Université de Montréal

Distribution asymptotique du nombre de diviseurs premiers distincts inférieurs ou égaux à m

par

Roberto Persechino

Département de mathématiques et de statistique

Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures

en vue de l'obtention du grade de

Maître ès sciences (M.Sc.)
en Discipline

mai 2011

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé

Distribution asymptotique du nombre de diviseurs premiers distincts inférieurs ou égaux à m

présenté par

Roberto Persechino

a été évalué par un jury composé des personnes suivantes :

Sabin Lessard
(président-rapporteur)

Anatole Joffe
(directeur de recherche)

Andrew Granville
(membre du jury)

Mémoire accepté le:
mai 2011

SOMMAIRE

Le sujet principal de ce mémoire est l'étude de la distribution asymptotique de la fonction f_m qui compte le nombre de diviseurs premiers distincts parmi les nombres premiers $p_1, ..., p_m$.

Au premier chapitre, nous présentons les sept résultats qui seront démontrés au chapitre 4. Parmi ceux-ci figurent l'analogue du théorème d'Erdös-Kac et un résultat sur les grandes déviations.

Au second chapitre, nous définissons les espaces de probabilités qui serviront à calculer les probabilités asymptotiques des événements considérés, et éventuellement à calcluer les densités qui leur correspondent.

Le troisième chapitre est la partie centrale du mémoire. On y définit la promenade aléatoire qui, une fois normalisée, convergera vers le mouvement brownien. De là, découleront les résultats qui formeront la base des démonstrations de ceux chapitre 1.

Mots clés :
Diviseurs premiers, théorème central limite d'Erdös-Kac, grandes déviations, promenade aléatoire, mouvement brownien, convergence faible.

SUMMARY

The main topic of this masters thesis is the study of the asymptotic distribution of the fonction f_m which counts the number of distinct prime divisors among the first m prime numbers, i.e. $p_1, \dots, p_m$.

The first chapter provides the seven main results which will later on be proved in chapter 4. Among these we find the analogue of the Erdös-Kac central limit theorem and a result on large deviations.

In the following chapter, we define several probability spaces on which we will calculate asymptotic probabilities of specific events. These will become necessary for calculating their corresponding densities.

The third chapter is the main part of this masters thesis. In it, we introduce a random walk which, when suitably normalized, will converge to the Brownian motion. We will then obtain results which will form the basis of the proofs of those of chapiter 1.

Key words :

Prime divisors, Erdös-Kac central limit theorem, large deviations, random walk, Brownian motion, weak convergence.

TABLE DES MATIÈRES

LISTE DES SYMBOLES UTILISÉS

$\mathbb{N}$: L'ensemble des nombres entiers positifs, $\{1, 2, 3, \ldots\}$.

$\mathbb{R}$: L'ensemble des nombres réels.

$\mathbb{P}$: L'ensemble des nombres premiers $\{p_1, p_2, \ldots, p_n, \ldots\}$ tels que $p_i < p_{i+1}$ pour tout $i \in \mathbb{N}$.

$|A|$: La cardinalité de la partie A lorsque celle-ci contient un nombre fini d'éléments.

$[N]$: L'ensemble des entiers positifs de 1 à N, inclusivement.

A_N: L'intersection de la partie A de $\mathbb{N}$ avec $[N]$.

$\mathcal{D}(A)$: La densité arithmétique de la partie A de $\mathbb{N}$.

$\pi(m)$: La cardinalité des nombres premiers inférieurs ou égaux à m.

$\mathbb{I}_E(n)$: La fonction indicatrice de l'ensemble E, prenant la valeur 1 lorsque $n \in E$ et 0 sinon.

$a \mid b$: Le nombre a est un diviseur du nombre b, où $a, b \in \mathbb{N}$.

$\lfloor x \rfloor$: Le plus grand entier inférieur ou égal à x.

$\mathfrak{P}(E)$: La classe des parties de E.

$O(f(n))$: Si $g(n) = O(f(n))$ alors il existe $K > 0$ tel que $|g(n)| \leq Kf(n)$ pour tout $n \geq 1$.

$f(n) \sim g(n)$: Lorsque $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$.

$\mathcal{B}(p)$: Variable aléatoire de loi Bernoulli et de paramètre $0 < p < 1$.

$\mathcal{P}_\lambda$: Variable aléatoire de loi Poisson et de paramètre $\lambda > 0$.

$\mathcal{N}(0, 1)$: Variable aléatoire de loi normale, de moyenne 0 et de variance 1.

$P_n \Rightarrow P$: La suite de loi de probabilité $\{P_n\}_{n\in\mathbb{N}}$ converge faiblement vers la loi de probabilité P.

$X_n \xrightarrow{\mathcal{L}} X$: La suite de variables aléatoires $\{X_n\}_{n \in \mathbb{N}}$ converge en loi vers la variable aléatoire X.

REMERCIEMENTS

Je tiens tout d'abord à remercier Dieu le Père, Jésus, l'Esprit-Saint, la Sainte-Vierge et Saint-Michel Archange pour leur constante aide et support tout au long de cette grande et belle aventure. De même, je remercie le Père J. Réal Bleau pour ses bons conseils, sa grande disponibilité et ses prières.

Par la suite, je désire remercier mon directeur de maîtrise, M. Anatole Joffe, un professeur exceptionnel, doté d'une intuition remarquable et d'une grande patience. Ses conseils ont une valeur inestimable et sont toujours grandement appréciés. Je le remercie particulièrment de m'avoir suggéré le sujet de ce mémoire, pour sa précieuse aide et ses encouragements.

Je désire aussi remercier M. Andrew Granville pour m'avoir initié à la théorie des nombres ainsi qu'à la combinatoire additive, deux sujets parmi lesquels il excelle tant sur le plan professionnel que sur le plan pédagogique. De plus, je le remercie pour ses précieux commentaires et ses suggestions qui m'ont permis d'améliorer le contenu et la présentation de ce mémoire.

Ensuite, je tiens à remercier mes parents Bruno et Ida Persechino, mon frère Luigi, ma sœur Lina et son époux Gino Valentini, ma tante Hélène Arnone et mes cousins Anthony et Andrew Beattie pour leurs encouragements et leur soutien. La liste serait incomplète sans mon oncle Michel Arnone, qui dès mon enfance m'a initié aux mathématiques et à l'informatique : merci beaucoup.

Finalement, je remercie la famille Soulard, et en particulier Jacinthe, pour leurs nombreuses prières qui m'ont permis de finaliser ce mémoire.

INTRODUCTION

Dans son article, *Prime Numbers and Brownian Motion* [**2**], P. Billingsley présente quatre propriétés asymptotiques de la fonction $f(n)$, qui associe à tout entier positif n le nombre de ses diviseurs premiers distincts. Il aborde le sujet dans un contexte probabiliste en exploitant les propriétés des promenades aléatoires et celles du mouvement brownien. En particulier, il présente une nouvelle approche pour démontrer le *théorème central limite d'Erdös-Kac :*

$$\lim_{N\to\infty} \frac{\left|\left\{n \leq N : \dfrac{f(n) - \log\log N}{\sqrt{\log\log N}} \leq \nu\right\}\right|}{N} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\nu} e^{-\frac{t^2}{2}}\, dt,$$

lorsque $N \to \infty$; son objectif étant de développer l'intuition et d'illustrer la puissance des méthodes probabilistes. Pour la partie technique des démonstrations, il renvoie le lecteur à d'autres articles.

Inspirés par l'article de P. Billingsley, nous allons plutôt étudier la distribution asymptotique du nombre de diviseurs premiers inférieurs ou égaux à m, m étant un entier positif fixé d'avance, noté $f_m(n)$; les démonstrations sont élémentaires, car elles ne font qu'utiliser les théorèmes classiques de la théorie des probabilités.

Afin de rendre le mémoire complet, nous reproduirons la démonstration légèrement simplifiée du théorème central limite d'Erdös-Kac tel que présenté dans [**3**]. Une des raisons motivant cette décision est que les techniques utilisées dans la démonstration semblent être la clé pour démontrer un résultat en théorie des nombres sur les grandes déviations que nous conjecturons : Pour tout $\epsilon > 0$,

$$\frac{1}{\log\log N}\frac{\left|\left\{n \le N : f(n) > (1+\epsilon)\log\log N\right\}\right|}{N} \longrightarrow -(1+\epsilon)\log\bigl(1+\epsilon\bigr) + \epsilon,$$

lorsque $N \to \infty$.

Chapitre 1

LES RÉSULTATS

Le présent mémoire aura pour objet la démonstration des septs théorèmes suivants :

Définition 1.0.1. *Soit* $m \in \mathbb{N}$. *Posons*

$$f_m(n) := \sum_{i=1}^{\pi(m)} \mathbb{I}_{P_i}(n),$$

où $P_i := \left\{n \in \mathbb{N} : p_i \mid n, p_i \in \mathbb{P}\right\}$ *et* $1 \leq i \leq \pi(m)$.

Théorème 1.0.1 (théorème central limite)**.** *Soit* $u, v \in \mathbb{R}$ *avec* $u < v$. *Alors*

$$\lim_{m \to \infty} \mathcal{D}\left(n \in \mathbb{N} : u \leq \frac{f_m(n) - \log\log m}{\sqrt{\log\log m}} \leq v\right) = \frac{1}{\sqrt{2\pi}} \int_u^v e^{-\frac{1}{2}t^2}\, dt.$$

Théorème 1.0.2. *Pour tout* $x \geq 0$ *nous avons*

$$\lim_{m \to \infty} \mathcal{D}\left(n \in \mathbb{N} : \frac{1}{\sqrt{\log\log m}} \max_{1 \leq l \leq \pi(m)} \left(\sum_{i=1}^{l} \mathbb{I}_{P_i}(n) - \sum_{i=1}^{l} \frac{1}{p_i}\right) \geq x\right) = \frac{2}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{1}{2}t^2}\, dt.$$

Définition 1.0.2. *La fonction* f_m *est dite excessive lorsque* $f_m - \sum_{i=1}^{\pi(m)} \frac{1}{p_i} > 0$.

Définition 1.0.3. *Soit* $L_m : \Omega_m \to \mathbb{R}$ *telle que*[1]

$$L_m := \int_0^1 \mathbb{I}_{\mathcal{S}_\tau^{(m)} > 0}\, d\tau.$$

Alors L_m *représente la proportion du temps pour laquelle* f_m *est excessive.*

[1]La variable aléatoire $\mathcal{S}_\tau^{(m)}$ sera définie à la page 15.

Théorème 1.0.3. *La proportion du temps où f_m est excessive est inférieur à x a pour limite :*

$$\int_0^x \frac{1}{\pi} \frac{1}{\sqrt{t(1-t)}} \mathbb{I}_{(0,1)}(t)\,dt,$$

lorsque $m \to \infty$.

