fonction de covariance empirique. Lorsque les ordres du modèle ARMA (p et q) sont supposés connus, l'estimation du processus consiste à trouver $r = p + q + 1$ paramètres $(\phi, \theta$ et $\sigma_Z^2)$. Une méthode directe d'estimation de ces paramètres consiste à trouver l'ensemble des r paramètres les plus vraisemblables, compte tenu des observations $(Y_1 \dots Y_n)$. Pour cela, la technique classique du maximum de vraisemblance est

applicable : calculer la vraisemblance L (likelihood) des observations $(Y_1 \dots Y_n)$ en fonction des r paramètres $((\phi, \theta \text{ et } \sigma_Z^2))$ et trouver l'ensemble des paramètres qui maximise L .

2.4.7 Validation d'un modèle : critères de contrôle sur le résidu

Au cours de la prédiction d'un processus ARMA, le résidu est défini comme la série temporelle :

$$\hat{W}_t = \frac{Y_t - \hat{Y}_t(\hat{\phi}, \hat{\theta})}{\sqrt{r_{t-1}(\hat{\phi}, \hat{\theta})}}$$

Lorsque le processus est estimé correctement, la série des résidus suit asymptotiquement à peu près la même loi que le processus générateur Z_t (bruit blanc faible en général). On peut donc tester la qualité de l'estimation en observant la fonction de corrélation empirique $\hat{\rho}(h)$ de \hat{W}_t . Celle-ci vaut 1 pour $h = 0$, et pour $h \geq 1$, elle est comprise entre $-\frac{1.96}{\sqrt{n}}$ et $\frac{1.96}{\sqrt{n}}$ avec une probabilité de 95%. C'est le test de "blancheur" de Bartlett. Lorsque l'estimation du processus a été obtenue par maximisation de la vraisemblance, les intervalles de confiance peuvent être affinés (c'est-à-dire diminués) pour les h petits. L'application de ce critère permet de dire si le modèle ARMA choisi (avec ses ordres et ses paramètres) modélise correctement le processus. Il est nécessaire pour le calculer de mettre en oeuvre une prédiction du processus, comme décrit dans l'approche générale.

2.4.8 Méthodes de sélection

On décrit ici des méthodes permettant de comparer des modèles ARMA entre eux :

- le critère AICC (Akaike Information Corrected Criterion) qui permet, par une méthode de pénalisation, d'éviter le phénomène d'overfitting,
- une méthode générale de sélection en "retenant" une partie des données pour choisir parmi plusieurs modèles.

Prédiction linéaire à h pas

Les techniques de prédiction à un pas étudiées précédemment (par filtrage de Kalman, ou comme combinaison linéaire des Y_i ou par l'algorithme des innovations) s'étendent à la prédiction à h pas. Il s'agit de prédire la valeur de Y_{n+h} à partir des valeurs de Y_1, \dots, Y_n . Quand h augmente pour n fixé, \hat{Y}_{n+h} tend vers 0 et la variance de l'erreur d'estimation tend vers σ_Y^2 .

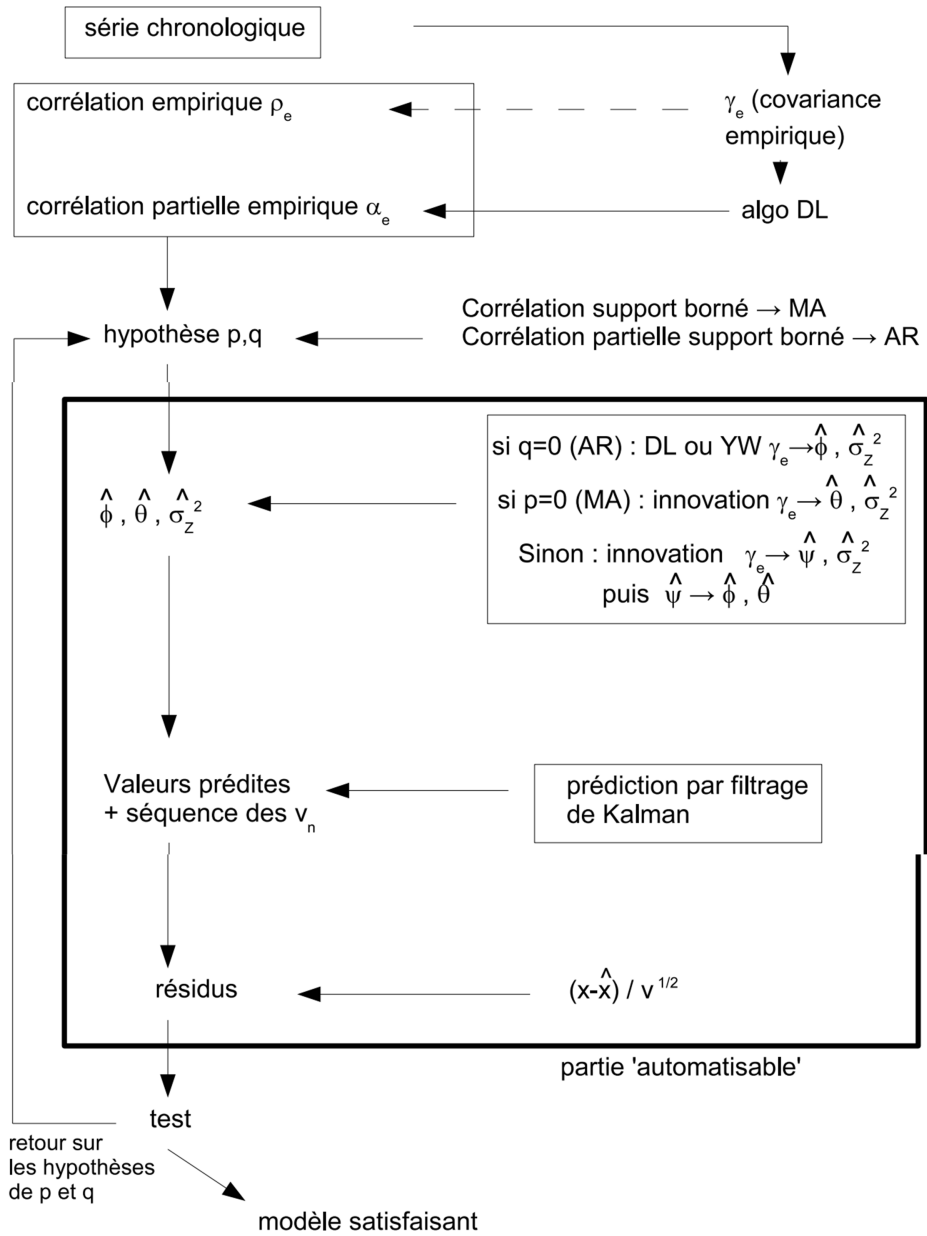


FIGURE 2.2 – Schéma de prédiction par filtre de Kalman

CHAPITRE 3

APPLICATION SUR LA PRÉVISION DU COVID19

Les virus Corona ont été découverts dans les années 1960, et les premiers virus découverts étaient le virus de la bronchite infectieuse chez les poulets et deux virus de la cavité nasale de patients humains atteints d'un rhume, appelés virus corona humain 229E et virus corona humain OC43. Depuis, d'autres éléments de cette famille ont été identifiés, notamment : le SARS Coronavirus en 2003, le Human Coronavirus NL63 en 2004, le Human Coronavirus HKU1 en 2005, le Coronavirus Mers en 2012 et le New Corona Virus 2019-nCoV, et la plupart de ces virus Il a un rôle à jouer dans une infection respiratoire grave et même la mort.

Les coronavirus sont une large famille de virus qui peuvent provoquer des maladies chez les animaux et les humains. Il est connu qu'un certain nombre de virus corona chez l'homme provoquent des maladies respiratoires allant du rhume à des maladies plus graves telles que le syndrome respiratoire du Moyen-Orient (MERS) et le syndrome respiratoire aigu sévère (SRAS). Le coronavirus récemment découvert provoque la maladie Covid-19. Ce virus est apparu dans la ville chinoise de Wuhan en décembre 2019 et se caractérise par sa propagation rapide parmi les personnes. Les gens peuvent attraper l'infection Covid-19 par d'autres personnes infectées par le virus. La maladie se transmet principalement d'une personne à l'autre par de petites gouttelettes qu'une personne atteinte de Covid-19 sécrète par le nez ou la bouche lorsqu'elle tousse, éternue ou parle. Ces gouttelettes ont un poids relativement lourd, car elles ne se déplacent pas vers un endroit lointain, mais tombent plutôt rapidement au sol. Les gens peuvent contracter la maladie Covid-19 s'ils respirent ces gouttelettes d'une personne infectée par le virus.

Le nom anglais de la maladie est dérivé comme suit : «CO» correspond aux deux premières lettres du mot «CORONA» (Le nom fait référence à l'apparence distinctive d'une couronne). Quant aux lettres «VI», elles sont dérivées des deux premières lettres du mot «virus» et la lettre «D» est la première lettre du mot «diseas», selon un rapport.

Publié par l'UNICEF des Nations Unies.

3.1 Analyse de COVID19 en algérie

La propagation d'un virus COVID 19 en Algérie, comme dans le reste des pays, où l'Algérie a enregistré le jour 26/02/2020. Pour assister à sa propagation rapide dans l'État de Blida, puis le reste dans les États du pays d'origine jusqu'à ce que les hôpitaux deviennent incapables d'accueillir tous les patients, nous allons donc aborder une prédiction sur le nombre d'infections.

