

TECHNIQUES DE PRÉVISION

RESPONSABLE DU COURS :
DR INES BEN HAMOUDA

OBJECTIFS DU COURS

Au terme de ce cours, l'apprenant sera capable de :

- Distinguer entre les différents types de séries temporelles
- Faire des prévisions en appliquant différentes méthodes de prévision
- Distinguer entre différents modèles de prévision
- Minimiser l'erreur de prévision
- Obtenir des prévisions optimales

**Public
cible**

Les étudiants de 3ème année
en Licence informatique
appliquée à la gestion

**Charge
hebdomadaire**

3h de travail

**Modalité
d'enseignement**

20% en ligne et 80% en
présentiel

PLANNING DU COURS

Titre du chapitre

Chapitre 1
Panorama des méthodes de prévision usuelles

Chapitre 2
Méthode Moyenne Mobile

Chapitre 3
Méthode du lissage exponentiel

Chapitre 4
Modèles autorégressifs et moyenne mobile

Chapitre 5
Régression linéaire

Description

- Identifier une série temporelle
- Distinguer entre les différents types de STP
- Identifier les différents indicateurs de calcul de l'erreur de mesure de prévision
- Appliquer la méthode MA pour différents types de STP
- Choisir l'ordre optimal de mesure de prévision par la méthode MA
- Identifier l'utilité de la prévision à travers MA

- Appliquer la méthode du lissage exponentiel
- Choisir l'ordre optimal de mesure de prévision par la méthode exp
- Identifier l'utilité de la prévision à travers exp
- Définir un bruit blanc
- Déterminer la stationnarité et l'inversibilité d'un modèle
- Identifier et étudier les caractéristiques d'un modèle AR
- Identifier et étudier les caractéristiques d'un modèle MA
- Identifier et étudier les caractéristiques d'un modèle ARMA
- Construire un modèle de régression linéaire simple
- Déterminer la nature de la relation entre deux variables (covariance et corrélation)
- Déterminer la qualité de l'ajustement
- Faire de prévisions

COMPOSITION DU COURS

- Support de cours ppt, pdf
- Exercices d'applications
- Vidéos

CHAPITRE 1

PANORAMA DES MÉTHODES DE PRÉVISION USUELLES

Au terme de ce chapitre, l'apprenant sera capable de:

- Identifier une série temporelle
- Distinguer entre les différents types de STP
- Identifier les différents indicateurs de calcul de l'erreur de mesure de prévision

CHAPITRE 2

MÉTHODE MOYENNE MOBILE

Au terme de ce chapitre, l'apprenant sera capable de:

- Appliquer la méthode MA pour différents types de STP
- Choisir l'ordre optimal de mesure de prévision par la méthode MA
- Identifier l'utilité de la prévision à travers MA

CHAPITRE 3

MÉTHODE DU LISSAGE EXPONENTIEL

Au terme de ce chapitre, l'apprenant sera capable de:

- Appliquer la méthode du lissage exponentiel
- Choisir l'ordre optimal de mesure de prévision par la méthode exp
- Identifier l'utilité de la prévision à travers exp

CHAPITRE 4

MODÈLES AUTORÉGRESSIFS ET MOYENNE MOBILE

Au terme de ce chapitre, l'apprenant sera capable de:

- Définir un bruit blanc
- Déterminer la stationnarité et l'inversibilité d'un modèle
- Identifier et étudier les caractéristiques d'un modèle AR
- Identifier et étudier les caractéristiques d'un modèle MA
- Identifier et étudier les caractéristiques d'un modèle ARMA

CHAPITRE 5

RÉGRESSION LINÉAIRE

Au terme de ce chapitre, l'apprenant sera capable de:

- Construire un modèle de régression linéaire simple
- Déterminer la nature de la relation entre deux variables (covariance et corrélation)
- Déterminer la qualité de l'ajustement
- Faire des prévisions

EVALUATION

- Un DS en présentiel d'une durée d'une heure
- Une note de participation
- La note de TD sera la moyenne de ces deux notes
- Examen final dans la session de Janvier



BONNE ANNÉE UNIVERSITAIRE

Chapitre 1

Panorama des méthodes de prévision usuelles

1. Concepts et définitions

La prévision peut être définie comme étant « une appréciation sur les valeurs futures d'une variable quantitative ».

On part d'une série d'observations à travers le temps portant sur une variable quelconque, de l'instant 1 jusqu'à l'instant T.

Il s'agit d'une série chronologique ou encore d'une série temporelle.

On cherche à prévoir la valeur qui sera atteinte par y à un instant futur $T+h$.

Notons h l'horizon de prévision et T l'origine de prévision.

On s'intéresse à la valeur future inconnue y_{t+h} .

On distingue trois formes de prévision :

1- La *prévision ponctuelle* notée \hat{y}_{t+h} :
c'est une valeur unique qui représente la meilleure estimation de y_{t+h} sur la base des données disponibles (ça peut être la moyenne, la médiane...)

2- *La distribution de la valeur future :*

C'est la loi de probabilité de la valeur future inconnue.

3- *L'intervalle de prévision :*

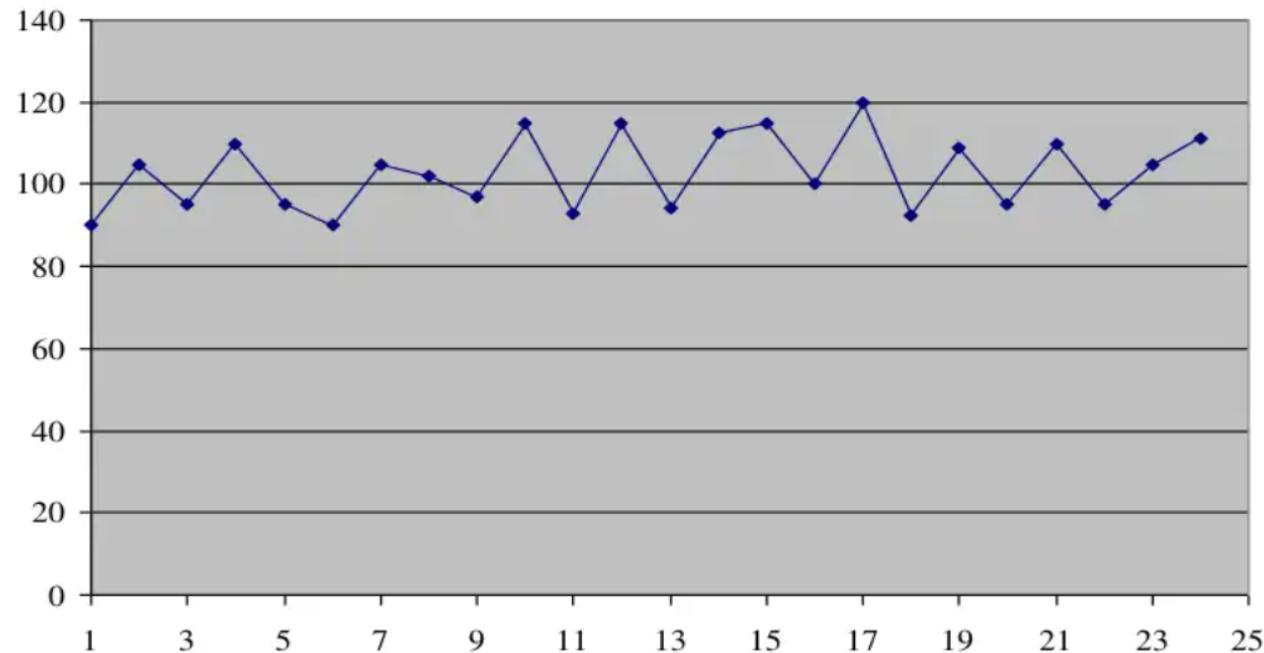
C'est un intervalle de valeurs possibles de la variable dans lequel la valeur future devrait se trouver avec une probabilité donnée.

7

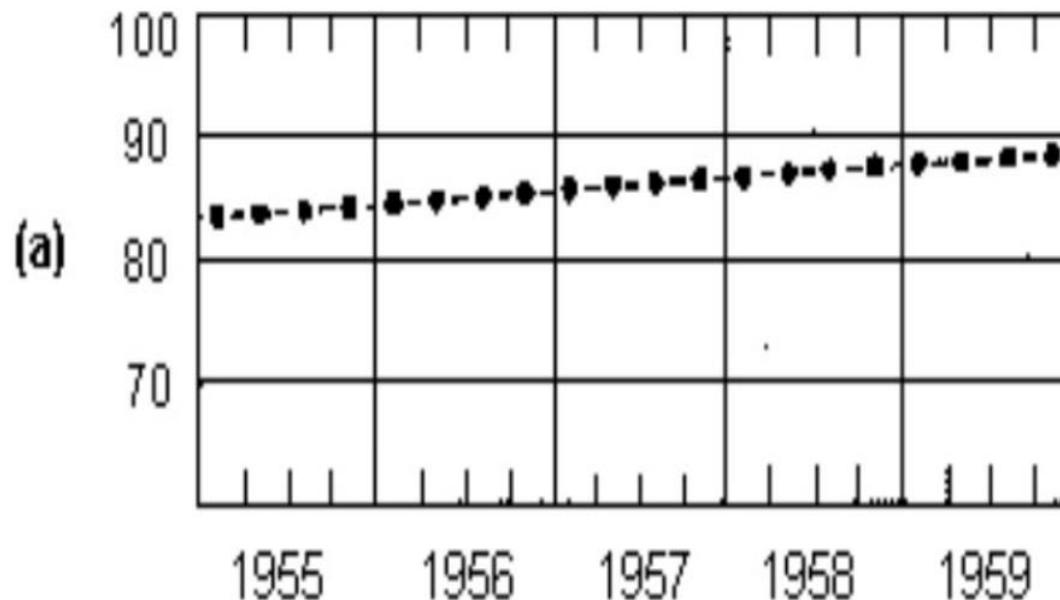
On distingue quatre composantes pour une série temporelle (STP)

- ▶ Une composante stationnaire ou cyclique
- ▶ Une composante de tendance
- ▶ Une composante saisonnière ou périodique
- ▶ Une composante aléatoire ou stochastique

Séries stationnaires : Le niveau moyen des observations est constant

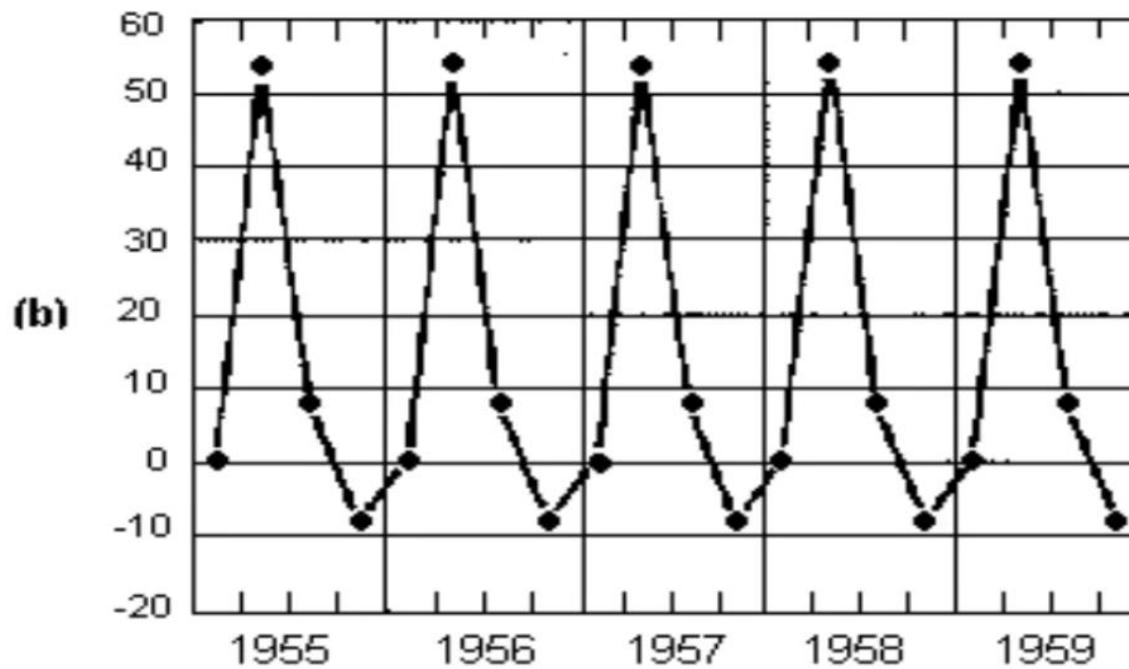


Tendance : Augmentation (ou une diminution) significative de la demande en fonction du temps.

