

Jeu de pile ou face et problème du collectionneur de coupons

Jean Vereecke & Gael Druez

30 mars 2022

On s'intéresse dans cette lecture dirigée à la loi des évènements rares ainsi qu'à la notion de marche aléatoire.

Table des matières

I	Loi des évènements rares	2
A	Somme de variables aléatoires réelles indépendantes	2
B	Distance en variation totale	3
C	Inégalité de Le Cam	6
D	Applications et exemples	9
D.1	Lois des évènements rares	9
D.2	Paradoxe des anniversaires	9
D.3	Production à la chaîne	10
II	Marches aléatoires et loi de l'arcsinus	11
A	Chemins de \mathbb{Z}	11
B	Marche aléatoire dans \mathbb{Z}	14
C	Loi de l'arcsinus pour le dernier retour en zéro	16
D	Preuve de la loi de l'arcsinus	19

I Loi des évènements rares

Le théorème limite central donne une propriété importante de toute suite de variables aléatoires indépendantes et identiquement distribuées. Cependant, dans le cas d'évènements de probabilité très faible (proche de 0) ou très forte (proche de 1), il se peut que cette convergence soit très lente et qu'il soit plus intéressant de s'intéresser à un autre modèle.

Nous allons, dans cette partie, nous intéresser à la convergence en loi des sommes partielles d'une suite de Bernoulli indépendantes, qui modélise l'évènement de probabilité très faible ou très forte en utilisant la notion de *distance en variation totale* dont la notion de convergence coïncide avec celle de convergence en loi et ensuite on va montrer l'*inégalité de Le Cam* qui, elle, permet de conclure quand à la vitesse de convergence des sommes partielles.

Dans un premier temps, nous allons toutefois nous intéresser, dans le cas général, à la somme de variables aléatoires réelles indépendantes.

A Somme de variables aléatoires réelles indépendantes

Les résultats de cette partie ne sont pas difficiles à obtenir mais sont cependant très utiles pour toute la suite du papier puisqu'ils permettent facilement d'obtenir la loi d'une somme de variables aléatoires réelles sans trop de difficulté, il suffit d'utiliser le produit de convolution. De plus, on va pouvoir montrer la stabilité de la loi de Poisson par somme.

Propriété I.1. Soient X et Y deux variables aléatoires réelles indépendantes de lois respectives \mathbb{P}_X et \mathbb{P}_Y , alors $X + Y$ est une variable aléatoire réelle de loi $\mathbb{P}_{X+Y} = \mathbb{P}_X * \mathbb{P}_Y$.

Preuve. Soit $A \in \mathcal{B}(\mathbb{R})$ un borélien de \mathbb{R} , on a

$$\begin{aligned}\mathbb{P}_{X+Y}(A) &= \mathbb{P}(X + Y \in A) = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_A(x + y) d\mathbb{P}_{(X,Y)}(x, y) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_A(x + y) \mathbb{P}_X(dx) \mathbb{P}_Y(dy) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbb{1}_{A-y}(x) \mathbb{P}_X(dx) \right) \mathbb{P}_Y(dy) \\ &= \int_{\mathbb{R}} \mathbb{P}_X(A - y) \mathbb{P}_Y(dy) = \mathbb{P}_X * \mathbb{P}_Y(A).\end{aligned}$$

Ainsi la loi de $X + Y$ est $\mu * \nu$.

Comme le produit de convolution est associatif, par itération de la propriété précédente, on obtient le résultat suivant :

Corollaire I.2. Soient $n \geq 0$, X_0, \dots, X_n des variables aléatoires réelles indépendantes de lois respectives μ_0, \dots, μ_n , alors $\sum_{i=0}^n X_i$ est une variable aléatoire réelle de loi $\mu_0 * \dots * \mu_n$.

Cela va nous servir dès que l'on cherchera la loi de la somme de Bernoulli indépendantes.

Pour nous faire la main sur la propriété, on va montrer que l'ensemble des variables de Poisson sont stables par somme (dès que les opérandes sont indépendantes).

Corollaire I.3. Soient X et Y deux variables aléatoires réelles indépendantes suivant des lois de Poisson de paramètres respectifs λ et μ , alors $X + Y$ est une variable aléatoire réelle suivant une loi de Poisson de paramètres $\lambda + \mu$.

Preuve. Avec la propriété précédente, on a pour $k \in \mathbb{N}$:

$$\begin{aligned} \mathbb{P}_{X+Y}(\{k\}) &= \mathbb{P}_X * \mathbb{P}_Y(\{k\}) = \sum_{i=0}^k \mathbb{P}_X(\{i\}) \mathbb{P}_Y(\{k-i\}) \\ &= \sum_{i=0}^k e^{-(\lambda+\mu)} \frac{\lambda^i}{i!} \frac{\mu^{k-i}}{(k-i)!} = e^{-(\lambda+\mu)} \frac{(\lambda + \mu)^k}{k!}. \end{aligned}$$

B Distance en variation totale

Les résultats de la partie précédente sont importants pour tout ce qui va être développé dans la suite, cependant puisque l'on va s'intéresser à la convergence en loi de sommes de variables indépendantes, on voudrait également avoir une notion de vitesse de convergence. Pour cela, on va métriser la topologie de la convergence. Ce faisant, on va se placer dans un cadre général.

Soit E un ensemble au plus dénombrable (par la suite, on considérera $E \subset \mathbb{N}$). On munit E de la tribu et de la topologie discrètes.

Proposition - Définition I.4. Pour tout couple (μ, ν) de lois sur $(E, \mathcal{P}(E))$, le nombre

$$d_{VT}(\mu, \nu) := \sup_{A \subset E} |\mu(A) - \nu(A)|$$

est bien défini et l'application d_{VT} est alors une distance sur l'ensemble des lois sur E , ensemble qui, muni de cette distance, est donc un espace métrique.

Preuve. Pour tout $A \subset E$, $|\mu(A) - \nu(A)| \leq 1$ et donc le nombre $\sup_{A \subset E} |\mu(A) - \nu(A)|$ est correctement défini.

Montrons que d_{VT} est une distance sur les lois de E .

Soient μ, ν et λ trois lois sur $(E, \mathcal{P}(E))$.

— **Symétrie :**

$$d_{VT}(\mu, \nu) = \sup_{A \subset E} |\mu(A) - \nu(A)| = \sup_{A \subset E} |\nu(A) - \mu(A)| = d_{VT}(\nu, \mu).$$

— **Séparation :**

$$d_{VT}(\mu, \nu) = 0 \Leftrightarrow \forall A \subset E, |\mu(A) - \nu(A)| = 0 \Leftrightarrow \mu = \nu.$$

— **Inégalité triangulaire :**

$$\begin{aligned} d_{VT}(\mu, \lambda) &= \sup_{A \subset E} |\mu(A) - \lambda(A)| = \sup_{A \subset E} |\mu(A) - \lambda(A) - (\lambda(A) - \nu(A))| \\ &\leq \sup_{A \subset E} |\mu(A) - \lambda(A)| + \sup_{A \subset E} |\lambda(A) - \nu(A)| \\ &\leq d_{VT}(\mu, \nu) + d_{VT}(\nu, \lambda). \end{aligned}$$

Ainsi d_{VT} est bien une distance sur les lois sur $(E, \mathcal{P}(E))$.

