Processus stochastiques discrets – Méthode de moments et marche aléatoire

Valentin Féray, Université de Lorraine

octobre 2024 – janvier 2025

Première partie

Méthodes de moments

Par définition, le moment d'ordre k d'une variable aléatoire X est $\mathbb{E}[X^k]$ (il peut être non défini, ou égal à $+\infty$). Dans de nombreux modèles, il est possible de calculer (ou d'estimer asymptotiquement) les moments alors qu'on ne sait pas calculer la distribution de X (ou sa fonction caractéristique), notamment grâce à la linéarité de l'espérance.

Il est donc important de savoir retrouver de l'information sur X (ou sur le comportement asymptotique d'une suite de variables aléatoire X_n) à partir de ses moments. Ce type de techniques est très utile pour l'étude d'objets combinatoires aléatoires (mots, permutations ; voir aussi le cours de Pascal Moyal au S10 sur les graphes aléatoires) et nous verrons de nombreuses applications dans ce cours.

1 Méthode du premier moment

Soit $\mathbb{N}$ l'ensemble des entiers naturels.

Theorème 1.1. *Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. Alors*

$$\mathbb{P}[X = 0] \geq 1 - \mathbb{E}[X].$$

En particulier, si X_n est une suite de v.a. a valeurs dans $\mathbb{N}$ telle que $\mathbb{E}[X_n]$ tend vers 0, alors $\mathbb{P}[X_n = 0]$ tend vers 1.

Première démonstration.

$$\mathbb{E}(X) = \sum_{k\geq 0} k\mathbb{P}[X = k] \geq \sum_{k\geq 1} \mathbb{P}[X = k] = 1 - \mathbb{P}[X = 0]. \qquad \square$$

Deuxième démonstration. On commence par rappeler l'inégalité de Markov : si $Y \geq 0$ p.s., on a, pour tout $a > 0$,

$$\mathbb{E}(Y) = \mathbb{E}(Y\,\mathbf{1}[Y \geq a]) + \mathbb{E}(Y\,\mathbf{1}[Y < a]) \geq a\mathbb{P}[Y \geq a],$$

ce qui se réécrit $\mathbb{P}[Y \geq a] \leq \frac{\mathbb{E}[Y]}{a}$.

On applique ce résultat à $Y = X$ et $a = 1$. On a

$$\mathbb{P}[X = 0] = 1 - \mathbb{P}[X \geq 1] \geq 1 - \mathbb{E}[X]. \qquad \square$$

On utilise souvent ce théorème dans le contexte suivant : X_n compte le nombre d'occurrences d'une certaine structure dans un objet aléatoire, et si $\mathbb{E}[X_n]$ tend vers 0, on sait que, avec probabilité $1 - o(1)$, cette structure n'apparait pas.

Exemple 1.2. Soit w_n un mot aléatoire uniforme dans $\{0,1\}^n$, i.e. on tire n variables indépendantes de loi Bern(1/2). Fixons $\varepsilon > 0$, et posons X_n le nombre de blocs de 0 consécutifs de taille $\ell_n := \lfloor (1+\varepsilon)\log_2(n) \rfloor$. Soit A_i l'évènement $w_i = \cdots = w_{i+\ell_n-1} = 0$, i.e. "$i$ est le début d'un blocs de ℓ_n zéros", alors

$$X_n = \sum_{i=1}^{n-\ell_n+1} \mathbf{1}[A_i].$$

On en déduit

$$\mathbb{E}[X_n] = \sum_{i=1}^{n-\ell_n+1} \mathbb{P}[A_i] = \sum_{i=1}^{n-\ell_n+1} 2^{-\ell_n} \leq n 2^{-\lfloor (1+\varepsilon)\log_2(n) \rfloor} \leq 2^{-\varepsilon \log_2(n)+1} = 2\,n^{-\varepsilon}.$$

La borne supérieure tend vers 0, donc $\mathbb{E}[X_n]$ tend vers 0. On en déduit que, avec probabilité tendant vers 1, w_n ne contient pas de blocs de zéros de taille $\lfloor (1+\varepsilon)\log_2(n) \rfloor$.

Nous allons voir maintenant un exemple d'application aux permutations. Pour cela, le lemme suivant sera utile. Notons S_n l'ensemble des permutations de n, i.e. l'ensemble des bijections de $\{1,\ldots,n\}$ dans $\{1,\ldots,n\}$.

Theorème 1.3. *Soit σ_n une permutation aléatoire uniforme de taille n. Soient $k \geq 1$ un entier et $(i_1,\ldots,i_k)$ et $(j_1,\ldots,j_k)$ deux listes d'entiers $\leq n$ sans répétitions (mais potentiellement avec une intersection non vide). Alors*

$$\mathbb{P}(\sigma(i_1) = j_1 \wedge \cdots \wedge \sigma(i_k) = j_k) = \frac{1}{n(n-1)\ldots(n-k+1)}. \qquad (1)$$

Démonstration. Fixons une liste $(i_1,\ldots,i_k)$ sans répétition. Je prétends que le membre de gauche de (1) ne dépend pas de $(j_1,\ldots,j_k)$. En effet, prenons deux suites $(j_1,\ldots,j_k)$ et $(j'_1,\ldots,j'_k)$ sans répétitions et choisissons une permutation τ telle que $\tau(j_1) = j'_1$, …, $\tau(j_k) = j'_k$ (note : τ n'est pas aléatoire). Les évènements

$$\sigma(i_1) = j_1 \wedge \cdots \wedge \sigma(i_k) = j_k,$$

et

$$\tau \circ \sigma(i_1) = j'_1 \wedge \cdots \wedge \tau \circ \sigma(i_k) = j'_k$$

sont alors identiques. Mais si σ est une permutation uniforme dans S_n, alors $\tau \circ \sigma$ est aussi uniforme. On a donc

$$\begin{aligned}\mathbb{P}(\sigma(i_1) = j_1 \wedge \cdots \wedge \sigma(i_k) = j_k) &= \mathbb{P}(\tau \circ \sigma(i_1) = j'_1 \wedge \cdots \wedge \tau \circ \sigma(i_k) = j'_k \\ &= \mathbb{P}(\sigma(i_1) = j'_1 \wedge \cdots \wedge \sigma(i_k) = j'_k),\end{aligned}$$

ou la dernière égalité vient du fait que σ et $\tau \circ \sigma$ ont la même distribution (distribution uniforme sur S_n). Comme il y a $n(n-1)\dots(n-k+1)$ listes $(j_1, \dots, j_k)$ sans répétition, on a

$$\mathbb{P}(\sigma(i_1) = j_1 \wedge \cdots \wedge \sigma(i_k) = j_k) = \frac{1}{n(n-1)\dots(n-k+1)},$$

comme annoncé. □

Voici un dernier exemple d'application de la méthode du premier moment. Une sous-suite croissante dans une permutation σ est une liste $(\sigma(i_1), \dots, \sigma(i_k))$ telle que

$$i_1 < \cdots < i_k \text{ et } \sigma(i_1) < \cdots < \sigma(i_k).$$

Par exemple, si $\sigma = 4\,7\,3\,5\,1\,8\,2\,6$ (j'utilise ici la notation "en ligne" des permutations consistant à écrire $\sigma(1)\,\sigma(2)\dots$ sans ponctuation), alors nous avons plusieurs sous-suites croissantes de taille 3, parmi lesquelles $1\,2\,6$, $3\,5\,8, \dots$, mais aucune de taille 4 ou plus.

