4
Simulation du hasard

I
Loi d'une variable aléatoire

1. Une **variable aléatoire réelle** est une fonction d'un ensemble Ω (qu'on ne cherche pas à connaître) dans l'ensemble $\mathbb{R}$ des nombres réels.

1.1 Pour étudier une fonction f de $\mathbb{R}$ dans $\mathbb{R}$, on doit connaître
– son ensemble de définition $I \subset \mathbb{R}$
– et son expression, c'est-à-dire la valeur de $f(x)$ pour chaque $x \in I$.

L'étude de la fonction f a pour objectif de faire apparaître des propriétés remarquables (monotonie, extrema locaux, convexité, limites, asymptotes...) qu'on peut synthétiser en traçant l'allure du graphe de f.

1.2 Pour une variable aléatoire $X : \Omega \to \mathbb{R}$, on ne cherche pas à connaître la valeur de $X(\omega)$. On se contente de considérer l'ensemble des valeurs prises par la fonction X et de savoir avec quelles fréquences sont prises ces valeurs. La famille de ces fréquences s'appelle la **loi (de distribution)** de la variable aléatoire.

I.1 Lois discrètes

2. Les **variables aléatoires discrètes** usuelles prennent leurs valeurs dans l'ensemble $\mathbb{N}$ des entiers. La plupart servent à compter un nombre d'événements aléatoires qui surviennent au cours d'une expérience.

3. La **loi** d'une variable aléatoire discrète X à valeurs dans $\mathbb{N}$ est la suite de terme général $\mathsf{P}(X = n)$.
L'étude de suite (ou sa représentation graphique par un diagramme en bâtons) permet de savoir quelles sont les valeurs les plus probables et les valeurs les moins probables pour X.

4. On considère une expérience aléatoire qui peut produire N résultats possibles :

$$x_1,\, x_2,\, \ldots,\, x_N.$$

4.1 Si on n'a aucune raison de penser qu'un résultat est plus probable qu'un autre, on modélise cette expérience par une **variable uniforme**.

$$\forall\, 1 \leqslant k \leqslant N, \quad \mathsf{P}(X = k) = \frac{1}{N}.$$

4.2 Dans le cas général, la loi de X est décrite par N réels positifs $p_1, \ldots, p_N$ dont la somme est égale à 1 et par les relations suivantes.

$$\forall\, 1 \leqslant k \leqslant N, \quad \mathsf{P}(X = k) = p_k.$$

5. Schéma de Bernoulli
On considère une expérience aléatoire répétée indéfiniment dans les mêmes conditions et qui, comme une variable booléenne, ne peut produire que deux résultats : un succès ou un échec.

5.1 Le résultat de chaque expérience est modélisé par une **variable de Bernoulli** : une telle variable ne peut prendre que les valeurs 1 (succès) et 0 (échec). Le paramètre

$$p = \mathsf{P}(X = 1)$$

est alors interprété comme la probabilité d'un succès et la probabilité

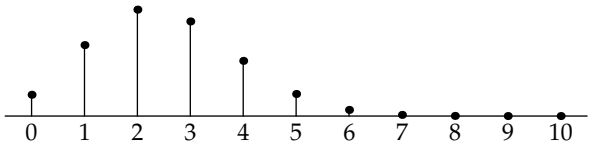
$$q = \mathsf{P}(X = 0)$$

est la probabilité d'un échec.

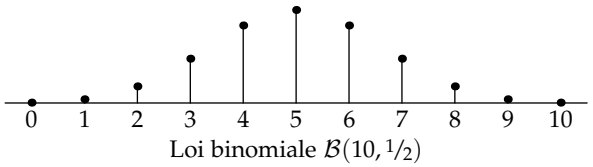
5.2 Le nombre N_n de succès obtenus en n tentatives est modélisé par une **variable binomiale** $\mathcal{B}(n, p)$.

$$\forall\, 0 \leqslant k \leqslant n, \quad \mathsf{P}(N_n = k) = \binom{n}{k} p^k q^{n-k}.$$

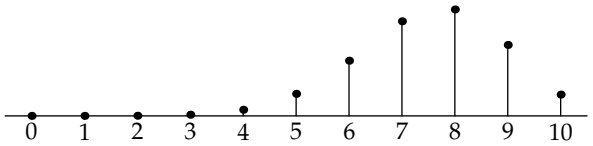
Comme on s'en doute, plus la probabilité de succès p est élevée, plus le nombre de succès en n tentatives est proche de n.