3.1.1 La prédiction sur le nombre d'infections par COVID19

Les données sont téléchargées depuis le site officiel de l'Organisation mondiale de la santé (OMS) du 01/01/2020 au 16/09/2020

statistique discrèptive

TABLE 3.1 – Discription de la série des cas d'infections en algérie

Min	1st Qu	Median	Mean	3rd Qu	Max
0.0	1.5	129.0	191.1	330.5	675

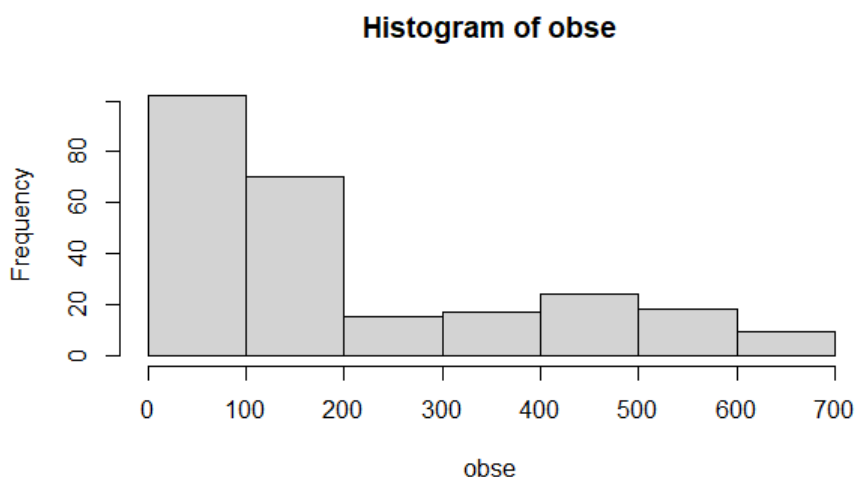


FIGURE 3.1 – Histogramme des observations des infectés

La série temporelle

obse : la série chronologique représente le nombre d'infections par COVID19

Analyse graphique de la série "cas d'infections par COVID19"

A partir des données de la série,et avec la commandes suivante : `plot.ts(obse)`.

Nous obtenons le graphique suivant :

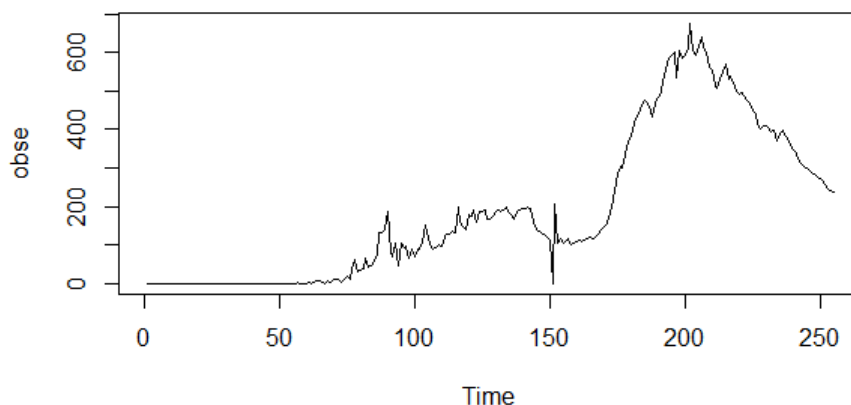


FIGURE 3.2 – Représentation graphique de la série cas de d’infection par COVID19

Le corrélogramme simple :

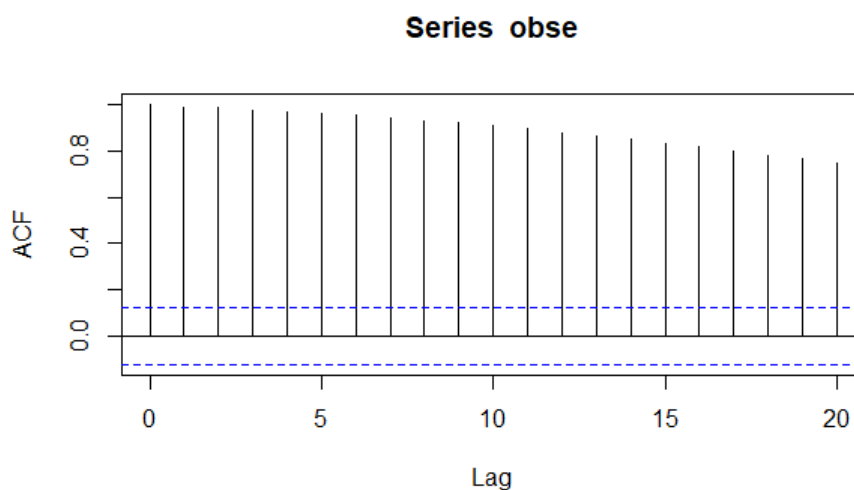


FIGURE 3.3 – Représentation graphique d’autocorrélation de la série cas de d’infection par COVID19

Et sa corrélogramme partielle :

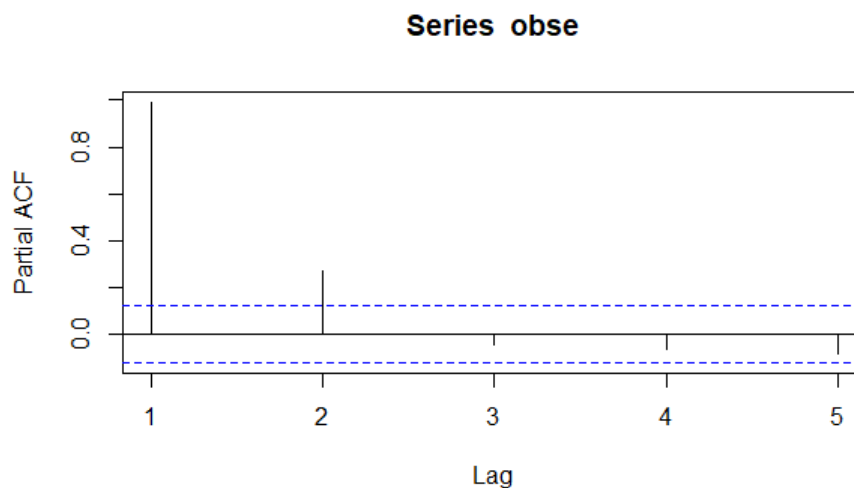


FIGURE 3.4 – Représentation graphique d'autocorrélation partielle de la série obse

La visualisation des corrélogrammes simple et partielle montre non stationnarité, l'ACF ne coupe pas.

On utilise le test de racine unitaire (test ADF) pour confirmer non stationnarité :

```
library(tseries)
```

```
adf.test(obse)
```

Augmented Dickey-Fuller Test

```
data : obse
```

```
Dickey-Fuller = -1.2053, Lag order = 6, p-value = 0.904
```

p-value = 0.904 > 0.05 donc la série n'est pas stationnaire (admet une racine unitaire)

On peut aussi utiliser la fonction **decompose** pour extraire les composantes d'une série temporelle

```
d=decompose(obse,type = "m")
```

```
plot(d)
```

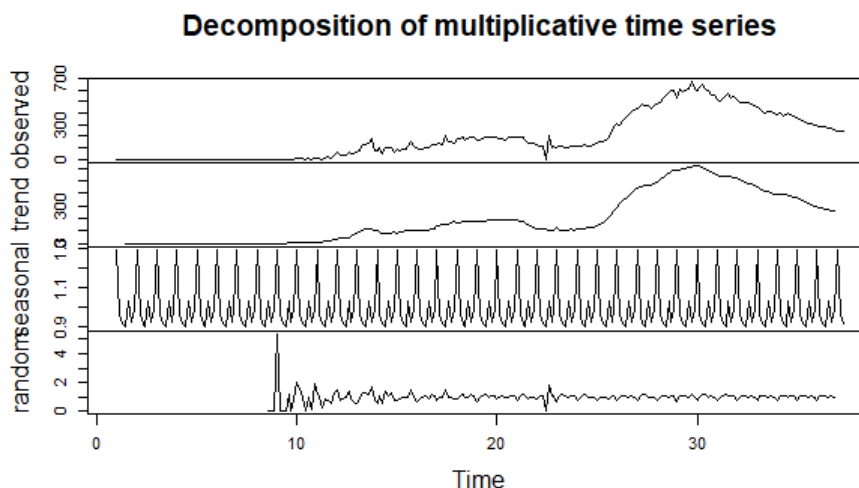


FIGURE 3.5 – les composantes d’une série temporelle

Transformation de la série

On utilise la commande **ndiffs** pour savoir l’ordre de différenciation qui élimine la tendance et la saisonnalité.

```
library(forecast)
```

```
ndiffs(obse)
```

```
[1] 1
```

Alors l’ordre de différenciation est 1.

```
d1=diff(obse,differences = 1)
```

```
d1
```

```
plot.ts(d1)
```