Avant de donner un énoncé, rappelons/introduisons quelques notations :

— $\binom{n}{k} := \frac{n!}{k!(n-k)!}$ est le coefficient binomial "k parmi n" ;

— $[\![1, n]\!] := \{1, \dots, n\}$ l'ensemble des entiers de 1 à n ;

— $\binom{[\![1,n]\!]}{k}$ l'ensemble des sous-parties de $[\![1, n]\!]$ à k elements.

Proposition 1.4. *Soit σ_n une permutation aléatoire uniforme de taille n, et L_n la longueur de sa plus longue sous-suite croissante. Alors*

$$\lim_{n \to +\infty} \mathbb{P}[L_n \leq 3\sqrt{n}] = 0.$$

Démonstration. Soit X_n le nombre de sous-suites croissantes de taille $\lfloor 3\sqrt{n} \rfloor$: on a

$$X_n = \sum_{I \in \binom{[\![1,n]\!]}{\lfloor 3\sqrt{n} \rfloor}} \mathbf{1}[\sigma/I \text{ croissante}],$$

ou la somme est prise sur les sous-ensembles de $[\![1, n]\!]$ de taille $\lfloor 3\sqrt{n} \rfloor$. Fixons $I \in \binom{[\![1,n]\!]}{\lfloor 3\sqrt{n} \rfloor}$. Alors,

$$\mathbb{P}[\sigma/I \text{ croissante}] = \frac{1}{n!} \#\{\sigma \in S_n : \sigma/I \text{ croissante}]\}.$$

Le numérateur se calcule ainsi : on commence par choisir l'ensemble des valeurs de σ/I ($\binom{n}{\lfloor 3\sqrt{n}\rfloor}$ choix) ; comme σ/I est croissante, cela détermine entièrement σ/I), et on complète ensuite la permutation (il reste $n - \lfloor 3\sqrt{n}\rfloor$ éléments à envoyer sur $n - \lfloor 3\sqrt{n}\rfloor$ valeurs, soit $(n - \lfloor 3\sqrt{n}\rfloor)!$ choix). Cela donne

$$\mathbb{P}[\sigma/I \text{ croissante}] = \frac{\binom{n}{\lfloor 3\sqrt{n}\rfloor}(n - \lfloor 3\sqrt{n}\rfloor)!}{n!} = \frac{1}{(\lfloor 3\sqrt{n}\rfloor)!}$$

et finalement

$$\mathbb{E}(X_n) = \binom{n}{\lfloor 3\sqrt{n}\rfloor}\frac{1}{(\lfloor 3\sqrt{n}\rfloor)!} = \frac{n!}{(n - \lfloor 3\sqrt{n}\rfloor)!(\lfloor 3\sqrt{n}\rfloor)!^2}.$$

En utilisant la formule de Stirling, on obtient

$$\mathbb{E}(X_n) \sim \frac{n^n e^{\lfloor 3\sqrt{n}\rfloor}}{(n - \lfloor 3\sqrt{n}\rfloor)^{n-\lfloor 3\sqrt{n}\rfloor}(9n)^{\lfloor 3\sqrt{n}\rfloor}(6\pi\sqrt{n})}.$$

En remarquant que

$$(n - \lfloor 3\sqrt{n}\rfloor)^{n-\lfloor 3\sqrt{n}\rfloor} = n^{n-\lfloor 3\sqrt{n}\rfloor}(1 - 3n^{-1/2})^{n-\lfloor 3\sqrt{n}\rfloor} = n^{n-\lfloor 3\sqrt{n}\rfloor}\, e^{-\lfloor 3\sqrt{n}\rfloor}\, O(1),$$

cela implique

$$\mathbb{E}(X_n) \sim \frac{(e^2)^{\lfloor 3\sqrt{n}\rfloor}}{9^{\lfloor 3\sqrt{n}\rfloor}(6\pi\sqrt{n})} \to 0.$$

On conclut par le théorème 1.1 que $X_n = 0$ avec probabilité $1 - o(1)$. En d'autres termes, avec probabilité $1 - o(1)$ une permutation aléatoire uniforme ne contient pas de sous-suites croissantes de taille $\lfloor 3\sqrt{n}\rfloor$ □

2 Méthode du second moment

La méthode du premier moment sert à prouver qu'une certaine variable X_n vaut 0 avec probabilité $1 - o(1)$. Pour prouver que ce n'est pas le cas, il n'est pas suffisant de regarder l'espérance de X_n, mais il faut regarder la variance. On utilise ensuite le résultat suivant, connu sous le nom d'inégalité de Bienaymé–Tchebychev.

Theorème 2.1. *Soit X une variable aléatoire de variance σ^2, supposée finie. Alors pour tout $\alpha > 0$, on a*

$$\mathbb{P}\left(|X - \mathbb{E}[X]| \geq \alpha\right) \leq \frac{\sigma^2}{\alpha^2}.$$

Démonstration. On applique l'inégalité de Markov ($\mathbb{P}[Y \geq a] \leq \frac{\mathbb{E}[Y]}{a}$ pour $Y, a \geq 0$) à $Y = (X - \mathbb{E}[X])^2$ et à $a = \alpha^2$. □

Ce type d'inégalité qui borne la probabilité que X soit loin de sa moyenne (parfois uniquement dans une direction) est appelée "inégalité de concentration". On en verra d'autres dans ce cours.

L'inégalité de Bienaymé–Tchebichev implique en particulier,

$$\mathbb{P}(X=0) \leq \mathbb{P}\left(|X-\mathbb{E}[X]| \geq \mathbb{E}[X]\right) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2}.$$

Conséquence : Prenons une suite de v.a. X_n avec un second moment fini. Si $\mathrm{Var}(X_n) = o(\mathbb{E}[X_n]^2)$, alors $\lim \mathbb{P}(X_n = 0) = 0$, i.e. $X_n \neq 0$ avec probabilité $1-o(1)$. Dans ce cas, on peut dire plus. Supposons que, pour tout n, $\mathbb{E}[X_n] \neq 0$. Alors $\frac{X_n}{\mathbb{E}[X_n]} \to 1$ en probabilité. En effet, pour tout $\varepsilon > 0$, on a

$$\mathbb{P}\Big[\Big|\frac{X_n}{\mathbb{E}[X_n]} - 1\Big| \geq \varepsilon\Big] = \mathbb{P}\Big[\big|X_n - \mathbb{E}[X_n]\big| \geq \varepsilon \mathbb{E}[X_n]\Big] \leq \frac{\mathrm{Var}(X_n)}{\varepsilon^2 \mathbb{E}[X_n]^2},$$

et cette borne supérieure tend vers 0 par hypothèse.

