

Modélisation stochastique (HAX603X)

HAX603X: Modélisation stochastique

Aspects numériques de la modélisation aléatoire et statistiques (cours de Licence 3). Les sections en haut de page regroupe le contenu pédagogique : les [slides](#) seront présentés en cours, et pour aller un peu plus loin le “poly” est disponible en format “html” sur la page [Cours](#).

Enseignants

- [Joseph Salmon](#): joseph.salmon@umontpellier.fr,
- [Benjamin Charlier](#): benjamin.charlier@umontpellier.fr

Ce cours est issu du travail antérieur de la part de:

- [Nicolas Meyer](#)
- [Benoîte de Saporta](#)

et de l'aide de [François-David Collin](#).

Prérequis

- Bases de probabilités (en particulier “HAX506X- Théorie des Probabilités”): probabilité, densité, espérance, fonction de répartition, mesure, intégration, analyse numérique élémentaire, etc. (Foata and Fuchs 1996; Barbe and Ledoux 2006; Ouvrard 2007, 2008)
- Programmation élémentaire (en Python): `if ... then... else ...`, `for`, `while`, fonctions, etc. [HLMA310 - Logiciels scientifiques](#), (Courant et al. 2013), [Cours de Python: Univ. Paris Diderot](#)

Description du cours

1. Générer l'aléa

- générateurs pseudo-aléatoires
- illustrations numériques et visualisation en Python (loi des grands nombres, théorème central limite)
- simulations de variables aléatoires (méthode de l'inverse, méthode du rejet, cas spécifiques, etc.)

2. Méthode de Monte-Carlo

- méthode de Monte-Carlo pour le calcul approché d'une intégrale
- réduction de la variance : variables antithétiques, variables de contrôle, échantillonnage préférentiel.

3. Compléments

- vecteurs gaussiens et lien avec les lois usuelles de la statistique inférentielle (student, chi2)
- construction d'intervalles de confiance.
- marche aléatoire simple, etc.

Modalité de contrôle des connaissances



- TP notés : Rendu en Quarto .qmd et .pdf.
 - TP noté 1 : rendre en fin de session
 - TP noté 2 : rendre en fin de session
- CC : devoir sur table d'une heure
- Coefficients:
 - Note Session 1 : (40% CC + 30% TP 1 + 30% TP 2)
 - Note Session 2 : (30% CC + 35% TP 1 + 35% TP 2)

Moodle







[Moodle: HAX603X Modélisation Stochastique](#)

Ressources supplémentaires

Niveau débutant

- Introduction à Python [Cours de Python](#) 
- [HLMA310 - Logiciels scientifiques](#) 
- Manuel d'algorithmique en Python (Courant et al. 2013) 
- Data Science: [Python Data Science Handbook, With Application to Understanding Data](#) by J. Van DerPlas, 2016;  **videos**: [Reproducible Data Analysis in Jupyter](#)
- [Math for journalists](#) by Naël Shiab 

Niveau avancé

- [Software Dev. for datascience](#) by J. Salmon and B. Charlier, 
- [Monte Carlo Methods and Applications](#) by Keenan Crane 
- Chaîne de Markov: [Markov Chains](#) by Ethan N. Epperly 
- [Advanced Data Analysis from an Elementary Point of View](#) by Cosma Shalizi; 
- [Maximum likelihood by numerical optimization](#) 
- Conditionnement, martingales et autres preuves de la loi des grands nombres: (Williams 1991) 

1 Perspectives historiques

Nous allons présenter ici quelques éléments historiques sur les méthodes de Monte-Carlo, dont les prémisses remontent au XVIIIème siècle.

1.1 L'aiguille de Buffon

[Georges-Louis Leclerc, Comte de Buffon](#)¹ proposa en 1733 une méthode qui s'avéra être utile pour estimer la valeur de π . On désigne de nos jours cette expérience sous le nom de **l'aiguille de Buffon**. C'est l'une des premières méthodes de Monte-Carlo référencée dans la littérature (la source du texte est disponible [ici](#) sur le site de la BNF).

La question initiale (simplifiée ici) posée par Buffon était la suivante: une aiguille de taille 1 tombe sur un parquet composé de lattes de largeur 1: quelle est alors la probabilité P que l'aiguille croise une ligne de la trame du parquet ?

Le contexte original était dans celui d'un jeu à deux joueurs: un joueur parie sur le fait que l'aiguille croise une ligne de la trame du parquet, l'autre sur le fait que l'aiguille ne croise pas une ligne de la trame du parquet. L'enjeu est alors de calculer la probabilité de succès de chacun des joueurs, et de voir si le jeu est équilibré ou non.

Voilà brièvement la question que s'est posée Buffon en 1733. La réponse est donnée par la formule suivante, qui montre que le jeu qu'il propose n'est pas équilibré:

$$P = \frac{2}{\pi} \approx 0.6366 .$$

Une preuve de ce résultat sera donnée ci-dessous.

L'idée sous-jacente de Buffon est que si l'on répète cette expérience un grand nombre de fois, on peut approcher la quantité P numériquement, par exemple en proposant un estimateur \hat{P}_n qui compte la proportion de chevauchement après avoir fait n répétition des lancers. Pour estimer π , il ne restera donc plus qu'à évaluer $\frac{2}{\hat{P}_n}$.

¹[Georges-Louis Leclerc, Comte de Buffon](#): (1707-1788) naturaliste, mathématicien et industriel français du siècle des Lumières

On peut faire cette expérience dans le monde réelle (c'est un peu long pour n grand!), mais on peut aussi utiliser une méthode numérique pour cela. Il s'agit alors de tirer aléatoire la position du centre de l'aiguille, puis de tirer aussi de manière aléatoire son angle de chute. On teste à la fin si l'aiguille croise une ligne de la trame du parquet ou non, et on recommence l'expérience un grand nombre de fois.

Cette méthode est donnée ci-dessous, avec un exemple interactif généré en Python.

On va fournir ici le calcul de la probabilité P . Pour cela on aura besoin de quelques éléments décrits dans le dessin ci-dessous.

- x : distance entre le centre de l'aiguille et la ligne de la trame du parquet la plus proche
- θ : angle entre l'aiguille et la ligne de la trame du parquet la plus proche
- l : longueur de l'aiguille (et donc la demi longueur est $\frac{l}{2}$)
- $\frac{l}{2} \sin(\theta)$: distance entre l'extrémité de l'aiguille et la ligne de la trame du parquet la plus proche

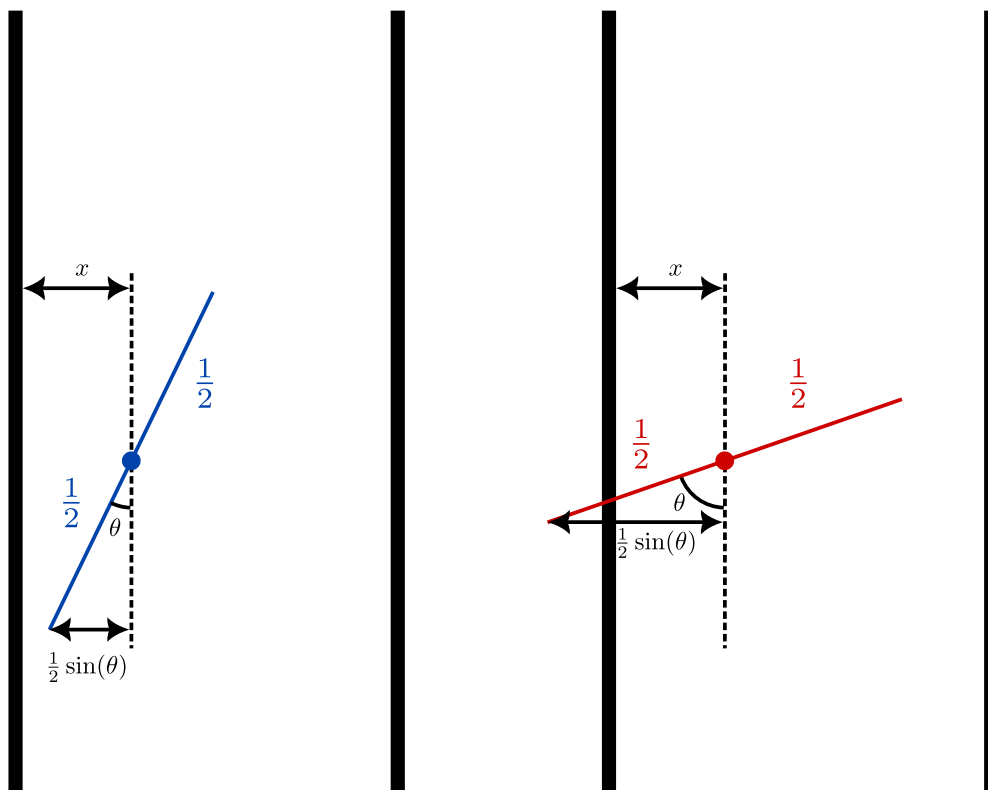
Avec les éléments ci-dessus, on voit qu'il y a chevauchement si et seulement si: $\frac{l}{2} \sin(\theta) \geq x$.

Maintenant par des arguments de symétrie on voit qu'on peut se restreindre à $\theta \in [0, \frac{\pi}{2}]$, et à $x \in [0, \frac{l}{2}]$. Les lois de générations des variables aléatoires X et Θ sont les suivantes:

- $X \sim \mathcal{U}([0, \frac{l}{2}])$, de densité $f_X(x) = 2\mathbb{1}_{[0, \frac{l}{2}]}(x)$
- $\Theta \sim \mathcal{U}([0, \frac{\pi}{2}])$ de densité $f_\Theta(\theta) = \frac{2}{\pi}\mathbb{1}_{[0, \frac{\pi}{2}]}(\theta)$

De plus on suppose que X et Θ sont indépendantes.

Maintenant pour calculer la probabilité P on procède comme suit:



(a) Sans croisement

(b) Avec croisement

Figure 1.1: Configuration sans croisement (à gauche) et avec croisement (à droite) de l'aiguille avec une ligne de la trame du parquet.

$$\begin{aligned}
P &= \mathbb{P} \left(\frac{1}{2} \sin(\Theta) \geq X \right) \\
&= \int_{\mathbb{R}^2} \mathbb{1}_{\{\frac{1}{2} \sin(\theta) \geq x\}} f_{\Theta}(\theta) f_X(x) d\theta dx \quad (\text{par indépendance}) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{\{\frac{1}{2} \sin(\theta) \geq x\}} \frac{2}{\pi} \mathbb{1}_{[0, \frac{\pi}{2}]}(\theta) \cdot 2 \mathbb{1}_{[0, \frac{1}{2}]}(x) d\theta dx \\
&= \int_0^{\frac{\pi}{2}} \int_0^{\frac{1}{2} \sin(\theta)} \frac{4}{\pi} dx d\theta \\
&= \frac{4}{\pi} \int_0^{\frac{\pi}{2}} \frac{1}{2} \sin(\theta) d\theta \\
&= \frac{2}{\pi} \left[-\cos(\theta) \right]_0^{\frac{\pi}{2}} \\
&= \frac{2}{\pi} .
\end{aligned}$$

Une démonstration alternative est donnée dans le livre de Borel and Deltheil (1923), inspirée de celle Barbier (1860) Elle utilise l'approche suivante. Au lieu de calculer la probabilité d'intersection entre une aiguille et les rainures d'un parquet, nous allons déterminer le nombre moyen d'intersections entre une courbe et les rainures du parquet. Nous démontrerons que ce nombre moyen est directement proportionnel à la longueur de la courbe et inversement proportionnel à la distance moyenne entre les rainures du parquet (nous prendrons cette distance égale à 1 pour simplifier).

Pour ce faire, nous allons d'abord considérer des lignes polygonales, pour lesquelles l'espérance d'intersection se calcule en sommant les espérances sur chaque segment. Nous étendrons ensuite ce résultat par continuité aux courbes régulières. La constante de proportionnalité de cette relation sera notée α .

Nous allons maintenant montrer que $\alpha = \frac{2}{\pi}$. Pour cela, prenons un cercle de rayon 2, donc de longueur 4π . Si ce cercle est tangent aux lattes du parquet, il y a deux points d'intersection. Sinon, il y a également deux points d'intersection. Ainsi, le nombre moyen d'intersection est de 2, ce qui nous donne l'équation $2 = \alpha \times 4\pi$. En résolvant cette équation, nous trouvons $\alpha = \frac{2}{\pi}$.

Ce résultat nous donne la probabilité d'intersection entre une aiguille et les lattes du parquet lorsque la taille de l'aiguille est inférieure à l'écartement des lattes.

Exercice: rendre le jeu équilibré?

En reprenant le même type de raisonnement que ci-dessus, trouver la distance entre les lattes du parquet qui rend le jeu équilibrer entre les deux joueurs introduit par Buffon (l'un pariant sur le fait que l'aiguille croise une ligne de la trame du parquet, l'autre pariant sur le fait que l'aiguille ne croise pas une ligne de la trame du parquet).

1.2 Méthode de Monte-Carlo

La méthode de Monte-Carlo, est une méthode de calcul numérique qui consiste à utiliser des nombres aléatoires pour résoudre des problèmes déterministes. Elle est utilisée dans de nombreux domaines, comme la physique, la chimie, la biologie, la finance, ou encore l'apprentissage automatique. Cette méthode basée sur la loi des grands nombres a été mis au point à Los Alamos, dans le cadre du projet Manhattan (dont l'objectif était le développement du nucléaire civil et militaire) par un groupe de scientifiques dont les plus connus sont: [John von Neumann](#)², [Nicholas Metropolis](#)³ ou encore [Stanisław Ulam](#)⁴

Dans le cadre du projet Manhattan, il s'agissait de calculer des intégrales de manière numérique pour modéliser l'évolution de particules, en utilisant des nombres aléatoires.

Eckhardt (1987) donne un bref aperçu historique, et mentionne les premières description de la méthode du rejet et de la méthode de l'inversion dans des lettres entre Von Neumann et Ulam datant de 1947. Ulam aurait une l'idée d'utiliser de telles méthodes pour résoudre le jeu du solitaire lors d'un séjour à l'hôpital en 1946, et éviter ainsi de faire des calculs combinatoires fastidieux. Rapidement, la possibilité d'appliquer cette approche pour des calculs en physique mathématiques (diffusion des neutrons notamment) lui serait apparue prometteuse. Le développement de l'informatique naissante allait permettre une mise en oeuvre pratique de ces idées, et c'est ainsi que la méthode de Monte-Carlo est née. Le nom Monte-Carlo est lui venu du besoin de confidentialité du projet, et provient du nom de la ville de Monte-Carlo, connue pour ses jeux de hasard, où l'oncle de Stanisław Ulam aimait se rendre pour assouvir sa soif de jeu. Ce serait N. Metropolis qui aurait proposé ce nom (cf. Metropolis 1987):

It was at that time that I suggested an obvious name for the statistical method—a suggestion not unrelated to the fact that Stan had an uncle who would borrow money from relatives because he “just had to go to Monte Carlo”.


²[John von Neumann](#): (1903-1957) mathématicien et physicien américano-hongrois, un des pères de l'informatique.

³[Nicholas Metropolis](#): (1915-1999), physicien gréco-américain, est des initiateurs de la méthode de Monte Carlo et du recuit simulé

⁴[Stanisław Ulam](#): (1909-1984)


1.3 Autres méthodes stochastiques populaires

1.3.1 Méthode d'Hasting-Metropolis

L'algorithme de Hasting-Metropolis est une méthode MCMC (: *Monte Carlo Markov Chains*) dont le but est d'obtenir un échantillonnage aléatoire d'une distribution de probabilité quand l'échantillonnage direct en est difficile (en particulier en grande dimension)

Un avantage est qu'il ne requiert la connaissance de loi de densité qu'à constante multiplicative près.

1.3.2 Recuit simulé

Le recuit simulé ( *simulated annealing*) est une méthode (empirique) d'optimisation, inspirée d'un processus, le recuit, utilisé en métallurgie. On alterne dans cette dernière des cycles de refroidissement lent et de réchauffage (recuit) qui ont pour effet de minimiser l'énergie du matériau. Cette méthode est transposée en optimisation pour trouver les extrema d'une fonction.

2 Notations et rappels

On considère un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, composé d'un ensemble Ω , d'une tribu \mathcal{F} , et d'une mesure de probabilité \mathbb{P} .

Definition 2.1 (Variable aléatoire). Soit (E, \mathcal{E}) un espace mesurable. Une *variable aléatoire* (ou v.a.) est une application mesurable

$$\begin{aligned} X : \Omega &\rightarrow E \\ \omega &\mapsto X(\omega). \end{aligned}$$

C'est-à-dire que pour tout $B \in \mathcal{E}$, $\{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$. Cet ensemble se réécrit souvent sous la forme

$$\begin{aligned} \{\omega \in \Omega : X(\omega) \in B\} &= X^{-1}(B) \\ &= \{X \in B\}. \end{aligned}$$

Cette définition permet de transposer l'aléa qui provient de Ω dans l'espace E . L'hypothèse $\{X \in B\} \in \mathcal{F}$ assure que cet ensemble est bien un évènement et donc que l'on peut calculer sa probabilité.

- Si $E = \mathbb{R}$, on prendra alors la tribu borélienne $\mathcal{F} = \mathcal{B}(\mathbb{R})$ et on parlera alors de v.a. réelle.
- Si $E = \mathbb{R}^d$, on parlera de *vecteurs aléatoires*.

Une fois que l'aléa a été transposé de Ω vers E , on souhaite également transposer la probabilité \mathbb{P} sur E . Ceci motive l'introduction de la notion de loi.

Definition 2.2 (Loi d'une variable aléatoire). Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E, \mathcal{E})$ une variable aléatoire. On appelle *loi de X* la mesure de probabilité

$$\begin{aligned} \mathbb{P}_X : \mathcal{E} &\rightarrow [0, 1] \\ B &\mapsto \mathbb{P}(X \in B). \end{aligned}$$

Les propriétés de \mathbb{P} assurent que \mathbb{P}_X est bien une loi de probabilité sur l'espace mesurable (E, \mathcal{E}) .

2.1 Loi discrètes

Les variables aléatoires discrètes sont celles à valeurs dans un ensemble E discret, le plus souvent \mathbb{N} , muni de la tribu pleine $\mathcal{F} = \mathcal{P}(E)$.

Exemple 2.1 (Loi de Bernoulli). La loi la plus simple est la *loi de Bernoulli* de paramètre $p \in [0, 1]$, définie sur $\{0, 1\}$ par $\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0) = p$ qui modélise une expérience aléatoire à deux issues (succès = 1 et échec = 0).

Exemple 2.2 (Loi binomiale). En sommant des variables aléatoires indépendantes de *loi de Bernoulli* on obtient une loi binomiale : $\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$, pour $k \in \{0, \dots, n\}$, qui modélise le nombre de succès parmi n lancers.

Exemple 2.3 (Loi géométrique). En observant le nombre d'expériences nécessaires avant d'obtenir un succès, on obtient une *loi géométrique* : $\mathbb{P}(X = k) = p(1-p)^{k-1}$, pour $k \geq 1$. C'est une loi de probabilité discrète qui décrit le comportement du nombre d'événements se produisant dans un intervalle de temps fixé, si ces événements se produisent avec une fréquence moyenne ou espérance connue, et indépendamment du temps écoulé depuis l'événement précédent (e.g., nombre de clients dans une file d'attente, nombre de mutations dans un gène, etc.).

Exemple 2.4 (Loi de Poisson). La loi de Poisson de paramètre $\lambda > 0$ est définie par $\mathbb{P}(X = k) = e^{-\lambda} \lambda^k / k!$, pour $k \in \mathbb{N}$.

2.2 Loïs continues

Parmi les variables aléatoires réelles non discrètes, beaucoup peuvent se représenter avec une densité, c'est-à-dire qu'il existe une fonction mesurable $f : \mathbb{R} \rightarrow [0, \infty[$ d'intégrale 1. La loi d'une telle variable aléatoire X est alors donnée pour tout $A \in \mathcal{B}(\mathbb{R})$ par

$$\mathbb{P}(X \in A) = \int_A f(x) \, dx .$$

Les propriétés de l'intégrale de Lebesgue assure que cette formule définit bien une loi de probabilité.

Exemple 2.5 (Loi uniforme). La loi uniforme sur un ensemble $B \in \mathcal{B}(\mathbb{R})$, s'obtient avec la densité définie par

$$f(x) = \mathbb{1}_B(x) / \lambda(B) ,$$

où $\lambda(B)$ représente la mesure de Lebesgue de l'ensemble B . En particulier pour la loi uniforme sur le segment $[0, 1]$ on obtient la fonction suivante:

$$f(x) = \mathbb{1}_{[0,1]}(x) .$$

Si une variable aléatoire U suit une telle loi on note $U \sim \mathcal{U}([0, 1])$.

Exemple 2.6 (Loi exponentielle). La loi exponentielle de paramètre $\gamma > 0$ est obtenue avec la densité donnée par

$$f(x) = \gamma e^{-\gamma x} \mathbb{1}_{\mathbb{R}_+}(x) .$$

Si une variable aléatoire X suit cette loi on note $X \sim \text{Exp}(\gamma)$.

Exemple 2.7 (Loi normale/gaussienne univariée). On obtient la loi normale de paramètre $\mu \in \mathbb{R}$ et $\sigma^2 > 0$ correspond à loi dont la densité est donnée par la fonction réelle:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} .$$

Si une variable aléatoire X suit une telle loi on note $X \sim \mathcal{N}(\mu, \sigma^2)$, μ correspondant à l'espérance de la loi, et σ^2 à sa variance. On nomme loi *normale centrée réduite* le cas correspondant à $\mu = 0$ et $\sigma^2 = 1$.

Exemple 2.8 (Loi normale multivariée). On peut étendre les lois normales au cas multidimensionnel. Fixons $d \in \mathbb{N}^*$ un entier non nul. Pour un vecteur $\mu \in \mathbb{R}^d$ et une matrice symétrique-définie positive $\Sigma \in \mathbb{R}^{d \times d}$, la densité normale multivariée associée est donnée par la fonction:


$$f(x) = \frac{1}{(2\pi)^{\frac{d}{2}} \det(\Sigma)} e^{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)}$$

Notons que μ est l'espérance de la loi et Σ la matrice de variance-covariance.

2.3 Fonction de répartition

La notion de variable aléatoire n'est pas facile à manipuler puisqu'elle part d'un espace Ω dont on ne sait rien. On souhaite donc caractériser la loi d'une variable aléatoire en ne considérant que l'espace d'arrivée (E, \mathcal{E}) .

Plusieurs outils existent : la fonction de répartition (pour des variables aléatoires réelles), la fonction caractéristique (pour des variables aléatoires dans \mathbb{R}^d), la fonction génératrice des moments (pour des variables aléatoires discrètes), etc. On se contente ici de la fonction de répartition qui nous sera utile pour simuler des variables aléatoires, ainsi que son inverse au sens de Levy.

Definition 2.3 (Fonction de répartition : *cumulative distribution function*). Soit X une variable aléatoire sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. La fonction de répartition de X est la fonction F_X définie sur \mathbb{R} par

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) \\ &= \mathbb{P}(X \in]-\infty, x]) . \end{aligned}$$

Example 2.9 (Cas discret). Soit $(x_i)_{i \in I}$ une suite ordonnée de réels, avec $I \subset \mathbb{N}$. Si X est une variable aléatoire discrète prenant les valeurs $(x_i)_{i \in I}$ et de loi $(p_i = \mathbb{P}(X = x_i))_{i \in I}$, alors

$$F_X(x) = \sum_{i \in I} p_i \mathbb{1}_{[x_i, \infty[}(x) .$$

Example 2.10 (Cas continu). Si X est une variable aléatoire de densité f , alors

$$F_X(x) = \int_{-\infty}^x f(t) dt .$$

Le graphe des fonctions de répartition des loi de Bernoulli, uniforme et normale sont représentées dans le widget ci-dessous. Notons que la fonction de répartition de la loi normale $\mathcal{N}(0, 1)$, souvent notée Φ , n'admet pas d'expression explicite autre que

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt ,$$

Les valeurs numériques de $\Phi(x)$ étaient autrefois reportées dans des tables¹. Par transformation affine, si $X \sim \mathcal{N}(\mu, \sigma^2)$ — ce que l'on peut aussi écrire : $X = \mu + \sigma Y$, avec $Y \sim \mathcal{N}(0, 1)$ — alors sa fonction de répartition est donnée par $F_X(x) = \Phi((x - \mu)/\sigma)$.

