Rapport de stage de L3 - Année 2021
Encadrant : Sary Drappeau

Le théorème d'Erdös-Kac

Par Bastien Lecluse

Table des matières

Dans toute la suite, et sauf indications contraires, la lettre p sera réservée aux nombres premiers, la notation log désignera le logarithme népérien, et la notation φ les fonctions caractéristiques de variables aléatoires.

1. Introduction

Pour un entier naturel $n \geqslant 2$, on pose

$$\omega(n) = \mathrm{Card}\,\{p : p|n\},$$

autrement dit $\omega(n)$ est le nombre de diviseurs premiers deux à deux distincts de n. Cette fonction a été introduite en 1917 par G. H. Hardy et S. Ramanujan dans un article de théorie probabiliste des nombres. Ainsi :

$$\omega(3) = 1 \quad \text{et} \quad \omega(5^2 \cdot 7 \cdot 17^2) = 3.$$

Il est clair que la fonction ω possède de grandes irrégularités, elle vaut 1 en toute puissance d'un nombre premier, tandis que

$$\omega(p_1 \cdots p_m) = m,$$

où $p_1, \ldots, p_m$ sont des nombres premiers deux à deux distincts. Nous allons donc plutôt nous intéresser à une estimation en moyenne de ω. Cette dernière est relativement facile à obtenir, et on obtient pour $x \geqslant 0$

$$\sum_{p \leqslant x} \omega(n) = x \log(\log(x)) + O(x)$$

(ceci sera démontré dans la section 2.3).
En 1939, les mathématiciens P. Erdös et M. Kac établissent le théorème suivant, concernant les fluctuations de ω par rapport à cette moyenne, connu sous le nom du théorème d'Erdös-Kac :

Théorème (Erdös-Kac). *Pour tout réel x,*

$$\frac{1}{n}\mathrm{Card}\left\{i \in [\![1, n]\!] : \frac{\omega(i) - \log(\log(i))}{\log(\log(i))^{1/2}} \leqslant x\right\} \underset{n\to+\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} \mathrm{d}t.$$

Nous démontrerons le théorème d'Erdös-Kac de deux manières différentes. Dans la partie 3, nous utiliserons la méthode des moments, qui permet de montrer des convergences en loi sans utiliser les fonctions génératrices. La méthode des moments se trouve être intéressante lorsque nous considérons des variables aléatoires dépendantes (ce qui est la cas ici). Dans la partie 4, nous utiliserons la méthode de Stein.

2. Résultats préliminaires

2.1. Quelques outils d'analyse

Nous aurons besoin dans la suite d'utiliser une généralisation de l'intégrale de Riemann : l'intégrale de Stieltjes.

Définition 1. Soient $a, b \in \mathbb{R}$ tels que $a < b$, $f, g : [a, b] \longrightarrow \mathbb{R}$ deux foncions réelles bornées, et une subdivision $a = x_0 < x_1 < \cdots < x_n = b$ de l'intervalle $[a, b]$ ($n \in \mathbb{N}^*$). Pour tout $i \in [\![1, n-1]\!]$, on se donne un réel $\xi_i \in [x_i, x_{i+1}]$. Si la quantité

$$\sum_{i=0}^{n-1} f(\xi_i)(g(x_{i+1}) - g(x_i))$$

admet une limite I lorsque que le pas de la subdivision tend vers 0, alors I est appelé l'intégrale de Stieltjes de f par rapport à g. On la note :

$$\int_a^b f(x)\mathrm{d}g(x).$$

Remarque 1. Si g est l'identité, alors l'intégrale de Stieltjes de f par rapport à g coïncident avec l'intégrale de Riemann.

Proposition 1. *Soient $a, b \in \mathbb{R}$ tels que $a < b$, $f, g : [a, b] \longrightarrow \mathbb{R}$ deux foncions réelles bornées. On suppose que g est de classe $\mathcal{C}^1$ et que l'intégrale de Stieltjes de f par rapport à g existe. Alors fg' est intégrable au sens de Riemann et*

$$\int_a^b f(x)\mathrm{d}g(x) = \int_a^b f(x)g'(x)\mathrm{d}x.$$

Démonstration. Pour $n \in \mathbb{N}^*$ on note

$$S_n = \sum_{i=0}^{n-1} f(\xi_i)(g(x_{i+1}) - g(x_i)),$$

les ξ_i étant choisi comme dans la définition 1 et

$$T_n = \sum_{i=0}^{n-1} f(\xi_i)g'(\xi_i)(x_{i+1} - x_i).$$

Montrons que

$$\lim_{n\to+\infty} S_n = \lim_{n\to+\infty} T_n.$$

Implicitement nous faisons tendre le pas des subdivisions vers 0 lorsque nous faisons tendre n vers $+\infty$. Pour $i \in [\![1, n-1]\!]$ on pose $\Delta x_i = x_{i+1} - x_i$, en gardant à l'esprit que cette quantité dépend de n. Soit $\varepsilon > 0$. Par continuité de g', il existe $N_1 \in \mathbb{N}$ tel que pour tout $n \geqslant N_1$ et pour tout $i \in [\![1, n]\!]$

$$\Delta x_i < \varepsilon \qquad \text{et} \qquad |g'(x_i) - g'(\xi_i)| < \varepsilon.$$

Soit $n \geqslant N_1$. Alors

$$\begin{aligned} |S_n - T_n| &\leqslant \sum_{k=0}^{n-1} \left| f(\xi_i)\left(g(x_{i+1}) - g(x_i) - g'(\xi_i)\Delta x_i\right)\right| \\ &\leqslant \sum_{k=0}^{n-1} |f(\xi_i)| \left|g(x_i + \Delta x_i) - g(x_i) - g'(x_i)(x_{i+1} - x_i)\right| + \sum_{k=0}^{n-1} |f(\xi_i)| \left|\Delta_{x_i}\left(g'(x_i) - g'(\xi_i)\right)\right| \end{aligned}$$

Or

$$\left|g(x_i+\Delta x_i)-g(x_i)-g'(\xi_i)(\Delta x_i)\right| = \underset{n\to+\infty}{o}(\Delta x_i),$$

donc il existe $N_2 \in \mathbb{N}$ tel que pour tout $n \geqslant N_2$ et pour tout $i \in [\![1,n]\!]$

$$\left|g(x_i+\Delta x_i)-g(x_i)-g'(\xi_i)(\Delta x_i)\right| < \varepsilon \Delta x_i.$$

Par hypothèse f est bornée sur $[a,b]$, il existe donc $M>0$ tel que pour tout $x \in [a,b]$ $|f(x)|<M$. Ainsi pour tout $n \geqslant \max(N_1,N_2)$

$$\begin{aligned}|S_n-T_n| &< 2\varepsilon \sum_{k=0}^{n-1} |f(\xi_i)|\Delta x_i \\ &< 2\varepsilon M \sum_{k=0}^{n-1} \Delta x_i,\end{aligned}$$

et

$$\sum_{k=0}^{n-1} \Delta x_i = b-a,$$

ainsi pour tout $n \geqslant \max(N_1,N_2)$

$$|S_n-T_n| < 2\varepsilon M(b-a),$$

donc

$$\lim_{n\to+\infty} S_n = \lim_{n\to+\infty} T_n.$$

□

Certaines propriétés propres à l'intégrale de Riemann se généralise très bien à l'intégrale de Stieltjes : intégration par partie, formule de la moyenne, linéarité.

Théorème 2 (intégration par parties)**.** *Soient $a,b \in \mathbb{R}$ tels que $a<b$, $f,g : [a,b] \longrightarrow \mathbb{R}$ deux foncions réelles bornées. On suppose que l'intégrale de Stieltjes de f par rapport à g existe. On suppose de plus que f (resp. g) est continue à droite en a (resp. à gauche en b). Alors l'intégrale de Stieltjes de g par rapport à f existe et on a :*

$$\int_a^b f(x)\mathrm{d}g(x) = \big[f(x)g(x)\big]_a^b - \int_a^b g(x)\mathrm{d}f(x).$$

Démonstration. Pour $n \in \mathbb{N}^*$ on note de nouveau.

$$S_n = \sum_{i=0}^{n-1} f(\xi_i)(g(x_{i+1})-g(x_i)).$$

Alors

$$S_n = f(\xi_{n-1})g(b) - f(\xi_{n-1})g(x_{n-1}) + f(\xi_0)g(x_1) - f(\xi_0)g(a) + \sum_{i=1}^{n-2} f(\xi_i)(g(x_{i+1})-g(x_i)),$$

puis après ré-indexation

$$S_n = f(\xi_{n-1})g(b) - f(\xi_0)g(a) - \sum_{k=0}^{n-1} g(x_i)(f(\xi_{i+1})-f(\xi_i)).$$

On obtient le résultat par passage à la limite. □

On énonce de plus deux résultats élémentaires :

Proposition 3 (formule de la moyenne)**.** *Soient $a, b \in \mathbb{R}$ tels que $a < b$, $f : [a, b] \longrightarrow \mathbb{R}$ une fonction continue et $g : [a, b] \longrightarrow \mathbb{R}$ une fonction monotone de classe $\mathcal{C}^1$. Alors il existe $c \in [a, b]$ tel que*

$$\int_a^b f(x)\mathrm{d}g(x) = f(c)(g(a) - g(b)).$$

Proposition 4. *Soient $a, b \in \mathbb{R}$ tels que $a < b$, $f, g, h : [a, b] \longrightarrow \mathbb{R}$ trois fonctions. On suppose que toutes les intégrales prochainement écrites existent. On a*

$$\int_a^b f\mathrm{d}(g + h) = \int_a^b f\mathrm{d}g + \int_a^b f\mathrm{d}h.$$

Le lemme suivant est un exemple d'utilisation de l'intégrale de Stieltjes qui nous servira dans la suite.

Lemme 5. *Soit $n \geqslant 2$. Alors*

$$\log(n!) = n\log(n) - n + \varepsilon_n \log(n),$$

avec $\varepsilon_n \in [0, 1]$.

Démonstration. Soit $n \geqslant 2$. La fonction log est continue sur $[1, n]$ est la fonction partie entière $\lfloor\cdot\rfloor$ est de classe $\mathcal{C}^1$ par morceaux. Ainsi l'intégrale de Stieltjes de log par rapport à $\lfloor\cdot\rfloor$ existe. On peut donc écrire

$$\log(n!) = \sum_{k=2}^{n} \log(k) = \int_1^n \log(t)\mathrm{d}\lfloor t\rfloor.$$

Par la proposition 4 il vient

$$\begin{aligned}\log(n!) - \int_1^n \log(t)\mathrm{d}t &= \int_1^n \log(t)\mathrm{d}\lfloor t\rfloor - \int_1^n \log(t)\mathrm{d}t\\ &= -\int_1^n \log(t)\mathrm{d}\{t\} \qquad \text{où } \{t\} := t - \lfloor t\rfloor \text{ désigne la partie fractionnaire,}\end{aligned}$$

puis en intégrant par parties

$$\begin{aligned}&= -\big[\log(t)\{t\}\big] + \int_1^n \{t\}\mathrm{d}\log(t)\\ &= \int_1^n \{t\}\mathrm{d}\log(t).\end{aligned}$$

Or log est une fonction croissante de classe $\mathcal{C}^1$, donc par la proposition 3 il existe une constante $c \in [1, n]$ telle que

$$\{c\}\log(n) = \int_1^n \{t\}\mathrm{d}\log(t),$$

on conclut en posant $\varepsilon_n = \{c\} \in [0, 1]$. □

2.2. Résultats de théorie des nombres

Certaines quantités se construisant sur l'ensemble des nombres premiers inférieurs à un réel x ont un comportement asymptotique relativement facile à trouver. Ces comportements asymptotiques nous aideront par la suite à déterminer le comportement en moyenne de ω. Nous commençons dans un premier temps par établir quelques résultats préliminaires.

Proposition 6. *Pour $n \geqslant 1$ on a*

$$\prod_{p \leqslant n} p \leqslant 4^n.$$

Démonstration. On raisonne par récurrence forte sur n. Les cas $n = 1$ et $n = 2$ sont clairs. Soit $n \geqslant 3$, on suppose le résultat vrai pour tout $k \leqslant n-1$. Montrons le résultat au rang n. On distingue deux cas.