Cette première définition demande, pour calculer la distance en variation totale, de regarder toutes les sous-parties de E (qui sont en nombre indénombrable) et de faire le calcul d'une borne supérieure, ce qui peut être laborieux selon les lois étudiés. Le théorème suivant permet de calculer la distance en variation totale d'une manière plus simple.

Théorème I.5. Si μ et ν sont des lois sur E alors

$$d_{VT}(\mu, \nu) = \frac{1}{2} \sup_{f: E \rightarrow [-1, 1]} \left| \int_E f d\mu - \int_E f d\nu \right| = \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)|.$$

De plus dans la définition initiale, la borne supérieure est atteinte en $A_* = \{x \in E : \mu(x) \geq \nu(x)\}$ et dans l'expression fonctionnelle précédente, la borne supérieure est atteinte en $f = \mathbb{1}_{A_*} - \mathbb{1}_{A_*^c}$.

Preuve. Pour $f : E \rightarrow [-1, 1]$,

$$\left| \int_E f d\mu - \int_E f d\nu \right| \leq \sum_{x \in E} |f(x)| |\mu(x) - \nu(x)| \leq \sum_{x \in E} |\mu(x) - \nu(x)|$$

et on obtient des égalités pour $f = \mathbb{1}_{A_*} - \mathbb{1}_{A_*^c}$.

En effet,

$$\begin{aligned} \left| \int_E (\mathbb{1}_{A_*} - \mathbb{1}_{A_*^c}) d\mu - \int_E (\mathbb{1}_{A_*} - \mathbb{1}_{A_*^c}) d\nu \right| &= |\mu(\mathbb{1}_{A_*}) - \mu(\mathbb{1}_{A_*^c}) - \nu(\mathbb{1}_{A_*}) - \nu(\mathbb{1}_{A_*^c})| \\ &= |\mu(\mathbb{1}_{A_*}) - \nu(\mathbb{1}_{A_*})| + |\mu(\mathbb{1}_{A_*^c}) - \nu(\mathbb{1}_{A_*^c})|. \end{aligned}$$

On obtient ainsi la deuxième égalité.

Pour $A \subset E$,

$$|\mu(A) - \nu(A)| = \frac{1}{2} \left| \int_E f_A d\mu - \int_E f_A d\nu \right|$$

où $f = \mathbb{1}_A - \mathbb{1}_{A^c}$ ainsi

$$|\mu(A) - \nu(A)| \leq \frac{1}{2} \sup_{f: E \rightarrow [-1,1]} \left| \int_E f d\mu - \int_E f d\nu \right| = \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)|$$

où l'égalité est obtenue en $A = A_*$, en effet

$$\begin{aligned} 2|\mu(A_*) - \nu(A_*)| &= \mu(A_*) - \nu(A_*) + \nu(A_*^c) - \mu(A_*^c) \\ &= \sum_{x \in A_*} |\mu(x) - \nu(x)| + \sum_{x \in A_*^c} |\mu(x) - \nu(x)| = \sum_{x \in E} |\mu(x) - \nu(x)| \end{aligned}$$

Avec ce théorème, la calcul de la distance en variation totale se réduit à une somme dénombrable qui est *a priori* plus simple à réaliser pour des lois particulières. Par exemple :

Exemple I.1. Pour $p \in (0, 1]$, calculons $d_{VT}(\mathcal{B}(p), \mathcal{P}(b))$.

Soit $\mu \sim \mathcal{B}(p)$ et $\nu \sim \mathcal{P}(p)$, on a alors

$$\begin{aligned} d_{VT}(\mu, \nu) &= \frac{1}{2} \sum_{k \in \mathbb{N}} |\mu(k) - \nu(k)| = \frac{1}{2} \left[|1 - p - e^{-p}| + |p - e^{-p}| + e^{-p} \sum_{k \geq 2} \frac{p^k}{k!} \right] \\ &= \frac{1}{2} \left[|1 - p - e^{-p}| + |p - e^{-p}| + e^{-p} \sum_{k \geq 2} \frac{p^k}{k!} \right] \\ &= \frac{1}{2} \left[-1 + p + e^{-p} + p - e^{-p} + e^{-p} \sum_{k \geq 2} \frac{p^k}{k!} \right] \\ &= \frac{1}{2} \left[2p - 1 - 2pe^{-p} + e^{-p} \sum_{k \geq 0} \frac{p^k}{k!} \right] = \frac{1}{2} [2p - 1 - 2pe^{-p} + 1] \end{aligned}$$

On a donc

$$d_{VT}(\mu, \nu) = p(1 - e^{-p}).$$

Notons qu'en particulier, on a $d_{VT}(\mathcal{B}(p), \mathcal{P}(b)) \leq p^2$.

On sait que la distance en variation totale fournit alors une topologie sur l'ensemble des lois sur E (qui est donc la topologie engendrée par ses boules). On peut alors s'intéresser à ce que signifie la notion de convergence dans cet ensemble. On a alors :

Théorème I.6 (Convergence en loi). Si l'on pose $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires sur E de lois respectives $(\mu_n)_{n \in \mathbb{N}}$, alors pour toute loi μ sur E , les propositions suivantes sont équivalentes :

1. X_n converge en loi vers μ .

2. Pour tout $x \in E$, $\lim_{n \rightarrow \infty} \mu_n(x) = \mu(x)$.
3. Pour tout $f : E \rightarrow \mathbb{R}$ bornée, $\lim_{n \rightarrow \infty} \int_E f d\mu_n = \int_E f d\mu$.
4. $\lim_{n \rightarrow \infty} d_{VT}(\mu_n, \mu) = 0$.

Remarquons avant de prouver ce théorème, qu'il permet de conclure deux choses : la convergence en loi est métrisable et la distance en variation donne une façon de métriser l'espace de lois sur E pour la convergence en loi.

Preuve. 1. \Leftrightarrow 2. : il s'agit de la définition.

