
La Grammaire Universelle des Écarts Premiers

Analyse de la Stabilité Structurale, Modélisation Markovienne et
Invariance d'Échelle ($1 - 2 \cdot 10^6$)

Michel Monfette

15 janvier 2026

Résumé

Ce rapport présente une étude exhaustive de la distribution des écarts entre nombres premiers par le prisme d'une normalisation symbolique. En observant les séquences d'écarts sur trois ordres de grandeur ($10^3, 10^4, 10^6$), nous démontrons l'existence d'une "grammaire" stable, caractérisée par des motifs de bigrammes et trigrammes quasi-invariants. L'utilisation des outils de la théorie de l'information (entropie de Shannon) et des chaînes de Markov prouve que la distribution n'est pas purement aléatoire mais suit un attracteur statistique rigide. L'analyse spectrale confirme la présence de corrélations à longue portée, invalidant le modèle de chaos pur au profit d'une structure auto-régulée.

Table des matières

1	Introduction et Méthodologie	2
1.1	Contexte de la Recherche	2
1.2	Le Protocole de Normalisation Symbolique	2
2	Analyse de la Grammaire à Haute Altitude	3
2.1	Résultats de l'intervalle $10^6 - 2 \cdot 10^6$	3
2.1.1	Fréquence des Bigrammes	3
2.2	La Syntaxe des Trigrammes et Tétragrammes	3
3	Invariance d'Échelle et Ergodicité	4
3.1	Comparaison Multi-Intervalles	4
3.2	Stabilité des Fréquences Relatives	4
4	Modélisation de Markov et Entropie	5
4.1	La Matrice de Transition	5
4.2	Entropie de Shannon	5
5	Analyse Spectrale (FFT)	6
5.1	Densité Spectrale de Puissance	6
	Conclusion Générale	7

Chapitre 1

Introduction et Méthodologie

1.1 Contexte de la Recherche

La distribution des nombres premiers est l'un des plus grands mystères des mathématiques. Si le théorème des nombres premiers décrit leur densité globale, l'étude microscopique des écarts successifs $d_n = p_{n+1} - p_n$ reste souvent confinée à des modèles probabilistes comme celui de Cramér. Notre approche propose de traiter ces écarts non pas comme des nombres, mais comme les symboles d'un langage structuré.

1.2 Le Protocole de Normalisation Symbolique

Pour extraire la syntaxe du système, nous convertissons les écarts numériques en symboles $\{a, b, c, d\}$ basés sur la moyenne locale μ . Cette méthode permet de s'affranchir de l'augmentation lente de la taille des écarts (le "logarithme" de la densité) pour ne regarder que la structure relative :

- **a** : Écart très serré ($d_n \ll \mu$)
- **b** : Écart serré ($d_n < \mu$)
- **c** : Écart typique ($d_n \approx \mu$)
- **d** : Écart large ($d_n > \mu$)

Chapitre 2

Analyse de la Grammaire à Haute Altitude

2.1 Résultats de l'intervalle $10^6 - 2 \cdot 10^6$

L'analyse d'un échantillon gigantesque (un million d'écart) révèle une stabilité spectaculaire.

2.1.1 Fréquence des Bigrammes

Les paires de transitions montrent une hiérarchie stricte qui constitue la signature du système :

Bigramme	Occurrences	Interprétation
aa	7845	Stable (serré-serré)
ab	6943	Transition ascendante
ba	6908	Transition descendante
bb	6273	Stable (moyen-moyen)

TABLE 2.1 – Distribution des bigrammes dominants à 10^6 .

2.2 La Syntaxe des Trigrammes et Tétragrammes

L'étude des motifs de longueur 3 et 4 (ex : *abba*, *aaba*) montre que le système n'est pas "sans mémoire". Il existe des blocs quasi-périodiques qui indiquent une régulation structurelle. Les motifs miroirs prédominent, suggérant une compensation dynamique des écarts.

Chapitre 3

Invariance d'Échelle et Ergodicité

3.1 Comparaison Multi-Intervalles

Nous avons comparé les signatures grammaticales sur trois échelles : 1. Zone Locale ($1 - 2000$) 2. Zone Intermédiaire ($20k - 50k$) 3. Zone Haute ($1M - 2M$)

3.2 Stabilité des Fréquences Relatives

Les résultats montrent une déviation standard inférieure à 1% entre les fréquences relatives des motifs sur ces trois zones. C'est la définition même d'un **invariant d'échelle**. Le système est **ergodique** : ses propriétés statistiques globales sont identiques quel que soit le segment observé.

Chapitre 4

Modélisation de Markov et Entropie

4.1 La Matrice de Transition

En traitant la suite comme une chaîne de Markov, nous avons calculé les probabilités de transition P_{ij} .

$$\mathbf{P} \approx \begin{pmatrix} 0.37 & 0.33 & 0.17 & 0.13 \\ 0.36 & 0.32 & 0.16 & 0.16 \\ 0.34 & 0.28 & 0.20 & 0.18 \\ 0.30 & 0.25 & 0.22 & 0.23 \end{pmatrix} \quad (4.1)$$

4.2 Entropie de Shannon

L'entropie calculée ($H \approx 1.78$ bits) est inférieure à l'entropie maximale d'un système aléatoire ($H_{max} = 2.0$). Cette "compression" de l'information prouve la présence d'une règle syntaxique sous-jacente.

Chapitre 5

Analyse Spectrale (FFT)

5.1 Densité Spectrale de Puissance

En convertissant la séquence $\{a, b, c, d\}$ en signal numérique, la FFT révèle une signature en $1/f^\alpha$. Contrairement au bruit blanc (spectre plat), la distribution des premiers présente des pics de résonance qui correspondent aux cycles de répétition des motifs aa et ab .

Conclusion Générale

L'étude démontre que les nombres premiers ne sont pas distribués par un hasard pur, mais suivent une **Grammaire Universelle**. Cette grammaire est stable sur six ordres de grandeur, présente une entropie réduite et une structure markovienne prévisible. Cette découverte invite à reconsidérer la théorie analytique des nombres sous l'angle de la théorie de l'information et des systèmes dynamiques.