(i) Si n est pair, alors n n'est pas premier (on rappelle que $n \geqslant 3$). Alors par hypothèse de récurrence
$$\prod_{p \leqslant n} p = \prod_{p \leqslant n-1} p \leqslant 4^{n-1} < 4^n.$$

(ii) Sinon on écrit $n = 2k+1$ avec $k \in \mathbb{N}^*$. $\binom{2k+1}{k} = \frac{(2k+1)!}{k!(k+1)!}$, donc pour tout nombre premier $p \in [\![k+2, 2k+1]\!]$, $p | \binom{2k+1}{k}$. Ainsi
$$\left(\prod_{k+1<p \leqslant 2k+1} p\right) \Bigg| \binom{2k+1}{k}.$$

Majorons ce coefficient binomial. On a
$$\begin{aligned} 2^{2k+1} &= (1+1)^{2k+1} \\ &= \sum_{i=0}^{2k+1} \binom{2k+1}{i} \\ &\geqslant \binom{2k+1}{k} + \binom{2k+1}{k+1} \\ &= 2\binom{2k+1}{k}, \end{aligned}$$

Finalement
$$\prod_{p \leqslant n} p = \prod_{p \leqslant k+1} p \prod_{k+1<p \leqslant 2k+1} p \leqslant 4^{k+1}\frac{1}{2}2^{2k+1} = 4^n.$$

□

Proposition 7. *Soient $n \in \mathbb{N}^*$ et p un nombre premier. Alors*
$$v_p(n!) = \sum_{k=1}^{+\infty} \left\lfloor \frac{n}{p^k} \right\rfloor,$$
où v_p désigne la valuation p-adique.

Remarque 2. La somme est en réalité finie.

Démonstration. C'est un simple calcul :
$$v_p(n!) = \sum_{k=1}^{n} v_p(k) = \sum_{k=1}^{n} \sum_{m=1}^{v_p(k)} 1 = \sum_{m=1}^{+\infty} \sum_{\substack{k \leqslant n \\ v_p(k) \geqslant m}} 1$$
et
$$\sum_{\substack{k \leqslant n \\ v_p(k) \geqslant m}} 1 = \left\lfloor \frac{n}{p^m} \right\rfloor,$$
car $\{k \leqslant n : v_p(k) \geqslant m\}$ est le nombre d'entier de $[\![1, n]\!]$ divisible par p^m. D'où le résultat. □

Corollaire 1. *Soit $n \geqslant 1$. Alors*
$$\frac{n}{p} - 1 < v_p(n!) \leqslant \frac{n}{p} - \frac{n}{p(p-1)}.$$

Démonstration. Pour tout $k \in \mathbb{N}^*$

$$\frac{n}{p^k} - 1 < \left\lfloor \frac{n}{p^k} \right\rfloor \leqslant \frac{n}{p^k}.$$

Alors

$$\sum_{k=1}^{+\infty} \left\lfloor \frac{n}{p^k} \right\rfloor \geqslant \left\lfloor \frac{n}{p} \right\rfloor > \frac{n}{p} - 1$$

et

$$\sum_{k=1}^{+\infty} \frac{n}{p^k} = \frac{n}{p-1} = \frac{n}{p} - \frac{n}{p(p-1)}.$$

□

Nous pouvons maintenant démontrer le premier théorème de Mertens, qui est un résultat central en théorie des nombres.

Théorème 8 (premier théorème de Mertens)**.** *Soit $x \geqslant 2$. On a*

$$\sum_{p \leqslant x} \frac{\log(p)}{p} = \log(x) + O(1).$$

Sur la figure suivante on a représenter la fonction log ainsi que la fonction $x \longmapsto \sum_{p \leqslant x} \log(p)/p$.

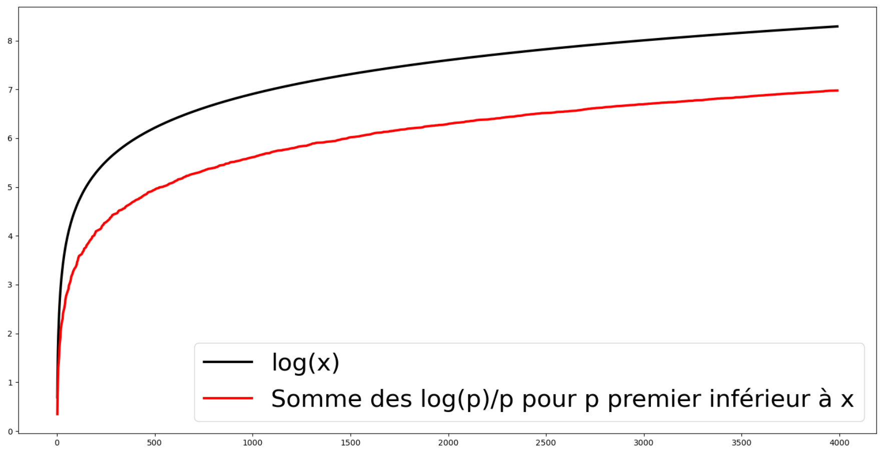

Figure 1 – Illustration graphique du premier théorème de Mertens.

Démonstration. Soit $x \geqslant 2$. On pose $n = \lfloor x \rfloor$. D'une part, par le lemme 5

$$\log(n!) = n \log(n) - n + 1 + \varepsilon_n \log(n)$$

avec $\varepsilon_n \in [0, 1]$.
D'autre part

$$n! = \prod_{p \leqslant n} p^{v_p(n!)},$$

donc

$$\log(n!) = \sum_{p \leqslant n} v_p(n!) \log(p).$$

On obtient grâce au corollaire 1 un encadrement de $\log(n!)$:

$$n\sum_{p\leqslant n}\frac{\log(p)}{p}-\sum_{p\leqslant n}\log(p)<\log(n!)<n\sum_{p\leqslant n}\frac{\log(p)}{p}+n\sum_{p\leqslant n}\frac{\log(p)}{p(p-1)}.$$

Par la proposition 6 on sait que la somme $\sum_{p\leqslant n}\log(p)$ est inférieur à $n\log(4)$. Ainsi

$$n\sum_{p\leqslant n}\frac{\log(p)}{p}-n\log(4)<n\log(n)-n+1+\log(n)\ \leqslant\ n\log(n)$$

de là

$$\sum_{p\leqslant x}\frac{\log(p)}{p}=\sum_{p\leqslant n}\frac{\log(p)}{p}<\log(n)+\log(4)\ \leqslant\ \log(x)+\log(4).$$

On a aussi

$$\begin{aligned}\sum_{p\leqslant n}\frac{\log(p)}{p(p-1)}&<\sum_{k=2}^{+\infty}\frac{\log(k)}{k(k-1)}\\&\leqslant\sum_{r=1}^{+\infty}\sum_{2^{r-1}<k\leqslant 2^r}\frac{r\log(2)}{k(k-1)}\\&=\sum_{r=1}^{+\infty}\frac{r\log(2)}{2^r}=\log(4).\end{aligned}$$

On en déduit que

$$n\sum_{p\leqslant n}\frac{\log(p)}{p}+n\log(4)>n\log(n)-n+1$$

et

$$\sum_{p\leqslant x}\frac{\log(p)}{p}>\log(n)+\frac{1}{n}-1-\log(4)\ \geqslant\ \log(x)-(1+\log(4)).$$

□

Remarque 3. On a de plus montrer que la constante implicite du $O(1)$ est inférieur en valeur absolue à $1+\log(4)$.

Le théorème suivant est une conséquence du premier théorème de Mertens.

Théorème 9. *Il existe une constante $C\in\mathbb{R}$ telle que pour $x\geqslant 2$ on a*

$$\sum_{p\leqslant x}\frac{1}{p}=\log(\log(x))+C+O\left(\frac{1}{\log(x)}\right).$$

Démonstration. Par le premier théorème de Mertens (théorème 8), on a pour $t\geqslant 2$:

$$R(t):=\sum_{p\leqslant t}\frac{\log(p)}{p}-\log(t)=O(1).$$

Soit $x\geqslant 2$. Alors

$$\sum_{p\leqslant x}\frac{1}{p}=\int_2^x\frac{1}{\log(t)}\mathrm{d}\left(\sum_{p\leqslant t}\frac{\log(p)}{p}\right).$$

Les fonctions $t\longmapsto\sum_{p\leqslant t}\frac{\log(p)}{p}$ et log sont de classe $\mathcal{C}^1$ par morceaux sur $\mathbb{R}$, la dernière intégrale est donc bien définie. On a donc

$$\sum_{p\leqslant x}\frac{1}{p}=\int_2^x\frac{\mathrm{d}t}{t\log(t)}+\int_2^x\frac{\mathrm{d}R(t)}{\log(t)}$$

En intégrant par parties on obtient

$$\sum_{p \leqslant x} \frac{1}{p} = \log(\log(x)) - \log(\log(2)) + \frac{R(x)}{\log(x)} - \frac{R(2^-)}{\log(2)} + \int_2^x \frac{R(t)}{t(\log(t))^2} \mathrm{d}t.$$

On pose $R = \sup_{t \geqslant 2} |R(t)|$. Alors

$$\begin{aligned} \left| \frac{R(x)}{\log(x)} - \int_x^{+\infty} \frac{R(t)}{t(\log(t))^2} \mathrm{d}t \right| &\leqslant \frac{R}{\log(x)} + \int_x^{+\infty} \frac{R}{t(\log(t))^2} \mathrm{d}t \\ &= \frac{2R}{\log(x)} \underset{x \to +\infty}{\longrightarrow} 0, \end{aligned}$$

ce qui achève la démonstration. □

2.3. Moyenne de ω

Nous pouvons désormais estimer aisément une moyenne de la fonction ω. En effet, pour $x \geqslant 0$ on écrit

$$\sum_{n \leqslant x} \omega(n) = \sum_{n \leqslant x} \sum_{p | n} 1 = \sum_{p \leqslant x} \sum_{\substack{k=1 \\ pk \leqslant x}}^{+\infty} 1,$$

il vient donc

$$\sum_{n \leqslant x} \omega(n) = \sum_{p \leqslant x} \left\lfloor \frac{x}{p} \right\rfloor = x \sum_{p \leqslant x} \frac{1}{p} + O(x).$$

Par le théorème 9, on obtient finalement

$$\boxed{\sum_{p \leqslant x} \omega(n) = x \log(\log(x)) + O(x).}$$

Ainsi, la moyenne de ω sur les entiers inférieurs ou égaux à un réel x est sensiblement égale à $\log(\log(x))$. Le théorème d'Erdös-Kac nous renseigne sur les fluctuations par rapport à cette moyenne.

2.4. Fonctions indicatrices des multiples de nombres premiers

Une idée pour simplifier l'étude de ω est de se rapprocher d'une somme de variables de Bernoulli indépendantes. On introduit dans ce but une famille de fonctions indicatrices comme suit.

Définition 2. Pour tout nombre premier p, on définit la fonction indicatrice des multiples de p :

$$\delta_p : \mathbb{N}^* \longrightarrow \{0, 1\},\ n \longmapsto \begin{cases} 1 & \text{si } p | n \\ 0 & \text{sinon.} \end{cases}$$

On a alors pour tout entier naturel $n \geqslant 2$

$$\omega(n) = \sum_{p \leqslant n} \delta_p(n).$$

On note $\mathbb{P}_n$ la probabilité uniforme sur $[\![1, n]\!]$ et $\mathbb{E}_n$ l'espérance associée. Il vient

$$\mathbb{P}_n(\delta_p = 1) = \frac{1}{n} \mathrm{Card}\ \{i \in [\![1, n]\!] : p | i\} = \frac{1}{n} \left\lfloor \frac{n}{p} \right\rfloor.$$

On remarque que

$$\mathbb{P}_n(\delta_p = 1) \underset{n\to+\infty}{\sim} \frac{1}{p},$$

ainsi, lorsque n est grand, la restriction de δ_p à $[\![1,n]\!]$ suit "presque" une loi de Bernoulli de paramètre $1/p$. De plus, si on se donne $p_1,\dots,p_r$ des nombres premiers deux à deux distincts, alors

$$\begin{aligned}\mathbb{P}_n(\delta_{p_1} = 1,\dots,\delta_{p_r} = 1) &= \frac{1}{n}\mathrm{Card}\,\{i \in [\![1,n]\!] : p_1|i,\dots,p_r|i\}\\ &= \frac{1}{n}\mathrm{Card}\,\{i \in [\![1,n]\!] : p_1\cdots p_r|i\}\\ &= \frac{1}{n}\left\lfloor \frac{n}{p_1\cdots p_r}\right\rfloor\end{aligned}$$

donc

$$\mathbb{P}_n(\delta_{p_1} = 1,\dots,\delta_{p_r} = 1) \underset{n\to+\infty}{\sim} \mathbb{P}_n(\delta_{p_1} = 1)\cdots\mathbb{P}_n(\delta_{p_r} = 1).$$

Lorsque n devient grand, les variables δ_p restreintes à $[\![1,n]\!]$ deviennent "presque indépendantes". Il est judicieux de considérer une suite $(X_p)_p$, indexée par les nombres premiers, de variables aléatoires de Bernoulli indépendantes de paramètre $1/p$. En notant

$$S_n := \sum_{p\leqslant n} X_p,$$

on définit une variable aléatoire qui a des chances de se comporter comme ω sur $[\![1,n]\!]$.

2.5. **Convolution de Dirichlet**

Dans cette section nous faisons quelques rappels de résultats classiques sur la convolution de Dirichlet et sur la fonction μ de Möbius qui nous serviront à démontrer le théorème d'Erdös-Kac par la méthode de Stein.