4. \Rightarrow 3. : On sait, grâce au théorème précédent que, pour tout $n \in \mathbb{N}$,

$$\begin{aligned} \frac{1}{2} \sup_{\substack{f: E \rightarrow \mathbb{R} \\ \text{bornée}}} \left| \int_E \left(\frac{f}{\sup|f|} \right) d\mu_n - \int_E \left(\frac{f}{\sup|f|} \right) d\mu \right| &= \frac{1}{2} \sup_{f: E \rightarrow [-1,1]} \left| \int_E f d\mu_n - \int_E f d\mu \right| \\ &= d_{VT}(\mu_n, \mu) \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

ce qui donne le résultat.

3. \Rightarrow 2. : Pour tout $x \in E$, $\mu_n(x) = \int_E \mathbb{1}_{\{x\}} d\mu_n \xrightarrow{n \rightarrow \infty} \int_E \mathbb{1}_{\{x\}} d\mu = \mu(x)$.

2. \Rightarrow 4. : Soit $A \subset E$,

$$\sum_{x \in E} |\mu_n(x) - \mu(x)| = \sum_{x \in A} |\mu_n(x) - \mu(x)| + \sum_{x \in A^c} |\mu_n(x) - \mu(x)|.$$

On vient de montrer que la convergence en loi est métrisable par d_{VT} , on va chercher à regarder la vitesse de convergence dans notre cadre particulier initial.

C Inégalité de Le Cam

L'objectif de cette partie est de démontrer le résultat suivant :

Théorème I.7 (Inégalité de Le Cam). Pour X_1, \dots, X_n de variables aléatoires indépendantes de lois de Bernoulli de paramètres p_1, \dots, p_n . Soit μ_n la loi de $S_n = X_1 + \dots + X_n$, et soit $\nu_n = \mathcal{P}(p_1 + \dots + p_n)$ la loi de Poisson de même moyenne que S_n . Alors :

$$d_{VT}(\mu_n, \nu_n) \leq p_1^2 + \dots + p_n^2$$

Preuve. La preuve utilise plusieurs résultats intermédiaires vus précédemment.

Etape 1 : On montre, en procédant par récurrence sur n , que si $\alpha_1, \dots, \alpha_n$ sont des lois de probabilités sur \mathbb{N} , alors :

$$d_{VT}(\alpha_1 * \dots * \alpha_n, \beta_1 * \dots * \beta_n) \leq d_{VT}(\alpha_1, \beta_1) + \dots + d_{VT}(\alpha_n, \beta_n)$$

Le cas $n = 1$ est immédiat, et par associativité du produit de convolution des lois sur \mathbb{N} , il suffit de montrer le résultat pour $n = 2$. On veut donc montrer que pour des lois de probabilité sur \mathbb{N} $\alpha_1, \alpha_2, \beta_1, \beta_2$, on a $d_{VT}(\alpha_1 * \alpha_2, \beta_1 * \beta_2) \leq d_{VT}(\alpha_1, \beta_1) + d_{VT}(\alpha_2, \beta_2)$.

D'après l'expression avec une somme de la distance en variation totale,

$$\begin{aligned} d_{VT}(\alpha_1 * \alpha_2, \beta_1 * \beta_2) &= \frac{1}{2} \sum_{k \in \mathbb{N}} \left| \sum_{l=0}^k \alpha_1(l) \alpha_2(k-l) - \beta_1(l) \beta_2(k-l) \right| \\ &= \frac{1}{2} \sum_{k \in \mathbb{N}} \left| \sum_{l=0}^k \alpha_1(l) \alpha_2(k-l) - \alpha_1(l) \beta_2(k-l) \right. \\ &\quad \left. + \alpha_1(l) \beta_2(k-l) - \beta_1(l) \beta_2(k-l) \right| \\ &\leq \frac{1}{2} \sum_{k \in \mathbb{N}} \left[\sum_{l=0}^k |\alpha_1(l) \alpha_2(k-l) - \alpha_1(l) \beta_2(k-l)| \right. \\ &\quad \left. + \sum_{l=0}^k |\alpha_1(l) \beta_2(k-l) - \beta_1(l) \beta_2(k-l)| \right] \end{aligned}$$

On a cela par inégalité triangulaire. De plus, comme les lois de probabilités sont des mesures positives, on a alors, en factorisant,

$$\begin{aligned} d_{VT}(\alpha_1 * \alpha_2, \beta_1 * \beta_2) &\leq \frac{1}{2} \sum_{k \in \mathbb{N}} \sum_{l=0}^k |\alpha_2(k-l) - \beta_2(k-l)| \alpha_1(l) + \frac{1}{2} \sum_{k \in \mathbb{N}} \sum_{l=0}^k |\alpha_1(l) - \beta_1(l)| \beta_2(k-l) \\ &\leq \frac{1}{2} \left(\sum_{k \in \mathbb{N}} |\alpha_2(k) - \beta_2(k)| \right) \left(\sum_{l \in \mathbb{N}} \alpha_1(l) \right) + \frac{1}{2} \left(\sum_{k \in \mathbb{N}} |\alpha_1(k) - \beta_1(k)| \right) \left(\sum_{l \in \mathbb{N}} \beta_1(l) \right) \end{aligned}$$

On obtient cette dernière ligne par produit de Cauchy.

Or α_1 et β_1 sont des lois de probabilité, donc

$$\sum_{l \in \mathbb{N}} \alpha_1(l) = \sum_{l \in \mathbb{N}} \beta_1(l) = 1.$$

On reconnaît dans les deux autres sommes la formule de la distance en variation totale, ce qui donne bien :

$$d_{VT}(\alpha_1 * \alpha_2, \beta_1 * \beta_2) \leq \frac{1}{2} \sum_{k \in \mathbb{N}} |\alpha_2(k) - \beta_2(k)| + \frac{1}{2} \sum_{k \in \mathbb{N}} |\alpha_1(k) - \beta_1(k)| = d_{VT}(\alpha_1, \alpha_2) + d_{VT}(\beta_1, \beta_2).$$

Etape 2 : Rappel des résultats intermédiaires

1. On a vu dans la propriété I.1, ainsi que dans son corollaire, que la loi d'une somme de variables aléatoires indépendantes de loi $\mathcal{B}(p_1), \dots, \mathcal{B}(p_n)$ est une variable de loi $\mathcal{B}(p_1) * \dots * \mathcal{B}(p_n)$, c'est-à-dire :

Si l'on note $X_1 \sim \mathcal{B}(p_1), \dots, X_n \sim \mathcal{B}(p_n)$, on a alors

$$\forall k \in \mathbb{N}, \mathbb{P}(X_1 + \dots + X_n = k) = \sum_{1 \leq i_1 < \dots < i_k \leq n} \prod_{j=1}^k p_{i_j} \times \prod_{\substack{j=1 \\ j \notin \{i_1, \dots, i_k\}}}^n (p_{i_j}).$$

2. On a également prouvé dans le corollaire I.3 que, pour $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ indépendantes, on a $X+Y \sim \mathcal{P}(\lambda+\mu)$. En itérant, on obtient que la loi de la somme de variables indépendantes de lois $\mathcal{P}(\lambda_1), \dots, \mathcal{P}(\lambda_n)$ est $\mathcal{P}(\lambda_1 + \dots + \lambda_n)$.
3. On a enfin montré dans l'exemple I.1 que $d_{VT}(\mathcal{B}(p), \mathcal{P}(b)) \leq p^2$.

