



PRÉCIS DE SIMULATION

P. Del Moral

Centre INRIA Bordeaux Sud-Ouest

∩ Institut de Mathématiques de Bordeaux

Université Bordeaux I, 351, cours de la Libération

33405 Talence, France

Table des matières

1	La notion de variable aléatoire	5
1.1	Introduction	5
1.2	Le principe d'indépendance	8
1.3	Les générateurs aléatoires	9
1.4	Quelques lois usuelles	11
1.4.1	Les lois uniformes	11
1.4.2	Les lois d'événements composés	13
1.4.3	Les lois géométriques	17
1.4.4	Les lois discrètes	20
1.4.5	Les lois continues	23
2	Quelques outils de simulation	27
2.1	La méthode d'inversion	27
2.1.1	Le principe de base	27
2.1.2	Retour aux lois géométriques	29
2.1.3	Les lois de Cauchy	30
2.1.4	Les lois de Rayleigh-Weibull	32
2.2	Changements de variables	33
2.2.1	Introduction	33
2.2.2	Lois uniformes sur des cubes	34
2.2.3	Lois uniformes sur des surfaces	36
2.2.4	Lois uniformes sur des disques	41
2.2.5	L'algorithme Box-Muller	44
2.3	Méthode d'acceptation-rejet	49
2.3.1	Les lois de référence	49
2.3.2	Les taux d'acceptation	50
2.3.3	Vérification mathématique	51
2.4	Files d'attentes exponentielles	53
2.4.1	La loi Gamma	54
2.4.2	La loi de Poisson	56
2.4.3	La statistique d'ordre uniforme	57

3	Chaînes de Markov discrètes	61
3.1	Introduction	61
3.2	Chaînes de Markov discrètes	62
3.2.1	Semigroupes de transitions	64
3.2.2	Processus historique	66
3.2.3	Interprétation matricielle	67
3.3	Quelques Exemples	71
3.3.1	Files d'attentes	71
3.3.2	Modèle d'urne	72
3.3.3	Marche aléatoire sur \mathbb{Z}	72
3.3.4	Marche aléatoire sur \mathbb{Z}^d	74
3.3.5	Marche aléatoire arrêtée	74
3.3.6	Processus de branchements	75
4	Chaînes de Markov abstraites	79
4.1	Description des modèles	79
4.1.1	Semigroupe des transitions	80
4.1.2	Équations de Chapman-Kolmogorov	82
4.1.3	Processus historique	83
4.2	Chaînes linéaires et gaussiennes	84
4.2.1	Formulation canonique	85
4.2.2	Formulation dynamique	85
4.3	Processus de Poisson	86
4.4	Évolutions dans des milieux absorbants	87
5	Chaînes de Markov non linéaires	91
5.1	Introduction	91
5.2	Description des modèles	92
5.3	Interprétations particulières en champ moyen	93
5.4	Champs moyens de type gaussien	94
5.5	Modèles simplifiés de gaz de McKean	96
5.6	Flots de mesures de Feynman-Kac	96
5.6.1	Description des modèles	96
5.6.2	Chaînes de Markov non linéaires	99
5.6.3	Champs moyens de type évolutionnaire	100
6	L'équation de la chaleur	103
6.1	Les fluctuations browniennes	103
6.2	La loi des grands nombres	106
6.3	Marches aléatoires	108
6.4	L'équation de la chaleur	109
6.5	Une formulation faible	110

7	Dynamiques de population avec branchements	113
7.1	Processus de branchements spatio-temporels	113
7.2	Algorithme génétique	115
7.2.1	Sélection/Adaptation	117
7.2.2	Mutation/Exploration	117
7.3	Modèles d'arbres généalogiques	119
7.3.1	Modèles non homogènes	119
7.3.2	Modèles trajectoriels	120
7.4	Chaînes renforcées	123

Chapitre 1

La notion de variable aléatoire

1.1 Introduction

Le mot latin *aléa*, et le mot arabe *hasard* se réfèrent tous deux au “jeu de dés”. Cette métaphore antique transmet ainsi l’idée que le résultat du lancer de dés est totalement inattendu et imprévisible. La formalisation mathématique de cette notion de hasard correspond à la notion de variable aléatoire. Ce nom probabiliste est un vrai pléonasme mathématique combinant la variabilité d’une quantité et les fluctuations aléatoires des résultats observés.

L’archétype des variables aléatoires est le simple jeu du pile ou face. Le résultat X du lancer d’une pièce de monnaie est clairement l’une des deux faces, mais il est bien difficile de prévoir laquelle. Si l’on désigne par le nombre 1 le côté pile, et par le nombre 2 le côté face, les valeurs possibles du lancer sont données un ensemble réduit à deux points $\{1, 2\}$. Lorsque la pièce n’est pas truquée, les deux résultats sont équiprobables. Autrement dit, la probabilité pour que X prenne la valeur 1 coïncide avec la probabilité pour que X prenne la valeur 2. Plus formellement, si l’on note respectivement par $\mathbb{P}(X = 1)$ et $\mathbb{P}(X = 2)$, ces deux probabilités, nous avons

$$\mathbb{P}(X = 1) = \mathbb{P}(X = 2) = 1/2$$

Cette formule mathématique exprime le fait que chaque face se réalise avec une probabilité 1/2.

Le lecteur profane est en droit de se poser au moins un couple de questions naturelles :

- Que signifie ces grandeurs numériques ?
- Existe-t-il des pièces si parfaites ?

Comme nous l’avons souligné dans l’introduction, les modèles probabilistes sont toujours des modèles simplifiés, mais suffisamment précis pour prédire les chances d’obtenir une réalisation donnée. Pour être plus précis, supposons que l’on lance de nombreuses fois la même pièce de monnaie. On note scrupuleusement la proportion de

fois où le cote pile s'est réalisé. Après un certain nombre de lancers, ces proportions se stabilisent autour d'une valeur déterministe comprise dans l'intervalle $[0, 1]$. Ce nombre correspond à la probabilité pour que le résultat du lancer soit "pile".

Une probabilité est donc par essence une notion asymptotique, évaluant le nombre de chances d'obtenir une réalisation d'un évènement aléatoire donné. Si les fréquences de réalisation du coté pile convergent vers $1/100$, lorsqu'on augmente le nombre de lancers, c'est que la pièce utilisée est truquée. Le coté "face" se réalise 99 fois sur 100. Cela ne veut bien évidemment pas dire que sur 100 expériences, on aura exactement 99 fois le coté "face" ! Néanmoins sur de longues séquences, on est à peu sur que le coté "face" se réalisera 99 fois plus souvent.

L'existence d'une pièce parfaite est aussi utopique que l'existence d'une pièce pour laquelle la probabilité de réalisation du coté "pile" est exactement égale à $1/\pi$. Une probabilité rationnelle, pouvant s'exprimer sous forme de fraction d'entiers, est elle plus vraisemblable qu'une quelconque probabilité réelle ?

Cette simple question concernant l'existence d'une telle pièce nous conduit vers une région mathématique encore plus mystérieuse, celle des nombres réels.

La continuité et l'extension infinie de notre univers restent deux des grandes questions des sciences modernes. L'existence des nombres réels semble dictée par l'analyse mathématique. Plus précisément, les nombres rationnels sont d'un usage courant, mais leur ensemble est trop troué pour définir convenablement certains objets limites, tels des vitesses ou des accélérations de mouvement, des volumes ou encore des aires de surfaces planes. Sans ces nombres ils serait par exemple totalement impossible de trouver le périmètre d'un cercle de un mètre diamètre !

Une fois admis l'existence de ces nombres réels, il devient très aisé d'imaginer des phénomènes aléatoires à valeurs dans des espaces continus : lancers de flechettes sur une disque, fluctuations de températures d'un liquide, perturbations de mesures dans des capteurs électroniques, évolutions de cyclones, instants d'arrivée de clients dans une file d'attente, etc.

L'exemple le plus simple de phénomène aléatoire réel est celui du choix d'un point au hasard dans l'intervalle réel $[0, 1]$. Pour poursuivre notre discussion, on note par U le résultat aléatoire d'une telle expérience. La probabilité de choisir un nombre entre 0 et $1/2$ est égale à la probabilité de choisir ce nombre entre $1/2$ et 1

$$\mathbb{P}(U \in [0, 1/2]) = \mathbb{P}(U \in [1/2, 1]) = 1/2$$

De même, la La probabilité de choisir un nombre entre 0 et $1/4$ est égale à la probabilité de choisir ce nombre entre $1/4$ et $1/2$, elle même égale à la probabilité de le choisir entre $1/2$ et $3/4$, ou encore entre $3/4$ et 1

$$\mathbb{P}(U \in [0, 1/4]) = \mathbb{P}(U \in [1/4, 1/2]) = \mathbb{P}(U \in [1/2, 3/4]) = \mathbb{P}(U \in [3/4, 1]) = 1/4$$

Plus généralement, la probabilité de choisir un nombre dans un intervalle donné $[a, b] \subset$

$[0, 1]$, dépend uniquement de la longueur de ce dernier

$$\mathbb{P}(U \in [a, b]) = \int_a^b dx = (b - a)$$

Ainsi la probabilité de choisir U dans $[1/2, 1/2 + 1/1000]$ est de $1/1000$, celle de le choisir dans $[1/2, 1/2 + 1/10000]$ est de $1/10000$. De proche en proche, on montre qu'il est de plus en plus improbable de le choisir de plus en plus proche de $1/2$

$$\mathbb{P}(U \in [1/2, 1/2 + \epsilon]) \xrightarrow{\epsilon \rightarrow 0} \mathbb{P}(U = 1/2) = 0$$

Ce résultat au coeur de la théorie de l'intégration de Lebesgue est assez déroutant. Si l'on mesure une réalisation d'un phénomène uniforme sur $[0, 1]$, tel le choix au hasard d'une pièce de monnaie truquée ou non, on observera clairement la réalisation d'un nombre quelconque entre 0 et 1. Cependant, la probabilité d'observer ce même nombre lors d'une telle expérience est tout simplement nulle ! Ces évènements élémentaires n'ont donc aucune chance de se réaliser. Cette propriété mathématique peut être rapprochée de la philosophie des "matérialistes quantiques" soutenant le fait que "l'Univers n'existe pas indépendamment de tout acte d'observation" (Fritz Rohrlich, *Science* 1983).

Ces pensées sont en partie soutenues par le principe d'incertitude d'Heisenberg énoncé dès 1927. Ce principe d'indétermination est à la base des sciences physiques modernes. Il souligne qu'il est *impossible* de connaître avec précision, et à un instant donné, le couple vitesse-position d'une particule à l'échelle atomique. Plus précisément, les incertitudes de mesures sur l'une des deux composantes sont inversement proportionnelles au degré de connaissance sur la seconde. En ce sens, certains aspects déroutant de la théorie des probabilités sont très probablement le reflet de certaines propriétés naturelles et physiques.

Les variables aléatoires uniformes sur l'intervalle $[0, 1]$, sont le ciment des méthodes de simulation numérique de lois de probabilités complexes. Plus précisément, on peut démontrer que **tout phénomène aléatoire à valeurs réelles s'exprime en terme des variables uniformes sur $[0, 1]$** . Autrement dit, on peut simuler toute loi de probabilité sur un espace multi-dimensionnel \mathbb{R}^d , avec $d \geq 1$, par une suite finie de variables aléatoires uniformes sur $[0, 1]$. Pour s'en convaincre, on peut noter que le simple jeu de pile ou face examiné plus haut, peut être décrit par la formule suivante :

$$X = \begin{cases} 1 & \text{si } U \in [0, 1/2[\\ 2 & \text{si } U \in [1/2, 1] \end{cases}$$

Autrement dit, lorsque l'évènement $U \in [0, 1/2[$ se réalise, on convient que la pièce de monnaie est tombée sur le coté pile, et sur le coté face dans le cas contraire. Par construction, la probabilité d'obtenir le coté "pile" est bien à nouveau égale à $1/2$:

$$\mathbb{P}(X = 1) = \mathbb{P}(U \in [0, 1/2]) = 1/2$$

1.2 Le principe d'indépendance

Les variables aléatoires uniformes sur $[0, 1]$ permettent de construire très simplement des phénomènes aléatoires dans l'espace. Un point $U = (U_1, U_2)$ choisi au hasard dans le carré unité $\mathcal{C} = ([0, 1] \times [0, 1])$ correspond tout simplement au choix de deux variables U_1 et U_2 uniformes et **indépendantes** sur $[0, 1]$. La probabilité pour que U soit choisi dans un petit pavé

$$\mathcal{P} = ([a_1, a_2] \times [b_1, b_2]) \subset \mathcal{C} = ([0, 1] \times [0, 1])$$

est le produit des probabilités pour que chacune des deux coordonnées U_1 et U_2 , tombe respectivement sur les faces $[a_1, a_2]$ et $[b_1, b_2]$. Par conséquent, la probabilité de choisir un point $U = (U_1, U_2)$ au hasard dans le carré \mathcal{P} est alors donnée par l'aire du pavé :

$$\begin{aligned} \mathbb{P}(U \in \mathcal{P}) &= \mathbb{P}(U_1 \in [a_1, a_2] \text{ et } U_2 \in [b_1, b_2]) \\ &= \mathbb{P}(U_1 \in [a_1, a_2]) \times \mathbb{P}(U_2 \in [b_1, b_2]) = (a_2 - a_1) \times (b_2 - b_1) \end{aligned}$$

On peut étendre ces constructions au cube unité

$$\mathcal{V} = ([0, 1] \times [0, 1] \times [0, 1]) = [0, 1]^3$$

Choisir au hasard un point $U = (U_1, U_2, U_3)$ dans ce cube revient trois variables U_1 , U_2 , et U_3 uniformes et **indépendantes** sur $[0, 1]$. La probabilité de choisir ce point dans un pavé

$$\mathcal{P} = ([a_1, a_2] \times [b_1, b_2] \times [c_1, c_2])$$

est alors donnée par l'aire du pavé :

$$\begin{aligned} \mathbb{P}(U \in \mathcal{P}) &= \mathbb{P}(U_1 \in [a_1, a_2] \text{ et } U_2 \in [b_1, b_2] \text{ et } U_3 \in [c_1, c_2]) \\ &= \mathbb{P}(U_1 \in [a_1, a_2]) \times \mathbb{P}(U_2 \in [b_1, b_2]) \times \mathbb{P}(U_3 \in [c_1, c_2]) \\ &= (a_2 - a_1) \times (b_2 - b_1) \times (c_2 - c_1) \end{aligned}$$

Dans notre discussion, nous avons utilisé sans y prêter trop d'attention une des notions fondamentales de la théorie des probabilités : la notion d'**indépendance** entre des évènements aléatoires. Intuitivement, deux évènements aléatoires sont physiquement indépendants lorsque la réalisation de l'un n'a aucune influence sur les possibilités de réalisation du second.

Reprenons l'exemple du jeu de pile ou face. On note X_1 et X_2 les résultats de deux lancers successifs. L'indépendance entre les lancers se traduit par le fait que le résultat du premier lancer n'influe pas sur la réalisation du second. Par exemple, le fait d'avoir "pile" au premier lancer, n'augmente pas les chances d'avoir à nouveau "pile" au second. Les deux évènements

$$A_1 = (X_1 = 0) \quad \text{et} \quad A_2 = (X_2 = 0)$$

sont donc physiquement indépendants. Examinons de plus près la caractérisation mathématique de cette notion d'indépendance. Nous commençons par noter que toutes les réalisations possibles et envisageables sont équiprobables. Plus précisément, nous avons autant de chance de voir deux fois le coté "pile", que deux fois le coté "face", ou encore deux cotés opposés dans un ordre quelconque :

$$\begin{aligned}\mathbb{P}((X_1, X_2) = (0, 0)) &= \mathbb{P}((X_1, X_2) = (0, 1)) \\ &= \mathbb{P}((X_1, X_2) = (1, 0)) = \mathbb{P}((X_1, X_2) = (1, 1)) = 1/4\end{aligned}$$

On a donc une chance sur quatre pour que les deux événements A_1 et A_2 se réalisent :

$$\mathbb{P}(A_1 \text{ et } A_2) = \mathbb{P}((X_1, X_2) = (0, 0)) = 1/4$$

D'autre part, puisque chacun d'entre eux à une chance sur deux de se réaliser, on a aussi

$$\mathbb{P}(A_1) \times \mathbb{P}(A_2) = \mathbb{P}(X_1 = 0) \times \mathbb{P}(X_2 = 0) = 1/2 \times 1/2 = 1/4$$

Par conséquent, la probabilité pour que A_1 et A_2 se produisent simultanément est égale au **produit** des probabilités pour que chacun d'eux se soit séparément réalisé

$$\mathbb{P}(A_1 \text{ et } A_2) = \mathbb{P}(A_1) \times \mathbb{P}(A_2)$$

En termes mathématiques, deux événements vérifiant cette propriété sont dit indépendants. Cette propriété élémentaire permet de calculer très simplement les probabilités d'évènements parfois complexes. Par exemple, choisir successivement, et au hasard une longue suite de points (U_1, U_2, \dots, U_n) dans le segment $[0, 1]$, s'effectue en n étapes **indépendantes**. On choisit tout d'abord la première coordonnée U_1 , puis la seconde U_2 , puis la troisième U_3 , jusqu'à la $n^{\text{ième}}$ U_n dans l'intervalle $[0, 1]$. La probabilité pour que ces variables tombent successivement dans des segments $[a_1, b_1], \dots, [a_n, b_n] \subset [0, 1]$ est tout simplement donnée par la formule produit

$$\begin{aligned}\mathbb{P}(U_1 \in [a_1, b_1] , U_2 \in [a_2, b_2] , \dots , U_n \in [a_n, b_n]) \\ &= \mathbb{P}(U_1 \in [a_1, b_1]) \times \mathbb{P}(U_2 \in [a_2, b_2]) \times \dots \times \mathbb{P}(U_n \in [a_n, b_n]) \\ &= (b_1 - a_1) \times (b_2 - a_2) \times \dots \times (b_n - a_n)\end{aligned}$$

1.3 Les générateurs aléatoires

Simuler sur ordinateur des variables aléatoires uniformes et indépendantes sur $[0, 1]$, revient à générer une séquence de nombres $u = (u_n)_{n \geq 0}$ compris entre 0 et 1 qui **ressemble** à une suite de variables aléatoires $U = (U_n)_{n \geq 0}$ indépendantes et uniformes sur $[0, 1]$. Cette suite u doit avoir les mêmes propriétés statistiques que U . Par exemple, la séquence de nombres u doit donc se dérouler dans le temps de façon

totalément imprévisible. Si la suite u était périodique et de période trop faible, on pourrait trop facilement prévoir ses valeurs après cette période, La suite u doit aussi être uniformément répartie sur l'intervalle $[0, 1]$. On s'attend donc à avoir la moitié des nombres dans $[0, 1/2]$ et plus généralement il devrait y avoir une proportion $(b - a)$ de nombres dans tout intervalle $[a, b] \subset [0, 1]$, etc.