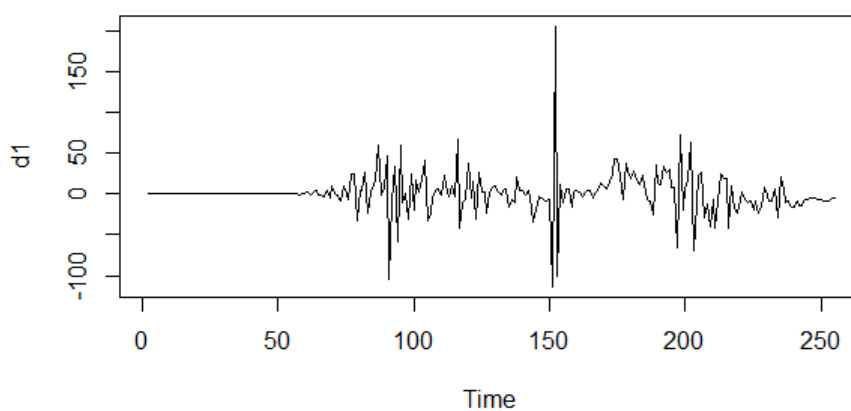


FIGURE 3.6 – Représentation graphique de la série obse différenciation

On applique l’ensemble des tests de la stationnarité on obtient :

```
library(tseries)
```

```
— kpss.test(d1)
```

KPSS Test for Level Stationarity

```
data : d1
```

KPSS Level = 0.22829, Truncation lag parameter = 5, p-value = 0.1

```
— PP.test(d1)
```

Phillips-Perron Unit Root Test

```
data : d1
```

Dickey-Fuller = -21.866, Truncation lag parameter = 5, p-value = 0.01

```
— adf.test(d1)
```

Augmented Dickey-Fuller Test

```
data : d1
```

Dickey-Fuller = -4.7187, Lag order = 6, p-value = 0.01

D'après les résultats des trois tests on résume que la série différenciée est stationnaire

Identification

les graphes de autorrélation simple et partiel de la série différenciée sont :

```
acf(d1)
```

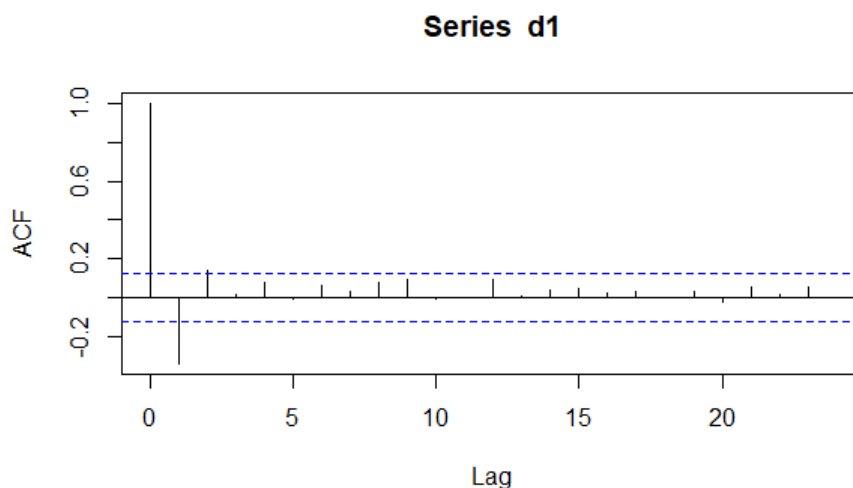


FIGURE 3.7 – Autorrélation simple de la série obsse différencier

```
pacf(d1)
```

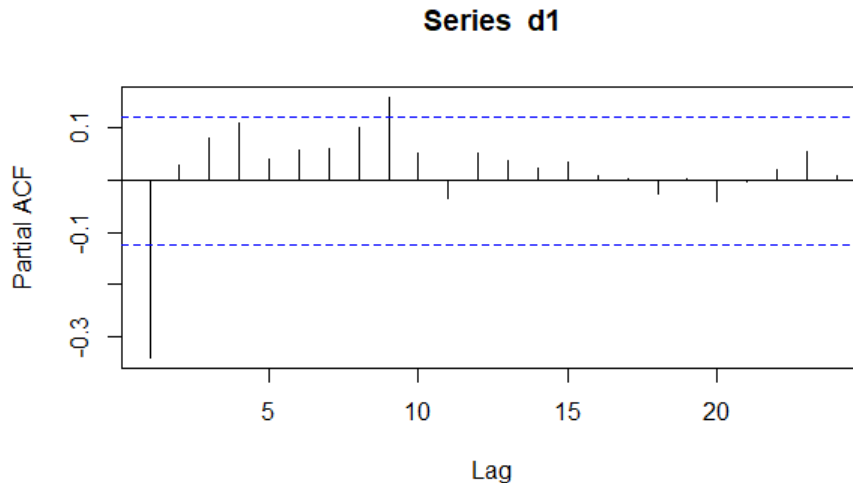


FIGURE 3.8 – Autorrélation partielle de la série différencier

nous voyons que ACF est coupé après décalage 1 et que PACF est coupé après décalage 1.

Nous proposons donc les modèles : ARIMA(1,1,1), ARIMA(2,1,1), ARIMA(1,1,2)

3.2 Test de tendance Mann-Kendall dans R

Un test de tendance de Mann-Kendall est utilisé pour déterminer si une tendance existe ou non dans les données de séries chronologiques. Il s'agit d'un test non paramétrique, ce qui signifie qu'aucune hypothèse sous-jacente n'est faite sur la normalité des données.

Les hypothèses du test sont les suivantes :

H_0 (hypothèse nulle) : aucune tendance n'est présente dans les données.

H_A (hypothèse alternative) : une tendance est présente dans les données. (Cela peut être une tendance positive ou négative)

Si la valeur p du test est inférieure à un certain niveau de signification (les choix courants sont 0,10, 0,05 et 0,01), alors il existe des preuves statistiquement significatives qu'une tendance est présente dans les données de la série chronologique.

Pour effectuer un test de tendance Mann-Kendall dans R, nous utiliserons la fonction `MannKendall()` de la bibliothèque `Kendall`, qui utilise la syntaxe suivante :

```
MannKendall(x)
```

On obtient les résultats suivantes :

```
MannKendall(x)
```

```
tau = 0.724 2-sided pvalue = 2.22e-16
```

La statistique du test est de 0,724 et la valeur p bilatérale correspondante est de $2.22e-16$. Parce que cette p -valeur est inférieure à 0,05, nous rejeterons l'hypothèse

nulle du test et concluons qu'une tendance est présente dans les données.

Pour visualiser la tendance, nous pouvons créer un graphique de série chronologique des précipitations annuelles par année et ajouter une ligne lisse pour représenter la tendance :

```
#Plot the time series data
#Add a smooth line to visualize the trend
lines(lowess(time(x),x), col='blue')
```

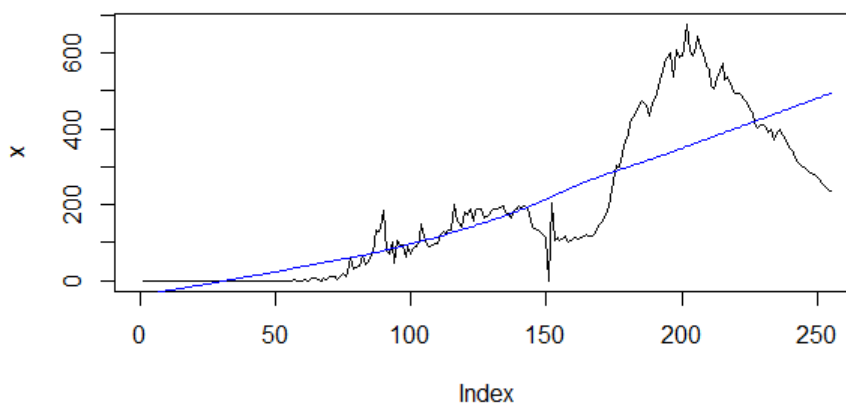


FIGURE 3.9 – Analyse de la tendance de la séries des infectés

Notez que nous pouvons également effectuer un test de tendance Mann-Kendall désaisonnalisé pour tenir compte de toute saisonnalité dans les données à l'aide de la commande

```
SeasonalMannKendall (x) :
```

3.3 Prévision par la méthode de Box-Jenkins

3.3.1 Identification du modèle

3.3.1.1 Estimation des paramètres

On utilise la commande **auto.arima** pour construire le meilleur modèle qui représente la série

```
library(forecast)
modobse=auto.arima(obse,trace = TRUE,test = "kpss",ic="aic")
on trouve
```

TABLE 3.2 – Résultats d’ajustements de la série des cas d’infections par COVID19

Modèle	AIC
ARIMA(2,1,2)(1,0,1)[7] with drift	2338.37
ARIMA(1,1,0)(1,0,0)[7] with drift	2339.681
ARIMA(0,1,0) with drift	2361
ARIMA(0,1,0)	2359.346
ARIMA(0,1,1) with drift	2337.403
ARIMA(0,1,1)(1,0,0)[7] with drift	2344.215
ARIMA(0,1,1)(1,0,1)[7] with drift	2345.213
ARIMA(1,1,1) with drift	2334.554
ARIMA(1,1,1)(1,0,0)[7] with drift	2341.662
ARIMA(1,1,1)(0,0,1)[7] with drift	2334.757
ARIMA(1,1,1)(1,0,1)[7] with drift	2342.913
ARIMA(1,1,0) with drift	2332.714
ARIMA(1,1,2)	2323.526
ARIMA(1,1,2)(1,0,0)[7]	2332.651
ARIMA(1,1,2)(1,0,1)[7]	2334.587
ARIMA(1,1,1)	2333.235
ARIMA(0,1,2)	2329.716
ARIMA(1,1,1)	2333.235