Définition 3. Soient f, g deux fonctions arithmétiques, ie définie sur $\mathbb{N}$ est à valeurs dans $\mathbb{C}$. On définit la convolution de Dirichlet de f et g par

$$\forall n \in \mathbb{N}^* \quad (f*g)(n) := \sum_{d|n} f(d)g\left(\frac{n}{d}\right).$$

On note :
(i) δ_1 la fonction indicatrice du singleton $\{1\}$;
(ii) $\mathbf{1}$ la fonction constante égale à 1 ;
(iii) $\mathbf{Id}$ la fonction identité.

L'ensemble des fonctions arithmétiques muni de la loi $*$ forme un anneau intègre, dont δ_1 est l'élément neutre.

Définition 4. On définit la fonction de Möbius μ par

$$\mu : \mathbb{N}^* \longrightarrow \{0,\pm1\},\ n \longmapsto \begin{cases}(-1)^{\omega(n)} & \text{si } n \text{ est un produit de nombres premiers distinct}\\ 0 & \text{sinon.}\end{cases}$$

Proposition 10. *La fonction de Möbius est l'inverse de la fonction* $\mathbf{1}$ *pour la convolution de Dirichlet, ie*

$$\mathbf{1} * \mu = \delta_1.$$

Démonstration. Soit $n \in \mathbb{N}^*$. Montrons que $(\mathbf{1}*\mu)(n) = \delta_1(n)$. On distingue les cas.

(*i*) Si $n = 1$, on a bien
$$(\mathbf{1} * \mu)(1) = \mathbf{1}(1)\mu(1) = 1 = \delta_1(1).$$

(*ii*) Si $n \geqslant 2$. On écrit
$$n = p_1^{\alpha_1} \cdots p_r^{\alpha_r},$$
avec $p_1, \ldots, p_r$ des nombres premiers deux à deux distincts et $\alpha_1, \ldots, \alpha_r$ des entiers. Montrons que $(\mathbf{1} * \mu)(n) = 0$. On a
$$(\mathbf{1} * \mu)(n) = \sum_{d|n} \mu(d).$$
Or par définition de la fonction μ, si d est divisible par un carré (différent de 1), alors $\mu(d) = 0$. On peut donc supposer sans perte de généralité que $n = p_1 \cdots p_r$. Un diviseur de n différent de 1 est donc de la forme
$$\prod_{i \in I} p_i,$$
avec I une partie de $[\![1, r]\!]$. Ainsi
$$(\mathbf{1} * \mu)(n) = 1 + (-1)^1 \binom{r}{1} + (-1)^2 \binom{r}{2} + \ldots + (-1)^r \binom{r}{r} = 0 = \delta_1(n)$$
par la formule du binôme.

□

Nous énonçons un lemme qui nous sera utile dans la suite. Pour $a, b \in \mathbb{N}$, on note (a, b) leur PGCD et on définit la fonction τ comme étant le nombre de diviseur d'un entier, ie
$$\tau(a) = \sum_{d|a} 1.$$

Lemme 11. *Soient* $1 \leqslant a \leqslant x$. *Alors*
$$\sum_{\substack{n \leqslant x \\ (a,n)=1}} 1 = x\frac{\varphi(a)}{a} + O(\tau(a)),$$
où φ *désigne la fonction indicatrice d'Euler.*

Démonstration. On a
$$\begin{aligned}
\sum_{\substack{n \leqslant x \\ (a,n)=1}} 1 &= \sum_{n \leqslant x} \delta_1((a, n)) \\
&= \sum_{n \leqslant x} \sum_{d|(n,a)} \mu(d) \\
&= \sum_{\substack{d \leqslant x \\ d|a}} \mu(d) \sum_{\substack{n \leqslant x \\ d|n}} 1 \\
&= \sum_{\substack{d \leqslant x \\ d|a}} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor \\
&= x \sum_{\substack{d \leqslant x \\ d|a}} \frac{\mu(d)}{d} + O(\tau(a)).
\end{aligned}$$

Il est connu que pour tout $m \in \mathbb{N}^*$

$$m = \sum_{d|m} \varphi(d),$$

autrement dit

$$\mathbf{Id} = \varphi * \mathbf{1}.$$

La formule d'inversion de Möbius donne donc

$$\varphi(a) = (\mathbf{Id} * \mu)(a) = \sum_{d|a} \frac{a}{d} \mu(d),$$

ce qui conclut la preuve. $\square$

3. Méthode des moments

Nous allons ici présenter une méthode permettant de montrer des convergences en loi sans passer par les fonctions caractéristiques. Nous commençons par quelques rappels et résultats préliminaires sur les convergences en loi.

3.1. Convergence en loi

Définition 5. Soient $(X_n)_n$ et X des variables aléatoires réelles. On note $\mathbb{P}_Y$ la loi d'une variable aléatoire réelle Y. On dit que $(X_n)_n$ converge en loi vers X et on note $X_n \xrightarrow{\mathcal{L}} X$, si l'un des quatre énoncés équivalents suivants est vérifié :

(i) pour toute fonction f continue et bornée de $\mathbb{R}$ dans $\mathbb{C}$,

$$E[f(X_n)] \underset{n\to+\infty}{\longrightarrow} E[f(X)];$$

(ii) pour tout $x \in \mathbb{R}$ tel que $\mathbb{P}_X(\{x\}) = 0$, on a

$$\mathbb{P}_{X_n}(]-\infty, x]) \underset{n\to+\infty}{\longrightarrow} \mathbb{P}_X(]-\infty, x]);$$

(iii) pour tout borélien B de $\mathbb{R}$ vérifiant $\mathbb{P}_X(\partial B) = 0$, on a

$$\mathbb{P}_{X_n}(B) \underset{n\to+\infty}{\longrightarrow} \mathbb{P}_X(B);$$

(iv) pour tout $t \in \mathbb{R}$, on a

$$\varphi_{X_n}(t) \underset{n\to+\infty}{\longrightarrow} \varphi_X(t).$$

Remarque 4. Le point (iv) est le théorème de Lévy.

La démonstration de l'équivalence des ces quatre énoncés se trouve dans la référence [2]. On définit aussi la convergence en probabilité de variables aléatoire, qui est une convergence plus forte que la convergence en loi.

Définition 6. Soient $(X_n)_n$ et X des variables aléatoires réelles. On dit que $(X_n)_n$ converge en probabilité vers X et on note $X_n \xrightarrow{\mathbb{P}}$ si

$$\forall \varepsilon > 0, \quad \mathbb{P}(|X_n - X| \geqslant \varepsilon) \underset{n\to+\infty}{\longrightarrow} 0.$$

Le lemme suivant va s'avérer fort utile dans la démonstration du théorème d'Erdös-Kac.

Lemme 12. *Soient* $(X_n)_n, (Y_n)_n, (Z_n)_n$ *des suites de variables aléatoires réelles et* X *une variable aléatoire réelle. On suppose que* $Y_n \xrightarrow{\mathbb{P}} 0$, $Z_n \xrightarrow{\mathbb{P}} 1$ *et que* $X_n \xrightarrow{\mathcal{L}} X$. *Alors :*

(i) $X_n + Y_n \xrightarrow{\mathcal{L}} X$,

(ii) $X_n Y_n \xrightarrow{\mathbb{P}} 0$,

(iii) $Z_n X_n + Y_n \xrightarrow{\mathcal{L}} X$.

Démonstration. On suppose que les variables aléatoires $(X_n)_n$, $(Y_n)_n$ et $(Z_n)_n$ sont définies sur des espaces probabilisés $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$ qui peuvent varier avec n.

(i) On va utiliser le théorème de Lévy. Soit $t \in \mathbb{R}$. Par l'inégalité triangulaire on a

$$|\varphi_{X_n+Y_n}(t) - \varphi_X(t)| \leqslant |\varphi_{X_n+Y_n}(t) - \varphi_{X_n}(t)| + |\varphi_{X_n}(t) - \varphi_X(t)|.$$

$(X_n)_n$ converge en loi vers X, donc par le théorème de Levy la quantité $|\varphi_{X_n}(t) - \varphi_X(t)|$ converge vers 0 lorsque n tend vers $+\infty$. Intéressons nous au premier terme. On a

$$|\varphi_{X_n+Y_n}(t) - \varphi_{X_n}(t)| = |\mathbb{E}[e^{itX_n}(e^{itY_n} - 1)]| \leqslant \mathbb{E}[|e^{itY_n} - 1|].$$

Soit $\varepsilon > 0$, alors

$$\begin{aligned}\mathbb{E}[|e^{itY_n} - 1|] &= \int_{|Y_n| \geqslant \varepsilon} |e^{itY_n} - 1| \mathrm{d}\mathbb{P}_n + \int_{|Y_n| < \varepsilon} |e^{itY_n} - 1| \mathrm{d}\mathbb{P}_n \\ &\leqslant 2 \int_{|Y_n| \geqslant \varepsilon} \mathrm{d}\mathbb{P}_n + 2 \int_{|Y_n| < \varepsilon} |\sin(Y_n t/2)| \mathrm{d}\mathbb{P}_n \\ &\leqslant 2\mathbb{P}_n(|Y_n| \geqslant \varepsilon) + \int_{|Y_n| < \varepsilon} |Y_n t| \mathrm{d}\mathbb{P}_n \\ &\leqslant 2\mathbb{P}_n(|Y_n| \geqslant \varepsilon) + \varepsilon |t|.\end{aligned}$$

Or $Y_n \xrightarrow{\mathbb{P}} 0$, donc $\limsup_n \mathbb{E}[|e^{itY_n} - 1|] \leqslant \varepsilon|t|$, ce qui permet de conclure.

(ii) Soit $\varepsilon > 0$. Pour tout $x > 0$ tel que $\mathbb{P}_X(\{\pm x\varepsilon\}) = 0$, on a

$$\mathbb{P}_n(|X_n| \geqslant x\varepsilon) \underset{n \to +\infty}{\longrightarrow} \mathbb{P}(|X| \geqslant x\varepsilon).$$

De plus

$$\begin{aligned}\mathbb{P}_n(|X_n Y_n| \geqslant \varepsilon) &= \mathbb{P}_n(||X_n Y_n| \geqslant \varepsilon \,|\, |Y_n| \geqslant 1/x) + \mathbb{P}_n(||X_n Y_n| \geqslant \varepsilon \,|\, |Y_n| < 1/x) \\ &\leqslant \mathbb{P}_n(|Y_n| \geqslant 1/x) + \mathbb{P}_n(|X_n| \geqslant x\varepsilon).\end{aligned}$$

Une variable aléatoire admet un nombre dénombrable d'atome, on peut donc supposer que x est judicieusement choisi, de tel sorte que $\mathbb{P}_X(\{\pm x\varepsilon\}) \neq 0$.
Ainsi

$$\limsup_n \mathbb{P}_n(|X_n Y_n| \geqslant \varepsilon) \leqslant \mathbb{P}(|X| \geqslant x\varepsilon),$$

en faisant tendre x vers $+\infty$ en évitant un nombre dénombrable de valeurs, on obtient finalement

$$\limsup_n \mathbb{P}_n(|X_n Y_n| \geqslant \varepsilon) = 0,$$

donc

$$X_n Y_n \xrightarrow{\mathbb{P}} 0.$$

(iii) Ce point ce déduit des deux précédant, en écrivant

$$Z_n X_n + Y_n = (Z_n - 1)X_n + X_n + Y_n$$

et en utilisant le fait que la somme de deux variables aléatoires convergeant en probabilité vers 0 converge elle aussi en probabilité vers 0.