Il est important de souligner qu'il n'existe à ce jour aucune méthode exacte pour générer une telle suite de nombres u . Pire encore, la plupart des générateurs de nombres uniformes connus sont basés sur des systèmes **déterministes** chaotiques. Ces algorithmes sont loin d'être aléatoire! Ils possèdent toujours une période mais cette dernière est suffisamment grande pour ne pas être vue par l'utilisateur. Ces systèmes s'expriment le plus souvent par une suite arithmétique linéaire congruentielle de la forme suivante :

$$u_n = x_n/m \quad \text{avec} \quad x_n = (a x_{n-1} + b) \bmod(m)$$

La valeur initiale u_0 est appelée la graine, et les paramètres (a, b, m) sont des entiers bien choisis. Lorsque

$$a = 5, \quad b = 1, \quad m = 2^4 \quad \text{et} \quad x_0 = 1$$

la suite $x_n = (5 x_{n-1} + 1) \bmod(16)$ obtenue par cet algorithme est donnée par

$$1, 6, 15, 12, 13, 2, 11, 8, 9, \mathbf{14}, 7, 4, 5, 10, 3, 0, 1, 6, 15, 12, 13, 2, 11, 8, 9, \mathbf{14}, \dots$$

La période de cette suite est égale à 16. Cette période est bien entendu maximale car l'opération "modulo 16" ne permet de générer que des nombres compris entre 0 et 16. Lorsque $b = 0$, $0 < x_0 < m$, et m est un nombre premier, le petit théorème de Fermat indique que la période de la suite est égale à $(m - 1)$.

Il existe une foule de générateurs pseudo-aléatoires fondés sur ces opérations arithmétiques, quasiment autant que de choix des valeurs (a, b, m) ! Pour donner quelques exemples, certains logiciels utilisent

$$(a, b, m) = (13^{13}, 0, 2^{59}) \quad \text{ou} \quad (a, b, m) = (7^5, 0, 2^{31} - 1)$$

Le générateur Unix rand48 correspond aux valeurs

$$a = 25214903917, \quad b = 11 \quad \text{et} \quad m = 2^{48}$$

et celui du CRAY

$$a = 44485709377909, \quad b = 0 \quad \text{et} \quad m = 2^{48}$$

On peut aussi trouver dans la littérature d'autres générateurs à programmer soi-même ([?, ?]).

Il existe de nouveaux types de générateurs basés sur les lois d'incertitudes inhérentes à la mécanique quantique. L'idée prometteuse de ces générateurs est à peu près la suivante. Le noyau de certaines particules telles le Krypton-85 contient une horloge aléatoire. Lorsque cette dernière sonne, un électron est expulsé. Les règles de la physique quantique montrent qu'il est impossible de prédire ces instants. Ces horloges physiques sont par essence aléatoires ! Les futurs ordinateurs quantiques seront capable de "simuler" de vrais aléas physiques. Mieux encore, la physique quantique prédit que les différents états quantiques cohérents se réalisent dans des dimensions multiples. Ces ordinateurs ultrarapides seraient donc en mesure de traiter en parallèle des informations variées dans chaque univers !

1.4 Quelques lois usuelles

Comme leur nom l'indique, les variables discrètes représentent des phénomènes aléatoires ne prenant qu'un nombre fini, ou au plus dénombrable de valeurs. Les archétypes de ces variables sont bien entendu les jeux de "pile ou face", le "lancé de dés", ou encore les célèbres "jeux de loto" nationaux, ou européens.

Ces variables discrètes permettent aussi de formaliser tout type de succession d'évènements aléatoires : tirage de boules dans une urne, séquences de pannes de machines, suites d'échecs ou de succès lors d'une expérience répétée d'évènements, sélection aléatoire d'objets, nombre de naissances ou de décès d'individus sur des sites donnés, nombres d'accidents d'avions par année, etc.

Ces phénomènes aléatoires discrets obéissent à des règles probabilistes simples et précises. Ces dernières sont bien souvent de nature combinatoire, ce qui n'a rien de très séduisant ni le profane, ni pour la plupart des probabilistes avisés.

1.4.1 Les lois uniformes

Les phénomènes aléatoires les plus simples à décrire sont ceux ne prenant qu'un **nombre fini de valeurs**, toutes **équiprobables**. Ces phénomènes formalisés correspondent à la notion de variables **discrètes et uniformes**. Aucune de ces valeurs possibles n'est privilégiée ! La description mathématique de leur lois est "simplement" de type combinatoire.

Examinons pour commencer le cas des variables aléatoires uniformes X sur un espace réduit à deux points $E = \{1, 2\}$. Le simple jeu du "pile ou face" (avec un pièce non truquée) peut bien entendu être décrit par de telles variables. Le résultat du lancer X peut alors prendre l'une des deux valeurs 1 ou 2, avec la même probabilité

$$\mathbb{P}(X = 1) = \mathbb{P}(X = 2) = 1/2$$

Pour générer sur ordinateur un tel choix, on utilise une variable uniforme U sur $[0, 1]$, et l'on pose simplement

$$X = 1 \times 1_{[0,1/2[}(U) + 2 \times 1_{[1/2,1]}(U) \tag{1.1}$$

où 1_A désigne la fonction indicatrice d'un ensemble A

$$1_A(u) = \begin{cases} 1 & \text{si } u \in A \\ 0 & \text{si } u \notin A \end{cases}$$

L'expression (1.1) permet de détecter lequel des deux événements

$$A_1 = (U \in [0, 1/2[) \quad \text{ou} \quad A_2 = (U \in [1/2, 1])$$

s'est produit lors de la simulation de la variable U . Dans le cas du jeu de "pile ou face", si le nombre U tombe dans l'intervalle $[0, 1/2)$, on convient que le résultat du lancer est "pile" (et "face" dans le cas contraire).

On peut éviter *de tester* si la simulation du nombre U est tombée soit à gauche, soit à droite de $1/2$, en posant tout simplement

$$X = 1 + [2U]$$

où $[a]$ désigne la partie entière d'un nombre réel a . Il est aisé de vérifier les équivalences entre les événements suivantes

$$(0 \leq U < 1/2) \iff (X = 1) \quad \text{et} \quad (1/2 \leq U < 1) \iff (X = 2)$$

Dans le le jeu du "lancé de dés" avec un dés non truqué, le résultat du lancer X peut prendre l'une des six valeurs 1, 2, 3, 4, 5, ou 6 avec la même probabilité

$$\mathbb{P}(X = 1) = \mathbb{P}(X = 2) = \mathbb{P}(X = 3) = \mathbb{P}(X = 4) = \mathbb{P}(X = 5) = \mathbb{P}(X = 6) = 1/6$$

Pour générer sur ordinateur un lancer de dés aléatoire, on utilise une seule variable uniforme U sur $[0, 1]$, et l'on pose simplement

$$X = 1 + [6U]$$

Comme précédemment, nous avons les équivalences événementielles

$$\left(U \in \left[0, \frac{1}{6} \right] \right) \iff (X = 1), \quad \left(U \in \left[\frac{1}{6}, \frac{2}{6} \right] \right) \iff (X = 2), \dots, \quad \left(U \in \left[\frac{5}{6}, 1 \right] \right) \iff (X = 6)$$

Plus généralement, la loi d'une variable aléatoire X prenant uniformément ses valeurs dans un ensemble fini E est déterminée pour chaque valeur possible x de E par la formule

$$\mathbb{P}(X = x) = 1/\text{Card}(E)$$

où $\text{Card}(E)$ désigne le cardinal de l'ensemble E , c'est à dire le nombre d'éléments dans l'ensemble E . Dans le jeu de pile ou face, et celui du lancer de dès, l'ensemble des valeurs possibles a pour cardinal respectivement 2, et 6

$$E_1 = \{1, 2\} \quad \text{et} \quad E_2 = \{1, 2, \dots, 6\} \implies \text{Card}(E_1) = 2 \quad \text{et} \quad \text{Card}(E_2) = 6$$

Pour générer sur ordinateur un tel phénomène aléatoire, on utilise à nouveau une seule variable uniforme U sur $[0, 1]$, et l'on pose

$$X = 1 + [\text{Card}(E) \times U]$$

Il est donc d'autant plus improbable de choisir un élément donné que le cardinal de l'ensemble est grand. On a par exemple une chance sur 10.000 de choisir exactement un élément donné dans un ensemble à 10.000 éléments. Dans un jeu de loto où l'on tire sans remise 6 boules parmi 49, le résultat du tirage X est à valeurs dans un ensemble ayant approximativement 14 millions d'éléments! Bien qu'il soit bien plus artistique de simuler physiquement, ou mécaniquement les tirages successifs des boules dans une urne, l'on peut numéroter chaque possibilité par un nombre compris entre 1 et 14 millions. Ce jeu de loto simplifié se résume alors à choisir un nombre entre 1 et 14 millions. Pour le simuler sur ordinateur, il suffit de poser

$$X = 1 + [14 \times 10^6 \times U]$$

Le joueur gagne s'il devine le nombre X choisi aléatoirement entre 1 et 14 millions par l'ordinateur.

1.4.2 Les lois d'événements composés

Introduction

Lors d'une expérience aléatoire, un événement donné, disons A , se réalise avec une probabilité p connue

$$\mathbb{P}(A) = p \in [0, 1]$$

Par exemple, lors d'un lancer de dés non truqué le chiffre 6 a exactement une chance sur 6 de se réaliser. Si l'on désigne par X le résultat du lancer, la probabilité de réalisation de l'évènement $A = \{X = 6\}$ est clairement égale à un sixième

$$\mathbb{P}(A) = \mathbb{P}(X = 6) = 1/6$$

L'expérience se répète dans le temps. Au fil du temps, l'évènement en question a de fortes chances de se réaliser au moins une fois. Il est par exemple bien improbable que le chiffre 6 ne soit jamais sorti lors d'un million de lancers de dés. Sur une infinité de lancers, il est même certain qu'il apparaisse au moins une fois!

Plusieurs questions naturelles se posent alors :

Combien de temps doit-on attendre pour que l'évènement qui nous intéresse se réalise avec une probabilité supérieure à $1/2$, voire supérieure à $99/100$? En reprenant l'exemple précédent, combien de fois doit-on lancer les dés pour que le chiffre 6 apparaisse avec une probabilité supérieure à $9/10$? Combien d'années doit-on jouer au jeu de loto décrit précédemment, pour avoir une chance de gagner supérieure à $1/2$? Dans un autre ordre d'idées, combien de joueurs doivent participer à ce loto, pour avoir au moins un gagnant avec une probabilité supérieure à $9/10$?

Associations évènementielles

Avant de répondre précisément aux questions précédentes, il convient de faire un certain nombre de remarques. Tout d'abord, il est important de souligner que **tout évènement A ayant une probabilité donnée de se réaliser peut s'exprimer en terme d'une variable uniforme U sur $[0, 1]$** . Plus précisément, on convient que l'évènement A se réalise lorsque la simulation de la variable uniforme U tombe dans le segment $[0, p]$. On met ainsi **en correspondance** deux évènements ayant la même probabilité de se réaliser

$$\{\text{"A se réalise"}\} = \{U \in [0, p]\}$$

Pour s'en convaincre, il suffit de noter que l'on a

$$\mathbb{P}(A) = p = \mathbb{P}(U \in [0, p])$$

Dans l'exemple du lancer de dés, on peut associer les deux évènements équiprobables $\{X = 6\}$ et $\{U \in [0, 1/6]\}$. Le chiffre 6 a autant de chances de se réaliser que la variable simulée U a de chances de tomber dans l'intervalle $[0, 1/6]$ de longueur $1/6$

$$\{X = 6\} = \{U \in [0, 1/6]\} \implies \mathbb{P}(X = 6) = \mathbb{P}(U \in [0, 1/6]) = 1/6$$

Description des modèles

En utilisant ces associations évènementielles, réaliser l'expérience plusieurs fois, revient à simuler une suite de variables aléatoires $(U_n)_{n \geq 1}$ uniformes et indépendantes sur $[0, 1]$. A chaque étape n , on vérifie si l'évènement A s'est réalisé en détectant si la variable U_n est tombée dans l'intervalle $[0, p]$. Dans une séquence de 4 lancers de dés, la suite d'évènements

$$\{U_1 \notin [0, 1/6]\}, \{U_2 \in [0, 1/6]\}, \{U_3 \in [0, 1/6]\}, \{U_4 \in [0, 1/6]\}$$

correspond à la situation où le chiffre 6 est sorti à chacun des instants, sauf au premier. De même, la suite d'évènements

$$\{U_1 \notin [0, 1/6]\}, \{U_2 \notin [0, 1/6]\}, \{U_3 \notin [0, 1/6]\}, \{U_4 \in [0, 1/6]\}$$

correspond à la situation où le chiffre 6 n'est sorti qu'au dernier instant. En utilisant les propriétés d'indépendance entre évènements, nous avons

$$\begin{aligned} & \mathbb{P}(\text{"le chiffre 6 ne sort jamais pendant les } n \text{ premiers lancers"}) \\ &= \mathbb{P}(U_1 \notin [0, 1/6], U_2 \notin [0, 1/6], \dots, U_n \notin [0, 1/6]) \\ &= \mathbb{P}(U_1 \notin [0, 1/6]) \times \mathbb{P}(U_2 \notin [0, 1/6]) \times \dots \times \mathbb{P}(U_n \notin [0, 1/6]) \\ &= (5/6)^n \end{aligned}$$

En passant aux évènements complémentaires, la probabilité pour que le chiffre 6 apparaisse au moins une fois est donné par

$$\begin{aligned} & \mathbb{P}(\text{“le chiffre 6 sort au moins une fois lors de } n \text{ lancers”}) \\ &= 1 - \mathbb{P}(\text{“le chiffre 6 ne sort jamais pendant } n \text{ lancers”}) \\ &= 1 - (5/6)^n \end{aligned}$$

Ces formules élémentaires nous permettent de répondre à toutes nos questions. Avec plus d’une chance sur deux, le chiffre 6 sort au moins une fois après 4 lancers. En effet, nous avons

$$\begin{aligned} 1 - (5/6)^n \geq 1/2 & \Leftrightarrow (5/6)^n \leq 1/2 \\ & \Leftrightarrow n \log(5/6) \leq \log(1/2) \\ & \Leftrightarrow n \log(6/5) \geq \log(2) \\ & \Leftrightarrow n \geq \log(2)/\log(6/5) \simeq 3,80 \end{aligned}$$

De même, avec plus de 99 chances sur 100, le chiffre 6 sort au moins une fois après 26 lancers

$$\begin{aligned} 1 - (5/6)^n \geq 99/100 & \Leftrightarrow (5/6)^n \leq 1/100 \\ & \Leftrightarrow n \log(5/6) \leq \log(1/100) \\ & \Leftrightarrow n \log(6/5) \geq \log(100) \\ & \Leftrightarrow n \geq \log(100)/\log(6/5) \simeq 25,26 \end{aligned}$$

Plus généralement, la probabilité pour qu’un évènement quelconque A se réalise au moins une fois, ou jamais lors de n expériences aléatoires indépendantes est donnée par la formule

$$\begin{aligned} & \mathbb{P}(\text{“}A \text{ se réalise au moins une fois lors de } n \text{ expériences”}) \\ &= 1 - \mathbb{P}(\text{“}A \text{ ne s’est jamais réalisé après } n \text{ expériences”}) \\ &= 1 - (1 - p)^n \end{aligned}$$

L’exemple du loto

Les probabilités de gain au loto sont bien sur plus faibles que celle de voir apparaître un 6 lors d’un lancer de dés. Néanmoins les raisonnements sont les mêmes, il suffit de remplacer la probabilité $1/6$ par la probabilité de gagner au loto égale à $1/(14 \times 10^6)$.