Théorème 1.0.4. *Soit* $y \in \mathbb{R} \setminus \{0\}$. *Alors*

$$\lim_{m\to\infty} \mathcal{D}\left(n \in \mathbb{N} : \frac{1}{\pi(m)} \left| \left\{ l \le \pi(m) : \sum_{i=1}^{l} \mathbb{I}_{P_i}(n) > \sum_{i=1}^{l} \frac{1}{p_i} \right\} \right| \le y \right) \\ = \frac{1}{2}\mathbb{I}_{(0,1)}(y) + \mathbb{I}_{[1,\infty)}(y).$$

Théorème 1.0.5. *Soit* $0 < \alpha < 1$. *Alors*

$$\lim_{m\to\infty} \mathcal{D}\big(n \in \mathbb{N} : f_m(n) - f_{m^\alpha}(n) = k\big) = e^{-\log(1/\alpha)} \frac{\big(\log(1/\alpha)\big)^k}{k!}.$$

Théorème 1.0.6 (La loi des grandes déviations)**.** *Pour tout* $\epsilon > 0$ *nous avons*

$$\lim_{m\to\infty} \frac{1}{\log\log m} \log \mathcal{D}\big(n \in \mathbb{N} : f_m(n) > (1+\epsilon)\log\log m\big) = -(1+\epsilon)\log\big(1+\epsilon\big) + \epsilon.$$

Théorème 1.0.7 (théorème central limite d'Erdös-Kac)**.**

$$\lim_{N\to\infty} \frac{\left| \left\{ n \le N : u \le \dfrac{f(n) - \log\log N}{\sqrt{\log\log N}} \le v \right\} \right|}{N} = \frac{1}{\sqrt{2\pi}} \int_u^v e^{-\frac{t^2}{2}}\,dt.$$

Chapitre 2

L'ÉTUDE DE f_m VIA UN ESPACE DE PROBABILITÉ PRODUIT

2.1. Les fonctions $f(n)$ et $f_m(n)$

Afin d'aborder l'étude de la factorisation première des nombres entiers, nous définissons la fonction $f(n)$ comme suit :

Définition 2.1.1. *Soit* $f : \mathbb{N} \to \mathbb{N}$ *la fonction telle que*

$$f(n) = |\{i \geq 1 : p_i \mid n\}|.$$

Remarquons que $f(n)$ croît très lentement, car, par exemple, il faudra attendre l'entier $n = 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 = 510510$ pour que $f(n)$ atteigne pour la première fois la valeur 7. De plus, comme il existe un nombre infini de nombres premiers, il s'ensuit que cette fonction prendra toute valeur entière positive une infinité de fois. En 1917, Hardy et Ramanujan [**5**] ont démontré que "presque tous" les nombres entiers positifs inférieurs ou égaux à n possèdent $\log\log n$ diviseurs premiers. Par exemple, si $n = 10^{70}$ alors la presque totalité des nombres inférieurs à n n'ont que $\log\log 10^{70} \approx 5$ diviseurs permiers, ce qui est remarquable.

L'étude des propriétés asympotiques de f_m se fera par l'entremise de la densité arithmétique :

Définition 2.1.2. *Soit* A *une partie de* $\mathbb{N}$*. La densité arithmétique de* A *est définie par*

$$\mathcal{D}(A) := \lim_{N\to\infty} \frac{|A_N|}{N},$$

lorsque la limite de droite existe, où $A_N = A \cap [N]$.

Considérons l'espace de probabilité $\big([N], \mathfrak{P}([N]), \mathbf{Q}_N\big)$, où $\mathbf{Q}_N$ est la loi uniforme. La densité arithmétique de A devient :

$$\mathcal{D}(A) = \lim_{N\to\infty} \mathbf{Q}_N(A_N),$$

lorsque la limite du membre de droite existe.

Le sujet des limites de probabilité étant un domaine bien connu des probabilistes, nous comprenons pourquoi la théorie des probabilités s'avérera utile pour déterminer les densités arithmétiques des ensembles considérés.

Notons que cette densité n'est pas une mesure de probabilité, car bien qu'additive elle ne vérifie pas généralement la propriété de la σ-additivé. Par exemple, lorsque $A_n = \{n\}$,

$$\mathcal{D}\Bigg(\bigcup_{n\geq 1} A_n\Bigg) = \mathcal{D}(\mathbb{N}) = 1,$$

mais

$$\sum_{n\geq 1} \mathcal{D}(A_n) = \sum_{n\geq 1} 0 = 0.$$

De plus, nous pouvons démontrer que la classe des ensembles possédant une densité arithmétique n'est pas une algèbre, et encore moins une tribu, voir **[8]** (*p. 16*).

Rappelons la définition de f_m :

$$f_m(n) = \sum_{i=1}^{\pi(m)} \mathbb{I}_{P_i}(n),$$

où $P_i := \big\{n \in \mathbb{N} : p_i \mid n, p_i \in \mathbb{P}\big\}$ et $1 \leq i \leq \pi(m)$. Dans le contexte probabiliste, les fonctions $\mathbb{I}_{P_1} \ldots, \mathbb{I}_{P_{\pi(m)}}$ deviennent une suite de variables aléatoires, et donc il est tout à fait naturel d'étudier leur loi de probabilité.

Lemme 2.1.1. *Soit* $m \in [N]$ *fixé et* $i_1, \ldots, i_k \in [\pi(m)]$. *Alors*

$$\mathbf{Q}_N(\mathbb{I}_{P_{i_1}} = 1, \ldots, \mathbb{I}_{P_{i_k}} = 1) = \frac{1}{N}\left\lfloor \frac{N}{p_{i_1}\cdots p_{i_k}} \right\rfloor, \tag{2.1.1}$$

et

$$\mathbf{Q}_N(\mathbb{I}_{P_{i_1}} = \epsilon_1, \ldots, \mathbb{I}_{P_{i_k}} = \epsilon_k) = \prod_{j=1}^{k} \left(\frac{1}{p_i}\right)^{\epsilon_j} \left(1 - \frac{1}{p_i}\right)^{1-\epsilon_j} + O(2^k/N), \quad (2.1.2)$$

où $\epsilon_1, \ldots, \epsilon_k \in \{0, 1\}$.

Démonstration. Pour démontrer l'équation (2.1.1), remarquons que :

$$\begin{aligned}
\{\mathbb{I}_{P_{i_1}} = 1, \ldots, \mathbb{I}_{P_{i_k}} = 1\} &= \{n \in [N] : p_{i_1} \mid n, \ldots, p_{i_k} \mid n\} \\
&= \{n \in [N] : p_{i_1} \cdots p_{i_k} \mid n\} \\
&= \left\{1 \cdot (p_{i_1} \cdots p_{i_k}), \ldots, \left\lfloor \frac{N}{p_{i_1} \cdots p_{i_k}} \right\rfloor \cdot (p_{i_1} \cdots p_{i_k})\right\},
\end{aligned}$$

d'où

$$\mathbf{Q}_N(\mathbb{I}_{P_{i_1}} = 1, \ldots, \mathbb{I}_{P_{i_k}} = 1) = \frac{1}{N} \left\lfloor \frac{N}{p_{i_1} \cdots p_{i_k}} \right\rfloor,$$

ou encore

$$\mathbf{Q}_N(\mathbb{I}_{P_{i_1}} = 1, \ldots, \mathbb{I}_{P_{i_k}} = 1) = \frac{1}{p_{i_1} \cdots p_{i_k}} + O(1/N).$$

Soit $\epsilon_1, \ldots, \epsilon_k \in \{0, 1\}$ et $A_j = \{\mathbb{I}_{P_{i_j}} = 1\}$ pour $j \in [k]$. De plus, posons $I_1 = \{j \in [k] : \epsilon_j = 1\}$ et $I_2 = \{j \in [k] : \epsilon_j = 0\}$, on aura

$$\begin{aligned}
\mathbf{Q}_N(\mathbb{I}_{P_{i_1}} = \epsilon_1, \ldots, \mathbb{I}_{P_{i_k}} = \epsilon_k) &= \mathbf{Q}_N\left(\bigcap_{j \in I_1} A_j \cap \bigcap_{l \in I_2} A_l^c\right) \\
&= \mathbb{E} \prod_{j \in I_1} \mathbb{I}_{A_j} \prod_{l \in I_2} \mathbb{I}_{A_l^c} \\
&= \mathbb{E} \prod_{j \in I_1} \mathbb{I}_{A_j} \prod_{l \in I_2} (1 - \mathbb{I}_{A_l}) \\
&= \sum_{b=0}^{|I_2|} (-1)^b \sum_{1 \leq l_1 < \ldots < l_a \leq b} \mathbb{E}\left(\prod_{j \in I_1} \mathbb{I}_{A_j}\right) \mathbb{I}_{A_{l_1}} \cdots \mathbb{I}_{A_{l_a}},
\end{aligned}$$

où $l_1, \ldots, l_a \in I_2$. Sachant que

$$\begin{aligned}
\mathbb{E}\left(\prod_{j \in I_1} \mathbb{I}_{A_j}\right) \mathbb{I}_{A_{l_1}} \cdots \mathbb{I}_{A_{l_a}} &= \mathbf{Q}_N\left(\bigcap_{j \in I_1} \mathbb{I}_{A_j} \cap \mathbb{I}_{A_{l_1}} \cap \cdots \cap \mathbb{I}_{A_{l_a}}\right) \\
&= \left(\prod_{j \in I_1} \frac{1}{p_j}\right) \frac{1}{p_{l_1} \cdots p_{l_a}} + O(1/N),
\end{aligned}$$

alors

$$\begin{aligned}\mathbf{Q}_N(\mathbb{I}_{P_{i_1}} = \epsilon_1, \dots, \mathbb{I}_{P_{i_k}} = \epsilon_k) &= \left(\prod_{j\in I_1} \frac{1}{p_j}\right) \sum_{b=0}^{|I_2|} (-1)^b \sum_{1\le l_1<\dots<l_a\le b} \frac{1}{p_{l_1}\cdots p_{l_a}} \\ &\quad + O(2^{|I_2|}/N) \\ &= \left(\prod_{j\in I_1} \frac{1}{p_j}\right) \prod_{l\in I_2}\left(1-\frac{1}{p_l}\right) + O(2^k/N),\end{aligned}$$

ce qui démontre l'équation (2.1.2). □

Par le lemme 2.1.1, nous obtenons une classe de parties de $\mathbb{N}$ dont chaque élément possède une densité. Cette classe servira à établir les résultats du chapitre 1.

Définition 2.1.3. *Soit* $m \in \mathbb{N}$. *Pour tout* $k \le \pi(m)$, *considérons toutes les parties de* $\mathbb{N}$ *qui sont des multiples de* $p_{i_1}\cdots p_{i_k}$, *où* $2 \le p_{i_1} < \cdots < p_{i_k} \le p_{\pi(m)}$. *Soit* $\mathcal{T}_m$ *la tribu engendrée par ces ensembles.*

Lemme 2.1.2. *Toute partie* A *appartenant à la tribu* $\mathcal{T}_m$ *a une densité.*

Démonstration. Soit $A = \{n \in \mathbb{N} : \mathbb{I}_{P_1}(n) = \epsilon_1, \dots, \mathbb{I}_{P_{\pi(m)}}(n) = \epsilon_{\pi(m)}\}$, où $\epsilon_1, \dots, \epsilon_{\pi(m)} \in \{0,1\}$. Par le lemme 2.1.1, $\mathcal{D}(A)$ existe. Le résultat découle maintenant du fait que la tribu engendrée par les parties disjointes de la forme A est précisément la tribu $\mathcal{T}_m$, et que cette dernière contient un nombre fini d'éléments. □

2.2. L'espace de probabilité produit $(\Omega, \mathcal{F}, \mathbf{P})$

Nous allons introduire un espace de probabilité produit sur lequel nous définirons deux mesures de probabilités, $\mathbf{P}^{(m)}$ qui sera associée à une promenade aléatoire définie sur cet espace, et $\mathbf{Q}_m^{(N)}$, la probabilité image de $\mathbf{Q}_N$, qui intéresse ceux qui travaillent en théorie des nombres. Nous nous intéresserons à comparer ces deux mesures de probabilité.

Pour $i \in \mathbb{N}$, considérons l'espace de probabilité $\Lambda_i := (E_i, \mathfrak{P}(E_i), \mathbf{P}_i)$ tel que $E_i = \{0,1\}$, $\mathbf{P}_i(1) = \frac{1}{p_i}$ et $\mathbf{P}_i(0) = 1 - \frac{1}{p_i}$, et définissons la suite de fonctions

$\{X_i\}_{i\in\mathbb{N}}$ telle que

$$X_i : \Omega \to \{0,1\} \qquad \text{et} \qquad X_i(\omega) := \omega_i, \tag{2.2.1}$$

où $\Omega = \prod_{j=1}^{\infty} E_i$ et $\omega = (\omega_1, \omega_2, \ldots)$. De plus, soit $\mathcal{F} = \sigma(X_i : i \geq 1)$, la plus petite tribu sur laquelle la suite de fonctions $\{X_i\}_{i\in\mathbb{N}}$ est mesurable. Par le théorème A.0.1 (voir Annexe A), il existe une unique mesure de probabilité $\mathbf{P}$ sur $(\Omega, \mathcal{F})$ telle que pour tout $k \in \mathbb{N}$ et pour tout $x_1, \ldots, x_k \in \{0,1\}$,

$$\mathbf{P}(X_{i_1} = x_1, \ldots, X_{i_k} = x_k) = \prod_{j=1}^{k} \left(\frac{1}{p_{i_j}}\right)^{x_j} \left(1 - \frac{1}{p_{i_j}}\right)^{1-x_j},$$

c'est-à-dire X_i est de loi $\mathbf{P}_i$ et la suite de fonctions $\{X_i\}_{i\in\mathbb{N}}$ est indépendante.