□

Si $p \geqslant 1$ et si X est une variable aléatoire réelle admettant un moment d'ordre p alors on note

$$\|X\|_p := (\mathbb{E}[|X|^p])^{\frac{1}{p}}.$$

Théorème 13. *Soient $(X_n)_{n \geqslant 1}$ une suite de variables aléatoires réelles, X une variable aléatoire réelle, $\varepsilon > 0$ et $p \in \mathbb{N}^*$. On suppose que*

(i) $X_n \xrightarrow{\mathcal{L}} X$,

(ii) *il existe une constante $C \in \mathbb{R}$ telle que pour tout $n \in \mathbb{N}^*$, $\mathbb{E}[|X_n|^{p+\varepsilon}] \leqslant C$.*

Alors

$$\mathbb{E}[|X|^p] < +\infty \quad \textit{et} \quad \mathbb{E}[X_n^p] \underset{n \to +\infty}{\longrightarrow} \mathbb{E}[X^p].$$

Démonstration. Pour tout $A > 0$, on pose

$$\varphi_A : \mathbb{R} \longrightarrow \mathbb{R},\, x \longmapsto \begin{cases} A^p & \text{si } x \geqslant A \\ x^p & \text{si } |x| \leqslant A \\ (-A)^p & \text{si } x \leqslant -A. \end{cases}$$

La fonction $x \mapsto \varphi_A(|x|)$ est continue et bornée sur $\mathbb{R}$, donc par définition de la convergence en loi :

$$\mathbb{E}[\varphi_A(|X_n|)] \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[\varphi_A(X)].$$

On sait que les normes $\|\cdot\|_p$ croissent avec p (ce résultat découle de l'inégalité de Jensen), donc il existe une constante $C' > 0$ telle que pour tout $n \geqslant 1$

$$\mathbb{E}[|X_n|^p] = \|X_n\|_p^p \leqslant \|X_n\|_{p+\varepsilon}^p \leqslant C'.$$

De plus pour tout $n \geqslant 1$

$$\varphi_A(|X_n|) = \begin{cases} A^p & \text{si } |X_n| \geqslant A \\ |X_n|^p & \text{si } |X_n| \leqslant A, \end{cases}$$

donc

$$\varphi(|X_n|) \leqslant |X_n|^p,$$

et par conséquent

$$\mathbb{E}[\varphi_A(|X_n|)] \leqslant C',$$

d'où en passant à la limite quand n tend vers $+\infty$,

$$\mathbb{E}[\varphi_A(|X|)] \leqslant C'.$$

Le lemme de Fatou donne alors

$$\mathbb{E}[|X|^p] = \mathbb{E}[\liminf_{A\to+\infty} \varphi_A(|X|)] \leqslant \liminf_{A\to+\infty} \mathbb{E}[\varphi_A(|X|)] < +\infty,$$

ainsi X admet bien un moment d'ordre p.
Montrons maintenant que la suite $(\mathbb{E}[\varphi_A(X_n)])_{n\geqslant 1}$ converge uniformément en n vers $\mathbb{E}[X_n^p]$ lorsque A tend vers $+\infty$, ce qui justifiera l'égalité :

$$\lim_{A\to+\infty} \lim_{n\to+\infty} \mathbb{E}[\varphi_A(X_n)] = \lim_{n\to+\infty} \lim_{A\to+\infty} \mathbb{E}[\varphi_A(X_n)]. \qquad (1)$$

Pour cela on fixe $n \in \mathbb{N}^*$ et on écrit

$$\begin{aligned} |\mathbb{E}[X_n^p] - \mathbb{E}[\varphi_A(X_n)]| &\leqslant \mathbb{E}[|X_n|^p \mathbb{1}_{|X_n|>A}] \qquad (2) \\ &\leqslant A^{-\varepsilon}\mathbb{E}[|X_n|^{p+\varepsilon}] \\ &\leqslant CA^{-\varepsilon}, \end{aligned}$$

ce qui assure la convergence uniforme en n. L'inégalité (2) se vérifie facilement en discutant selon la parité de p. Par convergence dominée dans l'égalité (1) on obtient

$$\lim_{A\to+\infty} \mathbb{E}[\varphi_A(X)] = \lim_{n\to+\infty} \mathbb{E}[X_n^p].$$

En utilisant à nouveau la théorème de convergence dominée dans l'égalité, et en utilisant le fait que X admet un moment d'ordre p, on obtient finalement :

$$\mathbb{E}[X_n^p] \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[X^p].$$

□

3.2. Lois de probabilités caractérisées par leurs moments

Dans cette section, toutes les mesures de probabilités sont définies sur la tribu borélienne de $\mathbb{R}$, et on supposera également qu'elles admettent toujours des moments de tout ordre.

Définition 7. Une probabilité $\mathbb{P}$ est dite caractérisée par ses moments si toute probabilité $\mathbb{P}'$ vérifiant

$$\int_{\mathbb{R}} x^n \mathrm{d}\mathbb{P}(x) = \int_{\mathbb{R}} x^n \mathrm{d}\mathbb{P}'(x) \quad \text{pour tout } n \in \mathbb{N},$$

est égale à $\mathbb{P}$.

Le théorème suivant donne une condition suffisante pour qu'une probabilité soit caractérisée par ses moments.

Théorème 14. *Soit $(u_n)_{n\in\mathbb{N}}$ une suite de complexes telle que le rayon de convergence de la série entière $\sum \frac{u_n}{n!} z^n$ soit non nulle. Alors il existe au plus une probabilité $\mathbb{P}$ telle que*

$$\int_{\mathbb{R}} x^n \mathrm{d}\mathbb{P}(x) = u_n \quad \textit{pour tout } n \in \mathbb{N}.$$

Exemple 1. Illustrons ce théorème par un exemple avec la loi de Gauss, qui a pour densité

$$f : x \longmapsto \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

La loi de Gauss est caractérisée par ses moments. Vérifions le à l'aide du théorème. Pour tout $n \in \mathbb{N}$, on pose

$$u_n = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^n e^{-\frac{x^2}{2}} \mathrm{d}x.$$

Les moments d'ordre impair sont tous nuls. On fixe $k \in \mathbb{N}^*$, calculons u_{2k}.

$$\begin{aligned}
u_{2k} &= \sqrt{\frac{2}{\pi}} \int_0^{+\infty} x^{2k} e^{-\frac{x^2}{2}} \mathrm{d}x \\
&= \frac{2^k}{\sqrt{\pi}} \int_0^{+\infty} t^{k-\frac{1}{2}} e^{-t} \mathrm{d}t \quad \text{en posant } t = x^2/2 \\
&= \frac{2^k}{\sqrt{\pi}} \Gamma\left(k + \frac{1}{2}\right) \\
&= \frac{2^k}{\sqrt{\pi}} \left(k - \frac{1}{2}\right)\left(k - \frac{3}{2}\right) \cdots \frac{1}{2} \Gamma\left(\frac{1}{2}\right) \\
&= \frac{(2k!)}{2^k k!}.
\end{aligned}$$

Ainsi le rayon de convergence de la série $\sum \frac{u_n}{n!} z^n$ est infini, ce qui conclut.

Pour démontrer le théorème, nous allons d'abord établir le lemme suivant :

Lemme 15. *En reprenant les notations du théorème précédent et en notant $\mathbb{P}$ une telle probabilité, si on pose pour tout $n \in \mathbb{N}$*

$$v_n = \int_{\mathbb{R}} |x|^n \mathrm{d}\mathbb{P}(x),$$

alors les séries entières $\sum \frac{u_n}{n!} z^n$ et $\sum \frac{v_n}{n!} z^n$ ont le même rayon de convergence.

Démonstration. On note R le rayon de convergence de la série entière $\sum \frac{u_n}{n!} z^n$. Soit $n \in \mathbb{N}$. On a déjà $|u_n| \leqslant v_n$, et $u_{2n} = v_{2n}$. De plus

$$v_{2n+1} = \int_{\mathbb{R}} |x|^{2n+1} \mathrm{d}\mathbb{P}(x) \leqslant \int_{\mathbb{R}} \left(1 + x^{2n+2}\right) \mathrm{d}\mathbb{P}(x) = 1 + u_{2n+2},$$

d'où

$$\frac{v_{2n+1}}{(2n+1)!} \leqslant \frac{1}{(2n+1)!} + (2n+2)\frac{u_{2n+2}}{(2n+2)!}.$$

Soient $r \in]0, R[$ et $s \in]r, R[$. Il existe alors une constante $C > 0$ telle que pour tout $n \geqslant 0$

$$\left| \frac{u_{2n+2}}{(2n+2)!} s^{2n+1} \right| \leqslant C \quad \text{et} \quad (2n+2) r^{2n+1} \leqslant C s^{2n+1}.$$

Ainsi

$$\left| (2n+2) \frac{u_{2n+2}}{(2n+2)!} r^{2n+1} \right| \leqslant C^2,$$

la suite $\left(\frac{v_{2n+1}}{(2n+1)!} r^{2n+1} \right)_{n \in \mathbb{N}}$ est donc bornée, ce qui achève la preuve. $\square$

Passons désormais à la démonstration du théorème.

Démonstration. On rappelle que R désigne le rayon de convergence de la série $\sum \frac{u_n}{n!} z^n$. On note φ la fonction caractéristique de $\mathbb{P}$, définie par

$$\varphi(t) = \int_{\mathbb{R}} e^{itx} \mathrm{d}\mathbb{P}(x), \text{ pour } t \in \mathbb{R}.$$

On fixe $t, h \in \mathbb{R}$ tels que $|h| < R$. On a

$$\varphi(h+t) = \int_{\mathbb{R}} e^{i(t+h)x} \mathrm{d}\mathbb{P}(x) = \int_{\mathbb{R}} e^{itx} \sum_{n=0}^{+\infty} \frac{(ihx)^n}{n!} \mathrm{d}\mathbb{P}(x),$$

or on peut intervertir la somme et l'intégrale car d'après le lemme 15 :

$$\sum_{n=0}^{+\infty} \int_{\mathbb{R}} \left| e^{itx} \frac{(ihx)^n}{n!} \right| \mathrm{d}\mathbb{P}(x) = \sum_{n=0}^{+\infty} \frac{v_n}{n!} |h|^n < +\infty$$

(car $|h| < R$). Ainsi

$$\varphi(t+h) = \sum_{n=0}^{+\infty} \frac{(ih)^n}{n!} \int_{\mathbb{R}} x^n e^{itx} \mathrm{d}\mathbb{P}(x),$$

la fonction φ est donc analytique sur $\mathbb{R}$. En $t = 0$ on obtient

$$\varphi(h) = \sum_{n=0}^{+\infty} \frac{u_n}{n!} (ih)^n.$$

Deux probabilités ayant pour moments les éléments de la suite $(u_n)_{n \in \mathbb{N}}$ auront des fonctions caractéristiques analytiques sur $\mathbb{R}$ coïncident dans un voisinage de 0, elles sont donc égales. Sachant qu'une fonction caractéristique caractérise la loi, la démonstration est terminée. $\square$

3.3. Topologie faible $*$ et théorème fondamental

On note $\mathcal{C}_0$ l'espace des fonctions continues de $\mathbb{R}$ dans $\mathbb{C}$ tendant vers 0 en $\pm\infty$, et $\mathcal{C}_0^*$ l'espace dual des formes linéaires associées. On admet le théorème suivant, dû à F. Riesz (voir [10] p.160 pour la démonstration).

Théorème 16. *Pour toute forme linéaire $\Phi \in \mathcal{C}_0^*$, il existe une unique mesure borélienne μ sur $\mathbb{R}$ telle que*

$$\Phi(f) = \int_{\mathbb{R}} f(x) \mathrm{d}\mu(x), \text{ pour tout } f \in \mathcal{C}_0.$$

De plus, la forme linéaire Φ est positive si et seulement si la mesure μ l'est, et dans ce cas

$$\|\Phi\| = \mu(\mathbb{R}),$$

où la norme $\|\cdot\|$ désigne la norme sur $\mathcal{C}_0^*$ subordonnée à la norme infinie.

Grâce à ce théorème, tout mesure de probabilité sur $\mathbb{R}$ peut être vue comme un élément de la boule unité de $(\mathcal{C}_0^*, \|\cdot\|)$. Nous allons maintenant introduire la notion de topologie faible $*$. Au lieu de munir l'espace $\mathcal{C}_0^*$ de la topologie définie par la norme $\|\cdot\|$, nous utiliserons la topologie faible $*$. Voici une définition générale de cette dernière.

Définition 8. Soit X un $\mathbb{R}$-espace de Banach. La topologie faible $*$ sur X^* est la topologie la moins fine telle que toutes applications

$$\left| \begin{array}{ccc} X^* & \longrightarrow & \mathbb{R} \\ f & \longmapsto & \langle f, x \rangle \end{array} \right.$$

avec $x \in X$ soient continues.

Pour $X = \mathcal{C}_0$ et $\varepsilon > 0$, une base de voisinages de 0 est constituée des ensembles

$$\{\Phi \in \mathcal{C}_0^* : |\Phi(f_i)| \text{ pour } 1 \leqslant i \leqslant p\}$$

où p parcourt $\mathbb{N}^*$ et les fonctions $f_1, \ldots, f_p$ appartiennent à $\mathcal{C}_0$. Ainsi, si $(\Phi)_{n \geqslant 1}$ est une suite de $\mathcal{C}_0^*$ et si $\Phi \in \mathcal{C}_0$, alors

$$\Phi_n \overset{*}{\rightharpoonup} \Phi \Longleftrightarrow \forall\, f \in \mathcal{C}_0,\, \Phi_n(f) \longrightarrow \Phi(f)$$

où $\overset{*}{\rightharpoonup}$ désigne la convergence pour la topologie faible $*$.
Énonçons deux propriétés justifiant son intérêt. Pour les démonstrations, voir [3].