TABLE 3.3 – Résultats d’ajustements de la série des cas d’infections par COVID19

Modèles	AIC
ARIMA(2,1,2)(1,0,1)[7] with drift	2338.37
ARIMA(1,1,0)(1,0,0)[7] with drift	2339.681
ARIMA(0,1,0) with drift	2361
ARIMA(0,1,0)	2359.346
ARIMA(0,1,1) with drift	2337.403
ARIMA(0,1,1)(1,0,0)[7] with drift	2344.215
ARIMA(0,1,1)(1,0,1)[7] with drift	2345.213
ARIMA(1,1,1) with drift	2334.554
ARIMA(1,1,1)(1,0,0)[7] with drift	2341.662
ARIMA(1,1,1)(0,0,1)[7] with drift	2334.757
ARIMA(1,1,1)(1,0,1)[7] with drift	2342.913
ARIMA(1,1,0) with drift	2332.714
ARIMA(1,1,0)(0,0,1)[7] with drift	2332.79
ARIMA(1,1,0)(1,0,1)[7] with drift	2340.915
ARIMA(2,1,0) with drift	2335.486
ARIMA(2,1,1) with drift	2327.738
ARIMA(2,1,1)(1,0,0)[7] with drift	2336.751
ARIMA(2,1,1)(0,0,1)[7] with drift	2329.602
ARIMA(2,1,2) with drift	2327.4
ARIMA(2,1,2)(1,0,0)[7] with drift	2336.486
ARIMA(2,1,2)(0,0,1)[7] with drift	2329.335
ARIMA(1,1,2) with drift	2325.471
ARIMA(1,1,2)(1,0,0)[7] with drift	2334.593
ARIMA(1,1,2)	2323.526
ARIMA(1,1,2)(1,0,0)[7]	2332.651
ARIMA(1,1,2)	2323.526
ARIMA(1,1,2)(1,0,0)[7]	2332.651
ARIMA(1,1,2)(1,0,1)[7]	2334.587
ARIMA(1,1,1)	2333.235

Méilleur modèles est : **ARIMA(1,1,2)**

La commande **summary(modobse)**

TABLE 3.4 – Estimation des coefficients du modèle d’ajustement

	ar1	ma1	ma2
coef	0.9429	-1.3004	0.3992
s.e	0.0353	0.0664	0.0553
AIC	2329.45		

3.3.2 Validation

Nous avons diagnostiqué notre modèle à partir des tests suivants pour montrer que le modèle choisis est valide.

```

— Test ADF
adf.test(modobse.residuals)
Augmented Dickey-Fuller Test
data : modobse$residuals
Dickey-Fuller = -6.2286, Lag order = 6, p-value = 0.01
— Test de Box-pierce
Box.test(modobse$residuals)
Box-Pierce test
data : modobse$residuals
X-squared = 0.14209, df = 1, p-value = 0.7062
— test de la stationnarité graphiquement
    
```

3.3.3 Prévision

D'après le modèle choisis précédent on peut prédire des résultats pour les prochaines jours

On utilise l'ensemble des commandes suivantes :

```

nse=auto.arima(obse)
ff=forecast(nse,h=10)
summary(ff)
    
```

TABLE 3.5 – Résultats du prévision par la méthode de Box-Jenkins

jour	prévision	Lo 80	Hi 80	Lo 95	Hi 95
256	233.6638	203.5801	263.7476	187.65467	279.6730
257	229.4237	193.6658	265.1816	174.73676	284.1107
258	225.4259	183.8568	266.9950	161.85144	289.0004
259	221.6565	174.1675	269.1454	149.02842	294.2845
260	218.1023	164.6094	271.5953	136.29198	299.9127
261	214.7513	155.1911	274.3115	123.66179	305.8408
262	211.5917	145.9188	277.2645	111.15366	312.0297
263	208.6125	136.7970	280.4281	98.78018	318.4449
264	205.8036	127.8287	283.7785	86.55130	325.0559
265	203.1552	119.0155	287.2948	74.47472	331.8356

— Graphe représente la prévision avec l'intervalle de confiance : avec la commande plot(ff)

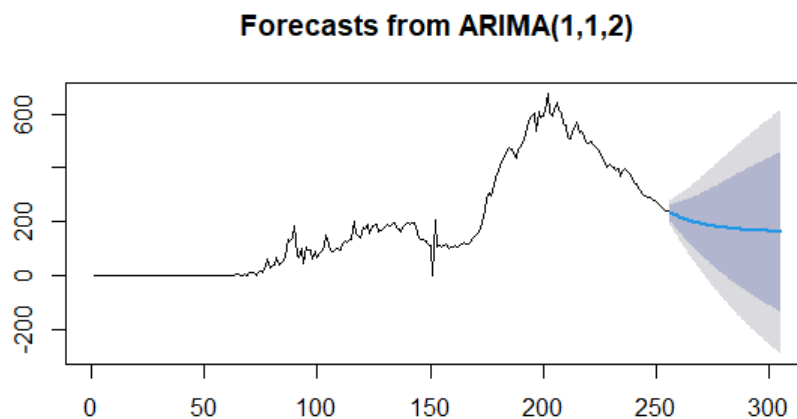


FIGURE 3.10 – Représentation graphique de la prévision par Box-Jenkins

3.4 Prévision par filtre de Kalman

Nous utilisons le package (TSPred) pour la prévision par filtre de Kalman
`library(TSPred)`

La représentation graphique de la série , nous utilisons la commande suivante :
`plot(obse,type='l',lwd=1,xlim=c(0,256),ylim=c(0,800),xlab="Time",ylab="ARIMA")`

Nous obtenons le graphe suivant :

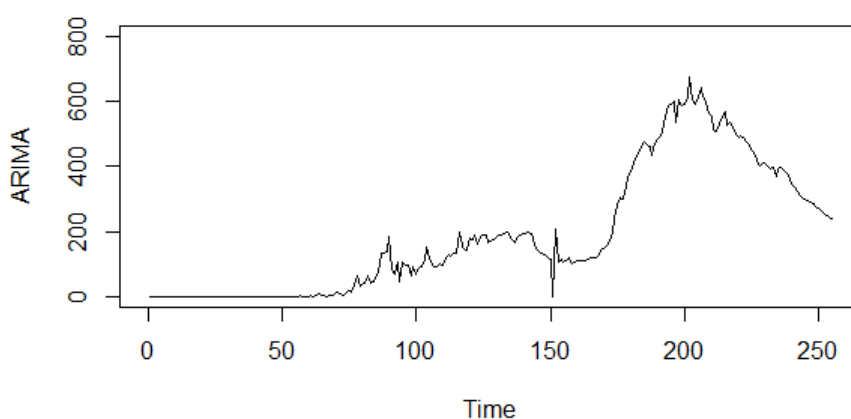


FIGURE 3.11 – Représentation graphique des observations

nous utilisons les commandes suivantes :

`fPolyRKF= fittestPolyRKF(obse,h=10)`

`fs<-KFAS::KFS(fPolyRKF,model,filtering=c("state","mean"),smoothing=c("state","mean"))`

Série chronologique filtrée par Kalman

```
f<-fitted(fs, filtered=TRUE)
```

```
s<-lines(f,col='red',lty=1,lwd=1)
```

Nous obtenons

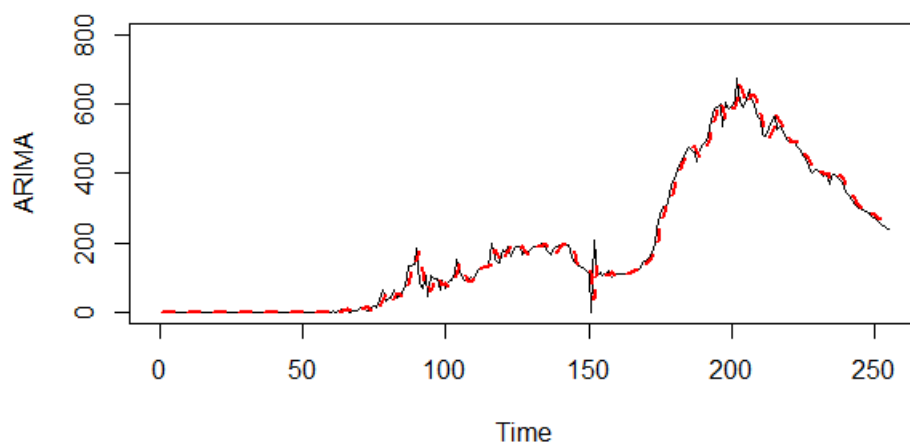


FIGURE 3.12 – Représentation graphique des observations

#Série chronologique lissée par Kalman

```
s=fitted(fs)
```

```
lines(s,col='green',lty=1,lwd=1)
```

Nous obtenons :