Soit $\Omega_m := \prod_{j=1}^{\pi(m)} E_i$ et $\mathcal{F}_m := \mathfrak{P}(\Omega_m)$ où $m \in \mathbb{N}$. Introduisons deux mesures de probabilité sur l'espace $(\Omega_m, \mathcal{F}_m)$:

Définition 2.2.1. *Soit* $p_m : \Omega \to \Omega_m$ *et* $d_m : [N] \to \Omega_m$ *deux applications telles que* $p_m(\omega_1, \omega_2, \ldots) := (\omega_1, \omega_2, \ldots, \omega_{\pi(m)})$ *et*

$$d_m(n) := \left(\mathbb{I}_{P_1}(n), \ldots, \mathbb{I}_{P_{\pi(m)}}(n)\right). \tag{2.2.2}$$

Proposition 2.2.1. *Soit*

$$\mathbf{P}^{(m)} := \mathbf{P} \circ p_m^{-1} \qquad \textit{et} \qquad \mathbf{Q}_m^{(N)} := \mathbf{Q}_N \circ d_m^{-1}. \tag{2.2.3}$$

Alors $(\Omega_m, \mathcal{F}_m, \mathbf{P}^{(m)})$ *et* $(\Omega_m, \mathcal{F}_m, \mathbf{Q}_m^{(N)})$ *sont des espaces de probabilité.*

En identifiant les variables aléatoires $X_1, X_2, \ldots$ définies en (2.2.1) comme suit :

$$X_i : \Omega_m \to \{0,1\} \qquad \text{et} \qquad X_i(\omega_1, \omega_2, \ldots, \omega_{\pi(m)}) := \omega_i,$$

nous obtenons que $X_1, X_2, \ldots$ sont des variables alétoires sur $(\Omega_m, \mathcal{F}_m)$ et

$$\mathbf{P}^{(m)}(X_{i_1} = x_1, \ldots, X_{i_k} = x_k) = \prod_{j=1}^{k} \left(\frac{1}{p_{i_j}}\right)^{x_j} \left(1 - \frac{1}{p_{i_j}}\right)^{1-x_j}, \tag{2.2.4}$$

lorsque $i_1, \ldots, i_k \leq \pi(m)$.

Notons que la mesure de probabilité $\mathbf{P}^{(m)}$ est celle associée à la promenade aléatoire $\{\sum_{i=1}^{l} X_i\}_{l\in[\pi(m)]}$ sur l'espace produit, tandis que $\mathbf{Q}_m^{(N)}$ représentera la mesure de probabilité de la *théorie des nombres*.

Soulignons que le résultat qui suit sera la clé pour calculer les densités impliquant f_m.

Corollaire 2.2.1. *Soit* $m \in [N]$ *un nombre fixé et* $i_1, \ldots, i_k \in [\pi(m)]$. *Alors*

$$\mathbf{Q}_m^{(N)}\big(X_{i_1} = \cdots = X_{i_k} = 1\big) = \frac{1}{N}\left\lfloor \frac{N}{p_{i_1} \cdots p_{i_k}} \right\rfloor, \tag{2.2.5}$$

et

$$\lim_{N\to\infty} \mathbf{Q}_m^{(N)}(X_{i_1} = \epsilon_1, \ldots, X_{i_k} = \epsilon_k) = \mathbf{P}^{(m)}(X_{i_1} = \epsilon_1, \ldots, X_{i_k} = \epsilon_k),$$

pour tout $\epsilon_{i_1}, \ldots, \epsilon_{i_k} \in \{0, 1\}$.

DÉMONSTRATION. Les deux résultats sont des conséquences du lemme 2.1.1.

□

Théorème 2.2.1. *Si* $A \in \mathcal{T}_m$ *alors* $\mathcal{D}(A) = \mathbf{P}^{(m)}(B)$, *où* $B = d_m(A_N) \in \mathcal{F}_m$.

DÉMONSTRATION. Par le lemme 2.1.2, il suffit de montrer que le résultat est vérifié pour $A = \{n \in \mathbb{N} : \mathbb{I}_{P_1}(n) = \epsilon_1, \ldots, \mathbb{I}_{P_{\pi(m)}}(n) = \epsilon_{\pi(m)}\}$, où $\epsilon_1, \ldots, \epsilon_{\pi(m)} \in \{0, 1\}$. Par le corollaire 2.2.1,

$$\mathcal{D}(A) = \lim_{N\to\infty} \mathbf{Q}_N(A_N) = \lim_{N\to\infty} \mathbf{Q}_N\big(d_m^{-1}(B)\big) = \lim_{N\to\infty} \mathbf{Q}_m^{(N)}(B) = \mathbf{P}^{(m)}(B),$$

où $B = \{X_1 = \epsilon_1, \ldots, X_{\pi(m)} = \epsilon_{\pi(m)}\}$. □

C'est grâce au théorème 2.2.1 que nous allons étudier le comportement de la fonction f_m en utilisant les propriétés de la promenade aléatoire $\{\sum_{i=1}^{l} X_i\}_{l\in[\pi(m)]}$. L'étude de cette promenade aléatoire sera l'objet du chapitre 3.

Chapitre 3

LOIS LIMITES DE LA PROMENADE ALÉATOIRE

3.1. Le théorème central limite

Notons que $\mathbb{E}X_i = \frac{1}{p_i}$ et $\mathbb{V}\mathrm{ar}X_i = \frac{1}{p_i}\left(1 - \frac{1}{p_i}\right)$. Considérons la suite de variables aléatoires $Y_1, \ldots, Y_m$ telle que $Y_i := X_i - \mathbb{E}X_i$ et leur somme partielle $S_l^{(m)} = \sum_{i=1}^{l} Y_i$, lorsque $1 \leq l \leq \pi(m)$. Alors

$$\mathbb{E}S_{\pi(m)}^{(m)} = 0 \qquad \text{et} \qquad \sigma_{\pi(m)}^2 = \sum_{i=1}^{\pi(m)} \frac{1}{p_i} - \sum_{i=1}^{\pi(m)} \frac{1}{p_i^2}.$$

Par [7], (*exercice 1.10.9, p. 50*),

$$\sum_{i=1}^{\pi(m)} \frac{1}{p_i} = \log\log m + C + O(1/\log m), \tag{3.1.1}$$

d'où $\sigma_{\pi(m)}^2 \sim \log\log m$. Les variables aléatoires $Y_1, \ldots, Y_m$ sont bornées et satisfont aux théorème central limite :

$$\frac{S_{\pi(m)}^{(m)}}{\sigma_{\pi(m)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1), \quad \text{lorsque } m \to \infty,$$

c'est-à-dire

$$\lim_{m\to\infty} \mathbf{P}^{(m)}\left(u \leq \frac{\sum_{i=1}^{\pi(m)} X_i - \mu_m}{\sqrt{\mu_m - \theta_m}} \leq v\right) = \frac{1}{\sqrt{2\pi}} \int_u^v e^{-\frac{1}{2}t^2}\, dt,$$

pour tout u, v, où $\mu_m = \sum_{p\leq m} \frac{1}{p}$ et où $\theta_m = \sum_{p\leq m} \frac{1}{p^2}$.

Nous obtenons le résultat suivant :

Lemme 3.1.1. *Soit* $u, v \in \mathbb{R}$ *avec* $u < v$. *Alors*

$$\lim_{m\to\infty} \mathbf{P}^{(m)}\left(u \leq \frac{\sum_{i=1}^{\pi(m)} X_i - \log\log m}{\sqrt{\log\log m}} \leq v\right) = \frac{1}{\sqrt{2\pi}} \int_u^v e^{-\frac{1}{2}t^2}\, dt,$$

lorsque $m \to \infty$.

Démonstration. Posons

$$B_m = \left\{ u \leq \frac{\sum_{i=1}^{\pi(m)} X_i - \log\log m}{\sqrt{\log\log m}} \leq v \right\}.$$

Alors

$$B_m = \left\{ \frac{u}{\alpha_m} - \frac{\gamma_m}{\alpha_m\sqrt{\log\log m}} \leq \frac{\sum_{i=1}^{\pi(m)} X_i - \mu_m}{\sqrt{\mu_m - \theta_m}} \leq \frac{v}{\alpha_m} - \frac{\gamma_m}{\alpha_m\sqrt{\log\log m}} \right\},$$

où

$$\gamma_m = C + O(1/\log m) \qquad \text{et} \qquad \alpha_m = \sqrt{\frac{\log\log m + \gamma_m - \theta_m}{\log\log m}}.$$

Étant donné que

$$\alpha_m \to 1 \qquad \text{et} \qquad \frac{\gamma_m}{\alpha_m\sqrt{\log\log m}} \to 0,$$

lorsque $m \to \infty$, pour $\epsilon > 0$ et pour m assez grand nous aurons,

$$\left\{ u + \epsilon \leq \frac{\sum_{i=1}^{\pi(m)} X_i - \mu_m}{\sqrt{\mu_m - \theta_m}} \leq v - \epsilon \right\} \subset B_m$$

et

$$B_m \subset \left\{ u - \epsilon \leq \frac{\sum_{i=1}^{\pi(m)} X_i - \mu_m}{\sqrt{\mu_m - \theta_m}} \leq v + \epsilon \right\},$$

d'où

$$\frac{1}{\sqrt{2\pi}} \int_{u+\epsilon}^{v-\epsilon} e^{-\frac{1}{2}t^2}\, dt \leq \varliminf_{m\to\infty} \mathbf{P}^{(m)}(B_m) \leq \varlimsup_{m\to\infty} \mathbf{P}^{(m)}(B_m) \leq \frac{1}{\sqrt{2\pi}} \int_{u-\epsilon}^{v+\epsilon} e^{-\frac{1}{2}t^2}\, dt.$$

Nous obtenons le résultat en faisant tendre ϵ vers 0. □

3.2. La convergence faible vers le mouvement brownien

Pour les détails sur le mouvement brownien et les notions de *convergence faible* sur le mouvement brownien nous renvoyons à **[1]**. Dans ce qui suit, nous travaillerons sur l'espace métrique des fonctions continues sur $[0, T]$, noté $C[0, T]$, et sa tribu borélienne.

Définition 3.2.1. *Soit* $P, P_1, P_2, \ldots$ *des mesures de probabilité sur l'espace* $C[0,T]$. *Nous dirons que la suite* $\{P_m\}_{m\in\mathbb{N}}$ *convergera faiblement vers* P, *et nous écrirons* $P_m \Rightarrow P$, *si pour toute fonction bornée, continue à valeur réelle,* f,

$$\int_{C[0,T]} f\, dP_m \to \int_{C[0,T]} f\, dP,$$

lorsque $m \to \infty$.

Définition 3.2.2. *Une suite de variables aléatoires* $\{X_m\}_{m\in\mathbb{N}}$ *convergera en loi vers la variable aléatoire* X, *et nous écrirons* $X_n \xrightarrow{\mathcal{L}} X$, *si la loi* P_n *de* X_n *converge faiblement vers la loi* P *de* X.

