

Distance entre deux distributions de probabilité sur \mathbb{N}

1 — Nombre de points fixes d'une permutation

Q 1. D'après le cours, $\#\mathfrak{S}_n = n!$. On en déduit que $0 \leq \frac{d_n}{n!} \leq 1$ pour tout $n \in \mathbb{N}$. La série entière $\sum 1 \times x^n$ étant de rayon de convergence 1, $\sum \frac{d_n}{n!} x^n$ est de rayon de convergence $R \geq 1$.

Q 2. Une permutation de $\llbracket 1, n \rrbracket$ est entièrement et univoquement déterminée par l'ensemble de ses points fixes et par le dérangement qu'elle induit sur l'ensemble des autres points. Si l'on veut qu'il y ait exactement k points fixes, il y a $\binom{n}{k}$ choix possibles de l'ensemble F des points fixes et, par définition, d_{n-k} dérangements de $\llbracket 1, n \rrbracket \setminus F$, d'où $\binom{n}{k} d_{n-k}$ éléments de \mathfrak{S}_n possédant exactement k points fixes. On en déduit que

$$\mathbf{P}_n(X_n = k) = \frac{1}{n!} \binom{n}{k} d_{n-k} = \frac{d_{n-k}}{k! (n-k)!}.$$

Q 3. On a vu à la question 1 que $s(x)$ est (au moins) définie pour $x \in]-1, 1[$. Pour un tel réel x , on effectue le produit de Cauchy des deux séries entières, ce qui est possible car l'exponentielle est de rayon de convergence infini (donc au moins 1). Il vient

$$\begin{aligned} e^x s(x) &= \left(\sum_{n=0}^{\infty} \frac{1}{n!} x^n \right) \left(\sum_{n=0}^{\infty} \frac{d_n}{n!} x^n \right) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^n \frac{1}{k!} \times \frac{d_{n-k}}{(n-k)!} \right) x^n \\ &= \sum_{n=0}^{\infty} \left(\sum_{k=0}^n \mathbf{P}_n(X_n = k) \right) x^n \stackrel{(1)}{=} \sum_{n=0}^{\infty} 1 \times x^n = \frac{1}{1-x}, \end{aligned}$$

en notant (1) que $X_n(\Omega) \subset \llbracket 0, n \rrbracket$, ce qui permet de remplacer la somme des probabilités par 1.

Le rayon de convergence d'un produit de Cauchy de deux séries entières est minoré par le plus petit des rayons de convergence des séries dont on fait le produit, d'où $R \leq 1$. L'inégalité $R \geq 1$ ayant été déjà établie, on a $R = 1$.

Q 4. Pour $x \in]-1, 1[$, la relation précédente donne $(1-x)s(x) = e^{-x}$, d'où par unicité du développement en série entière quand le rayon de convergence est non nul, puis par sommation télescopique,

$$\begin{aligned} (1-x)s(x) &\stackrel{(l.h.s.)}{=} \sum_{n=0}^{\infty} \frac{d_n}{n!} x^n - \sum_{n=0}^{\infty} \frac{d_n}{n!} x^{n+1} = d_0 + \sum_{n=1}^{\infty} \left(\frac{d_n}{n!} - \frac{d_{n-1}}{(n-1)!} \right) x^n = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} x^n \stackrel{(r.h.s.)}{=} e^{-x} \iff \\ &d_0 = 1 \quad \& \quad \forall n \geq 1: \frac{d_n}{n!} - \frac{d_{n-1}}{(n-1)!} = \frac{(-1)^n}{n!} \quad \therefore \\ \forall n \geq 1: \frac{d_n}{n!} &= d_0 + \sum_{k=1}^n \left(\frac{d_k}{k!} - \frac{d_{k-1}}{(k-1)!} \right) = 1 + \sum_{k=1}^n \frac{(-1)^k}{k!} = \sum_{k=0}^n \frac{(-1)^k}{k!}. \end{aligned}$$

Alternativement, on peut écrire $s(x) = \frac{e^{-x}}{1-x}$ et effectuer le produit de Cauchy :

$$\sum_{n=0}^{\infty} \frac{d_n}{n!} x^n = \left(\sum_{n=0}^{\infty} \frac{(-1)^n}{n!} x^n \right) \left(\sum_{n=0}^{\infty} x^n \right) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^n \frac{(-1)^k}{k!} \times 1 \right) x^n \quad \therefore \quad \frac{d_n}{n!} = \sum_{k=0}^n \frac{(-1)^k}{k!}$$

par unicité du développement en série entière, le rayon de convergence étant non nul ainsi qu'établi à la première question.

Q 5. D'après les questions précédentes,

$$\mathbf{P}_n(X_n = k) \stackrel{(Q2)}{=} \frac{d_{n-k}}{k!(n-k)!} \stackrel{(Q4)}{=} \frac{1}{k!} \sum_{i=0}^{n-k} \frac{(-1)^i}{i!}.$$

La somme est non nulle, sauf pour $k = n - 1$. Ainsi, $X_n(\mathfrak{S}_n) = \llbracket 0, n \rrbracket \setminus \{n - 1\}$.

Q 6. La variable aléatoire U_i prend (au plus, pour l'instant) les valeurs 0 et 1, donc suit une loi de Bernoulli de paramètre $\mathbf{P}(U_i = 1)$. Soient $\sigma \in \mathfrak{S}_n$ et $i \in \llbracket 1, n \rrbracket$. Alors, $\sigma(i) = i$ si, et seulement si, σ induit une permutation de $\llbracket 1, n \rrbracket \setminus \{i\}$. Le nombre de telles permutations étant $(n-1)!$, il vient $\mathbf{P}(U_i = 1) = \frac{(n-1)!}{n!} = \frac{1}{n}$. Ainsi, $U_i \sim \mathcal{B}\left(\frac{1}{n}\right)$.

Soit $(i, j) \in \llbracket 1, n \rrbracket^2$ avec $i \neq j$. Alors, $U_i U_j$ est une variable aléatoire de Bernoulli, prenant la valeur 1 si, et seulement si, i et j sont des points fixes. Le même raisonnement donne

$$\mathbf{P}_n(U_i U_j = 1) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)} \quad \therefore \quad U_i U_j \sim \mathcal{B}\left(\frac{1}{n(n-1)}\right).$$

Les variables U_i ne sont **pas** indépendantes (sinon, les v.a. $U_i U_j$ suivraient une loi de Bernoulli de paramètre $1/n^2$).

Q 7. On a $X_n = \sum_{i=1}^n U_i$, d'où, par linéarité de l'espérance,

$$\mathbf{E}(X_n) = \sum_{i=1}^n \mathbf{E}(U_i) = \sum_{i=1}^n \frac{1}{n} = 1.$$

Par défaut d'indépendance des U_i , le calcul de la variance de X_n nécessite de passer par un calcul de covariance. Par symétrie, $\mathbf{cov}(U_i, U_j)$ ne dépend pas de (i, j) si $i \neq j$ et vaut $\mathbf{E}(U_1 U_2) - \mathbf{E}(U_1)^2$.

$$\begin{aligned} \mathbf{V}(X_n) &= \sum_{i=1}^n \mathbf{V}(U_i) + 2 \sum_{1 \leq i < j \leq n} \mathbf{cov}(U_i, U_j) = n \mathbf{V}(U_1) + 2 \binom{n}{2} \mathbf{cov}(U_1, U_2) \\ &\stackrel{(Q6)}{=} n \times \frac{1}{n} \left(1 - \frac{1}{n}\right) + n(n-1) \left(\frac{1}{n(n-1)} - \frac{1}{n^2}\right) = \frac{n-1}{n} + \frac{n-1}{n(n-1)} = 1. \end{aligned}$$

De manière un peu différente, on peut écrire $\mathbf{V}(X_n) = \mathbf{E}(X_n^2) - \mathbf{E}(X_n)^2 = \mathbf{E}(X_n^2) - 1$. De plus,

$$\mathbf{E}(X_n^2) = \mathbf{E} \left[\sum_{i=1}^n X_i^2 + \sum_{1 \leq i \neq j \leq n} X_i X_j \right] = n \mathbf{E}(X_1^2) + (n^2 - n) \mathbf{E}(X_1 X_2) = n \times \frac{1}{n} + n^2 - (n^2 - n) \times \frac{1}{n(n-1)} = 2,$$

en utilisant que $X_1^2 = X_1$ et la linéarité de l'espérance. On retrouve $\mathbf{V}(X_n) = 1$.

Q 8. À k fixé,

$$\mathbf{P}_n(X_n = k) \stackrel{(Q5)}{=} \frac{1}{k!} \sum_{i=0}^{n-k} \frac{(-1)^i}{i!} \xrightarrow{n \rightarrow \infty} \frac{1}{k!} \sum_{i=0}^{\infty} \frac{(-1)^i}{i!} = \frac{1}{e k!} = y_k.$$

On reconnaît $y_k = \mathbf{P}(Y = k)$ pour $Y \sim \mathcal{P}(1)$: la suite $(X_n)_n$ converge en loi vers la loi de Poisson de paramètre 1.

Q 9. Quantifions la convergence établie à la question précédente. Pour tout $(n, k) \in \mathbb{N}^2$ avec $k \leq n$,

$$y_k - \mathbf{P}_n(X_n = k) = \frac{1}{e k!} - \frac{1}{k!} \sum_{i=0}^{n-k} \frac{(-1)^i}{i!} = \frac{1}{k!} \sum_{i=n-k+1}^{\infty} \frac{(-1)^i}{i!} \quad \therefore \quad |y_k - \mathbf{P}_n(X_n = k)| \leq \frac{1}{k!(n-k+1)!}$$

par le critère spécial des séries alternées, qui s'applique car la suite $(1/k!)_k$ est décroissante de limite nulle. Alors,

$$\begin{aligned} G_Y(s) - G_{X_n}(s) &\stackrel{(1)}{=} \sum_{k=0}^{\infty} y_k s^k - \sum_{k=0}^n \mathbf{P}(X_n = k) s^k = \sum_{k=0}^n (y_k - \mathbf{P}(X_n = k)) s^k + \sum_{k=n+1}^{\infty} \frac{s^k}{e k!} \quad \dots \\ |G_Y(s) - G_{X_n}(s)| &\stackrel{(2)}{\leq} \sum_{k=0}^n |y_k - \mathbf{P}(X_n = k)| |s|^k + \sum_{k=n+1}^{\infty} \frac{|s|^k}{e k!} \\ &\leq \sum_{k=0}^n \frac{|s|^k}{k! (n+1-k)!} + \sum_{k=n+1}^{\infty} \frac{|s|^k}{e k!} \stackrel{(3)}{=} \frac{(1+|s|)^{n+1} - |s|^{n+1}}{(n+1)!} + \sum_{k=n+1}^{\infty} \frac{|s|^k}{e k!} \end{aligned}$$

(1) Par définition d'une fonction génératrice, la variable aléatoire X_n prenant ses valeurs entre 0 et n .