Proposition 2.1 (Propriétés de la fonction de répartition). Soit X une variable aléatoire de fonction de répartition F_X .

1. F_X est une fonction croissante, de limite 0 en $-\infty$ et de limite 1 en $+\infty$.
2. F_X est continue à droite en tout point.
3. Pour tout $x \in \mathbb{R}$, on a $\mathbb{P}(X = x) = F_X(x) - F_X(x-)$, où $F_X(x-) = \lim_{\epsilon \rightarrow 0+} F_X(x - \epsilon)$.
4. Si X a pour densité f , alors F_X est dérivable λ -presque partout de dérivée f .

¹[Wikipedia: loi normale](#)

Pour les démonstrations, voir par exemple (Barbe and Ledoux 2006).

La propriété 3. est utile dans le cas discret : les valeurs prises par X correspondent aux points de discontinuité de F_X et les probabilités associées correspondent à la hauteur du saut.

La propriété 4. donne le lien entre la fonction de répartition d'une variable aléatoire à densité et sa densité. On peut donc retrouver la loi de X à partir de sa fonction de répartition. Le théorème suivant généralise ce résultat à toute variable aléatoire réelle (pas nécessairement discrète ou à densité).

Theorem 2.1 (Caractérisation de la loi d'une variable aléatoire réelle). *La fonction de répartition d'une variable aléatoire caractérise sa loi : deux variables aléatoires ont même loi si et seulement si elles ont même fonction de répartition.*

Démonstration: voir [Wikipedia](#)

On rappelle que la tribu des boréliens est engendrée par la famille d'ensembles $\{]-\infty, x], x \in \mathbb{R}\}$. Le théorème précédent assure que si on connaît la mesure \mathbb{P}_X sur cette famille d'ensembles alors on la connaît partout.


Example 2.11 (Loi exponentielle depuis une loi uniforme). On considère une variable aléatoire U de loi uniforme sur $[0, 1]$ et on pose $X = -\ln(1 - U)$. Déterminons la loi de X en calculant sa fonction de répartition. Pour tout $x \in \mathbb{R}$,

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) \\ &= \mathbb{P}(-\ln(1 - U) \leq x) \\ &= \mathbb{P}(U \leq 1 - e^{-x}) \\ &= \begin{cases} 0 & \text{si } x < 0, \\ 1 - e^{-x} & \text{si } x \geq 0, \end{cases} \end{aligned}$$

où on a utilisé l'égalité $\mathbb{P}(U \leq t) = t$ pour tout $t \in [0, 1]$. Ainsi la variable aléatoire X a la même fonction de répartition qu'une loi exponentielle de paramètre 1. On en conclut que $X \sim \mathcal{Exp}(1)$. Notons que l'on peut aussi montrer que $-\ln(X) \sim \mathcal{E}(1)$, sachant que U et $1 - U$ ont la même loi.

2.4 Fonction quantile, inverse généralisée à gauche

La fonction de répartition étant une fonction croissante on peut donner un sens à son inverse généralisée de la manière suivante.

Definition 2.4 (Fonction quantile/ inverse généralisée : *quantile distribution function*). Soit X une variable aléatoire sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et F_X sa fonction de répartition. La fonction quantile associée $F_X^{\leftarrow} :]0, 1[\rightarrow \mathbb{R}$ est définie par

$$F_X^{\leftarrow}(p) = \inf\{x \in \mathbb{R} : F_X(x) \geq p\} .$$

On parle parfois aussi d'inverse au sens de Levy pour cette inverse généralisée.

Dans le cas où la fonction de répartition F_X est bijective, alors l'inverse de la fonction de répartition coïncide avec la fonction quantile.

La **médiane** est égale à $F_X^{\leftarrow}(1/2)$, les premiers et troisièmes **quartiles** sont égaux à $F_X^{\leftarrow}(1/4)$ et $F_X^{\leftarrow}(3/4)$. Enfin, les déciles sont les quantiles $F_X^{\leftarrow}(k/10)$ pour $k = 1, \dots, 9$.

2.5 Visualisation: densité, fonction de répartition, quantiles, etc.

2.5.1 Cas des variables continues

2.5.2 Cas des variables discrètes

3 Théorèmes asymptotiques

3.1 Loi des grands nombres

Le premier résultat fondamental en probabilités concerne le comportement asymptotique de la moyenne empirique:

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n} .$$

quand on observe n variables aléatoires i.i.d X_1, \dots, X_n , ayant une espérance finie.

Theorem 3.1 (Loi forte des grands nombres). *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) dans $L^1(\Omega, \mathcal{F}, \mathbb{P})$. Notons $\mu = \mathbb{E}[X_1]$. Alors \bar{X}_n converge vers μ presque sûrement :*

$$\mathbb{P} \left(\frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{} \mu \right) = 1 .$$

Interprétation: Intuitivement, la probabilité d'un événement A correspond à la fréquence d'apparition de A quand on répète une expérience qui fait intervenir cet événement. Par exemple, si on dispose une pièce truquée, on estimera la probabilité d'apparition du côté pile en lançant la pièce un grand nombre de fois et en comptant le nombre de pile obtenu. La loi des grands nombres justifie a posteriori cette intuition : si X_1, \dots, X_n sont i.i.d. de loi de Bernoulli de paramètre p , alors

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{p.s.} p .$$

Le membre de gauche correspond au nombre empirique de pile obtenu, celui de droite à la valeur théorique.

Remarque: Bien qu'assez intuitif, ce théorème est difficile à démontrer, cf.(Ouvrard 2008; Barbe and Ledoux 2006) ou encore (Williams 1991) pour une version de preuve avec des martingales.

Pour aller plus loin:

Quant p varie, à n fixé...les signaux générés sont très très proches, ce qui ne devrait pas être le cas sans structuration particulière de la génération. L'aléa est imparfait!

3.2 Théorème central limite (TCL)

Une fois la loi des grands nombres établie, on peut se demander quel est l'ordre suivant dans le développement asymptotique de $\bar{X}_n - \mu$, ou de manière équivalente de $S_n - n\mu$, où $S_n = X_1 + \dots + X_n$. Le théorème suivant répond à cette question, en donnant une convergence en loi d'une transformation affine de la moyenne empirique:

Theorem 3.2 (Théorème central limite). *Soit X_1, \dots, X_n une suite de variables aléatoires i.i.d de variance $\sigma^2 = \text{var}(X_1) \in]0, \infty[$. On note $\mu = \mathbb{E}[X_1]$ leur espérance. Alors*

$$\sqrt{n} \left(\frac{\bar{X}_n - \mu}{\sigma} \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} N ,$$

où N suit une loi normale centrée réduite : $N \sim \mathcal{N}(0, 1)$.

Preuve: cf.(Ouvrard 2008; Barbe and Ledoux 2006).

On peut interpréter ce théorème grossièrement de la façon suivante: la moyenne empirique de variables aléatoires i.i.d de variance σ^2 se comporte asymptotiquement comme une loi normale $\mathcal{N}(\mu, \frac{\sigma^2}{n})$, ce que l'on écrit avec un abus de notation:

$$\bar{X}_n \approx \mathcal{N}(\mu, \frac{\sigma^2}{n}) .$$

En termes de somme cumulée empirique, la convergence se réécrit

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} N .$$

Les hypothèses de ce théorème sont plutôt faibles (il suffit de supposer une variance finie). Pourtant, le résultat est universel : la loi de départ peut être aussi farfelue que l'on veut, elle se rapprochera toujours asymptotiquement d'une loi normale.

On rappelle que la convergence en loi est équivalente à la convergence des fonctions de répartition en tout point de continuité de la limite. Ainsi, le théorème central limite se réécrit de la manière suivante : pour tout $a < b$, notons $\alpha_n = \mathbb{P} \left(\bar{X}_n \notin \left[\mu + \frac{a\sigma}{\sqrt{n}}, \mu + \frac{b\sigma}{\sqrt{n}} \right] \right)$.

Ainsi,

$$\begin{aligned}
 1 - \alpha_n &= \mathbb{P} \left(\bar{X}_n \in \left[\mu + \frac{a\sigma}{\sqrt{n}}, \mu + \frac{b\sigma}{\sqrt{n}} \right] \right) \\
 &= \mathbb{P} \left(\frac{\bar{X}_n - \mu}{\sigma} \in \left[\frac{a}{\sqrt{n}}, \frac{b}{\sqrt{n}} \right] \right) \\
 &= \mathbb{P} \left(a \leq \sqrt{n} \left(\frac{\bar{X}_n - \mu}{\sigma} \right) \leq b \right) \\
 &\xrightarrow{n \rightarrow \infty} \int_a^b \varphi(x) dx.
 \end{aligned}
 \tag{3.1}$$

où l'on note φ (resp. Φ) la densité (resp. la fonction de répartition) d'une loi normale centrée réduite, définie pour tout $x \in \mathbb{R}$ par $\varphi(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$ (resp. $\Phi(x) = \int_{-\infty}^x \varphi(u) du$).

Dans le cas classique d'un intervalle de confiance à 95%, c'est-à-dire quand $\alpha_n = 0.05$, et en prenant un intervalle de confiance symétrique (alors $a = -t$ et $b = q$) on obtient $1 - \alpha_n = \int_{-q}^q \varphi(x) dx = \Phi(q) - \Phi(-q) = 2\Phi(q) - 1 \implies q = \Phi^{-1}(1 - \frac{\alpha_n}{2})$ et q est donc le quantile de niveau $1 - \frac{\alpha_n}{2}$ de la loi normale centrée réduite. Numériquement on peut facilement évaluer q et vérifier que $q \approx 1.96$ avec `scipy`:

```
from scipy.stats import norm
q = norm.ppf(1-0.05/2)
print(f"Gaussienne centrée réduite,\nQuantile de niveau (1- /2):\nq = {q:.2f}")
```

```
Gaussienne centrée réduite,
Quantile de niveau (1- /2):
q = 1.96
```

Exemple 3.1 (Loi de Bernoulli). On considère des variables aléatoires X_1, \dots, X_n i.i.d. suivant une loi de Bernoulli de paramètre $p \in]0, 1[$, dont l'espérance et la variance sont respectivement p et $p(1 - p)$. Le théorème central limite donne alors

$$\sqrt{n} \left(\frac{\bar{X}_n - p}{p(1 - p)} \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} N,$$

avec $N \sim \mathcal{N}(0, 1)$. Cette convergence est illustrée dans le widget ci-dessous. Le contexte est le suivant. On répète t fois le processus, qui consiste à afficher $(\bar{X}_k)_{k \in [n]}$, où les n variables aléatoires sont i.i.d. et suivent une loi de Bernoulli de paramètre p .

Une autre illustration possible de la convergence donnée par le TCL est celle qui correspond au point de vue donnée par l'analyse. Pour cela supposons que l'on ait une suite de variables aléatoires réelles X_1, \dots, X_n , i.i.d. dont la fonction de densité commune est notée par f .

On rappelle quelques éléments de probabilités concernant les densités. Pour cela on rappelle la définition de la convolution deux fonctions. Pour cela prenons deux fonctions f et g définies sur \mathbb{R} et qui sont intégrables au sens de Lebesgue. La **convolution** de f par g est alors la fonction $f * g$ suivante:

$$\begin{aligned} \mathbb{R} &\mapsto \mathbb{R} \\ x &\mapsto \int_{-\infty}^{+\infty} f(x-y)g(y)dy . \end{aligned}$$

i Note

On peut aussi obtenir $f * g(x)$ en calculant $\int_{\mathbb{R}^2} f(u)g(v)\mathbb{1}_{u+v=x}dudv$.

Theorem 3.3 (Loi de la somme et convolutions). *Soient X et Y des v.a. indépendantes de densités f et g respectivement, la loi de $X + Y$ est donnée par la convolution $f * g$.*

Rappel: pour un scalaire $\alpha \neq 0$, la densité de αX est donnée par la fonction $x \mapsto \frac{1}{|\alpha|} \cdot f\left(\frac{x}{\alpha}\right)$.

Corollary 3.1 (Loi de la moyenne). *Soient X_1, \dots, X_n des v.a. i.i.d. de densité f , la densité de \bar{X}_n est donnée par la fonction $x \mapsto n \cdot [f * \dots * f](n \cdot x)$.*

Dessous, pour X_1, \dots, X_n , i.i.d., de densité f , on affiche la densité de la loi de \bar{X}_n .

Pour aller plus loin sur les convolutions, voir la vidéo de 3Blue1Brown à ce sujet: [Convolutions | Why X+Y in probability is a beautiful mess](#)

4 Simulation

Dans ce chapitre on se demande comment simuler en pratique des variables aléatoires i.i.d. L'idée est de commencer par le cas de variables aléatoires de loi uniforme et d'en déduire les autres lois.

4.1 Variables aléatoires uniformes

On rappelle qu'une variable aléatoire U suit une loi uniforme sur $[0, 1]$, noté $\mathcal{U}([0, 1])$ si sa fonction de répartition F_U est donnée par

$$F_U(x) = \begin{cases} 0, & \text{si } x < 0, \\ x, & \text{si } x \in [0, 1], \\ 1, & \text{si } x > 1. \end{cases}$$

Unable to display output for mime type(s): text/html



Figure 4.1: Fonction de répartition de la loi uniforme

Unable to display output for mime type(s): text/html

Figure 4.2

L'objectif est de simuler sur machine une suite U_1, \dots, U_n de variables aléatoires i.i.d. de loi $\mathcal{U}([0, 1])$. Plusieurs problèmes apparaissent alors :


- Une machine est déterministe.
- Les nombres entre 0 et 1 donnés par la machine sont de la forme $k/2^p$, pour $k \in \{0, \dots, 2^p-1\}$. On ne pourra donc jamais générer des nombres qui ne sont pas de cette forme.
- Vérifier qu'une suite est bien i.i.d. est un problème difficile.

Definition 4.1 (Générateur de nombres pseudo-aléatoires). Un **générateur de nombres pseudo-aléatoires** (: *Pseudo Random Number Generator*, PRNG), est un algorithme déterministe récursif qui renvoie une suite U_1, \dots, U_n dans $[0, 1]$ qui a un “comportement similaire” à une suite i.i.d. de loi $\mathcal{U}([0, 1])$. Pour être plus rigoureux, ces nombres sont en fait des nombres entiers générés uniformément sur un certain interval. Dans un second temps, une transformation simple (normalisation) permet d'obtenir des nombres flottants (: *floats*) entre 0 et 1.

i Pour aller plus loin

Parfois il est utile d'aller chercher dans le code source certaines information pour savoir comment les fonctions sont codées dans les packages que l'on utilise. Par exemple, pour `numpy` que l'on utilise fréquemment, on peut voir l'opération choisie ici: [Random: int -> float en numpy](#).


Un tel algorithme se construit de la manière suivante :

1. On part d'une graine (: *seed*) U_0 qui détermine la première valeur de manière la plus arbitraire possible.
2. La procédure récursive s'écrit $U_{n+1} = f(U_n)$, où f est une transformation déterministe, de sorte que U_{n+1} est le plus indépendant possible de U_1, \dots, U_n .
 - La fonction f est déterministe et prend ses valeurs dans un ensemble fini, donc l'algorithme est périodique. Le but est donc d'avoir la plus grande période possible.
 - Notons qu'une fois que la graine est fixée, alors l'algorithme donne toujours les mêmes valeurs. Fixer la graine peut donc être très utile pour répéter des simulations dans des conditions identiques et ainsi repérer des erreurs.

Exercice: bug ou feature?

Reprendre les widgets du chapitre [Théorèmes asymptotiques](#) et faites varier doucement le paramètre p (de Bernoulli). Que constatez-vous? Proposer une explication potentielle.

4.1.1 Générateur congruentiel linéaire

La plupart des PRNG s'appuient sur des résultats arithmétiques. Un des plus connus est celui appelé Générateur congruentiel linéaire (: *Linear congruential generator*, LCG). Il est défini comme suit: on construit récursivement une suite d'entiers X_i via la congruence

$$X_{n+1} = aX_n + b \mod m ,$$

où a, b, m sont des entiers bien choisis pour que la suite obtenue ait de bonnes propriétés. Il suffit alors de considérer X_n/m . Par exemple, la fonction `rand` sur `scilab` utilise cette congruence avec $m = 2^{31}$, $a = 843\,314\,861$, et $b = 453\,816\,693$.

4.1.2 Générateurs alternatifs

Les langages Python et R utilisent par défaut le générateur Mersenne-Twister qui s'appuie sur la multiplication vectorielle, mais d'autres générateurs sont aussi disponibles. Ce générateur a pour période $m = 2^{19937} - 1$, nombre qu'on peut raisonnablement considérer comme grand.

Pour `numpy` la méthode par défaut est PCG64 (cf. [documentation de numpy](#)), qui dispose de meilleures garanties statistiques (Voir le site <https://www.pcg-random.org> pour cela).

4.1.3 Usage en numpy

On suppose désormais disposer d'un générateur pseudo-aléatoire sur $[0, 1]$. En `numpy` depuis la version 1.17, une bonne manière d'utiliser des éléments aléatoires est d'utiliser un générateur que l'on définit soi-même:

```
seed = 12345 # Toujours être conscient qu'une graine existe
rng = np.random.default_rng(seed) #
print(rng.random()) ## un tirage uniforme sur [0,1]
print(rng.random(size=5)) ## cinq tirages uniformes sur [0,1]
print(rng.random(size=(3, 2))) ## matrice 3x2, à entrées unif. sur [0,1]
```

```
0.22733602246716966
[0.31675834 0.79736546 0.67625467 0.39110955 0.33281393]
[[0.59830875 0.18673419]
 [0.67275604 0.94180287]
 [0.24824571 0.94888115]]
```

Dans la suite on va voir comment générer d'autres lois à partir de la loi uniforme, mais il est clair que les logiciels modernes proposent un large éventail de distribution classique (gaussienne, exponentielle, etc.). Une liste exhaustive est donnée [ici](#) pour `numpy`.

i Pour aller plus loin

Une excellente discussion sur les bonnes pratiques aléatoires en `numpy`, et l'usage de `np.random.default_rng` est donnée dans ce [blog post d'Albert Thomas](#).

4.1.4 Propriété de la loi uniforme

On verra souvent apparaître la variable aléatoire $1 - U$ où $U \sim \mathcal{U}([0, 1])$. Il se trouve que $1 - U$ suit aussi une loi uniforme sur $[0, 1]$ comme le montre le calcul de sa fonction de répartition. Ainsi pour tout $x \in [0, 1]$ on obtient

$$\begin{aligned}\mathbb{P}(1 - U \leq x) &= \mathbb{P}(U \geq 1 - x), \\ &= 1 - (1 - x), \\ &= x.\end{aligned}$$

On peut démontrer facilement la même relation pour $x < 0$ et $x > 1$, d'où le résultat.

4.2 Méthode d'inversion

L'idée de la méthode d'inversion repose sur le résultat suivant :

4.3 Rappel sur la fonction quantile

Rappel : Pour F une fonction définie sur \mathbb{R} à valeurs dans $[0, 1]$, croissante, on note

$$\forall q \in]0, 1[, \quad F^{\leftarrow}(q) = \inf\{x \in \mathbb{R} : F(x) \geq q\} \quad (4.1)$$

Theorem 4.1 (Caractérisation des quantiles). *Soit F une fonction définie sur \mathbb{R} à valeurs dans $[0, 1]$, croissante et continue à droite, alors pour tout $q \in]0, 1[$, on a*

$$\{x \in \mathbb{R} : F(x) \geq q\} = \{x \in \mathbb{R} : x \geq F^{\leftarrow}(q)\} \quad (4.2)$$

Proof.

- Cas \subset : Soit $x \in \mathbb{R}$ t.q. $F(x) \geq q$, alors par définition de l'inf dans Equation 4.1, $x \geq F^{\leftarrow}(q)$.
- Cas \supset : Soit $x \in \mathbb{R}$ t.q. $x \geq F^{\leftarrow}(q)$ alors pour tout $\epsilon > 0$, $x + \epsilon > F^{\leftarrow}(q)$, donc $F(x + \epsilon) \geq q$. Puis, par continuité à droite de F , $F(x) \geq q$.

□

4.4 Méthode d'inversion

Theorem 4.2 (Méthode d'inversion). Soit X une v.a réelle, et $U \sim \mathcal{U}([0, 1])$, alors la variable aléatoire $F_X^{\leftarrow}(U)$ a même loi que X .

Proof. En utilisant le théorème précédent, on a $\mathbb{P}(F_X^{\leftarrow}(U) \leq x) = \mathbb{P}(U \leq F_X(x))$ pour tout $x \in \mathbb{R}$. Puis, comme U est une loi uniforme sur $[0, 1]$, $\mathbb{P}(U \leq F_X(x)) = F_X(x)$.

On en déduit donc que la loi de $F_X^{\leftarrow}(U)$ est la même que celle de X , car les deux v.a. ont la même fonction de répartition. \square

Example 4.1 (Simulation d'une loi exponentielle). On rappelle que la loi exponentielle de paramètre $\lambda > 0$ a pour densité

$$f_\lambda(x) = \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}_+}(x) .$$

et donc pour fonction de répartition

$$F_\lambda(x) = (1 - e^{-\lambda x}) \mathbb{1}_{\mathbb{R}_+}(x) .$$

On vérifie que F_λ est bijective de \mathbb{R}_+ dans $]0, 1[$ et que son inverse est donnée pour tout $u \in]0, 1[$ par

$$F_\lambda^{-1}(u) = -\frac{1}{\lambda} \log(1 - u) .$$

Example 4.2 (Simulation d'une loi de Weibull). La loi de Weibull de paramètre $\lambda > 0$ et $k > 0$ est caractérisée par la fonction de répartition

$$F(x) = (1 - e^{-(x/\lambda)^k}) \mathbb{1}_{\mathbb{R}_+}(x)$$

C'est une loi utilisée dans différents domaines, notamment en gestion des risques (hydrologie, finance, assurance, etc.). Le calcul de l'inverse généralisée F^{\leftarrow} est immédiat car F est bijective sur $]0, \infty[$:

$$F^{\leftarrow}(u) = \lambda(-\ln(1 - u))^{\frac{1}{k}} , \quad u \in]0, 1[.$$

La méthode d'inversion s'applique : si $U \sim \mathcal{U}([0, 1])$, alors la variable aléatoire $X = \lambda(-\ln(U))^{\frac{1}{k}}$ suit une loi de Weibull de paramètres λ et k .

Malheureusement, la fonction F n'est pas toujours inversible (penser aux lois discrètes) c'est donc pourquoi on utilise l'inverse généralisée ou fonction quantile introduite dans la section [Notations](#):

$$F^{\leftarrow}(p) = \inf\{x \in \mathbb{R} : F(x) \geq p\} .$$

Interprétation: Définir l'inverse d'une fonction de répartition F revient à résoudre l'équation $F(x) = \alpha$ d'inconnue x pour un α fixé. Si F n'est pas bijective, deux problèmes apparaissent :

- l'équation n'a aucune solution ce qui revient à dire que F n'est pas surjectif (graphiquement, F présente des sauts) ;
- l'équation a plusieurs solutions ce qui revient à dire que F n'est pas injective (graphiquement cela se matérialise par un plateau à la hauteur α). Un exemple classique est celui où F est la fonction de répartition d'une variable aléatoire discrète.

Le passage à l'inéquation $F(x) \geq u$ permet de contourner la non-surjectivité : on ne regarde non plus les droites horizontales $y = u$ mais la région $\{y \geq \alpha\}$. Le choix de l'inf dans la définition de F^{\leftarrow} permet de contourner la non-injectivité : vu qu'il y a possiblement plusieurs x tels que $F(x) \geq u$, on choisit le "premier". Ces considérations sont illustrées en Figure [4.3](#).