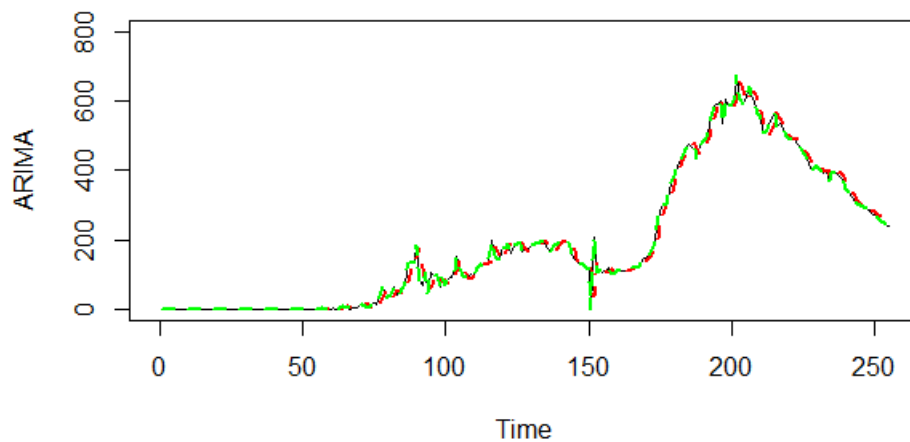


FIGURE 3.13 – Série chronologique ajustée par Filtre de Kalman

Nous utilisons les commandes suivantes pour la prévision

```
pred<-fPolyRKFpredmean
```

```
pred
```

Nous obtenons :

Time Series :

Start = 256

End = 265

TABLE 3.6 – Résultats du prévision pour dix jours avec les bornes de confiances

date	jour	La prévision	Borne inférieure	Borne supérieure
17/09/2020	256	231.4074	189.066601	273.7481
18/09/2020	257	224.6786	163.745674	285.6114
19/09/2020	258	217.8135	141.873512	293.7536
20/09/2020	259	210.8123	121.583139	300.0415
21/09/2020	260	203.6749	102.162255	305.1876
22/09/2020	261	196.4014	83.250236	309.5525
23/09/2020	262	188.9916	64.635231	313.3480
24/09/2020	263	181.4456	46.180518	316.7107
25/09/2020	264	173.7635	27.791877	319.7351
26/09/2020	265	165.9451	9.401146	322.4891

3.4.1 Représentation graphique de la prévision

```
lines(ts(pred,start=256),lwd=1,col='blue')
```

```
lines(ts(fPolyRKFpredupper,start=256),lwd=1,col='light blue')
```

```
lines(ts(fPolyRKFpredlower,start=256),lwd=1,col='light blue')
```

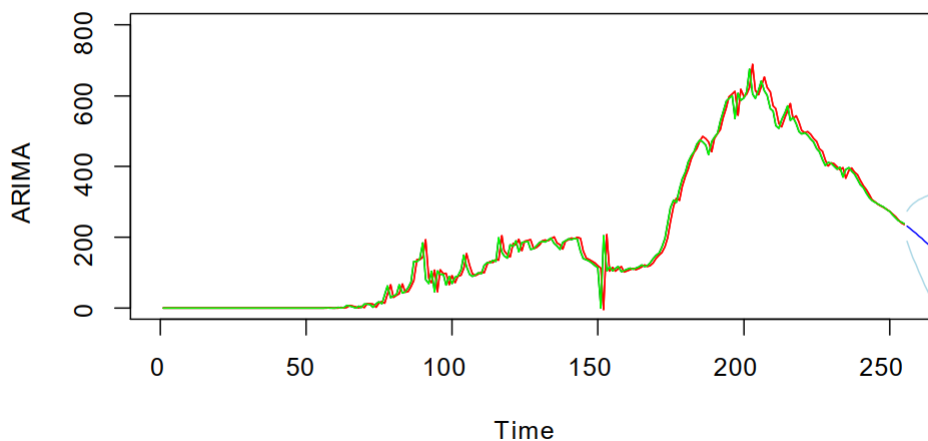


FIGURE 3.14 – Représentation graphique de la prévision par filtre de Kalman

3.5 Prédiction du nombre de décès dus à le virus COVID19 en Algérie

3.5.1 Résultats de la Statistique descriptive

TABLE 3.7 – Résultats descriptive des décès à cause de COVID-19 en Algérie

Min	1st Qu	Median	Mean	3rd Qu	Max
0.0	0.0	7.0	6.4	9.0	42.0

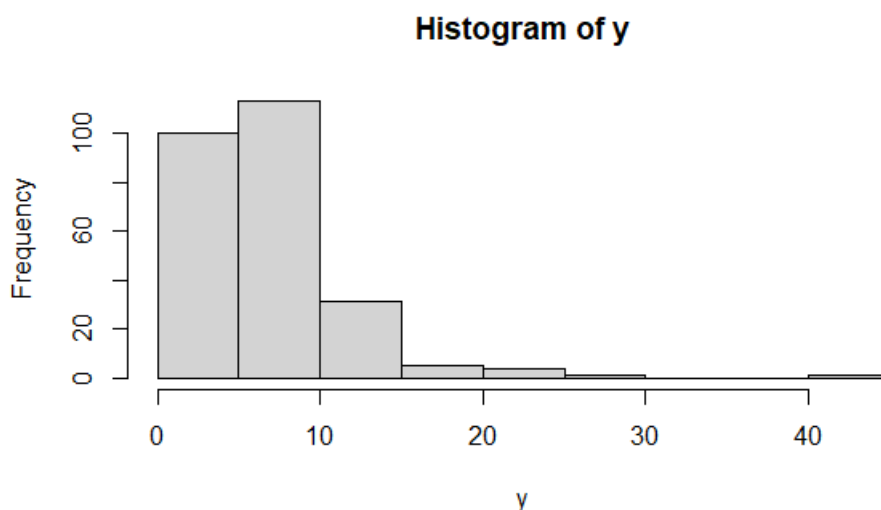


FIGURE 3.15 – Histogramme de la série des décès en Algérie par COVID-19

3.5.2 Analyse de la série temporelle

y : la série chronologique des décès par COVID19

Analyse graphique de la série "cas des décès par COVID19"

A partir des données de la série, et avec la commandes suivante : `plot.ts(y)`.

Nous obtenons le graphique suivant :

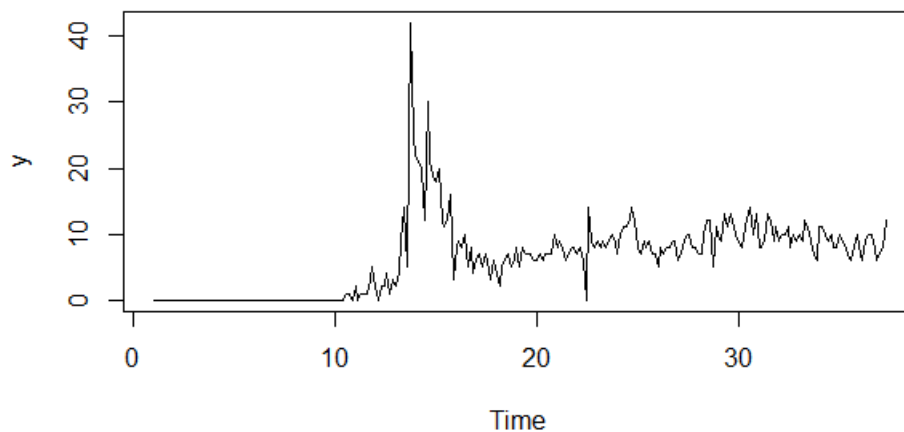


FIGURE 3.16 – Représentation graphique de la série des décès en Algérie par COVID-19

Le corrélogramme simple :

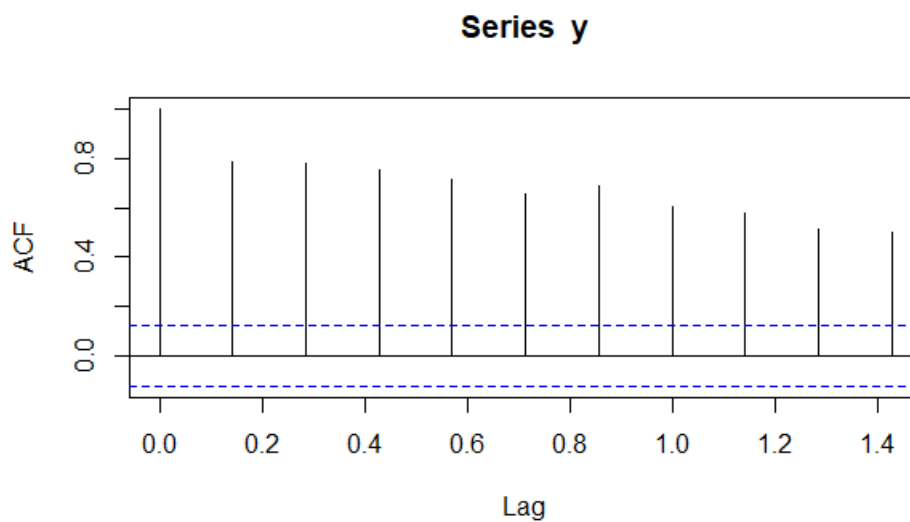


FIGURE 3.17 – Autocorrélation simple ACF de la série des décès en Algérie par COVID-19

Le corrélogramme partielle :