Etape 3 : Conclusion.

La loi μ_n étant la loi d'une somme de variables indépendantes de loi $\mathcal{B}(p_1), \dots, \mathcal{B}(p_n)$, par l'étape 2.1, on obtient que $\mu_n = \mathcal{B}(p_1) * \dots * \mathcal{B}(p_n)$.

Avec 2.2, on obtient également que la loi de ν_n (qui est, on le rappelle, $\mathcal{P}(p_1 + \dots + p_n)$) est la loi d'une somme de variables indépendantes de lois $\mathcal{P}(p_1), \dots, \mathcal{P}(p_n)$. Il vient alors, d'après le même corollaire que ci-dessus, que $\nu_n = \mathcal{P}(p_1) * \dots * \mathcal{P}(p_n)$.

L'inégalité de l'étape 1 donne alors :

$$\begin{aligned} d_{VT}(\mu_n, \nu_n) &= d_{VT}[\mathcal{B}(p_1) * \dots * \mathcal{B}(p_n), \mathcal{P}(p_1) * \dots * \mathcal{P}(p_n)] \\ &\leq d_{VT}[\mathcal{B}(p_1), \mathcal{P}(p_1)] + \dots + d_{VT}[\mathcal{B}(p_n), \mathcal{P}(p_n)] \end{aligned}$$

L'étape 2.3 fournit finalement l'inégalité voulue :

$$d_{VT}(\mu_n, \nu_n) \leq p_1^2 + \dots + p_n^2$$

D Applications et exemples

D.1 Lois des évènements rares

Pour $(u_n)_n = \left(\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)_n$, en appliquant, en particulier, l'inégalité de Le Cam avec

$$\forall n \in \mathbb{N}, k \in \llbracket 1, n \rrbracket, p_k = u_n = \frac{\lambda}{n} + o\left(\frac{1}{n}\right),$$

il vient :

$$d_{VT}[\mathcal{B}(n, u_n), \mathcal{P}(\lambda)] \leq np_n^2 = \frac{\lambda^2}{n} + o\left(\frac{1}{n}\right) \xrightarrow{n \rightarrow \infty} 0.$$

On retrouve donc bien la convergence, quand n tend vers $+\infty$, de la loi binomiale de probabilité petite, vers la loi de poisson de même moyenne.

Regardons quelques exemples concrets pour analyser cette convergence.

D.2 Paradoxe des anniversaires

Lors d'un cours dans le village, Panoramax demande aux enfants : "Les quatre garnisons de Baaborum, de Aquarium, de Laudanum et de Petitbonum sont composés au total de 500 légionnaires. Quelle est la chance que k de ces légionnaires soient nés le premier jour de l'année?"

Modélisons ce problème : On étudie la probabilité que k personnes soient nées le premier jour de l'année.

On suppose que les anniversaires sont uniformément répartis dans la population, et que les 500 personnes naissent avec probabilité égale à chaque jour de l'année : on peut donc appliquer le modèle de la loi de Bernoulli, de paramètres $n = 500$ et $p = \frac{1}{365}$.

Comme on a pris n "grand", on s'attend à ce que les probabilités soient proches de celles d'une loi de Poisson de paramètre $\lambda = np = \frac{500}{365}$.

On peut comparer les valeurs des probabilités p_k que k personnes aient leur anniversaire le jour du nouvel an :

k	0	1	2	3	4	5	6
$\mathcal{B}(n, p)$	0.2537	0.3484	0.2388	0.1089	0.0372	0.0101	0.0023
$\mathcal{P}(np)$	0.2541	0.3481	0.2385	0.1089	0.0373	0.0102	0.0023

Maintenant, se pose la question de la pertinence du modèle : on sait, par exemple, que les anni-

versaires ne sont pas répartis uniformément dans l'année, et que cette irrégularité évolue au fil des décennies. Ainsi, on pourrait modifier notre p en fonction du jour que l'on regarde pour être plus précis.

D.3 Production à la chaîne

On considère maintenant une usine qui produit des objets à la chaîne. Chaque objet a une faible probabilité p , par exemple $p = 0.015$, d'être défectueux. Ainsi, selon la loi binomiale, une boîte de 100 objets a une probabilité de ne contenir aucun objet défectueux égale à $(0,985)^{100} \simeq 0,22061$. La loi de Poisson qui approche cette loi binomiale et qui a pour paramètre $\lambda = np = 0,015 \times 100 = 1,5$, donne quant à elle une probabilité de $e^{-1,5} \simeq 0,22313$: les deux quantités sont extrêmement proches !

On peut alors se demander combien d'objets sont nécessaires dans une boîte pour avoir une grande probabilité d'en avoir 100 qui soient en bon état. Fixons cette "grande" probabilité à 0,8. Notons x le nombre d'objets que l'on ajoute dans chaque boîte. On utilise l'approximation par la loi de Poisson, de paramètre $\lambda_x = (100 + x)p$, et on recherche alors le plus petit x tel que :

$$e^{-\lambda_x} \left(1 + \dots + \frac{(\lambda_x)^x}{x!} \right) \geq 0,8.$$

En faisant le calcul pour $x = 1$ et $x = 2$, on obtient que $x = 2$ convient puisque dans ce cas :

$$e^{-\lambda_2} \left(1 + \lambda_2 + \frac{(\lambda_2)^2}{2!} \right) \simeq 0,8013\dots \geq 0,8.$$

De plus, la probabilité de trouver au moins 100 clous conformes dans une boîte de 102 est $\simeq 0,8022$: l'approximation par la loi de Poisson est ici précise !

II Marches aléatoires et loi de l'arcsinus

Intéressons-nous aux marches aléatoires dans \mathbb{Z} : il s'agit de considérer une suite de décisions (par exemple une puce qui fait un saut dans une direction ou dans son opposée) qui sont prises de manière aléatoire.

Pour cela, on va tout d'abord s'intéresser aux chemins à n étapes dans \mathbb{Z} dans le cas général avant d'y ajouter une notion de probabilité. Nous allons ensuite nous intéresser aux événements particuliers que sont les différents retours en zéro. Enfin, on va chercher une écriture asymptotique de la probabilité du dernier retour en zéro et montrer la loi de l'arcsinus.