Considérons la promenade aléatoire $\{S_l^{(m)}\}_{l\in[\pi(m)]}$. Remarquons qu'elle est de moyenne nulle et de variance d'ordre de grandeur $\sum_{i=1}^{l} \frac{1}{p_i}$ pour tout $l \in [\pi(m)]$. Il serait souhaitable que $\mathbb{V}\mathrm{ar} S_l^{(m)}$ soit approximativement l de sorte que nous puissions lui appliquer les normalisations analogues à celle de la promenade aléatoire symétrique, et ainsi obtenir la convergence faible vers le mouvement brownien. Procédons comme dans **[2]** : soit l'ensemble

$$E_m = \left\{ t_l \in \mathbb{R} : t_l = \sum_{i=1}^{l} \frac{1}{p_i}, l \in [\pi(m)] \right\},$$

et posons

$$\mathfrak{S}_{t_l}^{(m)} := S_l^{(m)}, \tag{3.2.1}$$

Cette promenade aléatoire $\{\mathfrak{S}_l^{(m)}\}_{l\in[\pi(m)]}$ prendra la valeur $S_l^{(m)}$ au temps t_l. De plus :

$$\mathbb{E}\mathfrak{S}_{t_l}^{(m)} = 0 \qquad \text{et} \qquad \mathbb{V}\mathrm{ar}\mathfrak{S}_{t_l}^{(m)} \sim t_l.$$

Afin d'étendre $\mathfrak{S}_t^{(m)}$ à toute valeur $t \in [0, t_{\pi(m)}]$, nous procédons par interpolation linéaire :

$$\mathfrak{S}_t^{(m)} := p_{l+1}(t - t_l)Y_{l+1} + \mathfrak{S}_{t_l}^{(m)} \quad \text{lorsque } t \in [t_l, t_{l+1}). \tag{3.2.2}$$

Nous pouvons maintenant normaliser cette variable aléatoire en posant

$$\mathcal{S}_\tau^{(m)} := \frac{1}{\sqrt{t_{\pi(m)}}} \mathfrak{S}_{\tau t_{\pi(m)}}^{(m)}, \quad \tau \in [0,1], \tag{3.2.3}$$

de sorte que $\mathbb{E}\mathcal{S}_{\tau}^{(m)} = 0$ et $\mathbb{V}\mathrm{ar}\mathcal{S}_{\tau_l}^{(m)} \sim \tau_l$, pour $l \in [\pi(m)]$, où $\tau_l = \frac{t_l}{t_{\pi(m)}}$. Notons que $\mathbb{V}\mathrm{ar}\mathcal{S}_{\tau_{\tau_m}}^{(m)} \sim 1$.

Ceci nous amène au lemme suivant :

Lemme 3.2.1. *Pour tout* $\tau \in [0,1]$,

$$\mathcal{S}_{\tau}^{(m)} \xrightarrow{\mathcal{L}} B_{\tau}, \quad 0 \le \tau \le 1,$$

lorsque $m \to \infty$, *où* $\left\{B_{\tau}\right\}_{0 \le \tau \le 1}$ *est le mouvement brownien.*

DÉMONSTRATION. On applique le théorème A.0.2 (voir Annexe A) avec $\xi_{m,i} := \frac{X_i - \frac{1}{p_i}}{\sqrt{t_{\pi(m)}}}$ pour $1 \le i \le \pi(m)$. En effet, par l'inégalité de Chebyshev on a

$$\mathbf{P}^{(m)}\left(\left|\frac{X_i - \frac{1}{p_i}}{\sqrt{t_{\pi(m)}}}\right| > \epsilon\right) \le \frac{\mathbb{V}\mathrm{ar}X_i}{t_{\pi(m)}\epsilon^2} \sim \frac{\frac{1}{p_i}\left(1 - \frac{1}{p_i}\right)}{\epsilon^2 \log\log m} \to 0,$$

lorsque $m \to \infty$. De plus, la condition de Lindeberg est satisfaite car les variables aléatoires $\xi_{m,i}$ sont bornées et $t_{\pi(m)} \to \infty$ lorsque $m \to \infty$. □

3.3. LE PRINCIPE D'INVARIANCE DE DONSKER

Soit $\left(C([0,1]), \rho\right)$ l'espace des fonctions continues sur $[0,1]$ muni de la métrique de la convergence uniforme, c'est-à-dire

$$\rho(f,g) = \max_{x \in [0,1]} |f(x) - g(x)|.$$

Le théorème A.0.3 (voir Annexe A) appliqué à $\{\mathcal{S}_{\tau}^{(m)}\}_{m \ge 1}$ affirme que

$$f(\mathcal{S}_{\tau}^{(m)}) \xrightarrow{\mathcal{L}} f(B_{\tau}), \quad 0 \le \tau \le 1,$$

lorsque $m \to \infty$, dès que la fonctionnelle $f : C([0,1]) \to \mathbb{R}$ est continue dans cette topologie, ou lorsqu'elle est continue sauf peut-être sur un ensemble de mesure de Wiener nulle.

Considérons la fonctionnelle $f_1(g) = \max_{0\leq t\leq 1} g(t)$. Elle est continue : Il existe $t_1, t_2 \in [0,1]$ tels que $f_1(g) = g(t_1)$ et $f_1(h) = g(t_2)$, d'où

$$\begin{aligned}|f_1(g) - f_1(h)| &= |g(t_1) - h(t_2)| \\ &\leq \max\big\{|g(t_1) - h(t_1)|, |g(t_2) - h(t_2)|\big\} \\ &\leq \max_{0\leq t\leq 1} |g(t) - h(t)|,\end{aligned}$$

car soit $|g(t_1) - h(t_2)| = g(t_1) - h(t_2) \leq g(t_1) - h(t_1) = |g(t_1) - h(t_1)|$ soit $|g(t_1) - h(t_2)| = h(t_2) - g(t_1) \leq h(t_2) - g(t_2) = |g(t_2) - h(t_2)|$.

De plus, il est connu que la fonctionnelle $f_2(g) = \int_0^1 \mathbb{I}_{g(t)>0}\, dt$ est continue sauf peut-être sur un ensemble de mesure de Wiener nulle, voir **[1]** (*pp. 230–232*). Ainsi, nous obtenons le lemme suivant :

Lemme 3.3.1.

$$\max_{0\leq\tau\leq 1} \mathcal{S}_\tau^{(m)} \xrightarrow{\mathcal{L}} \max_{0\leq\tau\leq 1} B_\tau,$$

et

$$\int_0^1 \mathbb{I}_{\mathcal{S}_\tau^{(m)}>0}\, d\tau \xrightarrow{\mathcal{L}} \int_0^1 \mathbb{I}_{B_\tau>0}\, d\tau,$$

lorsque $m \to \infty$.

3.4. Loi asymptotique de la proportion des sauts où la promenade aléatoire est positive

Nous désirons étudier la proportion des sauts de la promenade aléatoire normalisée $\{\mathcal{S}_{\tau_l}^{(m)}\}_{l\in\pi(m)}$ où elle est positive, et montrer que la suite de ces proportions converge faiblement vers une variable de loi $\mathcal{B}(1/2)$.

Définition 3.4.1. *Soit la variable aléatoire* $V_m : \Omega \to \mathbb{R}$ *telle que*

$$V_m := \frac{1}{\pi(m)} \sum_{l=1}^{\pi(m)} \mathbb{I}_{\mathcal{S}_{\tau_l}^{(m)}>0},$$

qui représente la proportion des sauts où la promenade aléatoire normalisée est positive.

Nous calculerons sa loi asymptotique. Commençons par un lemme :

Lemme 3.4.1. *Soit* $0 < \delta < 1$. *La proportion des sauts dans l'intervalle* $[1-\delta, 1]$ *converge vers 1 lorsque* $m \to \infty$.

DÉMONSTRATION. Nous montrerons que la proportion des sauts entre 0 et $1-\delta$ converge vers 0 lorsque $m \to \infty$.

Entre 0 et $1 = \tau_m$ il y a autant de sauts que de nombres premiers inférieurs ou égaux à m, c'est-à-dire

$$\pi(m) \sim \frac{m}{\log m}.$$

D'autre part, soit $\tau_l \leq 1-\delta < \tau_{l+1}$. Alors, le nombre de sauts entre 0 et $1-\delta$ correspond exactement au nombre de sauts entre 0 et τ_l, vu qu'il n'y a aucun saut entre τ_l et τ_{l+1}. Or, il y autant de sauts entre 0 et τ_l qu'il y a de nombres premiers entre 1 et p_l, soit l. Par (3.1.1), nous avons

$$\frac{\log\log p_l}{\log\log m} \leq 1-\delta,$$

ou encore

$$p_l \leq e^{(\log m)^{(1-\delta)}}.$$

Comme

$$l \leq \pi(e^{(\log m)^{(1-\delta)}}) \sim \frac{e^{(\log m)^{(1-\delta)}}}{(\log m)^{(1-\delta)}},$$

alors

$$\frac{l}{\pi(m)} \leq \frac{\frac{e^{(\log m)^{(1-\delta)}}}{(\log m)^{(1-\delta)}}}{\frac{m}{\log m}} = (\log m)^{\delta} e^{(\log m)^{(1-\delta)} - \log m}.$$

Sachant que

$$(\log m)^{(1-\delta)} - \log m = \log m \left(\frac{1}{(\log m)^{\delta}} - 1 \right),$$

et que pour m assez grand nous avons

$$\frac{1}{(\log m)^{\delta}} - 1 \leq -\frac{1}{2},$$

alors

$$\begin{aligned} \frac{l}{\pi(m)} &\leq (\log m)^{\delta} e^{-\frac{1}{2}\log m} \\ &= \frac{(\log m)^{\delta}}{m^{1/2}} \to 0, \end{aligned}$$

lorsque $m \to \infty$. $\square$

Ceci nous amène au résultat suivant :

Lemme 3.4.2.

$$V_m \xrightarrow{\mathcal{L}} \mathcal{B}(1/2),$$

lorsque $m \to \infty$.

DÉMONSTRATION. Par le lemme 3.3.1, nous avons que

$$\max_{1-\delta\leq\tau\leq 1} \mathcal{S}_\tau^{(m)} \xrightarrow{\mathcal{L}} \max_{1-\delta\leq\tau\leq 1} B_\tau.$$

Calculons $\mathbf{P}_W\left(\max_{1-\delta\leq\tau\leq 1} B_\tau < 0\right)$, où $\mathbf{P}_W$ est la mesure de probabilité de Wiener :

$$\mathbf{P}_W\Big(\max_{1-\delta\leq\tau\leq 1} B_\tau > 0\Big) = \frac{1}{\sqrt{2\pi(1-\delta)}}\int_{-\infty}^{\infty} e^{-\frac{z^2}{2\sqrt{1-\delta}}}\,\mathbf{P}_W\Big(\max_{1-\delta\leq\tau\leq 1} B_\tau > 0 \Big| B_{1-\delta} = z\Big)\,dz.$$

Si $z > 0$, alors

$$\mathbf{P}_W\Big(\max_{1-\delta\leq\tau\leq 1} B_\tau > 0 \Big| B_{1-\delta} = z\Big) = 1,$$

sinon $z < 0$, et par **[3]**, (*pp. 508–513*),

$$\mathbf{P}_W\Big(\max_{1-\delta\leq\tau\leq 1} B_\tau > 0 \Big| B_{1-\delta} = z\Big) = \frac{2}{\sqrt{2\pi\delta}}\int_{-z}^{\infty} e^{-\frac{x^2}{2\sqrt{\delta}}}\,dx.$$

Étant donné que $\mathbf{P}_W(B_{1-\delta} = 0) = 0$, il est inutile de traiter le cas $z = 0$. Alors

$$\begin{aligned}
\mathbf{P}_W\Big(\max_{1-\delta\leq\tau\leq 1} B_\tau > 0\Big) &= \frac{1}{\sqrt{2\pi(1-\delta)}}\int_{-\infty}^{0} e^{-\frac{z^2}{2\sqrt{1-\delta}}}\left(\frac{2}{\sqrt{2\pi\delta}}\int_{-z}^{\infty} e^{-\frac{x^2}{2\sqrt{\delta}}}\,dx\right)dz \\
&\quad + \frac{1}{\sqrt{2\pi(1-\delta)}}\int_{-\infty}^{0} e^{-\frac{z^2}{2\sqrt{1-\delta}}}\,dz \\
&= \frac{1}{2} + \frac{1}{\sqrt{2\pi(1-\delta)}}\int_{-\infty}^{0} e^{-\frac{z^2}{2\sqrt{1-\delta}}}\left(\frac{2}{\sqrt{2\pi\delta}}\int_{-z}^{\infty} e^{-\frac{x^2}{2\sqrt{\delta}}}\,dx\right)dz,
\end{aligned}$$

ce qui implique que

$$\mathbf{P}_W\Big(\max_{1-\delta\leq\tau\leq 1} B_\tau < 0\Big) = \frac{1}{2} - \frac{1}{\sqrt{2\pi(1-\delta)}}\int_{-\infty}^{0} e^{-\frac{z^2}{2\sqrt{1-\delta}}}\left(\frac{2}{\sqrt{2\pi\delta}}\int_{-z}^{\infty} e^{-\frac{x^2}{2\sqrt{\delta}}}\,dx\right)dz.$$

