

# Examen Maths

*Sous la direction de M. Henri LAUDE*

**Chaymae GASMI**

# Introduction

Dans ce document, nous allons évaluer 5 travaux de mes collègues en Maths, pour cela, nous allons introduire les 5 critères d'évaluation choisis:

- Exécution du Rmd
- Qualité de la rédaction
- Qualité et lisibilité du RMarkdow
- Aspect didactique du dossier
- Bibliographie

## Les 5 travaux choisis:

- La Marche Aléatoire
- Dossier Approche mathématiques - Émojis
- Dossier Statsmodels
- Algèbre tropicale
- Arbre de décision
- XGboost "Travail personnel"

# La Marche Aléatoire

## 1. Entête

Nous allons évaluer le document 'Marche aléatoire' élaboré par William & Marko Arsic.

En cliquant ici vous trouverez le lien menant au GitHub:

[https://github.com/ARSICMrk/ARSIC\\_PSBx/tree/main/Maths\\_BD/Marche\\_aleatoire](https://github.com/ARSICMrk/ARSIC_PSBx/tree/main/Maths_BD/Marche_aleatoire)

([https://github.com/ARSICMrk/ARSIC\\_PSBx/tree/main/Maths\\_BD/Marche\\_aleatoire](https://github.com/ARSICMrk/ARSIC_PSBx/tree/main/Maths_BD/Marche_aleatoire)).

## 2. Synthèse

La marche aléatoire est un modèle mathématique qui permet de prédire le mouvement d'un système en fonction de son point de départ. On peut l'appeler également par marche de l'ivrogne, car Selon Pearson:

*"Dans un pays ouvert, l'endroit le plus probable pour trouver un ivrogne encore capable de tenir sur ses pieds se trouve quelque part dans le voisinage de son point de départ."*

Les pas de la marche aléatoire sont totalement décorrélés les uns des autres. Cela signifie intuitivement qu'à chaque instant, le futur du système dépend de son état présent, mais pas de son passé, même le plus proche. Autrement dit, le système « perd la mémoire » à mesure qu'il évolue dans le temps.

La marche aléatoire est décomposée en une unité de base appelée *pas*, et sa longueur elle-même peut être constante, aléatoire ou fixée par le réseau ou les graphiques que nous voulons parcourir. Par conséquent, à chaque étape, nous avons une gamme de possibilités pour sélectionner au hasard la direction et la taille du pas. Cette gamme de possibilités peut être discrète (choisir parmi un nombre limité de valeurs) ou continue.

## 3. Extrait

Les auteurs ont réussi à expliquer le concept de l'algorithme avec un exemple illustrative, celui de la puce qui se trouve à l'instant  $t=0$  en un point appelé origine. Chaque seconde, elle peut faire un pas vers la gauche, ou vers la droite, ou vers le haut ou vers le bas, avec une égale probabilité.

Soit  $d \geq 1$  et soit  $(e_1, \dots, e_d)$  la base canonique de  $\mathbb{Z}^d$ . Soit  $(X_i)$  une suite de variable aléatoires indépendantes à valeurs dans  $\{(\pm e_1, \dots, \pm e_d)\}$ . On appelle marche aléatoire associée la suite de variables aléatoires  $(S_n)_{n \geq 1}$  où  $S_n$  est défini par

$$S_n = X_1 + \dots + X_n.$$

On considère une marche aléatoire sur  $\mathbb{Z}^d$  telle que

$$P(X_i = e_j) = P(X_i = -e_j) = \frac{1}{2d}$$

pour tout  $i \geq 1$  et tout  $j \in \{1, \dots, d\}$ . Alors:

- La marche aléatoire est récurrente si  $d = 1$  ou  $d = 2$
- La marche aléatoire est transiente si  $d \geq 3$

Les auteurs ont réussi à expliquer le concept de la marche aléatoire et le principe de l'indice PageRank, c'est à dire la notoriété d'une page/site sur le web.

## 4. Evaluation

### 1. Exécution du Rmd

le fichier Rmd s'exécute sans erreur.

## **2. Qualité de la rédaction**

Je trouve qu'il s'agit d'un très bon travail. Les auteurs ont réussi à bien expliquer l'algorithme en utilisant un exemple simple et clair pour bien illustrer le concept.

## **3. Qualité et lisibilité du RMarkdown**

Le fichier est globalement bien écrit, lisible et aéré.

## **4. Aspect didactique du dossier**

Le document est clair. Je trouve que la description du concept de l'algorithme est compréhensible et détaillée.

## **5. Bibliographie**

Il y a une bibliographie.

# **5. Conclusion**

Je trouve que c'est généralement un très bon travail. Les auteurs ont fourni un document clair, recherché et surtout riche.

# Dossier Approche mathématiques - Émojis

## 1. Entête

Le travail évalué dans cette partie est **Dossier Approche mathématiques - Émojis** élaboré par Imen Derrouiche.