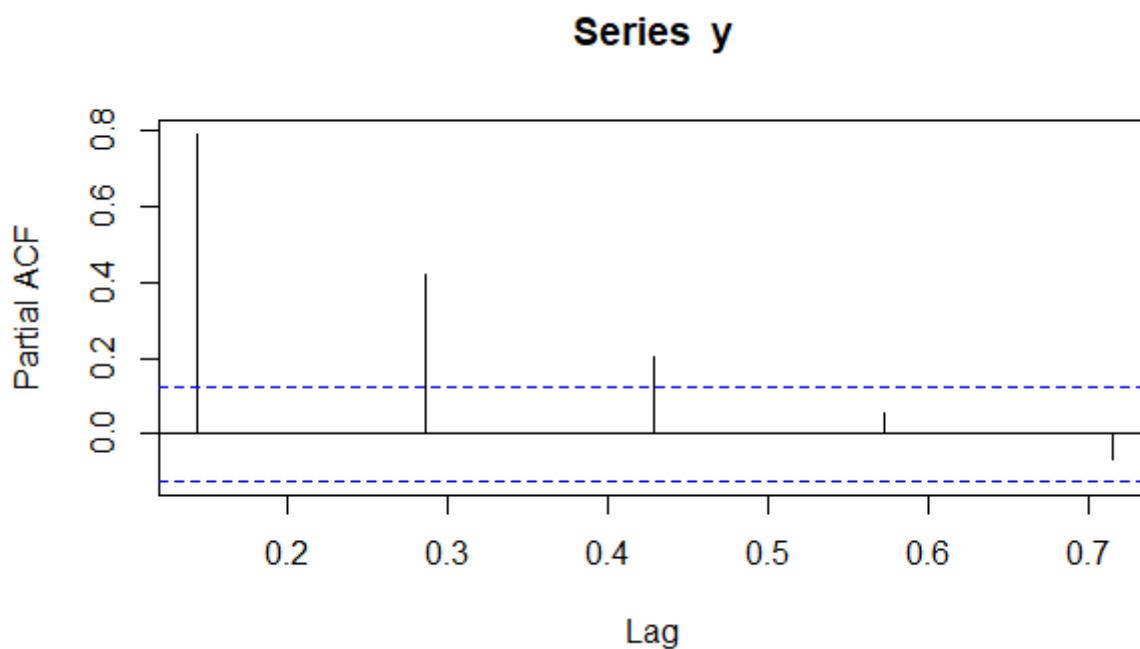


FIGURE 3.18 – Autocorrélation partielle PACF de la série des décès en Algérie par COVID-19

On peut aussi utiliser la fonction `decompose` pour extraire les composantes d'une série temporelle :

```
dec=decompose(y)
plot(dec)
```

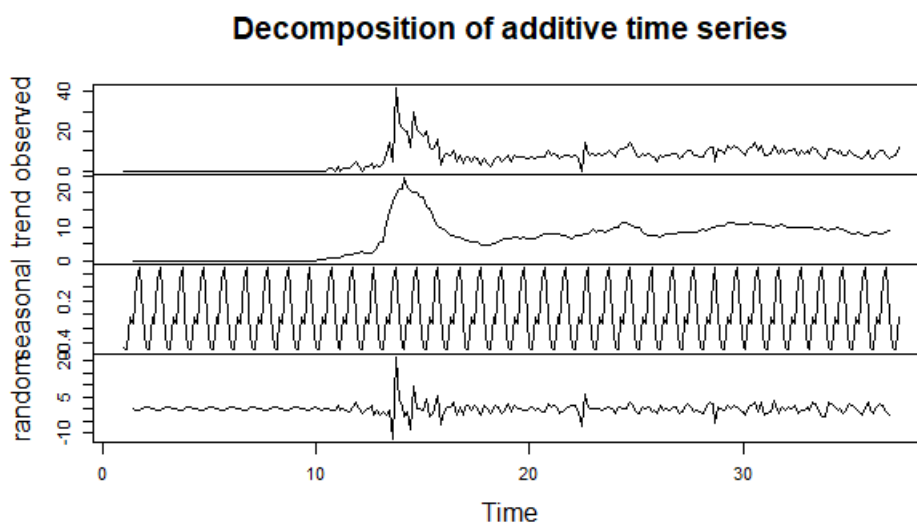


FIGURE 3.19 – Décomposition de la série des décès en Algérie par COVID-19

Transformation de la série

On utilise la commande **ndiffs** pour savoir l'ordre de différenciation qui élimine la tendance et la saisonnalité.

```
library(forecast)
ndiffs(Y)
[1] 1
df=diff(Y,differences = 1)
df
```

La représentation graphique est

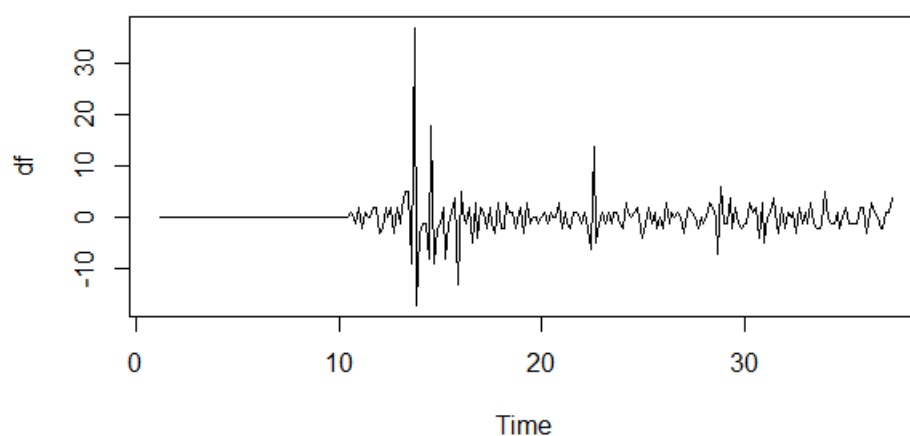


FIGURE 3.20 – Représentation graphique de la série différenciée

On applique l'ensemble des tests de la stationnarité on obtient :

```
library(tseries)
— kpss.test(df)
```

KPSS Test for Level Stationarity

data : df

KPSS Level = 0.027417, Truncation lag parameter = 5, p-value = 0.1

```
— PPtest(df)
```

Phillips-Perron Unit Root Test

data : df

Dickey-Fuller = -32.248, Truncation lag parameter = 5, p-value = 0.01

```
— adf.test(df)
```

Augmented Dickey-Fuller Test

data : df

Dickey-Fuller = -6.6778, Lag order = 6, p-value = 0.01

D'après les résultats des trois tests on résume que la série différenciée est stationnaire

Identification

les graphes de autorrélation simple et partielle de la série différenciée sont :

acf(df)

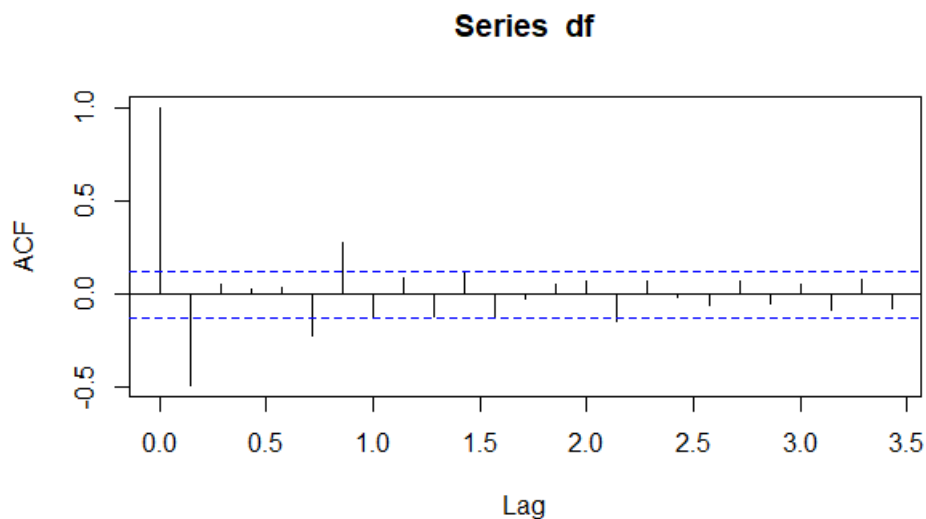


FIGURE 3.21 – Graphe de autorrélation simple de la série y différenciée

pacf(df)

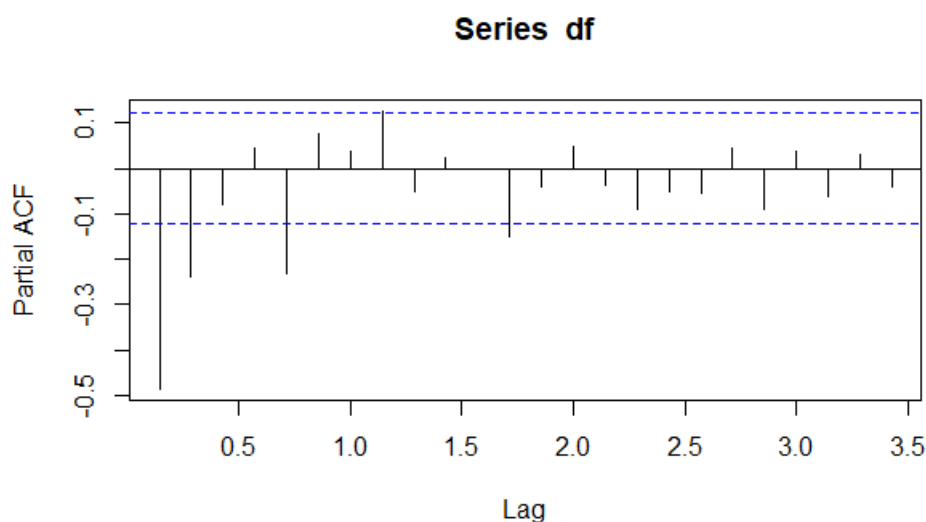


FIGURE 3.22 – Graphe de autorrélation partielle de la série y différenciée

Nous voyons que ACF est coupé après décalage 1 et que PACF est coupé après décalage 2.