(2) Par l'inégalité triangulaire.

(3) Par la formule du binôme de Newton.

Par croissances comparées, le premier terme de la majoration, qui est équivalent à $\frac{|s|^n}{n!}$, tend vers 0 et, par convergence de la série, le deuxième aussi. On a donc montré que la suite $(G_{X_n})_n$ converge simplement vers G_Y sur \mathbb{R} . On peut noter en passant que la majoration précédente donne en fait la convergence uniforme sur tout segment.

Une approche plus directe consiste à s'intéresser directement à la fonction génératrice de X_n . Pour tout $s \in \mathbb{R}$,

$$G_{X_n}(s) \stackrel{(Q5)}{=} \sum_{k=0}^n \frac{1}{k!} \sum_{i=0}^{n-k} \frac{(-1)^i}{i!} s^k = \sum_{\substack{0 \leq i, k \leq n \\ i+k \leq n}} \frac{(-1)^i s^k}{i! k!} = \sum_{m=0}^n \sum_{\substack{0 \leq i, k \leq n \\ i+k=m}} \frac{(-1)^i s^k}{i! k!}.$$

On reconnaît la somme partielle d'indice n du produit de Cauchy des séries absolument convergentes $\sum_{i=0}^{\infty} \frac{(-1)^i}{i!} = \frac{1}{e}$

et $\sum_{k=0}^{\infty} \frac{s^k}{k!} = e^s$, d'où $\lim_{n \rightarrow \infty} G_{X_n}(s) = \frac{1}{e} \times e^s = e^{s-1} = G_Y(s)$.

2 — Convergence en variation totale

Q 10. Soient x et y deux distributions sur \mathbb{N} .

• Par positivité de x et de y , l'inégalité triangulaire généralisée donne

$$0 \leq d_{\text{VT}}(x, y) = \frac{1}{2} \sum_{k=0}^{\infty} |x(k) - y(k)| \leq \frac{1}{2} \sum_{k=0}^{\infty} [x(k) + y(k)] = \frac{1}{2} \sum_{k=0}^{\infty} x(k) + \frac{1}{2} \sum_{k=0}^{\infty} y(k) = 1.$$

• Une série positive est de somme nulle si, et seulement si tous ses termes sont nuls. Ainsi, $d_{\text{VT}}(x, y) = 0$ si, et seulement si, $x = y$.

• La symétrie $d_{\text{VT}}(x, y) = d_{\text{VT}}(y, x)$ est évidente car $|x(k) - y(k)| = |y(k) - x(k)|$ pour tout $k \in \mathbb{N}$.

• Si z est aussi une distribution, l'inégalité triangulaire, la linéarité et le caractère croissant de la sommation des séries donnent

$$\begin{aligned} d_{\text{VT}}(x, z) &= \frac{1}{2} \sum_{k=0}^{\infty} |x(k) - z(k)| \leq \frac{1}{2} \sum_{k=0}^{\infty} (|x(k) - y(k)| + |y(k) - z(k)|) \\ &\leq \frac{1}{2} \sum_{k=0}^{\infty} |x(k) - y(k)| + \frac{1}{2} \sum_{k=0}^{\infty} |y(k) - z(k)| = d_{\text{VT}}(x, y) + d_{\text{VT}}(y, z). \end{aligned}$$

Les trois dernières propriétés montrent que d_{VT} définit une *distance* sur l'ensemble des distributions sur \mathbb{N} .

Q 11. Si $X \sim \mathcal{B}(\lambda)$, alors $p_X(0) = 1 - \lambda$ et $p_X(1) = \lambda$, les autres valeurs étant nulles. Alors, pour $Y \sim \mathcal{B}(\mu)$,

$$d_{\text{VT}}(p_X, p_Y) = \frac{1}{2} (|(1 - \lambda) - (1 - \mu)| + |\lambda - \mu|) = |\lambda - \mu|.$$

Q 12. Par définition, $\pi_\lambda(k) = \frac{e^{-\lambda} \lambda^k}{k!}$, d'où, en particulier, $\pi_\lambda(0) = e^{-\lambda}$ et $\pi_\lambda(1) = \lambda e^{-\lambda}$. Il s'ensuit

$$\begin{aligned} d_{\text{VT}}(p_X, \pi_\lambda) &= \frac{|1 - \lambda - e^{-\lambda}|}{2} + \frac{|\lambda(1 - e^{-\lambda})|}{2} + \frac{1}{2} \sum_{k=2}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!} \\ &= \frac{1}{2} [(e^{-\lambda} + \lambda - 1) + \lambda(1 - e^{-\lambda}) + e^{-\lambda}(e^\lambda - (1 + \lambda))] = \lambda(1 - e^{-\lambda}) \end{aligned}$$

en appliquant le caractère croissant de l'exponentielle et l'inégalité de convexité $e^x \geq 1 + x$ avec $x = -\lambda$ pour faire disparaître les valeurs absolues.

En appliquant une deuxième fois l'inégalité $e^{-\lambda} \geq 1 - \lambda$, il vient $d_{\text{VT}}(p_X, \pi_\lambda) \leq \lambda^2$. On peut alternativement noter que, pour $Z \sim \mathcal{P}(\lambda)$, l'inégalité de Markov donne $1 - e^{-\lambda} = \mathbf{P}(Z \geq 1) \leq \frac{\mathbf{E}(Z)}{1^2} = \lambda$.

Q 13. On applique simplement la définition et l'expression de la loi de X_n établie à la question 5.

$$\begin{aligned} 2 d_{\text{VT}}(p_{X_n}, \pi_1) &= \sum_{k=0}^{\infty} |\mathbf{P}_n(X_n = k) - \pi_1(k)| = \sum_{k=0}^n \left| \frac{1}{k!} \sum_{i=0}^{n-k} \frac{(-1)^i}{i!} - \frac{1}{k!} \sum_{i=0}^{\infty} \frac{(-1)^i}{i!} \right| + \sum_{k=n+1}^{\infty} \left| \frac{(-1)^k}{e k!} \right| \\ &= \sum_{k=0}^n \left| \frac{1}{k!} \sum_{i=n-k+1}^{\infty} \frac{(-1)^i}{i!} \right| + \sum_{k=n+1}^{\infty} \left| \frac{(-1)^k}{e k!} \right| = \sum_{k=0}^n \frac{1}{k!} \left| \sum_{i=n-k+1}^{\infty} \frac{(-1)^i}{i!} \right| + \frac{1}{e} \sum_{k=n+1}^{\infty} \frac{1}{k!}. \end{aligned}$$

Q 14. Soit $n \in \mathbb{N}$. Rappelons que, par convention, un produit vide vaut 1.

$$\begin{aligned} r_n &= \sum_{k=n+1}^{\infty} \frac{1}{k!} = \frac{1}{(n+1)!} \sum_{k=n+1}^{\infty} \prod_{j=n+2}^k \frac{1}{j} \leq \frac{1}{(n+1)!} \sum_{k=n+1}^{\infty} \frac{1}{(n+2)^{k-n-1}} \stackrel{(i=k-n-1)}{=} \frac{1}{(n+1)!} \sum_{i=0}^{\infty} \frac{1}{(n+2)^i} \quad \dots \\ &\frac{1}{(n+1)!} \leq r_n \leq \frac{1}{(n+1)!} \times \frac{1}{1 - \frac{1}{n+2}} = \frac{1}{(n+1)!} \times \frac{n+2}{n+1} \sim \frac{1}{(n+1)!}, \end{aligned}$$

d'où $r_n \sim \frac{1}{(n+1)!}$ par le théorème des gendarmes.

Q 15. Le deuxième terme de la majoration de la question 13 est équivalent à $\frac{1}{e(n+1)!}$ par la question 14. En utilisant la majoration du reste d'une série alternée, il vient

$$\begin{aligned} \sum_{k=0}^n \frac{1}{k!} \left| \sum_{i=n-k+1}^{\infty} \frac{(-1)^i}{i!} \right| &\leq \sum_{k=0}^n \frac{1}{k!(n-k+1)!} = \frac{1}{(n+1)!} \sum_{k=0}^n \binom{n+1}{k} = \frac{2^{n+1} - 1}{(n+1)!} \quad \dots \\ r_n &= \mathcal{O}\left(\frac{2^n}{(n+1)!}\right) + \mathcal{O}\left(\frac{1}{(n+1)!}\right) = \mathcal{O}\left(\frac{2^n}{(n+1)!}\right). \end{aligned}$$

3 — Autres estimations de distances en variation totale

Q 16. Soient x et y deux distributions de probabilité sur \mathbb{N} . Pour $k \in \mathbb{N}$, on définit leur *convolution* par la formule $x * y(k) = \sum_{i=0}^k x(i)y(k-i)$. Il est immédiat que $x * y(k) \geq 0$. De plus, la série $\sum_{k=0}^{\infty} x * y(k)$ est le produit de Cauchy des séries (absolument convergentes) de terme général respectifs $x(k)$ et $y(k)$, d'où $\sum_{k=0}^{\infty} x * y(k) = \sum_{k=0}^{\infty} x(k) \times \sum_{k=0}^{\infty} y(k) = 1$, ce qui achève de montrer que $x * y$ est une distribution de probabilité sur \mathbb{N} .

On peut aussi résoudre d'un seul coup cette question et la suivante : il existe deux v.a. X et Y indépendantes à valeurs dans \mathbb{N} de distributions respectives x et y . Alors, $x * y$ est la distribution de $X + Y$ et est donc une distribution de probabilités.

Q 17. Soient X et Y deux variables aléatoires indépendantes à valeurs dans \mathbb{N} . Alors, pour $n \in \mathbb{N}$,

$$\begin{aligned} (X + Y = n) &= \bigsqcup_{k=0}^n [(X = k) \cap (Y = n - k)] \quad \therefore \\ p_{X+Y}(n) &= \mathbf{P}(X + Y = n) = \sum_{k=0}^n \mathbf{P}[(X = k) \cap (Y = n - k)] = \sum_{k=0}^n \mathbf{P}(X = k) \mathbf{P}(Y = n - k) \\ &= \sum_{k=0}^n p_X(k) p_Y(n - k) = (p_X * p_Y)(n). \end{aligned}$$

On pouvait aussi passer par les séries génératrices en notant que $G_{X+Y} = G_X \times G_Y$ est, par unicité du DSE, équivalente à $p_{X+Y} = p_X * p_Y$.