Lorsque $z < 0$, nous avons

$$\lim_{\delta\to 0} \frac{2}{\sqrt{2\pi\delta}} \int_{-z}^{\infty} e^{-\frac{x^2}{2\sqrt{\delta}}}\, dx = 0.$$

Si $\varepsilon > 0$ alors il existe γ_ϵ tel que $0 \le \delta < \gamma_\epsilon$, d'où

$$0 < \mathbf{P}_W\Big(\max_{1-\delta\le\tau\le 1} B_\tau < 0\Big) - \frac{1}{2} < \epsilon.$$

Pour un tel $\delta > 0$, nous aurons pour tout m suffisamment grand

$$\left|\mathbf{P}^{(m)}\Big(\max_{1-\delta\le\tau\le 1} \mathcal{S}_\tau^{(m)} < 0\Big) - \frac{1}{2}\right| < 2\epsilon. \tag{3.4.1}$$

Maintenant, pour $y \in (0,1)$:

$$\begin{aligned}
&\mathbf{P}^{(m)}\big(V_m \le y\big) \\
&\qquad = \mathbf{P}^{(m)}\left(V_m \le y \,\middle|\, \max_{1-\delta\le\tau\le 1} \mathcal{S}_\tau^{(m)} < 0\right)\mathbf{P}^{(m)}\Big(\max_{1-\delta\le\tau\le 1} \mathcal{S}_\tau^{(m)} < 0\Big).
\end{aligned} \tag{3.4.2}$$

Lorsque $\max_{1-\delta\le\tau_l\le 1} \mathcal{S}_{\tau_l}^{(m)}(\omega) < 0$, nous obtenons

$$\begin{aligned}
V_m(\omega) &= \frac{1}{\pi(m)} \sum_{l=1}^{\nu_m} \mathbb{I}_{\mathcal{S}_{\tau_l}^{(m)}>0}(\omega) + \frac{1}{\pi(m)} \sum_{l=\nu_m+1}^{\pi(m)} \mathbb{I}_{\mathcal{S}_{\tau_l}^{(m)}>0}(\omega) \\
&= \frac{1}{\pi(m)} \sum_{l=1}^{\nu_m} \mathbb{I}_{\mathcal{S}_{\tau_l}^{(m)}>0}(\omega) + \frac{1}{\pi(m)} \sum_{l=\nu_m+1}^{\pi(m)} 0 \\
&= \frac{1}{\pi(m)} \sum_{l=1}^{\nu_m} \mathbb{I}_{\mathcal{S}_{\tau_l}^{(m)}>0}(\omega) \\
&\le \frac{\nu_m}{\pi(m)},
\end{aligned}$$

où $\nu_m = \frac{e^{(\log m)^{(1-\delta)}}}{(\log m)^{(1-\delta)}}$. Par le lemme 3.4.1, pour tout m suffisamment grand nous aurons $V_m(\omega) \le y$, ou encore

$$\mathbf{P}^{(m)}\left(V_m \le y \,\middle|\, \max_{1-\delta\le\tau\le 1} \mathcal{S}_\tau^{(m)} < 0\right) = 1.$$

En utilisant (3.4.1) et (3.4.2), il s'ensuit que

$$\lim_{m\to\infty} \mathbf{P}^{(m)}\big(V_m \le y\big) = \lim_{m\to\infty} \mathbf{P}^{(m)}\Big(\max_{1-\delta\le\tau\le 1} \mathcal{S}_\tau^{(m)} < 0\Big) = \frac{1}{2}.$$

Sachant que $0 \le V_m \le 1$, nous obtenons le résultat. $\square$

3.5. LES GRANDES DÉVIATIONS

Par la loi forte des grands nombres, nous savons que $\frac{1}{\log\log m}\sum_{i=1}^{\pi(m)} X_i \to 1$ presque sûrement lorsque $m \to \infty$. Ainsi, nous nous intéressons à déterminer la vitesse à laquelle la probabilité $\mathbf{P}^{(m)}\left(\sum_{i=1}^{\pi(m)} X_i > (1+\epsilon)\log\log m\right)$ converge à 0.

Lemme 3.5.1. *Pour tout* $\epsilon > 0$,

$$\lim_{m\to\infty} \frac{1}{\log\log m} \log \mathbf{P}^{(m)}\big(D_m > (1+\epsilon)\log\log m\big) = -(1+\epsilon)\log\big(1+\epsilon\big) + \epsilon,$$

où $D_m := \sum_{i=1}^{\pi(m)} X_i$.

DÉMONSTRATION. Nous procéderons en deux étapes :

1) Majoration

Pour tout $s > 0$, par l'inégalité de Markov, nous obtenons

$$\begin{aligned}\mathbf{P}^{(m)}\big(D_m > (1+\epsilon)\log\log m\big) &= \mathbf{P}^{(m)}(e^{sD_m} > e^{s(1+\epsilon)\log\log m})\\ &\le \frac{\mathbb{E}e^{sD_m}}{e^{s(1+\epsilon)\log\log m}}.\end{aligned}$$

Notons par $M_Z(s) = \mathbb{E}e^{sZ}$ la fonction génératrice d'une certaine variable aléatoire Z. Vu que $X_1, \ldots, X_{\pi(m)}$ sont indépendantes alors $M_{D_m}(s) = \prod_{i=1}^{\pi(m)} M_{X_i}(s)$, où $M_{X_i}(s) = 1 + (e^s - 1)\frac{1}{p_i}$. Alors

$$\mathbf{P}^{(m)}\big(D_m > (1+\epsilon)\log\log m\big) \le \frac{\prod_{i=1}^{\pi(m)}\big(1+(e^s-1)\frac{1}{p_i}\big)}{e^{s(1+\epsilon)\log\log m}}. \tag{3.5.1}$$

Comme $1 + x \le e^x$:

$$\mathbf{P}^{(m)}\big(D_m > (1+\epsilon)\log\log m\big) \le e^{\mu_m(e^s-1)-s(1+\epsilon)\log\log m}, \tag{3.5.2}$$

où $\mu_m = \sum_{i=1}^{\pi(m)} \frac{1}{p_i}$. Étant donné que cette inégalité est vérifiée pour tout $s > 0$, alors elle le sera pour s_m qui minimise $\mu_m(e^s - 1) - s(1+\epsilon)\log\log m$, à savoir $s_m = \log\big((1+\epsilon)\frac{\log\log m}{\mu_m}\big)$. Il s'ensuit que

$$\frac{1}{\log\log m}\log \mathbf{P}^{(m)}\big(D_m > (1+\epsilon)\log\log m\big) \le \frac{\mu_m}{\log\log m}(e^{s_m} - 1) - s_m(1+\epsilon).$$

Donc :

$$\overline{\lim_{m\to\infty}}\left(\frac{1}{\log\log m}\log \mathbf{P}^{(m)}\big(D_m > (1+\epsilon)\log\log m\big)\right) \leq (e^{s_0}-1) - s_0(1+\epsilon). \tag{3.5.3}$$

où $s_0 = \log(1+\epsilon)$.

2) Minoration

Introduisons les variables aléatoires $Z_1, \ldots, Z_{\pi(m)}$ telles que

$$dG_i(x) := \frac{e^{s_0 x}}{M_{X_i}(s_0)} dF_i(x),$$

où F_i et G_i sont respectivement les fonctions de distribution de X_i et Z_i, et $1 \leq i \leq \pi(m)$. La fonction génératrice de Z_i est :

$$M_{Z_i}(s) = \int_0^\infty e^{st}\, dG_i(t) = \frac{1}{M_{X_i}(s_0)}\int_0^\infty e^{(s+s_0)t}\, dF_i(t) = \frac{M_{X_i}(s+s_0)}{M_{X_i}(s_0)}.$$

Sachant que

$$\mathbb{E}Z_i = \frac{d}{ds}M_{Z_i}(s)\bigg|_{s=0},$$

nous obtenons

$$\mathbb{E}Z_i = \frac{M'_{X_i}(s_0)}{M_{X_i}(s_0)}.$$

Nous calculons :

$$\mathbb{E}\sum_{i=1}^{\pi(m)} Z_i = \sum_{i=1}^{\pi(m)} \frac{M'_{X_i}(s_0)}{M_{X_i}(s_0)} = \sum_{i=1}^{\pi(m)} \frac{e^{s_0}}{p_i + (e^{s_0}-1)} = (1+\epsilon)\sum_{i=1}^{\pi(m)} \frac{1}{p_i+\epsilon}.$$

L'égalité

$$\begin{aligned}(1+\epsilon)\sum_{i=1}^{\pi(m)} \frac{1}{p_i+\epsilon} &= (1+\epsilon)\mu_m + (1+\epsilon)\sum_{i=1}^{\pi(m)}\Big(\frac{1}{p_i+\epsilon} - \frac{1}{p_i}\Big)\\ &= (1+\epsilon)\mu_m - \epsilon(1+\epsilon)\sum_{i=1}^{\pi(m)} \frac{1}{p_i(p_i+\epsilon)},\end{aligned}$$

nous donne

$$\mathbb{E}\sum_{i=1}^{\pi(m)} Z_i = (1+\epsilon)\log\log m - a_{m,\epsilon},$$

où $a_{m,\epsilon} = \epsilon(1+\epsilon)\sum_{i=1}^{\pi(m)} \frac{1}{p_i(p_i+\epsilon)} - C + O(1/\log m)$ en utilisant (3.1.1).

Calculons $\mathbb{V}\mathrm{ar} \sum_{i=1}^{\pi(m)} Z_i$:

$$\mathbb{E}{Z_i}^2 = \frac{d^2}{ds^2} M_{Z_i}(s)\bigg|_{s=0} = \frac{M''_{X_i}(s+s_0)}{M_{X_i}(s_0)}\bigg|_{s=0} = \frac{M''_{X_i}(s_0)}{M_{X_i}(s_0)}.$$

Comme $M''_{X_i}(s_0) = M'_{X_i}(s_0)$, alors

$$\mathbb{E}{Z_i}^2 = \mathbb{E}Z_i = \frac{1+\epsilon}{p_i+\epsilon}.$$

Par l'indépendance de $Z_1, \dots, Z_{\pi(m)}$, nous obtenons

$$\begin{aligned}\mathbb{V}\mathrm{ar} \sum_{i=1}^{\pi(m)} Z_i &= \sum_{i=1}^{\pi(m)} \frac{(1+\epsilon)}{p_i+\epsilon}\left(\frac{p_i-1}{p_i+\epsilon}\right) \\ &\le (1+\epsilon) \sum_{i=1}^{\pi(m)} \frac{1}{p_i+\epsilon}.\end{aligned}$$

En notant ν_m^2 la variance de $\sum_{i=1}^{\pi(m)} Z_i$, nous obtenons que $\nu_m^2 \sim (1+\epsilon)\log\log m$. Maintenant,

$$\begin{aligned}&\mathbf{P}^{(m)}\big(D_m > (1+\epsilon)\log\log m\big) = \int_{\left\{D_m > (1+\epsilon)\log\log m\right\}} d\mathbf{P}^{(m)} \\ &= \underset{\left\{\sum_{i=1}^{\pi(m)} x_i > (1+\epsilon)\log\log m\right\}}{\int \cdots \int} dF_1(x_1) \cdots dF_{\pi(n)}(x_{\pi(m)}) \\ &= \prod_{i=1}^{\pi(m)} M_{X_i}(s_0) \\ &\quad \times \underset{\left\{\sum_{i=1}^{\pi(m)} x_i > (1+\epsilon)\log\log m\right\}}{\int \cdots \int} e^{-s_0(x_1+\ldots+x_{\pi(n)})}\, dG_1(x_1) \cdots dG_{\pi(n)}(x_{\pi(m)}) \\ &= \left(\prod_{i=1}^{\pi(m)} M_{X_i}(s_0)\right) \mathbb{E}_{Z_1,\ldots,Z_{\pi(m)}} e^{-s_0(Z_1+\ldots+Z_{\pi(m)})} \mathbb{I}_{\left\{\sum_{i=1}^{\pi(m)} Z_i > (1+\epsilon)\log\log m\right\}}.\end{aligned}$$