Loi binomiale $\mathcal{B}(10, 1/4)$



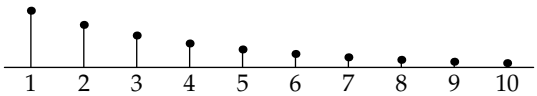
Loi binomiale $\mathcal{B}(10, 1/2)$



Loi binomiale $\mathcal{B}(10, 3/4)$

5.3 Le nombre d'expériences G à réaliser pour obtenir un premier succès est modélisé par une **variable géométrique**.

$$\forall\, k \geqslant 1, \quad \mathsf{P}(G = k) = q^{k-1} p.$$



Loi géométrique $\mathcal{G}(1/4)$

Loi géométrique $\mathcal{G}(1/2)$

Loi géométrique $\mathcal{G}(3/4)$

Ici encore, on se doute que plus la probabilité de succès p est élevée, plus le premier succès arrive vite. Plus précisément,

$$\forall\, n \geqslant \frac{-2}{\log q}, \qquad \mathsf{P}(G \leqslant n) \geqslant 99\%.$$

I.2 Lois à densité

6. Une **densité** est une fonction f positive et continue par morceaux sur $\mathbb{R}$, telle que

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1.$$

6.1 La loi d'une variable aléatoire réelle X est décrite par la densité f lorsque

$$\forall\, a \leqslant b, \quad \mathsf{P}(a \leqslant X \leqslant b) = \int_a^b f(x)\,\mathrm{d}x.$$

Dans ce cas, la probabilité $\mathsf{P}(X = a)$ est nulle pour tout $a \in \mathbb{R}$. On écrit parfois

$$\mathsf{P}\big(X \in [x, x+dx]\big) \approx f(x)\,\mathrm{d}x \quad \text{ou} \quad f(x) = \frac{\mathrm{d}\,\mathsf{P}}{\mathrm{d}x},$$

ce qui ne veut pas dire grand'chose mais qui a le mérite de rappeler que les valeurs de la densité f sont positives, mais ne sont pas des probabilités. Il est normal que la densité f prenne des valeurs supérieures à 1.

6.2 Le graphe de la densité f permet alors de savoir quelles sont les valeurs les plus probables (grandes valeurs de f) et les valeurs les moins probables (faibles valeurs de f) pour la variable aléatoire X.
Si la densité f est nulle sur un intervalle I, alors on peut considérer que les éléments de I sont des valeurs impossibles pour la variable X.

$$\mathsf{P}(X \in I) = \int_I f(x)\,\mathrm{d}x = 0$$

Densités usuelles

7. La **densité uniforme** sur $[A, B]$ est la fonction nulle sur $]-\infty, A[$ et sur $]B, +\infty[$ et telle que

$$\forall\, x \in [A, B], \quad f(x) = \frac{1}{B - A}.$$

7.1 Si la variable X suit la loi uniforme sur $[A, B]$, elle prend toutes ses valeurs entre A et B et la probabilité pour que X prenne une valeur dans un sous-intervalle $I \subset [A, B]$ est proportionnelle à la longueur de l'intervalle I.

7.2 Si une variable aléatoire X suit la loi uniforme sur $[0, 1]$, alors

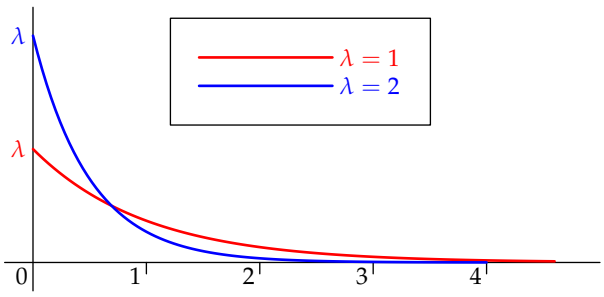
$$\forall\, 0 \leqslant \alpha < \beta \leqslant 1, \quad \mathsf{P}(\alpha \leqslant U \leqslant \beta) = (\beta - \alpha)$$

et la variable $Y = A + (B - A)X$ suit la loi uniforme sur $[A, B]$.

8. La **densité exponentielle** de paramètre $\lambda > 0$ est la fonction nulle sur $]-\infty, 0]$ telle que

$$\forall\, x \geqslant 0, \quad f(x) = \lambda e^{-\lambda x}.$$

8.1 Une variable aléatoire exponentielle ne prend donc que des valeurs positives. Elle sert en général à modéliser une durée aléatoire où $1/\lambda$ représente le temps d'attente moyen.