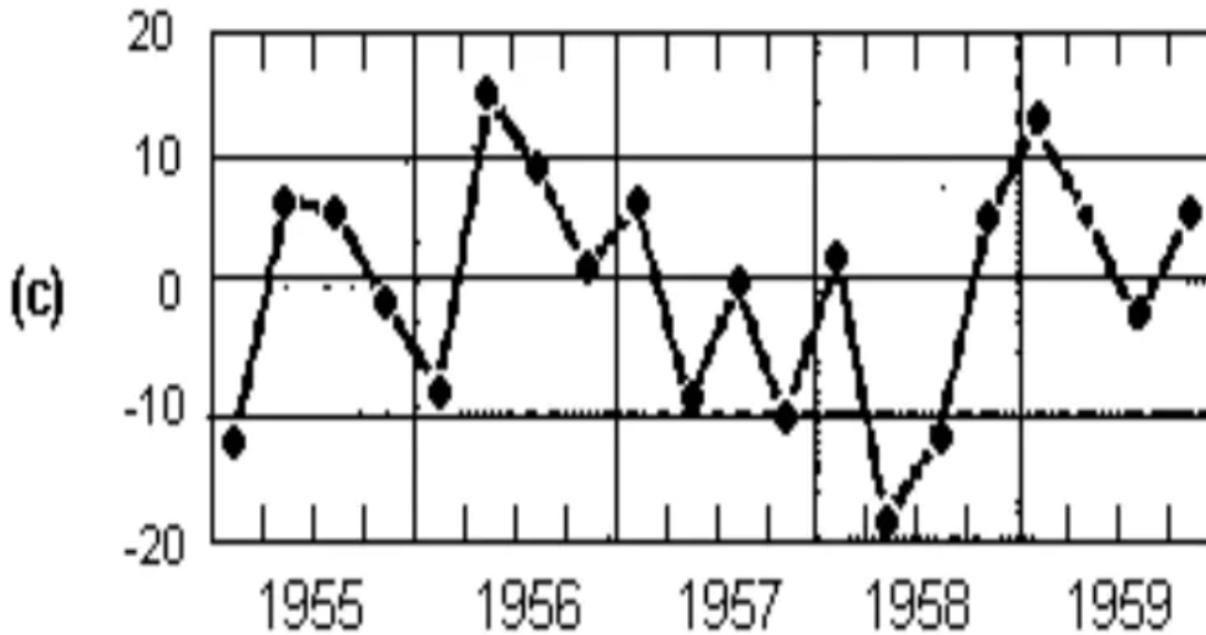


Saisonnalité: caractéristique d'un phénomène qui se répète à intervalles fixes, par exemple à tous les hivers, à tous les mois, etc.

Variations saisonnières



Aspect aléatoire: se dit d'un phénomène qui ne possède ni les caractéristiques de saisonnalité ni celles d'un cycle ou celle d'une tendance.



L'ensemble d'information représente les éléments dont le prévisionniste dispose afin de réaliser une prévision.

Les méthodes de prévision sont classées selon la quantité d'information requise.

Supposons qu'on doit prévoir y_{t+1}

- Méthodes de prévision exploratoires : courbes de croissance, méthodes de décomposition saisonnière, lissage exponentiel, méthode de Box et Jenkins pour les modèles ARIMA... utilisent le passé de la variable c'est-à-dire y_t, y_{t-1}, \dots

- Méthodes de prévision **explicatives** : régression linéaire simple ou multiple, méthode de Box et Jenkins pour les modèles de fonction de transfert... utilisent les valeurs passées d'une ou de plusieurs variables, y compris éventuellement, la variable étudiée : $x_t, x_{t-1}, \dots, y_t, y_{t-1} \dots$ (par exemple budget de la publicité)

- Méthodes de prévision systématiques et économétriques : elles considèrent le système dans sa totalité. Elles utilisent les relations entre les variables c'est-à-dire si y_{t+1} est fonction de x_{t+1} alors x_{t+1} peut aussi être fonction de y_{t+1} (par exemple prix proposé et quantité demandée dans le cadre de l'offre et de la demande).

Ces différentes méthodes sont relatives à de l'information quantitative.

Cependant, l'information qualitative peut être représentée principalement à l'aide de variables binaires.

Par exemple l'effet de la première crise pétrolière sera mesuré en introduisant une variable égale à 1 à partir de 1974, et à 0 auparavant.

On introduit parfois des variables binaires pour éviter des données aberrantes.

2. Critères usuels

Supposons qu'on dispose de n prévisions $\hat{y}_1, \dots, \hat{y}_n$, qui correspondent aux données y_1, \dots, y_n et donc de n erreurs de prévision e_1, \dots, e_n où $e_i = y_i - \hat{y}_i$.

Il peut s'agir de prévisions, soit pour toutes les données disponibles, soit pour un sous-ensemble de données récentes.

Les critères suivants sont utilisés pour juger de la validité de la méthode de prévision.

1. L'erreur Moyenne (mean error):

$$\bar{e} = \frac{1}{n} \sum_{i=1}^n e_i$$

2. La variance:

$$Var(e) = \frac{1}{n} \sum_{i=1}^n (e_i - \bar{e})^2$$

3. L'écart-type (standard deviation):

$$std(e) = \sqrt{\frac{1}{n} \sum_{i=1}^n (e_i - \bar{e})^2} = \sqrt{Var(e)}$$

4. L'écart absolu moyen (mean absolute deviation):

$$MED(e) = \frac{1}{n} \sum_{i=1}^n |e_i - \bar{e}|$$

5. Le carré moyen des erreurs (mean square error):

$$MSE(e) = \frac{1}{n} \sum_{i=1}^n e_i^2$$

Ce critère correspond à la fonction de coût quadratique.

6. L'erreur quadratique Moyenne (root mean square error):

$$RMSE(e) = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} = \sqrt{MSE(e)}$$

7. L'erreur absolue Moyenne (mean absolute error):

$$MAE(e) = \frac{1}{n} \sum_{i=1}^n |e_i|$$

Ce critère correspond à la fonction de coût valeur absolue.

Il pénalise moins les grandes erreurs que MSE. On dit qu'il est plus robuste.

8. L'erreur absolue moyenne en pourcentage (mean absolute percentage error):

$$MAPE(e) = \frac{1}{n} \sum_{i=1}^n \frac{|e_i|}{y}$$

qu'on exprime généralement en pour cent.

Ce critère correspond à la fonction de coût valeur absolue en pourcentage. C'est un nombre sans dimension sauf que la variable $y > 0$.

Ces différents critères ne considèrent qu'une méthode de prévision.

Il existe d'autres qui servent à comparer la méthode étudiée à une méthode de prévision de référence, souvent la méthode de prévision naïve dans laquelle la prévision est égale à la dernière donnée disponible.

Parmi ces critères : l'estimation non biaisée de la variance, le critère MAPE de la prévision naïve 1 (naïve forecast 1) ou encore le critère AIC (Akaike information criterion)...

MÉTHODE MOYENNE MOBILE

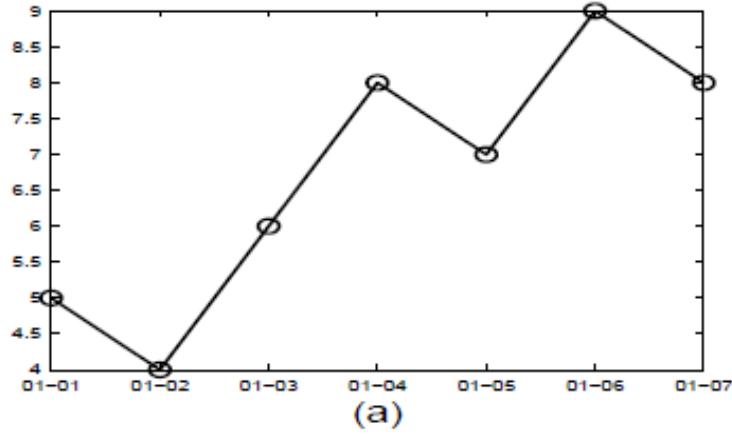
Définition

- Les *moyennes mobiles* sont utilisées à différentes fins. Elles sont employées pour lisser une série chronologique, dans le cadre des méthodes de décomposition saisonnière mais également comme méthodes de prévision.
- L'idée de base est de calculer une moyenne de k données autour de la date à laquelle on s'intéresse, en prenant par exemple 4 données, les 2 plus proches avant cette date (à gauche) et les 2 plus proches (à droite); et de procéder de même pour les autres dates. Le nombre k de données utilisées dans la moyenne est appelé l'ordre de la moyenne mobile.

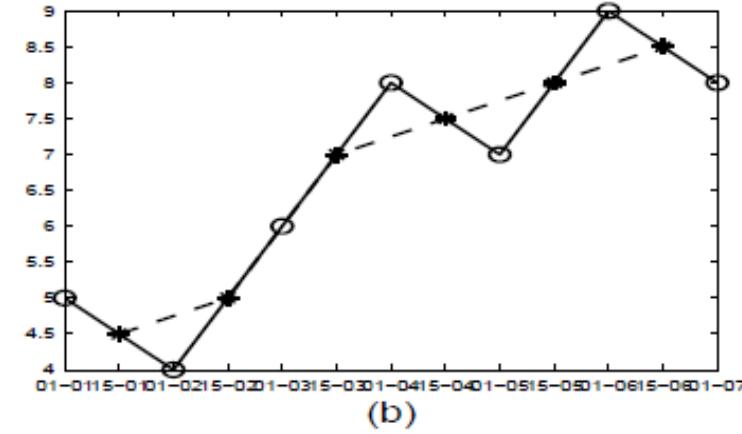
Exemple

Le tableau suivant présente les moyennes mobiles d'ordre 2, 3 et 4 d'une même série.

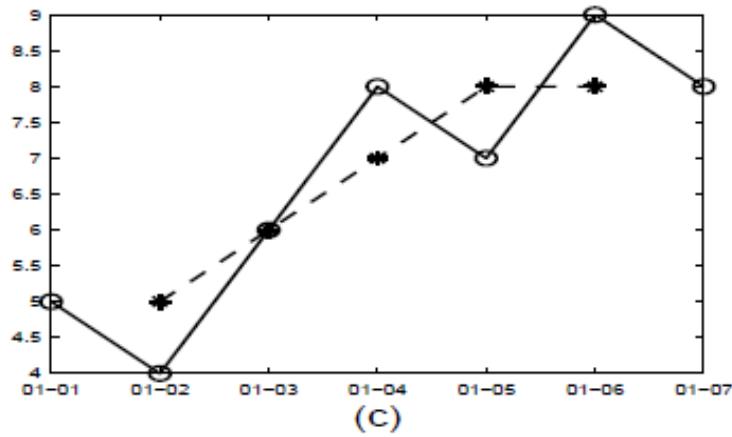
date t_i	date $\frac{t_i+t_{i+1}}{2}$	observ. y_i	$MM(2)$	$MM(3)$	$MM(4)$
01-01	15-01	5	$\frac{5+4}{2} = 4.5$		
01-02	15-02	4	$\frac{4+6}{2} = 5$	$\frac{5+4+6}{3} = 5$	$\frac{5+4+6+8}{4} = 5.75$
01-03	15-03	6	$\frac{6+8}{2} = 7$	$\frac{4+6+8}{3} = 6$	$\frac{4+6+8+7}{4} = 6.25$
01-04	15-04	8	$\frac{8+7}{2} = 7.5$	$\frac{6+8+7}{3} = 7$	$\frac{6+8+7+9}{4} = 7.50$
01-05	15-05	7	$\frac{7+9}{2} = 8$	$\frac{8+7+9}{3} = 8$	$\frac{8+7+9+8}{4} = 8.00$
01-06	15-06	9	$\frac{9+8}{2} = 8.5$	$\frac{7+9+8}{3} = 8$	
01-07		8			



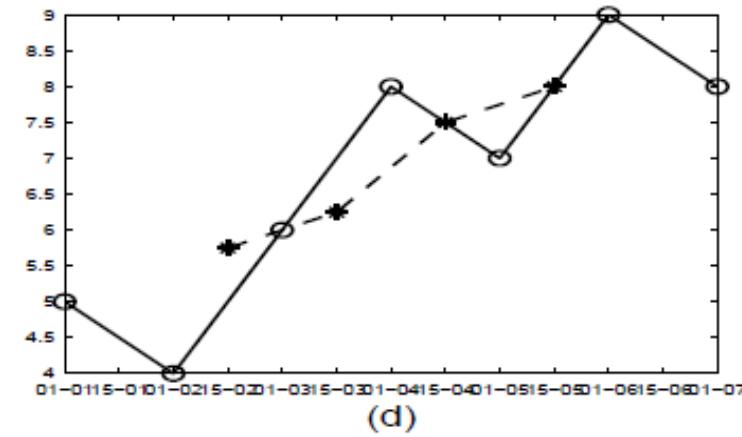
(a)



(b)



(c)



(d)

(a) la série observée ($y_i, 1 \leq i \leq 7$), (b) $MM(2)$, (c) $MM(3)$, (d) $MM(4)$

EXEMPLE I

Calculer les séries des moyennes-mobiles d'ordre 2, 3 et 4 de la série initiale (y_i) suivante.

t_i	1	2	3	4	5	6	7	8
y_i	30	15	5	30	36	18	9	36

t_i	9	10	11	12	13	14	15	16
y_i	45	15	10	60	48	16	8	72

On complétera le tableau suivant.