Exemple 2.2. On considère comme ci-dessus, w_n un mot aléatoire uniforme dans $\{0,1\}^n$. Fixons $\varepsilon > 0$, et posons Y_n le nombre de blocs de 0 consécutifs de taille $\ell_n := \lfloor (1-\varepsilon)\log_2(n) \rfloor$. Soit B_i l'évènement "i est le début d'un blocs de ℓ_n zéros", alors

$$Y_n = \sum_{i=1}^{n-\ell_n+1} \mathbf{1}[B_i].$$

On calcule l'espérance comme dans l'exemple précédent :

$$\mathbb{E}[Y_n] = \sum_{i=1}^{n-\ell_n+1} \mathbb{P}[B_i] = \sum_{i=1}^{n-\ell_n-1} 2^{-\ell_n} = (n-\ell_n-1)\, 2^{-\lfloor (1-\varepsilon)\log_2(n)} \sim 2^{\varepsilon \log_2(n)} = n^{\varepsilon}.$$

On voit que $\mathbb{E}[Y_n]$ tend vers $+\infty$, on ne peut rien conclure sur $\mathbb{P}[Y_n = 0]$. Regardons le second moment.

$$\mathrm{Var}[Y_n] = \sum_{1 \leq i,j \leq n-\ell_n+1} \mathrm{Cov}(B_i, B_j).$$

Si $|i-j| > \ell_n$, les évènements B_i et B_j concernent des lettres différentes du mot w_n et sont donc indépendants. En particulier $\mathrm{Cov}(B_i, B_j) = 0$ dans ce cas. Si $|i-j| \leq \ell_n$, comme $\mathbb{E}(B_i)\,\mathbb{E}(B_j) \geq 0$, on a

$$\mathrm{Cov}(B_i, B_j) \leq \mathbb{E}(B_i B_j) \leq \mathbb{E}(B_i) = 2^{-\ell_n}.$$

Il y a moins de $n(2\ell_n+1)$ paires (i,j) dans la somme avec $|i-j| \leq \ell_n$ (pour chaque i, il y a au plus $2\ell_n+1$ valeurs possibles pour j), donc

$$\mathrm{Var}[Y_n] \leq n(2\ell_n+1)2^{-\ell_n} \sim 2n^{\varepsilon}\log_2(n),$$

où la dernière estimée utilise que $\ell_n = \lfloor (1-\varepsilon)\log_2(n) \rfloor$.

Comme $\mathbb{E}[Y_n]^2 \sim n^{2\varepsilon}$, on a $\mathrm{Var}[Y_n] = o(\mathbb{E}[Y_n]^2)$ et $Y_n \neq 0$ avec probabilité $1-o(1)$. En d'autres termes, avec probabilité $1-o(1)$, w_n contient au moins une suite de 0 consécutifs de taille $(1-\varepsilon)\log_2(n)$.

3 Moments et convergence en distribution

3.1 Convergence des moments et convergence en distribution

Soit (X_n) une suite de v.a. et une v.a. Z (candidat limite pour X_n), on veut comparer les deux propriétés suivantes

Convergence en distribution $X_n \xrightarrow{d} Z$.

Convergence des moments Pour tout r entier positif, $\mathbb{E}[X_n^r] \to \mathbb{E}[Z^r]$.

De la convergence en distribution vers convergence des moments.

La convergence en distribution **n'implique pas** la convergence des moments ; par exemple, si X_n est définie par

$$\mathbb{P}(X_n = n) = 1/n = 1 - \mathbb{P}(X_n = 0),$$

alors X_n converges en distribution vers 0, mais $\mathbb{E}(X_n) = 1$ pour tout $n \geq 1$.

Cependant, avec des hypothèses supplémentaires, on peut obtenir la convergence des moments.

Theorème 3.1. *Soit r un entier positif et s un réel avec $r < s$. On suppose que $X_n \xrightarrow{d} Z$ et que $\mathbb{E}[|X_n|^s]$ est une suite bornée. Alors $\mathbb{E}[X_n^r]$ tend vers $\mathbb{E}[Z^r]$.*

En particulier, si tous les moments de X_n sont bornés, alors la convergence en distribution implique la convergence des moments.

Démonstration. Rappelons que la convergence en distribution de X_n vers Z signifie que pour toute fonction f continue bornée, on a $\mathbb{E}[f(X_n)] = \mathbb{E}[f(Z)]$. La difficulté que la fonction $g : x \mapsto x^r$ n'est pas bornée sur $\mathbb{R}$. On va la remplacer par une approximation bornée. Pour $A > 0$ on pose

$$g_A(x) = \begin{cases} x^r & \text{si } |x| \leq A\,; \\ (-A)^r & \text{si } x < -A\,; \\ A^r & \text{si } x > A. \end{cases}$$

Pour x dans $\mathbb{R}$, on a

$$|x^r - g_A(x)| \leq |x|^r \mathbf{1}_{[|x|>A]} \leq \frac{|x|^s}{A^{s-r}}.$$

En prenant l'espérance, on obtient

$$\big|\mathbb{E}[X_n^r] - \mathbb{E}[g_A(X_n)]\big| \leq \mathbb{E}\big[\,|X_n^r - g_A(X_n)|\,\big] \leq \frac{\mathbb{E}[|X_n|^s]}{A^{s-r}} \leq \frac{M}{A^{s-r}},$$

où $M = \sup \mathbb{E}[|X_n|^s]$ est fini par hypothèse. Fixons $\varepsilon > 0$ et choisissons A_0 tel que $\frac{M}{A_0^{s-r}} \leq \varepsilon$. On a

$$\big|\mathbb{E}[X_n^r] - \mathbb{E}[g_{A_0}(X_n)]\big| \leq \varepsilon.$$

On cherche une approximation similaire pour le moment d'ordre r de la variable limite Z. Par le théorème de représentation de Skorokhod [1] et le lemme de Fatou, on a

$$\mathbb{E}[|Z|^s] \leq \liminf_{n\to\infty} \mathbb{E}[|X_n|^s] \leq M.$$

À partir de là, le même argument que ci-dessus donne

$$\left|\mathbb{E}[Z^r] - \mathbb{E}\big(g_{A_0}(Z)\big)\right| \leq \varepsilon.$$

En mettant ensemble les deux approximations obtenues, on a

$$\left|\mathbb{E}[X_n^r] - \mathbb{E}[Z^r]\right| \leq \left|\mathbb{E}\big(g_{A_0}(X_n)\big) - \mathbb{E}\big(g_{A_0}(Z)\big)\right| + 2\varepsilon.$$

Comme $X_n \xrightarrow{d} Z$ et comme g_{A_0} est continue bornée, on a

$$\lim_{n\to\infty} \mathbb{E}\big(g_{A_0}(X_n)\big) - \mathbb{E}\big(g_{A_0}(Z)\big) = 0.$$

On en déduit que

$$\limsup_{n\to\infty} |\mathbb{E}[X_n^r] - \mathbb{E}[Z^r]| \leq 2\varepsilon.$$

Ceci est vrai pour tout $\varepsilon > 0$, donc la suite $\mathbb{E}[X_n^r]$ tend vers $\mathbb{E}[Z^r]$, comme annoncé. □

De la convergence des moments à la convergence en distribution

Dans l'autre sens, la convergence des moments implique-t-elle la convergence en distribution ? Encore une fois, il faut une hypothèse supplémentaire.

Definition 3.2. Soit X une v.a. avec des moments finis. On dit que X est déterminé par ses moments si, pour toute v.a. Y,

$$\Big(\forall\, r \geq 1,\, \mathbb{E}[Y^r] = \mathbb{E}[X^r]\Big) \Rightarrow (X \overset{d}{=} Y).$$

Theorème 3.3 (Méthode des moments)**.** *Soit X_n et X des v.a. telles que, pour tout $r \geq 1$, on ait $\mathbb{E}(X_n^r) \to \mathbb{E}(X^r)$ (en particulier, on suppose tous les moments finis). Supposons de plus que X est déterminé par ses moments. Alors $X_n \xrightarrow{d} X$.*

La condition "X déterminé par ses moments" est clairement nécessaire, donc ce résultat est optimal. Pour prouver le théorème nous aurons besoin de la notion de suite tendue de v.a.