Unable to display output for mime type(s): text/html

Figure 4.3

Remarques additionnelles:

- La fonction F étant croissante, la quantité $F^{\leftarrow}(u)$ correspond au premier instant où F dépasse α . Si F est bijective (ce qui équivaut dans ce cas à strictement croissante et injective), alors $F^{\leftarrow} = F^{-1}$.
- La fonction F^{\leftarrow} n'est rien d'autre que la fonction quantile : si $0 < \alpha < 1$, $q_{1-\alpha} = F^{\leftarrow}(1 - \alpha)$ est le quantile d'ordre $(1 - \alpha)$ de F . Par exemple, $F^{\leftarrow}(1/2)$ correspond à la médiane.
- Notons que si $u = 0$, on peut alors naturellement poser $F^{\leftarrow}(0) = -\infty$. De même, avec la convention la convention $\inf \emptyset = +\infty$, on peut alors étendre la définition de F^{\leftarrow} à $u = 1$ (mais $F^{\leftarrow}(1)$ n'est pas toujours égal à ∞ , voir les exemples ci-dessous).

Exemple 4.3 (Simulation d'une loi de Bernoulli). La fonction de répartition F d'une loi de Bernoulli de paramètre $p \in]0, 1[$ est donnée par

$$F(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1 - p & \text{si } 0 \leq x < 1, \\ 1 & \text{si } x \leq 1. \end{cases}$$

L'inverse généralisée de F peut ainsi être calculée via la formule de l'exemple ?? :

$$F^{\leftarrow}(u) = \begin{cases} 0 & \text{si } 0 < u \leq 1 - p, \\ 1 & \text{si } 1 - p < u \leq 1. \end{cases}$$

ce qui se réécrit plus simplement $F^{\leftarrow}(u) = \mathbb{1}_{\{1-p < u\}}$.

Ainsi, si U suit une loi uniforme sur $[0, 1]$ alors $\mathbb{1}_{\{1-p < U\}}$ suit une loi de Bernoulli de paramètre p . Comme $1 - U$ suit aussi une loi uniforme sur $[0, 1]$, on en déduit que $\mathbb{1}_{\{U < p\}}$ suit une loi de Bernoulli de paramètre p . Notons que l'on peut remplacer l'inégalité stricte par une inégalité large.

Proposition 4.1 (Loi à support fini). *Soit X une variable aléatoire discrète prenant uniquement les valeurs $x_1 < \dots < x_r$ (r modalités possibles) avec probabilité p_1, \dots, p_r (donc $p_1 + \dots + p_r = 1$). On vérifie que pour tout $u \in]0, 1[$,*

$$F^{\leftarrow}(u) = \begin{cases} x_1 & \text{si } 0 < u \leq p_1, \\ x_2 & \text{si } p_1 < u \leq p_1 + p_2, \\ \vdots & \\ x_r & \text{si } \sum_{i=1}^{r-1} p_i < u < 1. \end{cases}$$

Sur cet exemple, on peut prolonger la définition de F^{\leftarrow} à $u = 1$ en posant $F^{\leftarrow}(1) = x_r$. L'inverse généralisée se réécrit alors sous la forme

$$F^{\leftarrow}(u) = \sum_{k=1}^r x_k \mathbb{1}_{\{\sum_{i=1}^{k-1} p_i < u \leq \sum_{i=1}^k p_i\}},$$

où on a posé $p_0 = 0$.

L'expression précédente s'étend directement au cas où X prend un nombre (infini) dénombrable de valeurs, la somme devenant alors une série.

La méthode est illustrée ci-dessous pour quelques lois intéressantes:

4.5 Méthode de rejet

L'idée de la méthode de rejet est la suivante. On souhaite simuler une variable aléatoire X de densité f , appelée **loi cible**, mais f est trop compliquée pour que la simulation puisse se faire directement. On dispose cependant d'une autre densité g possédant les propriétés suivantes :

- on sait simuler Y de loi g ,
- il existe $m > 0$ tel que $f(x) \leq m \cdot g(x)$,
- on sait évaluer le **rapport d'acceptation** $r(x) = \frac{f(x)}{mg(x)}$.

Remarquons d'ores et déjà que la constante m est nécessairement plus grande que 1 car

$$1 = \int_{\mathbb{R}} f(x) dx \leq m \int_{\mathbb{R}} g(x) dx = m.$$

L'idée est alors de considérer deux suites i.i.d. de variables aléatoires indépendantes entre elles:

- $(Y_n)_{n \geq 1}$ de loi g ,
- $(U_n)_{n \geq 1}$ de loi uniforme sur $[0, 1]$.

En pratique, Y_n correspond à une proposition et U_n permettra de décider si on accepte la proposition ou non. Si oui, alors on conserve Y_n , sinon on simule Y_{n+1} . Le rapport d'acceptation, c'est-à-dire la proportion de Y_n acceptées, correspond à $r(x)$.

Autrement dit, pour simuler X de densité f , il suffit de simuler Y de densité g et U uniforme jusqu'à ce que $U \leq r(Y)$. La proposition suivante assure que cette méthode donne bien le résultat voulu.

Proposition 4.2 (Méthode de rejet). *Soit $T = \inf\{n \geq 1 : U_n \leq r(Y_n)\}$ le premier instant où le tirage est accepté. Alors :*

- T suit une loi géométrique de paramètre $1/m$,
- la variable aléatoire $X = Y_T$ a pour densité f et est indépendante de T .

Proof. Il s'agit d'étudier la loi du couple (X, T) . Pour $x \in \mathbb{R}$ et $n \in \mathbb{N}^*$, on écrit $\mathbb{P}(X \leq x, T = n) = \mathbb{P}(U_1 > r(Y_1), \dots, U_{n-1} > r(Y_{n-1}), U_n \leq r(Y_n), Y_n \leq x)$. Les n tirages étant iid, on obtient

$$\mathbb{P}(X \leq x, T = n) = \mathbb{P}(U_1 > r(Y_1))^{n-1} \mathbb{P}(U_n \leq r(Y_n), Y_n \leq x).$$

Concernant le premier terme, les variables aléatoires Y_1 et U_1 sont indépendantes donc leur

loi jointe correspond au produit des densités :

$$\begin{aligned}
\mathbb{P}(U_1 > r(Y_1)) &= \mathbb{P}((U_1, Y_1) \in \{(u, y) \in \mathbb{R}^2 : u > r(y)\}) \\
&= \int_{\mathbb{R}^2} \mathbb{1}_{\{u > r(y)\}} (\mathbb{1}_{[0,1]}(u) g(y)) \, du dy \\
&= \int_{\mathbb{R}} \left(\int_0^1 \mathbb{1}_{\{u > r(y)\}} \, du \right) g(y) \, dy \\
&= \int_{\mathbb{R}} (1 - r(y)) g(y) \, dy,
\end{aligned}$$

ce qui se réécrit, comme f et g sont des densités et que $r(y) = f(y)/(m \cdot g(y))$:

$$\begin{aligned}
\mathbb{P}(U_1 > r(Y_1)) &= \int_{\mathbb{R}} g(y) \, dy - \int_{\mathbb{R}} \frac{f(y)}{m} \, dy \\
&= 1 - \frac{1}{m}.
\end{aligned}$$

Le deuxième terme se calcule de manière analogue :

$$\begin{aligned}
\mathbb{P}(U_n \leq r(Y_n), Y_n \leq x) &= \int_{\mathbb{R}^2} \mathbb{1}_{\{u \leq r(y)\}} \mathbb{1}_{\{y \leq x\}} (\mathbb{1}_{[0,1]}(u) g(y)) \, du dy \\
&= \int_{\mathbb{R}} \left(\int_0^1 \mathbb{1}_{\{u \leq r(y)\}} \, du \right) \mathbb{1}_{\{y \leq x\}} g(y) \, dy,
\end{aligned}$$

c'est-à-dire

$$\begin{aligned}
\mathbb{P}(U_n \leq r(Y_n), Y_n \leq x) &= \int_{\mathbb{R}} r(y) \mathbb{1}_{\{y \leq x\}} g(y) \, dy \\
&= \int_{-\infty}^x \frac{f(y)}{m} \, dy \\
&= \frac{F(x)}{m},
\end{aligned}$$

où F est la fonction de répartition de la loi de densité f . On peut ainsi conclure que

$$\mathbb{P}(X \leq x, T = n) = \left(1 - \frac{1}{m}\right)^{n-1} \frac{F(x)}{m}.$$

Il ne reste plus qu'à étudier les lois marginales. D'une part, par continuité monotone croissante,

$$\mathbb{P}(T = n) = \lim_{q \rightarrow \infty} \mathbb{P}(X \in] - \infty, q], T = n),$$

ce qui donne

$$\begin{aligned} \mathbb{P}(T = n) &= \lim_{q \rightarrow \infty} \left(1 - \frac{1}{m}\right)^{n-1} \frac{F(q)}{m} \\ &= \left(1 - \frac{1}{m}\right)^{n-1} \frac{1}{m}. \end{aligned}$$

On en déduit que T suit une loi géométrique de paramètre $1/m$. D'autre part, par σ -additivité,

$$\begin{aligned} \mathbb{P}(X \leq x) &= \mathbb{P}(X \leq x, T \in \mathbb{N}^*) \\ &= \sum_{n=1}^{\infty} \mathbb{P}(X \leq x, T = n), \end{aligned}$$

ce qui donne

$$\begin{aligned} \mathbb{P}(X \leq x) &= \sum_{n=1}^{\infty} \left(1 - \frac{1}{m}\right)^{n-1} \frac{F(x)}{m} \\ &= \frac{1}{1 - (1 - 1/m)} \frac{F(x)}{m} \\ &= F(x), \end{aligned}$$

ce qui prouve que X a pour loi F .

Enfin, la loi du couple (X, T) est égale au produit des lois

$$\begin{aligned} \mathbb{P}(X \leq x, T = n) &= \left(1 - \frac{1}{m}\right)^{n-1} \frac{F(x)}{m} \\ &= \mathbb{P}(T = n) \mathbb{P}(X \leq x), \end{aligned}$$

□

```
def accept_reject(n, f, g, g_sampler, m):
    """
    n: nombre de simulations
    f: densité cible
    g: densité des propositions, g_sampler: simulateur selon g
    m: constante pour la majoration
    """
    x_samples = np.zeros(n)
    u_samples = np.zeros(n)
    accepted = np.zeros(n)
    n_accepted = 0
    while n_accepted < n:
        x = g_sampler()
        u = np.random.uniform()
        alpha = u * m * g(x)
        u_samples[n_accepted] = alpha
        x_samples[n_accepted] = x
        if alpha <= f(x):
            accepted[n_accepted] = 1
        n_accepted += 1
    return x_samples, u_samples, accepted
```

i En pratique...

On simule U_1 et Y_1 . Si $U_1 \leq r(Y_1)$ c'est gagné, on pose $X = Y_1$. Sinon, on simule U_2 et Y_2 et on teste à nouveau l'inégalité $U_2 \leq r(Y_2)$. Et ainsi de suite. Comme T suit une loi géométrique de paramètre $1/m$, son espérance vaut m : il faut en moyenne m tentatives pour obtenir une simulation de la loi de densité f . L'objectif est alors de choisir un couple (g, m) de sorte que m soit le plus proche possible de 1.

Exemple 4.4 (Rejet d'une loi polynomiale). Donnons un exemple jouet (on étudiera des exemples plus pertinents en TD). On considère la densité $f(x) = 4x^3 \mathbb{1}_{[0,1]}(x)$. Comme f est majorée par 4, on peut choisir pour g la densité de la loi uniforme sur $[0, 1]$ et $m = 4$. Alors, $r(x) = f(x)/(mg(x)) = x^3$, pour $x \in [0, 1]$. On simule donc (Y_1, U_1) et on teste si $U_1 \leq Y_1^3$, etc.

Bien évidemment, on privilégiera ici une simulation via F^{\leftarrow} qui permet de générer des variables aléatoires de loi f plus rapidement.

Nous pouvons facilement améliorer la proportion de point acceptés en proposant par exemple g définie par $g(x) = 2x \mathbb{1}_{[0,1]}(x)$, et $m = 2$.

Unable to display output for mime type(s): text/html

Figure 4.4: Visualisation des zones d'acceptations/rejet (g uniforme)

Unable to display output for mime type(s): text/html

Figure 4.5: Visualisation des zones d'acceptations/rejet (g triangulaire)

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

```
//| echo: false
md`Le taux d'acceptation est passé de  $\frac{1}{2}$  à  $\frac{1}{3}$  en utilisant une loi triangulaire au lieu d'une loi uniforme.`
```

Example 4.5 (Rejet d'une loi de densité d'Andrews). Considérons la densité d'Andrews définie par $f(x) = \frac{1}{S} \frac{\sin(\pi \cdot x)}{\pi \cdot x} \mathbb{1}_{[-1,1]}(x)$, avec $S = \int_{-1}^1 \frac{\sin(\pi \cdot x)}{\pi \cdot x} dx$. Dans ce contexte, on ne connaît pas la valeur exacte de S , et on va donc utiliser la méthode de rejet pour simuler des variables aléatoires de loi f sans cette information. On peut l'adapter le test de la manière suivante: si l'on prend $m = 2/S$ et $g(x) = \frac{1}{2} \mathbb{1}_{[-1,1]}(x)$, on observe que tester $u \leq \frac{f(x)}{m \cdot g(x)}$ est équivalent à tester $u \leq r(x) = \frac{1}{S} \frac{\sin(\pi \cdot x)}{\pi \cdot x} \cdot \frac{1}{\frac{1}{2} g(x)} = \frac{\sin(\pi \cdot x)}{\pi \cdot x} \cdot \frac{1}{2 \cdot g(x)}$, ce qui peut se faire sans connaissance de S . De plus on peut vérifier que $g(x) = \frac{1}{2} \mathbb{1}_{[-1,1]}(x)$ définit une densité et que $f(x) \leq m \cdot g(x)$ pour tout $x \in \mathbb{R}$.

```
n = 10000
g = lambda x: np.ones_like(x) / 2
g_sampler = lambda: 2 * np.random.uniform() - 1
m = 2

x_samples, u_samples, accepted = accept_reject(n, np.sinc, g, g_sampler, m)
ratio = np.sum(accepted) / n
# Note: https://stackoverflow.com/questions/70804891/how-to-vectorize-a-function-in-python-t
```

<IPython.core.display.HTML object>

On peut approcher numériquement la valeur exacte de S en utilisant une méthode de calcul approchée, ce qui permet de comparer ici notre méthode de rejet avec la densité sous-jacente:


```
from scipy import integrate
S = integrate.quad(np.sinc, -1, 1)[0]
print(f"En utilisant la méthode de rejet, on trouve que S = {S:.3f}")
```

En utilisant la méthode de rejet, on trouve que $S = 1.179$

Enfin, on peut visualiser la qualité l'approximation de la densité par la méthode de rejet en comparant la densité approchée (avec un histogramme) avec la densité exacte:

Unable to display output for mime type(s): text/html

Figure 4.6: Méthode de rejet pour simuler une loi de densité de type Andrews, sans connaissance de la valeur exacte de la constante de normalisation.

4.5.1 Cas multidimensionnel

Commençons par un cas de dimension deux.

Pour cela on va utiliser la méthode de rejet pour simuler une loi de densité f sur \mathbb{R}^2 . En particulier, un exemple classique est de tirer des points dans le disque unité, c'est-à-dire de simuler une loi uniforme sur le disque unité. Pour cela, on va utiliser la méthode de rejet avec g la densité de la loi uniforme sur le carré $[-1, 1]^2$.

Mais prenons un autre exemple, à savoir tirer des points uniformément dans la surface délimitée par une cardioïde. Pour cela, on va utiliser la méthode de rejet avec g la densité de la loi uniforme sur le carré $[-2, 2]^2$.

Unable to display output for mime type(s): text/html

Figure 4.7: Méthode de rejet pour simuler une loi uniforme sur un disque unité.

Aire estimée: 3.144

Unable to display output for mime type(s): text/html

Figure 4.8: Méthode de rejet pour simuler une loi uniforme sur une surface délimitée par une cardioïde.

Aire estimée: 4.745

EXERCICE loi uniforme sur un cylindre

Proposer une méthode pour simuler une loi uniforme sur un cylindre de rayon 1 et de hauteur 10.

4.6 Autres méthodes

4.6.1 Sommation de variables aléatoires

Pour simuler une variable aléatoire de loi binomiale $\mathcal{B}(n, p)$, on peut utiliser la méthode d'inversion. Cependant, cela nécessite le calcul de l'inverse généralisée de F , donc de coefficients binomiaux et de puissances de p et $1 - p$. À la place, on utilisera plutôt la relation bien connue suivante : si X_1, \dots, X_n est une suite iid de variables aléatoires de loi de Bernoulli de paramètre p , alors

$$X = X_1 + \dots + X_n \sim \mathcal{B}(n, p).$$

Pour simuler des variables aléatoires de Bernoulli, on utilise la méthode d'inversion (voir Exemple ??). Ainsi, si U_1, \dots, U_n sont des variables aléatoires iid de loi uniforme sur $[0, 1]$, alors

$$\sum_{i=1}^n \mathbb{1}_{\{U_i \leq p\}} \sim \mathcal{B}(n, p).$$

4.6.2 Loi de Poisson

Rappelons qu'une variable aléatoire X suit une loi de Poisson de paramètre $\lambda > 0$, notée $X \sim \mathcal{P}(\lambda)$ si

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}.$$

Une méthode pour simuler une variable aléatoire de loi de Poisson est donnée par la proposition suivante.

Proposition 4.3 (Génération de v.a. de loi de Poisson). *Soit $(E_n)_{n \geq 1}$ des variables aléatoires i.i.d. de loi exponentielle de paramètre $\lambda > 0$. On pose $S_k = E_1 + \dots + E_k$. Alors, pour tout $n \in \mathbb{N}$*

$$\mathbb{P}(S_n \leq 1 < S_{n+1}) = e^{-\lambda} \frac{\lambda^n}{n!}.$$

Ainsi, la variable aléatoire T définie par

$$T \triangleq \sup\{n \in \mathbb{N} : S_n \leq 1\}$$

suit une loi de Poisson de paramètre λ : $T \sim \mathcal{P}(\lambda)$.

La preuve repose sur le lemme suivant.

Lemma 4.1 (Loi de Erlang). Soit n variables aléatoires E_1, \dots, E_n i.i.d. de loi exponentielle de paramètre $\lambda > 0$. La somme $E_1 + \dots + E_n$ suit une loi d'Erlang de paramètres (n, λ) , donnée par la fonction de répartition

$$F_{n,\lambda}(t) = 1 - \sum_{k=0}^{n-1} e^{-\lambda t} \frac{(\lambda t)^k}{k!}.$$

Proof. On montre le résultat pour $n = 2$. La généralisation à k quelconque se fait par récurrence. Soit $t > 0$, et $f_\lambda(x) = \mathbb{1}_{\{x \geq 0\}} \lambda e^{-\lambda x}$ la densité d'une loi exponentielle de paramètre λ . Les variables aléatoires E_1 et E_2 étant indépendantes et suivant des lois exponentielles de paramètre λ_1 et λ_2 , on a

$$\begin{aligned} \mathbb{P}(E_1 + E_2 \leq t) &= \int_{\mathbb{R}^2} \mathbb{1}_{\{x_1+x_2 \leq t\}} f_\lambda(x_1) f_\lambda(x_2) dx_1 dx_2 \\ &= \int_{\mathbb{R}^2} \mathbb{1}_{\{x_1+x_2 \leq t\}} \lambda^2 e^{-\lambda(x_1+x_2)} \mathbb{1}_{\{x_1 \geq 0\}} \mathbb{1}_{\{x_2 \geq 0\}} dx_1 dx_2 \\ &= \int_{\mathbb{R}^2} \mathbb{1}_{\{0 \leq x_1 \leq t\}} \mathbb{1}_{\{0 \leq x_2 \leq t-x_1\}} \lambda^2 e^{-\lambda x_1} e^{-\lambda x_2} dx_1 dx_2 \\ &= \int_0^t \lambda e^{-\lambda x_1} \left(\int_0^{t-x_1} \lambda e^{-\lambda x_2} dx_2 \right) dx_1. \end{aligned}$$

La première intégrale se calcule alors facilement :

$$\int_0^{t-x_1} \lambda e^{-\lambda x_2} dx_2 = 1 - e^{-\lambda(t-x_1)}.$$

On obtient alors

$$\begin{aligned} \mathbb{P}(E_1 + E_2 \leq t) &= \int_0^t \lambda e^{-\lambda x_1} dx_1 - \int_0^t e^{-\lambda t} dx_1 \\ &= 1 - e^{-\lambda t} - \lambda t e^{-\lambda t}. \end{aligned}$$

Si $t < 0$, alors comme les E_i ne prennent que des valeurs positives on trouve $\mathbb{P}(E_1 + E_2 \leq t) = 0$. Ceci prouve le résultat pour $n = 2$. \square

On peut désormais prouver le résultat de la Proposition 4.3.

Proof. Pour $n \in \mathbb{N}$, on décompose la probabilité $\mathbb{P}(S_n \leq 1 < S_{n+1})$ via

$$\begin{aligned}\mathbb{P}(S_n \leq 1 < S_{n+1}) &= \mathbb{P}(\{S_n \leq 1\} \cap \{S_{n+1} \leq 1\}) \\ &= \mathbb{P}(S_n \leq 1) - \mathbb{P}(S_{n+1} \leq 1).\end{aligned}$$

Le lemme précédent donne

$$\mathbb{P}(S_n \leq 1) = 1 - \sum_{k=0}^{n-1} e^{-\lambda} \frac{\lambda^k}{k!}$$

et

$$\mathbb{P}(S_{n+1} \leq 1) = 1 - \sum_{k=0}^n e^{-\lambda} \frac{\lambda^k}{k!}.$$

On obtient alors le résultat souhaité :

$$\mathbb{P}(S_n \leq 1 < S_{n+1}) = e^{-\lambda} \frac{\lambda^n}{n!}.$$

On conclut la preuve de la proposition en remarquant que

$$\mathbb{P}(T = n) = \mathbb{P}(S_n \leq 1 < S_{n+1}).$$

\square

La simulation d'une variable aléatoire de Poisson repose donc sur la simulation de lois exponentielles qui se fait via la méthode d'inversion, comme vu dans Exemple 4.1. En pratique, on simule E_1 et on teste si $E_1 > 1$. Si oui, on pose alors $T = 0$. Si non, on simule E_2 et on teste si $E_1 + E_2 > 1$. Si oui, on pose $T = 1$. Sinon on continue la procédure.

4.6.3 Bibliographie et pour aller plus loin

- [Generating Random Floating-Point Numbers by Dividing Integers: a Case Study](#) par Frédéric Goualard
- [Generating Pseudo-random Floating-Point Values](#) par Allen Downey.

5 Loi normale: cas univarié

5.1 Définitions et propriétés de la loi normale

On rappelle que la loi normale de paramètres $\mu \in \mathbb{R}$ et $\nu > 0$ a une densité donnée pour tout $x \in \mathbb{R}$ par

$$\varphi_{\mu,\nu}(x) = \frac{1}{\sqrt{2\pi\nu}} \exp\left(-\frac{(x-\mu)^2}{2\nu}\right).$$

On note $X \sim \mathcal{N}(\mu, \nu)$, si X est une variable aléatoire ayant pour densité $\varphi_{\mu,\nu}$. Notons que si $X \sim \mathcal{N}(\mu, \nu)$, alors X a pour **espérance** μ et pour **variance** ν . Le cas particulier $\mu = 0$ et $\nu = 1$ correspond à une variable aléatoire dite **centrée réduite**.

On parle aussi souvent de loi gaussienne, en hommage au mathématicien [Carl Friedrich Gauss, le prince des mathématiciens](#)¹.

La loi normale vérifie la propriété de **stabilité par transformation affine** : si $X \sim \mathcal{N}(\mu, \nu)$ et si $(a, b) \in \mathbb{R}^* \times \mathbb{R}$, alors la variable aléatoire $aX + b$ suit une loi normale $\mathcal{N}(a\mu + b, a^2\nu)$. On peut donc facilement passer d'une loi normale centrée réduite à une loi normale quelconque via une transformation affine :

- si $X \sim \mathcal{N}(0, 1)$, alors $\sqrt{\nu}X + \mu \sim \mathcal{N}(\mu, \nu)$,
- si $X \sim \mathcal{N}(\mu, \nu)$, alors $(X - \mu)/\sqrt{\nu} \sim \mathcal{N}(0, 1)$.