A Chemins de \mathbb{Z}

Considérons la situation suivante : le Grand Schtroumpf demande au Schtroumpf gourmand d'aller cueillir des feuilles de Salsepareille, la plante préférée de tous les Schtroumpfs, pour tout le village. Connaissant sa gourmandise, il demande au Schtroumpf à lunettes de bien vouloir aller le surveiller et lui donne une consigne :

"Tâche de ne pas te faire voir, et schtroumpfe simplement ceci dans ton carnet : Pour chaque feuille que le Schtroumpf gourmand cueille, s'il mange la feuille, note 1, s'il la met dans son panier, note -1."

Après que les deux Schtroumpfs aient accompli leurs missions, pour pouvoir savoir, s'il peut féliciter le Schtroumpf gourmand, le Grand Schtroumpf demande alors au Schtroumpf à lunette : "Schtroumpfe-moi la somme des plus ou moins un, pour savoir s'il a plus mangé ou s'il a plus cueilli."

Pour modéliser cette situation, on considère alors $n \in \mathbb{N}$, $\epsilon \in \{-1, 1\}^n$ et $s = (\epsilon_1 + \dots + \epsilon_k)_{k \in [1, n]}$.

Pour avoir une illustration de la situation, on associe à s une représentation graphique formée d'une ligne brisée (appelée *chemin*) dont les sommets ont pour coordonnées (k, s_k) .

On a alors facilement qu'il existe 2^n telles ϵ et s et si on note p le nombre de 1 et q le nombre de -1, on a : $n = p + q$ et $x = p - q = s_{p+q}$.

Dans le cas général, il existe plusieurs chemins définis de la manière précédente commençant en $(0, 0)$ et finissant en (n, x) , notons $N_{n, x}$ ce nombre.

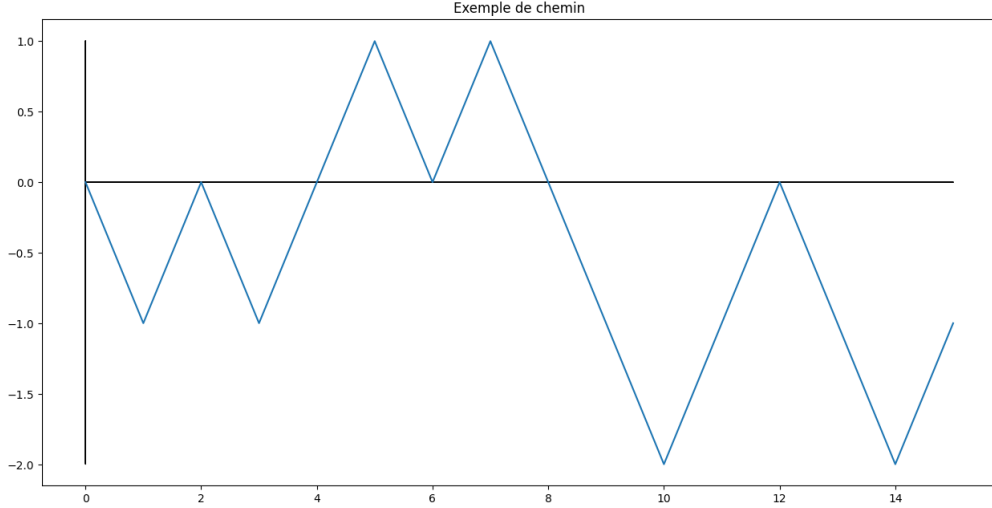


FIGURE 1 – Exemple de chemin

On obtient le résultat suivant :

$$\begin{cases} N_{n,x} = \binom{p+q}{p} = \binom{p+q}{q} & , \text{ si } n = p + q \text{ et } x = p - q \text{ pour } p, q \geq 0 \\ N_{n,x} = 0 & , \text{ sinon} \end{cases}$$

En effet, si de tels p et q n'existent pas (c'est-à-dire si $|x| > n$ ou si x et n n'ont pas la même parité), on a $N_{n,x} = 0$. Sinon il s'agit du nombre de manières de placer p fois le nombre $+1$ ou de placer q fois le nombre -1 parmi les $p + q$ nombres.

Le grand avantage de la représentation graphique est de nous permettre d'intuiter un résultat très intéressant pour la suite, comme :

Théorème II.1. Principe de réflexion

Soient un point $A(k, a)$, son symétrique $A'(k, -a)$ et un autre point $B(n, b)$, avec a de même signe que b et $k \leq n$. Alors il existe autant de chemins de A à B qui passent par zéro que de chemins de A' à B .

Preuve. La preuve de ce résultat est immédiate quand on regarde l'interprétation graphique (*cf.* Figure 2) : à tout chemin de A à B qui passe par zéro, on associe le chemin qui va de A' au premier retour en zéro en étant le symétrique du chemin de A à B , puis qui suit le premier chemin.

Concrètement, si $(s_k = a, s_{k+1}, \dots, s_{k+i} = 0, s_{k+i+1}, \dots, s_n = b)$ est un chemin de A à B passant par zéro, alors $(s'_k = -a, s'_{k+1} = -s_{k+1}, \dots, s'_{k+i} = 0, s'_{k+i+1} = s_{k+i+1}, \dots, s'_n = b)$ est un chemin de A' à B .

On peut faire la construction réciproque, d'où le résultat.

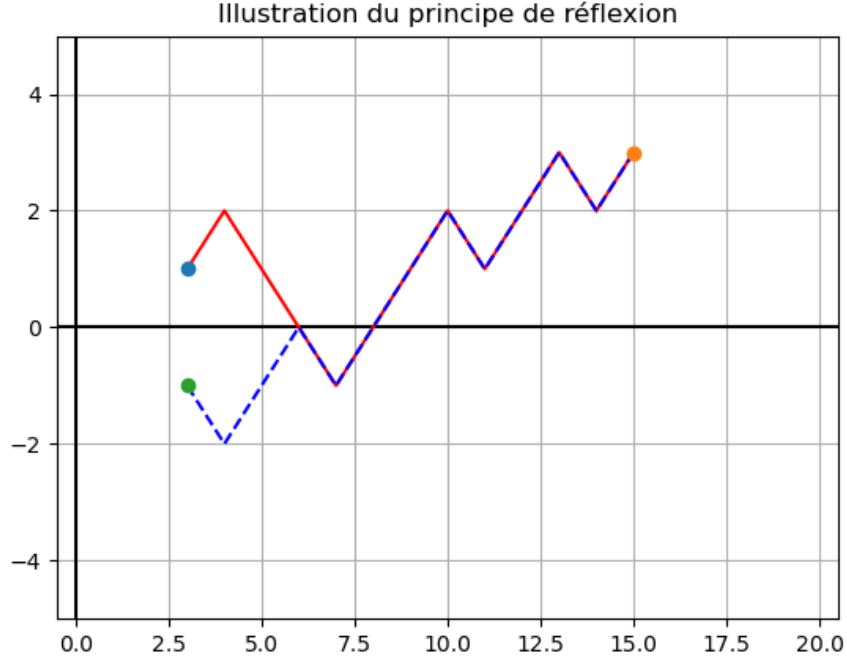


FIGURE 2 – Illustration du principe de réflexion avec $A(3, 1)$, $A'(3, -1)$ et $B(15, 3)$

Théorème II.2. Théorème du scrutin

Soient n et x des entiers strictement positifs. Alors il existe exactement $\frac{x}{n} N_{n,x}$ chemins de l'origine à (n, x) tels que $s_1 > 0, \dots, s_n > 0$.