Proposition 17. *On note B la boule unité fermée de $(\mathcal{C}_0^*, \|\cdot\|)$.*

(i) *La topologie induite sur B par la topologie faible $*$ est métrisable.*

(ii) *B muni de la topologie induite par la topologie faible $*$ est un espace compact.*

Remarque 5. Le point (i) est vrai car $\mathcal{C}_0^*$ est séparable.
Voici deux conséquences de ces propriétés.
• Soit $(\mathbb{P}_n)_{n \geqslant 1}$ une suite de mesure de probabilité. Alors par le théorème de représentation de Riesz il existe une unique suite $(\Phi)_{n \geqslant 1}$ de $\mathcal{C}_0^*$ telles que pour tout $n \in \mathbb{N}$

$$\Phi_n(f) = \int_{\mathbb{R}} f(x) \mathrm{d}\mathbb{P}_n(x), \text{ pour tout } f \in \mathcal{C}_0.$$

Or les mesures de probabilité $\mathbb{P}_n$ sont positives, donc on a pour tout entier n non nul

$$\|\Phi_n\| = \mathbb{P}(\mathbb{R}) = 1,$$

donc les éléments de la suite $(\Phi)_{n \geqslant 1}$ appartiennent à la boule B. Par compacité, il existe une extraction $\psi : \mathbb{N} \longrightarrow \mathbb{N}$ et une forme linéaire $\Phi \in \mathcal{C}_0^*$ telles que

$$\Phi_{\psi(n)} \overset{*}{\rightharpoonup} \Phi.$$

Toujours par le théorème de Riesz, il existe donc une mesure positive $\mathbb{P}$ telle que

$$\int_{\mathbb{R}} f \mathrm{d}\mathbb{P}_{\psi(n)} \underset{n \to +\infty}{\longrightarrow} \int_{\mathbb{R}} f \mathrm{d}\mathbb{P} \text{ pour toute } f \in \mathcal{C}_0.$$

La mesure $\mathbb{P}$ n'est en général pas une probabilité, on a seulement $\mathbb{P}(\mathbb{R}) \leqslant 1$. C'est une probabilité si la suite $(\mathbb{P}_n)_{n \geqslant 1}$ est dite tendue, ie si pour tout $\varepsilon > 0$, il existe $A > 0$ tel que pour tout $n \geqslant 1$

$$\mathbb{P}_n([-A, A]) \geqslant 1 - \varepsilon.$$

• Soient $(\mathbb{P}_n)_{n \geqslant 1}$ une suite de probabilité et $\mathbb{P}$ une probabilité telles que pour toute extraction ψ et pour toute fonction $f \in \mathcal{C}_0$ on a

$$\int_{\mathbb{R}} f \mathrm{d}\mathbb{P}_{\psi(n)} \underset{n \to +\infty}{\longrightarrow} \int_{\mathbb{R}} f \mathrm{d}\mathbb{P},$$

alors

$$\int_{\mathbb{R}} f \mathrm{d}\mathbb{P}_n \underset{n\to+\infty}{\longrightarrow} \int_{\mathbb{R}} f \mathrm{d}\mathbb{P}.$$

En effet dans un espace métrique compact une suite converge si et seulement si elle admet une unique valeur d'adhérence.

Nous pouvons maintenant énoncer et démontrer le théorème central de la méthode des moments.

Théorème 18. *Soit $(X_n)_{n\geqslant 1}$ et X des variables aléatoires réelles admettant des moments de tout ordre. On suppose que :*
(i) la loi de X est caractérisée par ses moments,
(ii) pour tout entier $k \in \mathbb{N}^*$, $\mathbb{E}[X_n^k] \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[X^k]$.

Alors $X_n \xrightarrow{\mathcal{L}} X$.

Démonstration. Pour tout $n \in \mathbb{N}^*$, on note $\mathbb{P}_n$ la loi de la variable aléatoire X_n et $\mathbb{P}$ la loi de X. Par ce qui précède, il existe une mesure positive $\mathbb{P}'$ et une extraction ψ telles pour toute $f \in \mathcal{C}_0$

$$\int_{\mathbb{R}} f \mathrm{d}\mathbb{P}_{\psi(n)} \underset{n\to+\infty}{\longrightarrow} \int_{\mathbb{R}} f \mathrm{d}\mathbb{P}'.$$

Montrons que $\mathbb{P}'$ est en réalité la loi de X. Premièrement, $\mathbb{P}'$ est une probabilité car la suite $(\mathbb{P}_n)_{n\geqslant 1}$ est tendu. En effet, pour $A > 0$, par l'inégalité de Markov on a pour tout $n \in \mathbb{N}^*$

$$\mathbb{P}(|X_n| \leqslant A) \geqslant 1 - \frac{\mathbb{E}[X_n^2]}{A^2} \geqslant 1 - \frac{C}{A^2},$$

où $C > 0$ est une constante (la suite $(\mathbb{E}[X_n^2])_{n\geqslant 1}$ est bornée).
Soit X' une variable aléatoire de loi $\mathbb{P}'$. On se donne un entier $k \geqslant 1$, un entier pair $N > p$ et on note $\varepsilon = N - p > 0$. Par hypothèse, la suite $(\mathbb{E}[|X_n|^{p+\varepsilon}])_{n\geqslant 1}$ est convergente, elle est donc bornée. De plus $X_{\psi(n)} \xrightarrow{\mathcal{L}} X'$, donc d'après le théorème 13

$$\mathbb{E}[|X'^p|] < +\infty \quad \text{et} \quad \mathbb{E}[X_{\psi(n)}^p] \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[X'^p].$$

Par unicité de la limite $\mathbb{E}[X^p] = \mathbb{E}[X'^p]$, donc comme X est caractérisée par ses moments $X = X'$. En vertu du deuxième point précédent on obtient finalement

$$X_n \xrightarrow{\mathcal{L}} X.$$

□

3.4. Un théorème central limite non identiquement distribué

Nous allons énoncé dans cette section un théorème central limite pour des variables aléatoires indépendantes mais non identiquement distribuées. Nous le démontrerons plus tard par la méthode de Stein.

Théorème 19. *Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes, centrés, et uniformément bornées en valeur absolue par une constante C. Pour $n \in \mathbb{N}^*$, on note σ_n^2 la variance de X_n et on pose :*

$$S_n = \sum_{k=1}^{n} X_k,\ s_n = \left(\sum_{k=1}^{n} \sigma_k^2\right)^{1/2} \text{ et } S_n^* = \frac{S_n}{s_n}.$$

Si $\lim\limits_{n\to+\infty} s_n = +\infty$, alors

$$S_n^* \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Remarque 6. Les variables aléatoires X_k sont toutes bornées, elles admettent donc un moment à tout ordre. Il est donc correct d'introduire la suite $(\sigma_k^2)_{k\in\mathbb{N}^*}$ des variances.

3.5. Démonstration du théorème d'Erdös-Kac

Nous pouvons maintenant démontrer le théorème d'Erdös-Kac. Rappelons d'abord son énoncé.

Théorème 20 (Erdös-Kac). *Pour tout réel x,*

$$\frac{1}{n}\text{Card}\left\{i \in [\![1,n]\!] \Big/ \frac{\omega(i)-\log(\log(i))}{\log(\log(i))^{1/2}} \leqslant x\right\} \underset{n\to+\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\mathrm{d}t. \qquad (1)$$

Nous établirons ce théorème sous la forme plus maniable suivante :

$$\mathbb{P}_n\left(\left\{i \in [\![1,n]\!] \Big/ \frac{\omega(i)-\log(\log(n))}{\log(\log(n))^{1/2}} \leqslant x\right\}\right) \underset{n\to+\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\mathrm{d}t. \qquad (2)$$

où $x \in \mathbb{R}$. Montrons que (2) implique (1). Pour $n \geqslant 3$ on pose $\lambda_n = \log(\log(n))$. Alors pour $i \in [\![1,n]\!]$

$$\frac{\omega(i)-\lambda_n}{\sqrt{\lambda_i}} = \frac{\sqrt{\lambda_n}}{\sqrt{\lambda_i}}\left(\frac{\omega(i)-\lambda_i}{\sqrt{\lambda_n}} + \frac{\lambda_n-\lambda_i}{\sqrt{\lambda_n}}\right) =: U_n(i)(X_n(i)+Y_n(i)).$$

Il nous faut montrer que $U_n \xrightarrow{\mathbb{P}} 1$ et $Y_n \xrightarrow{\mathbb{P}} 0$. On se donne $\varepsilon > 0$. On a

$$\mathbb{P}_n(|U_n-1|) \geqslant \varepsilon) = \mathbb{P}_n(\sqrt{\lambda_n}-\sqrt{\lambda_i} \geqslant \varepsilon\sqrt{\lambda_i}).$$

Or si $\sqrt{n} \leqslant i \leqslant n$

$$\sqrt{\lambda_n}-\sqrt{\lambda_i} = \frac{\lambda_n-\lambda_i}{\sqrt{\lambda_n}+\sqrt{\lambda_i}} \leqslant \frac{\log(2)}{\sqrt{\lambda_n}},$$

et la quantité $\frac{\log(2)}{\sqrt{\lambda_n}}$ est inférieur à $\varepsilon\sqrt{\lambda_i}$ pour n suffisamment grand. Ainsi pour n assez grand

$$\mathbb{P}_n(|U_n-1| \geqslant \varepsilon) \leqslant \mathbb{P}_n([\![1,\lfloor\sqrt{n}\rfloor+1]\!]) = \frac{\lfloor\sqrt{n}\rfloor}{n} \underset{n\to+\infty}{\longrightarrow} 0.$$

On raisonne de manière similaire pour montrer que $\lim\limits_{n\to+\infty}\mathbb{P}_n(|Y_n| > \varepsilon) = 0$.

Nous pouvons maintenant nous concentrer sur la démonstration du point (2). La preuve se déroule en trois étapes. Nous introduisons préalablement quelques notations. Soit $n \in \mathbb{N}^*$. On pose :

$$S_n := \sum_{p\leqslant\alpha_n} X_p$$

où $(\alpha_n)_n$ est une suite divergeant vers $+\infty$ et $(X_p)_p$ une suite de variable de Bernoulli indépendantes indexée par les nombres premiers et telle que pour tout nombre premier p

$$\mathbb{P}(X_p = 1) = \frac{1}{p}.$$

On pose de plus

$$\omega_n = \sum_{p\leqslant\alpha_n}\delta_p,$$

$$e_n := \mathbb{E}[S_n] = \sum_{p\leqslant\alpha_n}\frac{1}{p} \text{ et } \sigma_n^2 := \mathrm{Var}\,(S_n) = \sum_{p\leqslant\alpha_n}\frac{1}{p}\left(1-\frac{1}{p}\right).$$

Nous avons vu précédemment que les variables aléatoires δ_p restreintes à l'ensemble $[\![1,n]\!]$ suivent "presque" les lois de Bernoulli de paramètre $1/p$ lorsque n devient grand. L'idée générale de la preuve est "d'imiter" la somme $\sum_{p\leqslant n}\delta_p$ (égale à ω sur $[\![1,n]\!]$) par la variable S_n.

Étape 1

Nous allons montrer que si la suite $(\alpha_n)_n$ est judicieusement choisie, alors il suffit de montrer que pour $x \in \mathbb{R}$

$$\mathbb{P}_n\left(\frac{\omega_n - e_n}{\sigma_n} \leqslant x\right) \underset{n\to+\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\mathrm{d}t.$$

On note d'abord que

$$\mathbb{E}_n\left[\frac{\omega - \omega_n}{\sigma_n}\right] = \frac{1}{\sigma_n}\sum_{\alpha_n < p \leqslant n}\mathbb{E}_n[\delta_p] = \frac{1}{\sigma_n}\sum_{\alpha_n < p \leqslant n}\frac{1}{n}\left\lfloor\frac{n}{p}\right\rfloor \leqslant \frac{1}{\sigma_n}\sum_{\alpha_n < p \leqslant n}\frac{1}{p}.$$

On choisit $(\alpha_n)_n$ telle que

$$\sum_{\alpha_n < p \leqslant n}\frac{1}{p} = o\left(\log(\log(n))^{1/2}\right),$$

on souhaite donc qu'elle ne tende pas trop lentement vers $+\infty$.
On a alors

$$\sigma_n^2 = \sum_{p \leqslant \alpha_n}\frac{1}{p} + O(1) = \sum_{p \leqslant n}\frac{1}{p} + o\left(\log(\log(n))^{1/2}\right),$$

donc par le théorème 9

$$\sigma_n \sim \log(\log(n))^{1/2}.$$

Ainsi

$$\mathbb{E}_n\left[\frac{\omega - \omega_n}{\sigma_n}\right] \underset{n\to+\infty}{\longrightarrow} 0.$$

Soit $\varepsilon > 0$, alors par l'inégalité de Markov

$$\mathbb{P}_n\left(\left|\frac{\omega - \omega_n}{\sigma_n}\right| \geqslant \varepsilon\right) \leqslant \frac{1}{\varepsilon}\mathbb{E}_n\left[\frac{\omega - \omega_n}{\sigma_n}\right],$$

donc

$$\frac{\omega - \omega_n}{\sigma_n} \xrightarrow{\mathbb{P}} 0.$$

Or pour $n \in \mathbb{N}^*$ on peut écrire

$$\frac{\omega - e_n}{\sigma_n} = \frac{\omega - \omega_n}{\sigma_n} + \frac{\omega_n - e_n}{\sigma_n},$$

donc par le lemme 12

$$\frac{\omega - e_n}{\sigma_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \iff \frac{\omega_n - e_n}{\sigma_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