Ainsi, savoir simuler une loi normale centrée réduite, permet de simuler n'importe quelle loi normale.

Proposition 5.1 (Fonction caractéristique de la loi normale). *La fonction caractéristique d'une variable aléatoire $X \sim \mathcal{N}(\mu, \nu)$ est donnée pour tout $t \in \mathbb{R}$ par*

$$\begin{aligned}\phi_{\mu,\nu}(t) &\triangleq \mathbb{E}(e^{itX}) \\ &= \exp\left(i\mu t - \frac{\nu t^2}{2}\right).\end{aligned}$$

¹[Carl Friedrich Gauss](#): (1777-1855) mathématicien, astronome et physicien né à Brunswick, directeur de l'observatoire de [Göttingen](#) de 1807 jusqu'à sa mort en 1855

Proof. on remarque d'abord que si $X \sim \mathcal{N}(0, 1)$ alors pour tout $z \in \mathbb{R}$, on a

$$\begin{aligned}\mathbb{E}[e^{zX}] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} e^{zx} dx \\ &= \frac{e^{\frac{1}{2}z^2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-z)^2} dx \\ &= \frac{e^{\frac{1}{2}z^2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy \\ &= e^{\frac{1}{2}z^2}.\end{aligned}$$

En utilisant le théorème de prolongement analytique (voir par exemple (Théorème I.10., Queffélec and Zuily 2013)) on peut donc étendre cette formule à tout $z \in \mathbb{C}$, et particulier au cas $z = it$ pour $t \in \mathbb{R}$. On obtient alors $\phi_{\mu,\nu}(t) = e^{-\frac{t^2}{2}}$. Enfin, on utilise la linéarité de l'espérance pour obtenir le résultat pour $X \sim \mathcal{N}(\mu, \nu)$. En effet, si $X \sim \mathcal{N}(0, 1)$, alors $X = \mu + \sqrt{\nu}Z$ avec $Z \sim \mathcal{N}(0, 1)$, et donc $\phi_{\mu,\nu}(t) = e^{i\mu t} \phi_{0,1}(\sqrt{\nu}t) = e^{i\mu t - \frac{\nu t^2}{2}}$. \square

5.1.1 Simulation d'une loi normale

i Une mauvaise piste pour simuler une loi normale

On peut simuler une loi normale à partir de variables aléatoires uniformes U_1, \dots, U_n iid en appliquant le théorème central limite à

$$\frac{U_1 + \dots + U_n - n/2}{\sqrt{n/12}}.$$

Cependant, cette méthode ne donne qu'une approximation d'une loi normale. Par ailleurs, la vitesse de convergence étant relativement lente (de l'ordre de \sqrt{n}), il faudra simuler beaucoup de variables aléatoires uniformes pour avoir une approximation correcte, ce qui demande un temps de calcul assez élevé.

5.1.1.1 Changement de variables

Le théorème suivant permet de passer de la loi d'un couple (X, Y) à celle de $(U, V) = \phi(X, Y)$, où ϕ est un C^1 -difféomorphisme, c'est-à-dire une application bijective dont la

réciroque est également de classe C^1 .

Pour cela rappelons que la **jacobienne** de ϕ^{-1} correspond à la matrice (application linéaire) des dérivées partielles. Ainsi, si $\phi(x, y) = (u, v) \iff (x, y) = \phi^{-1}(u, v)$, alors

$$J_{\phi^{-1}} : (u, v) \mapsto \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} .$$

Theorem 5.1 (Caractérisation de la loi d'une variable aléatoire réelle). *Soit (X, Y) un vecteur aléatoire de densité $f_{(X,Y)}$ définie sur l'ouvert $A \subset \mathbb{R}^2$ et $\phi : A \rightarrow B \subset \mathbb{R}^2$ un C^1 -difféomorphisme. Le vecteur aléatoire $(U, V) = \phi(X, Y)$ admet alors pour densité $f_{(U,V)}$ définie sur B pour tout $(u, v) \in \mathbb{R}^2$ par*

$$(u, v) \mapsto f_{(X,Y)}(\phi^{-1}(u, v)) |\det(J_{\phi^{-1}}(u, v))| \mathbb{1}_B(u, v) .$$

On a énoncé le résultat en dimension 2 par simplicité. Il s'étend bien évidemment à une dimension d quelconque. En particulier, pour $d = 1$, on retrouve le changement de variable classique dans le cas de l'intégration d'une fonction à valeurs réelles.

Proof. On rappelle que la loi de (U, V) est caractérisée par les quantités $\mathbb{E}[h(U, V)]$ pour tout $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ mesurable bornée. On considère donc une telle fonction h et on applique la formule de transfert :

$$\begin{aligned} \mathbb{E}[h(U, V)] &= \mathbb{E}[h(\phi(X, Y))] \\ &= \int_{\mathbb{R}^2} h(\phi(x, y)) f_{(X,Y)}(x, y) dx dy \\ &= \int_A h(\phi(x, y)) f_{(X,Y)}(x, y) dx dy . \end{aligned}$$

On applique alors la formule du changement de variables vu en théorie de l'intégration avec $(u, v) = \phi(x, y) \iff \phi^{-1}(u, v) = (x, y)$:

$$\begin{aligned}
& \mathbb{E}[h(U, V)] \\
&= \int_B h(u, v) f_{(X, Y)}(\phi^{-1}(u, v)) |\det(J_{\phi^{-1}}(u, v))| du dv \\
&= \int_{\mathbb{R}^2} h(u, v) f_{(X, Y)}(\phi^{-1}(u, v)) |\det(J_{\phi^{-1}}(u, v))| \mathbb{1}_B(u, v) du dv.
\end{aligned}$$

ce qui donne le résultat voulu. \square

Exemple 5.1 (Exemple : loi de $\cos(X)$). Donnons un exemple dans le cas réel. On considère une variable aléatoire X de loi uniforme sur $]0, \pi[$. Sa densité est donnée par $f_X(x) = \mathbb{1}_{]0, \pi[}(x)/\pi$. On pose $U = \cos(X)$ et on souhaite déterminer la loi de U .

On applique le théorème précédent avec la fonction $\phi^{-1}(u) = \arccos(u)$ sur $] -1, 1[$. La densité de U est alors donnée pour tout $u \in \mathbb{R}$ par

$$\begin{aligned}
f_U(u) &= \frac{\mathbb{1}_{]0, \pi[}(\arccos(u))}{\pi} \left| \frac{-1}{\sqrt{1-u^2}} \right| \mathbb{1}_{]-1, 1[}(u) \\
&= \frac{1}{\pi \sqrt{1-u^2}} \mathbb{1}_{]-1, 1[}(u).
\end{aligned}$$

5.2 Méthode de Box-Müller

Un cas particulier fondamental de la formule de changement de variables concerne le passage en coordonnées polaires. Cette transformation est définie via l'application

$$\begin{aligned}
\phi^{-1} :]0, \infty[\times]0, 2\pi[&\rightarrow \mathbb{R}^2 \setminus ([0, \infty[\times \{0\}) \\
\begin{pmatrix} r \\ \theta \end{pmatrix} &\mapsto \begin{pmatrix} r \cos(\theta) \\ r \sin(\theta) \end{pmatrix}.
\end{aligned}$$

L'expression de ϕ ne nous sera pas utile. On peut tout de même la donner au passage :

$$\begin{aligned}
\phi : \mathbb{R}^2 \setminus ([0, \infty[\times \{0\}) &\rightarrow]0, \infty[\times]0, 2\pi[\\
\begin{pmatrix} x \\ y \end{pmatrix} &\mapsto \begin{pmatrix} \sqrt{x^2 + y^2} \\ 2 \arctan\left(\frac{y}{x + \sqrt{x^2 + y^2}}\right) \end{pmatrix}.
\end{aligned}$$

Ici, le jacobien de ϕ^{-1} est la matrice

$$J_{\phi^{-1}}(r, \theta) = \begin{pmatrix} \cos(\theta) & -r \sin(\theta) \\ \sin(\theta) & r \cos(\theta) \end{pmatrix},$$

qui vérifie $|\det(J_{\phi^{-1}}(r, \theta))| = r$. Ainsi, si (X, Y) a pour densité $f_{(X,Y)}$, alors $(R, \Theta) = \phi(X, Y)$ a pour densité

$$f_{(R,\Theta)}(r, \theta) = r \cdot f_{(X,Y)}(r \cos(\theta), r \sin(\theta)) \cdot \mathbb{1}_{]0, \infty[}(r) \mathbb{1}_{]0, 2\pi[}(\theta).$$

Dans le cas où X et Y sont des variables aléatoires gaussiennes indépendantes, on obtient le résultat suivant.

Theorem 5.2 (Méthode de Box-Müller). *Soit X et Y deux variables aléatoires indépendantes de loi normales centrées réduites : $X, Y \sim \mathcal{N}(0, 1)$. Le couple de variables aléatoires polaires $(R, \Theta) = \phi(X, Y)$ a pour densité*

$$f_{R,\Theta}(r, \theta) = \left(r \cdot e^{-\frac{r^2}{2}} \mathbb{1}_{]0, \infty[}(r) \right) \left(\frac{\mathbb{1}_{]0, 2\pi[}(\theta)}{2\pi} \right).$$

Autrement dit, elles sont indépendantes, l'angle Θ suit une loi uniforme sur $]0, 2\pi[$ et la distance à l'origine R suit une loi de Rayleigh donnée par la densité

$$f_R(r) = r \cdot e^{-r^2/2} \mathbb{1}_{]0, \infty[}(r), \quad r > 0.$$

Proof. La densité du couple (X, Y) est donnée par

$$f_{(X,Y)}(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}, \quad x, y \in \mathbb{R}.$$

Le théorème précédent donne alors la densité de (R, Θ) :

$$\begin{aligned} f_{(R,\Theta)}(r, \theta) &= r \cdot f_{(X,Y)}(r \cos(\theta), r \sin(\theta)) \cdot \mathbb{1}_{]0, \infty[}(r) \mathbb{1}_{]0, 2\pi[}(\theta) \\ &= r \cdot \frac{1}{2\pi} e^{-\frac{r^2}{2}} \cdot \mathbb{1}_{]0, \infty[}(r) \mathbb{1}_{]0, 2\pi[}(\theta), \end{aligned}$$

ce qui conclut la preuve. □

Si R suit une loi de Rayleigh alors $\sqrt{-2 \log(U)}$ a la même loi que R , où U est une variable aléatoire de loi uniforme sur $]0, 1[$. Pour cela il suffit de remarquer que pour tout $x > 0$, $F_R(x) = \mathbb{P}(R \leq x) = 1 - \exp(-\frac{x^2}{2})$, et donc que pour tout $q \in]0, 1[$, $F_R^{\leftarrow}(q) = \sqrt{-2 \log(1 - q)}$ et donc $\sqrt{-2 \log(1 - U)}$, puis $\sqrt{-2 \log(U)}$ on donc la même loi que R .

L'algorithme de Box-Müller s'en suit: si U et V sont des v.a. indépendantes de loi uniforme sur $[0, 1]$ et qu'on définit X et Y par

$$\begin{cases} X = \sqrt{-2 \log(U)} \cos(2\pi V) \\ Y = \sqrt{-2 \log(U)} \sin(2\pi V) . \end{cases}$$

alors X et Y des variables aléatoires gaussiennes centrées réduites indépendantes.

Note

Cet algorithme n'est en fait pas souvent utilisé en pratique : il fait appel à des fonctions dont l'évaluation est coûteuse (logarithme, cosinus, sinus). Pour s'affranchir des fonctions trigonométriques, une version modifiée de l'algorithme de Box-Müller a été proposée : la méthode de Marsaglia, qui s'appuie sur des variables aléatoires uniformes sur le disque unité (voir l'exercice dédié en TD). Une autre alternative est la méthode de Ziggurat.

5.3 Lois autour de la loi normale

5.3.1 Loi du χ^2

Concernant la prononciation, on prononce "khi-deux" le nom de cette loi.

Definition 5.1 (Loi du χ^2). Soit X_1, \dots, X_k des variables aléatoires i.i.d. de loi normale centrée réduite. La loi de la variable aléatoire $X = X_1^2 + \dots + X_k^2$ est appelée **loi du χ^2 à k degrés de liberté**. Sa densité est donnée par

$$f(x) = \frac{1}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-x/2}, \quad x \geq 0,$$

où Γ désigne la fonction gamma d'Euler :

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt .$$

On note alors $X \sim \chi^2(k)$.

Au vu de sa définition, la simulation d'une loi du χ^2 est claire : on simule k variables aléatoires gaussiennes centrées réduites indépendantes et on somme leur carrés.

Proof. Montrons pour $k = 1$ que la densité est bien de la forme précédente, c'est-à-dire

$$f(x) = \frac{1}{\sqrt{2\pi}} \frac{e^{-x/2}}{\sqrt{x}}, \quad x \geq 0,$$

où on a utilisé la relation $\Gamma(1/2) = \sqrt{\pi}$ (intégrale de Gauss).

Soit $h : \mathbb{R} \rightarrow \mathbb{R}$ une fonction mesurable bornée. On a

$$\begin{aligned} \mathbb{E}[h(X_1^2)] &= \int_{\mathbb{R}} h(x^2) \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx \\ &= \int_{-\infty}^0 h(x^2) \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx + \int_0^{\infty} h(x^2) \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx. \end{aligned}$$

En effectuant le changement de variable $x = -\sqrt{u}$ dans la première intégrale et $x = \sqrt{u}$ dans la deuxième, on obtient

$$\mathbb{E}[h(X_1^2)] = \int_{\infty}^0 h(u) \frac{e^{-u/2}}{\sqrt{2\pi}} \frac{du}{2\sqrt{u}} + \int_0^{\infty} h(u) \frac{e^{-u/2}}{\sqrt{2\pi}} \frac{du}{2\sqrt{u}}.$$

Les deux intégrales étant égales, on conclut que

$$\mathbb{E}[h(X_1^2)] = \int_0^{\infty} h(u) \frac{e^{-u/2}}{\sqrt{2\pi}\sqrt{u}} du,$$

ce qui prouve le résultat pour $k = 1$.

La généralisation à k quelconque se fait par récurrence: on utilise la formule de convolution de la loi pour obtenir la loi pour $k + 1$:

$$\begin{aligned} X_1^2 + \dots + X_k^2 + X_{k+1}^2 &= (X_1^2 + \dots + X_k^2) + X_{k+1}^2 \\ &= \chi^2(k) + X_{k+1}^2. \end{aligned}$$

Ainsi,

$$\begin{aligned}
f_{\chi^2(k+1)}(x) &= \int_0^\infty f_{\chi^2(k)}(x-y) f_{X_{k+1}^2}(y) dy \\
&= \int_0^x \frac{1}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} (x-y)^{\frac{k}{2}-1} e^{-\frac{x-y}{2}} \frac{e^{-\frac{y}{2}}}{\sqrt{2\pi y}} dy \\
&= \frac{e^{-\frac{x}{2}}}{2^{\frac{k+1}{2}} \Gamma(\frac{k}{2}) \Gamma(\frac{1}{2})} \int_0^x (x-y)^{\frac{k}{2}-1} \frac{1}{\sqrt{y}} dy \\
&= \frac{e^{-\frac{x}{2}}}{2^{\frac{k+1}{2}} \Gamma(\frac{k}{2}) \Gamma(\frac{1}{2})} x \int_0^1 (x-ux)^{\frac{k}{2}-1} \frac{1}{\sqrt{xu}} du
\end{aligned}$$

avec le changement de variable $y = ux$. Ensuite,

$$f_{\chi^2(k+1)}(x) = \frac{x^{\frac{k+1}{2}} e^{-\frac{x}{2}}}{2^{\frac{k+1}{2}} \Gamma(\frac{k}{2}) \Gamma(\frac{1}{2})} \int_0^1 (1-u)^{\frac{k}{2}-1} u^{1/2-1} du .$$

Or rappelons que si $B(a, b) = \int_0^1 (1-u)^{a-1} u^{b-1} du$, alors pour tout $a, b \in [0, +\infty[$, $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$. En effet, en faisant le changement de variable dans l'intégrale double qui suit:

$$\begin{aligned}
\psi : \mathbb{R}^+ \times \mathbb{R}^+ &\rightarrow \mathbb{R}^+ \times]0, 1[\\
(s, t) &\mapsto \left(s+t, \frac{t}{s+t} \right),
\end{aligned}$$

c'est-à-dire $\psi^{-1}(r, w) = (r(1-w), rw)$, et le jacobien est donné par $J_{\psi^{-1}}(r, w) = \begin{pmatrix} 1-w & -r \\ w & r \end{pmatrix}$, et donc $J_{\psi^{-1}}(r, w) = r$, on obtient:

$$\begin{aligned}
\Gamma(a)\Gamma(b) &= \int_0^\infty t^{a-1} e^{-t} dt \int_0^\infty s^{b-1} e^{-s} ds \\
&= \int_0^\infty \int_0^\infty e^{-t-s} t^{a-1} s^{b-1} dt ds \\
&= \int_0^1 \int_0^\infty e^{-r} (rw)^{a-1} (r(1-w))^{b-1} r dr dw \\
&= \int_0^1 w^{a-1} (1-w)^{b-1} \int_0^\infty e^{-r} r^{a+b-1} dr dw
\end{aligned}$$

et donc $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$. En appliquant cette relation pour $a = \frac{k}{2}$ et $b = 1/2$, on obtient

$$f_{\chi^2(k+1)}(x) = \frac{x^{\frac{k+1}{2}} e^{-\frac{x}{2}}}{2^{\frac{k+1}{2}} \Gamma(\frac{k+1}{2})}.$$

Le résultat est donc prouvé par récurrence. □

5.3.2 Loi de Student

Définition 5.2 (Loi de Student). Soit $X \sim \mathcal{N}(0, 1)$ et $Y \sim \chi^2(k)$ deux variables aléatoires indépendantes. La loi de la variable aléatoire $V = \frac{X}{\sqrt{Y/k}}$ est appelée **loi de Student à k degrés de liberté**. Elle admet pour densité

$$f_V(t) = \frac{1}{\sqrt{k\pi}} \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \left(1 + \frac{t^2}{k}\right)^{-\frac{k+1}{2}}, \quad t \in \mathbb{R}.$$

La loi de Student correspond donc au ratio d'une loi normale par la racine carrée d'une loi du $\chi^2(k)$ normalisée. Ce ratio apparaît souvent en statistique lors de la construction d'intervalles de confiance. Cette loi a été décrite en 1908 par [William Gosset](#)².

²[William Gosset](#): (1876-1937) statisticien et chimiste anglais. Il était employé à la brasserie Guinness à Dublin, chargé du contrôle qualité. Son employeur lui refusant le droit de publier sous son propre nom, W. Gosset choisit un pseudonyme, *Student* (🇫🇷: étudiant).

Au vu de la proposition précédente, simuler une loi de Student est assez simple : on simule $k + 1$ loi normales indépendantes X_1, \dots, X_{k+1} et on considère

$$V = \frac{\sqrt{k}X_{k+1}}{\sqrt{X_1^2 + \dots + X_k^2}}.$$

Formule de la densité. On applique pour cela la formule du changement de variables avec la transformation

$$\begin{aligned} \phi : \mathbb{R}^* \times]0, \infty[&\rightarrow \mathbb{R}^* \times \mathbb{R}^* \\ (x, y) &\mapsto \left(x, \frac{x}{\sqrt{y/k}}\right), \end{aligned}$$

c'est-à-dire

$$\phi^{-1}(u, v) = \left(u, k \frac{u^2}{v^2}\right).$$

La fonction ϕ^{-1} a pour matrice jacobienne

$$J_{\phi^{-1}}(u, v) = \begin{pmatrix} 1 & 0 \\ \frac{2k}{v^2} & -\frac{2ku^2}{v^3} \end{pmatrix},$$

dont le déterminant vaut $\frac{-2ku^2}{v^3}$. Par ailleurs, les variables aléatoires X et Y étant indépendantes, la densité du couple (X, Y) correspond au produit des densités :

$$f_{(X,Y)}(x, y) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \frac{1}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} y^{\frac{k}{2}-1} e^{-\frac{y}{2}} \mathbb{1}_{]0, \infty[}(y) \quad x, y \in \mathbb{R}.$$

Tout est prêt pour appliquer la théorème du changement de variables qui assure que la densité du couple (U, V) est donnée par

$$f_{(U,V)}(u, v) = \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} \frac{1}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} \left(\frac{ku^2}{v^2}\right)^{\frac{k}{2}-1} e^{-\frac{1}{2} \frac{ku^2}{v^2}} \frac{2ku^2}{(v^2)^{\frac{3}{2}}} \mathbb{1}_{\mathbb{R}^*}(v).$$

Il suffit alors de marginaliser pour obtenir la densité de V , ce qui s'effectue en calculant l'intégrale

$$\int_{\mathbb{R}} f_{(U,V)}(u, v) du.$$

Les termes en u de l'expression précédente s'intègre en

$$\begin{aligned}\int_{-\infty}^{\infty} e^{-\frac{u^2}{2}(1+\frac{k}{v^2})} (u^2)^{\frac{k}{2}} du &= \int_0^{\infty} e^{-s} \left(\frac{2s}{1+\frac{k}{v^2}} \right)^{\frac{k}{2}} \sqrt{\frac{2}{1+\frac{k}{v^2}}} \frac{ds}{2\sqrt{s}} \\ &= \frac{2^{\frac{k+1}{2}}}{2} \frac{1}{(1+\frac{k}{v^2})^{\frac{k+1}{2}}} \int_0^{\infty} e^{-s} s^{\frac{k}{2}-\frac{1}{2}} ds ,\end{aligned}$$

où la première égalité résulte du changement de variable

$$s = \frac{u^2}{2} \left(1 + \frac{k}{v^2} \right) \iff \sqrt{\frac{2s}{1+\frac{k}{v^2}}} = u .$$

On reconnaît dans l'intégrale la valeur de $\Gamma(\frac{k+1}{2})$ ce qui conduit à

$$f_V(v) = \frac{1}{\sqrt{2\pi}} \frac{1}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} \left(\frac{k}{v^2} \right)^{\frac{k}{2}-1} \frac{2k}{(v^2)^{\frac{3}{2}}} \frac{2^{\frac{k+1}{2}}}{2} \frac{\Gamma(\frac{k+1}{2})}{(1+\frac{k}{v^2})^{\frac{k+1}{2}}} .$$

On réécrit alors les termes en k/v^2 via

$$\begin{aligned}\left(\frac{k}{v^2} \right)^{\frac{k}{2}-1} \frac{k}{(v^2)^{\frac{3}{2}}} \frac{1}{(1+\frac{k}{v^2})^{\frac{k+1}{2}}} &= \left(\frac{k}{v^2} \right)^{\frac{k}{2}-1} \frac{1}{\sqrt{k}} \left(\frac{k}{v^2} \right)^{\frac{3}{2}} \frac{1}{(1+\frac{k}{v^2})^{\frac{k+1}{2}}} \\ &= \frac{1}{\sqrt{k}} \left(1 + \frac{v^2}{k} \right)^{-\frac{k+1}{2}} ,\end{aligned}$$

ce qui permet de conclure :

$$f_V(v) = \frac{1}{\sqrt{k\pi}} \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \left(1 + \frac{v^2}{k} \right)^{-\frac{k+1}{2}} .$$

□

5.3.3 Loi de Cauchy

Définition 5.3 (Loi de Cauchy). Une variable aléatoire X suit une **loi de Cauchy** standard si sa densité est donnée par

$$f_X(x) = \frac{1}{\pi(1+x^2)} , \quad x \in \mathbb{R} .$$

On note alors $X \sim \mathcal{C}(0, 1)$ dans ce cas. Plus généralement on dit que Y suit une loi de Cauchy de paramètres $(\mu, \sigma) \in \mathbb{R} \times]0, +\infty[$ si $Y = \mu + \sigma X$ où X suit une loi de Cauchy standard. Sa densité est alors donnée par

$$f_Y(y) = \frac{1}{\sigma\pi(1 + \frac{(y-\mu)^2}{\sigma^2})}, \quad y \in \mathbb{R}.$$

Pour rappel cette loi est importante comme exemple de loi qui n'admet ni espérance, ni variance a fortiori. En effet, si X suit une loi de Cauchy standard,

$$\int_{\mathbb{R}} |x| f_X(x) dx = \int_{\mathbb{R}} \frac{|x|}{\pi(1+x^2)} dx.$$

Or $\frac{|x|}{\pi(1+x^2)} \sim \frac{1}{\pi x}$ quand $|x| \rightarrow \infty$, et $x \mapsto \frac{|x|}{\pi(1+x^2)}$ n'est donc pas intégrable sur \mathbb{R} au sens de Lebesgue. C'est donc un exemple de loi pour laquelle la loi des grands nombres et le théorème central limite ne s'appliquent pas.