Nous proposons donc les modèles : ARIMA(2,1,1)

3.5.3 Application de la méthode de Box-Jenkins

3.5.3.1 Estimation des paramètres

On utilise la commande `auto.arima` pour construire le meilleur modèle qui représente la série

`library(forecast)`

`ym=auto.arima(Y,trace = TRUE,test = "kpss",ic="aic")`

on trouve

TABLE 3.8 – Résultats d’ajustement de la série des décès par un modèle temporelle

Modèle	AIC
ARIMA(2,1,2)(1,0,1)[7] with drift :	1322.827
ARIMA(0,1,0) with drift :	1387.721
ARIMA(1,1,0)(1,0,0)[7] with drift :	1331.33
ARIMA(0,1,1)(0,0,1)[7] with drift :	1308.11
ARIMA(0,1,0) :	1385.762
ARIMA(0,1,1) with drift :	1306.184
ARIMA(0,1,1)(1,0,0)[7] with drift :	1315.199
ARIMA(0,1,1)(1,0,1)[7] with drift :	1317.161
ARIMA(1,1,1) with drift :	1307.127
ARIMA(0,1,2) with drift :	1306.047
ARIMA(0,1,2)(1,0,0)[7] with drift :	1315.123
ARIMA(0,1,2)(0,0,1)[7] with drift :	1308.031
ARIMA(0,1,2)(1,0,1)[7] with drift :	1316.927
ARIMA(1,1,2) with drift :	1309.047
ARIMA(0,1,3) with drift :	1308.045
ARIMA(1,1,3) with drift :	1311.045
ARIMA(0,1,2) :	1304.222
ARIMA(0,1,2)(1,0,0)[7] :	1313.3
ARIMA(0,1,2)(0,0,1)[7] :	1306.203
ARIMA(0,1,2)(1,0,1)[7] :	1315.056
ARIMA(0,1,1) :	1304.384
ARIMA(1,1,2) :	1307.223
ARIMA(0,1,3) :	1306.22
ARIMA(1,1,1) :	1305.307
ARIMA(1,1,3) :	1309.22

ARIMA(0,1,2) : 1307.26

Best model : ARIMA(0,1,2)

On utilise la commande

`summary(ym)`

on trouve

Series : y

ARIMA(0,1,2)

Coefficients :

	ma1	ma2
	-0.6262	0.0908
s.e	0.0624	0.0609

3.5.3.2 Validation

Nous avons diagnostiqué notre modèle à partir des tests suivants pour montrer que le modèle choisis est valide

— Test ADF

```
library(tseries)
```

```
adf.test(yr$residuals)
```

Augmented Dickey-Fuller Test

```
data : yr$residuals
```

Dickey-Fuller = -6.2286, Lag order = 6, p-value = 0.01

— Test de Box-pierce

```
Box.test(yr$residuals)
```

Box-Pierce test

```
data : yr$residuals
```

X-squared = 0.002659, df = 1, p-value = 0.9589

3.5.3.3 Prévision

D'après le modèle choisis précédent on peut prédire des résultats pour les prochaines jours

On utilise l'ensemble des commandes suivantes :

```
mm=auto.arima(y)
```

```
ab=forecast(mm,h=10)
```

```
summary(ab)
```

on trouve

```
plot(ab)
```

TABLE 3.9 – Résultats de la prévision des décès par la méthode de Box-Jenkins

jours	prévision	Lo 80	Hi 80	Lo 95	Hi 95
256	9.210569	5.180428	13.24071	3.046999479	15.37414
257	9.625807	5.323307	13.92831	3.045701375	16.20591
258	9.625807	4.933470	14.31814	2.449496747	16.80212
259	9.625807	4.573624	14.67799	1.899160402	17.35245
260	9.625807	4.237758	15.01386	1.385497149	17.86612
261	9.625807	3.921633	15.32998	0.902026558	18.34959
262	9.625807	3.622132	15.62948	0.443977990	18.80764
263	9.625807	3.336877	15.91474	0.007718691	19.24389
264	9.625807	3.064011	16.18760	-0.409593413	19.66121
265	9.625807	2.802048	16.44957	-0.810231529	20.06185

Forecasts from ARIMA(0,1,2)

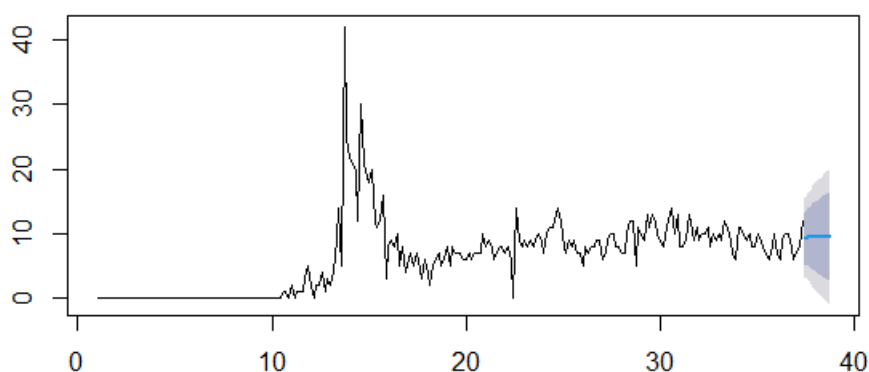


FIGURE 3.23 – Représentation graphique de la prévision des décès

3.5.4 La prévision par filter de Kalman

Nous utilisons le package (TSPred) pour la prévision par filter de Kalman

```
library(TSPred)
```

```
fArimaKF<-fittestArimaKF(y,h=10)
```

La représentation graphique de la série , nous utilisons la commande suivante :

```
plot(y,type='l',lwd=1,xlim=c(0,300),ylim=c(0,60),xlab="Time",ylab="ARIMAKF")
```

Nous obtenons

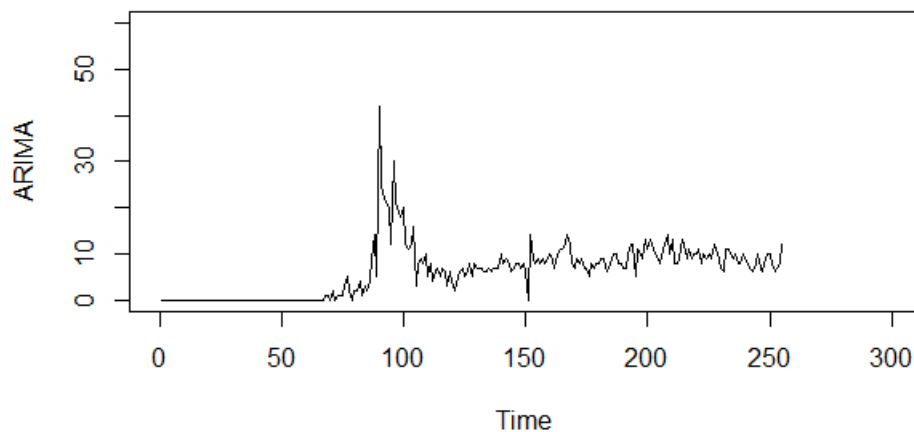


FIGURE 3.24 – Représentation graphique de la série des décès

nous utilisons les commandes suivantes :

```
fs<-KFAS::KFS(fArimaKFmodel,filtering=c("state","mean"),smoothing=c("state","mean"))  
f<-fitted(fs, filtered = TRUE)  
lines(f,col='red',lty=1,lwd=1)
```

Nous obtenons

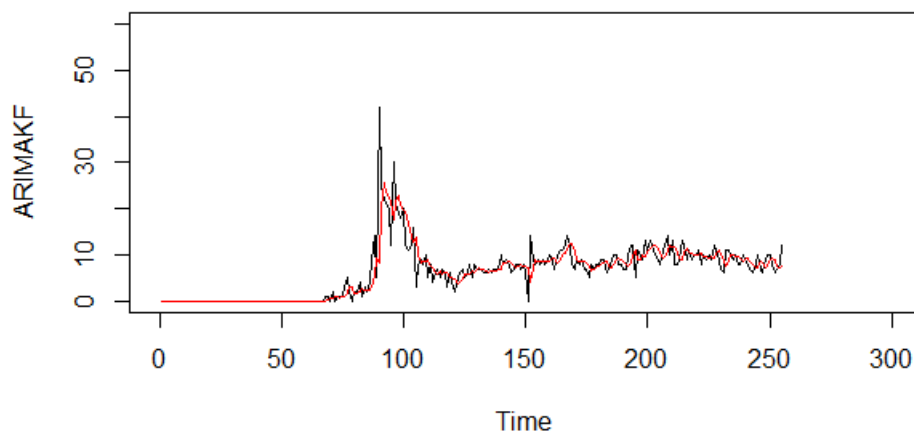


FIGURE 3.25 – Série chronologique filtrée par Kalman

Série chronologique lissée par Kalman

```
s<- fitted(fs)  
lines(s,col='green',lty=1,lwd=1)
```