Preuve. Le nombre de tels chemins est le même que le nombre de chemins du point $(1, 1)$ au point (n, x) qui ne passent jamais par zéro.

C'est-à-dire que c'est le nombre de chemins de $(1, 1)$ à (n, x) (c'est-à-dire $N_{n-1, x-1}$ en considérant $(1, 1)$ comme nouvelle origine), auquel on a retiré nombre de chemins de $(1, 1)$ à (n, x) passant par zéro (c'est-à-dire, par le principe de réflexion, le nombre de chemins de $(1, -1)$ à (n, x) , soit $N_{n-1, x+1}$ en changeant l'origine).

On a donc que le nombre de tels chemins vaut $N_{n-1, x-1} - N_{n-1, x+1}$, ou encore $\binom{p+q-1}{p-1} - \binom{p+q-1}{p}$ (où la définition p et q a été rappelée juste au-dessus : on a toujours $n = p + q$ et $x = p - q$).

Si on développe ces coefficients binomiaux, qu'on met en facteur ce qui va bien, on obtient

$$\frac{(p+q-1)!}{(p-1)!(q-1)!} \left(\frac{1}{q} - \frac{1}{p} \right) = \frac{(p+q)!}{(p-1)!(q-1)!} \frac{1}{p+q} \left(\frac{p-q}{pq} \right) = \frac{(p+q)!}{p!q!} \frac{p-q}{p+q} = \binom{p+q}{p} \frac{p-q}{p+q} = N_{n,x} \frac{x}{n}.$$

B Marche aléatoire dans \mathbb{Z}

Allons plus loin dans la modélisation du problème précédent, imaginons que le Schtroumpf gourmand sait qu'il sera observé mais veut tout de même profiter de la salsepareille. Pour se donner bonne conscience, il se dit :

"Je vais aller schtroumpfer une pièce au Schtroumpf financier et à chaque feuille que je cueille, je schtroumpfe la pièce, si c'est pile, je mets la feuille dans le panier, sinon je la mange."

Modélisons cette expérience : soit une suite $(\epsilon_n)_n$ de variables de Rademacher indépendantes, c'est-à-dire :

$$\forall k \in \mathbb{N}^*, \mathbb{P}(\epsilon_k = +1) = \mathbb{P}(\epsilon_k = -1) = \frac{1}{2}.$$

Posons alors, comme précédemment, pour $n \in \mathbb{N}$, $s_n = \epsilon_1 + \dots + \epsilon_n$.

Dans ce cadre, on peut s'intéresser à des évènements particuliers. Commençons par l'évènement "atteindre r en n étapes" $= \{s_n = r\}$, notons $p_{n,r} = \mathbb{P}(s_n = r)$ sa probabilité, on a alors :

$$p_{n,r} = N_{n,r} 2^{-n}.$$

(En effet, avec la modélisation choisie, les chemins sont choisis de manière équiprobable).

Un autre évènement qui va nous intéresser est le retour en zéro (qui correspond dans notre exemple au fait que le Schtroumpf gourmand a placé autant de feuilles dans son panier qu'au fond de son estomac). On note alors u_n la probabilité que cet évènement se produise en n .

On remarque très vite que pour n impair, $u_n = 0$ et que pour $k \in \mathbb{N}$,

$$u_{2k} = p_{2k,0} = \binom{2k}{k} 2^{-2k}$$

La formule de Stirling nous donne un équivalent de u_{2n} quand n devient grand :

$$u_{2n} = \frac{(2n)!}{(n!)^2} 2^{-2n} \sim \frac{\left(\frac{2n}{e}\right)^{2n} 2\sqrt{\pi n}}{\left(\frac{n}{e}\right)^{2n} 2\pi n} 2^{-2n} = \frac{1}{\sqrt{\pi n}}.$$

On a un résultat principal pour la suite de notre développement :

Théorème II.3. Main lemma

La probabilité qu'aucun retour en zéro n'arrive jusqu'à l'instant $2n$ est la même que la probabilité

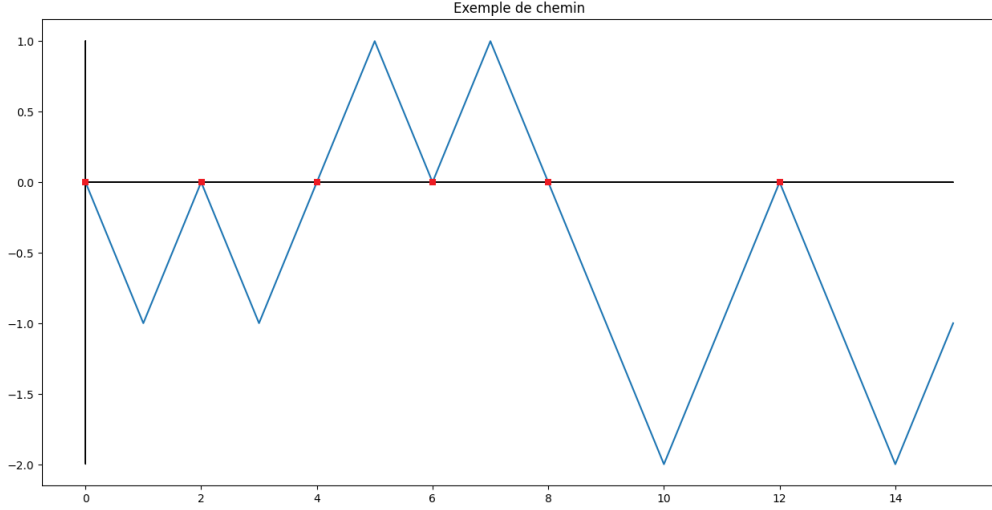


FIGURE 3 – Illustration du retour en zéro

qu'un retour en zéro arrive à l'insant $2n$. Plus formellement :

$$\mathbb{P}(s_1 \neq 0, \dots, s_{2n} \neq 0) = \mathbb{P}(s_{2n} = 0) = u_{2n}$$