(t_i)	(y_i)	$\left(\frac{t_i+t_{i+1}}{2} \right)$	$MM(2)$	$MM(3)$	$MM(4)$
1	30				
2	15	1.5			
3	5	2.5			
4	30				
5	36				
6	18				
7	9				
8	36				
9	45				
10	15				
11	10				
12	60				
13	48				
14	16				
15	8				
16	72				

Pour pallier l'inconvénient que les dates des moyennes mobiles d'ordre pair soient intermédiaires par rapport aux dates des données, on peut définir pour k pair les *moyennes mobiles centrées* (centered moving average) d'ordre k : CMA(k) comme les moyennes entre deux MA consécutives : c'est-à-dire $CMA(k) = MA(2)$ de $MA(k)$.

Par exemple, la première CMA(4) est donnée par :

$$\frac{\frac{y_1 + y_2 + y_3 + y_4}{4} + \frac{y_2 + y_3 + y_4 + y_5}{4}}{2} = \frac{y_1 + 2y_2 + 2y_3 + 2y_4 + y_5}{8}$$

On remarque qu'on trouve une moyenne pondérée de 5 observations consécutives avec des poids positifs et dont la somme est égale à 1.

Les CMA représentent en fait un cas particulier des *moyennes mobiles pondérées* définies par :

$$\bar{y}_t = \sum_{j=-p_1}^{p_2} \omega_j y_{t+j}$$

avec $\omega_j \geq 0$, $\sum_{j=-p_1}^{p_2} \omega_j = 1$, $j = -p_1, -p_1 + 1, \dots, p_2$

Les MA sont à la base d'une méthode de prévision qui consiste à utiliser la moyenne des k dernières données disponibles comme prévision pour la date suivante.

EXEMPLE 2

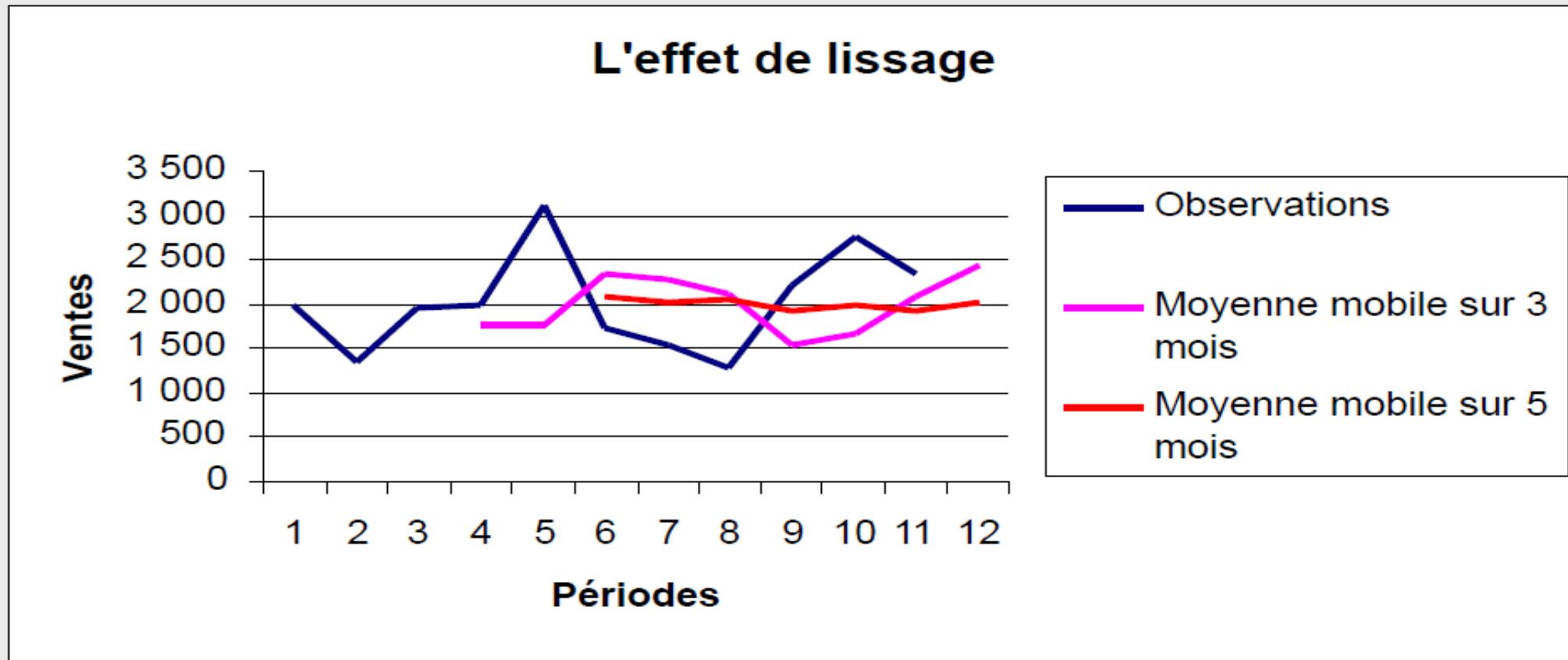
Considérons la demande observée pour chaque mois d'un bien électronique (les valeurs calculées sont arrondies)

Date	Janv	Févr	Mars	Avril	Mai	Juin	Juillet	Aout	Sept	Oct	Nov	Dec
Demande	2000	1350	1950	1975	3100	1750	1550	1300	2200	2770	2350	

- Effectuez des prévisions selon une moyenne mobile d'ordre 3 et 5
- Quel ordre de MA donne les meilleures prévisions?

- La valeur réelle observée est déterminée par une loi d'une part et par l'intervention du hasard d'autre part. Il existe un écart entre les valeurs prévues et les valeurs réellement observées. Un objectif commun à toutes les techniques est de minimiser ces écarts.
- On définit l'erreur de prévision comme étant la différence entre la valeur réelle y_i et la valeur prédite \hat{y}_i ; $e_i = y_i - \hat{y}_i$. Le choix de la technique repose sur le calcul de la moyenne de l'erreur absolue et du carré moyen de l'erreur.

Effet de lissage visualisé :



UTILITÉ DE LA MA

- La moyenne mobile est un indicateur reflétant la valorisation moyenne d'un titre sur une période donnée. Sa facilité d'utilisation et d'interprétation en font un outil très prisé des chartistes.
- La moyenne mobile donne la valeur moyenne des cours sur une période donnée. Cet indicateur permet de s'affranchir des aberrations des cours en les « lissant ». Une moyenne mobile à 20 jours donne le cours moyen sur les vingt dernières séances, une moyenne mobile à 50 semaines donne le cours moyen hebdomadaire sur les 50 dernières semaines et ainsi de suite.

- Le choix de la période de référence est primordial dans la mesure où il détermine son aptitude à réagir à des fluctuations de cours.
- Quelle durée choisir ? Il n'y a de bonne réponse, cela dépend de la volatilité de l'actif et de votre horizon d'investissement. Si vous souhaitez investir sur une semaine au maximum, il est inutile d'utiliser une moyenne mobile à 200 jours !
- Dans la pratique, il est fréquent de rencontrer des moyennes à 10, 20 et 50 semaines pour une moyenne de moyen/long terme et de 5,10 , 20 et 50 jours pour du court terme. Le cours le plus communément utilisé pour calculer une moyenne mobile est le cours de clôture.

- La Moyenne Mobile Exponentielle est **un type de Moyenne Mobile visant à affecter un coefficient dont la valeur est proportionnelle à la fraîcheur de la période étudiée.**
- Parmi les divers types de Moyennes Mobiles utilisées : la Moyenne Mobile Exponentielle (aussi abrégée MME, ou EMA pour Exponential Moving Average en anglais) est sans doute l'une des plus reconnues. En effet, sa formule de calcul permet de donner davantage de poids aux périodes récentes d'un cours, par rapport aux périodes plus anciennes.
- Ainsi, la MME offre une plus grande flexibilité, appréciée des Traders cherchant à détecter la tendance rapidement sur les marchés financiers.

- **La Moyenne Mobile Exponentielle** est considérée comme un meilleur outil de suivi de tendance. Elle accorde une plus forte pondération aux cours les plus récents ce qui la rend plus réactive.

$$\text{MME} = \text{MME}[n-1] + (2/(\text{Période} + 1)) * (\text{Cours}[n] - \text{MME}[n-1])$$

Où MME[n-1] : est la valeur de la moyenne mobile de la veille

Cours[n] : est le cours de clôture du jour

(2/(\text{Période} + 1)) : est le pourcentage exponentiel, qui dépend de la période de calcul de la moyenne mobile.

n le nombre de jours utilisés pour calculer la moyenne mobile

COMMENT UTILISER LES MOYENNES MOBILES ?

- La courbe de moyenne mobile peut être utilisée de diverses manières : seule pour identifier la tendance, avec les cours pour générer des signaux, avec une autre moyenne mobile.
- **Identifier une tendance** : Si la moyenne mobile est relativement plate on est en présence d'un range, si la moyenne mobile croît ou décroît il y a une tendance.

- **Générer des signaux** : Nous pouvons observer la position relative du cours par rapport à la moyenne mobile.

* Si le cours est au dessus d'une moyenne mobile croissante nous sommes en présence d'une tendance haussière. Si le cours passe alternativement au dessus et au dessous d'une moyenne mobile relativement plate, la configuration en range est évidente.

* Si les cours viennent à dépasser la moyenne à la hausse, il s'agit d'une indication d'achat. Ce signal est plus ou moins fort en fonction de la moyenne mobile. Le franchissement d'une moyenne mobile à long terme est synonyme d'un fort changement de tendance à long terme.

Inversement, la rupture d'une moyenne mobile est un signal de vente plus ou moins fort.

- **Utiliser plusieurs moyennes mobiles** : Nous pouvons observer la position relative d'une moyenne court terme par rapport à une moyenne long terme.
- * Si la moyenne court terme se trouve au dessus de la moyenne long terme et que leur espacement s'accroît, alors nous sommes en tendance haussière. Si les deux moyennes sont confondues nous sommes en range.
- * Une moyenne mobile à court terme franchissant une moyenne mobile à long terme est un fort signal d'achat à court terme.
- * La cassure d'une moyenne mobile à long terme par une moyenne mobile à court terme indique un renversement de tendance à court terme.

Méthode du lissage exponentiel

Définition

- Le but est d'effectuer au temps T une prévision pour le temps $T + 1$: $\hat{y}_T(1)$.
- L'idée de base du *lissage exponentiel* est très simple : il s'agit de combiner la dernière observation disponible y_T et la dernière prévision, celle qui a été calculée au temps $T - 1$ pour le temps T :

$$F_{T+1} = S_T = \hat{y}_T$$

- ▶ Par commodité, on notera cette prévision F_{T+1} (F pour forecast) ou S_T (S pour smoothing).
- ▶ On effectue une combinaison pour une moyenne pondérée :

$$F_{T+1} = \alpha y_T + (1 - \alpha)F_T \text{ ou } S_T = \alpha y_T + (1 - \alpha)S_{T-1}$$

α est appelée la constante de lissage, c'est une valeur généralement comprise entre 0 et 1.

Une valeur recommandée par Brown (1959), l'auteur de la méthode, est 0,3.

► La méthode du lissage exponentiel simple présente les avantages suivants :

- elle est simple à comprendre et à expliquer,
- simple à mettre en œuvre,
- nécessite peu de calculs, donc un temps de calcul réduit, peu d'informations à mémoriser, facilement automatisée donc un coût total très bas,
- très pratique lors des compétitions à grandes échelles à condition de corriger la saisonnalité et avec un choix optimal de α .

- Un autre moyen d'écrire le lissage exponentiel facile à appliquer et qui garde le même principe est le suivant: $S_{t+1} = S_t + \alpha(y_t - S_t)$
- Sous cette forme, la nouvelle prévision par lissage exponentiel est simplement l'ancienne prévision plus α fois l'erreur de l'ancienne prévision, car $y_t - S_t = e_t$ est l'erreur de cette dernière prévision.
- La nouvelle prévision apparaît donc comme l'ancienne prévision corrigée.

- ▶ On peut interpréter la constante du lissage α comme un paramètre qui traduit le sentiment du prévisionniste sur l'importance du passé dans l'évolution future de la série.
- ▶ C'est ce qu'on appelle « la correction par l'erreur » (error correction form) du lissage exponentiel simple.
- ▶ Cette méthode est adaptée notamment dans la gestion des stocks de plusieurs milliers d'articles différents, les cours en bourse, ventes d'un article dont la demande est stable dans le temps...

Exemple: Calculer les prévisions de la demande des jeans de cette société artisanale, un mois à l'avance, par la méthode du lissage exponentiel

Mois	Dde observée	$\alpha = 0,1$	$\alpha = 0,5$	$\alpha = 0,9$
Janvier	2000	-	-	-
Février	1350			
Mars	1950			
Avril	1975			
Mai	3100			
Juin	1750			
Juillet	1550			
Aout	1300			
Septembre	2200			
Octobre	2775			
Novembre	2350			
Décembre				

- ▶ La question qui se pose : pour ces trois valeurs de α , laquelle choisir ?
- ▶ Pour pouvoir répondre à cette question, on procède à l'analyse du carré moyen de l'erreur et l'écart absolu moyen pour comparer ces différents procédés de prévisions.