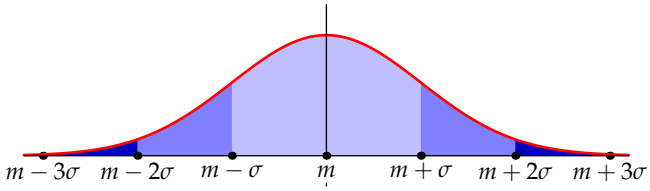
8.2 On voit que plus λ est grand, plus la durée aléatoire modélisée par G prend fréquemment des valeurs proches de 0. Plus précisément,

$$\forall\, x \geqslant \frac{2\,\ell\mathrm{n}\,10}{\lambda}, \qquad \mathsf{P}(G \leqslant x) \geqslant 99\%.$$

9. La **densité normale** de paramètres $m \in \mathbb{R}$ et $\sigma > 0$ est la fonction définie par

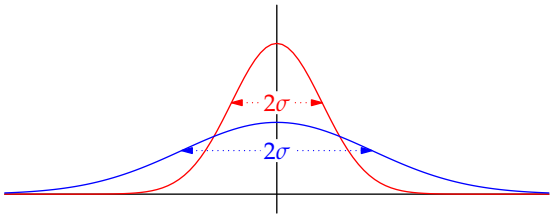
$$\forall\, x \in \mathbb{R}, \quad f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(\frac{-(x-m)^2}{2\sigma^2}\Big).$$

9.1 La loi normale sert en général à modéliser les fluctuations aléatoires autour de la valeur moyenne m.



Valeurs de référence
$\mathsf{P}\big(\lvert X - m\rvert \leqslant \sigma\big) \approx 68\,\%$
$\mathsf{P}\big(\lvert X - m\rvert \leqslant 2\sigma\big) \approx 95\,\%$
$\mathsf{P}\big(\lvert X - m\rvert \leqslant 3\sigma\big) \approx 99{,}7\,\%$

9.2 La dispersion des valeurs autour de la valeur moyenne est décrite par l'écart type σ. Plus l'écart type est petit, plus les valeurs sont concentrées au voisinage de la moyenne.



9.3 La loi normale est dite **centrée** pour $m = 0$ et **réduite** pour $\sigma = 1$.
Si une variable aléatoire N suit la loi normale centrée et réduite, alors la variable aléatoire $X = m + \sigma N$ suit la loi normale de paramètres m et σ.

Rôle central de la loi uniforme sur $[0, 1]$

10. Soit U, une variable aléatoire de loi uniforme sur $[0, 1]$.

10.1 On peut démontrer que, pour toute variable aléatoire réelle X, il existe une une fonction $f_X : [0, 1] \to \mathbb{R}$ telle que la loi de $f_X(U)$ soit égale à la loi de X.
Comme la fonction f_X est connue (en théorie), on peut simuler n'importe quelle variable aléatoire réelle dès lors qu'on sait simuler une variable aléatoire de loi uniforme sur $[0, 1]$.

10.2 **Une pièce truquée**
Soient $0 < p < 1$ et f_X, la fonction définie par

$$\forall\, u \in [0, 1], \quad f_X(u) = \mathbb{1}_{[u<p]}.$$

Alors $f_X(U)$ suit la loi de Bernoulli de paramètre p.

10.3 Un dé truqué
Soit $p = (p_k)_{1\leqslant k\leqslant 6}$, une loi de probabilité sur $\{1,\dots,6\}$.
On pose alors $R_0 = 0$ et $R_k = p_1 + \cdots + p_k$ pour $1 \leqslant k \leqslant 6$. En particulier, $R_6 = 1$.
Si on pose

$$\forall\, u \in [0,1], \quad f_X(u) = \sum_{k=0}^{5} \mathbb{1}_{[u>R_k]}$$

alors la loi de la variable aléatoire $f_X(U)$ est la loi discrète définie par

$$\forall\, 1 \leqslant k \leqslant 6, \quad \mathsf{P}\big(f_X(U) = k\big) = p_k.$$

II

Exemples de processus aléatoires

11. Sur les exemples les plus courants, on vérifie que tout processus aléatoire à temps discret $(X_n)_{n\in\mathbb{N}}$ peut être explicitement déduit d'une suite $(U_n)_{n\in\mathbb{N}}$ de variables aléatoires indépendantes de loi uniforme sur $[0,1]$.

II.1 Échantillons i.i.d.