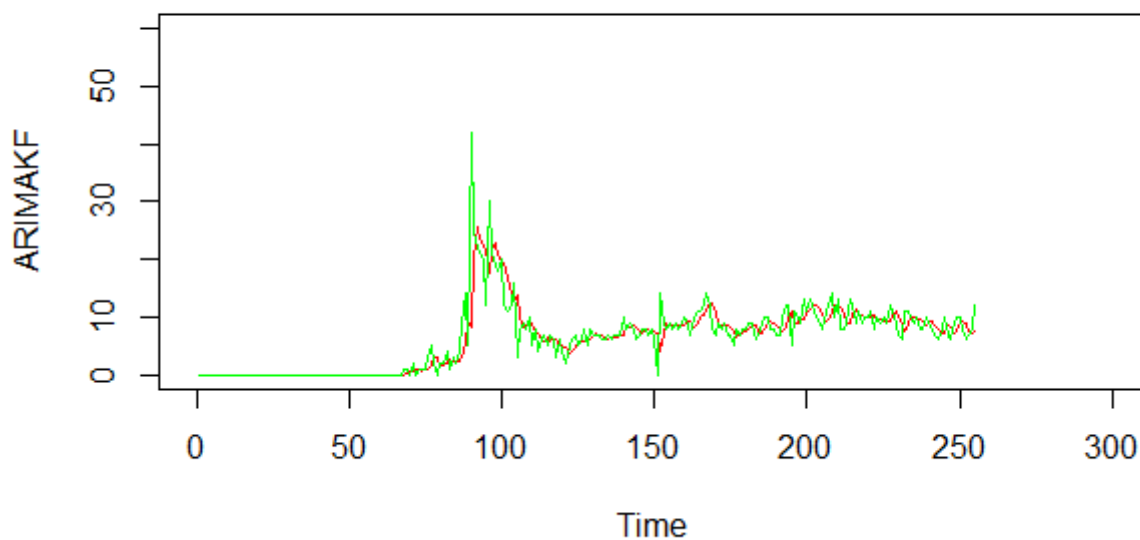



FIGURE 3.26 – Série chronologique ajustée par Kalman

Nous utilisons les commandes suivantes pour la prévision

```
pr=fArimaKF$pred$mean
```

```
pr
```

Nous obtenons : L'interval de confiance

TABLE 3.10 – Résultats de la prévision des décès par la méthode de Filtre de Kalman

jour	Prévision
256	9.277891
257	9.689314
258	9.566864
259	9.603308
260	9.592461
261	9.595690
262	9.594729
263	9.595015
264	9.594930
265	9.594955

```
fArimaKF$pred$lower
```

```
fArimaKF$pred$suppe
```

TABLE 3.11 – Bornes de confiances de la prévision des décès par la méthode de Filtre de Kalman

jour	Borne inférieure	Borne supérieure
256	4.1224544	14.43333
257	4.1771352	15.20149
258	3.5354980	15.59823
259	3.1482466	16.05837
260	2.7249391	16.45998
261	2.3433380	16.84804
262	1.9757686	17.21369
263	1.6266345	17.56339
264	1.2917239	17.89814
265	0.9699397	18.21997

TABLE 3.12 – Résultats de la prévision des décès par la méthode de Filtre de Kalman

jour	Prévision	Borne inférieure	Borne supérieure
256	9.277891	4.1224544	14.43333
257	9.689314	4.1771352	15.20149
258	9.566864	3.5354980	15.59823
259	9.603308	3.1482466	16.05837
260	9.592461	2.7249391	16.45998
261	9.595690	2.3433380	16.84804
262	9.594729	1.9757686	17.21369
263	9.595015	1.6266345	17.56339
264	9.594930	1.2917239	17.89814
265	9.594955	0.9699397	18.21997

3.5.4.1 Représentation graphique .

```
lines(ts(pred$mean,start=256),lwd=2,col='blue')
```

```
lines(ts(pred$upper,start=256),lwd=2,col='light blue')
```

```
lines(ts(pred$lower,start=256),lwd=2,col='light blue')
```

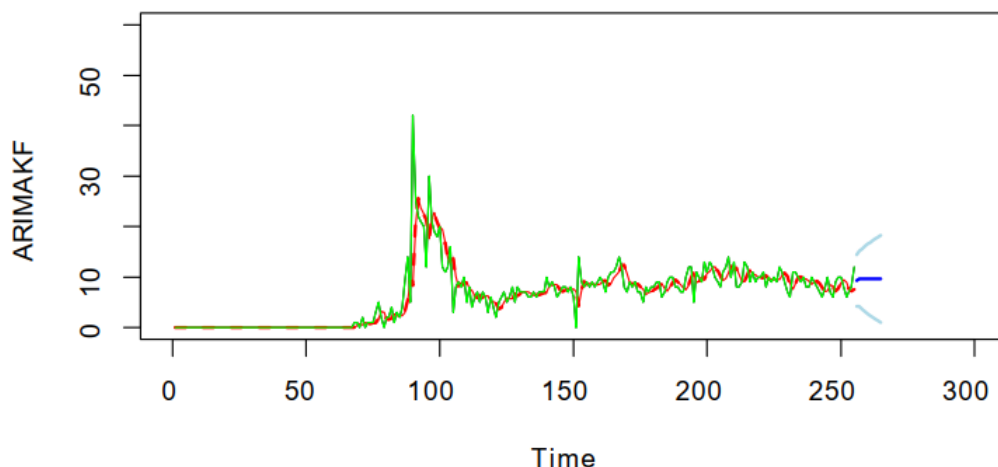


FIGURE 3.27 – Représentation graphique de prévision des décès par filtre de Kalman

3.5.5 Comparaison entre prévision par la méthode Box et filtre de Kalman

TABLE 3.13 – Comparaison de la prévision des infections par les deux méthodes avec le bilan réel

Date	Jour	Box	Kalman	Réelle
17/09/2020	256	233.6638	231.4074	228
18/09/2020	257	229.4237	224.6786	219
19/09/2020	258	225.4259	217.8135	210
20/09/2020	259	221.6565	210.8123	203
21/09/2020	260	218.1023	203.6749	197
22/09/2020	261	214.7513	196.4014	191
23/09/2020	262	211.5917	188.9916	186
24/09/2020	263	208.6125	181.4456	179
25/09/2020	264	205.8036	173.7635	175
26/09/2020	265	203.1552	165.9451	

TABLE 3.14 – Comparaison de la prévision des décès par les deux méthodes avec le bilan réel

Date	Jour	Box	Kalman	Réelle
17/09/2020	256	9.21056	9.277891	9
18/09/2020	257	9.625807	9.689314	6
19/09/2020	258	9.625807	9.566864	6
20/09/2020	259	9.625807	9.603308	7
21/09/2020	260	9.625807	9.592461	7
22/09/2020	261	9.625807	9.595690	10
23/09/2020	262	9.625807	9.595015	9
24/09/2020	263	9.625807	9.594930	5
25/09/2020	264	9.625807	9.594930	4
26/09/2020	265	9.625807	9.594955	4

On note à partir de ces résultats que filtre de Kalman est plus proche de la réalité.

Donc filtre de Kalman est meilleure que Box et Jenkins. Les prédictions de Kalman peuvent être utilisées pour prendre les meilleures décisions pour lutter contre la propagation du virus.

CONCLUSION GÉNÉRALE

L'objectif de ce travail a été d'introduire l'essentiel de la modélisation des séries chronologiques. Dans la première partie de ce travail, on a effectué une revue des principales notions de base concernant les séries chronologiques à une variable, cela a permis d'introduire, entre autres, la famille des processus ARIMA qui sont des modèles mathématiques efficaces pour modéliser des phénomènes temporels.

Ainsi que la procédure de sélection de modèles de Box-Jenkins et filtre de Kalman qui nous permet de modéliser et prédire le comportement d'un caractère statistique dans l'horizon. Le logiciel R facilite de manipuler et traiter le modèle temporel pour choisir le meilleur modèle d'ajustement, et donne les valeurs des prévisions numériquement et visuellement.

BIBLIOGRAPHIE

- [1] Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the institute of Statistical Mathematics*, 21(1), 243-247.
- [2] Bowerman, B. L., & O'Connell, R. T. (1979). *Time series and forecasting*. North Scituate, MA : Duxbury Press.
- [3] Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis : forecasting and control*. John Wiley & Sons.
- [4] Corinne, P. (2005). *Séries chronologiques : Quelques éléments du cours*. Paris.
- [5] Cheung, Y. W., & Lai, K. S. (1995). Lag order and critical values of the augmented Dickey&Fuller test. *Journal of Business & Economic Statistics*, 13(3), 277-280.
- [6] Eshel, G. (2003). The yule walker equations for the AR coefficients. *Internet resource*, 2, 68-73
- [6] Mechgoug, R. (2013). *La Prédiction des Séries Temporelles utilisant les Paradigmes de Soft Computing*. Algérie.
- [6] Melard, G. (1979). Modèles ARIMA pour des séries chronologiques non homogènes. *Statistique et analyse des données*, 4(2), 41-50.
- [7] MCLEOD, A. Ian. Diagnostic checking of periodic autoregression models with application. *Journal of Time Series Analysis*, 1994, vol. 15, no 2, p. 221-233.
- [8] Phillips, P. C., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2), 335-346.
- [9] Woodward, W. A. et Gray, H. L. (1996). On the relationship between the S-array , and the box-jenkins method of ARMA model identification, *journal of American Statistical Association*, vol 76, n 375
- [10] Yves, A. (2011). *Series temporelles avec R Méthodes et cas*, Springer-Verlang, France.