On écrit ensuite

$$\frac{\omega - \log(\log(n))}{\log(\log(n)^{1/2}} = \frac{\sigma_n}{\log(\log(n))^{1/2}}\left(\frac{\omega - e_n}{\sigma_n} + \frac{e_n - \log(\log(n))}{\sigma_n}\right),$$

en remarquant que

$$e_n - \log(\log(n)) = \sum_{p \leqslant n}\frac{1}{p} - \log(\log(n)) - \sum_{\alpha_n < p \leqslant n}\frac{1}{p} = o\left(\log(\log(n))^{1/2}\right) = o(\sigma_n),$$

on utilise à nouveau le lemme 12 pour conclure que

$$\frac{\omega - \log(\log(n))}{\log(\log(n))^{1/2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \iff \frac{\omega_n - e_n}{\sigma_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Étape 2

L'étape 2 consiste à montrer que pour tout entier $r \geqslant 1$

$$\mathbb{E}\left[\left(\frac{S_n - e_n}{\sigma_n}\right)^r\right] - \mathbb{E}_n\left[\left(\frac{\omega_n - e_n}{\sigma_n}\right)^r\right] \underset{n\to+\infty}{\longrightarrow} 0.$$

Soit n, r des entiers supérieurs ou égaux à 1. Nous allons dans un premier temps majorer la quantité $|\mathbb{E}[S_n^r] - \mathbb{E}_n[w_n]|$. On note $q_1, \dots, q_s$ les nombres premiers inférieurs ou égaux à α_n.
La formule du multinôme nous donne

$$S_n^r = (X_{q_1} + \cdots X_{q_s})^r = \sum_{\beta_1+\cdots\beta_s=r} \frac{r!}{\beta_1! \cdots \beta_s!} X_{q_1}^{\beta_1} \cdots X_{q_s}^{\beta_s},$$

nous allons l'écrire avec des exposants β_i strictement positifs :

$$S_n^r = \sum_{k=1}^{r} \sum_{\substack{\beta_1+\cdots+\beta_k=r \\ \forall\, i\, \beta_i>0}} \frac{r!}{\beta_1! \cdots \beta_k!} \sum_{p_1<\cdots<p_k \leqslant \alpha_n} X_{q_1}^{\beta_1} \cdots X_{q_k}^{\beta_k},$$

où les p_i sont des nombres premiers. Les variables aléatoires X_p sont des variables de Bernoulli indépendantes donc par le lemme des coalitions

$$\mathbb{E}\left[S_n^r\right] = \sum_{k=1}^{r} \sum_{\substack{\beta_1+\cdots+\beta_k=r \\ \forall\, i\, \beta_i>0}} \frac{r!}{\beta_1! \cdots \beta_k!} \sum_{p_1<\cdots<p_k \leqslant \alpha_n} \mathbb{E}[X_{q_1}^{\beta_1}] \cdots \mathbb{E}[X_{q_k}^{\beta_k}],$$

or pour tout k, $\beta_k > 0$ donc

$$\mathbb{E}[X_{p_k}^{\beta_k}] = \mathbb{E}[X_{p_k}] = \frac{1}{p_k},$$

on en déduit que

$$\mathbb{E}[S_n^r] = \sum_{k=1}^{r} \sum_{\substack{\beta_1+\cdots+\beta_k=r \\ \forall\, i\, \beta_i>0}} \frac{r!}{\beta_1! \cdots \beta_k!} \sum_{p_1<\cdots<p_k \leqslant \alpha_n} \frac{1}{p_1 \cdots p_t}.$$

De la même façon

$$\begin{aligned}
\mathbb{E}_n[\omega_n^r] &= \sum_{k=1}^{r} \sum_{\substack{\beta_1+\cdots+\beta_k=r \\ \forall\, i\, \beta_i>0}} \frac{r!}{\beta_1! \cdots \beta_k!} \sum_{p_1<\cdots<p_k \leqslant \alpha_n} \mathbb{E}[\delta_{p_1} \cdots \delta_{p_k}] \\
&= \sum_{k=1}^{r} \sum_{\substack{\beta_1+\cdots+\beta_k=r \\ \forall\, i\, \beta_i>0}} \frac{r!}{\beta_1! \cdots \beta_k!} \sum_{p_1<\cdots<p_k \leqslant \alpha_n} \frac{1}{n} \left\lfloor \frac{n}{\delta_{p_1} \cdots \delta_{p_k}} \right\rfloor.
\end{aligned}$$

Comme

$$\left| \frac{1}{p_1 \cdots p_k} - \frac{1}{n} \left\lfloor \frac{n}{p_1 \cdots p_k} \right\rfloor \right| \leqslant \frac{1}{n}$$

on a

$$|\mathbb{E}[S_n^r] - \mathbb{E}[\omega_n^r]| \leqslant \frac{1}{n} \sum_{k=1}^{r} \sum_{\substack{\beta_1+\cdots+\beta_k=r \\ \forall\, i\, \beta_i>0}} \frac{r!}{\beta_1! \cdots \beta_k!} \sum_{p_1<\cdots<p_k \leqslant \alpha_n} 1 = \frac{1}{n} \left(\sum_{p \leqslant \alpha_n} 1 \right)^r$$

d'où

$$|\mathbb{E}[S_n^r] - \mathbb{E}[\omega_n^r]| \leqslant \frac{\alpha_n^r}{n}.$$

Le binôme de Newton donne alors

$$\mathbb{E}\left[(S_n-e_n)^r-(w_n-e_n)^r\right]| \leqslant \left|\sum_{k=0}^{r}\binom{r}{k}\left(\mathbb{E}[S_n^k]-\mathbb{E}[\omega_n^k]\right)(-e_n)^{r-k}\right| \leqslant \sum_{k=0}^{r}\binom{r}{k}\frac{\alpha_n^r}{n}e_n^{r-k},$$

puisque $e_n \leqslant \alpha_n$ on obtient finalement

$$\mathbb{E}\left[(S_n-e_n)^r-(w_n-e_n)^r\right]| \leqslant \frac{(2\alpha_n)^r}{n}.$$

Si on impose à la suite $(\alpha_n)_n$ la condition supplémentaire

$$\alpha_n = o(n^{\varepsilon}) \quad \text{pour tout } \varepsilon > 0,$$

alors $\alpha_n^r/n \to 0$ et donc

$$\mathbb{E}\left[\left(\frac{S_n-e_n}{\sigma_n}\right)^r\right]-\mathbb{E}_n\left[\left(\frac{\omega_n-e_n}{\sigma_n}\right)^r\right] \underset{n\to+\infty}{\longrightarrow} 0.$$

Cette dernière condition exprime le fait que la suite $(\alpha_n)_n$ ne tend pas trop vite vers $+\infty$. Les deux conditions sur la suite sont simultanément réalisables, on peut par exemple définir $(\alpha_n)_n$ par

$$\alpha_n = n^{\varepsilon_n} \quad \text{pour tout } n \in \mathbb{N}^*.$$

avec $\log(1/\varepsilon_n) = \log(\log(n))^{1/3}$. En effet on aurait alors

$$\sum_{\alpha_n < p \leqslant n}\frac{1}{p} = \log\left(\frac{\log(n)}{\log(\alpha_n)}\right)+O(1) = \log\left(\frac{1}{\varepsilon_n}\right)+O(1).$$

Étape 3

Nous allons finalement montrer que tout entier $r \geqslant 1$:

$$\mathbb{E}\left[\left(\frac{S_n-e_n}{\sigma_n}\right)^r\right] \underset{n\to+\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} x^r e^{-\frac{x^2}{2}}\mathrm{d}x =: \mu_r,$$

où μ_r est le r-ième moment moment de la loi normale. La loi normale étant caractérisée par ses moments, le théorème 18 affirmera que

$$\frac{\omega_n-e_n}{\sigma_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1),$$

ce qui conclura la démonstration du théorème d'Erdös-Kac.
D'après le théorème central limite 19,

$$\frac{S_n-e_n}{\sigma_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Montrons que

$$\sup_{n\geqslant 1}\mathbb{E}\left[\left(\frac{S_n-e_n}{\sigma_n}\right)^r\right] < +\infty,$$

la convergence des moments de $\frac{S_n-e_n}{\sigma_n}$ vers ceux de la loi normale serait alors assurée par le théorème 13.
Pour p premier on pose $Y_p = X_p - \frac{1}{p}$. On définit ainsi une suite de variables aléatoires centrées et indépendantes. On reprend les notations de l'étape 2. On a

$$\mathbb{E}[(S_n-e_n)^r] = \sum_{k=1}^{k}\sum_{\substack{\beta_1+\cdots+\beta_k=r \\ \forall i\,\beta_i>0}}\frac{r!}{\beta_1!\cdots\beta_k!}\sum_{p_1<\cdots<p_k\leqslant\alpha_n}\mathbb{E}[Y_{p_i}^{\beta_1}]\cdots\mathbb{E}[Y_{p_t}^{\beta_t}].$$

Pour tout nombre premier p, $|Y_p| \leqslant 1$, donc pour tout $\beta \geqslant 2$, $\mathbb{E}[Y_p^\beta] \leqslant \mathbb{E}[Y_p^2]$ et $\mathbb{E}[Y_p] = 0$. Il vient

$$\begin{aligned}
\mathbb{E}[(S_n - e_n)^r] &\leqslant \sum_{k=1}^{r} \sum_{\substack{\beta_1+\cdots+\beta_k=r \\ \forall\, i\, \beta_i \geqslant 2}} \frac{r!}{\beta_1! \cdots \beta_k!} \sum_{p_1<\cdots<p_k \leqslant \alpha_n} \mathbb{E}[Y_{p_i}^2] \cdots \mathbb{E}[Y_{p_t}^2] \\
&\leqslant \sum_{k=1}^{r} \sum_{\substack{\beta_1+\cdots+\beta_k=r \\ \forall\, i\, \beta_i \geqslant 2}} \frac{r!}{\beta_1! \cdots \beta_k!} \left(\sum_{p \leqslant \alpha_n} \mathbb{E}[Y_p^2] \right)^k \\
&= \sum_{k=1}^{r} \sum_{\substack{\beta_1+\cdots+\beta_k=r \\ \forall\, i\, \beta_i \geqslant 2}} \frac{r!}{\beta_1! \cdots \beta_k!} \sigma_n^{2k}.
\end{aligned}$$

Dans la deuxième somme pour chaque k-uplet, on a $2 \geqslant 2k$, de plus pour n assez grand $\sigma_n \geqslant 1$, donc pour n assez grand $\sigma_n^{2k} \leqslant \sigma_n^r$. Ainsi on a une majoration de la forme

$$\mathbb{E}[(S_n - e_n)^r] \leqslant C(r)\sigma_n^r,$$

où $C(r)$ est une constante ne dépendant que de r, ce qui conclut la preuve du théorème d'Erdös-Kac par la méthode des moments.

4. Méthode de Stein

Introduite par C. Stein en 1972, la méthode de Stein permet de quantifier l'erreur dans l'approximation d'une variable aléatoire par une autre variable aléatoire. Nous commençons donc par introduire la notion de distance entre variable aléatoire, et plus généralement la notion de distance entre probabilités.

4.1. Métriques usuelles

Soit $\mathcal{H}$ une classe de fonction. On définit une distance entre probabilités. Pour $\mathbb{P}, \mathbb{Q}$ des probabilités on pose

$$d_{\mathcal{H}}(\mathbb{P}, \mathbb{Q}) := \sup_{h \in \mathcal{H}} \left| \int_{\mathbb{R}} h \mathrm{d}\mathbb{P} - \int_{\mathbb{R}} h \mathrm{d}\mathbb{Q} \right|.$$

Il faut que $\mathcal{H}$ soit assez "riche" pour que $d_{\mathcal{H}}$ définisse une distance. Si X et Y sont des variables aléatoires alors on mesure leur distance par

$$d_{\mathcal{H}}(X, Y) = \sup_{h \in \mathcal{H}} |\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]|.$$

Donnons quelques exemples de métriques usuelles. Soient X et Y des variables aléatoires réelles.

Exemples 1. (*i*) Si $\mathcal{H} = \{\mathbb{1}_{]-\infty,t]},\ t \in \mathbb{R}\}$, alors la distance $d_{\mathcal{H}}$ est notée d_{K} et est appelée la distance de Kolmogorov. On a

$$d_{\mathrm{K}}(X, Y) = \sup_{x \in \mathbb{R}} |\mathbb{P}(X \leqslant x) - \mathbb{P}(Y \leqslant x)|.$$

(*ii*) Si $\mathcal{H}$ est l'ensemble des fonctions 1-lipschitziennes, alors la distance induite est la distance de Wasserstein notée d_{W}.

