**MASTERE DATA & IA**

**FONDEMENTS MATHEMATIQUES DE L’IA :**

**~MODEL ARIMA~**

**DOUKPENI D. B. KEVIN**



Professeur

**Hakim HORAIRY**

Sommaire

[**QU’EST-CE QUE ARIMA ?** 3](#_Toc137847215)

[**CONNAISSANCES DE BASE** 3](#_Toc137847216)

[**I.** **INTRODUCTION A ARIMA** 3](#_Toc137847217)

[**II.** **Paramètres ARIMA p, d, q** 4](#_Toc137847218)

[**III.** **IMPLEMENTATION DU MODELE ARIMA** 5](#_Toc137847219)

[**A.** **EXPLORER LE DATASET** 5](#_Toc137847220)

[**B.** **IDENTIFICATION DES ORDRES (p, q) DU MODELE** 6](#_Toc137847221)

[**C.** **ESTIMATION DES PARAMETRES DU MODELE** 8](#_Toc137847222)

[**D.** **PREDICTIONS ET EVALUATIONS** 10](#_Toc137847223)

[**CONCLUSION** 11](#_Toc137847224)

# **QU’EST-CE QUE ARIMA ?**

ARIMA est une classe générale de modèles statistiques pour la prévision de l’analyse des séries chronologiques. Il signifie **Auto-Regressive Integrated Moving Average**.

Lors de l’application des modèles ARIMA, nous utilisons les valeurs passées d’une série chronologique et/ou les erreurs de prévision pour prédire ses valeurs futures.

Dans la suite de ce rapport, nous allons explorer les concepts clés du modèle ARIMA ainsi que les étapes nécessaires à sa mise en œuvre.

# **CONNAISSANCES DE BASE**

Avant d’examiner les détails des modèles ARIMA, récapitulons quelques définitions dans l’analyse des séries chronologiques.

Une **série chronologique** est une séquence de points de données collectés dans l’ordre temporel. En fonction des caractéristiques des séries chronologiques, nous pouvons les diviser en différentes catégories, telles que univariée vs multivariée ou linéaire vs non linéaire.

Avant de construire des modèles ARIMA, nous devons vérifier si la série chronologique est **stationnaire**. Sinon, nous pouvons utiliser une technique appelée différenciation pour le transformer en stationnaire. Les modèles ARIMA couvrent cette transformation.

# **INTRODUCTION A ARIMA**

Comme mentionné précédemment, ARIMA est une classe générale de modèles. Pour différentes séries chronologiques, nous pourrions être en mesure de les adapter à l’une des combinaisons des trois composants d’ARIMA:

* **AR (Auto-Regressive) :** la série chronologique est régressive linéairement sur ses propres valeurs passées
* **I (Integrated) :** si elle n’est pas stationnaire, la série chronologique peut être différenciée pour devenir stationnaire, c’est-à-dire calculer les différences entre des observations consécutives
* **MA (Moving Average) :** la série chronologique est « régressée » sur les erreurs de prévision passées

Voici la relation de ces trois composantes dans ARIMA: AR et MA sont deux modèles qui fonctionnent généralement sur des séries temporelles stationnaires - si la série temporelle n’est pas stationnaire, nous pouvons utiliser la partie I pour la rendre stationnaire.

# **Paramètres ARIMA p, d, q**

Donc, correspondant aux trois composants d’ARIMA mentionnés ci-dessus, il y a trois paramètres **p**, **d** et **q**. Ils prennent des valeurs entières non négatives, indiquant quel modèle ARIMA spécifique est utilisé.