- On voit clairement que $\alpha = 0,1$ donne des prévisions bien meilleures que les grandes valeurs de α . Ceci est vrai tant pour l'écart absolu moyen que pour le carré moyen de l'erreur.
- En résumé, pour les moyennes mobiles et le lissage exponentiel il n'y a pas de règles idéales pour déterminer la pondération appropriée :
 - Pour la MA, on doit décider et spécifier le nombre d'observations à inclure dans la moyenne.
 - Pour le lissage exponentiel, il s'agit de choisir une valeur de α . La plupart du temps, on procède expérimentalement, en essayant 2 ou 3 valeurs différentes pour déterminer la plus appropriée.

Régression linéaire

Modèle linéaire simple

- On s'intéresse à la régression quand on veut savoir jusqu'à quel point on peut prédire la valeur d'une variable en connaissant la valeur d'une autre variable.
- La régression linéaire simple permet d'évaluer s'il existe une relation fonctionnelle linéaire entre une variable explicative quantitative X et une variable expliquée quantitative Y.

On veut calculer Y à partir de X, c'est-a-dire : $Y_i = f(X_i)$

- Variable X : variable expliquée, variable indépendante, variable prédite, variable de réponse.
- Variable Y : variable explicative, variable dépendante, variable prédicatrice, variable de contrôle.

- Le modèle de régression simple s'écrit :

$$y_i = ax_i + b \quad \forall i = 1, 2, \dots, n$$

n : est la taille de l'échantillon,

b : est le terme constant du modèle, c'est-à-dire la valeur moyenne de Y quand X vaut 0.

a : est la pente de la droite de régression.

- Le modèle tel qu'on vient de spécifier n'est qu'une caricature de la réalité. En effet ne retenir que X pour expliquer Y est insuffisant.
- Il existe une multitude d'autres facteurs susceptibles d'expliquer Y c'est pourquoi on ajoute ε_i qui synthétise l'ensemble des phénomènes explicatifs de Y et non liés à X .

- ε_i : quantifie les écarts entre les valeurs réellement observées et les valeurs prédites par le modèle.
- La droite de régression devient donc :

$$y_i = ax_i + b + \varepsilon_i \quad \forall i = 1, 2, \dots, n$$

- Le terme ε_i regroupe :

- une erreur de spécification : le fait que la seule variable explicative n'est pas suffisante Pour rendre compte de la totalité du phénomène expliqué.
- une erreur de fluctuation d'échantillonnage : d'un échantillon à un autre les observations, et par là les estimations, peuvent être légèrement différentes (si on change d'échantillon, on peut obtenir un résultat différent).

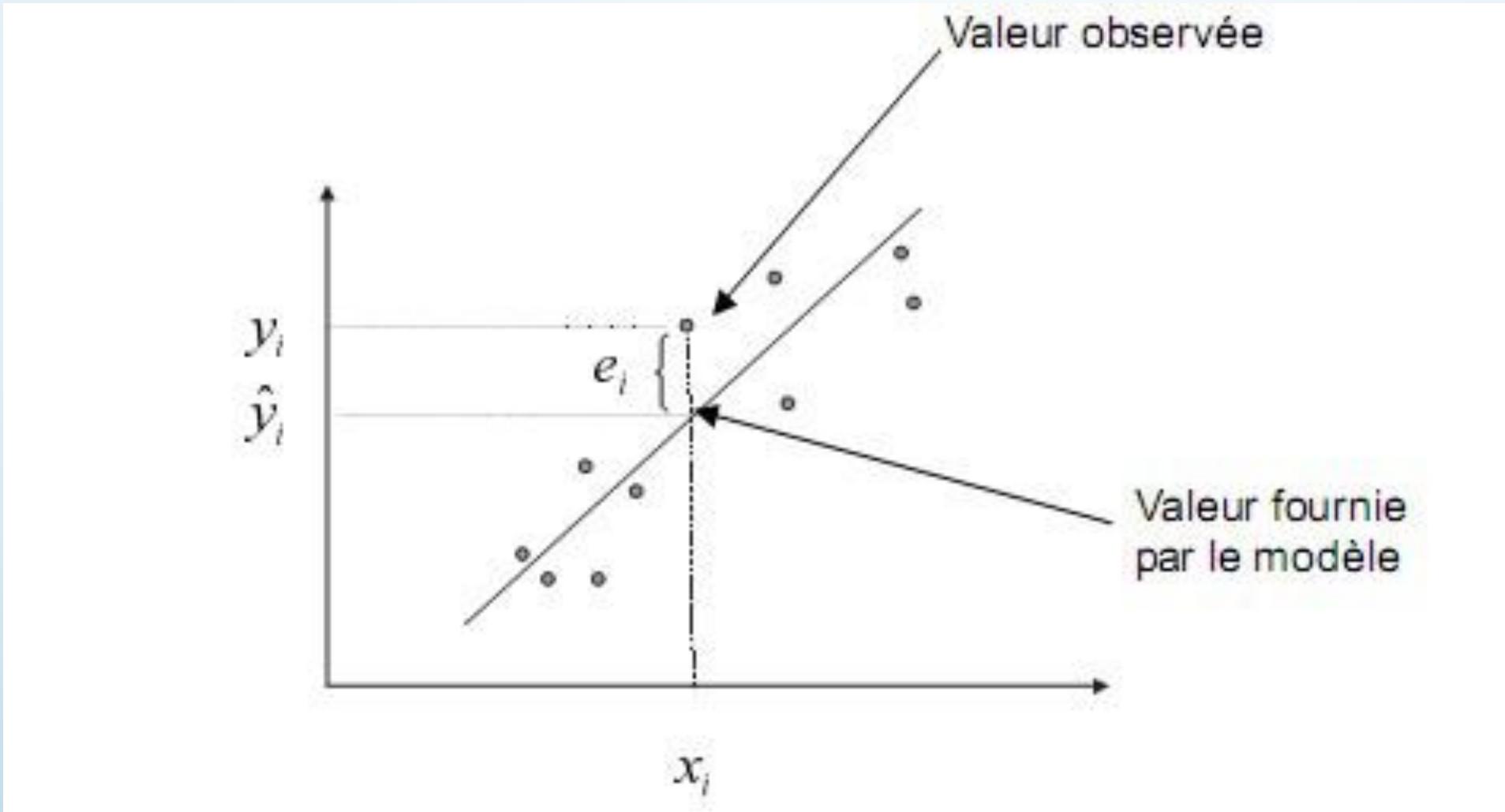
- Pour estimer les paramètres de ce modèle, on applique la méthode des moindres carrés ordinaires. Cette méthode repose sur la minimisation des carrés des erreurs (des résidus).

La méthode des Moindres Carrés Ordinaires (MCO)

- La méthode MCO consiste à ajuster le nuage de points par une droite de manière à minimiser la somme des carrés des distances entre les points du nuage et cette droite. Ceci revient à minimiser la somme des carrés des résidus.

$$\varepsilon_i = Y_i - (aX_i + b) \Rightarrow \varepsilon_i^2 = (Y_i - (aX_i + b))^2$$

$$\text{minimiser} \sum_{i=1}^n \varepsilon_i^2 = \text{minimiser} \sum_{i=1}^n (Y_i - (aX_i + b))^2 = f(a, b)$$



La résolution du système d'équations donne les résultats suivants:

$$\hat{a} = \frac{COV(X, Y)}{V(X)} = \frac{\sum_{i=1}^n Y_i X_i - n \bar{Y} \bar{X}}{\sum_{i=1}^n X_i^2 - n \bar{X}^2}$$

$$\hat{b} = \bar{Y} - \hat{a} \bar{X}$$

Enfin, l'équation du modèle linéaire est donnée par : $Y_i = \hat{a} X_i + \hat{b}$

Remarques :

- La droite de régression passe par le point moyen de coordonnées (\bar{X}, \bar{Y}) . En effet, comme $\hat{b} = \bar{Y} - \hat{a}\bar{X}$, alors on a : $\bar{Y} = \hat{a}\bar{X} + \hat{b}$
- L'étude de la droite de régression permet de prévoir Y en fonction de X :

$$\hat{Y}_i = \hat{a}X_i + \hat{b}$$

- On peut exprimer X en fonction de Y . Dans ce cas, X est la variable expliquée ou endogène et Y est la variable explicative ou exogène. Dans ce cas l'estimation par la méthode MCO des paramètres du modèle donne :

$$\hat{a}' = \frac{COV(X, Y)}{V(Y)} = \frac{\sum_{i=1}^n Y_i X_i - n \bar{Y} \bar{X}}{\sum_{i=1}^n Y_i^2 - n \bar{Y}^2}$$

$$\hat{b}' = \bar{X} - \hat{a}' \bar{Y}$$

Hypothèses du modèle linéaire simple

- H_1 : le modèle est linéaire en ses paramètres.
- H_2 : le terme résiduel est d'espérance nulle $E(\varepsilon_i) = 0$
- H_3 : homoscédasticité des résidus : la variance des résidus est constante et ne varie pas suivant les observations $V(\varepsilon_i) = \sigma^2, \forall i$
- H_4 : pas de corrélations entre les erreurs $Cov(\varepsilon_i, \varepsilon_j) = 0$
- H_5 : le terme résiduel ne doit pas être corrélé avec les variables explicatives $Cov(X_i, \varepsilon_i) = 0$

Covariance

- La covariance est égale à la moyenne des écarts des couples (x_i, y_i) de X et Y par rapport au point moyen (\bar{X}, \bar{Y}) :

$$Cov(X, Y) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

- La covariance indique le sens de la relation entre les variables X et Y .

- Ainsi, on peut distinguer les cas suivants :
 - Si $Cov(X, Y) > 0$, alors la relation entre les deux variables est positive.
Dans ce cas, les deux variables varient dans le même sens.
 - Si $Cov(X, Y) < 0$, alors la relation entre les deux variables est négative.
Dans ce cas, les deux variables varient en sens inverse.
 - Si $Cov(X, Y) = 0$, alors il n'y a pas de relation entre les deux variables.
Dans ce cas, les variations de l'une n'entraînent pas la variation de l'autre.

Coefficient de corrélation

- Le coefficient de corrélation linéaire est un nombre sans dimension qui permet de mesurer le degré ou l'intensité de la liaison linéaire entre deux variables statistiques. Ce coefficient est défini par :

$$r_{X,Y} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

- La covariance indique le sens de la relation entre les variables X et Y .
Ainsi, on distingue les cas suivants :
 - Si $r_{X,Y} > 0$, les deux variables varient dans le même sens.
 - Si $r_{X,Y} < 0$, les deux variables varient en sens inverse.
 - Si $r_{X,Y} = 0$, les deux variables sont linéairement indépendantes.
- $-1 \leq r_{X,Y} \leq 1$

Interprétation de la valeur de $r_{X,Y}$

- $r_{X,Y} = 1$, on dit qu'il y a une parfaite corrélation linéaire positive entre les deux variables.
- $r_{X,Y} = -1$, on dit qu'il y a une parfaite corrélation linéaire négative entre les deux variables.
- $r_{X,Y} = 0$, on dit qu'il y a absence de corrélation linéaire entre les deux variables
- On dit qu'il y a une forte corrélation linéaire entre les deux variables (ou forte dépendance linéaire) si $r_{X,Y}$ est proche de ± 1 . En revanche, si $r_{X,Y}$ est proche de 0, on dit qu'il y a une faible corrélation linéaire entre les deux variables.

Coefficient de détermination

- Afin de mesurer la qualité de l'ajustement linéaire du modèle, on définit le coefficient de détermination noté R^2 , par la part de la variance expliquée dans la variance totale.

$$R^2 = \frac{\text{Variance Expliquée}}{\text{Variance Totale}}$$

- On peut retenir le coefficient de détermination comme étant le carré du coefficient de corrélation linéaire entre X et Y .

$$R^2 = r_{X,Y}^2 = \left(\frac{Cov(X, Y)}{\sigma_X \sigma_Y} \right)^2$$

- Le coefficient de détermination est aussi égal au produit des pentes des deux droites de régression de Y en X et de X en Y .

$$R^2 = a * a'$$

Exemple 1

Soit la série statistique suivante relative aux deux variables X (les dépenses en recherche) et Y (le profit) :

Dépenses en recherche X	Profit Y
40	50
42	60
26	40
45	50
60	80

Travail à faire:

- 1.** Déterminer le coefficient de corrélation linéaire entre X et Y.
Interpréter.
- 2.** Déterminer la droite de régression qui permet d'estimer Y en fonction de X.
- 3.** En faisant l'hypothèse que dans l'avenir, le phénomène va suivre la même évolution ; estimer le profit pour une dépense en recherche de 80.

Exemple 2

Considérons un échantillon de 10 employés du même âge d'une certaine entreprise. Soit X le nombre d'années d'études effectuées et Y le revenu mensuel (en dinars)

X	Y
6	410
8	720
9	480
10	820
10	610
11	860
12	1020
14	780
16	1120
18	1080

Travail à faire:

- 1) Déterminer le coefficient de corrélation entre le nombre d'années d'études (X) et le revenu mensuel (Y). Interpréter.
- 2) Donner la droite de régression qui permet d'estimer Y en fonction de X (une droite du type : $Y_i = \alpha X_i + \beta + \varepsilon_i$.)
- 3) Calculer le coefficient de détermination. Interpréter.