12. Lorsqu'on effectue des tirages successifs *avec* remise dans une urne, la composition de l'urne est la même pour chaque tirage. L'expérience aléatoire qui consiste à tirer une boule dans cette urne est donc réalisée à chaque fois dans les mêmes conditions.

13. Plus généralement, toute expérience aléatoire répétée indéfiniment dans les mêmes conditions est modélisée par un **processus i.i.d.**, c'est-à-dire une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires indépendantes et de même loi.
13.1 L'indépendance signifie que la connaissance des premiers résultats ne permet pas de prévoir les résultats à venir.
13.2 Le fait que les variables X_n aient toute la même loi traduit l'hypothèse que l'expérience soit répétée dans les mêmes conditions : chaque résultat a toujours la même probabilité d'apparaître.

14. Jeu de pile ou face
On modélise un jeu de pile ou face par une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires de Bernoulli indépendantes et de même paramètre p.
Si l'événement $[X_n = 1]$ représente le fait d'obtenir *Pile* au n-ième lancer, alors

$$\forall\, n \in \mathbb{N}, \qquad p = \mathsf{P}(X_n = 1)$$

et le paramètre p représente la probabilité d'obtenir *Pile* lors d'un lancer de la pièce.
Si la pièce est normale et est lancée de manière normale, la symétrie des données conduit à faire l'hypothèse naturelle que *Pile* et *Face* ont la même probabilité d'apparaître et donc à $p = 1/2$.
Une valeur $p \neq 1/2$ s'interprète comme l'utilisation d'une pièce truquée : l'un des résultats est plus probable que l'autre, il n'y a plus de symétrie.

15. Jeu de dé
On modélise un jeu de dé par une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires indépendantes et de même loi.
Un dé normal possède six faces et aucune face n'a de raison d'apparaître plus souvent que les autres. Dans ce cas,

$$\forall\, n \in \mathbb{N},\ \forall\, k \in \{1,\dots,6\}, \quad \mathsf{P}(X_n = k) = \frac{1}{6}.$$

Truquer un dé consiste à rompre la symétrie qui existe naturellement entre les faces de ce dé. On modélise donc un jeu truqué par une loi autre que la loi uniforme sur l'ensemble $\{1,\dots,6\}$.

16. Réalisation d'un processus i.i.d.
Soient $(X_n)_{n\in\mathbb{N}}$, un processus i.i.d. quelconque et $(U_n)_{n\in\mathbb{N}}$, une suite de variables aléatoires indépendantes de loi uniforme sur l'intervalle $[0,1]$.
16.1 On sait **[10]** qu'il existe une fonction $f : [0,1] \to \mathbb{R}$ telle que les variables X_0 et $f(U_0)$ aient même loi.
16.2 Le processus $(f(U_n))_{n\in\mathbb{N}}$ est alors un processus i.i.d. où $f(U_n)$ a même loi que X_n pour tout $n \in \mathbb{N}$.
16.3 Si on sait simuler convenablement le processus $(U_n)_{n\in\mathbb{N}}$ et calculer facilement la fonction f, on peut simuler convenablement le processus $(X_n)_{n\in\mathbb{N}}$.

II.2 Chaînes de Markov

17. Processus de Lévy
Soient X, une variable aléatoire et $(X_n)_{n\in\mathbb{N}}$, un processus i.i.d. tel que X_n et X aient même loi pour tout $n \geqslant 1$, la loi de X_0 étant quelconque.
17.1 Le **processus de Lévy** de pas X est le processus aléatoire $(S_n)_{n\in\mathbb{N}}$ défini par la donnée initiale

$$S_0 = X_0$$

et la relation de récurrence

$$\forall\, n \in \mathbb{N}, \quad S_{n+1} = S_n + X_{n+1}.$$

17.2 On peut construire n'importe quel processus de Lévy à partir d'une suite $(U_n)_{n\in\mathbb{N}}$ de variables aléatoires indépendantes de loi uniforme sur $[0,1]$.
On sait **[10]** qu'il existe deux fonctions

$$f : [0,1] \to \mathbb{R} \quad \text{et} \quad g : [0,1] \to \mathbb{R}$$

telles que, d'une part, les variables X_0 et $f(U_0)$ aient même loi et que, d'autre part, les variables X_1 et $g(U_1)$ aient même loi.
Dans ces conditions, le processus de Lévy $(S_n)_{n\in\mathbb{N}}$ a même loi que le processus $(Y_n)_{n\in\mathbb{N}}$ défini par

$$Y_0 = f(U_0) \quad \text{et} \quad \forall\, n \in \mathbb{N}, \quad Y_{n+1} = Y_n + g(U_n).$$