* **p :** le nombre de valeurs antérieures incluses dans le modèle AR

A black text on a white background

Description automatically generated with low confidence

où c’est une constante, **φ1**, ..., **φp** sont des paramètres, et **εt** est un bruit blanc

* **d :** Le nombre de fois que la série chronologique est différente. Par exemple, la différenciation du premier ordre est calculée comme suit :

A picture containing font, white, design, typography

Description automatically generated

* **q :** le nombre d’erreurs de prévision passées incluses dans le modèle MA, ou la taille de la fenêtre de moyenne mobile. Il s’appelle le modèle MA puisque chaque yt peut être considéré comme une moyenne mobile pondérée des erreurs de prévision passées

**A picture containing font, handwriting, line, text

Description automatically generated**

Ainsi, les modèles ARIMA sont en effet une classe générale de modèles, y compris AR, MA et ARMA. Par exemple, ARIMA(p, 0, 0) est équivalent à AR(p), ARIMA(0, 0, q) est équivalent à MA(q).

L’équation complète du modèle ARIMA(p, d, q) est:

A picture containing font, line, text, typography

Description automatically generated

où ∇yt est la série chronologique différente, qui peut être plus d’une différenciation temporelle

# **IMPLEMENTATION DU MODELE ARIMA**

Pour mettre en œuvre le modèle ARIMA, plusieurs étapes doivent être suivies.

## **EXPLORER LE DATASET**

Ce jeu de données décrit un décompte mensuel du nombre de taches solaires observées pendant un peu plus de 230 ans (1749-1983).

Les unités sont un comptage et il y a 2.820 observations.

A picture containing text, screenshot, plot, line

Description automatically generated

## **IDENTIFICATION DES ORDRES (p, q) DU MODELE**

La première étape consiste à analyser les données temporelles et à identifier les propriétés du processus. L'autocorrélation **ACF** et l'autocorrélation partielle **ACPF** peuvent être utilisées pour estimer les ordres p , d et q du modèle ARIMA.

Voici l’autocorrélation et l'autocorrélation partielle que nous avons obtenus pour notre jeu de données :

A picture containing text, screenshot, line, plot

Description automatically generated

A picture containing text, screenshot, plot, line

Description automatically generated

## **ESTIMATION DES PARAMETRES DU MODELE**

Il est maintenant temps de décider des deux autres paramètres p et q. Ils déterminent les modèles que nous utiliserons d’ARIMA.

Nous pouvons, encore une fois, utiliser les parcelles ACF et PACF. Nous allons modéliser en fonction de la première série chronologique de différence. Alors regardons les intrigues de nouveau fonction de coût.

A picture containing text, screenshot, line, rectangle

Description automatically generated

A picture containing text, screenshot, line, plot

Description automatically generated

* Si le graphique PACF a un pic significatif au décalage p, mais pas au-delà; la parcelle ACF se dégrade plus progressivement. Cela peut suggérer un modèle ARIMA(p, d, 0)
* Si le graphique ACF a un pic significatif au décalage q, mais pas au-delà; la parcelle PACF se dégrade plus progressivement. Cela peut suggérer un modèle ARIMA(0, d, q)

## **PREDICTIONS ET EVALUATIONS**

Avant d’utiliser ce modèle pour faire des prédictions de séries chronologiques, nous devons nous assurer que notre modèle a capturé des informations adéquates à partir des données. Nous pouvons vérifier cela en regardant les résidus. Si le modèle est bon, ses résidus devraient ressembler à du bruit blanc.

Nous allons tracer les résidus et leur densité.

A picture containing text, plot, diagram, screenshot

Description automatically generated

Calculons les prédictions en Python et traçons-les avec la série réelle, y compris l’ensemble de test.

A picture containing text, screenshot, plot, diagram

Description automatically generated

# **CONCLUSION**

Le modèle ARIMA est un outil puissant pour modéliser et prédire les séries temporelles. En combinant les termes autorégressifs et les termes moyens mobiles, il permet de capturer les structures complexes des données. Son implémentation nécessite une compréhension approfondie des termes AR et I et MA, ainsi que des étapes clés telles que l'identification du modèle, l'estimation des paramètres, le diagnostic et l'évaluation.