Preuve. Main lemma

Par équiprobabilité des chemins,

$$\mathbb{P}(s_1 \neq 0, \dots, s_{2n} \neq 0) = 2\mathbb{P}(s_1 > 0, \dots, s_{2n} > 0)$$

et on veut donc montrer que $\mathbb{P}(s_1 > 0, \dots, s_{2n} > 0) = \frac{1}{2}u_{2n}$. On peut décomposer cet évènement selon les valeurs de s_{2n} :

$$\mathbb{P}(s_1 > 0, \dots, s_{2n} > 0) = \sum_{r=1}^n \mathbb{P}(s_1 > 0, \dots, s_{2n-1} > 0, s_{2n} = 2r).$$

D'après l'expression trouvée lors de la preuve du théorème du scrutin, le nombre de chemins correspondant aux évènements du terme de droite est

$$N_{2n-1,2r-1} - N_{2n-1,2r+1}.$$

Les probabilités correspondantes sont donc

$$N_{2n-1,2r-1}2^{-2n} - N_{2n-1,2r+1}2^{-2n} = \frac{1}{2} \left(N_{2n-1,2r-1}2^{-(2n-1)} - N_{2n-1,2r+1}2^{-(2n-1)} \right) = \frac{1}{2} (p_{2n-1,2r-1} - p_{2n-1,2r+1})$$

(d'après l'expression de $p_{n,r}$ donnée au début).

La somme est donc une somme télescopique, et on obtient

$$\mathbb{P}(s_1 > 0, \dots, s_{2n} > 0) = \frac{1}{2} (p_{2n-1,1} - p_{2n-1,2n+1}) = \frac{1}{2} p_{2n-1,1}.$$

Il suffit finalement de montrer que $p_{2n-1,1} = u_{2n}$.

Pour cela, on étudie les deux cas possibles pour un retour en zéro à l'instant $2n$: $u_{2n} = \frac{1}{2} p_{2n-1,1} + \frac{1}{2} p_{2n-1,-1}$. Par équiprobabilité, $p_{2n-1,1} = p_{2n-1,-1}$, d'où le résultat.

Parmi les retours en zéro, le premier retour en zéro va en particulier avoir une importance spécifique. C'est l'évènement $\{s_1 \neq 0, \dots, s_{2k-1} \neq 0, s_{2k} = 0\}$. On note $f_{2k} = \mathbb{P}(s_1 \neq 0, \dots, s_{2k-1} \neq 0, s_{2k} = 0)$.

Il est très intéressant de relier les u_{2k} et les f_{2k} , en distinguant un retour en zéro en $2n$ comme l'union des retours en zéro en $2n$ dont le premier retour en zéro a eu lieu en $2k$, pour $k \leq n$, suivi d'un retour en zéro sur le chemin allant de $2n - 2k$ à $2n$.

Plus concrètement, on a

$$\begin{aligned} \{s_{2n} = 0\} &= \bigsqcup_{k=0}^n \{s_1 \neq 0, \dots, s_{2k-1} \neq 0, s_{2k} = 0, s_{2n} = 0\} \\ &= \bigsqcup_{k=0}^n \{s_1 \neq 0, \dots, s_{2k-1} \neq 0, s_{2k} = 0\} \cap \{\epsilon_{2k+1} + \dots + \epsilon_{2n} = 0\}. \end{aligned}$$

On a alors, par indépendance et les ϵ_k ayant la même distribution :

$$u_{2n} = f_2 u_{2n-2} + f_4 u_{2n-4} + \dots + f_{2n} u_0.$$

C Loi de l'arcsinus pour le dernier retour en zéro

Un retour en zéro particulier va nous intéresser : le dernier en retour en zéro. On note

$$\alpha_{2k,2n} = \mathbb{P}(s_{2k} = 0, s_{2k+1} \neq 0, \dots, s_{2n} \neq 0)$$

la probabilité que jusqu'à l'instant $2n$ (inclus), le dernier retour en zéro ait eu lieu en $2k$ (avec $k \leq n$).

Ce dernier retour en zéro est en effet le point d'étude du théorème principal de cette partie :

Théorème II.4. Loi de l'arcsinus pour le dernier retour en zéro La probabilité que, sur

un chemin de longueur $2n$, le dernier retour en zéro ait lieu en $2k$ vaut :

$$\alpha_{2k,2n} = u_{2k}u_{2n-2k}$$

Nous prouverons ce résultat dans la prochaine partie, commençons par quelques remarques.

Le nom de loi de l'arcsinus vient du fait que la fonction de densité d'une variable aléatoire suivant la loi dite de l'arcsinus, caractérisée par sa fonction de répartition F telle que

$$\forall x \in [0, 1], F(x) = \frac{2}{\pi} \arcsin(\sqrt{x}),$$

approche très bien les $\alpha_{2k,2n}$ dès que n devient assez grand (ce qui, dans les faits, arrive relativement rapidement). Pour cela, on va juste se ramener à l'intervalle $[0, 1]$ en regardant $\frac{k}{n}$.

Regardons la pertinence de cette approximation.

	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
	$k = 10$	$k = 9$	$k = 8$	$k = 7$	$k = 6$	
$\alpha_{2k,20}$	0.1762	0.0927	0.0736	0.0655	0.0617	0.0606
$\frac{1}{\pi\sqrt{\frac{k}{n}(1-\frac{k}{n})}}$		0.1061	0.0795	0.0695	0.0649	0.0636

On voit que pour $n = 10$, qui reste une valeur relativement "petite", on a déjà une très bonne approximation dès que k n'est ni très proche de zéro, ni très proche de n . Cette précision s'améliore encore pour des n plus grands.

Étudions maintenant les propriétés de cette fonction de densité f , telle que $f(x) = \frac{1}{\pi\sqrt{x(1-x)}}$. En effet, elle a plusieurs propriétés intéressantes à remarquer.

Tout d'abord, notons qu'elle est symétrique par rapport à $\frac{1}{2}$. À la fois les probabilités et la fonction de densité sont symétriques par rapport au milieu de leur intervalle de définition :

$$\alpha_{2k,2n} = \alpha_{2n-2k,2n} \text{ et } f(x) = f(1-x).$$

Ainsi, en reprenant l'exemple des marches aléatoires, le dernier retour en zéro a autant de chances de se produire à un instant, à une proximité quelconque de la fin, qu'à l'instant de même proximité mais avec le début. Cela signifie qu'avec probabilité $\frac{1}{2}$, le dernier retour en zéro se produit dans la première moitié de l'expérience, peu importe la longueur de celle-ci.