On obtient ainsi les formules suivantes

$$\begin{aligned}
 & \mathbb{P}(\text{“La grille choisie sort au moins une fois lors de } n \text{ tirages de loto”}) \\
 &= 1 - \mathbb{P}(\text{“La grille choisie ne sort jamais lors de } n \text{ tirages de loto”}) \\
 &= 1 - \left(1 - \frac{1}{14 \times 10^6}\right)^n
 \end{aligned}$$

On en conclut qu’il nous faut jouer plus de 10 millions de fois pour que notre grille ait une chance sur deux de sortir au moins une fois

$$\begin{aligned}
 1 - \left(1 - \frac{1}{14 \times 10^6}\right)^n \geq 1/2 & \Leftrightarrow \left(1 - \frac{1}{14 \times 10^6}\right)^n \leq 1/2 \\
 & \Leftrightarrow n \log \left(1 - \frac{1}{14 \times 10^6}\right) \leq \log(1/2) \\
 & \Leftrightarrow n \log \left(\frac{14 \times 10^6}{14 \times 10^6 - 1}\right) \geq \log(2) \\
 & \Leftrightarrow n \geq \log 2 / \log \left(\frac{14 \times 10^6}{14 \times 10^6 - 1}\right) \simeq 9,8 \times 10^6
 \end{aligned}$$

En supposant que le joueur avisé joue une seule fois par semaine, il lui faudrait jouer 10 millions de semaines, soit plus de 192 mille ans pour espérer gagner au moins une fois avec une chance sur deux !

Par des raisonnements analogues, on vérifie qu’il est nécessaire de jouer plus de 65 millions de fois pour que notre grille ait plus de 99 chances sur 100 de sortir au moins une fois. De façon équivalente, il faudrait à peu près 65 millions de joueurs de loto acharnés pour être sur que l’un d’entre eux puisse gagner avec une probabilité 99/100

$$\begin{aligned}
 1 - \left(1 - \frac{1}{14 \times 10^6}\right)^n \geq 99/100 & \Leftrightarrow \left(1 - \frac{1}{14 \times 10^6}\right)^n \leq 1/100 \\
 & \Leftrightarrow n \log \left(1 - \frac{1}{14 \times 10^6}\right) \leq \log(1/100) \\
 & \Leftrightarrow n \log \left(\frac{14 \times 10^6}{14 \times 10^6 - 1}\right) \geq \log(100) \\
 & \Leftrightarrow n \geq \log(100) / \log \left(\frac{14 \times 10^6}{14 \times 10^6 - 1}\right) \simeq 65,7 \times 10^6
 \end{aligned}$$

Ces chiffres pour le moins astronomiques n’ont jamais détourné les joueurs de ces jeux de loto. Ces derniers semblent toujours apprécier l’espoir de gagner, avec de faibles mises des sommes d’argent parfois inversement proportionnelles aux chances de gain.

1.4.3 Les lois géométriques

Introduction

Les phénomènes aléatoires les plus intéressants se déroulent dans le temps. Ils se composent d'une succession d'évènements élémentaires aussi complexes qu'imprévisibles. Pour donner quelques exemples d'actualité, les formations de cyclones, ou de gaz au niveau atomique, ou encore l'évolution d'actifs financiers sont le fruit de fluctuations aléatoires météorologiques, quantiques, ou financières se déroulant dans le temps.

La formalisation mathématique de ces phénomènes aléatoires complexes est fondée sur des études physiques ou statistiques précises. Les modèles probabilistes formalisés s'expriment alors sous la forme de processus stochastiques diffusifs ou à sauts. Nous reviendrons sur ces modèles lorsque nous aborderons la simulation de chaînes de Markov.

Avant d'étudier en profondeur ces phénomènes aléatoires, il est souvent très utile de restreindre notre analyse à certaines questions plus élémentaires, tout en simplifiant à l'extrême ces modèles :

Quelle est la probabilité pour qu'un cyclone se forme à une date donnée ? Quelle sont nos "chances" pour qu'un appareil électronique tombe en panne après une centaine d'utilisations ? Quelle est la probabilité pour qu'une action financière s'effondre à une date donnée ?

La réponse précise à ces questions nécessite bien évidemment une analyse fine des processus de turbulence météorologique, des fluctuations thermiques et quantiques des composants électroniques, ou encore une étude profonde des fluctuations du marché financier. Néanmoins, une première étude statistique naïve permettent d'estimer très grossièrement les probabilités pour que l'évènement qui nous intéresse se réalise.

Supposons donc qu'un instrument électronique ait une chance sur mille de tomber en panne à chaque utilisation. Quelle est la probabilité pour qu'il tombe en panne après une centaine d'utilisations ? Quelle la loi de probabilité complète de ses instants de pannes ? De même, si un cyclone peut se former chaque jour sur le golfe du Mexique avec une chance sur mille, quelle est la probabilité d'avoir une catastrophe cyclonique exactement après six mois ? Enfin, si un actif financier peut s'effondrer tous les jours avec un chance sur mille, quelle est la loi exacte des instants de faillites ? Dans tous les cas, on supposera que les réalisations ou non des évènements en questions à chaque étape sont mutuellement indépendants.

Nous insisterons sur le fait que les lois de probabilités réelles de ces phénomènes aléatoires dépendent du paramètre temporel. De plus, dans la plupart des phénomènes physiques, les états du système sont loins d'être indépendants dans le temps. Les mécanismes de dépendance sont souvent dictés par des lois fondamentales de la physique. Par conséquent, les modèles statistiques que nous venons de décrire sont très peu réalistes, mais ils suffisent pour se donner une première idée du comportement

des systèmes.

Description des modèles

Toutes les questions précédentes correspondent à la donnée d'un événement A pouvant se réaliser à chaque unité de temps, avec une probabilité fixée

$$\mathbb{P}(A) = p \in [0, 1]$$

On simule la réalisation ou non de cet événement en convenant que l'évènement A se réalise, si variable uniforme U générée sur $[0, 1]$ par l'ordinateur tombe dans l'intervalle $[0, p[$, de longueur p . Cette stratégie est résumée par la formule

$$A = \{U \in [0, p[\}$$

L'évènement contraire A^c , correspondant à la situation où A ne se réalise pas, est donnée par

$$A^c = \{U \notin [0, p[\} = \{U \in [p, 1] \}$$

Dans l'exemple de l'instrument électronique, l'évènement A correspond au disfonctionnement du système. On convient que la panne est effective lorsque la variable générée par l'ordinateur tombe dans le segment $[0, p[$, avec $p = 1/1000$. Autrement dit, nous avons

$$A = \{ \text{"L'appareil tombe en panne"} \} = \{U \in [0, 1/1000[\}$$

et

$$A^c = \{ \text{"L'appareil fonctionne"} \} = \{U \in [1/1000, 1] \}$$

On vérifie que les probabilités d'occurrence de ces événements correspondent bien aux données du problème

$$\mathbb{P}(\text{"L'appareil tombe en panne"}) = 1/1000$$

et

$$\mathbb{P}(\text{"L'appareil fonctionne"}) = 1 - 1/1000 = 999/1000$$

La loi géométrique

Pour simuler l'évolution temporelle du phénomène aléatoire, on génère sur l'ordinateur une suite $(U_n)_{n \geq 1}$ de variables uniformes sur $[0, 1]$, tout en vérifiant au fil du temps si l'une des variables tombe dans $[0, p[$. On note T **la première fois** où cela se produit. Cet instant aléatoire T correspond au plus petit entier n , pour lequel U_n tombe dans $[0, p[$. On note ce temps aléatoire

$$T = \inf \{n \geq 1 : U_n \in [0, p[\}$$

Plus précisément, si la première simulation U_1 tombe dans $[0, p[$, on convient que l'évènement A s'est réalisé dès le premier instant. On pose dans ce cas

$$T = 1$$

Lorsque U_1 est dans $[p, 1]$, on simule une nouvelle variable uniforme U_2 sur $[0, 1]$. Si cette variable tombe dans $[0, p[$, alors on convient que A s'est réalisé lors de la seconde expérience, et on pose

$$T = 2$$

Dans le cas contraire, les variables U_1 et U_2 sont tombées dans l'intervalle $[p, 1]$. On génère alors une nouvelle variable U_3 sur $[0, 1]$. Si cette variable tombe dans $[0, p[$, alors on convient que A s'est réalisé après trois expériences, et on pose

$$T = 3$$

On continue ce procédé jusqu'à ce que l'évènement souhaité se réalise enfin. Ces premiers instants T de réalisation de l'évènement A sont des variables aléatoires à valeurs entières. D'après la discussion précédente, nous avons les correspondances évènementielles suivantes :

$$\begin{aligned} \{T = 1\} &= \{U_1 \in [0, p[\} \\ \{T = 2\} &= \{U_1 \notin [0, p[, U_2 \in [0, p[\} \\ \{T = 3\} &= \{U_1 \notin [0, p[, U_2 \notin [0, p[, U_3 \in [0, p[\} \\ &\vdots \\ \{T = n\} &= \{U_1 \notin [0, p[, U_2 \notin [0, p[, \dots, U_{n-1} \notin [0, p[, U_n \in [0, p[\} \\ &\vdots \end{aligned}$$

À titre d'exemple, la situation où notre appareil électronique tombe en panne après quatre utilisations est décrite par l'évènement

$$\{T = 4\} = \{U_1 \notin [0, 1/1000[, U_2 \notin [0, 1/1000[, U_3 \notin [0, 1/1000[, U_4 \in [0, 1/1000[\}$$

D'après les propriétés d'indépendance entre les évènements, la probabilité pour que cet évènement se réalise est donnée par

$$\begin{aligned} \mathbb{P}(T = 4) &= \mathbb{P}(U_1 \notin [0, 1/1000[, U_2 \notin [0, 1/1000[, U_3 \notin [0, 1/1000[, U_4 \in [0, 1/1000[) \\ &= \mathbb{P}(U_1 \notin [0, 1/1000[) \times \mathbb{P}(U_2 \notin [0, 1/1000[) \times \mathbb{P}(U_3 \notin [0, 1/1000[) \\ &\quad \times \mathbb{P}(U_4 \in [0, 1/1000[) \\ &= \mathbb{P}(U \notin [0, 1/1000[)^3 \times \mathbb{P}(U \in [0, 1/1000[) \\ &= \frac{1}{1000} \times \left(1 - \frac{1}{1000}\right)^3 \end{aligned}$$

Plus généralement, la probabilité pour qu'il tombe en panne après n utilisations est donnée par la formule

$$\mathbb{P}(T = n) = \frac{1}{1000} \times \left(1 - \frac{1}{1000}\right)^{n-1}$$

On a donc approximativement neuf chances sur 10.000 pour que notre appareil tombe en panne **exactement** après une centaine d'utilisations

$$\mathbb{P}(T = 100) = \frac{1}{1000} \times \left(1 - \frac{1}{1000}\right)^{99} \simeq 9 \times 10^{-4}$$

Ces lois des premiers instants où un évènement se produit sont appelées des **lois géométriques**. Pour des évènements quelconques A ayant une probabilité $\mathbb{P}(A) = p$ de se réaliser, le premier instant de réalisation

$$T = \inf \{n \geq 1 : U_n \in [0, p]\}$$

est distribué selon la loi géométrique de paramètre p définie par la formule

$$\forall n \geq 1, \quad \mathbb{P}(T = n) = p (1 - p)^{n-1}.$$

1.4.4 Les lois discrètes

Introduction

Il arrive bien souvent que certaines réalisations d'expériences aléatoires soient bien plus probables que d'autres. Pour illustrer cette remarque, reprenons l'exemple réchauffé du choix d'une boule dans une urne contenant des boules blanches et noires. Supposons que l'urne contienne 100 boules noires, et une seule boule blanche. Dans ce cas, nous avons 99 chances sur 100 de choisir une boule noire, et seulement une chance sur 100 de trouver la boule blanche.

Il va de même, lors d'un lancer de dés truqué où les six faces n'ont plus la même fréquence d'apparition. Nous profiterons de cet exemple du lancer de dés pour souligner que certains évènements aléatoires sont par essence plus probables que d'autres. Lors d'un lancer de **deux dès non truqués**, il est plus probable que la somme des chiffres obtenus soit égale à 9, plutôt qu'à 10. Pour vérifier cette assertion, il suffit de noter que l'on peut obtenir 10 de trois façons

$$5 + 5 = 6 + 4 = 4 + 6 = 10$$

et le chiffre 9 de 4 façons distinctes

$$4 + 5 = 5 + 4 = 6 + 3 = 3 + 6 = 9$$

On a donc

$$\mathbb{P}(\text{la somme des deux lancers vaut } 10) = 3/36 = 1/12$$

et

$$\mathbb{P}(\text{la somme des deux lancers vaut } 9) = 4/36 = 1/9$$

Les processus de naissances et morts d'individus dans des environnements plus ou moins accueillants sont aussi sujet à de telles variations. Par exemple, les naissances dans des régions désertiques sont moins fréquentes que celles des capitales européennes. Ces taux de naissances sont souvent reliés à la qualité d'adaptation d'un individu dans des sites donnés.

Dans un tout autre registre, l'élaboration de portefeuilles financiers est souvent fondée sur des critères de qualité rendant compte des tendances à la hausse ou à la baisse d'une collection d'actions. Ainsi, un bon agent financier cherchera à miser le plus souvent sur des actifs ayant de fort potentiel de gain. Le comportement d'un agent financier peut ainsi être décrit par des sélections aléatoires d'actifs, certains plus probables que d'autres en fonction d'une conjoncture financière, ou d'un climat économique donné.

Description des modèles

D'un point de vue pratique, comment simuler sur un ordinateur de tels phénomènes aléatoires? L'idée est relativement simple. Nous sommes en présence d'une variable aléatoire X pouvant prendre un certain nombre de valeurs, que nous noterons $\{x_1, \dots, x_d\}$, avec une probabilité donnée

$$\forall i \in \{1, \dots, d\} \quad \mathbb{P}(X = x_i) = p_i \in [0, 1]$$

L'interprétation de ces états dépend du problème étudié. Dans l'exemple de l'urne, la variable X représente la couleur de la boule. Deux seules couleurs sont présentes dans l'urne, et l'espace d'état est donc réduit à deux points $\{x_1, x_2\}$. Dans cette situation, on peut convenir que x_1 représente le choix d'une boule blanche, et x_2 celui d'une boule noire. En reprenant la composition de l'urne décrite ci-dessus, nous avons

$$\mathbb{P}(X = x_1) = 1 - \mathbb{P}(X = x_2) = 1/100$$

Pour réaliser cette expérience X en terme d'une variable uniforme U sur $[0, 1]$, on décompose cet intervalle en deux segments de longueurs respectives $1/100$ et $99/100$

$$[0, 1] = [0, 1/100] \cup]1/100, 1]$$

On met ensuite en correspondance les événements équiprobables

$$\{X = x_1\} = \{U \in [0, 1/100]\} \quad \text{et} \quad \{X = x_2\} = \{U \in]1/100, 1]\}$$

Autrement dit, l'on convient que l'on choisit une boule blanche lorsque le nombre simulé U tombe dans le sous-intervalle $[0, 1/100]$. Cette expérience aléatoire X peut s'écrire sous la forme synthétique suivante

$$X = x_1 1_{[0,1/100]}(U) + x_2 1_{]1/100,1]}(U)$$

On rappelle que $1_I(U)$ représente l'indicatrice de l'ensemble I

$$1_I(U) = \begin{cases} 1 & \text{si } U \in I \\ 0 & \text{si } U \notin I \end{cases}$$

Dans l'exemple du lancer de dès, la variable X représente le résultat de l'expérience. Dans cette situation, la variable X peut prendre l'une des valeurs

$$x_1 = 1, \quad x_2 = 2, \quad x_3 = 3, \quad x_4 = 4, \quad x_5 = 5 \quad \text{ou} \quad x_6 = 6$$

Un exemple de dès truqué est donné par

$$\mathbb{P}(X = 1) = 1/2, \quad \mathbb{P}(X = 2) = 1/4$$

et

$$\mathbb{P}(X = 3) = 1/8, \quad \mathbb{P}(X = 4) = 1/16, \quad \mathbb{P}(X = 5) = 1/32 \quad \text{et} \quad \mathbb{P}(X = 6) = 1/32$$