Remarquons que les variables aléatoires $Z_1, \dots, Z_{\pi(m)}$ sont indépendantes et bornées, car $X_1, \dots, X_{\pi(m)}$ le sont aussi. Par le théorème central limite, il s'ensuit que

$$\frac{Z_1 + \cdots + Z_{\pi(m)} - (1+\epsilon)\log\log m + a_{m,\epsilon}}{\nu_m} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Si $K > 0$, alors

$$\mathbb{I}_{\left\{\sum_{i=1}^{\pi(m)} Z_i > (1+\epsilon)\log\log m\right\}} = \mathbb{I}_{\left\{\frac{\sum_{i=1}^{\pi(m)} Z_i - (1+\epsilon)\log\log m + a_{m,\epsilon}}{\nu_m} > \frac{a_{m,\epsilon}}{\nu_m}\right\}},$$

et donc

$$\mathbb{I}_{\left\{\sum_{i=1}^{\pi(m)} Z_i > (1+\epsilon)\log\log m\right\}} > \mathbb{I}_{\left\{\frac{a_{n,\epsilon}}{\nu_m} < \frac{\sum_{i=1}^{\pi(m)} Z_i - (1+\epsilon)\log\log m + a_{m,\epsilon}}{\nu_m} < K\right\}}.$$

Nous obtenons

$$\begin{aligned}
\mathbf{P}^{(m)}\big(D_m > (1+\epsilon)\log\log m\big) &> \left(\prod_{i=1}^{\pi(m)} M_{X_i}(s_0)\right) \\
&\quad \times \mathbb{E}_{Z_{p_1},\ldots,Z_{p_{\pi(m)}}} e^{-s_0(Z_{p_1}+\ldots+Z_{p_{\pi(m)}})} \\
&\quad \times \mathbb{I}_{\left\{\frac{a_{n,\epsilon}}{\nu_m} < \frac{\sum_{i=1}^{\pi(m)} Z_i - (1+\epsilon)\log\log m + a_{m,\epsilon}}{\nu_m} < K\right\}} \\
&> \left(\prod_{i=1}^{\pi(m)} M_{X_i}(s_0)\right) e^{-s_0((1+\epsilon)\log\log m - a_{m,\epsilon} + K\nu_m)} \\
&\quad \times \mathbb{E}_{Z_{p_1},\ldots,Z_{p_{\pi(m)}}} \mathbb{I}_{\left\{\frac{a_{n,\epsilon}}{\nu_m} < \frac{\sum_{i=1}^{\pi(m)} Z_i - (1+\epsilon)\log\log m + a_{m,\epsilon}}{\nu_m} < K\right\}} \\
&= \left(\prod_{i=1}^{\pi(m)} M_{X_i}(s_0)\right) e^{-s_0((1+\epsilon)\log\log m - a_{m,\epsilon} + K\nu_m)} \\
&\quad \times \mathbf{P}^{(m)}\left(\frac{a_{m,\epsilon}}{\nu_m} < \frac{\sum_{i=1}^{\pi(m)} Z_i - (1+\epsilon)\log\log m + a_{m,\epsilon}}{\nu_m} < K\right),
\end{aligned}$$

ou encore

$$\begin{aligned}
&\frac{1}{\log\log m}\log \mathbf{P}^{(m)}\big(D_m > (1+\epsilon)\log\log m\big) \\
&\quad > \frac{1}{\log\log m}\sum_{i=1}^{\pi(m)} \log M_{X_i}(s_0) - s_0(1+\epsilon) - \frac{a_{m,\epsilon}}{\log\log m} + K\frac{\nu_m}{\log\log m} \\
&\quad + \frac{1}{\log\log m}\log \mathbf{P}^{(m)}\left(\frac{a_{m,\epsilon}}{\nu_m} < \frac{\sum_{i=1}^{\pi(m)} Z_i - (1+\epsilon)\log\log m + a_{m,\epsilon}}{\nu_m} < K\right).
\end{aligned} \tag{3.5.4}$$

Comme $a_{m,\epsilon} \leq \epsilon(1+\epsilon)\sum_{i\geq 1}\frac{1}{p_i^2} + C'$, alors $\lim_{n\to\infty}\frac{a_{m,\epsilon}}{\log\log m} = 0$. D'autre part, un calcul immédiat montre que

$$\lim_{m\to\infty}\frac{1}{\log\log m}\sum_{i=1}^{\pi(m)}\log\Big(1+\frac{\epsilon}{p_i}\Big) = \epsilon = e^{s_0}-1.$$

Aussi, pour m assez grand, $\frac{a_{m,\epsilon}}{\nu_m} < 1/2$, et donc

$$\mathbf{P}^{(m)}\bigg(\frac{a_{m,\epsilon}}{\nu_m} < \frac{\sum_{i=1}^{\pi(m)} Z_i - (1+\epsilon)\log\log m + a_{m,\epsilon}}{\nu_m} < K\bigg)$$
$$\geq \mathbf{P}^{(m)}\bigg(\frac{1}{2} < \frac{\sum_{i=1}^{\pi(m)} Z_i - (1+\epsilon)\log\log m + a_{m,\epsilon}}{\nu_m} < K\bigg),$$

et

$$\mathbf{P}^{(m)}\bigg(\frac{1}{2} < \frac{\sum_{i=1}^{\pi(m)} Z_i - (1+\epsilon)\log\log m + a_{m,\epsilon}}{\nu_m} < K\bigg) = \frac{1}{4},$$

pour une valeur de $K > 0$ appropriée ; ceci implique que

$$\lim_{m\to\infty}\frac{1}{\log\log m}\log\mathbf{P}^{(m)}\bigg(\frac{1}{2} < \frac{\sum_{i=1}^{\pi(m)} Z_i - (1+\epsilon)\log\log m + a_{m,\epsilon}}{\nu_m} < K\bigg) = 0.$$

Par (3.5.4), nous obtenons

$$\underline{\lim}_{m\to\infty}\bigg(\frac{1}{\log\log m}\log\mathbf{P}^{(m)}\big(D_m > (1+\epsilon)\log\log m\big)\bigg) \geq (e^{s_0}-1) - s_0(1+\epsilon). \tag{3.5.5}$$

Il suffit alors de remarquer que $(e^{s_0}-1) - s_0(1+\epsilon) = -(1+\epsilon)\log\big(1+\epsilon\big)+\epsilon$, et par la formule (3.5.3) le théorème est démontré. □

Remarquons que pour des petites valeurs de $\epsilon > 0$ nous obtenons l'approximation $-(1+\epsilon)\log\big(1+\epsilon\big)+\epsilon \approx -\frac{\epsilon^2}{2}$.

3.6. La loi asymptotique de $\sum_{i=\pi(m^\alpha)}^{m} X_i$

Soit m un nombre entier positif et α un réel appartenant à $(0,1)$. Posons $D_{m,\alpha} := \sum_{i=\pi(m^\alpha)+1}^{\pi(m)} X_i$. Nous démontrerons le lemme suivant :

Lemme 3.6.1. *Pour tout* $\alpha \in (0,1)$

$$D_{m,\alpha} \xrightarrow{\mathcal{L}} \mathcal{P}_{\log(1/\alpha)},$$

lorsque $m \to \infty$.

DÉMONSTRATION. Premièrement, remarquons que

$$\begin{aligned}
\mathbb{E}D_{m,\alpha} &= \mathbb{E}\Big(\sum_{i=1}^{\pi(m)} X_i\Big) - \mathbb{E}\Big(\sum_{i=1}^{\pi(m^\alpha)} X_i\Big) \\
&= \sum_{i=1}^{\pi(m)} \frac{1}{p_i} - \sum_{i=1}^{\pi(m^\alpha)} \frac{1}{p_i} \\
&= \log\log m + C + O(1/\log m) \\
&\quad - \big(\log\log m^\alpha + C + O(1/\log m^\alpha)\big) \qquad \text{, par (3.1.1)} \\
&= \log\Big(\frac{1}{\alpha}\Big) + O(1/\log m).
\end{aligned}$$

D'où $\lim_{n\to\infty} \mathbb{E}D_{m,\alpha} = \log\big(\frac{1}{\alpha}\big)$.

$D_{m,\alpha}$ étant une somme de variables aléatoires indépendantes, sa fonction génératrice est le produit des fonction génératrices de $X_{\pi(m^\alpha)+1}, \ldots, X_{\pi(m)}$:

$$G_{D_{m,\alpha}}(s) = \prod_{i=\pi(m^\alpha)+1}^{\pi(m)} \big(1 - \theta_i + \theta_i s\big),$$

où $\theta_i = \frac{1}{p_i}$. Alors

$$\begin{aligned}
\log G_{D_{m,\alpha}}(s) &= \sum_{i=\pi(m^\alpha)+1}^{\pi(m)} \log\big(1 + (s-1)\theta_i\big) \\
&= \sum_{i=\pi(m^\alpha)+1}^{\pi(m)} \Big((s-1)\theta_i + \sum_{k\geq 2} (-1)^{k+1} \frac{(s-1)^k \theta_i^k}{k}\Big) \\
&= (s-1)\mathbb{E}D_{m,\alpha} + \sum_{k\geq 2} (-1)^{k+1} \frac{(s-1)^k}{k} \sum_{i=\pi(m^\alpha)+1}^{\pi(m)} \theta_i^k,
\end{aligned}$$

car $\theta_i = \mathbb{E}X_i$. Comme

$$\sum_{i=\pi(m^\alpha)+1}^{\pi(m)} \theta_i^k \leq \max_{\pi(m^\alpha)+1,\ldots,\pi(m)} \theta_i^{k-1} \sum_{i=\pi(m^\alpha)+1}^{\pi(m)} \theta_i \leq \frac{\log\big(\frac{1}{\alpha}\big) + 1}{p_{\pi(m^\alpha)+1}},$$

pour tout m assez grand :

$$\begin{aligned}\left|\sum_{k\geq 2}(-1)^{k+1}\frac{(s-1)^k}{k}\sum_{i=\pi(m^\alpha)+1}^{\pi(m)}\theta_i^k\right| &\leq \frac{\log\left(\frac{1}{\alpha}\right)+1}{p_{\pi(m^\alpha)+1}}\sum_{k\geq 2}|s-1|^k \\ &\leq \frac{1}{p_{\pi(m^\alpha)+1}}\left(\frac{\log\left(\frac{1}{\alpha}\right)+1}{1-|s-1|}\right),\end{aligned}$$

et alors

$$\begin{aligned}\lim_{m\to\infty}\log G_{D_{m,\alpha}}(s) &= (s-1)\lim_{m\to\infty}\mathbb{E}D_{m,\alpha} \\ &= (s-1)\log\left(\frac{1}{\alpha}\right).\end{aligned}$$

Il s'ensuit que

$$\lim_{m\to\infty} G_{D_{m,\alpha}}(s) = e^{(s-1)\log(\frac{1}{\alpha})},$$

qui est la fonction génératrice de la loi Poisson de paramètre $\log(\frac{1}{\alpha})$. Par le théorèmes A.0.4 (voir Annexe A), il s'ensuit que $D_{m,\alpha} \xrightarrow{\mathcal{L}} \mathcal{P}_{\log 1/\alpha}$, lorsque $m \to \infty$. $\square$

Chapitre 4

DÉMONSTRATION DES SEPT THÉORÈMES

4.1. Théorème 1.0.1

Démonstration. Posons $A_{u,v} = \left\{n \in \mathbb{N} : u \leq \frac{f_m(n) - \log\log m}{\sqrt{\log\log m}} \leq v\right\}$. Alors par le théorème 2.2.1,

$$\mathcal{D}\big(A_{u,v}\big) = \mathbf{P}^{(m)}\left(u \leq \frac{\sum_{i=1}^{\pi(m)} X_i - \log\log m}{\sqrt{\log\log m}} \leq v\right).$$

En prenant les limites de chaque côté, nous obtenons le résultat. □

Par exemple, si $u = -1,96, v = 1,96$ et $m \approx 10^{70}$ alors $\log\log m \approx 5$ et

$$\mathcal{D}\big(1 \leq f_m \leq 9\big) \approx 0,95;$$

c'est-à-dire 95% des nombres ont au plus 9 facteurs premiers parmi $p_1, \ldots, p_m$.