Q 18. Pour $(x, y, u, v) \in \mathcal{D}_{\mathbb{N}}^4$ et $(i, j) \in \mathbb{N}^2$,

$$\begin{aligned} (x * y)(k) - (u * v)(k) &= \sum_{i+j=k} [x(i)y(j) - u(i)v(j)] = \sum_{i+j=k} [(y(j)(x(i) - u(i)) + u(i)(y(j) - v(j)))] \quad \therefore \\ |(x * y)(k) - (u * v)(k)| &\leq \sum_{i+j=k} y(j)|x(i) - u(i)| + \sum_{i+j=k} u(i)|y(j) - v(j)| \end{aligned}$$

en vertu de l'inégalité triangulaire et de la positivité de u et de y .

Q 19. De l'inégalité précédente, on déduit

$$\begin{aligned} 2 \, d_{\text{VT}}(x * y, u * v) &= \sum_{k=0}^{\infty} |(x * y)(k) - (u * v)(k)| \stackrel{(\text{Q18})}{\leq} \sum_{k=0}^{\infty} \sum_{i+j=k} y(j)|x(i) - u(i)| + \sum_{k=0}^{\infty} \sum_{i+j=k} u(i)|y(j) - v(j)| = \\ &= \left(\sum_{j=0}^{\infty} y_j \right) \left(\sum_{i=0}^{\infty} |x(i) - u(i)| \right) + \left(\sum_{i=0}^{\infty} u_i \right) \left(\sum_{j=0}^{\infty} |y(j) - v(j)| \right) = 2 \, d_{\text{VT}}(x, u) + 2 \, d_{\text{VT}}(y, v) \end{aligned}$$

en reconnaissant à nouveau des produits de Cauchy de séries (absolument) convergentes.

Q 20. Raisonnons par récurrence sur n . Pour $n = 1$, l'inégalité est celle de la question 12. Supposons l'inégalité vérifiée à l'ordre $n - 1$. Pour $U \sim \mathcal{B}(n, \lambda)$, on peut écrire U comme une somme de n v.a.i.i.d. X_i de loi $\mathcal{B}(\lambda)$, d'où en posant $U_1 = X_1 + X_2 + \dots + X_{n-1}$, $U = U_1 + X_n$ avec $U_1 \perp\!\!\!\perp X_n$ en vertu du lemme des coalitions. La question 17 donne alors $p_U = p_{U_1} * p_{X_n}$. On sait également qu'une somme de deux v.a. suivant une loi de Poisson suit également une loi de Poisson dont le paramètre est la somme des deux paramètres, ce qui donne en particulier $\pi_{n\lambda} = \pi_{(n-1)\lambda} * \pi_{\lambda}$. On peut alors conclure :

$$\begin{aligned} d_{\text{VT}}(p_U, \pi_{n\lambda}) &= d_{\text{VT}}(p_{U_1} * p_{X_n}, \pi_{(n-1)\lambda} * \pi_{\lambda}) \stackrel{(\text{Q19})}{\leq} d_{\text{VT}}(p_{U_1}, \pi_{(n-1)\lambda}) + d_{\text{VT}}(p_{X_n}, \pi_{\lambda}) \\ &\stackrel{(\text{Q12} \& \text{HR}(n-1))}{\leq} (n-1)\lambda^2 + \lambda^2 = n\lambda^2. \end{aligned}$$

On peut aussi raisonner directement en étendant l'inégalité de la question 19 « par une récurrence immédiate » à

$$(\diamond) \quad d_{\text{VT}}(x_1 * x_2 * \dots * x_n, y_1 * y_2 * \dots * y_n) \leq \sum_{k=1}^n d_{\text{VT}}(x_i, y_i).$$

Bon, dit comme ça, cela risque d'être considéré comme un peu rapide (le rapport du jury, qui est une bonne approximation de l'ensemble vide, n'en dit rien) alors, détaillons :

$$\forall (x, y, z) \in \mathcal{D}_{\mathbb{N}}^3: ((x * y) * z)(\ell) = \sum_{k=0}^{\ell} (x * y)(k) z(\ell - k) = \sum_{k=0}^{\ell} \sum_{i=0}^k x(i) y(k - i) z(\ell - k) = \sum_{i+j+k=\ell} x(i) y(j) z(k).$$

Le résultat étant symétrique en x, y, z , il s'ensuit que l'opération $*$ est associative et que l'on peut en conséquence définir sans ambiguïté un produit de convolution fini quelconque de mesures de probabilités (le produit est également

clairement commutatif, ce qui n'est pas indispensable ici). On peut alors démontrer par récurrence l'inégalité (\diamond). La question Q19 la prouve pour $n = 2$. Si celle-ci est vraie pour $n - 1$ et pour 2, alors

$$\begin{aligned} d_{\text{VT}}(x_1 * x_2 * \cdots * x_n, y_1 * y_2 * \cdots * y_n) &= d_{\text{VT}}(x_1 * (x_2 * \cdots * x_n), y_1 * (y_2 * \cdots * y_n)) \\ &\stackrel{\text{(HR(2))}}{\leq} d_{\text{VT}}(x_1, y_1) + d_{\text{VT}}(x_2 * \cdots * x_n, y_2 * \cdots * y_n) \stackrel{\text{(HR(n-1))}}{\leq} d_{\text{VT}}(x_1, y_1) + \sum_{i=2}^n d_{\text{VT}}(x_i, y_i) = \sum_{i=1}^n d_{\text{VT}}(x_i, y_i). \end{aligned}$$

On applique cette inégalité avec $x_1 = x_2 = \cdots = x_n = x$, où x est la distribution associée à la loi de Bernoulli de paramètre λ et $y_1 = y_2 = \cdots = y_n = \pi_\lambda$. D'après la question 17 et le fait qu'une somme de v.a.i.i.d. de loi de Bernoulli suit une loi binomiale et qu'une somme de v.a. suivant une loi de Poisson suit aussi une loi de Poisson, il vient

$$d_{\text{VT}}(p_U, \pi_{n\lambda}) = d_{\text{VT}}(\underbrace{x * x * \cdots * x}_{\#n}, \underbrace{\pi_\lambda * \pi_\lambda * \cdots * \pi_\lambda}_{\#n}) \leq n d_{\text{VT}}(x, \pi_\lambda) \stackrel{\text{(Q12)}}{\leq} n\lambda^2.$$

Q 21. On applique la question précédente avec $\lambda = \frac{\alpha}{n}$ (notons que l'hypothèse $n > \lfloor \alpha \rfloor$ assure que $\lambda < 1$). Il vient

$$d_{\text{VT}}(p_{B_n}, \pi_\alpha) \leq n \left(\frac{\alpha}{n}\right)^2 = \frac{\alpha^2}{n}. \text{ Corrélativement, pour } k \in \mathbb{N},$$

$$|\mathbf{P}(B_n = k) - \pi_\alpha(k)| \leq \sum_{i=0}^{\infty} |\mathbf{P}(B_n = i) - \pi_\alpha(i)| = 2 d_{\text{VT}}(p_{B_n}, \pi_\alpha) \stackrel{\text{(Q20)}}{n \rightarrow \infty} \rightarrow 0,$$

soit $\lim_{n \rightarrow \infty} \mathbf{P}(B_n = k) = e^{-\alpha} \frac{\alpha^k}{k!}$. Majorer le terme général d'une série à termes positifs par sa somme est assez grossier. Le résultat demandé se montre aussi directement (*cf.* cours sur la loi de Poisson) :

$$\mathbf{P}(B_n = k) = \binom{n}{k} \left(\frac{\alpha}{n}\right)^k \left(1 - \frac{\alpha}{n}\right)^{n-k} = \frac{\alpha^k}{k!} \times \underbrace{\frac{n!}{n^k(n-k)!}}_{\rightarrow 1} \times \underbrace{\left(1 - \frac{\alpha}{n}\right)^n}_{\rightarrow e^{-\alpha}} \times \underbrace{\left(1 - \frac{\alpha}{n}\right)^{-k}}_{\rightarrow 1} \xrightarrow{n \rightarrow \infty} e^{-\alpha} \frac{\alpha^k}{k!}.$$

Q 22. Notons $B_n \sim \mathcal{B}\left(n, \frac{\alpha}{n}\right)$ et $B'_n \sim \mathcal{B}\left(n, \frac{\beta}{n}\right)$. Alors,

$$\begin{aligned} d_{\text{VT}}(\pi_\alpha, \pi_\beta) &\stackrel{\text{(Q10)}}{\leq} d_{\text{VT}}(\pi_\alpha, p_{B_n}) + d_{\text{VT}}(p_{B_n}, p_{B'_n}) + d_{\text{VT}}(p_{B'_n}, \pi_\beta) \\ &\stackrel{\text{(Q20)}}{\leq} \frac{\alpha^2}{n} + d_{\text{VT}}(p_{B_n}, p_{B'_n}) + \frac{\beta^2}{n} \stackrel{\text{(Q11 \& Q19)}}{\leq} \frac{\alpha^2 + \beta^2}{n} + n \left| \frac{\alpha}{n} - \frac{\beta}{n} \right|, \end{aligned}$$

d'où $d_{\text{VT}}(\pi_\alpha, \pi_\beta) \leq |\alpha - \beta|$ en faisant tendre n vers l'infini.