Dans cet exemple, la chiffre 1 apparait très souvent avec une chance sur 2, alors que la face 6 est plus improbable, elle se réalise en général une seule fois sur 32. Comme précédemment, on réalise ce lancer de dès truqué en terme d'une variable uniforme U sur $[0, 1]$, en décomposant l'intervalle $[0, 1]$ en six segments de longueurs respectives $1/2, 1/4, 1/8, 1/16, 1/32$

$$\begin{aligned} [0, 1] &= I_1 \cup I_2 \cup I_3 \cup I_4 \cup I_5 \cup I_6 \\ &= \left[0, \frac{1}{2}\right] \cup \left[\frac{1}{2}, \frac{3}{4}\right] \cup \left[\frac{3}{4}, \frac{7}{8}\right] \cup \left[\frac{7}{8}, \frac{15}{16}\right] \cup \left[\frac{15}{16}, \frac{31}{32}\right] \cup \left[\frac{31}{32}, 1\right] \end{aligned}$$

On effectue ensuite les associations évènementielles suivantes

$$\{X = i\} = \{U \in I_i\}$$

pour chaque $i \in \{1, 2, \dots, 6\}$. Le résultat du lancer de dés X peut s'écrire sous la forme synthétique suivante

$$X = 1 \times 1_{I_1}(U) + 2 \times 1_{I_2}(U) + \dots + 6 \times 1_{I_6}(U)$$

Plus généralement, pour simuler une variable X distribuée dans un ensemble fini $\{x_1, \dots, x_d\}$ avec la loi de probabilité donnée

$$\forall i \in \{1, \dots, d\} \quad \mathbb{P}(X = x_i) = p_i \in [0, 1]$$

on décompose tout d'abord l'intervalle $[0, 1]$ en d petits segments de longueurs respectives p_1, p_2, \dots, p_d

$$\begin{aligned} [0, 1] &= I_1 \cup I_2 \cup I_3 \cup \dots \cup I_d \\ &= [0, p_1] \cup]p_1, p_1 + p_2] \cup]p_1 + p_2, p_1 + p_2 + p_3] \cup \dots \cup]p_1 + \dots + p_{d-1}, 1] \end{aligned}$$

Pour chaque indice $i \in \{1, \dots, d\}$, on fait les associations évènementielles suivantes :

$$\{X = x_i\} = \{U \in I_i\}$$

Autrement dit, on convient que la variable aléatoire X prend la valeur x_i , lorsque le nombre simulé U dans l'intervalle $[0, 1]$ prend une valeur dans l'intervalle I_i . En termes plus synthétiques, la variable X peut s'écrire sous la forme suivante

$$X = \sum_{i=1}^d x_i 1_{I_i}(U)$$

1.4.5 Les lois continues

Introduction

Dans le chapitre précédent nous avons présenté des techniques de simulation d'expériences aléatoires prenant un nombre de valeurs au plus dénombrable. Ces expériences sont parfois l'expression simplifiée de phénomènes aléatoires plus complexes, à valeurs dans des espaces continus, ou de dimensions multiples. C'est le cas des arrivées de clients dans des files d'attente, les instants de désintégration de particules quantiques, la répartition de la chaleur, les impacts de météorites sur le globe terrestre, les fluctuations de prix d'actifs financiers dans des marchés boursiers, les explorations aléatoires de régions par des populations d'individus, etc.

Tous ces aléas sont de nature très variée : répartitions gaussiennes, inter temps exponentiels, statistiques d'ordre uniformes, distributions log-normales, densités de type Gamma, lois de Cauchy, de Rayleigh, de Pareto, de Weibull, distributions uniformes sur des surfaces, ou des volumes quelconques, etc.

Chacune de ses lois traduit une répartition statistique particulière des aléas d'une expérience. La loi gaussienne peut s'interpréter comme la répartition aléatoire de la chaleur dans une droite. Cette distribution en forme de cloche souligne les différents degrés de chaleur en chaque site. Les régions éloignées du point de chauffage initial sont plus froides. Les températures autour de la source de chaleur se répartissent en forme de cloche. Les distributions uniformes sur des sphères peuvent s'interpréter comme des lieux d'impact de météorites sur une planète. Ces lois peuvent être étonnamment reliées à des projections de variables aléatoires gaussiennes sur chaque axe du repère considéré ! Les lois exponentielles peuvent s'interpréter comme des horloges aléatoires représentant des instants d'arrivée d'individus dans des files d'attente, des durées de traitement

d'information dans des réseaux de télécommunication, des durées de fonctionnement de machines, des désintégrations de particules, etc.

Comment simuler de tels phénomènes aléatoires sur un ordinateur? Ces aléas continus peuvent-ils toujours s'exprimer en terme de variables uniformes sur l'intervalle $[0, 1]$? Existe-t-il des procédés de simulation de lois universels? Avant de répondre à ces questions, examinons tout d'abord l'expression mathématique des phénomènes continus. Pour des expériences aléatoires réelles X , telles les répartition de température sur une barre rectiligne, les instants d'arrivées, de pannes, ou de ruptures, la probabilité pour que le résultat aléatoire soit compris entre deux valeurs a , et b , correspond à **une aire** entre le graphe d'une fonction positive

$$p : x \in \mathbb{R} \longrightarrow p(x) \in [0, \infty[$$

et le segment $[a, b]$.

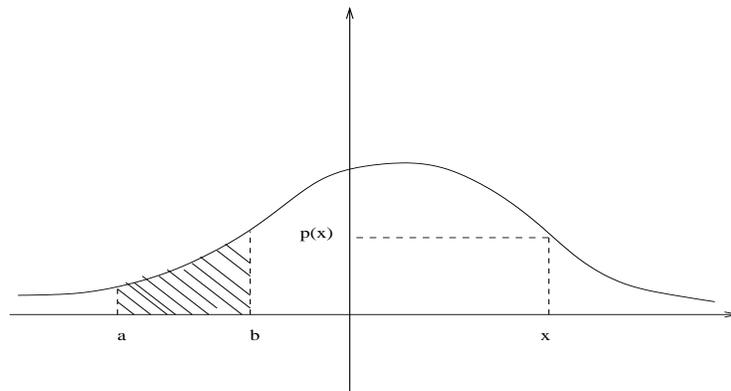


FIG. 1.1 – Intégrale entre a et b

Cette aire est appelé l'**intégrale** de la densité de probabilité p entre le point a et le point b , et on la note

$$\mathbb{P}(a \leq X \leq b) = \int_a^b p(x) dx = \int_{[a,b]} p(x) dx$$

Cette interprétation probabiliste offre une interprétation physique à la théorie de l'intégration. Par exemple, la somme des aires sur deux intervalles disjoints $[a_1, b_1]$, et $[a_2, b_2]$ correspond à la probabilité pour que le résultat de l'expérience X prenne ses

valeurs dans le premier intervalle $[a_1, b_1]$, ou dans le second $[a_2, b_2]$

$$\begin{aligned} \mathbb{P}(X \in [a_1, b_1] \text{ ou } X \in [a_2, b_2]) &= \int_{[a_1, b_1] \cup [a_2, b_2]} p(x) dx \\ &= \int_{a_1}^{b_1} p(x) dx + \int_{a_2}^{b_2} p(x) dx \\ &= \mathbb{P}(X \in [a_1, b_1]) + \mathbb{P}(X \in [a_2, b_2]) \end{aligned}$$

Horloges exponentielles

Comme nous l'avons souligné dans l'introduction, l'évolution aléatoire du nombre de clients dans une file d'attente est clairement le fruit d'un processus d'arrivées aléatoires d'individus au fil du temps. Ces dates d'arrivées peuvent être représentées par des variables aléatoires T à valeurs dans la demi-droite $[0, \infty[$. Leur nature est souvent exponentielle. Plus précisément, la probabilité pour qu'une personne arrive dans une file d'attente à un instant T compris entre t_1 et t_2 est donnée par l'intégrale d'une fonction exponentielle entre les dates en question

$$\mathbb{P}(t_1 \leq T \leq t_2) = \int_{t_1}^{t_2} \lambda e^{-\lambda s} ds = e^{-\lambda t_1} - e^{-\lambda t_2}$$

En particulier, nous avons

$$\mathbb{P}(T \geq t) = \lim_{s \uparrow \infty} \mathbb{P}(t \leq T \leq s) = e^{-\lambda t}$$

On utilise souvent la notation synthétique suivante

$$\mathbb{P}(T \in dt) \stackrel{\text{déf.}}{=} \mathbb{P}(T \in [t, t + dt[) = \lambda e^{-\lambda t} dt$$

D'après la remarque précédente, la probabilité pour qu'une personne arrive entre deux instants très rapprochés t et $t + dt$, sachant qu'elle n'est pas encore arrivée avant t , est égale au produit λdt

$$\begin{aligned} \mathbb{P}(T \in dt \mid T \geq t) &\stackrel{\text{déf.}}{=} \mathbb{P}(T \in [t, t + dt[\mid T \geq t) = \frac{\mathbb{P}(T \in [t, t + dt[\text{ et } T \geq t)}{\mathbb{P}(T \geq t)} \\ &= \frac{\mathbb{P}(T \in [t, t + dt[)}{\mathbb{P}(T \geq t)} = \lambda dt \end{aligned}$$

Le paramètre $\lambda > 0$ est appelé l'intensité de la variable aléatoire. Notons que pour $t_1 = 0$, et lorsque $t_2 \uparrow \infty$, nous avons

$$\mathbb{P}(T < \infty) = 1 - \lim_{t_2 \rightarrow \infty} e^{-\lambda t_2} = 1$$

Les amateurs de l'intégration par parties, pourrons noter que l'instant moyen d'arrivée d'un client dans une file d'attente est inversement proportionnel à l'intensité

$$\begin{aligned}\mathbb{E}(T) &= \int_0^{\infty} s \times \mathbb{P}(T \in ds) = \int_0^{\infty} s \lambda e^{-\lambda s} ds = - \int_0^{\infty} s \frac{\partial}{\partial s}(e^{-\lambda s}) ds \\ &= -[s e^{-\lambda s}]_0^{\infty} + \frac{1}{\lambda} \int_0^{\infty} \lambda e^{-\lambda s} ds = \frac{1}{\lambda}\end{aligned}$$

Le paramètre λ correspond donc à la fréquence d'arrivée des clients par unité de temps, disons par minute. Lorsque $\lambda = 10$ clients par minute, l'instant d'arrivée moyen d'un client est de 1/10 de minute. Autrement dit, les clients arrivent en moyenne toutes les six minutes dans la file d'attente.

La simulation de ces instants d'arrivée aléatoire est à nouveau l'expression d'une variable uniforme U sur $[0, 1]$! Pour être plus précis, commençons par poser

$$T = -\frac{1}{\lambda} \log U$$

On vérifie sans trop de peine que l'on a

$$\{t_1 \leq T \leq t_2\} = \{e^{-\lambda t_2} \leq U \leq e^{-\lambda t_1}\}$$

Cette association événementielle nous conduit au résultat recherché

$$\mathbb{P}(t_1 \leq T \leq t_2) = \mathbb{P}(e^{-\lambda t_2} \leq U \leq e^{-\lambda t_1}) = e^{-\lambda t_1} - e^{-\lambda t_2}$$

Pour simuler l'arrivée d'un client dans une file d'attente, on prend donc tout simplement le logarithme d'un nombre uniforme sur $[0, 1]$ fourni par l'ordinateur, et l'on multiplie le résultat par $-1/\lambda$.

Chapitre 2

Quelques outils de simulation

2.1 La méthode d'inversion

2.1.1 Le principe de base

Derrière la méthode de simulation d'horloges exponentielles décrite dans la section précédente, se cache une technique universelle. Pour être plus précis, nous commencerons par noter que la fonction de répartition de la loi exponentielle de paramètre λ est donnée par la formule explicite

$$\begin{aligned} F(x) = \mathbb{P}(T \leq x) &= \int_0^x \lambda e^{-\lambda t} dt \\ &= - \int_0^x \frac{\partial}{\partial t} (e^{-\lambda t}) dt = -[e^{-\lambda t}]_0^x = 1 - e^{-\lambda x} \end{aligned}$$

L'inverse de cette fonction est facilement calculable en chaque point $u \in]0, 1[$

$$\begin{aligned} F(x) = u &\iff e^{-\lambda x} = (1 - u) \\ &\iff x = F^{-1}(u) = -\frac{1}{\lambda} \log(1 - u) \end{aligned}$$

Cet inverse ressemble étrangement à l'expression de la variable exponentielle en terme d'une variable uniforme présentée dans la section précédente.

$$T = -\frac{1}{\lambda} \log U$$

En effet, les variables U et $(1 - U)$ sont toutes deux uniformes sur l'intervalle $[0, 1]$! Il semble donc qu'en posant

$$T = F^{-1}(U)$$

l'on obtienne un procédé de simulation universel d'une variable aléatoire ayant pour fonction de répartition F . Pour vérifier cette intuition, notons F la fonction de

répartition d'une variable réelle quelconque X définie par

$$F(x) = \mathbb{P}(X \leq x).$$

Lorsque cette fonction est inversible, on a clairement pour chaque $x \in \mathbb{R}$

$$\mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = \int_0^{F(x)} du = F(x).$$

Il en découle que la variable $F^{-1}(U)$ est bien distribuée selon la même loi que la variable aléatoire X .

Inversement, si X est une variable aléatoire réelle ayant une fonction de répartition strictement croissante $F : \mathbb{R} \rightarrow [0, 1]$ alors $F(X)$ est une variable uniforme sur $[0, 1]$. On vérifie aisément ce résultat en notant que pour chaque $u \in [0, 1]$, nous avons

$$\mathbb{P}(F(X) \leq u) = \mathbb{P}(X \leq F^{-1}(u)) = F \circ F^{-1}(u) = u$$

Cette propriété est souvent utilisée en pratique non pas pour simuler des variables uniformes mais plutôt pour vérifier si un échantillon de données statistiques indépendantes suit bien une loi donnée!

Illustrons graphiquement cette technique d'inversion. La nature de la fonction de répartition décrit les accroissement de masse de probabilité entre les points. Pour des variables aléatoires réelles admettant une densité de probabilité p , la accroissement de la fonction de répartition

$$F(x) = \mathbb{P}(X \leq x) = \int_{]-\infty, x]} p(y) dy$$

entre les points x_1 et x_2 correspondent à l'intégrale de la densité de probabilité sur le segment $[x_1, x_2]$

$$F(x_2) - F(x_1) = \mathbb{P}(X \in [x_1, x_2]) = \int_{x_1}^{x_2} p(x) dx$$

La fonction de répartition d'une densité de probabilité p ayant deux modes séparés présentera deux paliers de croissance sur chacun d'entre eux. De plus les variations d'aires sont toujours plus faibles sur les bords de l'axe des réels. Ceci se traduit par des fonctions de répartition $F(x)$ convergeant vers 0, lorsque $x \downarrow -\infty$, et vers 1, lorsque $x \uparrow +\infty$.

La figure précédente illustre bien la répartition sur les masses de probabilités d'une série de points choisis au hasard par le méthode d'inversion

$$X_1 = F^{-1}(U_1), X_2 = F^{-1}(U_2), \dots, X_n = F^{-1}(U_n)$$

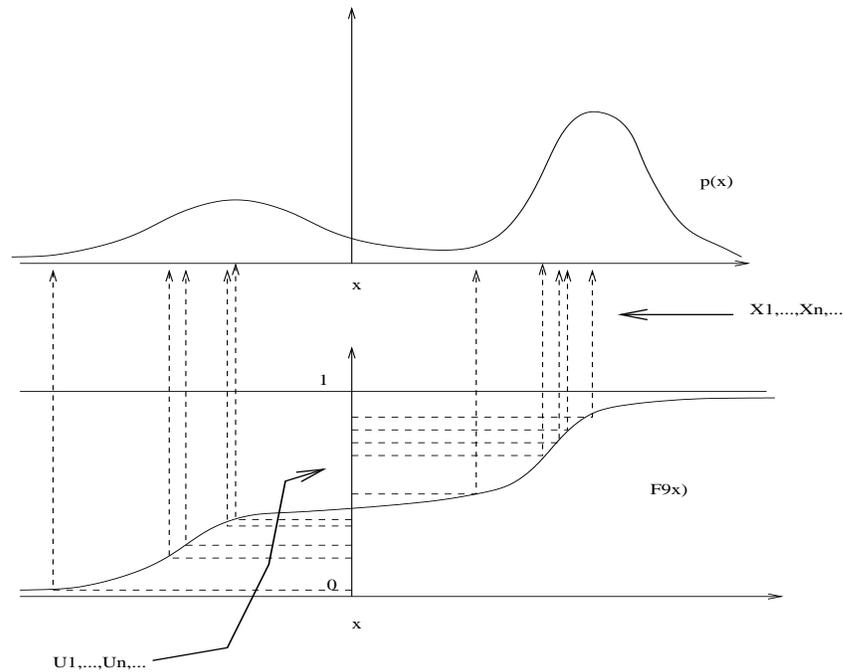


FIG. 2.1 – Méthode d'inversion

où U_1, U_2, \dots, U_n désignent une suite de variables uniformes sur $[0, 1]$.