En cliquant ici vous trouverez le lien menant au GitHub:

<https://github.com/imenderrouiche/PSBX/blob/main/Approche%20mathématiques%20-%20Émojis.ipynb>

(<https://github.com/imenderrouiche/PSBX/blob/main/Approche%20mathématiques%20-%20Émojis.ipynb>).

## 2. Synthèse

La thèse étudiée traite sur la recommandation automatique et adaptative d'emojis.

Le but de ce travail est d'étudier une formule mathématiques sur la "Recommandation automatique et adaptative d'emojis", cette formule est celle de la distance de Levenshtein.

## 3. Extrait

L'autrice a expliqué la formule distance de Levenshtein qui permet de calculer le nombre d'étapes minimales nécessaires pour transformer la chaîne de caractères a en b. Le résultat de cette formule est donc un nombre entier.

La distance est calculée en répétant les étapes suivantes:

- Construire une matrice M de longueur\_mot\_2+1 lignes et longueur\_mot\_1+1 colonnes;
- Initialiser la première ligne par la matrice ligne [ 0,1,....., longueur\_mot\_2-1, longueur\_mot\_2] et la première colonne par la matrice colonne [ 0,1,....., longueur\_mot\_1-1, longueur\_mot\_1];

Ensuite la matrice M est remplie en utilisant la règle suivante  $M[i, j]$  est égale au minimum entre les éléments suivants :

- L'élément directement au-dessus et on ajoute 1 :  $M[i-1, j] + 1$ . (effacement)
- L'élément directement avant et on ajoute 1 :  $M[i, j-1] + 1$ . (insertion)
- L'élément diagonal précédent plus le coût :  $M[i-1, j-1] + \text{Cost}(i-1, j-1)$ . (substitution)

Enfin, la distance de Levenshtein entre les chaînes 1 et 2 se retrouve en  $M[\text{longueur\_mot\_1}, \text{longueur\_mot2}]$ .

## 4. Evaluation

### 1. Exécution du Rmd

le fichier Rmd s'exécute sans erreur.

### 2. Qualité de la rédaction

Selon moi, il s'agit d'un bon travail. L'autrice a bien expliqué l'algorithme en utilisant un exemple pour mieux illustrer le concept.

### 3. Qualité et lisibilité du RMarkdown

Le fichier est organisé, lisible et aéré.

### 4. Aspect didactique du dossier

Le document est facile à lire et comprendre .je trouve que la description du concept de l'algorithme est compréhensible et détaillée .

## **5. Bibliographie**

Il y a une bibliographie.

## **5. Conclusion**

Je trouve que c'est globalement un bon travail.La thèse est bien expliquée , facile à comprendre et convient à un étudiant avec un niveau moyen à faible en mathématiques.

# Dossier Statsmodels

## 1. Entête

Nous allons evaluer le document de **Regression Linéaire simple et multiple** fait par Nina Zoumanigui .

En cliquant ici vous trouverez le lien menant au GitHub: <https://github.com/Nina809/PSBX/blob/main/Regression.Rmd>  
(<https://github.com/Nina809/PSBX/blob/main/Regression.Rmd>).

## 2. Synthèse

L'auteur ici traite la regression linéaire(simple ou multiple) qui est une technique de prédiction d'une variable dépendante par rapport à une ou plusieurs variables indépendantes.

La régression se base sur des hypothèses stochastiques et structurelles.

## 3. Extrait

Le modèle de régression multiple est une généralisation du modèle de régression simple lorsque les variables explicatives sont en nombre fini. Nous supposons donc que les données collectées suivent le modèle suivant :

$$y_i = b_0 + b_1x_{i1} + b_2x_{i2} + \dots + b_px_{ip} + \epsilon_i, i = 1, \dots, n.$$

- $y_i$  : variable à prédire
- $b_0$  : point à l'origine
- $b_1 \dots b_p$  : variables dépendantes
- $\epsilon_i$  : erreur résiduelle hypothèse que sa distribution soit normale.

## 4. Evaluation

### 1. Exécution du Rmd

le fichier Rmd s'exécute sans erreur.

### 2. Qualité de la rédaction

Il s'agit d'un bon travail. L'auteur a bien expliqué le concept de la regression.

### 3. Qualité et lisibilité du RMarkdown

Le fichier est bien écrit et organisé.

### 4.Aspect didactique du dossier

Le document est facile à lire et comprendre .

### 5. Bibliographie

Il n'y a pas de bibliographie.

## 5. Conclusion

Je trouve que c'est globalement un bon travail. Le document présente une introduction sur la régression linéaire (simple et multiple). On passe en revue les hypothèses faites et on a sa forme générale.



# 1. Entête

Le document évalué est **Algèbre tropicale** de Marion DANYACH sur un papier de Dominique CASTELLA et Stephane GAUBERT intitulé “Algèbre de groupe en caractéristique 1 et distances invariantes sur un groupe fini” .

En cliquant ici vous trouverez le lien menant au GitHub: <https://github.com/MarionD436/MATHS>  
(<https://github.com/MarionD436/MATHS>).