SÉRIES TEMPORELLES

Série chronologique ou série temporelle (STP)

Une série chronologique provient de la réalisation d'une famille de variables aléatoires $\{y_t, t \in I\}$, où l'ensemble I est un intervalle de temps qui peut être discret ou continu. Dans la suite, nous utilisons l'ensemble $I = \{0, 1, \dots, T\}$, où T est le nombre total d'observations.

Exemples: PNB, X, M, séries financières, taux de change, cours boursiers...

L'analyse des STP permet l'étude des situations passées et présentes et peut extrapoler l'événement dans un futur relativement proche.

La prévision se fonde sur la connaissance du passé et du présent.

Parmi les conditions nécessaires pour la prévision des STP est l'identification du processus stochastique qui engendre la STP en question.

On distingue 4 composantes pour une STP:

- Composante de tendance : T
- Composante cyclique : C
- Composante périodique : P
- Composante stochastique (aléatoire) : ε

$$T + C + P = DT \text{ (*composante déterministe*)}$$

$$\varepsilon = DS \text{ (*composante stochastique*)}$$

$$y_t = DT_t + DS_t$$

Bruit blanc

- On dit que la suite de variables aléatoires $\{\varepsilon_t\}$ constitue un **bruit blanc faible** si elle possède les propriétés suivantes :

$$E(\varepsilon_t) = 0 \text{ pour tout } t \in Z$$

$$E(\varepsilon_t^2) = \sigma_\varepsilon^2 \neq 0 \text{ et constante}$$

$$\text{Cov}(\varepsilon_s, \varepsilon_t) = 0 \text{ si } t \neq s.$$

- En d'autres termes, les variables aléatoires ε_t sont de moyenne nulle, de variance constante et non corrélées.

On dit que $\{\varepsilon_t\}$ est un *bruit blanc fort* s'il est un bruit blanc faible et que les variables aléatoires ε_t sont i.i.d. (indépendantes et identiquement distribuées).

Notation:

- Si $\{\varepsilon_t\}$ bruit blanc faible, alors $\{\varepsilon_t\} \sim \text{WN}(0, \sigma_\varepsilon^2)$;
- Si $\{\varepsilon_t\}$ bruit blanc fort, alors $\{\varepsilon_t\} \sim \text{I. I. D}(0, \sigma_\varepsilon^2)$.

Présentation des STP

- Un modèle défini le mécanisme par lequel les observations sont construites.

Exemple: modèle simple

$$MA(1): y_t = \varepsilon_t + \theta \varepsilon_{t-1} \quad (\textcolor{red}{a}) \quad \text{avec } \varepsilon_t \sim WN$$

Un ensemble particulier de valeurs de $(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_T)$ réalise une série correspondante (y_0, y_1, \dots, y_T) . Par conséquent, le modèle **(a)** peut définir une infinité de réalisations.

■ Le modèle définit une distribution conjointe de (y_0, y_1, \dots, y_T) .

A l'instant t , on a:

$$\mu_t = E(y_t)$$

$$V(y_t) = E(y_t - \mu_t)^2$$

$$cov(y_t, y_{t-s}) = E(y_t - \mu_t)(y_{t-s} - \mu_{t-s})$$

Dans les STP, on considère une seule réalisation.

Stationnarité

- Une propriété importante des séries chronologiques est la stationnarité. Cette propriété est nécessaire pour appliquer certains théorèmes sur la causalité.
- *La stationnaire au sens large* exige que ces moments soient indépendants de t :

$$E(y_t) = \mu ; \sigma^2 = V(y_t) = \gamma(0) ; \gamma(s) = E(y_t - \mu)(y_{t-s} - \mu)$$

$$\mu_t = \mu \qquad \qquad \sigma^2(t) = \sigma^2 \qquad \qquad \gamma_t(s) = \gamma(s)$$

Estimateurs

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T y_t$$

$$\hat{\gamma}(0) = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{\mu})^2$$

$$\hat{\gamma}(s) = \frac{1}{T} \sum_{t=s+1}^T (y_t - \hat{\mu})(y_{t-s} - \hat{\mu})$$

Ces estimateurs convergent vers les moments théoriques si y_t est ergodique.

L'ergodicité implique que les observations assez distantes dans le temps ne sont pas corrélées.

En général, on considère les modèles pour lesquels stationnarité implique ergodicité.

- *La stationnarité au sens strict* implique que la densité conjointe reste inchangée à travers le temps

$$F(y_{t_1}, y_{t_2}, \dots, y_{t_K}) = F(y_{t_1+s}, y_{t_2+s}, \dots, y_{t_{K+s}})$$

Fonction d'autocorrélation (FAC)

- $\gamma(s)$: covariance d'ordre s

$$\gamma(s) = E(y_t - \mu)(y_{t-s} - \mu) = COV(y_t, y_{t-s})$$

- $\rho(s)$: coefficient de corrélation d'ordre s

$$\rho(s) = \frac{\gamma(s)}{\gamma(0)}$$

La FAC montre comment les observations du processus sont corrélées entre elles.

$\rho(s)$ en fonction de s et $\rho(0) = 1$.

- $\rho(s)$ coefficients théoriques inconnus estimés par:

$r(s) = \frac{\hat{\gamma}(s)}{\hat{\gamma}(0)}$: coefficient de corrélation empirique

Opérateur de retard

- L'opérateur de retard B se définit de la manière suivante :

$$By_t = y_{t-1} ,$$

$$B^2y_t = y_{t-2} ,$$

$$B^n y_t = y_{t-n} \text{ pour tout } n \in \mathbb{N}.$$

Opérateur de différence

- On définit l'opérateur de différence Δ comme suit :

$$\Delta = 1 - B$$

$$\Delta y_t = (1 - B)y_t = y_t - y_{t-1}$$

$$\Delta^2 y_t = (1 - B)^2 y_t = y_t - 2y_{t-1} + y_{t-2}$$

- Ces opérateurs peuvent être utilisés afin de transformer un processus de moyenne non nulle en un processus de moyenne nulle.
- On peut aussi s'en servir pour enlever la composante saisonnière de la série.
- On les utilise aussi pour mieux représenter les modèles.



LE PROCESSUS AUTORÉGRESSIF

Définition, Stationnarité, Inversibilité

Elaboré par: BEN KHALIFA Tasnim
MOUEDDEB Amal
BETTAIB Arij

C'est quoi un processus AR?

- Un processus autorégressif est un modèle de régression pour séries temporelles dans lequel la série est expliquée par ses valeurs passées plutôt que par d'autres variables.
- Un processus autorégressif d'ordre p, noté AR(p) est donné par :

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t.$$

- Où $\varphi_1, \dots, \varphi_p$ Sont des paramètres, ε_t un bruit blanc

$$\{\varepsilon_t\} \sim WN(0, \sigma_\varepsilon^2)$$

- De la même façon on peut écrire le processus AR(p) en utilisant L comme opérateur des retards, on peut l'écrire:

$$(1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p) X_t = \varepsilon_t.$$

- Où l'opérateur de retard L est défini comme $L(X_t) = X_{t-1}$
et pour tout $n \in \mathbb{N}$ on a $L^{[n]}(X_t) = X_{t-n}$

Moments d'un processus AR(1)

Pour calculer les différents moments d'un processus AR(1), soit son espérance, sa variance, son autocovariance et son autocorrélation, on va supposer que les bruits blancs sont indépendamment et identiquement distribués, d'espérance nulle et de variance σ^2

Espérance

$$\mathbf{E}[X_t] = \varphi^t X_0 + c \sum_{i=0}^{t-1} \varphi^i$$

Variance

$$\text{Var}[X_t] = \sum_{i=0}^t \varphi^{2i} \sigma^2$$

Autocovariance

$$\text{Cov}[X_t, X_{t-j}] = \varphi^j \sum_{i=0}^{\infty} \varphi^{2i} \sigma^2$$

Autocorrélation

$$\text{Corr}[X_t, X_{t-j}] \equiv \frac{\text{Cov}[X_t, X_{t-j}]}{\sqrt{\text{Var}(X_t) \text{Var}(X_{t-j})}} = \varphi^j \sqrt{\frac{1 - \varphi^{2(t-j)+2}}{1 - \varphi^{2t+2}}}$$

Moments d'un Processus AR(p)

$$\mathbb{E}(X_t) = \frac{c}{1 - \varphi_1 - \varphi_2 - \dots - \varphi_p}$$

$$\text{Var}(X_t) = \varphi_1 \gamma_1 + \varphi_2 \gamma_2 + \dots + \varphi_p \gamma_p + \sigma^2$$

$$\text{Cov}(X_t, X_{t-j}) = \varphi_1 \gamma_{j-1} + \varphi_2 \gamma_{j-2} + \dots + \varphi_p \gamma_{j-p}$$

Les formules de la variance et de la covariance correspondent aux équations dites de Yule et walker

Stationnarité et inversibilité d'un processus AR

◦ Résoudre $\varphi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p = 0$

Si $|racines| > 1$ (en dehors du disque unitaire) Alors le processus est stationnaire.

Équivalent: $f(z) = z^p - \varphi_1 z^{p-1} - \varphi_2 z^{p-2} - \dots - \varphi_{p-1} z - \varphi_0 = 0$

Si $|racines| < 1$ (à l'intérieur du disque unitaire) Alors le processus est inversible.

Dans les deux cas, stationnarité et inversibilité pour un processus AR(p) implique stabilité.

o

Explication:

soit $X_t \sim AR(p)$

$$\underbrace{\varphi(B)}_{en dehors} X_t = \varepsilon_t \Leftrightarrow X_t = \overbrace{\frac{1}{\varphi(B)}}^{\text{à l'intérieur}} \varepsilon_t$$

Sources

- <https://openclassrooms.com/fr/courses/4525371-analysez-et-modelisez-des-series-temporelles/5001211-de-couvrez-les-processus-stationnaires>
- <https://openclassrooms.com/fr/courses/4525371-analysez-et-modelisez-des-series-temporelles/5001216-les-processus-ar-ma-et-arma>
- [https://fr.wikipedia.org/wiki/Processus_autor%C3%A9gressif#Processus_AR\(p\)](https://fr.wikipedia.org/wiki/Processus_autor%C3%A9gressif#Processus_AR(p))

PROCESSUS AUTORÉGRESSIF: FAC ET FACP

Élaboré par :

Yasmine Faddey

Ghofran Bousbih

DEFINITION

Dans un modèle autorégressif (AR) la variable dépendante dépend de ses propres valeurs décalées jusqu'à un décalage de temp p et d'une fraiche réalisation d'un processus aléatoire d'où la notation AR(p).

Ce modèle est utilisé pour estimer la composante irrégulière d'un série temporelle et éventuellement sa composante saisonnière.

On dit qu'un processus (X_t) (stationnaire) est un processus AR (Autorégressive) d'ordre p , noté AR(p) , si :

$$X_t + \sum_{i=1}^p \phi_i X_{t-i} = \varepsilon_t, \quad \forall t \in \mathbf{Z}$$

avec $\phi_i \in \mathbb{R}$ et ε_t un bruit blanc de variance σ^2 .

Cette relation s'écrit également:

$$\Phi(L)X_t = \varepsilon_t$$

$$\text{ où } \Phi(L) = 1 + \phi_1 L + \dots + \phi_p L^p$$

La modélisation de X_t se résume à une relation linéaire liant aux p derniers instants.

Un bruit blanc discret se définit tout simplement comme **un processus aléatoire stationnaire** dont la **fonction d'autocorrélation** est nulle partout sauf en zéro.

Rappelons que l'autocorrélation est maximale en 0 et que sa valeur en ce point est égal à **la puissance du bruit blanc centré**.

- $E[\epsilon_k] = 0$;
- $E[\epsilon_k^2] = \sigma^2$;
- $E[\epsilon_k \epsilon_l] = 0 \quad \forall k \neq l$.

FONCTION D'AUTOCORRÉLATION D'UN PROCESSUS AR(P)

soit un processus stationnaire AR(p) défini par

$$\Phi(L)X_t = X_t + \sum_{i=1}^p \phi_i X_{t-i} = \varepsilon_t$$

dont les racines sont de modules > 1.

Alors la fonction d'autocorrélation:

$$\rho(h) + \sum_{j=1}^p \phi_j \rho(h-j) = 0, \quad h = 1, 2, \dots$$

est donc définie comme la solution d'une suite récurrente d'ordre p, donc racines non-nulles de l'équation:

$$\lambda^p + \sum_{j=1}^p \phi_j \lambda^{p-j} = 0$$

Soit:

$$1 + \sum_{j=1}^p \phi_j \lambda^{-j} = 0$$

c'est à dire que λ est racine de Φ donc $|\lambda| < 1$. On sait donc que $\rho(h)$ est une combinaison de différentes composantes: exponentielles décroissantes pour les racines réelles, sinusoïdes amorties pour les racines complexes, exponentielles décroissantes et termes polynomiaux pour les racines multiples. Par conséquent, $\rho(h)$ tend exponentiellement vers 0 avec h .