Exercice 1 — Oral X

Q 23. Avant toute chose, on peut vérifier que $\frac{k}{k+\theta} + \frac{\theta}{k+\theta} = 1$. L'événement $(K_n = 1)$ traduit que les n premiers convives (sauf le premier d'entre eux) n'ont pas choisi de s'installer à une nouvelle table, ce qui, d'après la formule des probabilités composées donne

$$\mathbf{P}(K_n = 1) = 1 \times \frac{1}{1+\theta} \times \frac{2}{2+\theta} \times \cdots \times \frac{n-1}{n-1+\theta} = \prod_{k=1}^{n-1} \frac{k}{k+\theta}.$$

Q 24. Comme $K_1 = 1$, on a $G_1(x) = x$. Par ailleurs, K_n et $K_{n+1} - K_n$ sont indépendantes (s'installer à une nouvelle table est indépendant du nombre de tables comptant au moins un commensal) et $K_{n+1} - K_n \sim \mathcal{B}\left(\frac{\theta}{n+\theta}\right)$, d'où

$$\begin{aligned} G_{n+1}(x) &= \mathbf{E}(x^{K_{n+1}}) = \mathbf{E}(x^{K_{n+1}-K_n} \times x^{K_n}) = \mathbf{E}(x^{K_{n+1}-K_n}) \times \mathbf{E}(x^{K_n}) = \left(\frac{n}{n+\theta} + \frac{\theta x}{n+\theta}\right) G_n(x) \\ &= G_1(x) \prod_{k=1}^n \frac{\theta x + k}{\theta + k} = \prod_{k=0}^n \frac{\theta x + k}{\theta + k} = \frac{L_n(\theta x)}{L_n(\theta)}. \end{aligned}$$

Q 25. On peut calculer l'espérance en utilisant la série génératrice. En effet, à partir de l'expression obtenue à la question 2, on obtient

$$G'_{n+1}(x) = \frac{\theta L'_n(\theta x)}{L_n(\theta)} = \frac{\theta L_n(\theta x)}{L_n(\theta)} \sum_{k=0}^n \frac{1}{\theta x + k} \quad \therefore \quad \mathbf{E}(K_{n+1}) = G'_{n+1}(1) = \frac{\theta L'_n(\theta)}{L_n(\theta)} = \sum_{k=0}^n \frac{\theta}{k + \theta}.$$

La décroissance de la fonction $t \mapsto \frac{1}{t + \theta}$ donne

$$\int_k^{k+1} \frac{dt}{t + \theta} \stackrel{(k \geq 0)}{\leq} \frac{1}{k + \theta} \stackrel{(k \geq 1)}{\leq} \int_{k-1}^k \frac{dt}{t + \theta} \quad \therefore$$

$$\theta [\ln(n+1 + \theta) - \ln(\theta)] = \theta \int_0^{n+1} \frac{dt}{t + \theta} \leq \mathbf{E}(K_{n+1}) \leq 1 + \theta \int_0^n \frac{dt}{t + \theta} = 1 + \theta [\ln(n + \theta) - \ln(\theta)] = \theta \ln n + \mathcal{O}(1),$$

soit $\mathbf{E}(K_{n+1}) \sim \theta \ln n$ par le théorème des gendarmes.

On pourrait aussi calculer ainsi la variance, mais le plus simple est de poser $K_0 = 0$ et d'écrire la décomposition télescopique $K_{n+1} = \sum_{j=0}^n (K_{j+1} - K_j)$ en somme de v.a.i. Avec $K_{j+1} - K_j \sim \mathcal{B}\left(\frac{\theta}{j + \theta}\right)$, on obtient alors immédiatement à nouveau $\mathbf{E}(K_{n+1})$ ainsi que, par l'indépendance,

$$\mathbf{V}(K_{n+1}) = \sum_{j=0}^n \mathbf{V}(K_{j+1} - K_j) = \sum_{j=0}^n \frac{\theta}{j + \theta} \times \frac{j}{j + \theta} = \theta \sum_{j=0}^n \frac{j}{(\theta + j)^2}.$$

La comparaison série intégrale donne avec un peu de travail $\mathbf{V}(K_{n+1}) \sim \theta \ln n$: la fonction $f(x) = \frac{x}{(x + \theta)^2}$ a pour dérivée $f'(x) = \frac{\theta - x}{(x + \theta)^3}$ et est donc décroissante à partir de θ ; de plus,

$$\int \frac{x}{(\theta + x)^2} dx = \int \frac{\theta + x - x}{(\theta + x)^2} dx = \ln(\theta + x) + \frac{\theta}{\theta + x} \underset{x \rightarrow +\infty}{\sim} \ln x.$$

Q 26. Pour $\beta > 0$, l'inégalité de Bienaymé-Tchebychev donne

$$\mathbf{P}\left(\left|\frac{K_n}{\mathbf{E}(K_n)} - 1\right| > \beta\right) \leq \frac{\mathbf{V}(K_n / \mathbf{E}(K_n))}{\beta^2} = \frac{\mathbf{V}(K_n)}{(\beta \mathbf{E}(K_n))^2} \sim \frac{1}{\beta^2 \theta \ln n} \xrightarrow{n \rightarrow \infty} 0.$$

Soit $\varepsilon > 0$. De $\mathbf{E}(K_n) \sim \theta \ln n$, il s'ensuit que

$$\exists n_0 \in \mathbb{N}, \forall n \geq n_0 : (1 - \varepsilon)\theta \ln n \leq \mathbf{E}(K_n) \leq (1 + \varepsilon)\theta \ln n.$$

Alors, en supposant $n \geq n_0$ et $\beta < 1$, il vient

$$\begin{aligned} \left|\frac{K_n}{\mathbf{E}(K_n)} - 1\right| \leq \beta &\implies (1 - \beta)(1 - \varepsilon)\theta \ln n \leq (1 - \beta) \mathbf{E}(K_n) \leq K_n \leq (1 + \beta) \mathbf{E}(K_n) \leq (1 + \beta)(1 + \varepsilon)\theta \ln n \\ &\implies \theta(-\beta - \varepsilon + \beta\varepsilon) \leq \frac{K_n}{\ln n} - \theta \leq \theta(\beta + \varepsilon + \beta\varepsilon) \implies \left|\frac{K_n}{\ln n} - \theta\right| \leq \theta(\beta + \varepsilon + \beta\varepsilon). \end{aligned}$$

Pour $\alpha > 0$, si ε et β sont choisis tels que $\theta(\beta + \varepsilon + \beta\varepsilon) \leq \alpha$, il s'ensuit

$$\left(\left|\frac{K_n}{\ln n} - \theta\right| > \alpha\right) \subset \left(\left|\frac{K_n}{\mathbf{E}(K_n)} - 1\right| > \beta\right),$$

événement dont la probabilité tend vers 0 quand n tend vers l'infini. Il y a bien convergence en probabilité vers θ .

Exercice 2 — Oral CCINP

Q 27. Si $r = 1$, il y a une unique boule blanche dans l'urne et X désigne le moment de son tirage. On a donc $X \sim \mathcal{U}(\llbracket 1, N \rrbracket)$ et $\mathbf{E}(X) = \frac{N+1}{2}$. Si $r = N$, toutes les boules sont blanches et X est la v.a. certaine égale à N . On a alors $\mathbf{E}(X) = N$.

Q 28. Pour $r \in \llbracket 2, N-1 \rrbracket$, il faut au moins r tirages pour obtenir toutes les boules blanches et, toujours au plus N , donc $X(\Omega) \subset \llbracket r, N \rrbracket$. Soit $r \leq k \leq N$. Si l'on commence par tirer $r-1$ boules blanches, puis $k-r \in \llbracket 0, N-r \rrbracket$ boules noires, puis la dernière boule blanche, on a obtenu une réalisation de $(X = k)$, d'où $X(\Omega) = \llbracket r, N \rrbracket$. Notons que ce raisonnement vaut aussi pour $r \in \{1, N\}$.

Soit alors $k \in \llbracket r, N \rrbracket$. L'événement $(X = k)$ peut se décrire par le fait que, en ayant vidé l'urne boule par boule, le tirage d'indice k a produit une boule blanche et que, parmi les $k-1$ tirages précédents, on avait tiré $r-1$ boules blanches. On peut considérer l'ensemble des tirages sans discerner les boules de même couleur. Il y a alors $\binom{N}{r}$ tirages possibles (r boules blanches à placer parmi N), dont $\binom{k-1}{r-1}$ sont favorables, ce qui donne bien

$$\mathbf{P}(X = k) = \binom{k-1}{r-1} \binom{N}{r}^{-1}.$$

Q 29. En utilisant les factorielles, on a (pour $p \geq q \geq 1$),

$$\binom{p}{q} = \frac{p!}{q!(p-q)!} = p(p-1)q(q-1)\dots((p-1)-(q-1))! = \frac{p}{q} \binom{p-1}{q-1}.$$

L'interprétation combinatoire classique de cette identité consiste à constituer une équipe de q joueurs avec un capitaine dans un réservoir de p joueurs. Il y a $q \binom{p}{q} = p \binom{p-1}{q-1}$ façons de le faire, selon que l'on choisit l'équipe, puis le capitaine au sein de l'équipe, ou bien le capitaine, puis le reste de l'équipe. En utilisant les résultats déjà obtenus,

$$\begin{aligned} \mathbf{E}(X) &= \sum_{k=r}^N k \mathbf{P}(X = k) = \sum_{k=r}^N \frac{k \binom{k-1}{r-1}}{\binom{N}{r}} = \sum_{k=r}^N \frac{r \binom{k}{r}}{\binom{N}{r}} = \frac{r}{\binom{N}{r}} \sum_{k=r}^N \binom{k}{r} \\ &= \frac{r}{\binom{N}{r}} \sum_{k=r}^N \left[\binom{k+1}{r+1} - \binom{k}{r+1} \right] = \frac{r}{\binom{N}{r}} \left[\binom{N+1}{r+1} - \binom{r}{r+1} \right] = \frac{r \binom{N+1}{r+1}}{\binom{N}{r}} = \frac{r(N+1)}{r+1}. \end{aligned}$$

Notons que le résultat reste valable pour $r \in \{1, N\}$. On pouvait aussi (ce n'était pas la piste suggérée par l'énoncé), utiliser la formule donnant l'espérance d'une v.a. à valeurs dans \mathbb{N} à partir de sa fonction de répartition. On obtient facilement la fonction de répartition de X : $\mathbf{P}(X \leq k) = \binom{k}{r} \binom{N}{r}^{-1}$ (on a tiré toutes les boules blanches parmi les k premières). Il vient alors, par un calcul classique de somme (en utilisant par exemple la formule de Pascal pour la rendre télescopique),

$$\begin{aligned} \mathbf{E}(X) &= \sum_{k=1}^{\infty} \mathbf{P}(X \geq k) = \sum_{k=1}^N [1 - \mathbf{P}(X < k)] = N - \binom{N}{r}^{-1} \sum_{k=1}^N \binom{k-1}{r} \\ &= N - \binom{N}{r}^{-1} \sum_{k=1}^N \left[\binom{k}{r+1} - \binom{k-1}{r+1} \right] = N - \binom{N}{r}^{-1} \binom{N}{r+1} = N - \frac{N-r}{r+1} = \frac{r(N+1)}{r+1}. \end{aligned}$$

Sujet 2 — Problème CCMP 2020, MP (4h)

Nombre de sites visités par une marche aléatoire

On considère une variable aléatoire X à valeurs dans \mathbb{Z}^d , $(X_k)_{k \in \mathbb{N}^*}$ une suite de variables aléatoires mutuellement indépendantes suivant chacune la loi de X et définies sur un même espace probabilisé. La suite de variables aléatoires $(S_n)_{n \in \mathbb{N}}$ est définie par

$$S_0 = 0_d \quad \& \quad \forall n \in \mathbb{N}^* : S_n = \sum_{k=1}^n X_k.$$

La suite $(S_n)_{n \in \mathbb{N}}$ est une *marche aléatoire de pas* X , à valeurs dans \mathbb{Z}^d .