18. Marches aléatoires
La notion de *marche aléatoire* est une généralisation des processus de Lévy qui permet par exemple de modéliser les tirages *avec* remise.
18.1 On se restreint ici aux marches aléatoires à valeurs dans un espace d'états $E \subset \mathbb{N}$ qu'on suppose *fini* : chaque variable aléatoire X_n est une variable aléatoire discrète.
18.2 Une **marche aléatoire discrète** est un processus $(X_n)_{n\in\mathbb{N}}$ qui vérifie la **propriété de Markov** : pour tout indice $n \in \mathbb{N}$, quels que soient les valeurs x_0, x_1, …, x_n dans E,

$$\begin{aligned}\mathsf{P}(X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = x_n)\\ = \mathsf{P}(X_0 = x_0) \times \prod_{k=0}^{n-1} \mathsf{P}(X_{k+1} = x_{k+1} \mid X_k = x_k).\end{aligned}$$

18.3 Supposons connue une suite $(U_n)_{n\in\mathbb{N}}$ de variables aléatoires indépendantes de loi uniforme sur $[0,1]$.
Il existe une fonction $f : [0,1] \to E$ telle que les variables aléatoires X_0 et $f(U_0)$ aient même loi.
Pour chaque $x \in E$ et chaque $k \in \mathbb{N}$, il existe une fonction

$$g_k(x, \cdot\,) : [0,1] \to E$$

telle que la loi de la variable aléatoire $g_k(x, U_{k+1})$ soit la loi conditionnelle de X_{k+1} sachant l'événement $[X_k = x]$:

$$\forall\, y \in E, \quad \mathsf{P}\big(g_k(x, U_{k+1}) = y\big) = \mathsf{P}(X_{k+1} = y \mid X_k = x).$$

La loi de la marche aléatoire $(X_n)_{n\in\mathbb{N}}$ est alors la loi du processus $(Y_n)_{n\in\mathbb{N}}$ défini par

$$Y_0 = f(U_0) \quad \text{et} \quad \forall\, n \in \mathbb{N}, \quad Y_{n+1} = g_n(Y_n, U_{n+1}).$$

18.4 La chaîne de Markov est dite **homogène** lorsque les fonctions g_n sont toutes identiques.

III

Étude statistique d'une variable aléatoire

19. Point de vue statistique
Modéliser le résultat d'une expérience par une variable aléatoire X signifie que :

1. le résultat de l'expérience n'est pas toujours le même et n'est pas prévisible ;
2. mais que la *fréquence d'apparition* de chacun des résultats possibles est assez bien déterminée.

Ce sont les fréquences respectives de chacun de ces résultats qui constituent la **loi** de la variable aléatoire X.

20. Loi des grands nombres
On appelle **loi des grands nombres** tout théorème qui affirme qu'on peut retrouver une approximation de la loi d'une variable aléatoire X à partir d'une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires indépendantes de même loi que X.

21. Si X est une variable aléatoire discrète à valeurs dans $\mathbb{N}$, sa **loi** est la famille des probabilités $p_k = \mathsf{P}(X = k)$ lorsque k parcourt $\mathbb{N}$.

21.1 Si $(X_n)_{n\in\mathbb{N}}$ est une suite de variables aléatoires indépendantes de même loi que X, alors pour tout $k \in \mathbb{N}$, les variables aléatoires définies par

$$\forall\, i \in \mathbb{N}, \quad B_{i,k} = \mathbb{1}_{[X_i=k]}$$

sont des variables aléatoires de Bernoulli indépendantes, de paramètre p_k.

21.2 La variable aléatoire définie par

$$M_{n,k} = \frac{1}{n}\sum_{i=0}^{n-1} B_{i,k}$$

doit être interprétée comme la **fréquence** d'apparition de la valeur k lors des n premières réalisations de l'expérience modélisée par X.

21.3 L'inégalité de Bienaymé-Tchebychev montre alors que, pour tout réel $\alpha > 0$ et tout entier $n \geqslant 1$,

$$0 \leqslant \mathsf{P}\big(|M_{n,k} - p_k| \geqslant \alpha\big) \leqslant \frac{1}{4n\alpha^2}.$$

En particulier, pour $\alpha = 10^{-1}$ et $n = 10^3$, la probabilité de l'événement

$$[M_n - \alpha \leqslant p_k \leqslant M_n + \alpha]$$

est supérieure à 99%.
On peut ainsi déduire des résultats obtenus lors de 1 000 réalisations de l'expérience des valeurs approchées de chacune des probabilités p_k avec une erreur absolue de l'ordre de $0,1$, ce qui reste assez grossier, mais une incertitude inférieure à 1%, ce qui est excellent.