Dans les sections suivantes, nous illustrerons la performance de méthode d'inversion de la fonction de répartition avec les algorithmes de simulation des lois géométriques, de Cauchy, de Rayleigh, et de Weibull. Nous essayerons dans chaque situation d'offrir une preuve mathématique rigoureuse des formules d'inversion associée à ces diverses lois. Le lecteur ayant aucune connaissance sur les techniques de changement de variable dans les intégrales à de fortes chances de s'ennuyer, voire de s'ennerver! Afin d'éviter de telles crises, notre recommandation dans ce cas est d'éviter tout simplement ces démonstrations, et de passer directement à la description des algorithmes de simulation de ces lois.

2.1.2 Retour aux lois géométriques

Les lois géométriques correspondent au nombre d'expériences nécessaires pour faire apparaître un événement aléatoire donné. Nous avons définie et étudiée la simulation de ces lois dans la section 1.4.3.

Ces distributions temporelles ou de longueurs discrètes sont des modèles simplifiés de phénomènes aléatoires naturels souvent plus complexes. En biologie, ces lois sont parfois utilisés pour simuler des phénomènes de morphogenèse dans la croissance des conifères, des longueurs d'exons de gènes. En physique, la loi géométrique est plutôt utilisée

pour décrire grossièrement des instants de désintégration de particules radioactives, ou d'autre type d'absorption de particules dans des puits de potentiels. En économie, ces lois peuvent représenter des dates de hausse ou de chute des valeurs de produits financiers.

Cette distribution géométrique est l'analogie de la loi exponentielle de paramètre

$$\lambda = -\log(1 - p).$$

En effet, si X est une variable exponentielle de paramètre λ alors, pour tout $n \geq 1$, nous avons

$$\mathbb{P}(n - 1 \leq X < n) = \lambda \int_{n-1}^n e^{-\lambda x} dx = e^{-(n-1)\lambda} (1 - e^{-\lambda}) = p (1 - p)^{n-1}.$$

Par conséquent la variable $Y = 1 + [X]$ est distribuée selon une loi géométrique de paramètre p . En utilisant les techniques de simulation de lois exponentielles décrites précédemment, on vérifie sans trop de peine que la variable

$$Y = 1 + [\log(U)/\log(1 - p)]$$

avec U uniforme sur $[0, 1]$, suit encore une loi géométrique de paramètre p . On pourra noter que ce procédé de simulation est bien plus rapide que celui présenté dans la section 1.4.3. Une seule variable uniforme permet de décrire l'apparition d'un événement aléatoire de probabilité donnés, sans simuler les échecs !

2.1.3 Les lois de Cauchy

La loi de Cauchy est la hantise des probabilistes. Cette distribution de probabilités particulière n'admet aucune moyenne, et a fortiori aucun moments ! Il est donc impossible d'effectuer des calculs élémentaires, et y développer une quelconque intuition.

Cette distribution d'aléas est souvent associée à des phénomènes aléatoires très oscillants. En optique, cette loi de probabilité est parfois utilisée pour modéliser les variations de l'indice de réfraction d'un milieu liées à la fréquence du rayonnement monochromatique passant de l'air à ce milieu. En économie, ces lois représentent plutôt les fluctuations d'amplitudes des logarithmes des rapports journaliers de prix d'actifs financiers.

Pour être plus précis, une variable de Cauchy X de paramètre $\sigma > 0$, est distribuée sur tout segment $[a, b]$ de la droite réelle selon la loi

$$\mathbb{P}(X \in [a, b]) = \int_a^b \frac{\sigma}{\pi (x^2 + \sigma^2)} dx.$$

La fonction de répartition F de X est définie pour chaque $x \in \mathbb{R}$ par la formule

$$F(x) = \int_{-\infty}^x \frac{\sigma}{\pi(y^2 + \sigma^2)} dy = \frac{1}{\sigma} \int_{-\infty}^x \frac{dy}{\pi((y/\sigma)^2 + 1)}$$

En effectuant le changement de variable

$$z = y/\sigma \in] - \infty, x/\sigma] \implies dz = dy/\sigma$$

la fonction F est définie plus simplement par la formule

$$F(x) = \int_{-\infty}^{x/\sigma} \frac{dz}{\pi(z^2 + 1)}$$

Il reste à poser

$$z = \tan \theta \stackrel{\text{déf.}}{=} \frac{\sin \theta}{\cos \theta} \quad \text{avec} \quad \theta \in] - \pi/2, \arctan(x/\sigma)]$$

Dan ce cas, nous avons

$$dz = \frac{\partial z}{\partial \theta} d\theta = \frac{\cos^2 \theta + \sin^2 \theta}{\cos^2 \theta} d\theta = (1 + \tan^2 \theta) d\theta \implies \frac{dz}{1 + z^2} = d\theta$$

et par conséquent

$$F(x) = \frac{1}{\pi} \int_{-\pi/2}^{\arctan(x/\sigma)} d\theta = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x}{\sigma}\right)$$

Cette fonction de répartition est inversible, et son inverse F^{-1} est donné pour chaque $u \in [0, 1]$ par la formule

$$F^{-1}(u) = \sigma \tan\left(\pi\left(u - \frac{1}{2}\right)\right).$$

La simulation d'une loi de Cauchy est donc claire, on simule tout d'abord une variable uniforme U sur l'intervalle $]0, 1[$, et on pose

$$X = F^{-1}(U) = \sigma \tan\left(\pi\left(U - \frac{1}{2}\right)\right)$$

D'après ce qui précède, cette variable est bien une variable aléatoire de Cauchy de paramètre $\sigma > 0$. On pourra noter que la variable aléatoire angulaire

$$V = \pi\left(U - \frac{1}{2}\right)$$

est uniforme sur l'intervalle $] - \pi/2, \pi/2[$. L'algorithme de simulation peut donc s'interpréter comme la tangente d'un angle choisi au hasard dans l'intervalle $] - \pi/2, \pi/2[$

$$X = \sigma \tan V$$

Le paramètre de dilatation σ vient augmenter ou diminuer l'amplitude de cette tangente.

2.1.4 Les lois de Rayleigh-Weibull

Les lois de Rayleigh-Weibull sont des distributions de probabilités sur la demi-droite des réels positifs. Tout comme les lois exponentielles, ces distributions permettent de décrire des instants, des durées, des longueurs, ou des tailles de files d'attente aléatoires. Cependant, les lois de Rayleigh sont plutôt associées à des phénomènes aléatoires temporels bien plus rapides, ou à des longueurs aléatoires de plus faibles amplitudes. Elle sont ainsi utilisées pour décrire des bruits de mesure dans des récepteurs de transmission, des instants de dégradation de matériel, des durées d'appels dans certains réseaux de télécommunication, des temps d'accès sur des sites internet, etc.

Une variable aléatoire X de Rayleigh de paramètre $\sigma > 0$ est distribuée sur la demi droite $[0, \infty[$ selon la loi

$$\mathbb{P}(X \in dx) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} 1_{[0, \infty[}(x) dx.$$

La fonction de répartition F de X est définie pour chaque $x \in [0, \infty[$ par la formule explicite suivante

$$F(x) = \mathbb{P}(X \leq x) = - \int_0^x \frac{\partial}{\partial y} \left(e^{-\frac{y^2}{2\sigma^2}} \right) dy = 1 - \exp \left[-\frac{x^2}{2\sigma^2} \right]$$

Cette fonction est clairement inversible, et son inverse est donné par la fonction

$$u \in [0, 1[\longrightarrow F^{-1}(u) = \sigma \sqrt{-\log(1-u)} \in [0, \infty[$$

La simulation d'une variable de Rayleigh de paramètre $\sigma > 0$ consiste donc à prendre la racine du logarithme d'une variable U uniforme sur $[0, 1]$

$$X = \sigma \sqrt{-\log U}$$

Les lois de Weibull sont de simples extensions des lois de Rayleigh. Plus précisément, une variable aléatoire X de Weibull de paramètres $a, b > 0$, est distribuée selon la loi

$$\mathbb{P}(X \in dx) = \frac{a}{b^a} x^{a-1} e^{-(x/b)^a} 1_{[0, \infty[}(x) dx.$$

Avec ces notations, une loi de Rayleigh de paramètre σ est donc une loi de Weibull de paramètres

$$a = 2 \quad \text{et} \quad b = \sigma$$

Par des raisonnements analogues aux précédents, la fonction de répartition F de X est définie explicitement pour chaque $x \in [0, \infty[$ par la formule suivante

$$F(x) = \int_0^x \frac{\partial}{\partial z} (-e^{-(z/b)^a}) dz = 1 - e^{-(x/b)^a}$$

Cette fonction est clairement inversible et son inverse F^{-1} est donné pour chaque $u \in [0, 1[$ par

$$F^{-1}(u) = b (-\log(1 - u))^{1/a}.$$

Si U est une variable aléatoire uniforme sur $]0, 1]$ alors

$$X = b \times (-\log U)^{1/a}$$

est une variable aléatoire de Weibull de paramètres $a, b > 0$. Pour conclure, on pourra noter que les variables exponentielles de paramètre $\lambda = 1$ données par

$$Y = -\log U$$

et les variables de Weibull de paramètres (a, b) sont reliées par la formule suivante

$$X = b \times Y^{1/a}$$

En terme de fiabilité, la variable exponentielle de référence correspond à des instants de pannes purement accidentelles. Les paramètres (a, b) des variables de Weibull représentent les degrés d'usure d'un appareil. Par exemple, les durées de fonctionnement d'un appareil plus usé sont d'autant plus courtes que le paramètre b est faible, ou a est grand.

Tout une panoplie de variables réelles peuvent être simulées en utilisant ce principe d'inversion de la fonction de répartition. Son application nécessite néanmoins d'avoir une description explicite de l'inverse de cette fonction. Pour plus d'information, nous renvoyons le lecteur à l'ouvrage [?].

2.2 Changements de variables

2.2.1 Introduction

Il existe une variété d'expériences aléatoires X prenant leurs valeurs dans des espaces multi-dimensionnels \mathbb{R}^d , avec $d \geq 1$. C'est le cas des répartitions de température sur une surface ou dans un volume, les évolutions de photons dans des tissus lors d'examens d'imagerie médicale, les perturbations de mesures sur chaque composante d'un capteur électronique, les impacts de météorites sur une surface, etc. Dans la plupart de ces situations, la probabilité pour que le résultat aléatoire soit compris dans un pavé $A \subset \mathbb{R}^d$ correspond à **un volume** entre le pavé en question, et le graphe d'une fonction positive

$$p : x \in \mathbb{R}^d \longrightarrow p(x) \in [0, \infty[$$

Cette aire est appelé l'**intégrale** de la densité de probabilité p sur la pavé A , et on la note

$$\mathbb{P}(X \in A) = \int_A p(x) dx$$

2.2.2 Lois uniformes sur des cubes

Comment choisir un mot au hasard dans un texte, ou dans un manuscrit de 200 pages d'épaisseur ? Existe-t-il un procédé de simulation permettant de colorier un point choisi au hasard dans un tableau ? Les pixels d'une image numérique peuvent-ils être allumés au hasard, et indépendamment les uns des autres ? Peut-on faire apparaître un personnage de jeu vidéo dans une pièce virtuelle, de façon totalement imprévisible ? Un adepte du pointillisme peut-il peindre par petites touches totalement hasardeuses un tableau ? Peut-on choisir au hasard un point dans un glaçon cubique ?

D'un point de vue probabiliste, toutes ces petites questions reviennent à trouver un algorithme de simulation d'une variable aléatoire uniforme sur des rectangles, des cubes, ou tout autre pavé d'un espace euclidien \mathbb{R}^d , de dimension $d \geq 1$.

Nous commencerons par examiner le cas des pavés dans le plan. La densité de probabilité d'une variable aléatoire uniforme $X = (X_1, X_2)$ sur un rectangle quelconque

$$A = ([a_1, b_1] \times [a_2, b_2])$$

est donnée par la fonction

$$p : x \in C \mapsto p(x) = \frac{1}{\text{Aire}(A)} 1_A(x)$$

La probabilité pour que le point aléatoirement choisi X tombe dans un sous-ensemble régulier $B \subset A$ est simplement donnée par le rapport des aires

$$\mathbb{P}(X \in B) = \frac{\text{Aire}(B)}{\text{Aire}(A)}$$

La fonction p étant une fonction indicatrice sur le carré, de hauteur $1/\text{Aire}(A)$, on retrouve le fait que la probabilité de l'évènement $\{X \in B\}$ est le volume du pavé de base B et de hauteur $1/\text{Aire}(A)$.

Il est très aisé de simuler le point $X = (X_1, X_2)$ à l'aide d'un couple de variables uniformes et indépendants (U_1, U_2) sur le segment $[0, 1]$. Il suffit de poser

$$X_1 = a_1 + (b_1 - a_1) U_1 \quad \text{et} \quad X_2 = a_2 + (b_2 - a_2) U_2$$

Pour vérifier cette assertion, on considère la fonction θ qui à un point $(u_1, u_2) \in [0, 1]^2$ associe le point

$$(x_1, x_2) = \theta(u_1, u_2) = (a_1 + (b_1 - a_1)u_1, a_2 + (b_2 - a_2)u_2) \in A = ([a_1, b_1] \times [a_2, b_2])$$

Cette transformation envoie le carré unité $[0, 1]^2$ sur le carré

$$A = \theta([0, 1]^2)$$

Les points (U_1, U_2) choisis sur $[0, 1]^2$ sont ainsi envoyés sur des points $(X_1, X_2) = \theta(U_1, U_2)$ du carré A . Mieux encore, la fonction θ est une bijection entre ces deux

ensembles. Autrement dit, θ est une correspondance bi-univoque entre $[0, 1]^2$ et A . L'inverse de θ est donné par la formule suivante

$$(u_1, u_2) = \theta^{-1}(x_1, x_2) = \left(\frac{x_1 - a_1}{b_1 - a_1}, \frac{x_2 - a_2}{b_2 - a_2} \right)$$

Ceci nous permet simplement de vérifier que l'on a

$$\mathbb{P}(X \in B) = \mathbb{P}(\theta(U_1, U_2) \in B) = \mathbb{P}((U_1, U_2) \in \theta^{-1}(B)) = \text{Aire}(\theta^{-1}(B))$$

Il nous reste cependant à vérifier que l'on a

$$\text{Aire}(\theta^{-1}(B)) = \frac{\text{Aire}(B)}{\text{Aire}(A)}$$

Avant de vérifier cette assertion, nous commencerons par noter que la transformation θ est loin de préserver les aires des carrés. Par exemple, le carré unité $[0, 1]^2$ d'aire 1 est envoyé sur un carré

$$A = \theta([0, 1]^2) = ([a_1, b_1] \times [a_2, b_2])$$

dont l'aire est donnée par la formule

$$\text{Aire}(A) = (b_1 - a_1)(b_2 - a_2)$$

Pour être plus précis, nous avons les associations d'aires infinitésimales suivantes

$$du_1 du_2 \rightsquigarrow \left| \frac{\partial u_1}{\partial x_1} \frac{\partial u_2}{\partial x_2} - \frac{\partial u_1}{\partial x_2} \frac{\partial u_2}{\partial x_1} \right| \times dx_1 dx_2 \quad (2.1)$$

Pour plus détails concernant ces associations d'aires, nous renvoyons le lecteur à la section 2.2.3 consacrée aux variations d'aires engendrées par des transformations planes. Dans notre cas, nous avons

$$\frac{\partial u_1}{\partial x_1} = \frac{1}{(b_1 - a_1)}, \quad \frac{\partial u_2}{\partial x_2} = \frac{1}{(b_2 - a_2)}, \quad \text{et} \quad \frac{\partial u_1}{\partial x_2} = 0 = \frac{\partial u_2}{\partial x_1}$$

Ceci nous conduit aux associations d'aires infinitésimales

$$du_1 du_2 \rightsquigarrow \frac{1}{\text{Aire}(A)} \times dx_1 dx_2$$

On obtient finalement le résultat recherché

$$\begin{aligned} \text{Aire}(\theta^{-1}(B)) &= \mathbb{P}((U_1, U_2) \in \theta^{-1}(B)) \\ &= \int_{\theta^{-1}(B)} du_1 du_2 = \frac{1}{\text{Aire}(A)} \times \int_B dx_1 dx_2 = \frac{\text{Aire}(B)}{\text{Aire}(A)} \end{aligned}$$

La simulation d'une variable uniforme sur le cube unité $[0, 1]^d$ consiste à prendre un vecteur

$$(U_1, \dots, U_d)$$

formé de d variables aléatoires indépendantes $(U_k)_{k=1, \dots, d}$ de loi uniforme sur $[0, 1]$. Par exemple, choisir un point au hasard (U_1, U_2, U_3) dans un cube $[0, 1]^3$ revient à choisir chacune des coordonnées U_k uniformément sur chacune des trois arêtes de base $[0, 1]$. De même, pour choisir un point uniformément distribué sur un pavé de la forme

$$[a_1, b_1] \times \dots \times [a_n, b_n]$$

il suffit de poser

$$(X_1, \dots, X_n) = (a_1 + (b_1 - a_1)U_1, \dots, a_n + (b_n - a_n)U_n)$$

Pour vérifier que chacune des composantes X_k est uniformément distribuée sur chaque arête $[a_k, b_k]$, on note simplement que pour tout segment $[\alpha_k, \beta_k] \subset [a_k, b_k]$, nous avons

$$\mathbb{P}(X_k \in [\alpha_k, \beta_k]) = \mathbb{P}\left(U_k \in \left[\frac{\alpha_k - a_k}{b_k - a_k}, \frac{\beta_k - a_k}{b_k - a_k}\right]\right) = \frac{\beta_k - \alpha_k}{b_k - a_k}$$

2.2.3 Lois uniformes sur des surfaces

Les règles de l'intégration traduisent les aires globales de surfaces $A \subset \mathbb{R}^2$ en termes d'aires infinitésimales

$$\text{Aire}(A) = \int_A du_1 du_2$$

La mesure $du_1 du_2$ désigne la mesure de Lebesgue sur le plan, cette mesure s'exprime simplement par le fait que l'aire d'un rectangle est donnée par la formule classique apprise au lycée

$$\text{Aire}([a, b] \times [c, d]) = \int_{[a, b] \times [c, d]} du_1 du_2 = (b - a) \times (d - c)$$

On doit interpréter $du_1 du_2$ comme une aire virtuelle d'un petit carré infinitésimal dont les cotés sont de longueurs respectives du_1 et du_2 .