4.2. Théorème 1.0.2

Démonstration. Par le lemme 3.3.1, nous savons que

$$\max_{0 \leq \tau \leq 1} \mathcal{S}_\tau^{(m)} \xrightarrow{\mathcal{L}} \max_{0 \leq \tau \leq 1} B_\tau,$$

lorsque $m \to \infty$. Or, l'équation (3.2.2) implique que

$$\max_{0 \leq \tau \leq 1} \mathcal{S}_\tau^{(m)} = \max_{1 \leq l \leq \pi(m)} \mathcal{S}_{\tau_l}^{(m)},$$

car pour $\tau \in [\tau_l, \tau_{l+1}]$ nous avons que $\mathcal{S}_\tau^{(m)} \leq \mathcal{S}_{\tau_l}^{(m)}$ ou bien $\mathcal{S}_\tau^{(m)} \leq \mathcal{S}_{\tau_{l+1}}^{(m)}$. D'autre part, par (3.2.1) et (3.2.3), nous obtenons

$$S_l^{(m)} = \sqrt{t_{\pi(m)}}\,\mathcal{S}_{\tau_l}^{(m)},$$

lorsque $1 \leq l \leq \pi(m)$, et donc

$$\frac{1}{\sqrt{t_{\pi(m)}}} \max_{1 \leq l \leq \pi(m)} S_l^{(m)} = \max_{1 \leq l \leq \pi(m)} \mathcal{S}_{\tau_l}^{(m)}.$$

Il s'ensuit que

$$\frac{1}{\sqrt{t_{\pi(m)}}} \max_{1 \leq l \leq \pi(m)} S_l^{(m)} \xrightarrow{\mathcal{L}} \max_{0 \leq \tau \leq 1} B_\tau,$$

lorsque $m \to \infty$, et nous obtenons :

$$\lim_{m \to \infty} \mathbf{P}^{(m)}\left(\frac{1}{\sqrt{t_{\pi(m)}}} \max_{1 \leq l \leq \pi(m)} S_l^{(m)} \geq x\right) = \frac{2}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{1}{2}t^2}\, dt.$$

Sachant que $t_{\pi(m)} \sim \log\log m$, alors

$$\lim_{m \to \infty} \mathbf{P}^{(m)}\left(\frac{1}{\sqrt{\log\log m}} \max_{1 \leq l \leq \pi(m)} S_l^{(m)} \geq x\right) = \frac{2}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{1}{2}t^2}\, dt.$$

Posons

$$A_x = \left\{ n \in \mathbb{N} : \frac{1}{\sqrt{\log\log m}} \max_{1 \leq l \leq \pi(m)} \left(\sum_{i=1}^{l} \mathbb{I}_{p_i}(n) - \sum_{i=1}^{l} \frac{1}{p_i} \right) \geq x \right\}.$$

Alors par le théorème 2.2.1,

$$\mathcal{D}\big(A_x\big) = \mathbf{P}^{(m)}\left(\frac{1}{\sqrt{\log\log m}} \max_{1 \leq l \leq \pi(m)} S_l^{(m)} \geq x\right).$$

Il suffit de prendre les limites de chaque côté. $\square$

Par exemple, lorsque $m \approx 10^{70}$ et $x = 1,7$, alors

$$\mathcal{D}\big(B_{3,8}\big) \approx 0,0982;$$

que nous interprétons comme suit : approximativement 10% des nombres seront *"très composés"*, lorsque nous considérons uniquement les facteurs premiers $p_1, \ldots, p_m$.

4.3. Théorème 1.0.3

Commençons par justifier la définition 1.0.3. Nous dirons que $n \in [N]$ *est excessif au temps* l lorsqu'il possède plus de diviseurs premiers parmi $p_1, \dots, p_l$ que le nombre moyen, c'est-à-dire lorsque

$$\sum_{i=1}^{l} \mathbb{I}_{P_i}(n) - \sum_{i=1}^{l} \frac{1}{p_i}(n) > 0.$$

Nous voulons déterminer la proportion du temps où n est excessif. Précisons ce que nous entendons par *proportion du temps*. Pour cela nous avons recours à la marche aléatoire normalisé $\{\mathcal{S}^{(m)}_{\tau_l}\}_{l \in [\pi(m)]}$. En effet, par (2.2.2) :

$$\begin{aligned} \sum_{i=1}^{l} \mathbb{I}_{P_i}(n) - \sum_{i=1}^{l} \frac{1}{p_i} > 0 \iff & \ S^{(m)}_l\big(d_m(n)\big) > 0 \\ \iff & \ \mathcal{S}^{(m)}_{\tau_l}\big(d_m(n)\big) > 0. \end{aligned}$$

En vertu de (3.2.3), nous pouvons attribuer la valeur

$$\int_{\tau_l}^{\tau_{l+1}} \mathbb{I}_{\mathcal{S}^{(m)}_\tau(d_m(n))>0}\, d\tau,$$

à la proportion du temps où n est excessif au temps l, d'où

$$\int_0^1 \mathbb{I}_{\mathcal{S}^{(m)}_\tau(d_m(n))>0}\, d\tau,$$

représente la proportion recherchée. Il s'ensuit que la fonction $L_m : \Omega_m \to \mathbb{R}$ telle que

$$L_m = \int_0^1 \mathbb{I}_{\mathcal{S}^{(m)}_\tau>0}\, d\tau.$$

représente bien la proportion du temps pour laquelle f_m est excessive.

Démonstration. Par le lemme 3.3.1, nous avons

$$L_m \xrightarrow{\mathcal{L}} L_B,$$

lorsque $m \to \infty$, où $L_B = \int_0^1 \mathbb{I}_{B_\tau>0}\, d\tau$ suit la loi de l'Arc Sinus, voir **[1]** (*pp.80–82*). Il s'ensuit que

$$\mathbf{P}^{(m)}(a \leq L_m \leq b) \longrightarrow \frac{1}{\pi}\int_a^b \frac{dv}{\sqrt{v(1-v)}},$$

lorsque $m \to \infty$. $\square$

Par exemple,

$$\mathbf{P}^{(m)}(L_m \geq 0,9) = \mathbf{P}^{(m)}(L_m \leq 0,1) \approx 0,2048,$$

et

$$\mathbf{P}^{(m)}(0,45 \leq L_m \leq 0,55) \approx 0,0638;$$

ce qui se traduit comme suit : pour 20% des nombres, f_m sera excessive pendant au moins 90% du temps ou au plus 10% du temps, et pour 6% des nombres, elle sera excessive entre 45% et 55% du temps.

4.4. THÉORÈME 1.0.4

Définition 4.4.1. *Soit* $U_m : [N] \to \mathbb{R}$ *tel que*

$$U_m := \frac{1}{\pi(m)} \left| \left\{ l \leq \pi(m) : \sum_{i=1}^{l} \mathbb{I}_{P_i} > \sum_{i=1}^{l} \frac{1}{p_i} \right\} \right|.$$

Alors U_m *représente la proportion des nombres premiers pour laquelle* f_m *est excessive.*

DÉMONSTRATION. Notons que, par la définition 3.4.1, nous avons

$$V_m := \frac{1}{\pi(m)} \left| \left\{ l \leq \pi(m) : \sum_{i=1}^{l} X_i > \sum_{i=1}^{l} \frac{1}{p_i} \right\} \right|.$$

Posons

$$A_y = \left\{ n \in \mathbb{N} : U_m(n) \leq y \right\},$$

où $y \in \mathbb{R} \setminus \left\{0\right\}$. On a $\mathcal{D}(A_1) = 1$, et par le théorème 2.2.1,

$$\mathcal{D}(A_y) = \mathbf{P}^{(m)}(V_m \leq y),$$

lorsque $y \notin \left\{0, 1\right\}$. En prenant les limites de chaque côté, et par le lemme 3.4.2, nous obtenons que $\lim_{m\to\infty} \mathcal{D}(A_y) = \frac{1}{2}\mathbb{I}_{(0,1)}(y) + \mathbb{I}_{[1,\infty)}(y)$. $\square$

4.5. THÉORÈME 1.0.5

DÉMONSTRATION. Posons

$$A_k = \{n \in \mathbb{N} : f_m(n) - f_{m^\alpha}(n) = k\}.$$

Alors par le théorème 2.2.1,

$$\mathcal{D}(A_k) = \mathbf{P}^{(m)}(D_{m,\alpha} = k).$$

En prenant les limites de chaque côtés, et par le lemme 3.6.1, nous obtenons le résultat. □

4.6. THÉORÈME 1.0.6

DÉMONSTRATION. Posons

$$A_\epsilon = \{n \in \mathbb{N} : f_m(n) > (1+\epsilon) \log\log m\}.$$

Alors par le théorème 2.2.1,

$$\mathcal{D}(A_\epsilon) = \mathbf{P}^{(m)}(D_m > (1+\epsilon)\log\log m),$$

et donc

$$\frac{1}{\log\log m} \log \mathcal{D}(A_\epsilon) = \frac{1}{\log\log m} \log \mathbf{P}^{(m)}(D_m > (1+\epsilon)\log\log m).$$

Il suffit de prendre les limites de chaque côté et d'utliser le lemme 3.5.1. □

4.7. THÉORÈME 1.0.7

DÉMONSTRATION. Soit $D_m := \sum_{i=1}^{\pi(m)} X_i$. Il suffit de montrer que $\frac{D_N - \log\log N}{\sqrt{\log\log N}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ lorsque $N \to \infty$, car alors par (2.2.2),

$$\begin{aligned}
\mathbf{Q}_N\left(\frac{f - \log\log N}{\sqrt{\log\log N}} \le v\right) &= \mathbf{Q}_N\left(d_N^{-1}\left(\frac{D_N - \log\log N}{\sqrt{\log\log N}} \le v\right)\right) \\
&= \mathbf{Q}_N^{(N)}\left(\frac{D_N - \log\log N}{\sqrt{\log\log N}} \le v\right).
\end{aligned}$$

Soit $r_N := N^{\frac{1}{\log\log N}}$ pour $N \geq 1$. Nous montrerons que $\frac{D_N - D_{r_N}}{\sqrt{\log\log N}} \xrightarrow{\mathcal{L}} 0$. Déterminons une borne supérieure de $\mathbb{E}_{\mathbf{Q}_N^{(N)}} |D_N - D_{r_N}|$:

$$\mathbb{E}_{\mathbf{Q}_N^{(N)}} |D_N - D_{r_N}| = \sum_{r_N < p \leq N} \frac{1}{N} \left\lfloor \frac{N}{p} \right\rfloor \leq \sum_{r_N < p \leq N} \frac{1}{p} \sim \log\log\log N,$$

en utilisant (3.1.1). Par l'inégalité de Markov, nous obtenons

$$\mathbf{Q}_N^{(N)}\left(|D_N - D_{r_N}| \geq \epsilon\sqrt{\log\log N}\right) \leq \frac{\mathbb{E}_{\mathbf{Q}_N^{(N)}} |D_N - D_{r_N}|}{\epsilon\sqrt{\log\log N}} \sim \frac{\log\log\log N}{\epsilon\sqrt{\log\log N}},$$

où le dernier quotient tend vers 0 lorsque $N \to \infty$. Alors $\frac{D_N - D_{r_N}}{\sqrt{\log\log N}} \xrightarrow{\mathcal{L}} 0$ lorsque $N \to \infty$, et donc il suffit de montrer que

$$\lim_{N\to\infty} \mathbf{Q}_N^{(N)}\left(\frac{D_{r_N} - \mu_{r_N}}{\sqrt{\mu_{r_N}}} \leq \nu\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\nu} e^{-\frac{t^2}{2}}\, dt,$$

où $\mu_{r_N} := \log\log N$.