1. Si X_n est une suite de v.a. à valeurs dans un espace Polonais E (i.e. un espace séparable métrique comple) convergeant en loi vers une variable X, alors il existe un espace de probabilités $(\Omega, \mathcal{A}, \mathbb{P})$ et des variables Y_n et Y sur Ω telles que (i) pour tout $n \geq 1$, X_n et Y_n ont la même loi, (ii) X et Y ont la même loi, (iii) Y_n converge p.s. vers Y. Voir la page wikipedia dédiée (écrite par Philippe Chassaing, que je remercie).

Definition 3.4. Une suite (X_n) de v.a. à valeurs dans un espace E (supposé métrique, complet et séparable) est tendu si pour tout $\varepsilon > 0$, il existe une partie compacte $K \subset \mathbb{E}$ tel que $\mathbb{P}(X_n \in K) \geq 1 - \varepsilon$ pour tout n.

Note : K dépend de ε mais pas de n. Si $E = \mathbb{R}$ (cas qui va nous intéresser dans ce chapitre), le compact K peut être choisi de la forme $[-A, A]$, et la condition est équivalente à

$$\forall \varepsilon > 0, \exists A > 0 \text{ t.q. } \mathbb{P}(|X_n| \leq A) \geq 1 - \varepsilon.$$

(Dans ce cas, on parle aussi de suites stochastiquement bornées.) La tension est une sorte de notion de compacité pour des suites de v.a. vis-à-vis de la convergence en distribution. En particulier, on a le résultat suivant (connu sous le nom de théorème de Prokhorov, admis dans ce cours).

Theorème 3.5. *Soit X_n une suite tendue de v.a. Alors il existe une sous-suite $X_{\varphi(n)}$ de X_n qui converge en distribution.*

Corollaire 3.6. *Soit X_n une suite tendue de v.a. et Z un candidat limite. On suppose que toute sous-suite $X_{\varphi(n)}$ de X_n qui converge en distribution converge vers Z. Alors X_n converge vers Z.*

Démonstration. On raisonne par l'absurde. Si X_n ne converge pas vers Z, alors il existe une fonction f continue bornée, $\varepsilon > 0$ et une sous suite $X_{\psi(n)}$ de X_n telles que $|\mathbb{E}(f(X_{\psi(n)}) - \mathbb{E}(f(Z))| \geq \varepsilon$ pour n assez grand. Comme X_n est tendue, $X_{\psi(n)}$ tendue. Il existe donc une sous-suite $X_{\psi(\psi'(n))}$ de $X_{\psi(n)}$ qui converge en distribution. Par hypothèse, elle converge nécessairement vers Z. Mais c'est impossible car $|\mathbb{E}(f(X_{\psi(n)}) - \mathbb{E}(f(Z))| \geq \varepsilon$, et donc en particulier $|\mathbb{E}(f(X_{\psi(\psi'(n))}) - \mathbb{E}(f(Z))| \geq \varepsilon$ pour n assez grand. □

Démonstration du théorème 3.3. On prouve d'abord que X_n est tendue. Par l'inégalité de Markov

$$\mathbb{P}[|X_n| \geq A] = \mathbb{P}[X_n^2 \geq A^2] \leq \frac{\mathbb{E}(X_n^2)}{A^2}.$$

Or $\mathbb{E}(X_n^2)$ converge (vers $\mathbb{E}(X)$, par hypothèse) et est donc bornée. Le fait que X_n est tendu en découle immédiatement.

Considérons maintenant une sous-suite $X_{\varphi(n)}$ de X_n qui converge en distribution vers Y. On veut prouver $Y \stackrel{d}{=} X$. Pour tout entier pair $s > 0$, la suite $\mathbb{E}[|X_{\varphi(n)}|^s] = \mathbb{E}[X_{\varphi(n)}^s]$ est borné (car elle converge vers $\mathbb{E}(X^s)$ par hypothèse). Donc, d'après le théorème 3.1, les moments de $X_{\varphi(n)}$ converge vers ceux de Y. Mais comme $X_{\varphi(n)}$ est une sous-suite de X_n, ils convergent aussi vers les moments de X. Par unicité des moments, on en déduit que Y et X ont les même moments, i.e. $\mathbb{E}[Y^r] = \mathbb{E}[X^r]$ pour tout $r \geq 1$. Comme X est déterminé par ses moments, $Y \stackrel{d}{=} X$. On a donc montré que toute sous-suite de X_n qui converge en distribution converge nécessairement vers X.

D'après le corollaire 3.6 (rappelons que X_n est tendue), on conclut que $X_n \xrightarrow{d} X$. □

Quand une mesure est-elle déterminée par ses moments ?

Il y a une condition suffisante facile à vérifier. Commençons par un lemme.

Lemme 3.7. *Soit X une v.a. aléatoire réelle avec des moments finis. Alors les conditions suivantes sont équivalentes :*

(i) il existe $C > 0$ tel que $|\mathbb{E}(X^r)| \leq C^r r!$ pour tout entier $r \geq 1$;

(ii) il existe $C > 0$ tel que $\mathbb{E}(|X|^r) \leq C^r r!$ pour tout entier $r \geq 1$;

(iii) il existe $u > 0$ tel que $\mathbb{E}(e^{u|X|}) < +\infty$;

(iv) il existe $\varepsilon > 0$ tel que $\mathbb{E}(e^{uX}) < +\infty$ pour tout $|u| < \varepsilon$;

Démonstration. (ii) implique (i) est trivial. Montrons (i) $\Rightarrow$ (ii). Il suffit de considérer le cas ou $r = 2k - 1$ est impair ($k \geq 2$). En utilisant l'inégalité $|x|^{2k-1} \leq 1 + x^{2k}$, valable pour tout x réel, on a

$$E(|X|^{2k-1}) \leq 1 + \mathbb{E}(X^{2k}) \leq 1 + C^{2k}(2k)! \leq (C')^{2k-1}(2k-1)!,$$

pour C' bien choisi.

Pour (ii) $\Leftrightarrow$ (iii), on utilise le théorème de Fubini–Tonelli pour écrire

$$\mathbb{E}(e^{u|X|}) = \mathbb{E}\left(\sum_{k\geq 0} \frac{u^k |X|^k}{k!}\right) = \sum_{k\geq 0} \frac{u^k}{k!}\mathbb{E}(|X|^k).$$

L'implication (ii) $\Rightarrow$ (iii) se prouve alors en choisissant $u < 1/C$ et en bornant $\frac{u^k}{k!}\mathbb{E}(|X|^k)$ par la série géométrique convergente $(uC)^k$. Dans le sens (iii) $\Rightarrow$ (ii), on utilise le fait que le terme général d'une série convergente tend vers 0, et est donc borné. On a donc $\frac{u^k}{k!}\mathbb{E}(|X|^k) \leq A$ pour $u > 0$ et $A > 0$ bien choisis, ce qui implique (ii).