Ajoutons à cela que la distribution prend ses valeurs les plus élevées aux extrémités, et ses

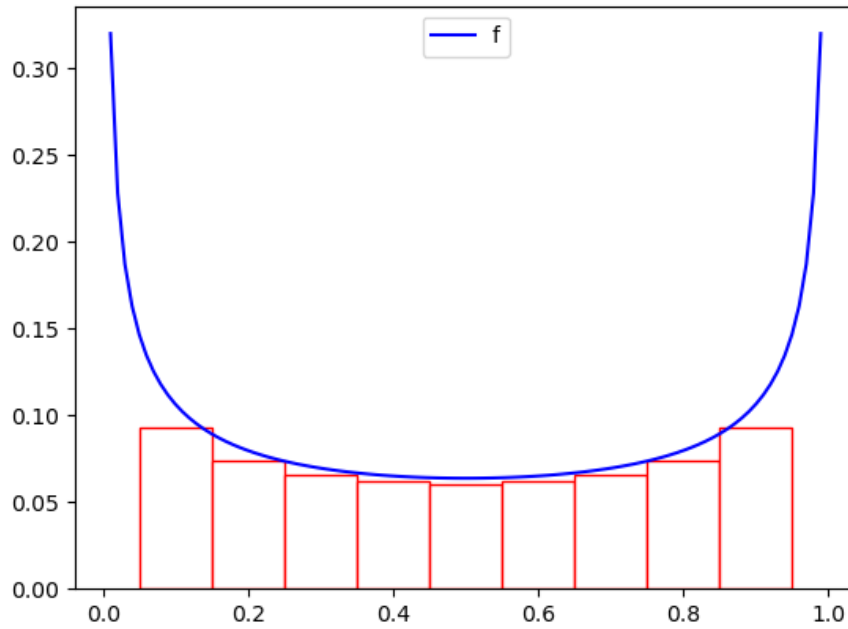


FIGURE 4 – Approximation de la fonction $f \mapsto \frac{1}{n} \frac{1}{\pi \sqrt{x(1-x)}}$ par les $\alpha_{2k,2n}$ pour $n = 10$

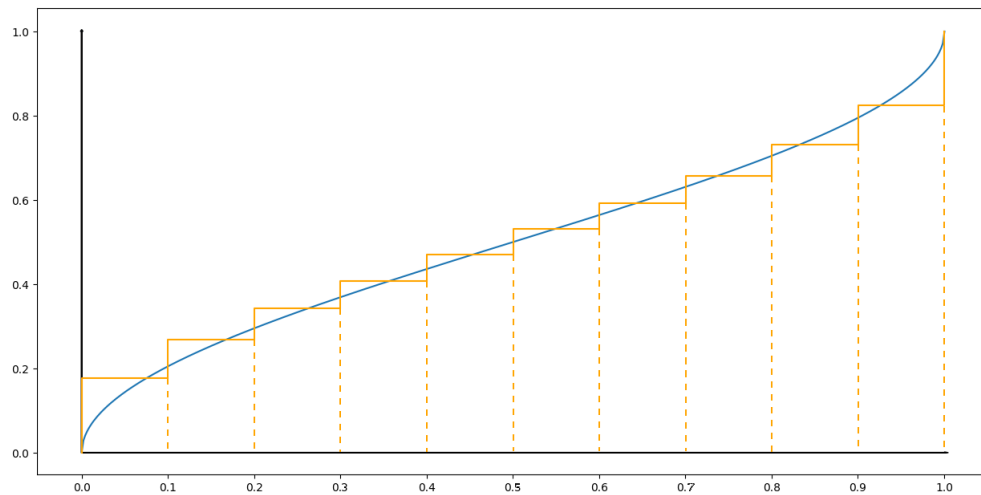


FIGURE 5 – Affichage des fonctions de répartition de la loi de l'arcsinus (en bleu) et du dernier retour en zéro pour $n = 10$ (en orange)

valeurs minimales au centre. Ces résultats sur la répétition d'une épreuve aléatoire ont parfois des conséquences contre-intuitives.

Par exemple, si on considère une situation où deux joueurs jouent à pile ou face et comptent le nombre de victoires d'écart, alors un des deux joueurs peut prendre la tête dès les premiers lancers et la garder jusqu'au bout, même si la partie est très longue.

On peut même regarder ce qu'il se passe concrètement en calculant numériquement les probabilités. Par exemple, considérons que le jeu de pile ou face est joué à une fréquence d'une manche par seconde, jour et nuit pendant un an (soit un total de plus de 31 millions de manches) et regardons le moment du dernier changement de leader (c'est-à-dire un retour en zéro de l'écart entre les deux scores).

Alors, la loi de l'arcsinus nous dit qu'avec une chance sur 10, ce changement de leader aura lieu avant le 9ème jour, soit avant que 2,5% du temps se soit écoulé. Dans une partie sur 100, c'est dans les 130 premières minutes qu'aura lieu ce dernier changement, soit dans les premiers 0,025% du temps.

Dans le même ordre d'idées, si on regarde deux élèves qui ont le même niveau et qui passent une succession d'évaluation sur une année, la loi de l'arcsinus nous dit qu'il n'est pas improbable qu'un des deux élèves ait toujours une meilleure moyenne générale que l'autre à partir des premières semaines.

Concrètement, si on fait un test d'"apprentissage" avec un groupe de 11 élèves n'impliquant que la chance et pas l'intelligence, et qu'on soumet ce groupe à ce test sur un an, alors un élève pourrait être premier pendant toute l'année sauf éventuellement la première semaine, et de même un élève pourrait être dernier tout du long sauf éventuellement au début.

D Preuve de la loi de l'arcsinus

Prouvons la loi de l'arcsinus.

Preuve. Loi de l'arcsinus pour le dernier retour en zéro

Pour n fixé et $k \leq n$, le dernier retour en zéro a lieu à l'instant $2k$ si $s_{2k} = 0, s_{2k+1} \neq 0, \dots, s_{2n} \neq 0$. Procédons par dénombrement.

On sait qu'il y a 2^{2n} chemins de longueur $2n$. On veut donc compter ceux qui vérifient les conditions ci-dessus. Pour les $2k$ premiers sommets, on doit ranger k valeurs $+1$ parmi $2k$ emplacements, soit $\binom{2k}{k}$ choix possibles.

Notons que comme $u_{2k} = \binom{2k}{k} 2^{-2k}$, cela nous donne $2^{2k} u_{2k}$ choix pour les $2k$ premiers sommets.

Pour les $2n - 2k$ suivants, en utilisant le Main lemma, on voit que vouloir des sommets tous strictements positifs, c'est aussi probable que de vouloir un retour en zéro en $2n$. En considérant le point $(2k, 0)$ comme nouvelle origine, cela nous indique donc par analogie avec la première étape qu'on a $2^{2n-2k}u_{2n-2k}$ choix possibles pour les $2n - 2k$ derniers sommets.

Ces deux phases de choix étant indépendantes, on a au final $2^{2n}u_{2k}u_{2n-2k}$ choix possibles. En divisant par le nombre de chemins possibles de longueur $2n$, c'est-à-dire par 2^{2n} , on obtient le résultat.