22. Variables à densité
Soient X, une variable aléatoire dont la loi est représentée par la densité f et $(X_n)_{n\in\mathbb{N}}$, une suite de variables aléatoires indépendantes de même loi que X.

22.1 Quels que soient les réels $a < b$, on peut obtenir une valeur approchée de la probabilité

$$p = \mathsf{P}(a \leqslant X \leqslant b)$$

en considérant la fréquence

$$M_n = \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{[a\leqslant X_k\leqslant b]}$$

avec laquelle le résultat produit lors des n premières réalisations se situe dans l'intervalle $[a, b]$.

22.2 Comme plus haut,

$$\mathsf{P}\big(p \in [M_n - \alpha,\, M_n + \alpha]\big) \geqslant 1 - \frac{1}{4n\alpha^2}.$$

IV

Générateur pseudo-aléatoire

23. Cahier des charges
Un générateur de nombres au hasard doit pouvoir produire de longues suites $(x_k)_{0\leqslant k<n}$ de réels compris entre 0 et 1 qui reproduisent le comportement statistique d'une suite $(U_k)_{k\in\mathbb{N}}$ de variables aléatoires indépendantes de loi uniforme sur $[0, 1]$.

23.1 D'après la loi des grands nombres, quels que soient les bornes α et β telles que $0 \leqslant \alpha < \beta \leqslant 1$, la fréquence

$$\frac{\#\big\{0 \leqslant k < n \,:\, \alpha \leqslant x_k \leqslant \beta\big\}}{n}$$

doit tendre vers $(\beta - \alpha)$ lorsque n tend vers l'infini.

23.2 Pour tout entier $d \geqslant 1$ et tout entier $k \in \mathbb{N}$, le vecteur aléatoires

$$X_k^d = (U_{kd}, U_{kd+1}, \ldots, U_{(k+1)d-1})$$

suit la loi uniforme sur $[0, 1]^d$ au sens où

$$\mathsf{P}\Big(\bigcap_{0\leqslant i<d} [\alpha_i \leqslant U_{kd+i} \leqslant \beta_i]\Big) = \prod_{0\leqslant i<d} (\beta_i - \alpha_i)$$

quels que soient $0 \leqslant \alpha_1 < \beta_1 \leqslant 1$, …, $0 \leqslant \alpha_{d-1} < \beta_{d-1} \leqslant 1$.
Comme les vecteurs aléatoires $(X_k^d)_{k\in\mathbb{N}}$ sont indépendants, la loi des grands nombres s'applique et on dit que le générateur est **uniformément distribué sur** d **dimensions** lorsque la fréquence

$$\frac{1}{n}\cdot\#\Big\{0 \leqslant k < n \,:\, X_k^d \in \prod_{0\leqslant i<d} [\alpha_i, \beta_i]\Big\}$$

converge vers

$$\prod_{0\leqslant k<d} (\beta_k - \alpha_k)$$

quels que soient $0 \leqslant \alpha_1 < \beta_1 \leqslant 1$, …, $0 \leqslant \alpha_{d-1} < \beta_{d-1} \leqslant 1$.

24. L'important est bien dans le comportement statistique des nombres et non pas dans le caractère imprévisible des résultats. Les générateurs usuels sont d'ailleurs complètement prévisibles puisque, d'un point de vue mathématique, il s'agit de suites récurrentes périodiques !
On doit donc parler de **générateurs pseudo-aléatoires**.

25. Le générateur utilisé par Python est un *Mersenne Twister* (Matsumoto et Nishimura, 1997) qui produit des flottants de 32 bits avec les propriétés statistiques suivantes.

25.1 La période de ce générateur est le nombre premier de Mersenne

$$2^{19\,937} - 1 \approx 2\,080! \approx 10^{6\,000}.$$

Pour conserver de bonnes propriétés statistiques, le nombre de flottants utilisés dans une simulation doit rester très inférieur à la période, ce qui n'est pas une limitation sérieuse pour les applications courantes.

25.2 Il est uniformément distribué sur d dimensions pour tout $d \leqslant 623$, ce qui permet de simuler des vecteurs aléatoires de dimensions assez élevées, mais ne suffit pas en cryptographie par exemple.