(iii) $\Leftrightarrow$ (iv) est facile en écrivant $e^{uX} \leq e^{\varepsilon|X|}$ pour $|u| \leq \varepsilon$, et, dans l'autre sens, $e^{u|X|} \leq e^{u\,X} + e^{-u\,X}$. □

Theorème 3.8. *Soit X une v.a. aléatoire réelle avec des moments finis. Si les conditions (i)-(iv) du lemme sont vérifiées, alors X est déterminé par ses moments.*

Première démonstration (sans analyse complexe). Soit Y une v.a. avec les mêmes moments que X, i.e. pour tout $r \geq 1$, on suppose que $\alpha_r := \mathbb{E}(X^r) = \mathbb{E}(Y^r) < \infty$. L'idée de la preuve consiste à exprimer la fonction caractéristique de X et Y en fonction de leurs moments, pour utiliser le fait que la fonction caractéristique détermine la loi. Informellement, on a

$$e^{itX} \sim \sum_{k=0}^{\infty} (itX)^k/k!$$

et donc, en prenant l'espérance :

$$\varphi_X(t) \sim \sum_{k=0}^{\infty} (it)^k\, \alpha_k/k!$$

Rigoureusement, on peut contrôler le terme d'erreur par une expansion de Taylor :

$$\left| e^{itX} - \sum_{k=0}^{N} (itX)^k/k! \right| \leq \frac{|tX|^{N+1}}{(N+1)!}$$

En prenant l'espérance

$$\left| \varphi_X(t) - \sum_{k=0}^{N} (it)^k \, \alpha_k/k! \right| \leq \frac{|t|^{N+1}}{(N+1)!} \mathbb{E}[|X|^{N+1}] \leq (C|t|)^{N+1},$$

où on a utilisé que X vérifie la condition (ii) ci-dessus. Pour $|t| < \rho := 1/C$, le membre de droite tend vers 0 et

$$\varphi_X(t) = \sum_{k=0}^{\infty} (it)^k \, \alpha_k/k!.$$

On peut faire le même raisonnement avec Y (comme Y a les mêmes moments que X, elle vérifie la condition (i) ci-dessus, et donc aussi la condition (ii), même si les moments de $|Y|$ ne sont a priori pas les mêmes que ceux de $|X|$). On obtient que $\varphi_X(t) = \varphi_Y(t)$ pour $|t| \leq \rho$. Ce n'est malheureusement pas suffisant pour conclure que $X \stackrel{d}{=} Y$, il faut étendre l'égalité à $\mathbb{R}$ tout entier.

Écrivons maintenant, pour t_0 dans $\mathbb{R}$ fixé,

$$\left| e^{it_0 X} \left(e^{itX} - \sum_{k=0}^{N} (itX)^k/k! \right) \right| \leq \frac{|tX|^{N+1}}{(N+1)!}$$

de sorte que, pour $|t| < \rho$

$$\varphi_X(t_0 + t) = \sum_{k=0}^{\infty} \mathbb{E}(e^{it_0 X} X^k) \, (it)^k/k!. \tag{2}$$

Notons $c_{t_0,k}(X)$ l'espérance dans le membre de droite, et adoptons une notation similaire avec X remplacé par Y. L'équation (3) est un développement de Taylor de φ_X autour de t_0 : les $c_{t_0,k}(X)$ (resp. $c_{t_0,k}(Y)$) sont donc les dérivées successives de φ_X (resp. φ_Y) en t_0, divisées par i^k. Pour $|t_0| < \rho$, on $\varphi_X \equiv \varphi_Y$ localement autour de t_0, et donc

$$c_{t_0,k}(X) = c_{t_0,k}(Y)$$

En revenant à (2) et à l'égalité analogue pour Y, on en déduit que pour $|t_0| < \rho$ et $|t| < \rho$,

$$\varphi_X(t_0 + t) = \varphi_Y(t_0 + t).$$

Autrement dit, φ_X et φ_Y coïncident sur $(-2\rho, 2\rho)$. En itérant le raisonnement, on montre que φ_X et φ_Y coincident sur $(-3\rho, 3\rho)$, puis sur $(-4\rho, 4\rho)$, etc. Finalement, les fonctions caractéristiques φ_X et φ_Y coincident sur toute la droite réelle, ce qui prouve $X \stackrel{d}{=} Y$. □

Deuxième démonstration (avec analyse complexe). Comme X vérifie la condition (ii) ci-dessus, il existe C tel que $\mathbb{E}[|X|^k] < C^k k!$. Notons que pour t_0 dans $\mathbb{R}$ et z dans $\mathbb{C}$, on a

$$e^{(z+it_0)X} = e^{it_0X}e^{zX} = \sum_{k\geq 0} \frac{z^k(e^{it_0X}X^k)}{k!}.$$

Lorsque $|z| < 1/C$ la série

$$\sum_{\boldsymbol{k}\geq 0} \mathbb{E}\left[\left|\frac{z^k(e^{it_0X}X^k)}{k!}\right|\right] \leq \sum_{\boldsymbol{k}\geq 0} \frac{|z|^k\mathbb{E}[|X|^k]}{k!} \leq \sum_{\boldsymbol{k}\geq 0}(C|z|)^k$$

est sommable et on peut écrire

$$\mathbb{E}[e^{(z+it_0)X}] = \sum_{k\geq 0} \frac{z^k\mathbb{E}[e^{it_0X}X^k]}{k!}. \tag{3}$$

Cela montre que la fonction $\omega \mapsto \mathbb{E}[e^{\omega X}]$ est bien définie et analytique sur le disque de centre it_0 et de rayon $1/C$. Ceci étant vrai pour tout t_0 réel, la fonction $\omega \mapsto \mathbb{E}[e^{\omega X}]$ est analytique sur la bande $\{|\Re\omega| < 1/C\}$.

Soit Y une v.a. avec les mêmes moments que X, i.e. pour tout $r \geq 1$, on suppose que $\alpha_r := \mathbb{E}(X^r) = \mathbb{E}(Y^r) < \infty$. Comme X verifie la condition (i) ci-dessus, Y aussi. On en deduit que Y verifie aussi la condition (ii) du lemme, i.e. il existe C' tel que $\mathbb{E}[|Y|^k] < (C')^k k!$. Le raisonnement ci-dessus implique que $\omega \mapsto \mathbb{E}[e^{\omega Y}]$ est analytique sur la bande $\{|\Re\omega| < 1/C'\}$.

De plus les développements en série de $\mathbb{E}[e^{\omega X}]$ et $\mathbb{E}[e^{\omega Y}]$ autour de 0 sont donnés par l'équation (3) pour $t_0 = 0$ (et son équivalent pour Y) :

$$\mathbb{E}[e^{zX}] = \sum_{k\geq 0} \frac{z^k\mathbb{E}[X^k]}{k!}, \quad \mathbb{E}[e^{zY}] = \sum_{k\geq 0} \frac{z^k\mathbb{E}[Y^k]}{k!}.$$

Les moments $\mathbb{E}[X^k]$ et $\mathbb{E}[Y^k]$ étant égaux, les fonctions $\mathbb{E}[e^{\omega X}]$ et $\mathbb{E}[e^{\omega Y}]$ coïncident sur un disque autour de 0. Comme ce sont des fonctions analytiques, elles coïncident sur l'intersection de leur domaine qui est connexe. En particulier elle coïncide sur $i\mathbb{R}$, i.e. $\mathbb{E}[e^{itX}] = \mathbb{E}[e^{itY}]$ pour tout t reel. Comme la fonction caractéristique d'une v.a. réelle détermine sa loi, on en déduit que X et Y ont la même loi, ce qui conclut la preuve du théorème. □

Notons que la méthode des moments ne s'applique pas aux lois dont certains moments sont infinis, comme la loi de Cauchy $\mathbb{P}(dx) = \frac{1}{\pi}\frac{1}{x^2+1}$ (dans ce cas, tous les moments sont infinis). Il existe aussi des mesures dont tous les moments sont finis, et qui ne sont pas déterminés par les moments (voir exercice).