Remarquons également qu'on a la relation matricielle (équation de Yule-Walker):

$$\begin{pmatrix} -\phi_1 \\ \vdots \\ -\phi_p \end{pmatrix} = [\rho(i-j)]^{-1} \begin{pmatrix} \rho(1) \\ \vdots \\ \rho(p) \end{pmatrix}$$

Ce qui donne un algorithme d'estimation des coefficients d'un processus AR(p) en remplaçant $\rho(h)$ par leur autocorrélation empirique.

Remarquons également que la variance de ce processus vaut:

$$\gamma(0) = - \sum_{j=1}^p \phi_j \gamma(j) + \sigma^2 = \frac{\sigma^2}{1 + \sum_{j=1}^p \phi_j \rho(j)}$$

FONCTION D'AUTOCORRÉLATION PARTIELLE

- Le problème lorsqu'on adopte un processus $AR(p)$ est de déterminer p .
- Une première indication c'est la FAC. On peut aussi se référer à la FAC partielle qui déterminera p .
- L'idée est de mesurer la relation entre $(X_t, X_{t+\tau})$ dénotée par $\psi_{\tau\tau}$: coefficient d'autocorrélation partielle d'ordre τ .
- On tient compte de la relation $[(X_t, X_{t+\tau}), X_{t+1}]$ par le coefficient ψ_{21}

L'idée est analogue à celle d'un coefficient de régression en économétrie.

Ainsi, pour un couple (X_t, X_{t+1}) , on estime

$$X_{t+1} = \psi_{11}X_t + \varepsilon_{t+1} ; \text{ dans ce cas, } \psi_{11} = \rho_1$$

Pour le couple (X_t, X_{t+2}) , $X_{t+2} = \psi_{21}X_{t+1} + \psi_{22}X_t + \varepsilon_{t+2}$

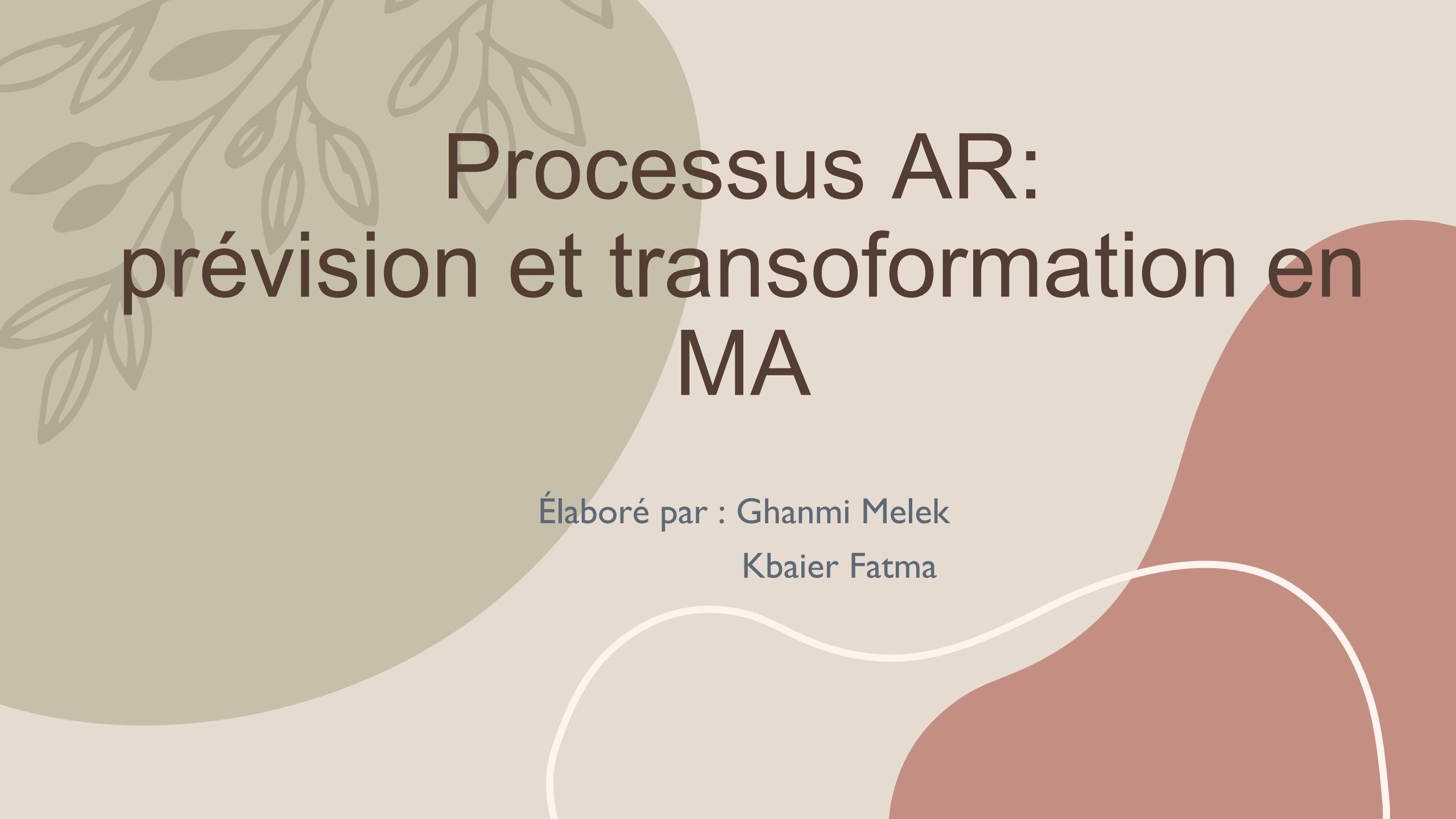
En utilisant l'ensemble d'équations récursives appelé Algorithme de Durbin, on peut calculer les coefficients $\psi_{\tau j}, \forall j = 1, 2, \dots, \tau$

$$\psi_{11} = \rho_1$$

$$\psi_{\tau\tau} = \frac{\rho_\tau - \sum_{j=1}^{\tau-1} \psi_{\tau-1,j} \rho_{\tau-j}}{1 - \sum_{j=1}^{\tau-1} \psi_{\tau-1,j} \rho_{\tau-j}} , \quad \forall \tau = 2, 3, \dots$$

$$\psi_{\tau j} = \psi_{\tau-1,j} - \psi_{\tau\tau} \psi_{\tau-1,\tau-j} ; \quad \forall \tau = 2, 3, \dots \text{ et } \forall j = 1, 2, \dots, \tau$$

MERCI POUR VOTRE ATTENTION



Processus AR: prévision et transformation en MA

Élaboré par : Ghanmi Melek
Kbaier Fatma



Plan

Définition d'un AR(p)

Ecriture moyenne mobile
infinie d'un AR(p)

Prédiction d'un AR(p)

Définition d'un AR(p)

définition on appelle processus autorégressif d'ordre p , usuellement noté AR(p), un processus stationnaire $(X_t)_{t \in \mathbf{Z}}$ vérifiant une relation du type:

$$X_t + \sum_{i=1}^p \phi_i X_{t-i} = \varepsilon_t, \quad \forall t \in \mathbf{Z}$$

avec $\phi_i \in \mathbf{R}$ et ε_t un bruit blanc de variance σ^2 .

Ecriture moyenne mobile infinie d'un AR(p)

Ecriture moyenne mobile infinie la définition du processsus AR(p) proposée n'implique pas forcément que cette équation a une solution unique stationnaire. Pour celà, il faut relier ce problème à la question de l'inversion du polynôme en L : $\Phi(L)$ vu précédemment.

- si Φ a toutes ses racines de module différent de 1 on peut inverser $\Phi(L)$ et l'équation a une unique solution avec une écriture moyenne mobile infinie:

$$X_t = \Phi(L)^{-1} \varepsilon_t = \sum_{j=-\infty}^{+\infty} h_j \varepsilon_{t-j}, \quad \sum_{j=-\infty}^{+\infty} |h_j| < +\infty$$

on voit que dans ce cas la valeur de X_t peut dépendre des valeurs passées, présentes et futures de ε_t . Dans le cas où les racines de Φ sont de module strictement supérieur à 1, $\Phi^{-1}(L)$ ayant un développement ne contenant que des puissances positives de L , la valeur de X_t ne dépend que des valeurs passées de ε_t :

$$X_t = \sum_{j=0}^{+\infty} h_j \varepsilon_{t-j}, \quad \sum_{j=0}^{+\infty} |h_j| < +\infty, \quad h_0 = 1$$

Prediction d'un AR(p)

Les AR(p) sont bien adaptés à la prédition. Les prévisions optimales $\hat{X}_t(h)$ pour $h > 0$ sont des combinaisons linéaires de $X_t, X_{t-1}, \dots, X_{t-p+1}$ comme d'habitude. On les calcule de façon récurrente.

A l'horizon 1 : on a

$$X_{t+1} = \phi_1 X_t + \phi_2 X_{t-1} + \dots + \phi_p X_{t+1-p} + \eta_{t+1}$$

que l'on prédit par

$$\hat{X}_t(1) = \phi_1 X_t + \phi_2 X_{t-1} + \dots + \phi_p X_{t+1-p}$$

A l'horizon 2 : on a

$$X_{t+2} = \phi_1 X_{t+1} + \phi_2 X_t + \dots + \phi_p X_{t+2-p} + \eta_{t+2}$$

que l'on prédit par

$$\hat{X}_t(1) = \phi_1 \hat{X}_t(1) + \phi_2 X_t + \dots + \phi_p X_{t+2-p}$$

et ainsi de suite.

Théorème 2.9 Soit X un AR(p) de bruit blanc d'innovation η et de polynôme canonique $\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$. Les prévisions optimales $\hat{X}_t(h)$ pour $h > 0$ sont des combinaisons linéaires de $X_t, X_{t-1}, \dots, X_{t-p+1}$:

$$\hat{X}_t(h) = \sum_{0 \leq j \leq p-1} a_j(h) X_{t-j}$$

qui s'obtiennent par récurrence selon

$$\hat{X}_t(h) = \sum_{1 \leq k \leq p} \phi_k \hat{X}_t(h-k)$$

avec les conditions initiales $\hat{X}_t(-j) = X_{t-j}$ pour $j \geq 0$.

On peut donner des formules explicites pour les coefficients $a_j(h)$ mais c'est un peu compliqué, surtout si Φ a des racines multiples. Du point de vue du calcul par un ordinateur, il est en général plus simple, et tout aussi efficace, d'évaluer $\hat{X}_t(h)$ par récurrence comme ci-dessus. Lorsque Φ n'a pas de racines multiples, on peut aussi procéder selon la remarque suivante.



Merci pour votre attention

Processus Moyennes Mobiles (MA)

élaboré par : Fehd Ounis et Mohamed Amine Hayouni

I. Définition

□ On appelle processus moyenne mobile d'ordre q, noté MA(q) pour Moving Average un processus $\{Y_t\}, \forall t \in \mathbb{Z}$ défini par :

$$Y_t = \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad \forall t \in \mathbb{Z}$$

où les θ_i sont des réels, $\theta_q \neq 0$ et $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ est un processus bruit blanc de variance σ^2 .

$$\iff Y_t = (1 - \theta_1 L - \dots - \theta_q L^q) \varepsilon_t \iff Y_t = \Theta(L) \varepsilon_t$$

II. Représentation stationnaire et causale

La définition d'un MA(q) est explicite et ne pose donc pas de problème : le processus Y_t est parfaitement défini et est automatiquement stationnaire.

La représentation est causale par définition.

III. Représentation invisible:

$\Theta(L)$ est un polynôme en L de degré q , que l'on peut factoriser en ayant calculé ses racines $z_i = 1/\lambda_i$, $i = 1, \dots, q$:

$$\Theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q = (1 - \lambda_1 L)(1 - \lambda_2 L) \dots (1 - \lambda_q L)$$

Si $\Theta(z)$ n'a pas de racine de module égal à 1, on peut calculer l'inverse de $\Theta(L)$, qui est alors donnée par :

$$\Theta(L)^{-1} = (1 - \lambda_1 L)^{-1}(1 - \lambda_2 L)^{-1} \dots (1 - \lambda_q L)^{-1} = \frac{k_1}{1 - \lambda_1 L} + \dots + \frac{k_q}{1 - \lambda_q L}$$

si toutes les racines de Θ (c'est-à-dire $1/\lambda_i$, $i = 1, \dots, q$) sont distinctes et où k_1, \dots, k_q sont des paramètres qui dépendent de $\lambda_1, \dots, \lambda_q$.

On obtient alors l'expression suivante :

$$\Theta(L)^{-1} Y_t = \sum_{i=-\infty}^{\infty} \pi_i Y_{t-i} = \varepsilon_t$$



où les π sont fonctions des paramètres θ_j , et on peut montrer que $\sum_{i=-\infty}^{\infty} |\pi_i| < \infty$.