(*iii*) Si $\mathcal{H} = \{\mathbb{1}_A,\ A \in \mathcal{B}(\mathbb{R})\}$, alors la distance induite est la distance de la variation totale et est notée d_{VT}.

Voici quelques résultats généraux sur ces distances, démontrés dans la référence [9].

Proposition 21. (*i*) *Si X, Y sont des variables aléatoires, alors*

$$d_{\mathrm{VT}}(X, Y) \leqslant d_{\mathrm{K}}(X, Y).$$

(*ii*) *Si Z est une variable aléatoire admettant une densité par rapport à la mesure de Lebesgue bornée par une constante C, alors pour toute variable aléatoire X*

$$d_{\mathrm{K}}(X, Z) \leqslant \sqrt{2C d_{\mathrm{W}}(X, Z)}.$$

(*iii*) *Soient X, Y des variables aléatoires à valeurs dans Ω.*

(*a*) *Si Ω est un espace discret alors*

$$d_{\mathrm{VT}}(X, Y) = \frac{1}{2} \sum_{x \in \Omega} |\mathbb{P}(X = x) - \mathbb{P}(Y = x)|.$$

(*b*) *Si Ω est un espace continue et si X et Y admettent des fonctions densités f_X et f_Y alors*

$$d_{\mathrm{VT}}(X, Y) = \frac{1}{2} \int_{\Omega} |f_X(x) - f_Y(x)| \mathrm{d}x.$$

Remarque 7. Comme la fonction de répartition caractérise la loi, le point (*ii*) de la proposition 21 assure que si $(X_n)_{n \geqslant 1}$ est une suite de variables aléatoires convergeant vers une variable aléatoire X pour la distance d_{W}, alors la suite $(X_n)_{n \geqslant 1}$ converge en loi vers X.

4.2. **Méthode de Stein pour la loi normale**

La notation $X \sim \mathcal{N}(\mu, \sigma^2)$ indiquera que la variable aléatoire X suit une loi normale de paramètre (μ, σ^2). Dans toute la suite on considérera $\mathcal{H}$ comme étant l'ensemble des fonctions $1-$lipschitziennes sur $\mathbb{R}$. La distance $d_{\mathcal{H}}$ est alors appelée la distance de Wasserstein et sera notée d_{W}.

Toute la méthode de Stein sur l'approximation des lois normales repose sur le lemme fondamental suivant.

Lemme 22. *Soient X une variable aléatoire réelle et $Z \sim \mathcal{N}(0,1)$. Sont équivalents :*
(*i*) $X \sim \mathcal{N}(0,1)$.
(*ii*) $\forall f \in \mathcal{C}^1$ *telle que* $\mathbb{E}[|f'(Z)|] < +\infty$, $\mathbb{E}[f'(X) - Xf(X)] = 0$.

Démonstration. On note Φ la fonction de répartition de la loi normale. " $\Longrightarrow$ " Condition nécessaire.
On suppose que $X \sim \mathcal{N}(0,1)$. On se donne f une fonction de classe $\mathcal{C}^1$ telle que $\mathbb{E}[|f'(Z)|] < +\infty$. On a

$$\begin{aligned}\mathbb{E}[f'(X)] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f'(t) e^{-\frac{t^2}{2}} \mathrm{d}t \\ &= \frac{1}{\sqrt{2\pi}} \int_{0}^{+\infty} f'(t) e^{-\frac{t^2}{2}} \mathrm{d}t + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} f'(t) e^{-\frac{t^2}{2}} \mathrm{d}t,\end{aligned}$$

puis en remarquant que

$$e^{-\frac{t^2}{2}} = \int_t^{+\infty} x e^{-\frac{x^2}{2}} \mathrm{d}x = \int_{-\infty}^{t} -x e^{-\frac{x^2}{2}} \mathrm{d}x$$

il vient

$$\mathbb{E}[f'(X)] = \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} f'(t) \left(\int_t^{+\infty} x e^{-\frac{x^2}{2}} \mathrm{d}x \right) \mathrm{d}t + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} f'(t) \left(\int_{-\infty}^{t} x - e^{-\frac{x^2}{2}} \mathrm{d}x \right) \mathrm{d}t.$$

En appliquant le théorème de Fubini-Lebesgue (en utilisant l'hypothèse $\mathbb{E}[|f'(X)|] < +\infty$) on obtient

$$\begin{aligned}\mathbb{E}[f'(X)] &= \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} x e^{-\frac{x^2}{2}} \left(\int_0^x f'(t) \mathrm{d}t \right) \mathrm{d}x + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} -x e^{-\frac{x^2}{2}} \left(\int_x^0 f'(t) \mathrm{d}t \right) \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} x e^{-\frac{x^2}{2}} \left(f(x) - f(0) \right) \mathrm{d}x + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} -x e^{-\frac{x^2}{2}} \left(f(0) - f(x) \right) \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x f(x) e^{-\frac{x^2}{2}} \mathrm{d}x - f(0) \int_{-\infty}^{+\infty} x e^{-\frac{x^2}{2}} \mathrm{d}x \\ &= \mathbb{E}[Xf(X)] - f(0)\mathbb{E}[X],\end{aligned}$$

or $\mathbb{E}[X] = 0$, on conclut donc par linéarité de l'espérance.

" $\Longleftarrow$ " Condition suffisante. On suppose que $\mathbb{E}[f'(X) - Xf(X)] = 0$ pour toute fonction f dérivable vérifiant $\mathbb{E}[|f'(Z)|] < +\infty$. On fixe $x \in \mathbb{R}$. On pose

$$h : \left| \begin{array}{ccc} \mathbb{R} & \longrightarrow & \mathbb{R} \\ t & \longmapsto & \mathbb{1}_{\{t \leqslant x\}} \end{array} \right.$$

On considère l'équation différentielle suivante d'inconnue f, appelée équation de Stein :

$$f'(t) - tf(t) = h(t) - \mathbb{E}[h(Z)].$$

On va montrer que l'unique solution bornée de cette équation, notée f_x, est donnée par

$$f_x : u \longmapsto e^{\frac{u^2}{2}} \int_u^{+\infty} e^{-\frac{t^2}{2}} \left(\Phi(x) - \mathbb{1}_{\{t \leqslant x\}}\right) \mathrm{d}t.$$

L'équation de Stein admet comme solution l'ensemble des fonctions de la forme

$$u \longmapsto e^{\frac{u^2}{2}} \int_u^{+\infty} e^{-\frac{t^2}{2}} \left(\Phi(x) - \mathbb{1}_{\{t \leqslant x\}}\right) \mathrm{d}t + Ce^{\frac{u^2}{2}},$$

où $C \in \mathbb{R}$. On suppose que $C = 0$, montrons que f_x est bornée. On remarque que

$$f_x(u) = \begin{cases} \sqrt{2\pi} e^{u^2/2} \Phi(u)(1 - \Phi(x)) & \text{si } u \leqslant x, \\ \sqrt{2\pi} e^{u^2/2} \Phi(x)(1 - \Phi(u)) & \text{si } u > x. \end{cases}$$

À l'aide d'une étude de fonctions on obtient pour tout $u > 0$

$$1 - \Phi(u) \leqslant \min\left\{\frac{1}{2}, \frac{1}{\sqrt{2\pi} u}\right\} e^{-\frac{u^2}{2}},$$

on en déduit donc que $\|f_x\|_\infty \leqslant \sqrt{\pi/2}$. De plus il est clair que si $C \neq 0$ alors f_x n'est pas bornée. f_x est donc l'unique solution bornée de l'équation de Stein. Ainsi par hypothèse

$$0 = \mathbb{E}[f_x'(X) - X f_x(X)] = \mathbb{E}[\mathbb{1}_{\{X \leqslant x\}}] - \Phi(x),$$

donc pour tout $x \in \mathbb{R}$

$$\mathbb{P}(X \leqslant x) = \Phi(x),$$

X et Z ont donc la même fonction de répartition : X et Z suivent donc la même loi (ie $X \sim \mathcal{N}(0,1)$). $\square$

Explicitons l'importance de ce lemme. On note $\mathcal{A}$ l'opérateur défini sur $\mathcal{C}^1$ par

$$\mathcal{A}f(x) := f'(x) - xf(x), \forall x \in \mathbb{R}.$$

Soit $h \in \mathcal{H}$ fixée. On note f_h l'unique solution bornée de l'équation différentielle

$$\mathcal{A}f = h - \mathbb{E}[h(Z)],$$

où $Z \sim \mathcal{N}(0,1)$. Ainsi, si X est une variable aléatoire réelle, comme

$$d_{\mathcal{H}}(X, Z) = \sup_{h \in \mathcal{H}} |\mathbb{E}[h(X)] - \mathbb{E}[h(Z)]|,$$

il vient

$$d_{\mathcal{H}}(X, Z) = \sup_{h \in \mathcal{H}} |\mathbb{E}[\mathcal{A}f_h(X)]|.$$

L'idée générale est donc maintenant de majorer le terme de droite (qui est souvent plus facile à majorer que le terme de gauche).

4.3. **Exemple d'application**

Avant de donner un exemple d'application de la méthode de Stein pour les lois normales, nous énonçons un résultat général sur les solutions de l'équation de Stein qui nous servira dans la suite. Ce résultat est démontré dans [4].

Proposition 23. *Soit $h \in \mathcal{H}$ et $Z \sim \mathcal{N}(0,1)$ une variable aléatoire. Alors l'unique solution bornée f_h de l'équation de Stein d'inconnue f*

$$\mathcal{A}f = h - \mathbb{E}[h(Z)]$$

est donnée par

$$f_h : x \longmapsto e^{\frac{x^2}{2}} \int_x^{+\infty} e^{-\frac{t^2}{2}} (\mathbb{E}[h(Z)] - h(t))\mathrm{d}t,$$

avec

$$\|f_h\|_\infty \leqslant 1, \quad \|f_h'\|_\infty \leqslant \sqrt{\frac{2}{\pi}}, \quad \text{et} \quad \|f_h''\|_\infty \leqslant 2.$$

Nous allons maintenant redémontrer le théorème central limite pour des variables non identiquement distribuées par la méthode de Stein. Nous rappelons son énoncé.

Théorème. *Soit $(X_n)_{n\geqslant 1}$ des variables aléatoires réelles centrées réduites, uniformément bornées en valeur absolue par une constante C. Pour $n \geqslant 1$ on note $S_n = \frac{1}{\sqrt{n}}\sum_{k=1}^n X_k$. On suppose que les variables X_k pour $k \geqslant 1$ sont indépendantes. Alors*

$$S_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Démonstration par la méthode de Stein. Soient $h \in \mathcal{H}$ une fonction test et $Z \sim \mathcal{N}(0,1)$. On note f l'unique solution de l'équation de Stein. Soit $n \in \mathbb{N}^*$. Pour $i \in [\![1,n]\!]$, on note

$$S_n^i := \frac{1}{\sqrt{n}} \sum_{\substack{k=1 \\ k\neq i}}^n X_k = \frac{1}{\sqrt{n}}(S_n - X_i).$$

Par le lemme des coalitions, les variables aléatoires S_n^i et X_i sont indépendantes, donc

$$\mathbb{E}[S_n^i X_i] = \mathbb{E}[S_n^i]\mathbb{E}[X_i] = 0.$$

De là on obtient

$$\begin{aligned}\mathbb{E}[S_n f(S_n)] &= \mathbb{E}\left[\frac{1}{\sqrt{n}}\sum_{k=1}^n \left(X_k f(S_n) - X_k f(S_n^k)\right)\right] \\ &= \mathbb{E}\left[\frac{1}{\sqrt{n}}\sum_{k=1}^n X_k \left(f(S_n) - f(S_n^k) - (S_n - S_n^k) f'(S_n)\right)\right] + \mathbb{E}\left[\frac{1}{\sqrt{n}}\sum_{k=1}^n X_k f'(S_n)\left(S_n - S_n^k\right)\right].\end{aligned}$$

On a donc

$$\begin{aligned}|\mathbb{E}[\mathcal{A}f(S_n)]| &= |\mathbb{E}[f'(S_n) - S_n f(S_n)]| \\ &\leqslant \left|\mathbb{E}\left[\frac{1}{\sqrt{n}}\sum_{k=1}^n X_k \left(f(S_n) - f(S_n^k) - (S_n - S_n^k) f'(S_n)\right)\right]\right| \\ &\quad + \left|\mathbb{E}\left[f'(S_n)\left(1 - \frac{1}{\sqrt{n}}\sum_{k=1}^n X_k\left(S_n - S_n^k\right)\right)\right]\right|\end{aligned}$$

Par l'inégalité triangulaire le premier terme peut être majoré par

$$\mathbb{E}\left[\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left|X_k\left(f(S_n)-f(S_n^k)-(S_n-S_n^k)f'(S_n)\right)\right|\right]$$

puis par une inégalité de Taylor on majore finalement par

$$\frac{\|f''\|_\infty}{2n^{3/2}}\sum_{k=1}^{n}\mathbb{E}[X_k^3] \leqslant \frac{\|f''\|_\infty C^3}{2\sqrt{n}},$$

en remarquant que $S_n - S_n^k = X_k/\sqrt{n}$. Par l'inégalité triangulaire le dernier terme peut être majoré par la quantité

$$\frac{\|f'\|_\infty}{n}\mathbb{E}\left[\left|\sum_{k=1}^{n}(1-X_k^2)\right|\right]$$

puis par l'inégalité de Cauchy-Schwarz on majore par

$$\frac{\|f'\|_\infty}{n}\sqrt{\mathrm{Var}\left(\sum_{k=1}^{n}X_k^2\right)}.$$

Finalement, en utilisant l'indépendance et le fait que $\mathrm{Var}\,(X_i^2) \leqslant \mathbb{E}[X_i^4] \leqslant C^4$, on obtient

$$|\mathbb{E}[\mathcal{A}f(S_n)]| \leqslant \frac{\|f''\|_\infty C^3}{2\sqrt{n}} + \frac{\|f'\|_\infty C^2}{\sqrt{n}}.$$

Or $h \in \mathcal{H}$, donc par la proposition 23,

$$|\mathbb{E}[\mathcal{A}f(S_n)]| \leqslant \frac{C^3}{\sqrt{n}} + \frac{\sqrt{2/\pi}C^2}{\sqrt{n}},$$

de plus le majorant est indépendant de h donc

$$d_{\mathrm{W}}(S_n, Z) \leqslant \frac{K}{\sqrt{n}},$$

où K est une constante réelle.
Ainsi

$$d_{\mathrm{W}}(S_n, Z) \underset{n\to+\infty}{\longrightarrow} 0,$$

ce qui prouve que $S_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ (par la remarque 7). □

Remarque 8. Au vu de la démonstration, on peut juste supposer que les moments d'ordre 4 soient finis.