3.2 Quelques applications

3.2.1 Théorème centrale limite

Comme première application de la méthode des moments, donnons une preuve du théorème centrale limite dans le cas de v.a. dont tous les moments

sont finis.

On commence par calculer les moments d'une variance Gaussienne.

Theorème 3.9. *Soit Z une v.a. Gaussienne centrée réduite. Alors*

$$\mathbb{E}[Z^k] = \begin{cases} 1 \cdot 3 \cdots (k-1) & \text{si } k \text{ est pair ;} \\ 0 & \text{sinon.} \end{cases}$$

Démonstration. On a

$$\mathbb{E}[Z^k] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^k e^{-x^2/2} dx.$$

Pour k impair, l'intégrale vaut 0 par symétrie. Pour k pair, on fait une intégration par partie : on pose $u = x^{k-1}$, ce qui implique $du = (k-1)x^{k-2}dx$, et $dv = xe^{-x^2/2}dx$, soit $v = -e^{-x^2/2}$. On trouve

$$\int_{-\infty}^{+\infty} x^k e^{-x^2/2} dx = (k-1) \int_{-\infty}^{+\infty} x^{k-2} e^{-x^2/2} dx,$$

(les termes aux bords sont nuls), ce qui implique $\mathbb{E}[Z^k] = (k-1)\mathbb{E}[Z^{k-2}]$. En utilisant l'égalité triviale $\mathbb{E}[Z^0] = 1$, une récurrence immédiate prouve que, pour k pair,

$$\mathbb{E}[Z^k] = (k-1)\mathbb{E}[Z^{k-2}] = (k-1)(k-3)\mathbb{E}[Z^{k-4}] = \cdots = (k-1)(k-3)\cdots 3 \cdot 1. \ \square$$

Theorème 3.10. *Soit X_1, X_2, ... des v.a. i.i.d. telles que, pour tout $r \geq 1$, $\mathbb{E}[|X_1|^r] < +\infty$. Posons $\sigma^2 = E[X_1^2]$ et $S_n = \sum_{i=1}^n X_i$. Alors $\tilde{S}_n := \frac{S_n - \mathbb{E}[S_n]}{\sigma\sqrt{n}}$ converge en distribution vers une loi Gaussienne centrée réduite Z.*

Comme vous l'avez probablement déjà vu dans un cours précécent, le théorème est valide sous l'hypothèse plus faible $\mathbb{E}[X_1^2] < +\infty$.

Démonstration. On écrit

$$\mathbb{E}[\tilde{S}_n^k] = \frac{1}{\sigma^k n^{k/2}} \sum_{i_1,\ldots,i_k} \mathbb{E}\left(\prod_{j=1}^k (X_{i_j} - \mathbb{E}(X_{i_j})) \right). \tag{4}$$

À une suite $\boldsymbol{i} = (i_1, \ldots, i_k)$, on associe de manière unique une partition $\pi_{\boldsymbol{i}}$ de l'ensemble $[\![1, k]\!]$ par la condition $i_j = i_h$ si et seulement si j et h sont dans la même part de π. De plus, pour une liste $\boldsymbol{i}$ et une partition π, on écrit $\boldsymbol{i} \in \mathcal{I}_\pi$ si $\pi_{\boldsymbol{i}} = \pi$. En notant $m_1, \ldots, m_\ell$ les tailles des parts de π, le caractère i.i.d. des X_i implique que, si $\boldsymbol{i} \in \mathcal{I}_\pi$,

$$\mathbb{E}\left(\prod_{j=1}^k (X_{i_j} - \mathbb{E}(X_{i_j})) \right) = \prod_{t=1}^{\ell} \mathbb{E}\left(X_1 - \mathbb{E}(X_1)\right)^{m_t}. \tag{5}$$

En particulier cela ne dépend que de π. On peut réécrire (4) sous la forme

$$\mathbb{E}[\tilde{S}_n^k] = \frac{1}{\sigma^k n^{k/2}} \sum_{\pi} \left[\#\{\boldsymbol{i} : \boldsymbol{i} \in \mathcal{I}_\pi\} \prod_{t=1}^{\ell} \mathbb{E}\left(X_1 - \mathbb{E}(X_1)\right)^{m_t} \right]. \tag{6}$$

Le comportement de la quantité entre crochets dépend de la combinatoire de la partition π.

— Si π contient un singleton, i.e. $m_t = 1$ pour un certain t, alors

$$\mathbb{E}\left(X_1 - \mathbb{E}(X_1)\right)^{m_t} = 0$$

et le terme correspondant à π dans (6) vaut 0.

— Le nombre de suites $\boldsymbol{i}$ telle que $\boldsymbol{i} \in \mathcal{I}_\pi$ est $n(n-1)\dots(n-\ell+1) = O(n^\ell)$. De plus on peut borner le produit des espérances en utilisant la croissance des normes L^p :

$$\left| \prod_{t=1}^{\ell} \mathbb{E}\left[\left(X_1 - \mathbb{E}(X_1)\right)^{m_t}\right] \right| \leq \prod_{t=1}^{\ell} \|X_1 - \mathbb{E}(X_1)\|_{m_t}^{m_t}$$
$$\leq \prod_{t=1}^{\ell} \|X_1 - \mathbb{E}(X_1)\|_k^{m_t} = \|X_1 - \mathbb{E}(X_1)\|_k^{\sum_{t=1}^{\ell} m_t} = \|X_1 - \mathbb{E}(X_1)\|_k^k.$$

Cette borne supérieure est indépendante de π et de n, le terme associé à π dans (6) est donc $O(n^\ell)$. Après division par $n^{k/2}$, les contributions des termes avec $\ell < k/2$ sont donc négligeables.

Il suffit donc de considérer les partitions sans singletons avec au moins $k/2$ parts, c'est à dire les partitions de $[\![1, k]\!]$ en paires. Si k est impair, il n'y en a pas et on a $\mathbb{E}[\tilde{S}_n^k] = o(1)$. Si k est pair, il y en a $(k-1)(k-3)\cdots 3 \cdot 1$, et la contribution de chacune de ces partitions est

$$n(n-1)\dots(n-\tfrac{k}{2}+1) \prod_{t=1}^{k/2} \mathbb{E}\left(X_1 - \mathbb{E}(X_1)\right)^2 \sim n^{k/2}\sigma^k.$$

On obtient donc

$$\lim_{n\to+\infty} \mathbb{E}[\tilde{S}_n^k] = (k-1)(k-3)\cdots 3 \cdot 1.$$

Dans les deux cas (k pair et k impair), on a prouvé

$$\lim_{n\to+\infty} \mathbb{E}[\tilde{S}_n^k] = \mathbb{E}[Z^k],$$

ce qui implique $\tilde{S}_n \to Z$, en distribution, par la méthode des moments. □

3.2.2 Exemple de convergence vers une loi de Poisson

Au lieu de montrer la convergence des moments, dans le cas d'une loi de Poisson, c'est souvent plus facile de montrer la convergence des moments factoriels, définis par $M_{(r)}(X_n) := \mathbb{E}\big[X(X-1)\dots(X-k+1)\big]$.