L'intégrale consiste simplement à sommer les aires tous les petits carrés infinitésimaux contenus dans la surface en question. Considérons désormais une transformation θ suffisamment régulière du plan dans lui même. Cette transformation peut s'interpréter comme une déformation physique et élastique d'une surface plane. Chaque point (u_1, u_2) est envoyé sur un nouveau point

$$(x_1, x_2) = \theta(u_1, u_2)$$

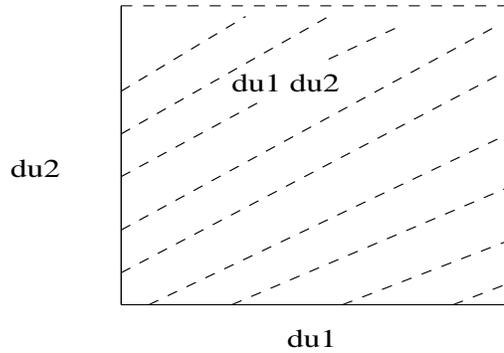


FIG. 2.2 – Aire de carrés infinitésimaux

Les petits carrés infinitésimaux

$$[u_1, u_1 + du_1] \times [u_2, u_2 + du_2]$$

de surface $du_1 du_2$ autour de points (u_1, u_2) sont alors transformés en petits carrés

$$[x_1, x_1 + dx_1] \times [x_2, x_2 + dx_2]$$

de surface $dx_1 dx_2$, autour des points transformés $(x_1, x_2) = \theta(u_1, u_2)$. Les variations d'étirement du plan subies par l'action de la transformation élastique θ se traduisent par une variation des longueurs des cotés des carrés. Plus précisément, dans la base des longueurs de référence (du_1, du_2) sur chaque coordonnées, nous avons

$$dx_1 = \frac{\partial x_1}{\partial u_1} du_1 + \frac{\partial x_1}{\partial u_2} du_2 \quad \text{et} \quad dx_2 = \frac{\partial x_2}{\partial u_1} du_1 + \frac{\partial x_2}{\partial u_2} du_2$$

L'objectif est de calculer l'aire des rectangles

$$[x_1, x_1 + dx_1] \times [x_2, x_2 + dx_2]$$

en fonction des variations de longueurs infinitésimales

$$\frac{\partial x_i}{\partial u_j}, \quad \text{avec } i, j \in \{1, 2\}$$

Autrement dit, il convient de calculer l'aire du parallélogramme engendré par les vecteurs

$$dx_1 = \begin{pmatrix} \frac{\partial x_1}{\partial u_1} \\ \frac{\partial x_1}{\partial u_2} \end{pmatrix} \quad \text{et} \quad dx_2 = \begin{pmatrix} \frac{\partial x_2}{\partial u_1} \\ \frac{\partial x_2}{\partial u_2} \end{pmatrix}$$

dans la base de référence (du_1, du_2) .

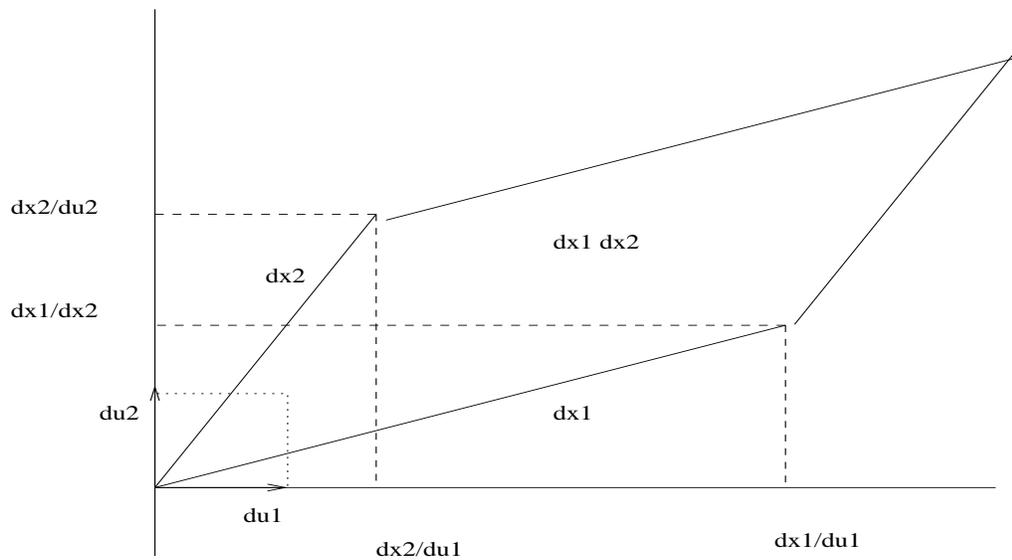


FIG. 2.3 – Aire infinitésimale $dx_1 dx_2$

Si $(\widehat{dx_1, dx_2})$ désigne l'angle formé par ces deux vecteurs, nous avons

$$\sin(\widehat{dx_1, dx_2}) = \frac{h}{\sqrt{\left(\frac{\partial x_2}{\partial u_1}\right)^2 + \left(\frac{\partial x_2}{\partial u_2}\right)^2}}$$

où h désigne la hauteur du parallélogramme. Il est aisé de vérifier que l'aire de ce parallélogramme est donnée par la formule produit classique

$$\begin{aligned} h \times (\text{base } dx_1) &= h \times \sqrt{\left(\frac{\partial x_1}{\partial u_1}\right)^2 + \left(\frac{\partial x_1}{\partial u_2}\right)^2} \\ &= \sqrt{\left(\frac{\partial x_1}{\partial u_1}\right)^2 + \left(\frac{\partial x_1}{\partial u_2}\right)^2} \times \sqrt{\left(\frac{\partial x_2}{\partial u_1}\right)^2 + \left(\frac{\partial x_2}{\partial u_2}\right)^2} \times \sin(\widehat{dx_1, dx_2}) \end{aligned}$$

Il nous reste donc à évaluer le sinus de l'angle $(\widehat{dx_1, dx_2})$. Pour simplifier nos calculs, on considère les vecteurs normalisés sur le cercle unité

$$dy_1 = \frac{dx_1}{\sqrt{\left(\frac{\partial x_1}{\partial u_1}\right)^2 + \left(\frac{\partial x_1}{\partial u_2}\right)^2}} \quad \text{et} \quad dy_2 = \frac{dx_2}{\sqrt{\left(\frac{\partial x_2}{\partial u_1}\right)^2 + \left(\frac{\partial x_2}{\partial u_2}\right)^2}}$$

On remarque que l'on a

$$(\widehat{dx_1, dx_2}) = (\widehat{dy_1, dy_2})$$

En identifiant le cercle unité à l'ensemble des nombres complexes de module 1, nous avons

$$\begin{aligned} dy_1 = e^{ia_1} &= \cos a_1 + i \sin a_1 = \frac{\frac{\partial x_1}{\partial u_1}}{\sqrt{\left(\frac{\partial x_1}{\partial u_1}\right)^2 + \left(\frac{\partial x_1}{\partial u_2}\right)^2}} + i \frac{\frac{\partial x_1}{\partial u_2}}{\sqrt{\left(\frac{\partial x_1}{\partial u_1}\right)^2 + \left(\frac{\partial x_1}{\partial u_2}\right)^2}} \\ dy_2 = e^{ia_2} &= \cos a_2 + i \sin a_2 = \frac{\frac{\partial x_2}{\partial u_1}}{\sqrt{\left(\frac{\partial x_2}{\partial u_1}\right)^2 + \left(\frac{\partial x_2}{\partial u_2}\right)^2}} + i \frac{\frac{\partial x_2}{\partial u_2}}{\sqrt{\left(\frac{\partial x_2}{\partial u_1}\right)^2 + \left(\frac{\partial x_2}{\partial u_2}\right)^2}} \end{aligned}$$

où a_1 , et a_2 désignent respectivement les angles entre l'axe des abscisses et les vecteurs dy_1 , et dy_2 .

Avec ces notations, nous avons

$$(\widehat{dx_1, dx_2}) = (\widehat{dy_1, dy_2}) = a_2 - a_1$$

Il reste alors à remarquer que l'on a

$$\cos(a_2 - a_1) + i \sin(a_2 - a_1) = e^{i(a_2 - a_1)} = e^{ia_2} e^{-ia_1} = dy_2 \times \overline{dy_1} \quad (2.2)$$

où $\overline{dy_1}$ désigne le conjugué de dy_1

$$\overline{dy_1} = e^{-ia_1} = \cos a_1 - i \sin a_1$$

En identifiant les parties complexes dans le produit (2.2), on en conclut que

$$\sin(a_2 - a_1) = \sin a_1 \cos a_2 - \cos a_1 \sin a_2$$

et finalement

$$\sin(\widehat{dx_1, dx_2}) = \frac{1}{\sqrt{\left(\frac{\partial x_1}{\partial u_1}\right)^2 + \left(\frac{\partial x_1}{\partial u_2}\right)^2}} \frac{1}{\sqrt{\left(\frac{\partial x_2}{\partial u_1}\right)^2 + \left(\frac{\partial x_2}{\partial u_2}\right)^2}} \left(\frac{\partial x_1}{\partial u_2} \frac{\partial x_2}{\partial u_1} - \frac{\partial x_1}{\partial u_1} \frac{\partial x_2}{\partial u_2} \right)$$

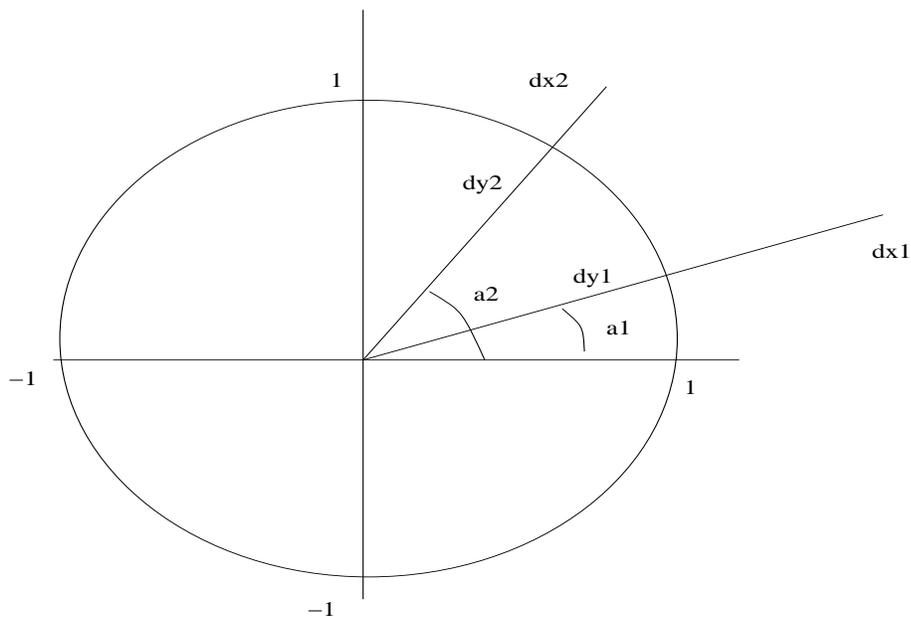


FIG. 2.4 – Angles a_1, a_2 sur le cercle trigonométrique

L'aire recherchée est donc donnée par la formule

$$h \times (\text{base } dx_1) = \left(\frac{\partial x_1}{\partial u_2} \frac{\partial x_2}{\partial u_1} - \frac{\partial x_1}{\partial u_1} \frac{\partial x_2}{\partial u_2} \right)$$

En conclusion, nous avons les associations d'aires infinitésimales

$$dx_1 dx_2 \rightsquigarrow \left(\frac{\partial x_1}{\partial u_2} \frac{\partial x_2}{\partial u_1} - \frac{\partial x_1}{\partial u_1} \frac{\partial x_2}{\partial u_2} \right) du_1 du_2$$

Le terme $du_1 du_2$ représente l'unité d'aire infinitésimale dans laquelle nous avons effectué nos calculs. Ces formules témoignent du fait que l'aire d'une surface \mathcal{D} suffisamment régulière

$$\text{Aire}(\mathcal{D}) = \int_{\mathcal{D}} du_1 du_2$$

est transformée par θ en une surface $\theta(\mathcal{D})$ d'aire

$$\text{Aire}(\theta(\mathcal{D})) = \int_{\theta(\mathcal{D})} dx_1 dx_2 = \int_{\mathcal{D}} \left| \frac{\partial x_1}{\partial u_2} \frac{\partial x_2}{\partial u_1} - \frac{\partial x_1}{\partial u_1} \frac{\partial x_2}{\partial u_2} \right| du_1 du_2$$

Plus généralement, pour des fonctions intégrables sur la surface $\theta(\mathcal{D})$, nous avons la formule de changement de variables

$$\int_{\theta(\mathcal{D})} f(x_1, x_2) dx_1 dx_2 = \int_{\mathcal{D}} f(\theta(u_1, u_2)) \left| \frac{\partial \theta^1}{\partial u_2} \frac{\partial \theta^2}{\partial u_1} - \frac{\partial \theta^1}{\partial u_1} \frac{\partial \theta^2}{\partial u_2} \right| du_1 du_2$$

avec la transformation

$$(u_1, u_2) \in \mathcal{D} \mapsto \theta(u_1, u_2) = (\theta^1(u_1, u_2), \theta^2(u_1, u_2)) = (x_1, x_2) \in \theta(\mathcal{D})$$

2.2.4 Lois uniformes sur des disques

Comment placer au hasard le diamant d'une platine sur un disque vinyl? Peut-on élaborer une pizza en plaçant tous les ingrédients de façon totalement hasardeuse? Comment simuler un mauvais joueur de flechette sur ordinateur? Un géomètre a-t-il la possibilité de choisir un point $X = (X_1, X_2)$ au hasard dans le disque unité

$$\mathcal{D} = \{(x_1, x_2) \in \mathbb{R}^2 ; x_1^2 + x_2^2 < 1\} ?$$

Comme dans le cas des choix de points au hasard dans des pavés examiné dans la section précédente, nous commencerons par noter que la probabilité de choisir (X_1, X_2) dans un sous ensemble régulier $B \subset \mathcal{D}$ correspond simplement au rapport des surfaces de B et de \mathcal{D}

$$\mathbb{P}((X_1, X_2) \in B) = \frac{\text{Aire}(B)}{\text{Aire}(\mathcal{D})} \quad \text{avec} \quad \text{Aire}(\mathcal{D}) = \int_{\mathcal{D}} dx_1 dx_2 = \pi$$