5.3.3.1 Fonction caractéristique

La fonction caractéristique de la loi de Cauchy standard est donnée par

$$\varphi_X(t) \triangleq \int_{\mathbb{R}} e^{itx} f_X(x) dx = e^{-|t|}.$$

et donc si $X \sim \mathcal{C}(\mu, \sigma)$, alors pour tout $t \in \mathbb{R}$, $\varphi_X(t) = e^{i\mu t - \sigma|t|}$. Pour la preuve voir par exemple (Exemple III.5.5., Barbe and Ledoux 2006) Une conséquence directe est que la somme de deux variables aléatoires indépendantes de loi de Cauchy reste une loi de Cauchy: Si $X_1 \sim \mathcal{C}(\mu_1, \sigma_1)$ et $X_2 \sim \mathcal{C}(\mu_2, \sigma_2)$ sont indépendantes, alors $X_1 + X_2 \sim \mathcal{C}(\mu_1 + \mu_2, \sigma_1 + \sigma_2)$ (car elles ont la même fonction caractéristique).

On en déduit que la moyenne de variables de Cauchy standard i.i.d suit la loi de Cauchy standard. En effet, si X_1, \dots, X_n sont i.i.d de loi de Cauchy standard, alors pour tout $t \in \mathbb{R}$, $\varphi_{\frac{1}{n} \sum_{i=1}^n X_i}(t) = e^{-|t|}$, et donc $\frac{1}{n} \sum_{i=1}^n X_i \sim \mathcal{C}(0, 1)$. Ainsi la moyenne empirique ne converge pas en probabilité vers une constante!

5.3.3.2 Fonction de répartition et simulation

La fonction de répartition de X correspond, à une constante près, à la fonction arctangente qui est bijective de \mathbb{R} sur $] -\frac{\pi}{2}, \frac{\pi}{2} [$. La méthode d'inversion permet donc de simuler une variable aléatoire de loi de Cauchy. La proposition suivante donne un autre moyen.

Proposition 5.2 (Loi de Cauchy et loi normale). *Soit X et Y deux variables aléatoires indépendantes de loi normale centrée réduite. Alors la variable aléatoire Y/X suit une loi de Cauchy.*

Notons que Y/X est bien définie puisque X est différent de 0 presque sûrement.

Proof. Comme pour la loi de Student, on démontre ce résultat avec un changement de variables. On considère l'application

$$\begin{aligned}\phi : \mathbb{R}^* \times \mathbb{R} &\rightarrow \mathbb{R}^2 \\ (x, y) &\mapsto \left(x, \frac{y}{x}\right),\end{aligned}$$

c'est-à-dire

$$\phi^{-1}(u, v) = (u, uv).$$

La matrice jacobienne de ϕ^{-1} est donnée par

$$J_{\phi^{-1}}(u, v) = \begin{pmatrix} 1 & 0 \\ v & u \end{pmatrix},$$

dont le déterminant vaut u . Rappelons également que la densité du couple (X, Y) vaut

$$f_{(X,Y)}(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}, \quad x, y \in \mathbb{R}.$$

La formule du changement de variables donne alors la densité de $f_{(U,V)}$:

$$f_{(U,V)} = \frac{1}{2\pi} e^{-\frac{u^2+u^2v^2}{2}} |u|.$$

On obtient alors la densité de V en intégrant par rapport à u :

$$\begin{aligned}
f_V(v) &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\frac{u^2+u^2v^2}{2}} |u| \, du \\
&= \frac{1}{\pi} \int_0^\infty e^{-\frac{u^2(1+v^2)}{2}} u \, du \\
&= \frac{1}{\pi} \left[-\frac{e^{-\frac{u^2(1+v^2)}{2}}}{1+v^2} \right]_0^\infty \\
&= \frac{1}{\pi(1+v^2)} .
\end{aligned}$$

□

On obtient ainsi une autre manière de simuler une loi de Cauchy, en prenant le ratio de deux gaussiennes indépendantes. Cependant, la simulation via la méthode d'inversion peut-être moins coûteuse puisqu'elle ne fait appel qu'à une variable aléatoire uniforme et à la fonction tangente.

5.4 Bibliographie

- Barbe, Philippe, and Michel Ledoux. 2006. *Probabilités*.
- Barbier, Emile. 1860. "Note Sur Le Problème de l'aiguille Et Le Jeu Du Joint Couvert." *Journal de Mathématiques Pures Et Appliquées* 5: 273–86. http://www.numdam.org/item/JMPA_1860_2_5_273_0.pdf.
- Borel, E., and R. Deltheil. 1923. *Probabilités, Erreurs*. Armand Colin. A. Colin.
- Ciarlet, P. G. 2006. *Introduction à l'analyse Numérique Matricielle Et à l'optimisation. Cours Et Exercices Corrigés*. Dunod.
- Courant, J., M. de Falco, S. Gonnord, J.-C. Filliâtre, S. Conchon, G. Dowek, and B. Wack. 2013. *Informatique Pour Tous En Classes Préparatoires Aux Grandes écoles: Manuel d'algorithmique Et Programmation Structurée Avec Python*. Eyrolles.
- Eckhardt, R. 1987. "Stan Ulam, John von Neumann, and the Monte Carlo Method." *Los Alamos Science*, no. 15: 131–37.
- Foata, D., and A. Fuchs. 1996. *Calcul Des Probabilités: Cours Et Exercices Corrigés*. Masson.
- Metropolis, Nicholas. 1987. "The Beginning of the Monte Carlo Method." *Los Alamos Science*, no. 15: 125–30.
- Ouvrard, J.-Y. 2007. *Probabilités : Tome 2, Licence - CAPES*. 2nd ed. Enseignement Des Mathématiques. Cassini.
- . 2008. *Probabilités : Tome 1, Licence - CAPES*. 2nd ed. Enseignement Des Mathématiques. Cassini.

- Queffélec, H., and C. Zuily. 2013. *Analyse Pour l'agrégation*. Dunod.
- Williams, D. 1991. *Probability with Martingales*. Cambridge Mathematical Textbooks. Cambridge: Cambridge University Press.

6 Loi normale: cas multivarié

On considère ici \mathbb{R}^d muni du produit scalaire euclidien $\langle \cdot, \cdot \rangle$ et de la norme euclidienne $\| \cdot \|$ associée.

6.1 Rappel sur les vecteurs aléatoires

Si $\mathbf{X} = (X_1, \dots, X_d) \in \mathbb{R}^d$ est un vecteur aléatoire, son espérance est définie coordonnée par coordonnée :

$$\mathbb{E}[\mathbf{X}] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d]) \in \mathbb{R}^d$$

quantité qui existe dès que chaque espérance est bien définie. De plus, si $\mathbb{E}[X_j^2] < \infty$ pour tout $j \in \llbracket 1, d \rrbracket$, on peut alors définir les covariances pour tout $(i, j) \in \llbracket 1, d \rrbracket^2$:

$$\text{cov}(X_i, X_j) = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])] ,$$

quantités que l'on rassemble dans une matrice appelée *matrice de variance-covariance* :

$$\Sigma = (\text{cov}(X_i, X_j))_{1 \leq i, j \leq d} \in \mathbb{R}^{d \times d} .$$

On peut montrer que cette matrice est symétrique et semi-définie positive. En particulier, si les X_j sont indépendants, alors Σ est une matrice diagonale.

Note

Un point important. Si (X, Y) est un vecteur aléatoire, il ne suffit pas de connaître les marginales X et Y pour caractériser entièrement le vecteur. Par exemple, si X et Y suivent toutes les deux une loi normale, alors on peut avoir par exemple $X = Y$, ou bien X indépendant de Y , et ces deux cas modélisent clairement deux vecteurs aléatoires de lois distinctes.

6.2 Vecteurs gaussiens

Definition 6.1 (Vecteur gaussien). Un vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_d)^\top \in \mathbb{R}^d$ est un *vecteur gaussien* si pour tout $\alpha = (\alpha_1, \dots, \alpha_d)^\top$, la variable aléatoire réelle

$$\langle \alpha, \mathbf{X} \rangle = \alpha_1 X_1 + \dots + \alpha_d X_d ,$$

suit une loi normale.

En particulier chaque (loi marginale) X_j suit alors une loi gaussienne (choisir ci-dessus $\alpha = e_j$, les autres égaux à 0).

Cependant, il ne suffit pas que les X_j soient des gaussiennes pour que le vecteur X soit un vecteur gaussien. Par exemple, si X suit une loi normale centrée réduite et ε une loi uniforme (discrète) sur $\{-1, 1\}$, alors on peut montrer que εX suit encore une loi normale centrée réduite. En effet, pour tout $t \in \mathbb{R}$, on a

$$\begin{aligned} \mathbb{P}(\varepsilon X \leq t) &= \mathbb{P}(X \leq t)\mathbb{P}(\varepsilon = 1) + \mathbb{P}(-X \leq t)\mathbb{P}(\varepsilon = -1) \\ &= \frac{1}{2}\mathbb{P}(X \leq t) + \frac{1}{2}\mathbb{P}(-X \leq t) = \mathbb{P}(X \leq t) . \end{aligned}$$

Cependant, $X + \varepsilon X$ prend la valeur 0 avec probabilité 1/2 donc ne suit pas une loi normale. Ainsi le vecteur $(X, \varepsilon X)^\top$ n'est pas un vecteur gaussien bi-dimensionnel.

Soit \mathbf{X} un vecteur gaussien. Notons μ son espérance et Σ sa matrice de variance-covariance. En reprenant les notations de la définition, la variable aléatoire $\langle \alpha, \mathbf{X} \rangle$ vérifie

$$\begin{aligned} \mathbb{E}[\langle \alpha, \mathbf{X} \rangle] &= \mathbb{E}[\alpha_1 X_1 + \dots + \alpha_d X_d] \\ &= \alpha_1 \mathbb{E}[X_1] + \dots + \alpha_d \mathbb{E}[X_d] = \langle \alpha, \mu \rangle , \end{aligned}$$

et

$$\begin{aligned} \text{var}(\langle \alpha, \mathbf{X} \rangle) &= \text{var}(\alpha_1 X_1 + \dots + \alpha_d X_d) \\ &= \text{cov}(\alpha_1 X_1 + \dots + \alpha_d X_d, \alpha_1 X_1 + \dots + \alpha_d X_d) \\ &= \sum_{1 \leq i, j \leq d} \alpha_i \text{cov}(X_i, X_j) \alpha_j = \alpha^\top \Sigma \alpha \end{aligned}$$

Ainsi, $\langle \alpha, \mathbf{X} \rangle \sim \mathcal{N}(\langle \alpha, \mu \rangle, \alpha^\top \Sigma \alpha)$.

6.2.1 Fonction caractéristique d'un vecteur gaussien

La fonction caractéristique d'un vecteur gaussien d'espérance μ et de matrice de variance-covariance Σ est donnée par

$$\phi_{\mathbf{X}}(\alpha) = \mathbb{E}[e^{i\langle \alpha, \mathbf{X} \rangle}] = \exp \left(i\langle \alpha, \mu \rangle - \frac{\alpha^\top \Sigma \alpha}{2} \right), \quad \alpha \in \mathbb{R}^d.$$

où on a utilisé l'expression de la fonction caractéristique d'une variable aléatoire de loi normale $\mathcal{N}(\langle \alpha, \mu \rangle, \alpha^\top \Sigma \alpha)$. Ainsi, $\phi_{\mathbf{X}}$ est entièrement déterminée par les quantités μ et Σ . Comme cette fonction caractérise la loi de \mathbf{X} , on en déduit que la loi d'un vecteur gaussien est entièrement caractérisée par son espérance et sa matrice de variance-covariance. On note alors

$$\mathbf{X} \sim \mathcal{N}(\mu, \Sigma).$$

En particulier, si les variables aléatoires X_1, \dots, X_d sont indépendantes de loi $\mathcal{N}(0, 1)$, alors $\mu = (0, \dots, 0)^\top$ et $\Sigma = \text{Id}_d$.

6.2.2 Densité de probabilité

Commençons par le cas $\mu = 0$ et $\Sigma = \text{Id}_d$ correspond à la loi gaussienne centrée réduite $\mathcal{N}(0, \text{Id}_d)$. La loi de $(X_1, \dots, X_n)^\top$ correspond alors à la loi produit de n lois gaussiennes centrées réduites indépendantes (pour les gaussiennes la décorrélation implique l'indépendance). Sa densité est donc donnée par

$$\varphi_{0, \text{Id}_d}(x) = \frac{1}{\sqrt{(2\pi)^d}} \exp \left(-\frac{1}{2} x^\top x \right).$$

Proposition 6.1 (Densité de la loi gaussienne multivariée). *Soient $\mu \in \mathbb{R}^d$ et $\Sigma \in \mathbb{R}^{d \times d}$ (symétrique et définie positive) et supposons que $X \sim \mathcal{N}(\mu, \Sigma)$. Alors la densité de probabilité de X est donnée pour tout $x \in \mathbb{R}^d$ par*

$$\varphi_{\mu, \Sigma}(x) = \frac{1}{\sqrt{(2\pi)^d |\det(\Sigma)|}} \exp \left(-\frac{1}{2} (x - \mu)^\top \Sigma^{-1} (x - \mu) \right).$$

Proof. Prenons une matrice $L \in \mathbb{R}^{d \times d}$ telle que $LL^\top = \Sigma$ (par exemple, la décomposition de Cholesky, que nous reverrons ci-dessous). Calculons alors la loi de $\mathbf{Y} = \psi(X) \triangleq L\mathbf{X} + \mu$, pour $X \sim \mathcal{N}(0, \text{Id}_d)$. Pour appliquer la formule du changement de variable nous donne, on calcule ψ^{-1} et son jacobien: $\psi^{-1}(y) = L^{-1}(y - \mu)$, et $\det(J_{\psi^{-1}}) = \det(L^{-1}) = \det(L)^{-1} = |\det(\Sigma)|^{-1/2}$.

En notant que $LL^\top (L^{-1})^\top L^{-1} = \text{Id}_d$, et donc que $(L^{-1})^\top L^{-1} = \Sigma^{-1}$, on en déduit que la densité de \mathbf{Y} est donnée par

$$\begin{aligned}\varphi_{\mu, \Sigma}(y) &= \varphi_{0, \text{Id}_d}(\psi^{-1}(y)) |\det(J_{\psi^{-1}})| \\ &= \frac{|\det(\Sigma)|^{-1/2}}{\sqrt{(2\pi)^d}} \exp\left(-\frac{1}{2}(y - \mu)^\top (L^{-1})^\top L^{-1}(y - \mu)\right) \\ &= \frac{1}{\sqrt{(2\pi)^d |\det(\Sigma)|}} \exp\left(-\frac{1}{2}(y - \mu)^\top \Sigma^{-1}(y - \mu)\right).\end{aligned}$$

□

Proposition 6.2 (Transformation affine de vecteurs gaussiens). *Soit $\mathbf{X} \sim \mathcal{N}(\mu, \Sigma)$ un vecteur gaussien sur \mathbb{R}^d , $\Omega \in \mathbb{R}^{d' \times d}$ et $\nu \in \mathbb{R}^{d'}$. Alors, le vecteur aléatoire $\mathbf{Y} = \Omega \mathbf{X} + \nu$ est un vecteur gaussien vérifiant*

$$\mathbf{Y} \sim \mathcal{N}(\Omega \mu + \nu, \Omega \Sigma \Omega^\top).$$

Cette proposition se prouve sans peine en utilisant la fonction caractéristique. On retrouve en particulier la stabilité par transformation affine établie en dimension 1.

6.3 La factorisation de Cholesky

Pour rappel, la factorisation de [Cholesky](#)¹ d'une matrice symétrique définie positive est donnée ci-dessous.

Theorem 6.1 (Factorisation de Cholesky). *Soit $\Sigma \in \mathbb{R}^{d \times d}$ une matrice symétrique définie positive. Alors il existe une matrice triangulaire inférieure $L \in \mathbb{R}^{d \times d}$ telle que $\Sigma = LL^\top$. La décomposition est unique si l'on impose que les éléments diagonaux de L soient strictement positifs.*

Proof. La factorisation de Cholesky est une conséquence directe de la méthode du pivot de Gauss. Le détail est donné par exemple dans (Th. 4.4.1, Ciarlet 2006). □

¹**André-Louis Cholesky, dit René**: (1875-1918) ingénieur topographe et géodésien dans l'armée française, mort des suites de blessures reçues au champs de bataille.

6.4 Simulation de vecteurs gaussiens

La simulation d'un vecteur gaussien de paramètres $\mu = (0, \dots, 0)^\top$ et $\Sigma = \text{Id}_d$ ne pose pas de problème : il suffit de simuler X_1, \dots, X_d , d variables aléatoires indépendantes de loi normale centrée réduite. En effet, le vecteur $\mathbf{X} = (X_1, \dots, X_d)^\top$ est alors un vecteur gaussien de loi $\mathcal{N}(0, \text{Id}_d)$.

Supposons maintenant que l'on veuille simuler un vecteur gaussien de loi $\mathcal{N}(\mu, \Sigma)$ dans \mathbb{R}^d , μ et Σ symétrique définie positive donnés.

6.4.1 Approche par la factorisation de Cholesky

La matrice Σ étant symétrique, elle peut s'écrire comme $\Sigma = LL^\top$ où L est une matrice triangulaire inférieure de taille $d \times d$. Grâce à la décomposition de Cholevsky et en reprenant les éléments de la preuve de la Proposition 6.1, on peut écrire $\mathbf{Y} = L\mathbf{X} + \mu$ où $\mathbf{X} \sim \mathcal{N}(0, \text{Id}_d)$ et vérifier que $\mathbf{Y} \sim \mathcal{N}(\mu, \Sigma)$.

6.4.2 Approche par la décomposition spectrale de Σ

La matrice Σ étant symétrique, elle se diagonalise en base orthonormée : il existe une matrice orthogonale P telle que

$$\Sigma = P \text{diag}(\lambda_1, \dots, \lambda_d) P^{-1} = P \text{diag}(\lambda_1, \dots, \lambda_d) P^\top,$$

où $\lambda_1, \dots, \lambda_d \geq 0$ sont les valeurs propres de Σ qui est semi-définie positive. On pose alors $R = P \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d})$ qui est une racine carrée matricielle de Σ au sens où $\Sigma = RR^\top$. On part alors d'un vecteur gaussien centrée réduit $\mathbf{X}_0 \sim \mathcal{N}(0, \text{Id}_d)$ que l'on sait simuler (par exemple avec la méthode de Box-Müller). La proposition Proposition 6.2 assure alors que le vecteur $\mathbf{X} = R\mathbf{X}_0 + \mu$ est un vecteur gaussien de loi $\mathcal{N}(\mu, \Sigma)$.

6.5 Vecteurs gaussiens : cas bidimensionnel

En dimension $p = 2$, la matrice de covariance Σ peut toujours s'écrire comme suit, et la visualisation suivante montre l'impact des différents paramètres sur la densité de probabilité.

$$\Sigma = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \cdot \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \cdot \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$$

où θ est l'angle de rotation et σ_1 et σ_2 les écarts-types des marginales (dans le repère orthonormal après rotation).

7 Méthode de Monte Carlo

De manière générale, ce qu'on appelle une méthode de **Monte-Carlo**¹ est une technique visant à calculer une quantité déterministe par le biais d'un procédé aléatoire. Une application très usuelle des méthodes Monte-Carlo est l'approximation numérique d'intégrales ou d'espérance, c'est ce que nous allons beaucoup développer dans ce chapitre.

Il existe également de nombreuses méthodes pour l'intégration numérique déterministe (méthodes des rectangles, trapèzes,...). Ces méthodes sont très rapides et efficaces en petite dimension, mais elles ne le sont plus du tout pour des dimensions plus grandes ou pour l'intégration de fonction peu régulières. Ainsi une intégration par Monte Carlo pour le cas de la grande dimension sera bien plus profitable.

7.1 Principe de la méthode

L'idée de la Méthode de Monte Carlo est d'exprimer l'intégrale qu'on cherche à approcher comme l'espérance d'une variable aléatoire, puis d'approcher cette espérance par une moyenne empirique. Si on veut calculer une intégrale de la forme

$$I = \int g(x)f(x)dx,$$

on la réécrit comme

$$I = \mathbb{E}[g(X)],$$

où X est une variable aléatoire de densité f . On suppose évidemment que $g(X)$ est bien intégrable, puis on approche l'espérance par la moyenne empirique

$$I_n = \frac{g(X_1) + g(X_2) + \dots + g(X_n)}{n}.$$

Il y a donc deux étapes : traduction sous forme d'espérance puis calcul de la moyenne empirique. Si l'expression à calculer est déjà sous forme d'espérance, il y a juste le calcul de la moyenne empirique à faire. Ce calcul se fait à partir de simulation de la variable aléatoire X sous-jacente.

¹Le nom viendrait du *casino de Monte-Carlo* inspiré par la passion pour le jeu de l'oncle du physicien Stanislaw Ulam... Source [Wikipedia](#).

Exemple 7.1. On peut approcher l'intégrale

$$I = \int_0^1 g(x) dx \simeq \sum_{i=0}^n w_i g(x_i),$$

où les w_i sont des nombres entre 0 et 1 dont la somme vaut 1, et les x_i sont des points de l'intervalle $[0, 1]$. Par exemple, pour la méthode des trapèzes, on prend $w_0 = w_n = \frac{1}{2n}$ et $w_i = \frac{1}{n}$ sinon et les $x_i = \frac{i}{n}$ sont régulièrement répartis. Si on considère la fonction $g(x) = e^{x^2}$ on approche $I = \int_0^1 e^{x^2} dx \simeq 1.46265 \dots$

```
import numpy as np

n= 10000
t = np.linspace(0, 1, n + 1)
w = np.ones_like(t) / n
w[0] /= 2
w[-1] /= 2
m0 = np.sum(w * np.exp(t**2))
print(f"La méthode des trapèzes donne {m0:.6f}...")
```

La méthode des trapèzes donne 1.462652...

La question est alors d'avoir une borne sur l'erreur commise (qui dépendra de n et de la régularité de la fonction g , voir par exemple [Wikipedia](#)).

Pour une méthode de Monte Carlo, on interprète l'intégrale comme $\mathbb{E}[g(X)]$ où X est une variable aléatoire de loi uniforme sur $[0, 1]$. Alors les w_i valent $\frac{1}{n+1}$ et les x_i sont tirés selon la loi uniforme sur $[0, 1]$.

```
X = np.random.uniform(0, 1, n)

# Direct method
gX = np.exp(X**2)
m1 = np.mean(gX)
print(f"La méthode naïve de Monte-Carlo {m1:.6f}...")
```

La méthode naïve de Monte-Carlo 1.459867...

Contrairement aux méthodes déterministes, la valeur de l'approximation change si on relance le calcul puisque les tirages sont aléatoires:

```

X = np.random.uniform(0, 1, n)

# Direct method
gX = np.exp(X**2)
m1 = np.mean(gX)
print(f"La méthode naïve de Monte-Carlo donne maintenant {m1:.6f}...")

```

La méthode naïve de Monte-Carlo donne maintenant 1.464366...

La question est alors de déterminer un intervalle de confiance contenant la vraie valeur I . La méthode de Monte Carlo est très facile à programmer si on dispose d'un simulateur de la loi de X , et n'impose aucune régularité sur g à part de la mesurabilité.

7.2 Théorèmes de convergence

On s'intéresse maintenant à la convergence de cet algorithme : sous quelles conditions est-ce que I_n converge vers I , en quel sens, et à la vitesse de cette convergence : combien de tirages faut-il faire pour atteindre une précision souhaitée.