Proposition 3.11. *Soit X_n et X des v.a. réelles, dont tous les moments sont finis. Supposons que X est déterminé par ses moments, et que, quelque soit $r \geq 1$, $M_{(r)}(X_n)$ converge vers $M_{(r)}(X)$. Alors X_n converge en distribution vers X.*

Démonstration. Soit $P_r(x) = x(x-1)\cdots(x-r+1)$. La famille $(P_r)_{r\geq 0}$ forme une base de l'espace des polynômes. Donc pour tout $k \geq 1$, x^k est une combinaison linéaire de P_r. Comme $\mathbb{E}[P_r(X_n)]$ converge vers $\mathbb{E}[P_r(X)]$ pour tout $r \geq 0$, par linéarité, on obtient que $\mathbb{E}[X_n^k]$ tend vers $\mathbb{E}[X^k]$ pour tout $k \geq 1$. On conclut par la méthode des moments. □

Lemme 3.12. *Soit Y une loi de Poisson de paramètre λ. Alors $M_{(r)}(Y) = \lambda^r$.*

Démonstration.

$$M_{(r)}(Y) = \sum_{n\geq 0} n(n-1)\cdots(n-r+1)\frac{\lambda^n e^{-\lambda}}{n!} = \lambda^r e^{-\lambda}\sum_{n\geq r}\frac{\lambda^{n-r}}{(n-r)!} = \lambda^r. \quad \square$$

Lemme 3.13. *Si $X = \sum_{i=1}^N \mathbf{1}[A_i]$ pour certains évènements A_i, alors*

$$M_{(r)}(X) = \sum_{\substack{i_1,\ldots,i_r\leq N \\ distincts}} \mathbb{P}[A_{i_1}\wedge\cdots\wedge A_{i_r}].$$

Démonstration. On part du membre de droite

$$\sum_{\substack{i_1,\ldots,i_r\leq N \\ \text{distincts}}} \mathbb{P}[A_{i_1}\wedge\cdots\wedge A_{i_r}] = \mathbb{E}\left[\sum_{\substack{i_1,\ldots,i_r\leq N \\ \text{distincts}}} \mathbf{1}[A_{i_1}\wedge\cdots\wedge A_{i_r}]\right].$$

Prenons un point ω de l'espace de probabilités tels que $X(\omega) = k$. Soit J l'ensemble $\{i, \omega \in A_i\}$ des indices des évènements réalisés. On a alors $\mathbf{1}[A_{i_1}\wedge\cdots\wedge A_{i_r}] = 1$ si et seulement si $(i_1,\ldots,i_r)$ est une suite d'élements distincts de J. Il y a exactement $k(k-1)\cdots(k-r+1)$ telles suites vérifiant la condition de contenir des éléments distincts imposée dans la somme. Donc si $X(\omega) = k$, la somme d'indicatrice ci-dessus vaut $k(k-1)\cdots(k-r+1)$. Autrement dit, cette somme vaut $X(X-1)\ldots(X-k+1)$, ce qui conclut la démonstration. □

Theorème 3.14. *Soit X_n le nombre de points fixes d'une permutation aléatoire σ_n. Alors X_n converge vers une loi de Poisson de paramètre 1.*

Démonstration. D'après les deux premiers lemmes ci-dessus, il suffit de montrer que, si Y suit une loi de Poisson de paramètre 1, on a, pour tout $k \geq 1$,

$$\mathbb{E}\big[X_n(X_n-1)\cdots(X_n-k+1)\big] \to \mathbb{E}\big[Y(Y-1)\cdots(Y-k+1)\big] = 1.$$

Mais, en utilisant le troisième lemme, on a

$$\mathbb{E}[X_n(X_n-1)\cdots(X_n-k+1)] = \sum_{\substack{i_1,\ldots,i_k\leq N \\ \text{distincts}}} \mathbb{P}[A_{i_1}\wedge\cdots\wedge A_{i_k}], \qquad (7)$$

où A_i est l'évènement « $\sigma_n(i) = i$ ». Or, quand i_1, …, i_k sont distincts,

$$\mathbb{P}[A_{i_1} \wedge \cdots \wedge A_{i_k}] = \mathbb{P}[\sigma(i_1) = i_1 \wedge \cdots \wedge \sigma(i_k) = i_k] = \frac{1}{n(n-1)\cdots(n-k+1)},$$

où la dernière égalité est une application du théorème 1.3. On en déduit que les $n(n-1)\ldots(n-k+1)$ termes de la somme de l'équation (7) sont tous égaux à $\frac{1}{n(n-1)\cdots(n-k+1)}$ et on obtient

$$\mathbb{E}[X_n(X_n-1)\cdots(X_n-k+1)] = n(n-1)\cdots(n-k+1)\,\frac{1}{n(n-1)\cdots(n-k+1)} = 1,$$

ce qui finit la démonstration. □

Note : dans cet exemple, les X_n et la loi limite Y prennent leurs valeurs dans un ensemble discret $\mathbb{Z}$, la convergence en distribution est donc équivalente à : pour tout j dans $\mathbb{Z}$, la probabilité $\mathbb{P}[X_n = j]$ tend vers $\mathbb{P}[Y = j] = e^{-1}/j!$.

3.2.3 Nombre de diviseurs d'un entier aléatoire

Soit $n \geq 1$, prenons x_n uniformément au hasard dans $[\![1, n]\!]$. On s'intéresse au nombre de diviseurs premiers de x_n, que l'on notera $\nu(x_n)$. On commence par un résultat technique.

Lemme 3.15. *La quantité* $\tilde{H}_n := \sum_{p \text{ premier}}^{p \leq n} \frac{1}{p}$ *tend vers l'infini.*

Esquisse de preuve. Supposons que $\tilde{H}_n$ soit borné. Alors la série

$$\sum_{p \text{ premier}} \log(1 - \frac{1}{p})$$

est convergente, et donc le produit $\prod_p \frac{1}{1-\frac{1}{p}}$ est convergent aussi. Mais, d'après la formule d'Euler pour la fonction zeta, $\prod_p \frac{1}{1-\frac{1}{p}} = \sum_n \frac{1}{n}$, et donc $\sum_n \frac{1}{n}$ serait convergent : on aboutit à une contradiction. □

Note : en fait, il est connu que $\tilde{H}_n = \log(\log(n)) + M + o(1)$, pour un certain M réel, appelé constante de Meissel-Mertens ($M \approx 0.2615$).

Theorème 3.16 (Erdős-Kac theorem)**.**

$$\frac{\nu(x) - \tilde{H}_n}{\sqrt{\tilde{H}_n}} \to_d \mathcal{N}(0, 1).$$

Démonstration. Posons $Y_n = \nu(x_n)$, on a $Y_n = \sum_{p \leq n} \mathbf{1}[A_p]$, où A_p est l'événement « p divise x_n ». Ici, et dans ce qui suit, les sommes sur des indices p sont implicitement prises sur les nombres *premiers* p. Notons que

$$\mathbb{P}[A_p] = \frac{1}{n}\sum_{x=1}^{n} \mathbf{1}[p|x] = \frac{1}{n}\left\lfloor \frac{n}{p} \right\rfloor = \frac{1}{p} + O(n^{-1}),$$

avec un grand O uniforme en p (l'erreur est toujours inférieur à $1/n$). En sommant sur p, on obtient $\mathbb{E}[Y_n] = \tilde{H}_n + O(1)$.