# 2. Synthèse

L'autrice ici traite 'algèbre tropicale, qui re-définit l'algèbre que nous avons appris depuis toujours, elle est utilisée dans la Modélisation, l'analyse, l'évaluation des performances et la synthèse de lois des commandes pour des classes bien répertoriées de systèmes à 'événements discrets d'déterministes ou stochastiques . Programmation fonctionnelle. elle trouve son utilité dans SQL par exemple par sa caractéristique qui fait qu'un script idempotent ne modifie pas la base sur laquelle on l'applique.

# 3. Extrait

l'autrice a donné la définition de l'algèbre tropicale ainsi que quelques définitions utiles: tel que L'idempotence qui signifie qu'une opération a le même effet qu'on l'applique une ou plusieurs fois, et Un corps qui est un ensemble muni de deux opérations binaires.

il y a deux équations dans ce document qu'on doit retenir :

- Si  $e$  et  $f$  sont deux idempotents centraux,  $e + f$  est un idempotent si et seulement si

$$ef = e + f$$

- si  $e$  est idempotent central alors il est irréductible s'il vérifie :

$$e = fg \text{ où } f \text{ et } g \text{ sont idempotents centraux et donc } e = f \text{ et } e = g.$$

# 4. Evaluation

## 1. Exécution du Rmd

Il n'y a pas de fichier Rmd.

## 2. Qualité de la rédaction

Il s'agit d'un bon travail. L'autrice a bien expliqué le concept de l'algèbre tropicale.

## 3. Qualité et lisibilité du RMarkdown

Le fichier est bien écrit, lisible, aéré et bien organisé.

## 4. Aspect didactique du dossier

les équations fournies dans le document sont un peu compliquées et elles nécessitent plus d'explications .

## 5. Bibliographie

Il y a une bibliographie.

## 5. Conclusion

Je trouve que c'est globalement un bon travail. L'autrice a réussi à fournir une explication plus ou moins claire au terme "algèbre tropicale". Cependant, le papier de recherche n'a pas été traité entièrement, et la partie analysée est traitée d'une manière superficielle.

# 1. Entête

Nous allons évaluer le document **Arbres de Décision** de Rindra LUTZ et Nicolas ALLIX.

En cliquant ici vous trouverez le lien menant au GitHub: <https://github.com/Nicolas-all/PSB1/blob/main/Arbres-de-D%C3%A9cision.pdf> (<https://github.com/Nicolas-all/PSB1/blob/main/Arbres-de-D%C3%A9cision.pdf>).

# 2. Synthèse

Les auteurs expliquent le concept des arbres de décision , en effet Les arbres de décision sont une catégorie d'arbres utilisée dans l'exploration de données et en informatique décisionnelle. Ils emploient une représentation hiérarchique de la structure des données sous forme des séquences de décisions en vue de la prédiction d'un résultat ou d'une classe. Chaque individu (ou observation), qui doit être attribué(e) à une classe, est décrit(e) par un ensemble de variables qui sont testées dans les nœuds de l'arbre. Les tests s'effectuent dans les nœuds internes et les décisions sont prises dans les nœuds feuille (noeuds terminaux).

# 3. Extrait

Les auteurs ont donné les définition des notions suivantes :

-La pureté d'un noeud se mesure avec l'indice de Gini, plus la valeur de l'indice est proche de 0, plus le noeud est pur.

$$G_i = 1 - \sum_{k=1}^n p_{i,k}^2$$

-Le coût du noeud va permettre de mesurer la pertinence du choix de la variable de décision. Ce coût est calculé via la formule suivante :

$$J(k) = \left(\frac{m_{\text{gauche}}}{m}\right)G_{\text{gauche}} + \left(\frac{m_{\text{droite}}}{m}\right)G_{\text{droite}}$$

# 4. Evaluation

## 1. Exécution du Rmd

Le fichier Rmd s'exécute sans erreur.

## 2. Qualité de la rédaction

Il s'agit d'un bon travail en terme de rédaction,

## 3. Qualité et lisibilité du RMarkdown

Le fichier est bien écrit,lisible, aéré et bien organisé.

## 4.Aspect didactique du dossier

le sujet est ordinaire,les notions introduits sont basiques,et bien expliquées.

## 5. Bibliographie

Il n'y a pas de bibliographie.

## 5. Conclusion

Je trouve que c'est globalement un bon travail. Les auteurs ont réussi à expliquer le concept de l'arbre des décisions avec des notions basiques et bien expliquées.

## #XGboost “Auto-évaluation”

Je trouve que le document était très riche en matière de notions mathématique, j'ai essayé de bien expliquer l'aspect mathématiques derrière ce package et les différentes étapes qui mènent au calcul des poids de pondération qui représentent des clés importants de ces types d'algorithmes. Cependant, il n'y avait pas d'exemple, ce qui rend peut-être le document un peu difficile à comprendre pour les personnes qui n'ont pas un bagage riche en mathématique.