7.2.1 Convergence de la méthode

Commençons par énoncer des conditions sous lesquelles la méthode converge.

Proposition 7.1. *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et de même loi que X , et g une fonction mesurable telle que $g(X)$ est intégrable. Alors la variable aléatoire*

$$I_n = \frac{g(X_1) + g(X_2) + \cdots + g(X_n)}{n}$$

converge presque presque sûrement vers $I = \mathbb{E}[g(X)]$ lorsque n tend vers l'infini.

Proof. Comme $g(X)$ est intégrable, et que les variables sont indépendantes et de même loi, il s'agit d'une application directe de la loi forte des grands nombres. \square

En particulier, si on sait simuler X , on peut en faire des tirages X_1, \dots, X_n et ainsi calculer I_n qui est une valeur approchée de I pour n assez grand.

Exemple 7.2. Par exemple, si l'on répète un très grand nombre de fois n une expérience aléatoire et que l'on s'intéresse à la réalisation de l'événement A alors la loi des grands nombres nous garantit que :

$$f_n(A) = \frac{1}{n} \times \text{Nbre de fois où } A \text{ s'est réalisé} \xrightarrow[n \rightarrow \infty]{ps} \mathbb{P}(A).$$

car $f_n(A) = (1/n) \sum_{i=1}^n X_i$ où X_i est une v.a. de Bernoulli de paramètre $p = \mathbb{P}(A) = \mathbb{E}[X]$. Autrement dit, la fréquence observée $f_n(A)$ de la réalisation de l'événement A au cours d'un grand nombre de répétitions de l'expérience se rapproche de la probabilité $P(A)$.

Exemple 7.3 (Le jeu des trois portes ou paradoxe de Monty Hall). Un jeu télévisé se déroule face à trois portes identiques. On a placé une voiture derrière l'une des portes, et une chèvre derrière chacune des deux autres. Le candidat est placé devant les trois portes, il en choisit une au hasard et se place devant.

La présentatrice (qui sait où se trouve la voiture) ouvre alors l'une des deux autres portes, derrière laquelle il y a une chèvre, puis elle demande au candidat : "Voulez-vous modifier votre choix ?"

Le candidat a alors deux possibilités :

1. soit il ouvre la porte qu'il avait choisie en premier ; (il conserve son premier choix)
2. soit il ouvre une autre porte (celle qui n'a pas été ouverte par la présentatrice).

Quelle est la meilleure stratégie pour le candidat ? Garder son premier choix ou changer de porte ?

Un raisonnement probabiliste peut fournir la solution. Mais vous pouvez vous aider en faisant des expériences pour faire une conjecture ! On peut par exemple lancer le programme ci-dessous en demandant de reproduire dix mille fois l'expérience ($n = 10000$).

```
import numpy as np

portes = np.array(["chevre", "chevre", "voiture"])
NbSimulations = 10000

gagne_avec_change = 0
gagne_avec_conserve = 0

for _ in range(NbSimulations):
    porte_choisie = np.random.choice(portes)
    if porte_choisie == "voiture":
        gagne_avec_conserve += 1
    else:
```

```

gagne_avec_change += 1

print(f"Gagne en changeant : {gagne_avec_change / NbSimulations * 100:.3f} %")
print(f"Gagne sans changer : {gagne_avec_conserve / NbSimulations * 100:.3f} %")

```

Gagne en changeant : 66.920 %
 Gagne sans changer : 33.080 %

On peut alors conjecturer que la probabilité de gagner en changeant de porte est égale à $2/3$ alors que si le candidat reste sur son premier choix la probabilité de gagner ne sera que de $1/3$.

7.2.2 Précision de la méthode

On souhaiterait maintenant savoir quand on peut considérer que n est assez grand pour que l'approximation soit satisfaisante.

Proposition 7.2. *Si $g(X)$ est de carré intégrable, alors un intervalle de confiance asymptotique de niveau de confiance $1 - \alpha$ pour l'intégrale I est*

$$\left(I_n - F_{\mathcal{N}(0,1)}^{-1}\left(\frac{\alpha}{2}\right) \sqrt{\frac{S_n^2}{n}} \leq I \leq I_n + F_{\mathcal{N}(0,1)}^{-1}\left(\frac{\alpha}{2}\right) \sqrt{\frac{S_n^2}{n}} \right),$$

où $F_{\mathcal{N}(0,1)}^{-1}$ est la fonction quantile de la loi normale centrée réduite et

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (g(X_i) - I_n)^2$$

est l'estimateur sans biais² de la variance empirique.

Proof. Comme $g(X)$ est de carré intégrable, et que les X_i sont iid, on peut appliquer le théorème central limite à la suite $(g(X_i))$, ce qui donne

$$\frac{\frac{1}{n} \sum_{i=1}^n g(X_i) - \mathbb{E}[g(X)]}{\sqrt{\frac{\text{var}(g(X))}{n}}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1),$$

²Ceci signifie que $\mathbb{E}[S_n^2] = \text{var}(g(X))$

ce qui se traduit dans notre contexte par

$$\frac{I_n - I}{\sqrt{\frac{\text{Var}(g(X))}{n}}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

En général, on ne connaît pas non plus $\text{Var}(g(X))$, donc on la remplace par un estimateur de la variance $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (g(X_i) - I_n)^2$. D'après la loi des grands nombres, cet estimateur converge ps, donc en probabilité (donc en loi) vers $\text{Var}(g(X))$ qui est une constante. D'après le lemme de Slutsky, le couple $(\frac{I_n - I}{\sqrt{n \text{Var}(g(X))}}, S_n^2)$ converge en loi vers le couple $(\mathcal{N}(0, 1), \delta_{\text{Var}(g(X))})$. En particulier, on a encore la convergence en loi du produit

$$\frac{I_n - I}{\sqrt{\frac{S_n^2}{n}}} = \frac{I_n - I}{\sqrt{\frac{\text{Var}(g(X))}{n}}} \times \frac{\sqrt{\text{Var}(g(X))}}{\sqrt{S_n^2}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1) \times 1.$$

On peut en déduire facilement la construction de l'intervalle de confiance asymptotique de l'énoncé. \square

Ce résultat a deux conséquences importantes. D'une part, la vitesse de convergence est en $n^{-1/2}$, ce qui est relativement lent, mais indépendant à la fois de la dimension de X et de la régularité de g . D'autre part, on peut calculer S_n^2 donc estimer l'erreur commise sans tirages aléatoires supplémentaires, avec le même échantillon qui sert pour estimer I . **Il est fortement recommandé de toujours faire une estimation de la précision en même temps que l'estimation Monte Carlo de l'espérance.**

Exemple 7.4. En finance, certains contrats (dont le coût est $C > 0$ et que l'on peut voir comme une assurance) permettent de vendre un actif acheté $K > 0$ à une date future N fixée par le contrat s'il dépasse à ce moment là une certaine valeur cible K également fixée dans le contrat. Si le prix de l'actif est plus bas que K , on récupère tout de même sa mise K . Considérons un modèle simplifié pour le cours de l'actif $(S_n)_{n \geq 1}$: pour $n \geq 1$, on pose $S_n = \sum_{i=1}^n X_i$ où les X_i sont des variables aléatoires indépendantes de loi normale centrée réduite. Le gain moyen de la détentrice ou du détenteur du contrat décrit ci-dessus est donc $\mathbb{E}[\max(S_N - K, 0)]$. Le contrat n'est pas gratuit... Ainsi, l'enjeu est de comparer le coût C et l'espérance $\mathbb{E}[\max(S_N - K, 0)]$.

Estimons cette quantité par la méthode de Monte Carlo.

```
import numpy as np

n = 100000
N = 60
```

```

K= 1
I = []

for i in range(n):
    X = np.random.normal(0, 1, N)
    S = max(np.sum(X) - K, 0)
    I.append(S)

I = np.array(I)
m = np.mean(I)
s = np.var(I)

print(f"[{m - 1.96 * np.sqrt(s / n):.3f}, {m:.3f}, {m + 1.96 * np.sqrt(s / n):.3f}]")

```

[2.596, 2.622, 2.648]

7.3 Réduction de variance

La proposition Proposition 7.2 donne deux leviers pour augmenter la précision de l'estimation I_n de I :

1. augmenter le nombre n de tirages,
2. diminuer la variance $\text{Var}(g(X))$.

Si le premier levier est facilement praticable pour des cas simples, il n'est pas envisageable lorsque une simulation de $g(X)$ est coûteuse en temps de calcul ou lorsque la variance est très grande, ce qui arrive couramment. Nous allons examiner le deuxième levier dans cette partie, c'est ce qu'on appelle les **méthodes de réduction de variance**.

Regardons d'abord un exemple pour bien comprendre l'intérêt de ces méthodes.

Exemple 7.5. Revenons sur l'exemple Example 7.2 où on cherche à estimer $p = \mathbb{P}(A)$ à partir de tirages de loi de Bernoulli. Avec n tirages, on a $I_n \simeq p + \sqrt{\frac{p(1-p)}{n}} Z$ avec Z de loi normale centrée réduite. La valeur de p est inconnue, mais on peut majorer $p(1-p)$ par $1/4$. Si on veut faire une erreur de l'ordre 0.01 sur p , il convient de choisir $n = \frac{1}{4 \times 0.01^2} = 0.25 \cdot 10^4 = 2500$.

Si p est proche de 0.5, une erreur de 10^{-2} peut être acceptable. Mais si on cherche à approcher une probabilité très petite, par exemple $p = 10^{-5}$, il faudra une précision d'au

moins 10^{-6} , et donc prendre $n = 0.25 \cdot 10^{12}$ tirages. On voit bien apparaître les limites de la méthode directe, et l'intérêt de toujours calculer l'ordre de grandeur de l'erreur commise.

Il existe de nombreuses techniques de réduction de variance pour améliorer la vitesse de convergence. Malheureusement, aucune ne marche de façon automatique, il faut les choisir et les adapter au cas par cas. Il s'agit essentiellement d'exploiter d'une façon ou d'une autre l'idée que l'intégrale I peut s'écrire de nombreuses façons différentes comme une espérance, et de choisir l'expression de variance minimale.

7.3.1 Variable de contrôle

Une première façon très simple de décomposer I est la suivante

$$I = \mathbb{E}[g(X)] = \mathbb{E}[g(X) - h(X)] + \mathbb{E}[h(X)].$$

Cette méthode est intéressante si on sait calculer explicitement $\mathbb{E}[h(X)]$ et si la variance diminue : $\text{Var}(g(X) - h(X)) < \text{Var}(g(X))$. Idéalement, on voudrait que la variance de $g(X) - h(X)$ soit très petite. Or rappelons qu'une variable aléatoire est de variance nulle si et seulement si elle est constante. Pour mettre en oeuvre cette méthode, on va donc chercher une fonction h qui soit une bonne approximation de g sur l'intervalle d'intégration pour que la différence $g - h$ soit presque constante.

Exemple 7.6. Supposons que l'on veuille calculer $I = \int_0^1 e^{x^2} dx \simeq 1.46265 \dots$ comme dans l'exemple [Exemple 7.1](#) à l'aide de la loi uniforme sur $[0, 1]$. Comme au voisinage de 0, on a $e^{x^2} \simeq 1 + x^2$, on propose de prendre $h(x) = 1 + x^2$. On a alors

$$\mathbb{E}[h(X)] = \int_0^1 1 + x^2 dx = \left[x + \frac{x^3}{3} \right]_0^1 = \frac{4}{3}.$$

On peut donc approcher I par

$$I_n^c = \frac{1}{n} \sum_{i=1}^n (e^{X_i^2} - 1 - X_i^2) + \frac{4}{3},$$

où les X_i sont iid de loi uniforme sur $[0, 1]$. Regardons ce que ça donne numériquement.

```
import numpy as np

n = 10000
X = np.random.uniform(0, 1, n)
```



```
# Direct method
gX = np.exp(X**2)
m1 = np.mean(gX)
s1 = np.var(gX)

print(f"Direct method confidence interval: [{m1 - 1.96 * np.sqrt(s1 / n):.3f}, {m1:.3f}, {m1 + 1.96 * np.sqrt(s1 / n):.3f}]")
print(f"Variance {s1:.3f}")
```

Direct method confidence interval: [1.460, 1.469, 1.479]
Variance 0.225

```
# Control variable
ghX = np.exp(X**2) - 1 - X**2
m2 = np.mean(ghX) + 4/3
s2 = np.var(ghX)

print(f"Control variable confidence interval: [{m2 - 1.96 * np.sqrt(s2 / n):.3f}, {m2:.3f}, {m2 + 1.96 * np.sqrt(s2 / n):.3f}]")
print(f"Variance {s2:.3f}")
```

Control variable confidence interval: [1.461, 1.465, 1.468]
Variance 0.034

On constate que l'intervalle de confiance est plus étroit donc la précision meilleure avec la variable de contrôle. On a divisé la variance (empirique) par 6.7. Pour obtenir le même gain en augmentant le nombre de simulations, il aurait fallu $6.7^2 \simeq 45$ fois plus de simulations.

7.3.2 Variables antithétiques

S'il existe une fonction h telle que X et $h(X)$ ont la même loi, on peut écrire I comme

$$I = \mathbb{E} \left[\frac{g(X) + g(h(X))}{2} \right].$$

On a en fait décomposé l'intégrale I comme somme de deux espérances. Notons que ces deux intégrales font apparaître la même variable aléatoire, donc on peut estimer les deux espérances avec les **mêmes tirages**. En fait on a introduit une certaine forme de dépendance puisqu'on utilise $g(X_i)$ et $g(h(X_i))$ dans l'estimateur.

Examinons l'effet de cette décomposition sur la variance.

Lemma 7.1. Supposons qu'il existe une fonction h telle que X et $h(X)$ ont la même loi de carré intégrable. Si $\text{Cov}(g(X), g(h(X))) \leq 0$ alors

$$\text{Var} \left(\frac{g(X) + g(h(X))}{2} \right) \leq \frac{\text{Var}(g(X))}{2},$$

Autrement dit la méthode des variables antithétiques permet de réduire la variance au moins de moitié si $g(X)$ et $g(h(X))$ sont négativement corrélées.

Proof. Calculons la variance $V = \text{Var}(g(X) + g(h(X)))$:

$$\begin{aligned} V &= \text{Var}(g(X) + g(h(X))) \\ &= \mathbb{E}[(g(X) - \mathbb{E}[g(X)] + g(h(X)) - \mathbb{E}[g(h(X))])^2] \\ &= \mathbb{E}[(g(X) - \mathbb{E}[g(X)])^2] + \mathbb{E}[(g(h(X)) - \mathbb{E}[g(h(X))])^2] \\ &\quad + 2\mathbb{E}[(g(X) - \mathbb{E}[g(X)])(g(h(X)) - \mathbb{E}[g(h(X))])] \\ &= \text{Var}(g(X)) + \text{Var}(g(h(X))) + 2\text{Cov}(g(X), g(h(X))). \end{aligned}$$

En particulier, si $\text{Cov}(g(X), g(h(X))) \leq 0$, on a

$$\begin{aligned} \text{Var} \left(\frac{g(X) + g(h(X))}{2} \right) &\leq \frac{\text{Var}(g(X)) + \text{Var}(g(h(X)))}{4} \\ &= \frac{\text{Var}(g(X))}{2}, \end{aligned}$$

car $g(X)$ et $g(h(X))$ ont la même loi donc la même variance. Dans le cas général, on a quand toujours $2\text{Cov}(g(X), g(h(X))) \leq \text{Var}(g(X)) + \text{Var}(g(h(X)))$ par l'inégalité de Cauchy-Schwartz, ce qui implique

$$\begin{aligned} \text{Var} \left(\frac{g(X) + g(h(X))}{2} \right) &\leq 2 \frac{\text{Var}(g(X)) + \text{Var}(g(h(X)))}{4} \\ &= \text{Var}(g(X)), \end{aligned}$$

donc même si on ne réduit pas effectivement la variance, on ne peut pas l'augmenter par cette méthode. On peut faire des calculs plus précis en faisant intervenir la covariance de $g(X)$ et $g(h(X))$ dans certains cas particuliers si on spécifie les lois. \square

Cette méthode est particulièrement simple à mettre en oeuvre pour les lois qui ont des propriétés de symétrie

Example 7.7. Revenons sur l'exemple Example 7.6 où on cherche à approcher $I = \int_0^1 e^{x^2} dx$ à l'aide de la loi uniforme sur $[0, 1]$. Cette loi a une propriété de symétrie : X et $1 - X$ ont la même loi. On a donc une approximation de I de qualité identique ou meilleure en prenant

$$I_n^a = \frac{1}{2n} \sum_{i=1}^n (e^{X_i^2} + e^{(1-X_i)^2}).$$

Regardons ce que cela donne numériquement.

```
import numpy as np

n = 10000
X = np.random.uniform(0, 1, n)

# Direct method
gX = np.exp(X**2)
m1 = np.mean(gX)
s1 = np.var(gX)

print(f"Direct method confidence interval: [{m1 - 1.96 * np.sqrt(s1 / n):.3f}, {m1:.3f}, {m1 + 1.96 * np.sqrt(s1 / n):.3f}]")
print(f"Variance {s1:.3f}")
```

Direct method confidence interval: [1.451, 1.460, 1.470]
Variance 0.225

```
# Antithetic variable method
gaX = (np.exp(X**2) + np.exp((1 - X)**2)) / 2
m3 = np.mean(gaX)
s3 = np.var(gaX)

print(f"Antithetic variable method confidence interval: [{m3 - 1.96 * np.sqrt(s3 / n):.3f}, {m3:.3f}, {m3 + 1.96 * np.sqrt(s3 / n):.3f}]")
print(f"Variance {s3:.3f}")
```

Antithetic variable method confidence interval: [1.458, 1.461, 1.465]
Variance 0.028

On constate que l'intervalle de confiance est plus étroit donc la précision meilleure avec la variable antithétique. On a divisé la variance (empirique) par 8. Pour obtenir le même gain en augmentant le nombre de simulations, il aurait fallu $8^2 = 64$ fois plus de simulations.

7.3.3 Échantillonnage préférentiel

L'idée derrière l'échantillonnage préférentiel est similaire à celle des variables de contrôle. Cette fois-ci, au lieu de retrancher et d'ajouter une quantité, on va multiplier et diviser. L'interprétation en terme d'espérance est cependant complètement différente.

Supposons que l'on sache simuler une variable aléatoire Y de densité h . On peut alors écrire I comme

$$\begin{aligned} I &= \mathbb{E}[g(X)] = \int g(x)f(x)dx \\ &= \int \frac{g(x)f(x)}{h(x)}h(x)dx = \mathbb{E}\left[\frac{g(Y)f(Y)}{h(Y)}\right]. \end{aligned}$$

On a gagné quelque chose si $\text{Var}\left(\frac{gf}{h}(Y)\right) < \text{Var}(g(X))$.

Comme pour la méthode de la variable de contrôle, on souhaite que la nouvelle variable $\frac{g(Y)f(Y)}{h(Y)}$ soit de variance aussi petite que possible, donc proche d'une constante. Si on prend $h(x) = \frac{g(x)f(x)}{\mathbb{E}[g(X)]}$ (qui est bien une densité si g est positive), alors la variance de $\frac{g(Y)f(Y)}{h(Y)}$ est nulle. Évidemment, ce résultat n'est pas utilisable en pratique puisqu'on cherche justement à calculer $\mathbb{E}[g(X)]$. Cependant, il guide la recherche d'une bonne fonction h vers une approximation de $|gf|$ normalisée pour obtenir une densité.

Exemple 7.8. Revenons une fois de plus sur l'exemple Exemple 7.6 où on cherche à approcher $I = \int_0^1 e^{x^2} dx$ à l'aide de la loi uniforme sur $[0, 1]$. On a donc $f(x) = 1$ sur l'intervalle d'intégration, on cherche donc à nouveau une approximation de $g(x) = e^{x^2}$. Utilisons encore l'approximation $g(x) \simeq 1 + x^2$, ce qui nous donne comme densité candidate $h(x) = \frac{3}{4}(1 + x^2)$. Regardons si on peut facilement simuler cette densité. Calculons sa fonction de répartition (sur $[0, 1]$)

$$F(t) = \int_0^t \frac{3}{4}(1 + x^2) = \frac{3}{4}\left[x + \frac{x^3}{3}\right]_0^1 = \frac{3}{4}\left(t + \frac{t^3}{3}\right).$$

Essayons de l'inverser

$$F(t) = u \Leftrightarrow u = \frac{4}{3}\left(t + \frac{t^3}{3}\right) \Leftrightarrow t^3 + 3t - 4u = 0$$

On trouve alors par la **méthode de Cardan** (voir par exemple la page [Equation du troisième degré / Méthode de Cardan](#))

$$F(t) = u \Leftrightarrow t = (\sqrt{1 + 4u^2} + 2u)^{1/3} - (\sqrt{1 + 4u^2} - 2u)^{1/3}.$$

Regardons ce que ça donne numériquement.

```
import numpy as np

n = 10000
X = np.random.uniform(0, 1, n)

# Direct method
gX = np.exp(X**2)
m1 = np.mean(gX)
s1 = np.var(gX)

print(f"Direct method confidence interval: [{m1 - 1.96 * np.sqrt(s1 / n):.3f}, {m1:.3f}, {m1 + 1.96 * np.sqrt(s1 / n):.3f}]")
print(f"Variance {s1:.3f}")
```

Direct method confidence interval: [1.449, 1.458, 1.467]
Variance 0.222

```
# Importance sampling
Y = (2*X + np.sqrt(1 + 4*X**2))**(1/3) - (np.sqrt(1 + 4*X**2) - 2*X)**(1/3)
gpX = np.exp(Y**2) / (3 * (1 + Y**2) / 4)
m4 = np.mean(gpX)
s4 = np.var(gpX)

print(f"Importance sampling confidence interval: [{m4 - 1.96 * np.sqrt(s4 / n):.3f}, {m4:.3f}, {m4 + 1.96 * np.sqrt(s4 / n):.3f}]")
print(f"Variance {s4:.3f}")
```

Importance sampling confidence interval: [1.458, 1.461, 1.464]
Variance 0.020

On constate que l'intervalle de confiance est plus étroit donc la précision meilleure avec l'échantillonnage préférentiel. On a divisé la variance (empirique) par environ 12. Pour obtenir le même gain en augmentant le nombre de simulations, il aurait fallu $12^2 = 144$ fois plus de simulations.

Il existe bien d'autres méthodes de réduction de variance, comme par exemple la méthode de stratification très utilisée en théorie des sondages.

8 Marches aléatoires

L'objet de ce chapitre est de faire une première introduction à un exemple intéressant de processus stochastique, soit la marche aléatoire simple symétrique, et ce avec un minimum de formalisme.

Dans la partie Section 8.1, on définit la marche aléatoire simple symétrique, et on étudie l'allure de ses trajectoires. La partie Section 8.2 est dédiée à son comportement en temps long. La partie Section 8.3 présente une première application vers les mathématiques financières avec l'étude d'un problème de ruine.

8.1 Trajectoires de la marche aléatoire

Un *processus stochastique* est une suite de variables aléatoires indexée par le temps. Il permet de modéliser l'évolution temporelle d'un phénomène comme la richesse d'une joueuse ou la valeur d'un portefeuille d'actions par exemple. La marche aléatoire simple symétrique est le modèle le plus simple de processus à temps discret.

Dans toute la suite, on se place sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$.

8.1.1 Définition

Commençons par définir la marche aléatoire simple symétrique.

Definition 8.1. Une **marche aléatoire simple symétrique** sur \mathbb{Z} est une suite $(S_n)_{n \in \mathbb{N}}$ de variables aléatoires telles que

1. $S_0 = x \in \mathbb{Z}$ est déterministe,
2. $S_{n+1} = S_n + X_{n+1}$ pour tout $n \in \mathbb{N}$,

où $(X_n)_{n \geq 1}$ est une suite de variables aléatoires indépendantes et de même loi de Rademacher de paramètre $\frac{1}{2}$:

$$\mathbb{P}(X_1 = -1) = \mathbb{P}(X_1 = 1) = \frac{1}{2}.$$

On parle de *marche* car ce processus peut représenter la position d'une personne qui se déplace en ligne droite par pas d'une unité avec équiprobabilité d'aller vers l'avant (+1) ou vers l'arrière (−1). Cette marche est dite *simple* car on ne fait que des pas d'amplitude 1 et *symétrique* car il y a la même probabilité de tirer 1 et −1 pour chaque incrément.