Pour calculer les moments d'ordre supérieur, il faut tronquer la somme, sinon les termes d'erreurs vont être trop grands. Soit ε_n une suite qui tendant vers l'infini, moins vite que $\sqrt{\tilde{H}_n}$. On approche Y_n par $Z_n := \sum_{p \leq n^{1/\varepsilon_n}} \mathbf{1}[A_p]$. On a $\mathbb{E}[Z_n] = \tilde{H}_{n^{1/\varepsilon_n}} + O(1)$.

Comme il ne peut y avoir plus de ε_n nombres premiers $p \geq n^{1/\varepsilon_n}$ divisant x_n, on a $Z_n \leq Y_n \leq Z_n + \varepsilon_n$. Cela implique

$$\mathbb{E}[Z_n] - \mathbb{E}[Y_n] = \tilde{H}_n - \tilde{H}_{n^{1/\varepsilon_n}} + O(1) \leq \varepsilon_n.$$

Comme $\varepsilon_n = o(\tilde{H}_n)$, cela signifie que

$$\mathbb{E}[Z_n] = \tilde{H}_{n^{1/\varepsilon_n}} + O(1) = \tilde{H}_n + O(\varepsilon_n).$$

Par conséquent, montrer le théorème revient à montrer que

$$\tilde{Z}_n := \frac{Z_n - \mathbb{E}[Z_n]}{\sqrt{\tilde{H}_{n^{1/\varepsilon_n}}}}$$

converge en distribution vers une Gaussienne.

On écrit

$$Z_n - \mathbb{E}[Z_n] = \sum_{p \leq n^{1/\varepsilon_n}} (\mathbf{1}[A_p] - \tfrac{1}{p}).$$

Soit B_p des variables de Bernoulli indépendantes de paramètre $1/p$, on considère aussi leur version centrée $B'_p = B_p - \frac{1}{p}$. Pour $p_1, ..., p_k \leq n$ donnés, on a

$$\begin{aligned}\mathbb{E}\Big[\mathbf{1}[A_{p_1}] \cdots \mathbf{1}[A_{p_k}]\Big] &= \frac{1}{n} \sum_{x=1}^{n} \mathbf{1}[(p_1|x) \wedge \cdots \wedge (p_k|x)] = \frac{1}{n} \lfloor \frac{n}{p_1 \cdots p_k} \rfloor \\ &= \frac{1}{p_1 \cdots p_k} + O(n^{-1}) = \mathbb{E}\Big[B_{p_1} \cdots B_{p_k}\Big] + O(n^{-1}).\end{aligned}$$

On en déduit une égalité similaire pour les versions centrées :

$$\mathbb{E}\Big[(\mathbf{1}[A_{p_1}] - \tfrac{1}{p_1}) \cdots (\mathbf{1}[A_{p_k}] - \tfrac{1}{p_k})\Big] = \mathbb{E}\Big[B'_{p_1} \cdots B'_{p_k}\Big] + O(n^{-1}).$$

Cela donne, pour $k \geq 1$,

$$\begin{aligned}\mathbb{E}\big[(Z_n - \mathbb{E}[Z_n])^k\big] &= \sum_{p_1, \ldots, p_k \leq n^{1/\varepsilon_n}} \mathbb{E}\Big[(\mathbf{1}[A_{p_1}] - \tfrac{1}{p_1}) \cdots (\mathbf{1}[A_{p_k}] - \tfrac{1}{p_k})\Big] \\ &= \sum_{p_1, \ldots, p_k \leq n^{1/\varepsilon_n}} \mathbb{E}\Big[B'_{p_1} \cdots B'_{p_k}\Big] + O(n^{-1} n^{k/\varepsilon_n}).\end{aligned}$$

Le terme d'erreur est négligeable (note : si on n'avait pas tronqué la somme, comme il y a environ $n/\log(n)$ nombres premiers $\leq n$, on aurait un terme d'erreur de l'ordre $O(n^{k-1}/\log(n)^k)$, bien supérieur au terme principal).

Pour la terme principal, comme dans la preuve du théorème 3.10, on coupe selon la partition π de $[\![1,k]\!]$ décrivant l'égalité des p. Pour π donnée avec ℓ parts de tailles $m_1,\dots,m_\ell$,

$$\sum_{\substack{p_1,\dots,p_k\le n^{1/\varepsilon_n}\\(p_1,\dots,p_k)\in\mathcal{I}_\pi}} \mathbb{E}\Big[B'_{p_1}\cdots B'_{p_k}\Big] = \sum_{\substack{q_1,\dots,q_\ell\le n^{1/\varepsilon_n}\\ \text{distinct}}} \mathbb{E}\Big[(B'_{q_1})^{m_1}\Big]\cdots\mathbb{E}\Big[(B'_{q_\ell})^{m_\ell}\Big].$$

Ici, q_1, …, q_ℓ sont des nombres premiers correspondant aux p_1,…,p_k de la somme précédente, après effacement des répétitions. Comme précédemment si un des m_i vaut 1, alors tous les termes de la somme valent 0. Pour $m\ge 2$, on peut estimer $\mathbb{E}\Big[(B'_q)^m\Big] = \frac{1}{q}(1-\frac{1}{q})^m - (1-\frac{1}{q})\frac{1}{q^m} \sim 1/q$ quand q tend vers l'infini, de telle sorte que

$$\sum_{q_1,\dots,q_\ell\le n^{1/\varepsilon_n}} \mathbb{E}\Big[(B'_{q_1})^{m_1}\Big]\cdots\mathbb{E}\Big[(B'_{q_\ell})^{m_\ell}\Big] \sim \sum_{q_1,\dots,q_\ell\le n^{1/\varepsilon_n}} \frac{1}{q_1,\dots,q_\ell} = \tilde{H}^\ell_{n^{1/\varepsilon_n}}.$$

De plus, on peut montrer que les suites $q_1,\dots,q_e ll$ avec répétition ont une contribution négligeable. On a donc

$$\mathbb{E}\big[(Z_n-\mathbb{E}[Z_n])^k\big] = \sum_\pi \tilde{H}^{\#\pi}_{n^{1/\varepsilon_n}}(1+o(1)) + O(n^{-1}n^{k/\varepsilon_n}),$$

où la somme est prise sur les partitions π de $[\![1,k]\!]$ et $\#\pi$ désigne le nombre de parts de π.. Les termes contribuant le plus sont les partitions sans singletons avec le plus de parts possibles, i.e. les partitions en paires. En se rappelant qu'il y a $1\cdot 3\cdots(k-1)$ telles partitions pour k pair, et aucune pour k impair, on obtient

$$\mathbb{E}\big[(Z_n-\mathbb{E}[Z_n])^k\big] = \begin{cases} 1\cdot 3\cdots(k-1)\,\tilde{H}^{k/2}_{1/\varepsilon_n} & \text{si } k \text{ est pair};\\ o(\tilde{H}^{k/2}_{1/\varepsilon_n}) & \text{si } k \text{ est impair.}\end{cases}$$

En divisant par $H^{k/2}_{1/\varepsilon_n}$, on voit que les moments de $\tilde{Z}_n$ tende vers ceux d'une Gaussienne centrée réduite, ce qui conclut la démonstration. □