On note R la variable aléatoire à valeurs dans $\mathbb{N}^* \cup \{+\infty\}$ définie par

$$R = \begin{cases} \min\{n \in \mathbb{N}^* ; S_n = 0_d\} & \text{si } \{n \in \mathbb{N}^* , S_n = 0_d\} \neq \emptyset, \\ +\infty & \text{sinon.} \end{cases}$$

Autrement dit, R est égal à $+\infty$ si la marche aléatoire $(S_n)_{n \in \mathbb{N}^*}$ ne revient jamais en 0_d et au premier instant auquel cette marche aléatoire y revient sinon.

Pour $n \in \mathbb{N}$, soit $N_n = \#\{S_k ; k \in \llbracket 0, n \rrbracket\}$. Le nombre N_n est donc le nombre de points de \mathbb{Z}^d visités par la marche aléatoire $(S_k)_{k \in \mathbb{N}}$ après n pas.

A. PRÉLIMINAIRES

Les cinq questions de cette partie sont indépendantes et utilisées dans les parties C et E .

Q 1. La formule du binôme donne d'une part $(X + 1)^{2n} = \sum_{k=0}^{2n} \binom{2n}{k} X^k$. D'autre part, par produit de Cauchy, en exploitant la convention $\binom{a}{b} = 0$ lorsque $b > a$, il vient

$$(X + 1)^n (X + 1)^n = \left(\sum_{k=0}^{+\infty} \binom{n}{k} X^k \right)^2 = \sum_{k=0}^{+\infty} \left(\sum_{\ell=0}^k \binom{n}{\ell} \binom{n}{k-\ell} \right) X^k.$$

Ces deux polynômes étant égaux, l'identification des coefficients de X^n dans ces deux expressions donne, en exploitant la symétrie des coefficients binomiaux,

$$\binom{2n}{n} = \sum_{\ell=0}^n \binom{n}{\ell} \binom{n}{n-\ell} = \sum_{\ell=0}^n \binom{n}{\ell}^2.$$

Q 2. La formule de Stirling donne $n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$, d'où

$$\binom{2n}{n} = \frac{(2n)!}{(n!)^2} \underset{n \rightarrow +\infty}{\sim} \frac{\left(\frac{2n}{e}\right)^{2n} \sqrt{2\pi 2n}}{\left(\frac{n}{e}\right)^{2n} 2\pi n} = \frac{4^n}{\sqrt{\pi n}} = c \frac{4^n}{\sqrt{n}} \quad \text{avec } c = \frac{1}{\sqrt{\pi}}.$$

Q 3. Pour $0 < \alpha < 1$, la fonction $t \mapsto t^{-\alpha}$ est décroissante et intégrable sur $]0, 1]$. La comparaison série-intégrale donne alors

$$\forall k \in \mathbb{N}^* : \int_k^{k+1} \frac{dt}{t^\alpha} \leq \frac{1}{k^\alpha} \leq \int_{k-1}^k \frac{dt}{t^\alpha},$$

Supposons que $\alpha \in]0, 1[$. Alors, en rappelant que $\int_0^1 \frac{dt}{t^\alpha}$ est convergente, il vient

$$\frac{n^{1-\alpha}}{1-\alpha} = \int_0^n \frac{dt}{t^\alpha} = \sum_{k=1}^n \int_{k-1}^k \frac{dt}{t^\alpha} \geq \sum_{k=1}^n \frac{1}{k^\alpha} \geq \sum_{k=1}^n \int_k^{k+1} \frac{dt}{t^\alpha} = \int_1^{n+1} \frac{dt}{t^\alpha} = \frac{(n+1)^{1-\alpha} - 1}{1-\alpha} \sim \frac{n^{1-\alpha}}{1-\alpha},$$

ce qui montre que $\sum_{k=1}^n \frac{1}{n^\alpha} \sim \frac{n^{1-\alpha}}{1-\alpha}$. De même, si $\alpha > 1$, la série de Riemann et l'intégrale convergent et la fonction $t \mapsto t^{-\alpha}$ est toujours décroissante, d'où

$$\frac{1}{(\alpha-1)n^{\alpha-1}} = \int_n^{+\infty} \frac{dt}{t^\alpha} = \sum_{k=n+1}^{\infty} \int_{k-1}^k \frac{dt}{t^\alpha} \geq \sum_{k=n+1}^{\infty} \frac{1}{k^\alpha} \geq \sum_{k=n+1}^{\infty} \int_k^{k+1} \frac{dt}{t^\alpha} = \int_{n+1}^{+\infty} \frac{dt}{t^\alpha} = \frac{(n+1)^{1-\alpha}}{\alpha-1} \sim \frac{1}{(\alpha-1)n^{\alpha-1}}.$$

On en déduit que $\sum_{k=n+1}^{\infty} \frac{1}{n^\alpha} \sim \frac{1}{(\alpha-1)n^{\alpha-1}}$.

Q 4. On réalise une intégration par parties (toutes les fonctions sont de classe \mathcal{C}^1 sur le segment $[2, x]$) :

$$I(x) = \int_2^x \underbrace{\frac{1}{\ln t}}_u \times \underbrace{1}_{v'} dt = \left[\frac{t}{\ln t} \right]_2^x - \int_2^x \left(-\frac{1}{t(\ln t)^2} \right) \times t dt = \frac{x}{\ln x} - \frac{2}{\ln 2} + \int_2^x \frac{dt}{(\ln t)^2}.$$

Q 5. C'est une intégration de relation de comparaison, connue en MP, mais hors programme en PSI (voir la feuille d'exercices sur les intégrales impropres pour la démonstration). On va procéder directement, les fonctions étant explicites. On majore l'intégrale en deux fois, en utilisant la relation de Chasles et la monotonie de $t \mapsto 1/(\ln t)^2$.

$$\begin{aligned} 0 &\leq \int_2^{\sqrt{x}} \frac{dt}{(\ln t)^2} dt \leq \frac{\sqrt{x}-2}{(\ln 2)^2} = \mathcal{O}(\sqrt{x}) = \mathfrak{o}\left(\frac{x}{\ln x}\right); \\ 0 &\leq \int_{\sqrt{x}}^x \frac{dt}{(\ln t)^2} dt \leq \frac{x-\sqrt{x}}{(\ln \sqrt{x})^2} = \mathcal{O}\left(\frac{x}{(\ln x)^2}\right) = \mathfrak{o}\left(\frac{x}{\ln x}\right). \end{aligned}$$

Ainsi, $\int_2^x \frac{dt}{(\ln t)^2} dt = \mathfrak{o}\left(\frac{x}{\ln x}\right)$, d'où $I(x) \underset{x \rightarrow +\infty}{\sim} \frac{x}{\ln x}$.

Q 6. On utilise le développement en série entière de $(1-x)^\alpha$ avec $\alpha = -\frac{1}{2}$. En notant b_n le coefficient de x^n dans ce DSE, il vient

$$\begin{aligned} b_n &= -\frac{1}{2} \left(-\frac{1}{2} - 1 \right) \cdots \left(-\frac{1}{2} - n + 1 \right) \frac{(-1)^n}{n!} = -\frac{1}{2} \left(-\frac{3}{2} \right) \cdots \left(-\frac{2n-1}{2} \right) \frac{(-1)^n}{n!} \\ &= (-1)^n \frac{1 \times 3 \times \cdots (2n-1)}{2^n n!} (-1)^n = \frac{(2n)!}{2^{2n} n!^2} = \frac{1}{2^{2n}} \binom{2n}{n} \quad \therefore \quad \forall x \in]-1, 1[: \frac{1}{\sqrt{1-x}} = \sum_{n=0}^{+\infty} \frac{1}{4^n} \binom{2n}{n} x^n. \end{aligned}$$

B. MARCHES ALÉATOIRES, RÉCURRENCE

On considère les fonctions F et G définies par les formules

$$\forall x \in]-1, 1[: F(x) = \sum_{n=0}^{+\infty} \mathbf{P}(S_n = 0_d) x^n \quad \& \quad G(x) = \sum_{n=0}^{+\infty} \mathbf{P}(R = n) x^n.$$

Q 7. La suite $(\mathbf{P}(S_n = 0_d) \times 1^n)_{n \in \mathbb{N}}$ est positive et bornée (par 1), donc le rayon de convergence de F , disons, R_F , vaut au moins 1. D'après le cours, toute série entière de rayon de convergence R est de classe \mathcal{C}^∞ sur $] -R, R[$. Ainsi, F est de classe \mathcal{C}^∞ sur $] -1, 1[\subset] -R_F, R_F[$. Le même raisonnement s'applique à G .

Q 8. Par additivité dénombrable,

$$G(1) = \sum_{n=1}^{\infty} \mathbf{P}(R = n) = \mathbf{P}\left(\bigcup_{n=1}^{+\infty} (R = n)\right) = \mathbf{P}(R < +\infty).$$

En particulier, la série ci-dessus converge. Corrélativement, la série entière définissant G converge normalement, donc uniformément, sur $[-1, 1]$, ce qui montre que G est définie et continue sur le segment $[-1, 1]$.