- Si les racines de $\Theta(z) = 0$ sont toutes de module supérieur à 1, on peut montrer que $\pi_i = 0 \quad \forall i < 0$ et ε_t s'interprète comme l'innovation du processus. On dit que le processus est inversible.
- Si les racines de $\Theta(z) = 0$ sont inférieures à 1 en module, mais qu'il n'y a pas de racines de module égal à 1, on peut inverser les racines, quitte à changer de bruit blanc, et supposer que le processus est inversible.

On a alors la forme $AR(\infty)$ suivante :

$$Y_t = - \sum_{i=1}^{\infty} \pi_i Y_{t-i} + \varepsilon_t \quad \sum_{i=1}^{\infty} |\pi_i| < \infty$$

les π_i s'obtenant par division croissante de 1 par Θ (avec $\pi_0 = 1$).

Ainsi, Y_t dépend de ε_t et de son passé, donc Y_{t-1} dépend de ε_{t-1} et de son passé, cet ensemble étant indépendant de ε_t (puisque ε_t est un BB), donc le passé de Y_t est indépendant de ε_t , donc ε_t est l'innovation de Y_t .

IV. Exemple du MA

Exemple du MA(1) : $Y_t = \varepsilon_t - \theta \varepsilon_{t-1} = (1 - \theta L)\varepsilon_t$ avec ε_t BB $(0, \sigma_\varepsilon^2)$

La racine de $1 - \theta z = 0$ est $z = 1/\theta$.

- Si $|\theta| = 1$ alors la forme $AR(\infty)$ n'existe pas, la représentation est stationnaire, causale mais non inversible, donc non canonique.
- Si $|z| > 1$ et donc $|\theta| < 1$ alors :

$$\varepsilon_t = (1 - \theta L)^{-1}Y_t = \sum_{i=0}^{\infty} \theta^i L^i Y_t = Y_t + \sum_{i=1}^{\infty} \theta^i Y_{t-i}$$

et la forme $AR(\infty)$ s'écrit :

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

La représentation est alors inversible, en plus d'être stationnaire et causale \implies elle est donc canonique.

- Si $|z| < 1$ et donc $|\theta| > 1$ alors :

$$\varepsilon_t = (1 - \theta L)^{-1} Y_t = - \sum_{i=1}^{\infty} \theta^{-i} L^{-i} Y_t = - \left(\frac{Y_{t+1}}{\theta} + \frac{Y_{t+2}}{\theta^2} + \dots \right)$$

Dans ce cas, on peut toujours (tant que $|\theta| \neq 1$) se ramener à une représentation canonique quitte à changer la représentation et surtout à changer de BB et à inverser les racines.

En effet, la fonction génératrice des moments du MA(1) initial est donnée par :

$$\begin{aligned}\Gamma_Y(z) &= (1 - \theta z)(1 - \theta z^{-1}) \\ &= \theta z \left(\frac{1}{\theta z} - 1 \right) \left(\frac{1}{\theta z^{-1}} - 1 \right) \theta z^{-1} \\ &= \theta^2 \left(1 - \frac{1}{\theta} z^{-1} \right) \left(1 - \frac{1}{\theta} z \right)\end{aligned}$$

soit la fonction génératrice des moments d'un MA(1) où la racine est $1/\theta$ et le nouveau BB (η_t) est de variance $\theta^2 \sigma_\varepsilon^2$, soit une représentation de la forme :

$$Y_t = \eta_t - \frac{1}{\theta} \eta_{t-1}$$



Une autre manière de faire est de passer par la fonction d'autocovariance, en identifiant les paramètres de la représentation canonique :

$$Y_t = \varepsilon_t - \theta \varepsilon_{t-1} = \eta_t - \alpha \eta_{t-1}$$

Ainsi,

$$\begin{aligned}\gamma(0) &= (1 + \theta^2)\sigma_\varepsilon^2 = (1 + \alpha^2)\sigma_\eta^2 \\ \gamma(1) &= -\theta\sigma_\varepsilon^2 = -\alpha\sigma_\eta^2\end{aligned}$$

Les valeurs solutions de α vérifient alors :

$$\alpha^2\theta - \alpha(1 + \theta^2) + \theta = 0$$

on obtient alors $\alpha = 1/\theta$ ou $\alpha = \theta$. On choisit la solution qui est inférieure à 1 en module, soit $\alpha = 1/\theta$ puisqu'on est dans le cas $|\theta| > 1$. On a alors $\alpha = 1/\theta$ et $\sigma_\eta^2 = \theta^2\sigma_\varepsilon^2$.

Conclusion : la représentation $Y_t = \Theta(L)\varepsilon_t$ est canonique si les racines de $\Theta(z)$ sont supérieures à 1 en module.



Merci pour votre attention

COMPARAISON ENTRE PROCESSUS MOYENNES MOBILES (MA) & AUTORÉGRESSIFS (AR)

Projet élaboré par
Boujneh Souha & Guesmi Nour

PLAN

01 **Processus autorégressif
AR**

- Définition
- Caractérisation

02 **Processus moyenne
mobile MA**

- Définition
- Caractérisation

03 **$MA(q) \approx AR(\infty)$**

04 **Distinction entre AR(1)
et MA(1)**

01-PROCESSUS AUTORÉGRESSIF (AR)

- Définition

On appelle processus autorégressif d'ordre p, usuellement noté AR(p), un processus stationnaire $(X_t) \forall t \in \mathbb{Z}$ vérifiant une relation du type:

$$X_t + \sum_{i=1}^p \phi_i X_{t-i} = \varepsilon_t, \quad \forall t \in \mathbb{Z}$$

Avec $\phi_i \in \mathbb{R}$ et ε_t un bruit blanc de variance σ^2 .

- **Caractérisation**

Si $(X_t) \forall t \in Z$ est un processus AR(p) alors ses autocorrélations partielles s'annulent à partir du rang p+1

$$\begin{cases} r(p) \neq 0 \\ \forall h \in \mathbb{N}, h \geq p + 1 : r(h) = 0 \end{cases}$$

Les autocorrélations simples, quant à elles, décroissent rapidement vers 0 (de manière exponentielle ou sinusoïdale amortie).

02-PROCESSUS MOYENNES MOBILES (MA)

- **Définition**

On dit qu'un processus $(X_t) \forall t \in \mathbb{Z}$ est un processus MA (Moving Average) d'ordre q , noté $\text{MA}(q)$, si :

$$\forall t \in \mathbb{Z} : X_t = \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i}$$

Où $(\theta_1, \dots, \theta_q) \in \mathbb{R}^q$ et $\theta_q \neq 0$.

On considère ici que le processus est la résultante d'une combinaison linéaire de perturbations décorrélées (un bruit blanc et son passé).

- **Caractérisation**

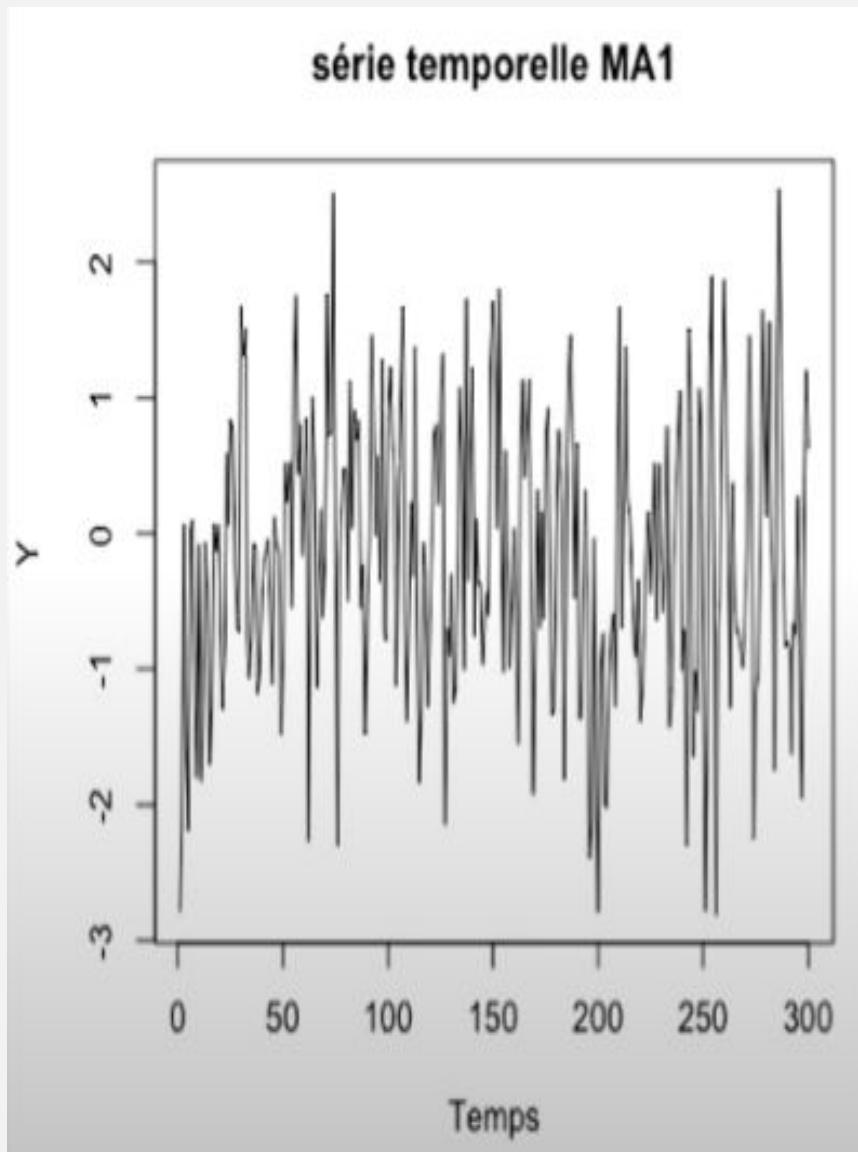
Si $(X_t) \forall t \in Z$ est un processus MA(q) alors ses autocorrélations simples s'annulent à partir du rang $q+1$:

$$\begin{cases} \rho(q) \neq 0 \\ \forall h \in \mathbb{N}, h \geq q+1 : \rho(h) = 0 \end{cases}$$

Il s'agit d'une condition nécessaire et suffisante pour qu'un processus $(X_t) \forall t \in Z$ soit un MA(q).

- Exemple:

- La série présentée à gauche est une simulation d'un processus MA(1)
- $y_t = \varepsilon_t + 0.5\varepsilon_{t-1}$
- Un évènement météorologique peut être un exemple de série qui suit une moyenne mobile :
 - humidité = pluie $_t$ + pluie $_{t-1}$
 - S'il pleut une journée, l'humidité sera forcément plus élevée. Le lendemain, il restera des flaques d'eau et de l'humidité.
 - La pluie représente ici un choc stochastique.



03- $MA(1) \approx AR(\infty)$

Soit Y_t un processus moyenne mobile d'ordre 1.

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

ε_t est un bruit blanc.

$$E(\varepsilon_t) = 0$$

$$V(\varepsilon_t) = \sigma^2$$

$$E(\varepsilon_t \varepsilon_k) = 0 \quad \forall t \neq k$$

Soit $\gamma_k = COV(Y_t, Y_{t-k})$

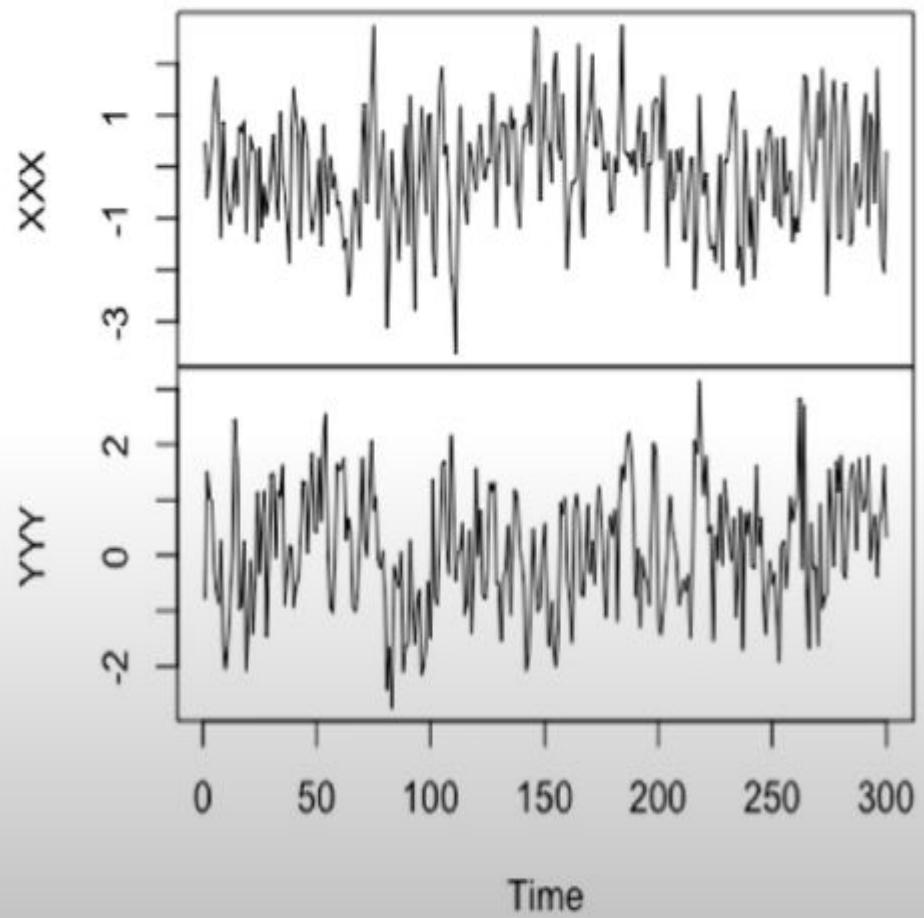
1- Exprimer ε_t en fonction de Y_t . Que peut-on conclure.