4.4. **Démonstration du théorème d'Erdös-Kac**

Dans cette partie la lettre q sera elle aussi réservée aux nombres premiers, et on fixe un entier $n \geqslant 2$. Pour p un nombre premier on pose

$$e_p := \frac{\delta_p - \mathbb{E}_n[\delta_p]}{\nu_n}$$

avec

$$\nu_n = \sum_{p \leqslant s_n} \mathbb{E}_n[(\delta_p - \mathbb{E}_n[\delta_p])^2].$$

Comme dans la méthode des moments, nous allons tronquer la variable aléatoire $\sum_p e_p$. On introduis pour cela la suite de terme général

$$s_n = n^{1/3(\log(\log(n)))^2},$$

et on pose

$$\widetilde{\omega}_n := \sum_{p \leqslant s_n} e_p.$$

Nous allons montrer que

$$\lim_{n \to +\infty} |\mathbb{P}_n(\widetilde{\omega}_n \leqslant x) - \Phi(x)| = 0,$$

ce qui démontrera le théorème d'Erdös-Kac. En effet, on vérifie facilement grâce au théorème 9 que

$$\nu_n \sim \log(\log(n)),$$

il suffit donc de faire une preuve analogue de l'étape 1 de la méthode des moments. Par la méthode de Stein, on peut montrer que

$$\begin{aligned} |\mathbb{P}(\widetilde{\omega}_n \leqslant x) - \Phi(x)| \leqslant\ & 2\sqrt{\mathbb{E}\left[\left(\sum_{p \leqslant s_n} e_p^2 - \mathbb{E}[e_p^2]\right)^2\right]} \\ & + \sqrt{\frac{\pi}{2}}\mathbb{E}\left[\sum_{p \leqslant s_n} |\mathbb{E}[e_p|\{e_q\}_{q \neq p, q \leqslant s_n}]|\right] \\ & + \sqrt{\sqrt{\frac{8}{\pi}}\mathbb{E}\left[\sum_{p \leqslant s_n} |e_p|^3\right]}. \end{aligned}$$

Ce résultat est expliqué et démontré dans [11], pp 105-110. Il nous reste donc à majorer les trois termes de droite.

1^{er} terme

Il nous faut majorer la quantité

$$2\sqrt{\mathbb{E}_n\left[\left(\sum_{p \leqslant s_n} e_p^2 - \mathbb{E}_n[e_p^2]\right)^2\right]}.$$

On a

$$\sum_{p \leqslant s_n} e_p^2 - \mathbb{E}_n[e_p^2] = \frac{1}{\nu_n}\sum_{p \leqslant s_n}(\delta_p - \mathbb{E}_n[\delta_p]) + \frac{2}{\nu_n}\sum_{p \leqslant s_n}(\mathbb{E}_n[\delta_p]^2 - \mathbb{E}_n[\delta_p]\delta_p),$$

donc par le théorème 9 et en utilisant le fait que $\nu_n \sim \log(\log(n))$ il est clair qu'il suffit de majorer

$$\mathbb{E}_n\left[\left(\sum_{p \leqslant s_n} \delta_p - \mathbb{E}_n[\delta_p]\right)^2\right].$$

En développant le carré on obtient

$$\mathbb{E}_n\left[\left(\sum_{p \leqslant s_n} \delta_p - \mathbb{E}_n[\delta_p]\right)^2\right] = \nu_n + 2\sum_{p < q \leqslant s_n} \mathbb{E}_n[(\delta_p - \mathbb{E}_n[\delta_p])(\delta_q - \mathbb{E}_n[\delta_q])],$$

on peut de plus montrer que la somme de droite est un $o(\nu_n^2)$ (en effet les variables $(\delta_p)_p$ sont "presque indépendantes", cf partie 2.4). Ainsi

$$2\sqrt{\mathbb{E}_n\left[\left(\sum_{p\leqslant s_n} e_p^2-\mathbb{E}_n[e_p^2]\right)^2\right]}=O\left(\frac{1}{\sqrt{\nu_n}}\right)=O\left(\frac{1}{\log(\log(n))^{1/2}}\right).$$

2$^{\text{ème}}$ terme

On cherche à majorer la quantité

$$\sqrt{\frac{\pi}{2}}\mathbb{E}_n\left[\sum_{p\leqslant s_n}|\mathbb{E}_n[e_p|\{e_q\}_{q\neq p,q\leqslant s_n}]|\right],$$

qui est égale à

$$\sqrt{\frac{\pi}{2}}\frac{1}{\nu_n}\mathbb{E}_n\left[\sum_{p\leqslant s_n}|\mathbb{E}_n[\delta_p|\{\delta_q\}_{q\neq p,q\leqslant s_n}]|\right].$$

On admet que

$$\sqrt{\frac{\pi}{2}}\mathbb{E}_n\left[\sum_{p\leqslant s_n}|\mathbb{E}_n[e_p|\{e_q\}_{q\neq p,q\leqslant s_n}]|\right]=O\left(\frac{1}{\log(\log(n))^{1/2}}\right).$$

Détaillons tout de même le calcul de la variable aléatoire $\mathbb{E}_n[\delta_p|\{\delta_q\}_{q\neq p,q\leqslant s_n}]$, que l'on note $D_{n,p}$. Soient p un nombre premier $\leqslant s_n$ et $m\in\mathbb{N}^*$. Alors

$$\begin{aligned}D_{n,p}(m)&=\mathbb{E}_n[\delta_p|\{\delta_q=\delta_q(m)\}_{q\neq p,q\leqslant s_n}]\\&=\frac{1}{\mathrm{Card}\,\{k\leqslant n:\forall q\leqslant s_n,q\neq p,\delta_q(k)=\delta_q(m)\}}\sum_{\substack{k\leqslant n\\ \forall q\leqslant s_n,q\neq p,\delta_q(k)=\delta_q(m)}}\delta_p(k)\\&=\frac{\mathrm{Card}\,\{k\leqslant n:p|k,\forall q\leqslant s_n,q\neq p,\delta_q(k)=\delta_q(m)\}}{\mathrm{Card}\,\{k\leqslant n:\forall q\leqslant s_n,q\neq p,\delta_q(k)=\delta_q(m)\}}\\&=\frac{\mathrm{Card}\,\{k\leqslant \frac{n}{p}:\forall q\leqslant s_n,q\neq p,\delta_q(k)=\delta_q(m)\}}{\mathrm{Card}\,\{k\leqslant n:\forall q\leqslant s_n,q\neq p,\delta_q(k)=\delta_q(m)\}}.\end{aligned}$$

On fixe un entier naturel a et on pose pour $m\in\mathbb{N}^*$

$$f_m(n):=\mathrm{Card}\,\{k\leqslant n:\forall q\leqslant s_a,q\neq p,\delta_q(k)=\delta_q(m)\}.$$

Déterminons une expression de $f_m(n)$ en fonction de m et de n. On a

$$\begin{aligned}f_m(n)&=\mathrm{Card}\,\{k\leqslant n:\forall q\leqslant s_a,q\neq p,\text{ si }\delta_q(m)=0,\text{ alors }(q,k)=1)\}\\&=\sum_{k\leqslant n}\mathbb{1}(d_{n,m}|k)\mathbb{1}((k,r_{n,m})=1)\end{aligned}$$

avec $d_{n,m}$ et $r_{n,m}$ des produits de nombres premiers $\leqslant s_a$, en effet $d_{n,m}=\prod_{q\leqslant s_a,q\neq p}q^{\delta_q(m)}$ et $r_{n,m}=\prod_{q\leqslant s_a,q\neq p}q^{1-\delta_q(m)}$. On a donc

$$f_m(n)=\sum_{\substack{k\leqslant n\\ d_{n,m}|k,\,(k,r_{n,m})=1}}1$$

ce qui nous donne par le lemme 11

$$f_m(n) = \frac{n}{d_{n,m}}\frac{\varphi(r_{n,m})}{r_{n,m}} + O(\tau(n/d_{n,m})) = n\frac{\varphi(r_{n,m})}{\prod_{q \leqslant s_a, q\neq p} q} + O(\tau(n/d_{n,m})).$$

On a donc (avec $a = n$) :

$$D_{n,p}(m) = \frac{f_m(n/p)}{f_m(n)} = \frac{\frac{n}{p}\frac{\varphi(r_{n,m})}{\prod_{q \leqslant s_n, q\neq p} q} + O(\tau(n/pd_{n,m}))}{n\frac{\varphi(r_{n,m})}{\prod_{q \leqslant s_n, q\neq p} q} + O(\tau(n/d_{n,m}))} = \frac{1}{p}\frac{\frac{\varphi(r_{n,m})}{\prod_{q \leqslant s_n, q\neq p} q} + O(\tau(1/pd_{n,m}))}{\frac{\varphi(r_{n,m})}{\prod_{q \leqslant s_n, q\neq p} q} + O(\tau(1/d_{n,m}))}.$$

3$^{\text{ème}}$ terme

Il nous reste donc à majorer la quantité

$$\sqrt{\sqrt{\frac{8}{\pi}}\mathbb{E}\left[\sum_{p \leqslant s_n} |e_p|^3\right]}.$$

On remarque que

$$\mathbb{E}_n[|e_p|^3] \leqslant \frac{\mathbb{E}_n[\delta_p^3] + \mathbb{E}_n[\delta_p]^3}{\nu_n^{3/2}}.$$

On sait que $\mathbb{E}_n[\delta_p] = \frac{1}{p} + O\left(\frac{1}{n}\right)$, donc par le théorème 9

$$\sum_{p \leqslant s_n} \mathbb{E}_n[|e_p^3|] = O\left(\frac{\log(\log(n))}{(\log(\log(n))^{3/2}}\right).$$

Ainsi

$$\sqrt{\sqrt{\frac{8}{\pi}}\mathbb{E}\left[\sum_{p \leqslant s_n} |e_p|^3\right]} = O\left(\frac{1}{\log(\log(n))^{1/4}}\right),$$

ce qui achève la preuve du théorème d'Erdös-Kac par la méthode de Stein.

Références

[1] A. Barbour et L.Chen *An introduction to Stein's method.*

[2] J-C. Breton *Fondements des PROBABILITÉS, Université de Rennes 1.*

[3] H. Brézis *Analyse fonctionnelle, théorie et applications.*

[4] L. H. Y. Chen, L. Goldstein et Q-M. Shao *Normal approximation by Stein's method.*

[5] D. Choimet *Combien les nombres entiers ont-ils de facteurs premiers ?.*

[6] A. J.Harper *Two new proofs of the Erdös-Kac Theorem, with bound on the rate of convergence, by Stein's method for distributional approximations.*

[7] H. Jeffreys et B. S. Jeffreys *Methods of Mathematical Physics, chapitre 1, p.26-36.*

[8] E. Konzu *Lois gaussiennes inverses (généralisées), lois de Kummer et méthode de Stein.*

[9] N. Ross *Fundamentals of stein's method.*

[10] W. Rudin *Analyse réelle et complexe, chapitre 6.*

[11] C. Stein *Approximate Computation of Expectations.*

[12] G. Tenenbaum *Introduction to Analytic and Probalistic Number Theory, chapitre I.1, p.9-19.*