Tout point (x_1, x_2) du disque peut être paramétrisé par la donnée d'un angle a et d'un rayon r . Plus précisément, nous avons la formule de changement de coordonnées polaires

$$\begin{cases} x_1 &= r \cos a \\ x_2 &= r \sin a \end{cases}$$

L'application

$$\begin{aligned} \theta : \mathcal{D}' = \mathcal{D} - ([0, 1[\times \{0\}) &\longrightarrow]0, 1[\times]0, 2\pi[\\ (x_1, x_2) &\mapsto \theta(x_1, x_2) = (r, a) \end{aligned}$$

transforme donc de façon bi-univoque le disque unité \mathcal{D} , auquel on a oté soigneusement le petit segment $([0, 1[\times \{0\})$, en un joli petit rectangle

$$\theta(\mathcal{D}') =]0, 1[\times]0, 2\pi[$$

Si (X_1, X_2) est uniforme sur \mathcal{D} , quelle est la loi du point

$$\theta(X_1, X_2) = (R, A)$$

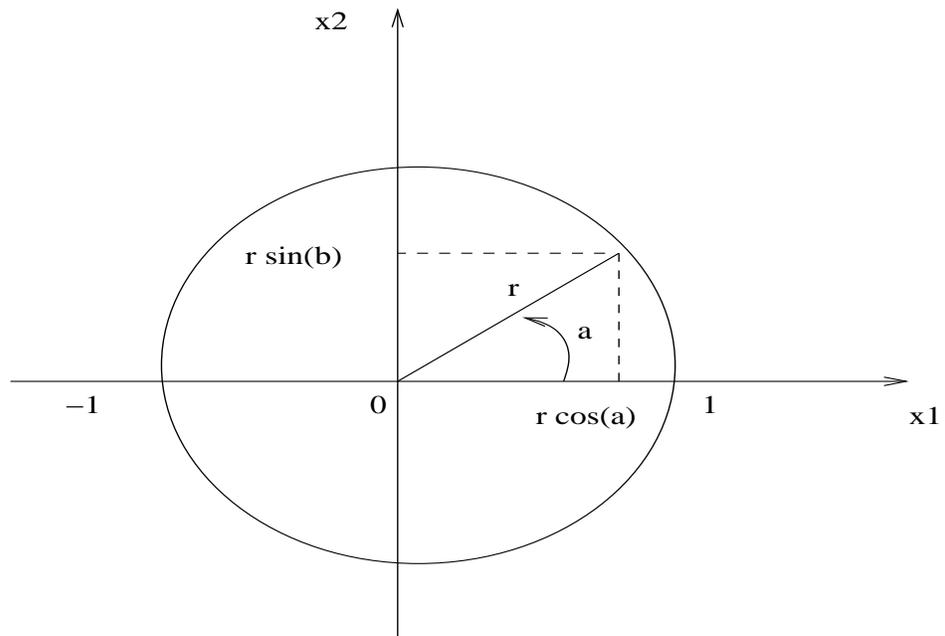


FIG. 2.5 – Coordonnées polaires

sur le pavé $]0, 1[\times]0, 2\pi[$? Pour répondre à cette question, il convient de noter les associations d'éléments de surface infinitésimale

$$dx_1 dx_2 \rightsquigarrow \left| \frac{\partial x_1}{\partial r} \frac{\partial x_2}{\partial a} - \frac{\partial x_1}{\partial a} \frac{\partial x_2}{\partial r} \right| \times dr da$$

Dans notre situation, nous avons

$$\frac{\partial x_1}{\partial r} = \cos a, \quad \frac{\partial x_2}{\partial r} = \sin a$$

et

$$\frac{\partial x_1}{\partial a} = -r \sin a, \quad \frac{\partial x_2}{\partial a} = r \cos a$$

Par conséquent, nous avons

$$\left| \frac{\partial x_1}{\partial r} \frac{\partial x_2}{\partial a} - \frac{\partial x_1}{\partial a} \frac{\partial x_2}{\partial r} \right| = r (\cos^2 a + \sin^2 a) = r$$

et donc

$$dx_1 dx_2 \rightsquigarrow r dr da$$

La densité de probabilité du point (X_1, X_2)

$$p^{X_1, X_2}(x_1, x_2) = \frac{1}{\pi} 1_{\mathcal{D}}(x_1, x_2) dx_1, dx_2$$

est donc transformée par θ en la densité de probabilité du point $\theta(X_1, X_2) = (R, A)$

$$p^{R, A}(r, a) = \frac{1}{\pi} 1_{\theta(\mathcal{D})}(r, a) r dr da$$

Pour approfondir notre étude du point (R, A) , il convient de noter que la fonction $p^{R, A}$ peut s'exprimer sous la forme d'un produit de densités

$$p^{R, A}(r, a) = [2r 1_{[0, 1]}(r) dr] \times \left[\frac{1}{2\pi} 1_{[0, 2\pi]}(a) da \right]$$

Ceci nous informe que les deux composantes de rayon R et d'angle A sont indépendantes, de densités respectives

$$p^R(r) = 2r 1_{[0, 1]}(r) dr \quad \text{et} \quad p^A(a) = \frac{1}{2\pi} 1_{[0, 2\pi]}(a) da$$

L'angle A est donc uniforme sur $[0, 2\pi]$, le rayon aléatoire R est distribuée selon une loi rendant plus probables les rayons proches de 1.

Si U désigne une variable aléatoire uniforme sur $[0, 1]$ alors la variable $(2\pi U)$ à la même loi que A , en effet pour tout $a \in [0, 2\pi]$

$$\mathbb{P}(2\pi U \leq a) = \mathbb{P}\left(U \leq \frac{a}{2\pi}\right) = \frac{a}{2\pi} = \mathbb{P}(A \leq a) = \frac{1}{2\pi} \int_0^a da'.$$

D'autre part, si U' désigne une variable uniforme sur $[0, 1]$ alors $\sqrt{U'}$ a la même loi que R . En effet pour chaque $r \in [0, 1]$ on a bien

$$\mathbb{P}\left(\sqrt{U'} \leq r\right) = \mathbb{P}(U' \leq r^2) = r^2 = \mathbb{P}(R \leq r) = \int_0^r \frac{\partial}{\partial s}(s^2) ds.$$

En choisissant U et U' indépendantes, on a donc montré les équivalences en loi

$$\theta(X_1, X_2) \stackrel{\text{loi}}{=} (2\pi U, \sqrt{U'})$$

Par conséquent, si U et U' sont deux variables uniformes et indépendantes sur $[0, 1]$ alors le point (X_1, X_2) de coordonnées

$$\begin{cases} X_1 &= \sqrt{U'} \cos(2\pi U) \\ X_2 &= \sqrt{U'} \sin(2\pi U) \end{cases}$$

est uniformément réparti sur le disque unité \mathcal{D} . Bien entendu, nous pouvons vérifier directement ce résultat en utilisant la transformation entre les variables uniformes

(u, u') sur $[0, 1]^2$, avec $u \notin \{\frac{1}{4}, \frac{3}{4}\}$, et les points (x_1, x_2) sur le disque donnée par la formule

$$u = \frac{1}{2\pi} \arctan \frac{x_2}{x_1} \quad \text{et} \quad u' = x_1^2 + x_2^2$$

Un simple calcul de dérivées permet d'obtenir les équations suivantes

$$\begin{aligned} \frac{\partial u}{\partial x_1} &= -\frac{x_2}{2\pi x_1^2} \frac{1}{1 + (x_2/x_1)^2} = -\frac{x_2}{2\pi x_1^2} \frac{1}{1 + (x_2/x_1)^2} \\ \frac{\partial u}{\partial x_2} &= \frac{1}{2\pi x_1} \frac{1}{1 + (x_2/x_1)^2} = \frac{1}{2\pi x_1} \frac{1}{1 + (x_2/x_1)^2} \end{aligned}$$

et

$$\frac{\partial u'}{\partial x_1} = 2x_1, \quad \frac{\partial u'}{\partial x_2} = 2x_2$$

On obtient alors aisément les les associations d'éléments de surface infinitésimale recherchés :

$$dudu' \rightsquigarrow \left| \frac{\partial u}{\partial x_1} \frac{\partial u'}{\partial x_2} - \frac{\partial u}{\partial x_2} \frac{\partial u'}{\partial x_1} \right| \times dx_1 dx_2 = \frac{1}{\pi} dx_1 dx_2$$

Notons pour conclure que l'on peut facilement décrire les loi des coordonnées cartésiennes X_1 et X_2 . Par exemple la loi du demi-cercle est la loi de X_1

$$\begin{aligned} \mathbb{P}(X_1 \in dx_1) &= \mathbb{P}(X_1 \in dx_1, X_2 \in]-1, 1[) = \frac{1}{\pi} \left(\int_{-1}^1 1_{\mathcal{D}}(x_1, x_2) dx_2 \right) 1_{]-1, 1[}(x_1) dx_1 \\ &= \left(\frac{1}{\pi} \int_{-\sqrt{1-x_1^2}}^{\sqrt{1-x_1^2}} dx_2 \right) 1_{]-1, 1[}(x_1) dx_1 = \frac{2}{\pi} \sqrt{1-x_1^2} 1_{]-1, 1[}(x_1) dx_1 \end{aligned}$$

D'après ce qui précède on en conclut que la variable aléatoire

$$X_1 = \sqrt{U'} \cos(2\pi U)$$

est distribuée selon la loi du demi-cercle.

2.2.5 L'algorithme Box-Muller

Les fluctuations aléatoires de nature gaussienne font vraisemblablement partie des phénomènes aléatoires les plus fréquemment observés dans la nature. Cette fréquence s'explique en grande partie par le théorème central de la limite. Ce fameux résultat probabiliste affirme que tout phénomène aléatoire résultant d'une accumulation de petites fluctuations indépendantes de même nature est nécessairement de nature gaussienne! On utilise donc de tels aléas gaussiens en ingénierie pour décrire les fluctuations d'erreurs de capteurs électroniques, ou tout autre type d'erreurs de modèles. En physique, les répartitions de chaleur, ou de fluides peuvent s'interpréter

comme des excitations et des collisions de particules. Les modèles cinétiques utilisés dans ce contexte sont souvent formés de particules browniennes évoluant dans l'espace selon des processus de diffusion de type gaussien.

Il existe une bonne demi-douzaine d'algorithmes permettant de simuler des phénomènes gaussiens. Le plus connu est sans nul doute l'algorithme de Box-Muller. La célébrité de cette technique provient de sa rapidité d'exécution. En effet, l'algorithme de Box-Muller permet de construire deux variables indépendantes et gaussiennes simplement à l'aide de deux variables uniformes et indépendantes sur $[0, 1]$.

Nous avons choisi de présenter cet algorithme en utilisant l'étude de la uniforme sur le disque unité \mathcal{D} décrite dans la section 2.2.4. On note (R, A) les coordonnées polaires d'un point choisi au hasard dans \mathcal{D} . Nous avons montré précédemment que ces variables aléatoires sont indépendantes, et distribuées selon la densité

$$p^{R,A}(r, a) = [2r \mathbb{1}_{]0,1[}(r)] \left[\frac{1}{2\pi} \mathbb{1}_{]0,2\pi[}(a) \right]$$

On opère maintenant une transformation non linéaire sur le rayon

$$S = \sqrt{-4 \log R}$$

On rappelle que l'on peut simuler le rayon R à l'aide d'une variable uniforme V sur $]0, 1[$ en posant

$$R = \sqrt{V}$$

Par conséquent, le point S peut être simulé en posant

$$S = \sqrt{-2 \log R^2} = \sqrt{-2 \log V}$$

La loi de S est donnée par un calcul élémentaire sur la fonction de répartition

$$\mathbb{P}(S \geq s) = \mathbb{P}(R \leq e^{-s^2/4}) = \int_0^{e^{-s^2/4}} \frac{\partial r^2}{\partial r} dr = e^{-s^2/2}$$

Par conséquent S est distribuée sur la demi-droite réelle $\mathbb{R}_+ = [0, \infty[$ avec la densité

$$p^S(s) = -\frac{\partial}{\partial s} \mathbb{P}(S \geq s) = s e^{-s^2/2} \mathbb{1}_{\mathbb{R}_+}(s)$$

On en déduit que les variables (S, A) sont indépendantes distribuées sur $(\mathbb{R}_+ \times]0, 2\pi[)$ avec la densité

$$p^{S,A}(s, a) = \left[s e^{-s^2/2} \mathbb{1}_{\mathbb{R}_+}(s) \right] \left[\frac{1}{2\pi} \mathbb{1}_{]0,2\pi[}(a) \right]$$

Examinons la distribution dans l'espace du point

$$\begin{cases} X_1 &= S \cos A \\ X_2 &= S \sin A \end{cases}$$

Pour déterminer cette loi de probabilité, il convient tout d'abord de noter que le point (X_1, X_2) est le transformé du point (S, A)

$$(X_1, X_2) = \theta(S, A)$$

avec la fonction

$$\theta : (s, a) \in \left(\mathbb{R}_+ \times \left(]0, 2\pi[- \left\{ \frac{\pi}{2}, \frac{3\pi}{2} \right\} \right) \right) \rightsquigarrow \theta(s, a) = (s \cos a, s \sin a) \in (\mathbb{R}^2 - \{ \{0\} \times \mathbb{R} \})$$

On obtient aisément les formules d'inversion

$$\theta(s, a) = (x_1, x_2) \iff \left(a = \arctan(x_2/x_1) \quad \text{et} \quad s = \sqrt{x_1^2 + x_2^2} \right)$$

Les variations des coordonnées (a, s) en fonction du couple (x_1, x_2) sont données ci-après

$$\frac{\partial a}{\partial x_1} = \frac{1}{1 + (x_2/x_1)^2} \times \left(-\frac{x_2}{x_1^2} \right), \quad \frac{\partial a}{\partial x_2} = \frac{1}{1 + (x_2/x_1)^2} \times \frac{1}{x_1}$$

et

$$\frac{\partial s}{\partial x_1} = \frac{x_1}{\sqrt{x_1^2 + x_2^2}}, \quad \frac{\partial s}{\partial x_2} = \frac{x_2}{\sqrt{x_1^2 + x_2^2}}$$

On en déduit les associations de densité de probabilité

$$\begin{aligned} \frac{1}{2\pi} s e^{-s^2/2} ds da &\rightsquigarrow \frac{1}{2\pi} \sqrt{x_1^2 + x_2^2} e^{-\frac{1}{2}(x_1^2 + x_2^2)} \left| \frac{\partial s}{\partial x_1} \frac{\partial a}{\partial x_2} - \frac{\partial s}{\partial x_2} \frac{\partial a}{\partial x_1} \right| dx_1 dx_2 \\ &= \frac{1}{2\pi} e^{-\frac{1}{2}(x_1^2 + x_2^2)} \end{aligned}$$

Autrement dit, X_1 et X_2 sont deux variables gaussiennes centrées et normées indépendantes. L'algorithme de Box-Muller correspond à l'expression de ce résultat en terme de variables uniformes sur $[0, 1]$. Si (U, V) sont deux variables indépendantes et uniformes sur $[0, 1]$ alors

$$X_1 = \sqrt{-2 \log U} \cos(2\pi V)$$

et

$$X_2 = \sqrt{-2 \log U} \sin(2\pi V)$$

sont deux variables gaussiennes centrées et normées indépendantes. Pour générer sur ordinateur un couple de variables indépendantes gaussiennes (X_1, X_2) de moyennes (m_1, m_2) et d'écart types (σ_1, σ_2) on posera tout simplement

$$X_1 = m_1 + \sigma_1 \sqrt{-2 \log U} \cos(2\pi V)$$

et

$$X_2 = m_2 + \sigma_2 \sqrt{-2 \log U} \sin(2\pi V).$$

Exercice 2.2.1 Nous proposons de montrer que la projection de la loi gaussienne dans \mathbb{R}^2 sur le cercle unité devient uniforme.

Soit (X_1, X_2) un couple de v.a. gaussiennes, indépendantes, centrées et normées sur \mathbb{R} . Montrer que la v.a. $R = \sqrt{X_1^2 + X_2^2}$ est indépendante du point aléatoire $P = \left(\frac{X_1}{\sqrt{X_1^2 + X_2^2}}, \frac{X_2}{\sqrt{X_1^2 + X_2^2}} \right)$, qui est uniformément distribué sur le cercle C de rayon 1 (centré en l'origine) de \mathbb{R}^2 . Calculer la loi exacte de R .

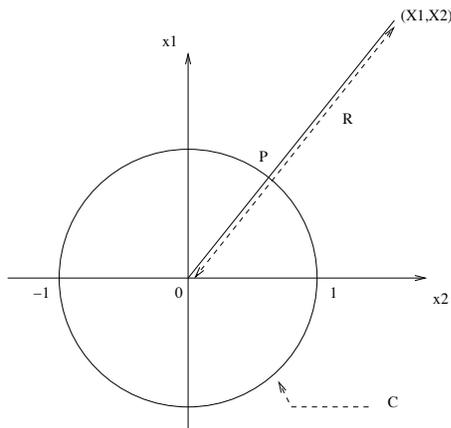


FIG. 2.6 – Point uniforme sur le cercle

Exercice 2.2.2 Soit (X_1, X_2, X_3) un triplé de v.a. gaussiennes, indépendantes, centrées et normées sur \mathbb{R} . Montrer que la v.a. $R = \sqrt{X_1^2 + X_2^2 + X_3^2}$ est indépendante du point aléatoire $P = \left(\frac{X_1}{\sqrt{X_1^2 + X_2^2 + X_3^2}}, \frac{X_2}{\sqrt{X_1^2 + X_2^2 + X_3^2}}, \frac{X_3}{\sqrt{X_1^2 + X_2^2 + X_3^2}} \right)$, qui est uniformément distribué sur la sphère S^2 de rayon 1, et centré en l'origine, de \mathbb{R}^3 . Calculer la loi exacte de R .