Q 9. Soient k et n des entiers naturels non nuls tels que $k \leq n$. Alors,

$$[(S_n = 0_d) \cap (R = k)] = \underbrace{[(S_1 \neq 0_d) \cap \dots \cap (S_{k-1} \neq 0_d) \cap (S_k = 0_d)]}_{(R=k)} \cap (S_n = 0_d) = [(R = k) \cap (S_n - S_k = 0_d)].$$

En vertu du lemme des coalitions, $S_n - S_k = f(X_{k+1}, \dots, X_n) \perp\!\!\!\perp (R = k) = g(X_1, \dots, X_k)$, d'où

$$\mathbf{P}[(S_n = 0_d) \cap (R = k)] = \mathbf{P}(R = k) \mathbf{P}(S_n - S_k = 0_d).$$

Enfin, $S_n - S_k = \sum_{i=k+1}^n X_i$ est une somme de $n - k$ v.a.i.i.d. de même loi que X , tout comme S_{n-k} , donc $S_n - S_k$ et S_{n-k} ont même loi. Finalement,

$$\mathbf{P}[(S_n = 0_d) \cap (R = k)] = \mathbf{P}(R = k) \mathbf{P}(S_{n-k} = 0_d)$$

Les événements $(S_n = 0_d)$ et $(R > n)$ sont incompatibles, d'où

$$\forall n \in \mathbb{N}^*: (S_n = 0_d) = (S_n = 0_d) \cap [(R = 1) \uplus \dots \uplus (R = n)] \quad \therefore$$

$$\mathbf{P}(S_n = 0_d) = \sum_{k=1}^n \mathbf{P}[(R = k) \cap (S_n = 0_d)] = \sum_{k=1}^n \mathbf{P}(R = k) \mathbf{P}(S_{n-k} = 0_d).$$

Q 10. Le point crucial est que la formule de la question 9 n'est valable que pour $n \in \mathbb{N}^*$ (elle donne $1 = 0$ pour $n = 0$). Avec $\mathbf{P}(S_0 = 0_d) = 1$ et $\mathbf{P}(R = 0) = 0$, il vient

$$\begin{aligned} \forall x \in]-1, 1[: F(x) &= \sum_{n=0}^{+\infty} \mathbf{P}(S_n = 0_d) x^n = 1 + \sum_{n=1}^{+\infty} \mathbf{P}(S_n = 0_d) x^n \stackrel{(\text{Q9})}{=} 1 + \sum_{n=1}^{+\infty} \left(\sum_{k=1}^n \mathbf{P}(R = k) \mathbf{P}(S_{n-k} = 0_d) \right) x^n \\ &= 1 + \sum_{n=0}^{+\infty} \left(\sum_{k=0}^n \mathbf{P}(R = k) \mathbf{P}(S_{n-k} = 0_d) \right) x^n = 1 + F(x)G(x), \end{aligned}$$

en reconnaissant le produit de Cauchy des séries entières définissant F et G , qui sont bien, en vertu de la question 7, de rayon de convergence au moins égal à 1.

Q 11. La relation de la question 10 exclut que $G(x)$ prenne la valeur 1 et se réécrit donc $F(x) = \frac{1}{1 - G(x)}$. La question 8 donne alors

$$\lim_{\substack{x \rightarrow 1 \\ x < 1}} F(x) = \begin{cases} \frac{1}{1 - G(1)} = \frac{1}{\mathbf{P}(R = +\infty)} & \text{si } \mathbf{P}(R \neq +\infty) \neq 1, \\ +\infty & \text{si } \mathbf{P}(R \neq +\infty) = 1. \end{cases}$$

Q 12. Soit $(c_k)_{k \in \mathbb{N}}$ une suite d'éléments de \mathbb{R}^+ telle que la série entière $\sum c_k x^k$ ait un rayon de convergence égal à 1 et que la série $\sum c_k$ diverge. La fonction $C : x \mapsto \sum_{k=0}^{\infty} c_k x^k$ est une série de fonctions croissantes sur $[0, 1[$ et est donc croissante. Elle admet donc une limite à gauche en 1. Alors,

$$\forall x \in [0, 1[, \forall N \in \mathbb{N} : C(x) \geq \sum_{k=0}^{\infty} c_k x^k \quad \therefore \quad \forall N \in \mathbb{N} : \lim_{\substack{x \rightarrow 1 \\ x < 1}} C(x) \geq \sum_{k=0}^N c_k.$$

On peut alors faire tendre N vers l'infini, ce qui donne $\lim_{\substack{x \rightarrow 1 \\ x < 1}} C(x) = +\infty$.

Q 13. Si la série $\sum \mathbf{P}(S_n = 0_d)$ diverge, le rayon de convergence de F , qui est déjà au moins égal à 1 par la question Q7, est aussi au plus égal à 1, donc égal à 1. La question Q12 assure alors que $\lim_{x \rightarrow 1^-} F(x) = +\infty$ donc, par la question 11, que $\mathbf{P}(R \neq +\infty) = 1$.

Supposons inversement, en reprenant les notations de la question Q12, que la série $\sum c_k$ converge. Alors, la série entière converge normalement, donc uniformément, sur $[-1, 1]$ et sa limite est ainsi continue. En particulier, C admet une limite finie à gauche en 1 et la question 11 montre cette fois que $\mathbf{P}(R \neq +\infty) < 1$. Finalement, la série $\sum \mathbf{P}(S_n = 0_d)$ est divergente si, et seulement si, $\mathbf{P}(R \neq +\infty) = 1$.

Q 14. Dans toute la question, on fixe $i \in \mathbb{N}^*$.

Soit Y_i la variable de Bernoulli indicatrice de l'événement $(S_i \notin \{S_k, 0 \leq k \leq i-1\})$. Soit $i \in \mathbb{N}^*$. Pour montrer que $\mathbf{P}(Y_i = 1) = \mathbf{P}(R > i)$, on commence par s'intéresser aux événements. On a bien $(Y_1 = 1) = (X_1 \neq 0) = (R > 1)$, mais $(Y_i = 1) \neq (R > i)$ en général : par exemple, pour la marche aléatoire symétrique sur \mathbb{Z} (cas étudié à la question 20), si $(X_1(\omega), X_2(\omega), X_3(\omega)) = (1, -1, -1)$, on a $Y_3(\omega) = 1$ et $R(\omega) = 2$. Toutefois, en regardant de plus près l'événement $(Y_i = 1)$,

$$(Y_i = 1) = \bigcap_{j=0}^{i-1} (S_i \neq S_j) = \bigcap_{j=0}^{i-1} (X_{j+1} + X_{j+2} + \cdots + X_i \neq 0),$$

ce qui suggère de considérer les sommes S « en partant de la fin ». Formellement, posons

$$X'_j = X_{i+1-j} \quad \& \quad S'_j = \sum_{k=1}^j X'_k = \sum_{k=i+1-j}^i X_k$$

ainsi que la variable aléatoire R' associée et reprenons l'écriture précédente :

$$(Y_i = 1) = \bigcap_{j=0}^{i-1} (X_{j+1} + X_{j+2} + \cdots + X_i \neq 0) = \bigcap_{j=0}^{i-1} (S'_{i-j} \neq 0) = \bigcap_{k=1}^i (S'_k \neq 0) = (R' > i).$$

Or, S'_k a même loi que S_k et R' que R , d'où $\mathbf{P}(Y_i = 1) = \mathbf{P}(R > i)$.

Q 15. Quand on rajoute des éléments un par un à un ensemble, on augmente son cardinal de 1 quand l'élément ajouté n'y est pas déjà et ce cardinal reste sinon inchangé. Ainsi, par définition,

$$N_n = \#\{S_k; k \in \llbracket 0, n \rrbracket\} = 1 + \sum_{i=1}^n Y_i.$$

Par linéarité de l'espérance pour la première égalité et la question précédente pour la deuxième, il vient

$$\forall n \in \mathbb{N}^*: \mathbf{E}(N_n) = 1 + \sum_{i=1}^n \mathbf{P}(R > i).$$

Notons que la formule montre aussi que $\lim_{n \rightarrow \infty} \mathbf{E}(N_n) = \mathbf{E}(R)$.

Q 16. De $(R = +\infty) = \bigcap_{i \geq 1} (R > i)$, on tire par continuité décroissante que $\mathbf{P}(R = +\infty) = \lim_{i \rightarrow \infty} \mathbf{P}(R > i)$. Le

théorème de Cesàro, appliqué à $u_n = \mathbf{P}(R > n)$, donne alors $\lim_{n \rightarrow +\infty} \frac{\mathbf{E}(N_n)}{n} = \mathbf{P}(R = +\infty)$.

C. LES MARCHES DE BERNOULLI SUR \mathbb{Z}

Dans cette question, $d = 1$ et on note donc simplement $0_d = 0$. Par ailleurs, $p \in]0, 1[$, $q = 1 - p$ et la loi de X est donnée par $\mathbf{P}(X = 1) = p$ et $\mathbf{P}(X = -1) = q$.

Q 17. Comme X ne prend que des valeurs impaires, S_n est de même parité que n . En particulier, $(S_{2n+1} = 0)$ est impossible, soit $\mathbf{P}(S_{2n+1} = 0) = 0$. Par ailleurs, $X' = \frac{X+1}{2} \sim \mathcal{B}(p)$, d'où

$$S'_{2n} = \sum_{i=1}^{2n} X'_i = \frac{S_{2n} + 2n}{2} \sim \mathcal{B}(2n, p) \quad \therefore \quad \mathbf{P}(S_{2n} = 0) = \mathbf{P}(S'_{2n} = n) = \binom{2n}{n} p^n q^n.$$

On peut aussi faire le calcul directement, ce qui revient peu ou prou à recalculer la loi d'une binomiale.

Q 18. Avec les notations de la partie B, on a

$$F(x) = \sum_{n=0}^{\infty} \mathbf{P}(S_{2n} = 0_d) x^{2n} = \sum_{n=0}^{\infty} \binom{2n}{n} (pqx^2)^n \stackrel{(Q6)}{=} \frac{1}{\sqrt{1 - 4pqx^2}}.$$

Il est manifeste que F ne s'annule pas sur $] - 1, 1[$. Son dénominateur ne s'annule pas non plus car, par Q7, F est bien définie sur $] - 1, 1[$. De l'expression de F , on déduit celle de G :

$$G(x) \stackrel{(Q10)}{=} 1 - \frac{1}{F(x)} = 1 - \sqrt{1 - 4pqx^2}.$$

Q19. Notons que, G étant continue sur $[-1, 1]$ en vertu de la question Q8, l'expression de G obtenue à la question précédente est en fait vraie pour tout $x \in [-1, 1]$. Alors,

$$\mathbf{P}(R = +\infty) = 1 - \mathbf{P}(R \neq +\infty) \stackrel{(Q8)}{=} 1 - G(1) \stackrel{(Q18)}{=} \sqrt{1 - 4pq}.$$

Or, $4pq = (p + q)^2 - (p - q)^2 = 1 - (p - q)^2$, d'où *in fine* $\mathbf{P}(R = +\infty) = |p - q|$.