Étant donné que $\mathbb{E}_{\mathbf{P}^{(N)}} D_{r_N} \sim \mu_{r_N}$ et $\mathbb{V}\mathrm{ar}_{\mathbf{P}^{(N)}} D_{r_N} \sim \mu_{r_N}$, alors par le théorème central limite :

$$\lim_{N\to\infty} \mathbf{P}^{(N)}\left(\frac{D_{r_N} - \mu_{r_N}}{\sqrt{\mu_{r_N}}} \leq \nu\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\nu} e^{-\frac{t^2}{2}}\, dt.$$

En utilisant le théorème A.0.5 (voir Annexe A),

$$\mathbb{E}_{\mathbf{P}^{(N)}} \left(\frac{D_{r_N} - \mu_{r_N}}{\sqrt{\mu_{r_N}}}\right)^k \to \mathbb{E}\mathcal{N}(0,1)^k,$$

pour tout $k \in \mathbb{N}$, lorsque $N \to \infty$. En vertu du théorème A.0.6 (voir Annexe A), il suffit de montrer que pour tout $k \geq 1$,

$$\mathbb{E}_{\mathbf{P}^{(N)}} \left(\frac{D_{r_N} - \mu_{r_N}}{\sqrt{\mu_{r_N}}}\right)^k - \mathbb{E}_{\mathbf{Q}_N^{(N)}} \left(\frac{D_{r_N} - \mu_{r_N}}{\sqrt{\mu_{r_N}}}\right)^k \to 0,$$

pour tout $k \in \mathbb{N}$, lorsque $N \to \infty$. Or,

$$\begin{aligned}
&|\mathbb{E}_{\mathbf{P}^{(N)}} \left(D_{r_N} - \mu_{r_N}\right)^k - \mathbb{E}_{\mathbf{Q}_N^{(N)}} \left(D_{r_N} - \mu_{r_N}\right)^k| \\
&= \left| \sum_{j=0}^{k} \binom{k}{j} \left[\mathbb{E}_{\mathbf{P}^{(N)}} D_{r_N}^j - \mathbb{E}_{\mathbf{Q}_N^{(N)}} D_{r_N}^j\right] (-\mu_{r_N})^{k-j} \right| \\
&\leq \sum_{j=0}^{k} \binom{k}{j} \left|\mathbb{E}_{\mathbf{P}^{(N)}} D_{r_N}^j - \mathbb{E}_{\mathbf{Q}_N^{(N)}} D_{r_N}^j\right| \mu_{r_N}^{k-j},
\end{aligned}$$

et

$$\begin{aligned}
&\left|\mathbb{E}_{\mathbf{P}^{(N)}} D_{r_N}^{j} - \mathbb{E}_{\mathbf{Q}_N^{(N)}} D_{r_N}^{j}\right| \\
&\leq \sum_{(i_1,\ldots,i_j)\in[\pi(r_N)]^j} \left|\mathbb{E}_{\mathbf{P}^{(N)}} X_{p_{i_1}}\cdots X_{p_{i_j}} - \mathbb{E}_{\mathbf{Q}_N^{(N)}} X_{p_{i_1}}\cdots X_{p_{i_j}}\right| \\
&\leq \sum_{(i_1,\ldots,i_j)\in[\pi(r_N)]^j} \frac{1}{N}, \qquad \text{, par (2.2.4) et (2.2.5)} \\
&\leq \frac{r_N^j}{N},
\end{aligned}$$

d'où

$$\begin{aligned}
\left|\mathbb{E}_{\mathbf{P}^{(N)}}\left(D_{r_N}-\mu_{r_N}\right)^k - \mathbb{E}_{\mathbf{Q}_N^{(N)}}\left(D_{r_N}-\mu_{r_N}\right)^k\right| &\leq \frac{1}{N}\sum_{j=0}^{k}\binom{k}{j} r_N^j \mu_{r_N}^{k-j} \\
&= \frac{1}{N}\left(r_N+\mu_{r_N}\right)^k \\
&\leq 2^k \frac{r_N^k}{N} \\
&\leq 2^k \frac{1}{N^{1-\frac{k}{\log\log N}}},
\end{aligned}$$

qui tend vers 0 lorsque $N \to \infty$. Le résultat est démontré. $\square$

CONCLUSION

Notons qu'au théorème 1.0.5 nous calculons d'abord la densité pour un $m \in \mathbb{N}$ fixé, et puis nous évaluons sa limite. Si m était une fonction de N le résultat serai faux en général : lorsque $1/2 \le \alpha < 1$,

$$\mathcal{D}\big(f_N - f_{N^\alpha} = k\big) = \mathbb{I}_{\{0\}}(k),$$

car

$$\frac{\big|\{n \le N : f_N(n) - f_{N^\alpha}(n) \ge 1\}\big|}{N} = \frac{\pi(N) - \pi(N^\alpha)}{N},$$

et donc

$$\mathcal{D}\big(f_N - f_{N^\alpha} \ge 1\big) = 0.$$

Clairement il n'existe aucun $\lambda > 0$ tel que $\mathcal{D}\big(f_N - f_{N^\alpha} = k\big) = e^{-\lambda}\frac{\lambda^k}{k!}$.

D'autre part, en ce qui concerne notre conjecture sur les grandes déviations, il est facile de montrer que

$$\overline{\lim_{N\to\infty}} \frac{1}{\log\log N}\left(\log \frac{\big|\{n \le N : f(n) > (1+\epsilon)\log\log N\}\big|}{N}\right) \le -(1+\epsilon)\log\big(1+\epsilon\big) + \epsilon.$$

En effet, il suffit de considérer l'événement qui lui correspond, suivant d_N, dans l'espace $\big(\Omega_m, \mathcal{F}_m\big)$, et de reproduire la première partie de la démonstration du lemme 3.5.1, en notant que les fonctions génératrices satisfont l'inégalité $\mathbb{E}_{\mathbf{Q}_N^{(N)}} e^{sD_m} \le \mathbb{E}_{\mathbf{P}^{(N)}} e^{sD_m}$.

Annexe A

RÉSULTATS DE SOURCES EXTERNES

Voici les résultats utilisés dans le mémoire :

Théorème A.0.1 (Kolmogorov, p.81 de **[8]**)**.** *Let* $(\Lambda_n : n \in \mathbb{N})$ *be a sequence of probability measures on* $(\mathbb{R}, \mathcal{T})$ *Define*

$$\Omega = \prod_{n \in \mathbb{N}} \mathbb{R},$$

so that a typical element ω *of* Ω *is a sequence* (ω_m) *in* $\mathbb{R}$*. Define*

$$X_n : \Omega \to \mathbb{R}, \qquad X_n(\omega) := \omega_n,$$

and let $\mathcal{F} := \sigma(X_n : n \in \mathbb{N})$*. Then there exists a unique probability measure* $\mathbf{P}$ *on* $(\mathbb{R}, \mathcal{F})$ *such that for* $r \in \mathbb{N}$ *and* $B_1, \ldots, B_r \in \mathcal{T}$*,*

$$\mathbf{P}\bigg(\Big(\prod_{1 \leq k \leq r} B_k\Big) \times \prod_{k > r} \mathbb{R}\bigg) = \prod_{1 \leq k \leq r} \Lambda_k(B_k).$$

Then the sequence $(X_n : n \in \mathbb{N})$ *is a sequence of independent random variables on* $(\Omega, \mathcal{F}, \mathbf{P})$*,* X_n *having law* Λ_n*.*

Théorème A.0.2 (Prokhorov, p.190 de **[6]**)**.** *Let* $\xi_{n,1}, \ldots, \xi_{n,k_n}$ *be a double sequence of random variable which are independent for each* n *and subject to the condition of "asymptotic negligibility" : for all* $\epsilon > 0$

$$\lim_{n \to \infty} \max_{1 \leq i \leq k_n} \mathbf{P}(|\xi_{n,i}| > \epsilon) = 0,$$

and

$$\mathbb{E}\xi_{n,i} = 0, \sigma^2_{n,i} = \mathbb{V}\mathrm{ar}\xi_{n,i} > 0, \sum_{i=1}^{k_n} \sigma^2_{n,i} = 1.$$

Assume that

$$\zeta_{n,0} = 0, \qquad \zeta_{n,i} = \xi_{n,1} + \ldots + \xi_{n,i} \qquad (1 \leq i \leq k_n),$$

$t_{n,i} = \mathbb{V}\mathrm{ar}\zeta_{n,i}$ *and* $\xi_n(t)$ *is a "random broken line l" with vertices* $(t_{n,i}, \zeta_{n,i})$*. Define* $\mathbf{P}_n$ *as its corresponding distribution in* $C[0,1]$ *and let* W *denote the "Wiener distribution" in* $C[0,1]$*. Then for the convergence*

$$\mathbf{P}_n \Rightarrow W,$$

the condition of Lindeberg is necessary and sufficient : For every $\lambda > 0$*,*

$$\lim_{n\to\infty} \sum_{i=1}^{k_n} \int_{|x|>\lambda} x^2 \, dF_{n,k}(x) = 0,$$

where $F_{n,k}(x) = \mathbf{P}(\xi_{n,i} < x)$*.*

Théorème A.0.3 (Principe d'invariance de Donsker, p.72 de **[1]**)**.** *If* h *is continuous on* $C[0,1]$ *— or continuous except at points forming a set of Wiener measure 0 — then* $X_n \xrightarrow{\mathcal{L}} W$ *implies*

$$h(X_n) \xrightarrow{\mathcal{L}} h(W),$$

where W is a Brownian motion.

Théorème A.0.4 (théorème de la continuité, p. 280 de **[4]**)**.** *Suppose that for every fixed* n *the sequence* $a_{0,n}, a_{1,n}, \ldots$ *is a probability distribution. In order that a limit*

$$a_k = \lim_{n\to\infty} a_{k,n},$$

exists for every $k \geq 0$ *it is necessary and sufficient that the limit*

$$A(s) = \lim_{n\to\infty} \sum_{k\geq 0} a_{k,n} s^k,$$

exists for each s *in the open interval* $0 < s < 1$*. In this case automatically*

$$A(s) = \sum_{k\geq 0} a_k s^k,$$

Théorème A.0.5 (Corollary 25.12, p. 338 de **[3]**)**.** *Let* r *be a positive integer. If* $X_n \xrightarrow{\mathcal{L}} X$ *and* $\sup_n \mathbb{E}|X_n|^{r+\epsilon} < \infty$*, where* $\epsilon > 0$*, then* $\mathbb{E}|X|^r < \infty$ *and* $\mathbb{E}X_n^r \to \mathbb{E}X^r$*.*

Théorème A.0.6 (Theorem 30.2, p. 390 de [**3**])**.** *Suppose that the distribution of* X *is determined by its moments, that the* X_n *have moments of all orders, and that* $\lim_{n\to\infty} \mathbb{E}X_n^r = \mathbb{E}X^r$ *for* $r = 1, 2, \ldots$. *Then* $X_n \xrightarrow{\mathcal{L}} X$.

BIBLIOGRAPHIE

[1] BILLINGSLEY, P., *Convergence of Probability Measures*, Wiley, New-York, (1968)

[2] BILLINGSLEY, P., Prime numbers and brownian motion, *The American Mathematical Monthly*, **Vol. 80, No. 10**, (1973), 1099–1115

[3] BILLINGSLEY, P., *Probability and Measure*, Wiley, New-York, (1995)

[4] FELLER, W., *An Introduction to Probability Theory and its Applications, vol. 1, third edition*, Wiley, New-York, (1968)

[5] HARDY, G. H. AND RAMANUJAN, S., The normal number of prime factors of a number n, *Quarterly Journal of Mathematics*, **48**, (1917), 76–92

[6] PROKHOROV, YU. V., Convergence of random processes and limit theorems in probability theory, *Theory of Probability and its Applications*, **Vol. 1, No. 2**, (1956), 157–214

[7] TAO, T. AND VU, VAN H., *Additive Combinatorics*, Cambridge University Press, (2010)

[8] WILLIAMS, D., *Probability with Martingales*, Cambridge University Press, (1991)