Ce processus peut également servir à modéliser la richesse d'une joueuse qui joue à pile ou face avec une pièce équilibrée. Partant d'une richesse initiale $S_0 = x > 0$, à chaque pas de temps elle perd ou gagne 1€ en fonction du tirage obtenu. Les variables X_n représentent alors le gain du n -ème jeu, et S_n la richesse totale de la joueuse après le n -ème tirage.

À partir de la définition, on obtient très facilement le comportement moyen de la marche.

Proposition 8.1. *Pour tout $n \in \mathbb{N}$, on a*

$$\mathbb{E}[S_n] = S_0, \quad \text{Var}(S_n) = n.$$

Proof. Remarquons d'abord que

$$\begin{aligned} \mathbb{E}[X_1] &= (-1) \times \frac{1}{2} + 1 \times \frac{1}{2} = 0, \\ \text{Var}(X_1) &= \mathbb{E}[X_1^2] = (-1)^2 \times \frac{1}{2} + 1^2 \times \frac{1}{2} = 1, \end{aligned}$$

les incréments sont donc centrés et réduits. Par linéarité de l'espérance, on obtient

$$\mathbb{E}[S_n] = \mathbb{E}[S_0 + \sum_{k=1}^n X_k] = S_0 + \sum_{k=1}^n \mathbb{E}[X_k] = S_0,$$

puisque S_0 est déterministe. Enfin, en utilisant l'indépendance des X_k , on obtient

$$\text{Var}(S_n) = \text{Var}(S_0 + \sum_{k=1}^n X_k) = \sum_{k=1}^n \text{Var}(X_k) = n,$$

d'où le résultat. □

Ainsi, en moyenne la marche reste constante, mais sa variance devient de plus en plus grande au cours du temps.

8.1.2 Représentation graphique

Les marches aléatoires, comme tous les processus à valeurs discrètes, ont une représentation graphique naturelle. On représente graphiquement un tirage de la marche $(S_n)_{n \in \mathbb{N}}$ sur l'intervalle de temps $\llbracket 0, N \rrbracket$ par une ligne brisée joignant les points $(0, S_0), (1, S_1), \dots, (N, S_N)$. On a donc en abscisse le temps, et en ordonnée la valeur de la marche, cf. **?@fig-marche**.

La marche aléatoire est un processus à temps discret, donc rigoureusement sa représentation graphique devrait être constituée des points (n, S_n) non reliés. Par exemple, $S_{1/2}$ n'est pas défini. Cependant, on relie les points pour une meilleure lecture des graphiques.

On s'intéresse maintenant aux représentations graphiques qui peuvent correspondre à des tirages de la marche aléatoire.

Definition 8.2. Soit $n > m$ deux entiers naturels et $(a, b) \in \mathbb{Z}^2$ deux entiers relatifs. On appelle *chemin* de (m, a) à (n, b) la représentation graphique d'un tirage de la marche aléatoire telle que $S_m = a$ et $S_n = b$.

Example 8.1. La figure XXX présente deux chemins de $(1, 1)$ à $(5, 3)$.

On peut maintenant se poser la question de savoir s'il existe toujours un chemin qui relie deux points quelconques, et s'il en existe, combien on peut construire de chemins différents reliant ces deux points. On peut facilement montrer que certains chemins sont impossibles en regardant la parité de la marche aléatoire et son amplitude maximale.

Proposition 8.2 (Parité). Pour tout $n \in \mathbb{N}$, S_{2n} a la même parité que S_0 et S_{2n+1} a la parité opposée à celle de S_0 .

Proof. Pour tout n , X_{n+1} prend les valeurs -1 ou 1 donc est impair. Comme $S_{n+1} = S_n + X_{n+1}$, la marche change de parité à chaque pas de temps. \square

On en déduit immédiatement que les chemins doivent respecter cette alternance de parité.

Corollary 8.1. Si $n - m$ et $b - a$ n'ont pas la même parité, alors il n'existe pas de chemin de (m, a) à (n, b) .

Proof. Supposons sans perte de généralité que $n - m$ est pair, et qu'il existe un chemin de (m, a) à (n, b) . Alors $S_m = a$ et $S_n = b$ ont nécessairement la même parité, donc $S_n - S_m = b - a$ est pair. \square

Example 8.2. Il n'existe pas de chemin de $(0, 0)$ à $(2, 1)$.

Proposition 8.3 (Amplitude). *Pour tout $n \in \mathbb{N}$, on a $S_0 - n \leq S_n \leq S_0 + n$.*

Proof. Pour tout n , X_n prend les valeurs -1 ou 1 donc en particulier

$$-1 \leq X_n \leq 1.$$

Comme $S_n = S_0 + \sum_{k=1}^n X_k$, en sommant les inégalités on obtient bien

$$S_0 - n \leq S_n \leq S_0 + n,$$

d'où le résultat. □

Corollary 8.2. *Si $n - m < |b - a|$, alors il n'existe pas de chemin de (m, a) à (n, b) .*

Proof. Supposons qu'il existe un chemin de (m, a) à (n, b) . Alors $S_m = a$ et $S_n = b$. En particulier, $S_n = S_m + \sum_{k=m+1}^n X_k$, d'où

$$\begin{aligned} S_m - (n - m) &\leq S_n \leq S_m + (n - m) \\ \Rightarrow a - (n - m) &\leq b \leq a + (n - m) \\ \Rightarrow -(n - m) &\leq b - a \leq n - m. \end{aligned}$$

On a donc nécessairement $|b - a| \leq n - m$. □

Exemple 8.3. Il n'existe pas de chemin de $(0, 0)$ à $(2, 4)$.

Une fois les conditions de parité et d'amplitude satisfaites, il existe toujours des chemins reliant deux points, et on peut les dénombrer explicitement.

Proposition 8.4 (Nombre de chemins). *Si $|b - a| \leq n - m$ et $n - m$ et $b - a$ ont la même parité, alors le nombre de chemins de (m, a) à (n, b) est le nombre de combinaisons*

$$\binom{n - m}{\frac{n - m}{2} + \frac{b - a}{2}}.$$

Proof. Un chemin de (m, a) à (n, b) correspond à $n - m$ pas de la marche aléatoire en partant de a et en arrivant en b . Supposons sans perte de généralité que $b > a$. Alors on peut décomposer les pas de la marche entre x pas vers le haut (tirages de $+1$) et y pas vers le bas (tirages de -1). On doit avoir un nombre total de pas de $x + y = n - m$, et une différence des ordonnées de $x - y = b - a$, ce qui donne le système

$$\begin{cases} x + y = n - m \\ x - y = b - a \end{cases} \Leftrightarrow \begin{cases} x = \frac{n-m}{2} + \frac{b-a}{2} \\ y = \frac{n-m}{2} - \frac{b-a}{2} \end{cases}.$$

Pour dénombrer le nombre total de chemins de longueur $x + y$ avec x montées et y descentes, il suffit de choisir les emplacements de x montées parmi les $x + y$ pas. On obtient donc $\binom{x+y}{x} = \binom{\frac{n-m}{2} + \frac{b-a}{2}}{\frac{n-m}{2}}$ chemins. \square

Remarquons que pour dénombrer le nombre de chemins d'un point à un autre, nous n'avons pas utilisé le fait que $\mathbb{P}(X = 1) = \mathbb{P}(X = -1)$, mais uniquement le fait que X ne peut prendre que les valeurs -1 et 1 . Le résultat ci-dessus est donc également valable pour les marches asymétriques où $p = \mathbb{P}(X = 1) = 1 - \mathbb{P}(X = -1) \neq \frac{1}{2}$.

8.1.3 Principe de réflexion

On souhaite maintenant dénombrer les chemins qui ont une propriété particulière supplémentaire, comme passer par un point fixé. Pour cela, on utilise les propriétés de symétrie de la marche.

Proposition 8.5 (Principe de réflexion). *Soit a et b deux entiers naturels non nuls et n, m deux entiers naturels tels que $n > m$, $|b - a| \leq n - m$ et $b - a$ et $n - m$ ont la même parité. Alors le nombre de chemins de (m, a) à (n, b) passant par 0 (i.e. touchant l'axe des abscisses) est égal au nombre de chemins de (m, a) à $(n, -b)$.*

Proof. Le principe est de construire une bijection entre les deux ensembles de chemins pour prouver qu'ils ont le même cardinal. Pour cela, on exploite la symétrie de la marche, comme illustré figure XXX.

Considérons un chemin (S_m, \dots, S_n) de (m, a) à (n, b) passant par 0 pour la première fois en p . On a alors $m < p < n$. On construit un chemin $(\tilde{S}_m, \dots, \tilde{S}_n)$ de (m, a) à $(n, -b)$ en gardant la première partie du chemin initial de (m, a) à $(p, 0)$, puis en prenant ensuite le symétrique du chemin initial par rapport à l'axe des abscisses de $(p, 0)$ à $(n, -b)$:

$$\tilde{S}_k = \begin{cases} S_k & \text{si } m \leq k \leq p, \\ -S_k & \text{si } p \leq k \leq n. \end{cases}$$

Cette application définit une bijection entre les deux ensembles de chemins. En effet, si on prend maintenant un chemin $(\tilde{S}_m, \dots, \tilde{S}_n)$ de (m, a) à $(n, -b)$, comme il fait des pas de 1, que $a > 0$ et $-b < 0$, ce chemin passe nécessairement par 0. Soit p le premier instant où

le chemin est en 0. On construit alors un chemin (S_m, \dots, S_n) de (m, a) à (n, b) passant par 0 en posant

$$S_k = \begin{cases} \tilde{S}_k & \text{si } m \leq k \leq p, \\ -\tilde{S}_k & \text{si } p \leq k \leq n. \end{cases}$$

Il y a donc le même nombre de chemins des deux types. □

A nouveau, nous n'avons pas utilisé le fait que $\mathbb{P}(X = 1) = \mathbb{P}(X = -1)$, mais uniquement le fait que X ne peut prendre que les valeurs -1 et 1 qui sont symétriques l'une de l'autre. Le résultat ci-dessus est donc également valable pour les marches asymétriques où $p = \mathbb{P}(X = 1) = 1 - \mathbb{P}(X = -1) \neq \frac{1}{2}$. Une première application de cette propriété est le résultat suivant connu sous le nom de *théorème du scrutin*.

Theorem 8.1. *Au cours d'une élection opposant deux candidates A et B , la candidate A (resp. B) a obtenu a (resp. b) voix, avec $a > b$. Alors la probabilité que A ait été majoritaire (au sens large) tout au long du dépouillement est*

$$p = 1 - \frac{b}{a + 1}.$$

Proof. Tous les dépouillement étant équiprobables, p s'obtient comme le rapport du nombre de dépouillements avec A en tête tout le long par le nombre total de dépouillements. On peut modéliser un dépouillement par une marche aléatoire (S_n) (pas nécessairement symétrique) où S_n est le nombre de voix d'avance de A sur B après le dépouillement du n -ème bulletin.

Il y a en tout $a + b$ bulletins et à la fin $a - b$ voix d'avance de A sur B , donc le nombre total de dépouillements est le nombre de chemins de $(0, 0)$ à $(a + b, a - b)$ et vaut $\binom{a+b}{a}$.

Le nombre de dépouillements avec A en tête tout le long correspond

- au nombre de chemins de $(0, 0)$ à $(a + b, a - b)$ ne prenant aucune valeur strictement négative, autrement dit le nombre de chemins de $(0, 0)$ à $(a + b, a - b)$ ne touchant pas -1 ;
- au nombre de chemins de $(0, 1)$ à $(a + b, a - b + 1)$ ne touchant pas 0, en décalant d'un cran les ordonnées;
- au nombre total de chemins de $(0, 1)$ à $(a + b, a - b + 1)$ moins le nombre de chemins de $(0, 1)$ à $(a + b, a - b + 1)$ touchant 0;
- au nombre total de chemins de $(0, 1)$ à $(a + b, a - b + 1)$ moins le nombre de chemins de $(0, 1)$ à $(a + b, -(a - b + 1))$ par le principe de réflexion.

Il vaut donc $\binom{a+b}{a} - \binom{a+b}{b-1}$. Ainsi

$$p = \frac{\binom{a+b}{a} - \binom{a+b}{b-1}}{\binom{a+b}{a}} = 1 - \frac{\binom{a+b}{b-1}}{\binom{a+b}{a}} = 1 - \frac{(a+b)!a!b!}{(b-1)!(a+1)!(a+b)!} = 1 - \frac{b}{a+1},$$

d'où le résultat. \square

8.2 Comportement asymptotique

On s'intéresse maintenant au comportement de la marche sans limite de temps. Parmi les questions d'intérêt, on peut se demander si elle revient à son point de départ x , combien de fois et en combien de temps.

Comme les $(X_n)_{n \geq 1}$ sont indépendantes et de même loi intégrable, on peut appliquer la loi des grands nombres pour obtenir que

$$\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{ps} \mathbb{E}[X_1] = 0,$$

ce qui ne donne aucune information sur le comportement en temps long de S_n .

8.2.1 Propriété de Markov et stationnarité

Les deux propriétés suivantes sont des propriétés du processus et non plus seulement des trajectoires. Elles énoncent une certaine forme d'invariance de la marche au cours du temps. La première est une propriété d'absence de mémoire : la position future de la marche ne dépend que de sa position actuelle et pas de la trajectoire qui l'a amenée à cette position.

Proposition 8.6 (Markov). *Pour tout $n \in \mathbb{N}$ et pour tous $(a_0, a_1, \dots, a_{n+1}) \in \mathbb{Z}^{n+2}$, on a*

$$\begin{aligned} & \mathbb{P}(S_{n+1} = a_{n+1} | S_n = a_n, S_{n-1} = a_{n-1}, \dots, S_1 = a_1, S_0 = a_0) \\ &= \mathbb{P}(S_{n+1} = a_{n+1} | S_n = a_n). \end{aligned}$$

Proof. Par définition des probabilités conditionnelles, on a

$$\begin{aligned} & \mathbb{P}(S_{n+1} = a_{n+1} | S_n = a_n, S_{n-1} = a_{n-1}, \dots, S_1 = a_1, S_0 = a_0) \\ &= \frac{\mathbb{P}(S_{n+1} = a_{n+1}, S_n = a_n, S_{n-1} = a_{n-1}, \dots, S_1 = a_1, S_0 = a_0)}{\mathbb{P}(S_n = a_n, S_{n-1} = a_{n-1}, \dots, S_1 = a_1, S_0 = a_0)} \\ &= \frac{\mathbb{P}(X_{n+1} = a_{n+1} - a_n, S_n = a_n, S_{n-1} = a_{n-1}, \dots, S_1 = a_1, S_0 = a_0)}{\mathbb{P}(S_n = a_n, S_{n-1} = a_{n-1}, \dots, S_1 = a_1, S_0 = a_0)}, \end{aligned}$$

puisque $S_{n+1} = S_n + X_{n+1}$. Remarquons maintenant que X_{n+1} est indépendante de X_1, \dots, X_n , donc de S_0, \dots, S_n . On obtient alors

$$\begin{aligned} \mathbb{P}(S_{n+1} = a_{n+1} | S_n = a_n, S_{n-1} = a_{n-1}, \dots, S_1 = a_1, S_0 = a_0) \\ = \frac{\mathbb{P}(X_{n+1} = a_{n+1} - a_n) \mathbb{P}(S_n = a_n, S_{n-1} = a_{n-1}, \dots, S_1 = a_1, S_0 = a_0)}{\mathbb{P}(S_n = a_n, S_{n-1} = a_{n-1}, \dots, S_1 = a_1, S_0 = a_0)} \\ = \mathbb{P}(X_{n+1} = a_{n+1} - a_n). \end{aligned}$$

Par ailleurs, par un raisonnement analogue on a

$$\begin{aligned} \mathbb{P}(S_{n+1} = a_{n+1} | S_n = a_n) &= \frac{\mathbb{P}(S_{n+1} = a_{n+1}, S_n = a_n)}{\mathbb{P}(S_n = a_n)} \\ &= \frac{\mathbb{P}(X_{n+1} = a_{n+1} - a_n, S_n = a_n)}{\mathbb{P}(S_n = a_n)} \\ &= \frac{\mathbb{P}(X_{n+1} = a_{n+1} - a_n) \mathbb{P}(S_n = a_n)}{\mathbb{P}(S_n = a_n)} \\ &= \mathbb{P}(X_{n+1} = a_{n+1} - a_n), \end{aligned}$$

d'où le résultat. □

Proposition 8.7. *Pour tout $n > m$ entiers positifs $S_n - S_m$ est indépendant de (S_0, S_1, \dots, S_m) .*

Proof. Comme les (X_n) sont indépendantes, on a directement que

$$S_n - S_m = \sum_{k=m+1}^n X_k,$$

qui est indépendant de $(S_0 = x, S_1 = x + X_1, \dots, S_m = x + \sum_{k=1}^m X_k)$. □

On a un résultat analogue à la propriété de Markov lorsqu'on regarde deux pas de temps plus éloignés.

Corollary 8.3. *Pour tout $n > m$ entiers positifs et pour tous $(a_0, a_1, \dots, a_m, a_n) \in \mathbb{Z}^{m+2}$, on a*

$$\begin{aligned} \mathbb{P}(S_n = a_n | S_m = a_m, S_{m-1} = a_{m-1}, \dots, S_1 = a_1, S_0 = a_0) \\ = \mathbb{P}(S_n = a_n | S_m = a_m). \end{aligned}$$

Proof. Par un raisonnement analogue à la preuve de la propriété de Markov, on obtient

$$\begin{aligned}
& \mathbb{P}(S_n = a_n | S_m = a_m, S_{m-1} = a_{m-1}, \dots, S_1 = a_1, S_0 = a_0) \\
&= \frac{\mathbb{P}(S_n = a_n, S_m = a_m, S_{m-1} = a_{m-1}, \dots, S_1 = a_1, S_0 = a_0)}{\mathbb{P}(S_m = a_m, S_{m-1} = a_{m-1}, \dots, S_1 = a_1, S_0 = a_0)} \\
&= \frac{\mathbb{P}(S_n - S_m = a_n - a_m, S_m = a_m, S_{m-1} = a_{m-1}, \dots, S_1 = a_1, S_0 = a_0)}{\mathbb{P}(S_m = a_m, S_{m-1} = a_{m-1}, \dots, S_1 = a_1, S_0 = a_0)} \\
&= \frac{\mathbb{P}(S_n - S_m = a_n - a_m) \mathbb{P}(S_m = a_m, S_{m-1} = a_{m-1}, \dots, S_1 = a_1, S_0 = a_0)}{\mathbb{P}(S_m = a_m, S_{m-1} = a_{m-1}, \dots, S_1 = a_1, S_0 = a_0)} \\
&= \mathbb{P}(S_n - S_m = a_n - a_m) \\
&= \mathbb{P}(S_n = a_n | S_m = a_m),
\end{aligned}$$

en utilisant la propriété des accroissements indépendants. \square

La seconde propriété dite de stationnarité signifie qu'on peut changer l'origine du repère sans influence sur le comportement futur de la marche.

Proposition 8.8 (Stationnarité). *Pour tout $n > m$ entiers positifs et $(a, b) \in \mathbb{Z}^2$, on a*

$$\mathbb{P}(S_n = b | S_m = a) = \mathbb{P}(S_{n-m} = b - a | S_0 = 0).$$

Proof. En utilisant l'indépendance des accroissements, on obtient

$$\begin{aligned}
\mathbb{P}(S_n = b | S_m = a) &= \frac{\mathbb{P}(S_n = b, S_m = a)}{\mathbb{P}(S_m = a)} \\
&= \frac{\mathbb{P}(S_n - S_m = b - a, S_m = a)}{\mathbb{P}(S_m = a)} \\
&= \frac{\mathbb{P}(S_n - S_m = b - a) \mathbb{P}(S_m = a)}{\mathbb{P}(S_m = a)} \\
&= \mathbb{P}(S_n - S_m = b - a) \\
&= \mathbb{P}\left(\sum_{k=m+1}^n X_k = b - a\right) \\
&= \mathbb{P}\left(\sum_{k=1}^{n-m} X_k = b - a\right) \\
&= \mathbb{P}(S_{n-m} = b - a | S_0 = 0),
\end{aligned}$$

car les (X_n) sont indépendantes et de même loi. \square

8.2.2 Premier retour en 0

On suppose maintenant que la marche part de 0 : $S_0 = 0$ et on cherche à savoir si la marche reviendra en 0 presque sûrement. Commençons par remarquer que si $S_n = 0$ alors n est nécessairement pair.

Proposition 8.9. *Pour tout $n \in \mathbb{N}$, on a*

$$\mathbb{P}(S_{2n} = 0) = \frac{1}{4^n} \binom{2n}{n}.$$

Proof. Cette probabilité correspond au nombre de chemins de $(0, 0)$ à $(2n, 0)$ divisé par le nombre total de chemins de longueur $2n$, puisque tous les chemins sont équiprobables. On a donc directement $\mathbb{P}(S_{2n} = 0) = \frac{\binom{2n}{n}}{4^n}$. \square

On s'intéresse maintenant au premier instant de retour en 0. On le note T_0 . Il s'agit donc de la variable aléatoire

$$T_0 = \inf\{n \geq 1; S_n = 0\},$$

avec la convention que $\inf \emptyset = +\infty$. C'est donc une variable aléatoire discrète à valeurs dans $\mathbb{N}^* \cup \{+\infty\}$. On peut calculer explicitement la loi de T_0 .

Proposition 8.10. *Pour tout $n \in \mathbb{N}^*$, on a*

$$\mathbb{P}(T_0 = 2n) = \frac{(2n-2)!}{2^{2n-1} n! (n-1)!}.$$

Proof. L'événement $(T_0 = 2n)$ correspond à $(S_2 \neq 0, \dots, S_{2n-2} \neq 0, S_{2n} = 0)$ puisqu'alors $2n$ est le premier temps où la marche retourne en 0. En particulier, entre l'instant 0 et l'instant $2n$ la marche ne change pas de signe, et garde le signe de S_1 . Ainsi, on a

$$\begin{aligned} \mathbb{P}(T_0 = 2n) &= \mathbb{P}(S_1 = 1, S_2 > 0, \dots, S_{2n-2} > 0, S_{2n} = 0) \\ &\quad + \mathbb{P}(S_1 = -1, S_2 < 0, \dots, S_{2n-2} < 0, S_{2n} = 0). \end{aligned}$$

Par symétrie, ces deux probabilités sont égales. De plus, si $S_{2n-2} > 0$ et $S_{2n} = 0$ alors nécessairement on a $S_{2n-1} = 1$. Ainsi il vient

$$\begin{aligned}
\mathbb{P}(T_0 = 2n) &= 2\mathbb{P}(S_1 = 1, S_2 > 0, \dots, S_{2n-2} > 0, S_{2n-1} = 1, S_{2n} = 0) \\
&= 2\mathbb{P}(S_{2n} = 0 | S_1 = 1, S_2 > 0, \dots, S_{2n-2} > 0, S_{2n-1} = 1) \\
&\quad \times \mathbb{P}(S_1 = 1, S_2 > 0, \dots, S_{2n-2} > 0, S_{2n-1} = 1) \\
&= 2\mathbb{P}(S_{2n} = 0 | S_{2n-1} = 1) \mathbb{P}(S_1 = 1, S_2 > 0, \dots, S_{2n-2} > 0, S_{2n-1} = 1) \\
&= 2\mathbb{P}(S_1 = -1 | S_0 = 0) \mathbb{P}(S_1 = 1, S_2 > 0, \dots, S_{2n-2} > 0, S_{2n-1} = 1) \\
&= \mathbb{P}(S_1 = 1, S_2 > 0, \dots, S_{2n-2} > 0, S_{2n-1} = 1),
\end{aligned}$$

en utilisant la propriété de Markov, la propriété de stationnarité puis le fait que

$$\mathbb{P}(S_1 = -1 | S_0 = 0) = \mathbb{P}(X_1 = -1) = \frac{1}{2}.$$

Cette dernière probabilité correspond au nombre de chemins de $(1, 1)$ à $(2n-1, 1)$ ne touchant pas 0 divisée par le nombre total de chemins de longueur $2n-1$. En utilisant le principe de réflexion, on obtient

$$\begin{aligned}
\mathbb{P}(T_0 = 2n) &= \frac{\binom{2n-2}{n-1} - \binom{2n-2}{n}}{2^{2n-1}} \\
&= \frac{1}{2^{2n-1}} \left(\frac{(2n-2)!}{(n-1)!(n-1)!} - \frac{(2n-2)!}{n!(n-2)!} \right) \\
&= \frac{1}{2^{2n-1}} \frac{(2n-2)!}{(n-1)!(n-2)!} \left(\frac{1}{n-1} - \frac{1}{n} \right) \\
&= \frac{(2n-2)!}{2^{2n-1} n! (n-1)!},
\end{aligned}$$

d'où le résultat. □

Nous pouvons maintenant énoncer le résultat principal de cette partie.