Pour obtenir la suite de la distribution de la loi de R , on développe $G(x)$ en série entière. On pourrait, comme à la question Q6, utiliser le DSE de $(1 + x)^\alpha$ avec, cette fois, $\alpha = 1/2$, mais on peut aussi remarquer que

$$\forall x \in] - 1, 1[: G'(x) = \frac{8pqx}{2\sqrt{1 - 4pqx^2}} = 4pqxF(x) = \sum_{n=0}^{+\infty} 4 \binom{2n}{n} (pq)^{n+1} x^{2n+1}.$$

Compte tenu de $G(0) = \mathbf{P}(R = 0) = 0$, il vient

$$G(x) = \sum_{n=0}^{+\infty} \frac{2}{n+1} \binom{2n}{n} (pq)^{n+1} x^{2n+2} = \sum_{n=1}^{+\infty} \frac{2}{n} \binom{2n-2}{n-1} (pq)^n x^{2n}.$$

Par unicité du développement en série entière, il vient $\mathbf{P}(R = 2n + 1) = 0$ pour tout $n \in \mathbb{N}$, $\mathbf{P}(R = 0) = 0$ et

$$\forall n \in \mathbb{N}^* : \mathbf{P}(R = 2n) = \frac{2}{n} \binom{2n-2}{n-1} (pq)^n.$$

Q20. On suppose que $p = q = 1/2$, soit $pq = 1/4$. On reprend l'expression précédente, qui donne

$$\mathbf{P}(R = 2n + 2) = \frac{1}{(2n+2)4^n} \binom{2n}{n} \stackrel{(Q2)}{\underset{n \rightarrow \infty}{\sim}} \frac{1}{2n\sqrt{\pi n}} = \frac{1}{2\sqrt{\pi} n^{3/2}} \sim \frac{1}{2\sqrt{\pi} (n-1)^{3/2}} \sim \mathbf{P}(R = 2n).$$

et l'on peut donc retenir la forme la plus simple, soit $\mathbf{P}(R = 2n) \sim \frac{1}{2\sqrt{\pi} n^{3/2}}$.

Pour fixer les idées, même si ce n'est pas ici nécessaire, on peut noter que

$$\lim_{n \rightarrow \infty} \frac{\mathbf{E}(N_n)}{n} \stackrel{(Q16)}{=} \mathbf{P}(R = +\infty) \stackrel{(Q19)}{=} 0 \quad \therefore \quad \lim_{n \rightarrow \infty} \mathbf{E}(N_n) \stackrel{(Q15)}{=} \mathbf{E}(R) = +\infty,$$

par divergence de la série $\sum 2n \mathbf{P}(R = 2n)$, vu que $2n \mathbf{P}(R = 2n) = \Theta\left(\frac{1}{\sqrt{n}}\right)$. On a donc $\lim_{n \rightarrow \infty} \mathbf{E}(N_n) = +\infty$ et $\mathbf{E}(N_n) = o(n)$. Pour déterminer plus précisément un équivalent simple de $\mathbf{E}(N_n)$ lorsque n tend vers $+\infty$, il n'y a d'autre choix que d'utiliser le théorème suivant :

Soient $(u_n)_n$ et $(v_n)_n$ deux suites strictement positives telles que $u_n \sim v_n$. Par le théorème de comparaison, les séries $\sum u_n$ et $\sum v_n$ sont de même nature. De plus,

- si $\sum u_n$ converge, alors $\sum_{k=n}^{\infty} u_k \sim \sum_{k=n}^{\infty} v_k$;
- si $\sum u_n$ diverge, alors $\sum_{k=1}^n u_k \sim \sum_{k=1}^n v_k$.

Preuve. Soit $\varepsilon > 0$. Alors,

$$\exists n_0 \in \mathbb{N}, \forall k \in \mathbb{N} : n \geq n_0 \implies (1 - \varepsilon)v_k \leq u_k \leq (1 + \varepsilon)v_k \quad (*) \quad \therefore$$

$$\forall n \in \mathbb{N} : n \geq n_0 \implies (1 - \varepsilon) \sum_{k=n}^{\infty} v_k \leq \sum_{k=n}^{\infty} u_k \leq (1 + \varepsilon) \sum_{k=n}^{\infty} v_k,$$

ce qui prouve le premier item. Le second est un petit peu plus délicat : pour $n \geq n_0$,

$$\begin{aligned} \left| \sum_{k=1}^n u_k - \sum_{k=1}^n v_k \right| &\stackrel{(1)}{\leq} \sum_{k=1}^{n_0-1} |u_k - v_k| + \left| \sum_{k=n_0}^n u_k - \sum_{k=n_0}^n v_k \right| \\ &\stackrel{(2)}{\leq} \sum_{k=1}^{n_0-1} |u_k - v_k| + \varepsilon \sum_{k=n_0}^n v_k \stackrel{(3)}{\leq} \sum_{k=1}^{n_0-1} |u_k - v_k| + \varepsilon \sum_{k=1}^n v_k \stackrel{(4)}{\leq} 2\varepsilon \sum_{k=1}^n v_k. \end{aligned}$$

(1) par l'inégalité triangulaire ;

(2) par l'encadrement (*) issu de la définition de l'équivalence ;

(3) par positivité de $(v_k)_k$;

(4) pour tout $n \geq n_1$, l'existence de n_1 étant assurée par le fait que $\lim_{n \rightarrow \infty} \sum_{k=1}^n v_k = +\infty$ et que $\sum_{k=1}^{n_0-1} |u_k - v_k|$ est constant (ne dépend pas de n).

Ainsi, en notant $U_n = \sum_{k=1}^n u_k$ et $V_n = \sum_{k=1}^n v_k$, on a $U_n - V_n = o(V_n)$, soit $U_n \sim V_n$, ce qui termine la démonstration.

On va maintenant appliquer le résultat sur la sommation des équivalents à la détermination d'un équivalent de $\mathbf{E}(N_n)$. Tout d'abord,

$$\mathbf{P}(R > i) = \sum_{j=i+1}^{\infty} \mathbf{P}(R = j) = \sum_{k=\lfloor \frac{i+1}{2} \rfloor}^{\infty} \mathbf{P}(R = 2k) \sim \frac{1}{2\sqrt{\pi}} \sum_{k=\lfloor \frac{i+1}{2} \rfloor}^{\infty} \frac{1}{k^{3/2}} \stackrel{(Q3)}{\sim} \frac{1}{\sqrt{\pi}} \left[\frac{i+1}{2} \right]^{-1/2} \sim \frac{\sqrt{2}}{\sqrt{\pi}} \frac{1}{\sqrt{i}}$$

en utilisant la loi de R et en appliquant le résultat ci-dessus à la série de Riemann convergente d'exposant $3/2 > 1$. Dans un deuxième temps, en appliquant cette fois la sommation d'équivalents à la série de Riemann divergente d'exposant $1/2 < 1$, il vient

$$\mathbf{E}(N_n) \sim \sum_{i=1}^n \mathbf{P}(R > i) \stackrel{(Q15)}{\sim} \frac{\sqrt{2}}{\sqrt{\pi}} \sum_{i=1}^n \frac{1}{\sqrt{i}} \stackrel{(Q3)}{\sim} \frac{2\sqrt{2}}{\sqrt{\pi}} \sqrt{n}.$$

Remarque. La question Q3 peut aussi se traiter avec le théorème de sommation des équivalents en calculant, pour $\alpha \in \mathbb{R}_+^* \setminus \{1\}$,

$$\frac{1}{k^{\alpha-1}} - \frac{1}{(k-1)^{\alpha-1}} = \frac{1}{k^{\alpha-1}} \left[1 - \left(1 - \frac{1}{k} \right)^{1-\alpha} \right] \underset{k \rightarrow \infty}{\sim} \frac{1-\alpha}{k^{\alpha}} \quad \therefore \quad \frac{1}{k^{\alpha}} \underset{k \rightarrow \infty}{\sim} \frac{1}{(\alpha-1)k^{1-\alpha}}.$$

On transforme ainsi la somme partielle (ou le reste) de la série de Riemann en une somme télescopique.

Le théorème donne aussi une preuve immédiate de l'équivalent de $I(x)$ à la question Q5, à partir de l'expression de la question Q4.

D. UN RÉSULTAT ASYMPTOTIQUE

Soient $(a_n)_{n \in \mathbb{N}}$ et $(b_n)_{n \in \mathbb{N}}$ deux suites d'éléments de \mathbb{R}_+^* . On suppose que $(a_n)_{n \in \mathbb{N}}$ est décroissante et que $\sum_{k=0}^n a_k b_{n-k} = 1$ pour tout $n \in \mathbb{N}$. Pour $n \in \mathbb{N}$, on pose $B_n = \sum_{k=0}^n b_k$.

Q 21. Pour tout $n \geq 1$,

$$1 = \sum_{k=0}^n a_k b_{n-k} \geq a_n \sum_{k=0}^n b_{n-k} = a_n B_n \quad \therefore \quad a_n \leq \frac{1}{B_n}.$$

Soient maintenant m et n deux entiers naturels tels que $m > n$. Alors,

$$1 = \sum_{k=0}^m a_k b_{m-k} = \sum_{k=0}^{n-1} a_k b_{m-k} + \sum_{k=n}^m a_k b_{m-k} \geq a_0 \sum_{k=0}^{n-1} b_{m-k} + a_n \sum_{k=n}^m b_{m-k} = a_0 (B_m - B_{m-n}) + a_n B_{m-n}.$$

Q 22. On suppose qu'il existe $(m_n)_{n \in \mathbb{N}}$ vérifiant $m_n > n$ pour n assez grand ainsi que

$$B_{m_n - n} \underset{n \rightarrow +\infty}{\sim} B_n \text{ et } B_{m_n} - B_{m_n - n} \xrightarrow[n \rightarrow +\infty]{} 0.$$

D'après la question Q 21,

$$1 \geq a_n B_n \sim a_n B_{m_n - n} \geq 1 - a_0(B_{m_n} - B_{m_n - n}) = 1 - \mathfrak{o}(1) \quad \therefore \quad \lim a_n B_n = 1.$$

Q 23. On suppose dans cette question qu'il existe $C > 0$ tel que $b_n \underset{n \rightarrow +\infty}{\sim} \frac{C}{n}$. Pour appliquer la question précédente, il faut trouver une suite $(m_n)_n$ qui vérifie le cahier des charges. Par sommation des équivalents et divergence de la série harmonique, on a $B_n \sim C \ln n$ (le cas n'est pas couvert par la question 3, mais la preuve est essentiellement la même, donc on ne détaille pas à ce stade du problème). La condition $\lim B_{m_n} - B_{m_n - n} = 0$ donne (sans se soucier de considérer des valeurs entières pour le moment)

$$\ln(m_n) - \ln(m_n - n) = -\ln\left(1 - \frac{n}{m_n}\right) \xrightarrow[n \rightarrow \infty]{} 0 \iff n = \mathfrak{o}(m_n).$$

Posons $m_n = nq_n$. La condition ci-dessus devient $\lim q_n = +\infty$. Par ailleurs,

$$\ln(m_n - n) = \ln n + \ln(q_n - 1) \sim \ln n \iff \ln(q_n) \sim \ln(q_n - 1) = \mathfrak{o}(\ln n)$$

Les suites $(m_n)_n$ qui répondent à la question sont donc les suites entières de la forme $m_n = nq_n$ avec $\lim q_n = +\infty$ et $\ln(q_n) = \mathfrak{o}(\ln n)$. Par exemple, la suite $m_n = \lfloor n \ln n \rfloor$ convient et l'on a donc $a_n \sim \frac{1}{C \ln n}$.