$$\begin{aligned}
 1/ \quad y_t &= \varepsilon_t + \theta B \varepsilon_t \Rightarrow y_t = (1 + \theta B) \varepsilon_t \\
 \Rightarrow \quad \varepsilon_t &= \frac{1}{1 + \theta B} = \frac{1}{1 - (-\theta B)} \\
 \Rightarrow \quad \varepsilon_t &= y_t (1 + (-\theta B) + (-\theta B)^2 + (-\theta B)^3 + \dots) \\
 \Rightarrow \quad \varepsilon_t &= y_t - \theta y_{t-1} + \theta^2 y_{t-2} - \theta^3 y_{t-3} + \theta^4 y_{t-4} + \dots
 \end{aligned}$$

Donc $MA(1) \Leftrightarrow AR(\infty)$

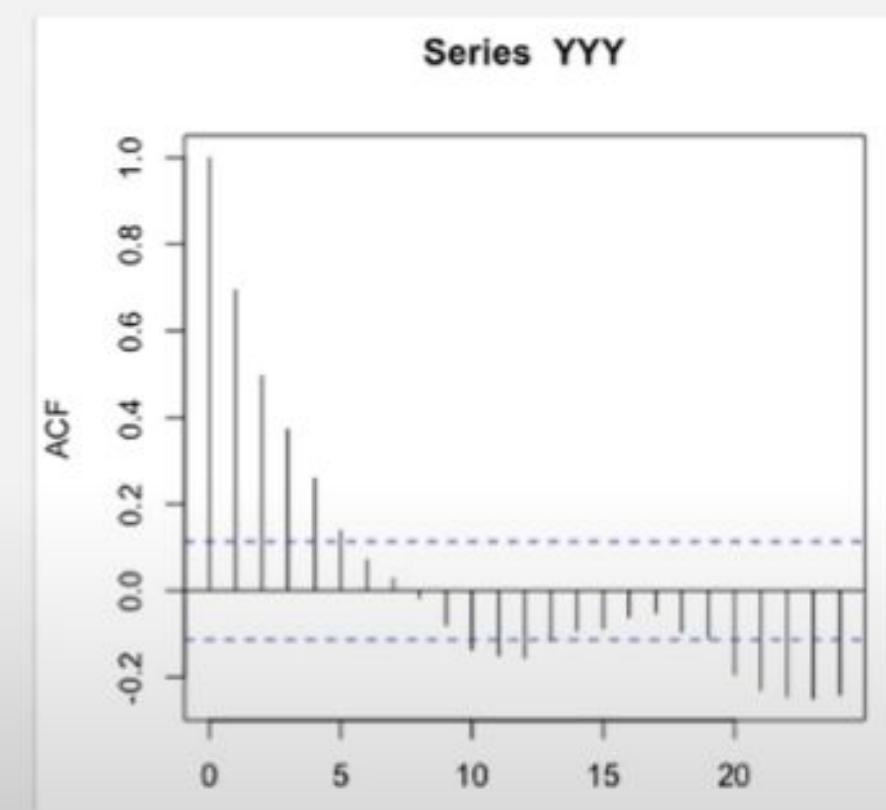
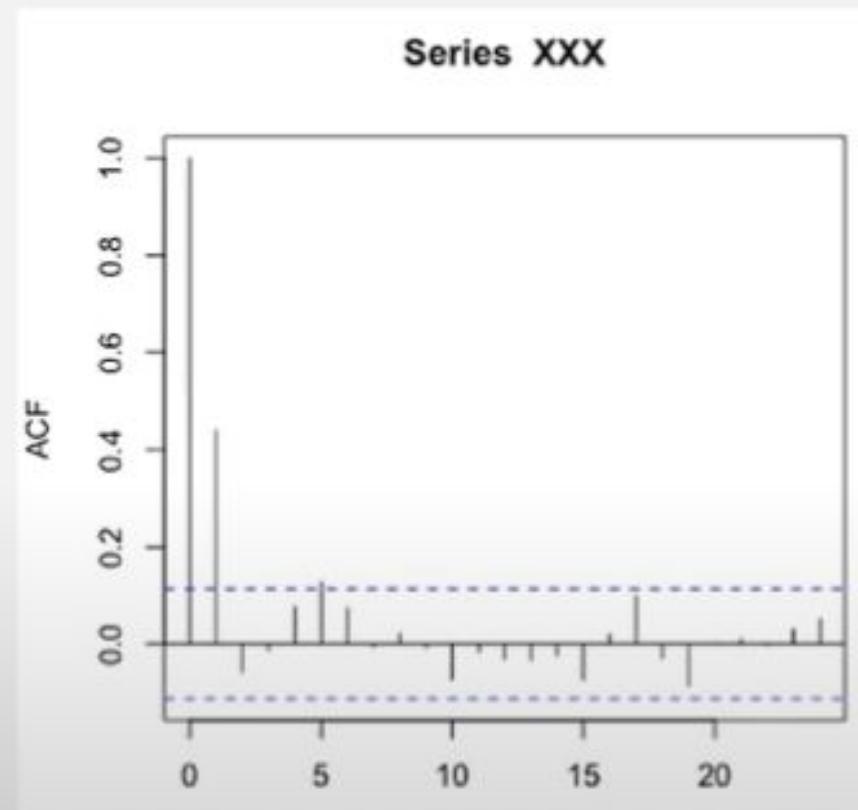
- Nous avons vu deux modèles, le modèle autorégressif et le modèle de moyenne mobile.
 - $y_t = \alpha y_{t-1} + k + \varepsilon_t$
 - $y_t = \varepsilon_t + \beta \varepsilon_{t-1}$
- On sait que la moyenne mobile devra être stationnaire (par définition du terme d'erreur).
- Mais autrement, pouvons nous facilement discerner ces deux séries?

04-DISTINCTION ENTRE AR(1) ET MA(1)



- Êtes vous capable de distinguer quelle série suit quel processus :
 - $y_t = 0.5y_{t-1} + \varepsilon_t$
 - $y_t = \varepsilon_t + 0.5\varepsilon_{t-1}$
- La série X est une moyenne mobile et la série Y est une série autorégressive !
- On ne pourra pas distinguer les processus AR et MA sans information additionnelle...

□ Fonction d'autocorrélation



- La fonction d'autocorrélation d'une série autorégressive se distingue très bien de celui d'une moyenne mobile.

RÉFÉRENCES

<https://www.youtube.com/watch?v=Ohg3iPlrmQ>

<https://www.youtube.com/watch?v=a-Jlw8gAdSI>

<https://openclassrooms.com/fr/courses/4525371-analysez-et-modelisez-des-series-temporelles/5001216-les-processus-ar-ma-et-arma>

Processus ARMA(p,q)

— Processus Moyenne mobile autorégressive —

Projet réalisé par :

Meddah sarra

Belgouthi halée

Définition

Soit le processus $\{Z_t\}$ tel que $E(Z_t^2) < \infty$ et supposons que . Le processus $\{Z_t\}$ est autorégr $E(Z_t) = 0$ aine mobile d'ordre (p,q) si il satisfait la relation:

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}, \forall t$$

- Le processus $\{a_t\}$ est un bruit blanc .
- Les paramètres $\phi_i, i = 1, \dots, p;$ $\theta_i, i = 1, \dots, q.$ sont des nombres réels.

Opérateur retard B (backward shift operator)

Soit le processus $\{Z_t\}$, l'opérateur retard B se définit comme suit:

$$Bz_t = z_{t-1}$$

$$B^2 z_t = B(Bz_t) = Bz_{t-1} = z_{t-2}$$

$$\vdots$$

$$B^m z_t = z_{t-m}$$

Réécriture du modèle ARMA à l'aide de l'opérateur retard

- On suppose également que $B^0 = I$ orte que $Iz_t = z_t$
- De plus: $(\alpha B^m)z_t = \alpha(B^m z_t) = \alpha z_{t-m}$
- L'opérateur retard est linéaire:

$$\begin{aligned}B(z_t + w_t) &= z_{t-1} + w_{t-1} = Bz_{t-1} + Bw_{t-1}, \\(\alpha B)z_t &= \alpha(Bz_t) = \alpha z_{t-1}.\end{aligned}$$

- Considérons l'opérateur polynomial B :

$$f(B) = \alpha_0 I + \alpha_1 B + \dots + \alpha_p B^p$$

- On a alors que:

$$\begin{aligned} f(B)z_t &= (\alpha_0 I + \alpha_1 B + \dots + \alpha_p B^p)z_t, \\ &= \alpha_0 z_t + \alpha_1 z_{t-1} + \dots + \alpha_p z_{t-p} \end{aligned}$$

- Posons: $\phi(B) = I - \phi_1 B - \dots - \phi_p B^p,$
 $\theta(B) = I - \theta_1 B - \dots - \theta_q B^q.$
- Il est économique d'écrire: $\phi(B)Z_t = \theta(B)a_t, \quad \forall t.$
- Si $E(Z_t) = \mu$, on dit que le processus $\{Z_t\}$ ARMA(p,q) si :
 $\phi(B)\dot{Z}_t = \theta(B)a_t, \quad \forall t.$
 $\dot{Z}_t = Z_t - \mu.$

Étude de la stationnarité d'un processus ARMA(p,q)

- Un processus $\{Z_t\}$ est dit stationnaire du second ordre si sa moyenne, sa variance et sa covariance sont indépendantes du temps et si sa variance est finie.

- Soit un processus $\{Z_t\}$ i est ARMA(p,q).
- Se demander si ce processus est stationnaire est de questionner s'il admet une représentation du genre:

$$Z_t = \sum_{j \geq 0} \psi_j a_{t-j}, \quad \forall t \quad \sum_{j \geq 0} |\psi_j| < \infty$$

- On rappelle que: $\phi(B)Z_t = \theta(B)a_t, \quad \forall t.$
- On aimeraient faire « disparaître » l'opérateur $\phi(B)$.
- On aimeraient multiplier par $\phi^{-1}(B)$ chaque côté

Un résultat stipule que pour avoir l'existence de l'opérateur $\phi^{-1}(B)$, il faut étudier les racines de l'équation :

$$\phi(z) = 0$$

- **Résultat fondamental:**

$\phi^{-1}(B)$ existe si et seulement si les racines de l'équation $\phi(z) = 0$ sont plus grandes que un en module.

Si $\phi^{-1}(B)$ est un opérateur qui existe, on a alors l'équation

$$\phi(B)Z_t = \theta(B)a_t$$

qui peut être multipliée de chaque côté par l'opérateur $, \phi^{-1}(B)$

ce qui nous donne: $\phi^{-1}(B)\phi(B)Z_t = \phi^{-1}(B)\theta(B)a_t$

$$Z_t = \phi^{-1}(B)\theta(B)a_t$$

$$Z_t = \Psi(B)a_t = \sum \psi_j a_{t-j}$$

Inversibilité d'un processus ARMA(p,q)

- la notion d'inversibilité consiste à déterminer s'il existe une représentation « MA » (respectivement pour ``AR'') équivalente pour « AR » (respectivement pour "MA")

- Soit un processus $\{Z_t\}$ st ARMA(p,q). Se demander si ce processus est inversible est se questionner si $\{Z_t\}$ admet une représentation du genre:

$$Z_t = \sum_{j \geq 1} \pi_j Z_{t-j} + a_t \quad \sum_{j \geq 1} |\pi_j| < \infty$$

- La discussion est sur la possibilité d'annulation de a_t . Dans , on veut multiplier de chaque côté par

$$\phi(B)Z_t = \theta(B)a_t$$

$$\theta^{-1}(B)$$

- L'opérateur $\theta^{-1}(B)$ existe si et seulement si les racines de l'équation plus grandes qu'un en module.
- Dans un tel cas:

$$\theta^{-1}(B)\phi(B)Z_t = \theta^{-1}(B)\theta(B)a_t$$

$$\theta^{-1}(B)\phi(B)Z_t = a_t$$

$$\Pi(B)Z_t = a_t$$

- On note que dans: $\Pi(B) = \theta^{-1}(B)\phi(B) = I - \sum_{j \geq 1} \pi_j B^j$

- Ainsi:

$$\begin{aligned}\Pi(B)Z_t &= a_t, \\ \left(I - \sum_{j \geq 1} \pi_j B^j\right)Z_t &= a_t, \\ Z_t &= \sum_{j \geq 1} \pi_j B^j Z_t + a_t, \\ Z_t &= \sum_{j \geq 1} \pi_j Z_{t-j} + a_t.\end{aligned}$$