

## Abstract

Ce document présente une approche empirique et intuitive pour analyser les écarts entre nombres premiers. En normalisant les écarts par  $\log(p)$ , on observe une structure stable qui peut être discrétisée en une grammaire symbolique. Cette grammaire, associée à un modèle Markovien simple, permet de générer des écarts simulés qui reproduisent la texture locale des écarts premiers. Les simulations montrent un taux de réussite d'environ 30–41%, soit environ 4 fois mieux que le hasard dans l'intervalle 1–2 millions.

## 1 Introduction

L'idée initiale provient d'une intuition géométrique simple : les nombres premiers au-delà de 5 n'occupent que 8 positions possibles modulo 30. Cette structure a inspiré une représentation mentale sous forme de cube  $3 \times 3 \times 3$ , où seules certaines positions sont autorisées.

A partir de cette intuition, une observation empirique a été faite : les écarts entre nombres premiers, une fois normalisés, semblent suivre des motifs répétitifs et stables.

## 2 Normalisation des écarts

Pour un nombre premier  $p_n$ , on définit :

$$\Delta_n = p_{n+1} - p_n, \quad g_n = \frac{\Delta_n}{\log(p_n)}.$$

Cette normalisation rend les écarts comparables sur de grands intervalles. Les valeurs  $g_n$  se regroupent naturellement en classes.

## 3 Construction d'une grammaire

Les valeurs normalisées sont discrétisées en symboles :

- a : écart très serré
- b : écart serré
- c : écart typique
- d : écart large

Les motifs les plus fréquents observés sont :

## **Bigrams**

bb, ab, ba, bc, cb, aa, ca

## **Trigrams**

bbb, bba, bab, abb, bcb, bbc

## **Tetragrammes**

bbbb, bbab, bccb, bbba, cbba, abbb

Ces motifs sont stables sur plusieurs intervalles : 1–2000, 20k–50k, 1M–2M.

## **4 Modèle Markovien**

A partir des bigrams, on construit une matrice de transition :

$$P(s_{n+1} \mid s_n).$$

Cette matrice est stable et permet de générer des séquences symboliques réalistes. Chaque symbole est ensuite associé à une valeur typique de  $g$ , ce qui permet de reconstruire des écarts simulés.

## **5 Filtrage arithmétique**

Pour rendre la simulation plus réaliste :

### **Résidus modulo 30**

Seuls les résidus :

$$1, 7, 11, 13, 17, 19, 23, 29$$

sont autorisés.

### **Filtre anti-multiples**

Les entiers divisibles par :

$$7, 11, 13, 17, 19, 23, 29$$

sont élimines.

## 6 Résultats expérimentaux

Sur l'intervalle 1–2 millions, avec 100 valeurs simulées :

- 41% des valeurs simulées sont des nombres premiers
- densité réelle : 7.24%
- performance relative : environ  $4.1 \times$  le hasard

## 7 Comparaison avec le modèle de Granville

Le modèle de Granville prédit :

$$P_{\text{Granville}}(n) \approx \frac{1.2}{\ln(n)} \approx 7\%.$$

Modèle	Résultat
Grammaire normalisée	41%
Granville	7%
Hasard	7%

## 8 Conclusion

Cette approche ne constitue pas une théorie, mais une structure empirique stable. Elle propose une nouvelle manière de regarder les écarts premiers : non pas comme une suite chaotique, mais comme un langage symbolique avec des motifs répétitifs. Ce document vise simplement à partager cette observation avec la communauté mathématique.