E. LA MARCHE ALÉATOIRE SIMPLE SUR \mathbb{Z}^2 : UN THÉORÈME D'ERDŐS ET DVORESTSKY

Q 24. Identifier un produit de Cauchy dans la somme de gauche incite à se lancer dans un calcul de séries génératrices.

Il est clair que $\sum_{k=0}^n \mathbf{P}(S_k = 0_d) \mathbf{P}(R > n - k) \in [0, n + 1]$, donc la série entière associée est de rayon de convergence au moins 1. Pour tout $x \in]-1, 1[$, on reconnaît divers produits de Cauchy et l'on a

$$\begin{aligned} \sum_{n=0}^{\infty} \sum_{k=0}^n \mathbf{P}(S_k = 0_d) \mathbf{P}(R > n - k) x^n &= F(x) \sum_{n=0}^{\infty} \mathbf{P}(R > n) x^n = F(x) \left[\sum_{n=0}^{\infty} x^n - \sum_{n=0}^{\infty} \mathbf{P}(R \leq n) x^n \right] \\ &= F(x) \left[\sum_{n=0}^{\infty} x^n - \sum_{n=0}^{\infty} \sum_{k=0}^n \mathbf{P}(R = k) x^n \right] = \frac{F(x)(1 - G(x))}{1 - x} \stackrel{(Q10)}{=} \frac{1}{1 - x}, \end{aligned}$$

d'où

$$\sum_{k=0}^n \mathbf{P}(S_k = 0_d) \mathbf{P}(R > n - k) = 1$$

pour tout $n \in \mathbb{N}$ par unicité du développement en série entière (le rayon de convergence est non nul).

On pouvait aussi raisonner directement en s'inspirant du raisonnement fait à la question Q 14 en posant

$$D = \max \{k \in \llbracket 0, n \rrbracket, S_k = 0_d\}.$$

L'ensemble $\{k \in \llbracket 0, n \rrbracket, S_k = 0_d\}$ est non vide puisque $S_0 = 0_d$, de sorte que D est bien définie, et est une variable aléatoire à valeurs dans $\llbracket 0, n \rrbracket$ en tant que fonction de (X_1, \dots, X_n) . Pour tout $k \in \llbracket 0, n - 1 \rrbracket$, on a

$$(D = k) = \{S_k = 0_d, S_{k+1} \neq 0_d, \dots, S_n \neq 0_d\} = \{S_k = 0_d, S_{k+1} - S_k \neq 0_d, \dots, S_n - S_k \neq 0_d\}.$$

Comme $S_{k+1} - S_k, \dots, S_n - S_k$ sont des fonctions de (X_{k+1}, \dots, X_n) , elles sont indépendantes de S_k (fonction de (X_1, \dots, X_k)) par le lemme des coalitions, de sorte que

$$\mathbf{P}(D = k) = \mathbf{P}(S_k = 0_d, S_{k+1} - S_k \neq 0_d, \dots, S_n - S_k \neq 0_d) = \mathbf{P}(S_k = 0_d) \mathbf{P}(S_{k+1} - S_k \neq 0, \dots, S_n - S_k \neq 0_d).$$

Comme vu à la question Q 14, $(S_{k+1} - S_k, \dots, S_n - S_k \neq 0_d)$ a même loi que (S_1, \dots, S_{n-k}) , si bien que

$$\mathbf{P}(D = k) = \mathbf{P}(S_k = 0_d) \mathbf{P}(S_1 \neq 0, \dots, S_{n-k} \neq 0_d) = \mathbf{P}(S_k = 0_d) \mathbf{P}(R > n - k).$$

La loi de D étant une probabilité sur $\llbracket 0, n \rrbracket$, il en découle bien

$$1 = \sum_{k=0}^n \mathbf{P}(D = k) = \sum_{k=0}^n \mathbf{P}(S_k = 0_d) \mathbf{P}(R > n - k).$$

Dans les deux dernières questions, on suppose que $d = 2$ et que la loi de X est donnée par

$$\mathbf{P}(X = (0, 1)) = \mathbf{P}(X = (0, -1)) = \mathbf{P}(X = (1, 0)) = \mathbf{P}(X = (-1, 0)) = \frac{1}{4}.$$

Q25. Soit $n \in \mathbb{N}$. La variable aléatoire S_{2n} compte $2n$ déplacements. Comme il y a à chaque étape quatre mouvements possibles, le nombre total de possibilités est donc 4^{2n} . L'événement $(S_{2n} = 0_2)$ traduit autant de déplacements vers le haut que vers le bas, disons $p \in \llbracket 0, n \rrbracket$ et autant de déplacements vers la gauche que vers la droite, soit $n - p$. Il y a $\binom{2n}{2p}$ façons de choisir les déplacements verticaux et, parmi ceux-là, $\binom{2p}{p}$ façons de choisir ceux qui seront faits vers le haut. Le même raisonnement s'appliquant aux déplacements horizontaux, il vient

$$\begin{aligned} \mathbf{P}(S_{2n} = 0_d) &= \frac{1}{4^{2n}} \sum_{p=0}^n \binom{2n}{2p} \binom{2p}{p} \binom{2n-2p}{n-p} = \frac{1}{4^{2n}} \sum_{p=0}^n \frac{(2n)! (2p)! (2n-2p)!}{(2p)! (2n-2p)! p!^2 (n-p)!^2} \\ &= \frac{1}{4^{2n}} \binom{2n}{n} \sum_{p=0}^n \frac{n!^2}{p!^2 (n-p)!^2} = \frac{1}{4^{2n}} \binom{2n}{n} \sum_{p=0}^n \binom{n}{p}^2 \stackrel{(Q1)}{=} \frac{1}{4^{2n}} \binom{2n}{n}^2. \end{aligned}$$

Les déplacements horizontaux et verticaux sont incompatibles et non indépendants. Il n'est donc nullement évident que la probabilité obtenue soit le carré de celle obtenue à la question 17 en dimension 1.

Toutefois, on peut s'y ramener par une astuce géométrique en utilisant le changement de base de matrice P avec $P^{-1} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. En notant $X'_i = (A_i, B_i)$ les coordonnées de X_i dans la nouvelle base, la formule $X' = P^{-1}X$, les déplacements $(0, 1)$, $(0, -1)$, $(1, 0)$ et $(-1, 0)$ deviennent alors respectivement $(1, 1)$, $(-1, -1)$, $(1, -1)$ et $(-1, 1)$. L'indépendance de $(X_i)_{i \geq 1}$ se transfère à $(X'_i)_{i \geq 1}$ par le théorème de transfert (attention à l'orthographe), donc $(X'_i)_{i \geq 1}$ est une suite de v.a.i.i.d. Par ailleurs, un calcul immédiat montre que A_i et B_i sont de même loi $\mathcal{U}(\{-1, 1\})$ (loi de Rademacher) et que $A_i \perp B_i$, ce qui corrige le défaut d'indépendance constaté sur les coordonnées de X .

Alors, pour $S'_n = \sum_{i=1}^n X'_i$,

$$\begin{aligned} \mathbf{P}(S_{2n} = 0_2) &= \mathbf{P}(S'_{2n} = 0_2) = \mathbf{P}[(A_1 + \dots + A_{2n} = 0) \cap (O_1 + \dots + O_{2n} = 0)] \\ &= \mathbf{P}(A_1 + \dots + A_{2n} = 0) \mathbf{P}(O_1 + \dots + O_{2n} = 0) = \mathbf{P}(S_{2n}^{(1)} = 0)^2 = \frac{1}{4^{2n}} \binom{2n}{n}^2, \end{aligned}$$

où $S_{2n}^{(1)}$ désigne la v.a. S_{2n} unidimensionnelle de la partie C.

Q 26. Il semble clair qu'il faille ici utiliser la partie D. On part ainsi *a priori* de $a_n = \mathbf{P}(R > n)$, qui est bien strictement positive et décroissante et de $b_n = \mathbf{P}(S_n = 0_2)$... qui est bien positive, mais pas strictement positive et qui ne risque pas d'être équivalente à Cn^{-1} pour n impair. On s'adapte et l'on part donc finalement de la relation

$$1 = \sum_{k=0}^{2n} \mathbf{P}(S_k = 0_2) \mathbf{P}(R > 2n - k) = \sum_{k=0}^n \mathbf{P}(S_{2k} = 0_2) \mathbf{P}(R > 2n - 2k) \stackrel{(j=n-k)}{=} \sum_{j=0}^n \underbrace{\mathbf{P}(R > 2j)}_{a_j} \underbrace{\mathbf{P}(S_{2n-2j} = 0_2)}_{b_{n-j}}.$$

Cette fois, les deux suites sont strictement positives et l'on a

$$b_n = \mathbf{P}(S_{2n} = 0_2) \stackrel{(Q25)}{=} 4^{-n} \binom{2n}{n}^2 \stackrel{(Q2)}{\sim} \frac{1}{\pi n}.$$

La question Q 23 s'applique et donne $a_n = \mathbf{P}(R > 2n) \sim \frac{\pi}{\ln n}$. Comme $\mathbf{P}(R > 2n+2) \leq \mathbf{P}(R > 2n+1) \leq \mathbf{P}(R > 2n)$ (en fait, $\mathbf{P}(R > 2n) = \mathbf{P}(R > 2n+1)$, mais on n'en a pas besoin) et que $\ln n \sim \ln(n+1)$, il s'ensuit, en utilisant à nouveau une comparaison série-intégrale (la fonction $1/\ln$ est décroissante) et une sommation d'équivalents,

$$\mathbf{P}(R > n) \sim \frac{\pi}{\ln \lfloor n/2 \rfloor} \sim \frac{\pi}{\ln n} \quad \therefore \quad \mathbf{E}(N_n) \stackrel{(Q15)}{=} 1 + \sum_{i=1}^n \mathbf{P}(R > i) \sim \pi \sum_{i=2}^n \frac{1}{\ln i} \sim \int_2^n \frac{dt}{\ln t} = I(n) \stackrel{(Q5)}{\sim} \frac{\pi n}{\ln n}.$$