Theorem 8.2. *La marche aléatoire partant de 0 revient en 0 en temps fini avec probabilité 1, i.e.*

$$\mathbb{P}(T_0 < +\infty) = 1.$$

Proof. On a $\mathbb{P}(T_0 < +\infty) = \sum_{n=1}^{\infty} \mathbb{P}(T_0 = 2n)$, il s'agit donc d'identifier la somme de cette série. Pour cela, on exprime différemment la probabilité de l'événement $(T_0 = 2n)$, en

repartant de l'expression du nombre de chemins de $(1, 1)$ à $(2n - 1, 1)$ ne touchant pas 0 divisée par le nombre total de chemins de longueur $2n - 1$. Ce nombre de chemins est égal à la différence entre le nombre de chemins de $(1, 1)$ à $(2n - 1, 1)$ et le nombre de chemins de $(1, 1)$ à $(2n - 1, -1)$ d'après le principe de réflexion. On obtient ainsi :

$$\begin{aligned}
\mathbb{P}(T_0 = 2n) &= \mathbb{P}(S_1 = 1, S_{2n-1} = 1) - \mathbb{P}(S_1 = 1, S_{2n-1} = -1) \\
&= \mathbb{P}(S_{2n-1} = 1 | S_1 = 1) \mathbb{P}(S_1 = 1) - \mathbb{P}(S_{2n-1} = -1 | S_1 = 1) \mathbb{P}(S_1 = 1) \\
&= \frac{1}{2} \mathbb{P}(S_{2n-2} = 0 | S_0 = 0) - \frac{1}{2} \mathbb{P}(S_{2n-2} = -2 | S_0 = 0) \\
&= \frac{1}{2} \mathbb{P}(S_{2n-2} = 0) - \frac{1}{2} \mathbb{P}(S_{2n-2} = -2),
\end{aligned}$$

par la propriété de stationnarité et puisque $S_0 = 0$. Par ailleurs, on a

$$\begin{aligned}
\mathbb{P}(S_{2n} = 0) &= \mathbb{P}(S_{2n} = 0, S_{2n-2} = -2) + \mathbb{P}(S_{2n} = 0, S_{2n-2} = 0) + \mathbb{P}(S_{2n} = 0, S_{2n-2} = 2) \\
&= 2\mathbb{P}(S_{2n} = 0, S_{2n-2} = -2) + \mathbb{P}(S_{2n} = 0, S_{2n-2} = 0),
\end{aligned}$$

par symétrie. Il vient

$$\begin{aligned}
\mathbb{P}(S_{2n} = 0) &= 2\mathbb{P}(S_{2n} = 0 | S_{2n-2} = -2) \mathbb{P}(S_{2n-2} = -2) + \mathbb{P}(S_{2n} = 0 | S_{2n-2} = 0) \mathbb{P}(S_{2n-2} = 0) \\
&= 2\mathbb{P}(S_2 = 2 | S_0 = 0) \mathbb{P}(S_{2n-2} = -2) + \mathbb{P}(S_2 = 0 | S_0 = 0) \mathbb{P}(S_{2n-2} = 0) \\
&= 2\frac{1}{4} \mathbb{P}(S_{2n-2} = -2) + \frac{1}{2} \mathbb{P}(S_{2n-2} = 0) \\
&= \frac{1}{2} \mathbb{P}(S_{2n-2} = -2) + \frac{1}{2} \mathbb{P}(S_{2n-2} = 0),
\end{aligned}$$

où on a utilisé le fait que

$$\begin{aligned}
\mathbb{P}(S_2 = 2 | S_0 = 0) &= \mathbb{P}(X_1 + X_2 = 0) \\
&= \mathbb{P}(X_1 = -1; X_2 = 1) + \mathbb{P}(X_1 = 1; X_2 = -1) \\
&= \mathbb{P}(X_1 = -1) \mathbb{P}(X_2 = 1) + \mathbb{P}(X_1 = 1) \mathbb{P}(X_2 = -1) \\
&= \frac{1}{2}.
\end{aligned}$$

Donc on a $\mathbb{P}(S_{2n-2} = -2) = 2\mathbb{P}(S_{2n} = 0) - \mathbb{P}(S_{2n-2} = 0)$.

Revenons maintenant à T_0 . On obtient

$$\begin{aligned}\mathbb{P}(T_0 = 2n) &= \frac{1}{2}\mathbb{P}(S_{2n-2} = 0) - \frac{1}{2}\mathbb{P}(S_{2n-2} = -2) \\ &= \frac{1}{2}\mathbb{P}(S_{2n-2} = 0) - \mathbb{P}(S_{2n} = 0) + \frac{1}{2}\mathbb{P}(S_{2n-2} = 0) \\ &= \mathbb{P}(S_{2n-2} = 0) - \mathbb{P}(S_{2n} = 0).\end{aligned}$$

Ainsi, la probabilité de l'événement $(T_0 = +\infty)$ peut s'écrire à l'aide d'une somme télescopique :

$$\begin{aligned}\mathbb{P}(T_0 = +\infty) &= 1 - \sum_{n=1}^{\infty} \mathbb{P}(T_0 = 2n) \\ &= 1 - \lim_{N \rightarrow \infty} \sum_{n=1}^N \mathbb{P}(T_0 = 2n) \\ &= 1 - \lim_{N \rightarrow \infty} \sum_{n=1}^N (\mathbb{P}(S_{2n-2} = 0) - \mathbb{P}(S_{2n} = 0)) \\ &= 1 - \lim_{N \rightarrow \infty} (\mathbb{P}(S_0 = 0) - \mathbb{P}(S_{2N} = 0)) \\ &= \lim_{N \rightarrow \infty} \mathbb{P}(S_{2N} = 0) \\ &= \lim_{N \rightarrow \infty} \frac{1}{4^N} \binom{2N}{N}.\end{aligned}$$

En utilisant la formule de Stirling, on obtient l'équivalence

$$\frac{1}{4^N} \binom{2N}{N} = \frac{1}{4^N} \frac{(2N)!}{N!N!} \sim \frac{1}{4^N} \frac{(2N)^{2N} \sqrt{4\pi N} e^{-2N}}{(N^N \sqrt{2\pi N} e^{-N})^2} = \frac{1}{\sqrt{\pi N}},$$

donc on a bien $\mathbb{P}(T_0 = +\infty) = 0$. □

8.2.3 Premier passage en $a \neq 0$

On peut obtenir par une méthode similaire la loi du temps d'atteinte de n'importe quel autre point. Soit $a \in \mathbb{Z}^*$. Le temps d'atteinte de a est la variable aléatoire

$$T_a = \inf\{n \geq 1; S_n = a\}.$$

Theorem 8.3. La marche aléatoire partant de 0 atteint tout état $a \in \mathbb{Z}$ en temps fini avec probabilité 1. Plus précisément, si $a \neq 0$, on a $\mathbb{P}(T_a = n) = 0$ si n et a n'ont pas la même parité ou si $|a| > n$, et sinon

$$\mathbb{P}(T_a = n) = \frac{|a|}{n2^n} \binom{n}{\frac{n+a}{2}}.$$

Proof. On suit les mêmes étapes que ce qu'on a fait pour T_0 . Les contraintes de parité et de longueur de chemin ont déjà été vues. Si n et a ont la même parité et que $|a| > n$, alors l'événement $(T_a = n)$ correspond à $(S_1 \neq a, \dots, S_{n-1} \neq a, S_n = a)$ puisqu'alors n est le premier temps où la marche atteint a . Supposons sans perte de généralité que $a > 0$. Donc la probabilité $\mathbb{P}(T_a = n)$ est égale au nombre de chemins de $(0, 0)$ à (n, a) ne touchant pas a avant n divisée par le nombre total de chemins de longueur n . Or, ce nombre de chemin est égal

- au nombre de chemins de $(0, 0)$ à $(n-1, a-1)$ ne touchant pas a
- au nombre de chemins de $(0, 0)$ à $(n-1, a-1)$ - le nombre de chemins de $(0, 0)$ à $(n-1, a-1)$ touchant a
- au nombre de chemins de $(0, 0)$ à $(n-1, a-1)$ - le nombre de chemins de $(0, -a)$ à $(n-1, -1)$ touchant 0
- au nombre de chemins de $(0, 0)$ à $(n-1, a-1)$ - le nombre de chemins de $(0, -a)$ à $(n-1, 1)$
- au nombre de chemins de $(0, 0)$ à $(n-1, a-1)$ - le nombre de chemins de $(0, 0)$ à $(n-1, a+1)$

Ainsi, on a

$$\begin{aligned} \mathbb{P}(T_a = n) &= \frac{\binom{n-1}{\frac{n-1}{2} + \frac{a-1}{2}} - \binom{n-1}{\frac{n-1}{2} + \frac{1+a}{2}}}{2^n} \\ &= \frac{1}{2^n} \left(\frac{(n-1)!}{\left(\frac{n+a-2}{2}\right)! \left(\frac{n-a}{2}\right)!} - \frac{(n-1)!}{\left(\frac{n+a}{2}\right)! \left(\frac{n-a-2}{2}\right)!} \right) \\ &= \frac{1}{2^n} \frac{(n-1)!}{\left(\frac{n+a-2}{2}\right)! \left(\frac{n-a-2}{2}\right)!} \left(\frac{1}{\frac{n-a}{2}} - \frac{1}{\frac{n+a}{2}} \right) \\ &= \frac{a}{2^n} \frac{(n-1)!}{\left(\frac{n+a}{2}\right)! \left(\frac{n-a}{2}\right)!} \\ &= \frac{a}{n2^n} \binom{n}{\frac{n+a}{2}} \end{aligned}$$

d'où le résultat.

Nous démontrerons dans la partie suivante que $\mathbb{P}(T_a < +\infty) = 1$. □

En particulier, la marche partant de 0 va atteindre tout point de \mathbb{Z} avec probabilité 1, donc elle est presque sûrement non bornée.

8.3 La ruine de la joueuse

On s'intéresse maintenant à un problème de ruine. Une joueuse dispose d'un capital initial x . Elle joue à pile ou face avec une pièce équilibrée, et gagne 1 si elle obtient pile, perd 1 si elle obtient face. La joueuse s'est fixé un objectif de gain $a \geq x$ et un plancher de perte $b \leq x$. Elle joue jusqu'à ce que sa richesse atteigne a ou b .

On modélise la fortune de la joueuse par une marche aléatoire $(S_n)_{n \in \mathbb{N}}$, avec $S_0 = x$ et $S_n = S_0 + \sum_{k=1}^n X_k$, où X_k représente le gain du k -ème tirage.

On sait qu'atteindre a ou b partant de x est équivalent à atteindre $a - x$ ou $b - x$ partant de 0. Le jeu s'arrête donc presque sûrement au bout d'un temps fini.

On note p_x la probabilité de ruine partant d'un capital initial x , et R_x l'événement *être ruinée en partant de x* , c'est-à-dire

$$p_x = \mathbb{P}(T_b < T_a) = \mathbb{P}(R_x).$$

On peut remarquer directement que $p_a = 0$ et $p_b = 1$ puisque dans ces deux situations, le jeu ne démarre pas, la ruine est impossible dans le premier cas, et certaine dans le second.

Nous allons obtenir une formule de récurrence sur les p_x . Si $b < x < a$, on a par la formule des probabilités totales

$$\begin{aligned} p_x &= \mathbb{P}(R_x) \\ &= \mathbb{P}(R_x | S_1 = x+1) \mathbb{P}(S_1 = x+1) + \mathbb{P}(R_x | S_1 = x-1) \mathbb{P}(S_1 = x-1) \\ &= \mathbb{P}(R_{x+1}) \frac{1}{2} + \mathbb{P}(R_{x-1}) \frac{1}{2}, \end{aligned}$$

en utilisant la stationnarité. On obtient ainsi que $(p_x)_{a < x < b}$ est une suite récurrente linéaire¹ d'ordre 2, de polynôme caractéristique $X^2 - 2X + 1 = (X - 1)^2$. Ainsi $p_x = \alpha + \beta x$.

¹Une suite $(u_n)_{n \in \mathbb{N}}$ est dite **récurrente linéaire d'ordre 2** s'il existe $(a, b) \in \mathbb{R} \times \mathbb{R}^*$ tel que, pour entier n , on a

$$u_{n+2} = au_{n+1} + bu_n.$$

Le polynôme $X^2 - aX + b$ est appelé le **polynôme caractéristique** de la suite. Si le polynôme caractéristique admet deux racines simples r_1 et r_2 , alors il existe deux réels α et β tels que pour tout entier n ,

$$u_n = \alpha r_1^n + \beta r_2^n.$$

Si le polynôme caractéristique admet une racine double r , alors il existe deux réels α et β tels que pour tout

On identifie α et β avec les conditions extrémales $p_a = 0$ et $p_b = 1$

$$\begin{cases} p_a = 0 \\ p_b = 1 \end{cases} = \begin{cases} \alpha + \beta a \\ \alpha + \beta b \end{cases} \Leftrightarrow \begin{cases} \alpha \\ \beta \end{cases} = \begin{cases} \frac{a}{a-b}, \\ -\frac{1}{a-b}. \end{cases}$$

Ainsi la probabilité de ruine partant de x vaut

$$p_x = \frac{a - x}{a - b}.$$

Un calcul analogue en cherchant la probabilité q_x de faire fortune partant d'une richesse x conduit à $q_x = \frac{x-b}{a-b} = 1 - p_x$, autrement dit il n'y a pas d'autre issue possible : le jeu va s'arrêter avec probabilité 1. Si on fait tendre b vers $-\infty$, on trouve que $\mathbb{P}(T_a < \infty | x_0 = a) = 1$. Donc partant de n'importe quel point, la marche atteindra n'importe quel autre point avec probabilité 1.

On peut montrer avec un raisonnement analogue que la durée du jeu d_x partant de x est solution de l'équation récurrente linéaire avec second membre²

$$d_x = \frac{1}{2}(d_{x+1} + d_{x-1}) + 1,$$

avec $d_a = d_b = 0$.

En effet, il est clair que $d_a = d_b = 0$, et si $a < x < b$, alors on a $d_x = \mathbb{E}(T_a \wedge T_b)$. Attention, la variable aléatoire $T_a \wedge T_b$ est à valeurs dans $\mathbb{N} \cup \{+\infty\}$. Si $\mathbb{P}(T_a \wedge T_b = +\infty | S_0 = x) > 0$, alors $\mathbb{E}(T_a \wedge T_b) = +\infty$. Par ailleurs, partant de $x + 1$, on a

$$\begin{aligned} \mathbb{P}(T_a \wedge T_b = +\infty | S_0 = x + 1) &\geq \mathbb{P}(T_a \wedge T_b = +\infty, S_1 = x | S_0 = x + 1) \\ &= \mathbb{P}(T_a \wedge T_b = +\infty | S_1 = x, S_0 = x + 1) \mathbb{P}(S_1 = x | S_0 = x + 1) \\ &= \frac{1}{2} \mathbb{P}(T_a \wedge T_b = +\infty | S_0 = x) \end{aligned}$$

par les propriétés de Markov et de stationnarité. Donc $\mathbb{P}(T_a \wedge T_b = +\infty | S_0 = x + 1) > 0$. Ainsi, les membres de gauche et de droite de l'équation $d_x = \frac{1}{2}(d_{x+1} + d_{x-1}) + 1$ valent tous

entier n ,

$$u_n = (\alpha + \beta n)r^n.$$

²Une suite $(u_n)_{n \in \mathbb{N}}$ est dite **récurrente linéaire d'ordre 2 avec second membre** s'il existe $(a, b) \in \mathbb{R} \times \mathbb{R}^*$ et une fonction f à valeurs dans \mathbb{R} tels que, pour entier n , on a

$$u_{n+2} = au_{n+1} + bu_n + f(n).$$

L'équation $u_{n+2} = au_{n+1} + bu_n$ s'appelle alors l'équation homogène associée. Les solutions de l'équation avec second membre sont la somme de la solution générique de l'équation homogène et d'une solution particulière de l'équation avec second membre.

les deux $+\infty$, et l'égalité est vraie. On montrerait de même que $\mathbb{P}(T_a \wedge T_b = +\infty | S_0 = x) > 0$ implique $\mathbb{P}(T_a \wedge T_b = +\infty | S_0 = x - 1) > 0$, et par contraposée et décalage d'indices que $\mathbb{P}(T_a \wedge T_b = +\infty | S_0 = x) = 0$ implique $\mathbb{P}(T_a \wedge T_b = +\infty | S_0 = x + 1) = \mathbb{P}(T_a \wedge T_b = +\infty | S_0 = x - 1) = 0$. Plaçons nous dans ce dernier cas. On a donc

$$\begin{aligned}
d_x &= \mathbb{E}(T_a \wedge T_b) \\
&= \sum_{k=1}^{\infty} k \mathbb{P}(T_a \wedge T_b = k | S_0 = x) \\
&= \sum_{k=1}^{\infty} k (\mathbb{P}(T_a \wedge T_b = k | S_1 = x + 1) \mathbb{P}(S_1 = x + 1) + \mathbb{P}(T_a \wedge T_b = k | S_1 = x - 1) \mathbb{P}(S_1 = x - 1)) \\
&= \frac{1}{2} \sum_{k=1}^{\infty} k \mathbb{P}(T_a \wedge T_b = k - 1 | S_0 = x + 1) + \frac{1}{2} \sum_{k=1}^{\infty} k \mathbb{P}(T_a \wedge T_b = k - 1 | S_0 = x - 1) \\
&= \frac{1}{2} \sum_{j=0}^{\infty} (j + 1) \mathbb{P}(T_a \wedge T_b = j | S_0 = x + 1) + \frac{1}{2} \sum_{j=0}^{\infty} (j + 1) \mathbb{P}(T_a \wedge T_b = j | S_0 = x - 1) \\
&= \frac{1}{2} \sum_{j=0}^{\infty} j \mathbb{P}(T_a \wedge T_b = j | S_0 = x + 1) + \frac{1}{2} \sum_{j=0}^{\infty} j \mathbb{P}(T_a \wedge T_b = j | S_0 = x - 1) + 1 \\
&= 1 + \frac{1}{2} \mathbb{E}(T_a \wedge T_b | S_0 = x + 1) + \frac{1}{2} \mathbb{E}(T_a \wedge T_b | S_0 = x - 1) \\
&= 1 + \frac{1}{2} (d_{x+1} + d_{x-1}).
\end{aligned}$$

% Cherchons une solution particulière sous la forme cn^2 (puisque les constantes et les multiples de n sont solutions de l'équation homogène) :

$$\begin{aligned}
cn^2 &= \frac{1}{2} (c(n + 1)^2 + c(n - 1)^2) + 1 \Leftrightarrow cn^2 = \frac{1}{2} (cn^2 + c + 2cn + cn^2 + c - 2cn) + 1 \\
&\Leftrightarrow cn^2 = cn^2 + c + 1 \\
&\Leftrightarrow c = -1.
\end{aligned}$$

% Ainsi, d est la somme d'une solution homogène et d'une solution particulière, donc

$$d_x = \alpha + \beta x - x^2.$$

En utilisant $d_a = d_b = 0$, on trouve α et β

$$\begin{cases} d_a = 0 \\ d_b = 0 \end{cases} = \begin{cases} \alpha + \beta a - a^2 \\ \alpha + \beta b - b^2 \end{cases} \Leftrightarrow \begin{cases} \alpha = -ab, \\ \beta = a + b. \end{cases}$$

Finalement, on obtient que $d_x = -ab + x(a + b) - x^2$.

En particulier, si $x = 0$ et b tend vers $-\infty$, on obtient que $\mathbb{E}[T_a] = +\infty$. On sait que la marche partant de 0 atteindra a presque sûrement en temps fini, cependant ce temps est en moyenne infini.

8.4 De la marche aléatoire vers le mouvement brownien

On s'intéresse maintenant à une autre forme de comportement asymptotique de la marche aléatoire. Nous allons la renormaliser pour la contraindre à rester dans l'intervalle de temps $[0, 1]$. Pour cela, pour tout $N \in \mathbb{N}^*$, on pose

$$B_t^N = \frac{1}{\sqrt{N}} S_{\lfloor Nt \rfloor} = \frac{1}{\sqrt{N}} \sum_{k=1}^{\lfloor Nt \rfloor} X_k.$$

Alors la suite (B_t^N) converge en loi lorsque N tend vers l'infini, et c'est également le cas lorsqu'on prend plusieurs temps t_i .

Proposition 8.11. *Pour tous $t_0 = 0 < t_1 < \dots < t_p$, le vecteur $(B_{t_1}^N, \dots, B_{t_p}^N)$ converge en loi vers un vecteur $(B_{t_1}, \dots, B_{t_p})$ tel que pour tout $i \in \{1, \dots, p\}$, $B_{t_i} - B_{t_{i-1}}$ est indépendant de $B_{t_{i-1}}$ et suit une loi normale $\mathcal{N}(0, t_i - t_{i-1})$*

Proof. On a

$$B_{t_i}^N - B_{t_{i-1}}^N = \frac{1}{\sqrt{N}} S_{\lfloor Nt_i \rfloor} - \frac{1}{\sqrt{N}} S_{\lfloor Nt_{i-1} \rfloor} = \frac{1}{\sqrt{N}} \sum_{k=\lfloor Nt_{i-1} \rfloor + 1}^{\lfloor Nt_i \rfloor} X_k$$

donc les accroissements du vecteur $(B_{t_1}^N, \dots, B_{t_p}^N)$ sont indépendants. Il suffit maintenant de montrer qu'ils convergent vers la loi normale. On a

$$B_{t_i}^N - B_{t_{i-1}}^N = \sqrt{\frac{\lfloor Nt_i \rfloor - \lfloor Nt_{i-1} \rfloor}{N}} \frac{1}{\sqrt{\lfloor Nt_i \rfloor - \lfloor Nt_{i-1} \rfloor}} \sum_{k=\lfloor Nt_{i-1} \rfloor + 1}^{\lfloor Nt_i \rfloor} X_k \rightarrow \sqrt{t_i - t_{i-1}} \mathcal{N}(0, 1),$$

par le théorème central limite appliqué à la suite (X_k) . □

Ce résultat permet de poser la définition suivante.

Définition 8.3 (Mouvement Brownien). On appelle mouvement Brownien standard, tout processus $(B_t)_{t \geq 0}$ satisfaisant les trois points suivants

1. $B_0 = 0$,
2. $\forall t \geq s, B_t - B_s$ independant de $(B_r)_{r \leq s}$,
3. $\forall t \geq s, B_t - B_s \sim \mathcal{N}(0, t - s)$.

Le théorème précédent ne donne pas l'existence d'un tel processus, mais si un tel processus existe, il donne la convergence des lois marginales fini-dimensionnelles de la marche aléatoire renormalisée vers les lois marginales fini-dimensionnelles du mouvement Brownien. On admettra qu'il existe bien un mouvement Brownien. On peut montrer que la marche aléatoire converge pour la norme infinie vers le mouvement Brownien dans l'espace des variables aléatoires à valeurs dans les fonctions continues de $[0, 1]$ dans \mathbb{R} . Ce résultat est connu sous le nom de théorème de Donsker. Pour simuler un mouvement Brownien, on simule en fait une marche aléatoire et on la renormalise comme ci-dessus avec N assez grand.