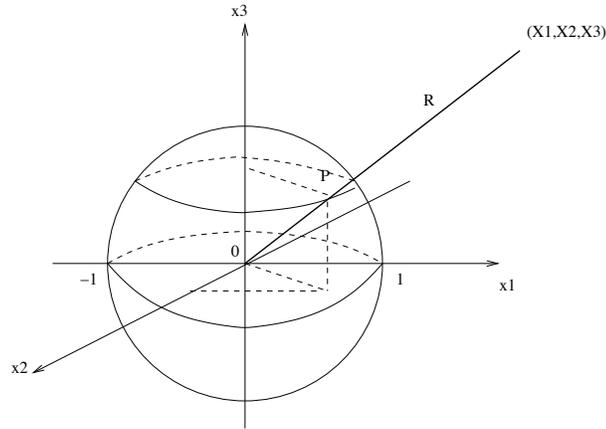


FIG. 2.7 – Point uniforme sur la sphère

Exercice 2.2.3 Soit (X_1, \dots, X_n) une suite de n v.a. gaussiennes, indépendantes, centrées et normées sur \mathbb{R} . Montrer que la v.a. $R_n = \sqrt{\sum_{i=1}^n X_i^2}$ est indépendante du point aléatoire $P_n = \left(\frac{X_i}{\sqrt{\sum_{j=1}^n X_j^2}}\right)_{1 \leq i \leq n}$, qui est uniformément distribué sur la sphère S^{n-1} de rayon 1, et centré en l'origine, de \mathbb{R}^n . Calculer la loi exacte de R_n , ainsi que le volume de la sphère S^{n-1} .

Exercice 2.2.4 1. Vérifier que la somme $\chi_n^2 = \sum_{i=1}^n X_i^2$, de n carrés de v.a. gaussiennes centrées et normées $(X_p)_{1 \leq p \leq n}$, suit une loi gamma de paramètres $(1/2, n/2)$

$$\mathbb{P}^{\chi_n^2}(dx) = 1_{\mathbb{R}_+}(x) \left(\frac{1}{2}\right)^{n/2} \frac{1}{\Gamma(n/2)} e^{-x/2} x^{\frac{n}{2}-1} dx$$

Montrer que $\mathbb{E}(\chi_n^2) = n$, et $\text{Var}(\chi_n^2) = 2n$. En statistique mathématique, les lois gamma de paramètres $(1/2, n/2)$, sont appelés des Chi-deux à n degrés de liberté.

2. Montrer que la v.a. χ_n^2 est indépendante du vecteur normalisé

$$(X_i/\sqrt{\chi_n^2})_{1 \leq i \leq n}$$

et vérifier que $\chi_n = \sqrt{\sum_{i=1}^n X_i^2}$ est distribué selon la loi de probabilité

$$\mathbb{P}^{\chi_n}(dx) = 1_{\mathbb{R}_+}(x) \frac{1}{2^{\frac{n-1}{2}} \Gamma(n/2)} x^{n-1} e^{-\frac{x^2}{2}} dx$$

2.3 Méthode d'acceptation-rejet

Nous allons décrire dans cette section est une méthode de simulation universelle. Cette technique est fondée sur un mécanisme d'apprentissage naturel et très simple. On propose dans un premier temps des variables aléatoires distribuées selon une loi de référence *dominante* et facile à simuler. Dans un second temps on accepte ou on refuse ces variables. Le critère d'acceptation est simplement basé sur une comparaison de cette mesure de référence avec la distribution de la variable que l'on souhaite simuler. Cette technique universelle permet notamment de simuler tout phénomène aléatoire à valeurs dans une régions compacte d'un espace topologique, et distribué sur cette région selon une densité de probabilité bornée. On illustrera ce procédé pour la simulation de lois telles que les lois conditionnelles ou la loi du demi-cercle et on présentera l'algorithme polaire de simulation d'une loi gaussienne.

2.3.1 Les lois de référence

Commençons par nous donner un couple de densités de probabilités (p, q) sur un espace \mathbb{R}^d , avec $d \geq 1$. On dit que la densité p **domine** q lorsque l'on a

$$q(x) \leq m \times p(x)$$

pour une certaine constante positive $m > 0$. Notons que dans cette situation, le support de q doit être contenu dans celui de p . Autrement dit, lorsque p domine q on a nécessairement

$$p(x) = 0 \implies q(x) = 0$$

Pour fixer les idées, on notera que les couples de densités suivantes vérifient cette propriété

- 1) $p(x) = \frac{1}{2} 1_{[-1,1]}$ $q(x) = \frac{2}{\pi} \sqrt{1-x^2} 1_{[-1,1]}(x)$
- 2) $p(x_1, x_2) = \frac{1}{4} 1_{[-1,1]^2}(x_1, x_2)$ $q(x_1, x_2) = \frac{1}{\pi} 1_{\{(y_1, y_2) \in \mathbb{R}^2 : y_1^2 + y_2^2 < 1\}}(x_1, x_2)$
- 3) $p(x) = \frac{1}{\pi(x^2+1)}$ $q(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

Le premier exemple correspond à la loi du demi-cercle q , et la densité uniforme p sur l'intervalle $[-1, 1]$. Dans cette situation, nous avons

$$m = \sup_{x \in [-1,1]} \frac{q(x)}{p(x)} = \frac{4}{\pi}$$

Dans le second cas, p est le densité uniforme sur le carré $[-1, 1]^2$, et q la mesure uniforme sur le disque unité. Dans cette situation, on peut noter que l'on a encore

$$m = \sup_{(x_1, x_2) \in [-1,1]^2} \frac{q(x_1, x_2)}{p(x_1, x_2)} = \frac{4}{\pi}$$

Le dernier exemple avec les densités gaussienne et de Cauchy, souligne le fait que la propriété précédente est automatiquement satisfaite pour des densités strictement positives. Dans cette situation, il convient de noter que

$$\frac{\partial}{\partial x} (q/p)(x) = \sqrt{\frac{\pi}{2}} x(1-x^2) e^{-x^2/2}$$

La fonction q/p est donc croissante sur $] -\infty, -1]$, décroissante sur $[-1, 0]$, puis à nouveau croissante sur $[0, 1]$, et enfin décroissante sur $[1, \infty[$. On a de plus

$$\lim_{x \rightarrow -/+ \infty} (q/p)(x) = 0, \quad (q/p)(-1) = \sqrt{\frac{2\pi}{e}} = (q/p)(1), \quad \text{et enfin} \quad (q/p)(0) = 1$$

Le maximum de (q/p) est donc atteint aux points $x \in \{-1, 1\}$. Autrement dit, nous avons

$$m = \sup_{x \in \mathbb{R}} (q/p)(x) = \sqrt{\frac{2\pi}{e}}$$

On peut aussi noter que les restrictions $q(x)$ d'une densité donnée $p(x)$ sur des régions $A \subset \mathbb{R}^d$ (telles que $\int_A q(y) dy > 0$) vérifient clairement cette propriété

$$p(x) = 0 \implies q(x) = \frac{1}{\int_A p(y) dy} 1_A(x) p(x) = 0$$

Un contre exemple est facile à construire, le couple de densités uniformes suivant

$$p(x) = \frac{1}{2} 1_{[-1,1]}(x) \quad \text{et} \quad q(x) = \frac{1}{4} 1_{[-2,2]}(x)$$

ne vérifie pas cette propriété simplement à cause du fait que le support $[-1, 1]$ de la densité p ne contient pas le support $[-2, 2]$ de p .

2.3.2 Les taux d'acceptation

Simulons une variable aléatoire X_1 distribué selon la densité dominante p . Par construction, nous avons

$$0 \leq \frac{1}{m} \times \frac{q(X_1)}{p(X_1)} \leq 1$$

Autrement dit, le segment $[0, 1]$ est partagé en deux intervalles

$$[0, 1] = \left[0, \frac{1}{m} \times \frac{q(X_1)}{p(X_1)} \right] \cup \left[\frac{1}{m} \times \frac{q(X_1)}{p(X_1)}, 1 \right]$$

On simule alors un nombre uniforme U_1 sur $[0, 1]$. Si ce nombre U_1 appartient au premier segment $\left[0, \frac{1}{m} \times \frac{q(X_1)}{p(X_1)} \right]$, on garde X_1 , et l'on pose

$$Y = X_1$$

Dans le cas contraire, on recommence toute l'opération. On simule une nouvelle variable aléatoire X_2 distribué selon la densité dominante p , et une variable uniforme U_2 sur $[0, 1]$. Si ce nombre U_2 appartient au premier segment $\left[0, \frac{1}{m} \times \frac{q(X_2)}{p(X_2)}\right]$, on garde X_2 , et l'on pose

$$Y = X_2$$

Dans le cas contraire, on reproduit entièrement le procédé. On simule une nouvelle variable aléatoire X_3 distribué selon la densité dominante p , et une variable uniforme U_3 sur $[0, 1]$. Si ce nombre U_3 appartient au premier segment $\left[0, \frac{1}{m} \times \frac{q(X_3)}{p(X_3)}\right]$, on garde X_3 , et l'on pose

$$Y = X_3$$

Dans le cas contraire, on réitère le procédé, jusqu'au premier instant T_m où l'on finit par accepter la variable simulée selon la loi dominante

$$Y = X_{T_m}$$

En terme mathématiques, ce procédé de simulation revient à produire une suite de variables X_n indépendantes et distribuées selon la densité dominante, ainsi qu'une suite U_n de variables indépendantes et uniformes sur $[0, 1]$. On note ensuite

$$T_m = \inf \left\{ n \geq 1 : U_n \leq \frac{1}{m} \times \frac{q(X_n)}{p(X_n)} \right\}$$

le premier instant où l'on a $U_n \leq \frac{1}{m} \times \frac{q(X_n)}{p(X_n)}$. La magie de ce procédé de simulation est résumée dans les deux lignes suivantes. Nous venons de construire un couple (X_{T_m}, T_m) de variables indépendantes telles que :

- La variable aléatoire X_{T_m} est distribuée selon la densité q .
- Le temps aléatoire d'acceptation T_m suit une loi géométrique de paramètre $\frac{1}{m}$. Autrement dit, nous avons pour tout $n \geq 1$

$$\mathbb{P}(T_m = n) = \frac{1}{m} \left(1 - \frac{1}{m}\right)^{n-1}.$$

En particulier le nombre moyen de boucles nécessaires pour obtenir une réalisation de la variable X_{T_m} est $\mathbb{E}(T_m) = m$.

2.3.3 Vérification mathématique

La vérification mathématique du tour de magie présenté dans la section précédente ne nécessite aucun raisonnement analytique. La dynamique de la preuve est purement ptobabiliste !

On commence par remarquer que pour chaque variable proposée selon la loi dominante, disons X_n , la probabilité pour que la variable uniforme U_n tombe dans le premier segment $\left[0, \frac{1}{m} \times \frac{q(X_n)}{p(X_n)}\right]$ est donnée par la probabilité conditionnelle

$$\mathbb{P}\left(U_n \leq \frac{1}{m} \times \frac{q(X_n)}{p(X_n)} \mid X_n\right) = \frac{1}{m} \times \frac{q(X_n)}{p(X_n)}$$

En intégrant sur toutes les valeurs que peut prendre X_n , on obtient

$$\mathbb{P}\left(U_n \leq \frac{1}{m} \times \frac{q(X_n)}{p(X_n)}\right) = \int \frac{1}{m} \times \frac{q(x_n)}{p(x_n)} p(x_n) dx_n = \frac{1}{m} \int q(x_n) dx_n = \frac{1}{m}$$

Autrement dit, lors du déroulement dynamique de l'algorithme stochastique, la probabilité d'un succès est égale à $1/m$. Inversement, la probabilité d'échec et de refus d'une variable est donnée par la formule suivante

$$\mathbb{P}\left(U_n \geq \frac{1}{m} \times \frac{q(X_n)}{p(X_n)}\right) = 1 - \frac{1}{m}$$

Dans la pratique, il est donc très important de bien choisir la loi dominante p telle que le nombre m soit le plus proche possible de 1. L'idéal étant bien entendu de choisir $p = q$, de sorte à avoir $m = 1$; mais la distribution cible q est souvent trop complexe à simuler. D'où l'intérêt de ce tour de magie.

Pour chaque région régulière $A \subset \mathbb{R}^d$, et pour chaque entier $n \geq 1$, nous avons aussi

$$\begin{aligned} \mathbb{P}(X_n \in A, U_n \leq \frac{1}{m} \times \frac{q(X_n)}{p(X_n)}) &= \mathbb{E}(1_A(X_n) \mathbb{P}(U_n \leq \frac{1}{m} \times \frac{q(X_n)}{p(X_n)} \mid X_n)) \\ &= \mathbb{E}\left(1_A(X_n) \frac{1}{m} \times \frac{q(X_n)}{p(X_n)}\right) \\ &= \int_A \frac{1}{m} \times \frac{q(x_n)}{p(x_n)} p(x_n) dx_n = \frac{1}{m} \int_A q(x_n) dx_n. \end{aligned}$$

On utilise maintenant les décompositions événementielles suivantes

$$\begin{aligned} \{X_{T_m} \in A\} &= \{\text{Il existe un instant } n \geq 1 \text{ tel que } T_m = n \text{ et } X_n \in A\} \\ &= \cup_{n \geq 1} \{X_n \in A, T_m = n\} \end{aligned}$$

et

$$\begin{aligned} &\{X_n \in A, T_m = n\} \\ &= \left\{X_n \in A, U_n \leq \frac{1}{m} \frac{q(X_n)}{p(X_n)} \text{ et pour les instants précédents } 1 \leq k < n \text{ on a } U_k > \frac{1}{m} \frac{q(X_k)}{p(X_k)}\right\} \\ &= \left\{X_n \in A, U_n \leq \frac{1}{m} \frac{q(X_n)}{p(X_n)}\right\} \cap \left(\bigcap_{k=1}^{n-1} \left\{U_k > \frac{1}{m} \frac{q(X_k)}{p(X_k)}\right\}\right) \end{aligned} \tag{2.3}$$

D'après les deux premières remarques, et en utilisant l'indépendance des couples de variables aléatoires (U_k, X_k) , on obtient

$$\mathbb{P}(X_{T_m} \in A) = \frac{1}{m} \left[\int_A q(x) dx \right] \sum_{n \geq 1} \left(1 - \frac{1}{m}\right)^{n-1} = \int_A q(x) dx$$

En prenant $A = \mathbb{R}^d$ dans (2.3) on trouve finalement

$$\mathbb{P}(T_m = n) = \mathbb{P}\left(U_n \leq \frac{1}{m} \frac{q(X_n)}{p(X_n)}\right) \prod_{k=1}^{n-1} \mathbb{P}\left(U_k > \frac{1}{m} \frac{q(X_k)}{p(X_k)}\right) = \frac{1}{m} \left(1 - \frac{1}{m}\right)^{n-1}.$$

L'indépendance entre les variables X_{T_m} et T_m provient simplement du fait que

$$\begin{aligned} \mathbb{P}(X_{T_m} \in A, T_m = n) &= \mathbb{P}(X_n \in A, T_m = n) \\ &= \left[\int_A q(x) dx \right] \frac{1}{m} \left(1 - \frac{1}{m}\right)^{n-1} \\ &= \mathbb{P}(X_{T_m} \in A) \mathbb{P}(T_m = n) \end{aligned}$$

En posant $z = \frac{1}{m}$, on a enfin

$$\begin{aligned} \mathbb{E}(T_m) &= \sum_{n \geq 1} n(1-z)^{n-1} z = -z \frac{\partial}{\partial z} \left(\sum_{n \geq 1} (1-z)^n \right) \\ &= -z \frac{\partial}{\partial z} \left(\frac{1}{z} - 1 \right) = \frac{1}{z} = m \end{aligned}$$

2.4 Files d'attentes exponentielles

Cette section concerne la simulation d'arrivées aléatoires de clients virtuels dans une file d'attente d'un magasin tout aussi virtuel. Ce type de phénomène aléatoire se prête à diverses interprétations. En effet, les instants d'arrivées des clients correspondent tout simplement à des dates aléatoires. Ces dernières peuvent alors s'interpréter comme des durées de fonctionnement de machines, ou des instants de panne, des durées de connexions, des temps de traitement d'information, des arrivées d'automobiles dans les bouchons du périphérique parisien, etc.

La simulation sur ordinateur de tels processus est assez claire. Il suffit de simuler les arrivées successives de chacun des clients. Les écarts temporels entre les arrivées de deux clients successifs peuvent être de nature très variée. Ces instants peuvent être distribués sur l'axe des temps selon des lois exponentielles, des lois de Rayleigh, de Weibull, etc.

La section 2.4.1 concerne la simulation de formations aléatoires des files d'attente exponentielles. Ces processus d'arrivées de clients à des dates aléatoires exponentielles

sont d'un usage très fréquent en pratique. Ces horloges aléatoires apparaissent notamment de façon naturelle dans la construction de processus markoviens à temps continu sautant de temps à autre, et à des dates exponentielles sur d'autres site de l'espace d'état. Dans la section 2.4.2, nous aborderons la distribution des tailles de files d'attente exponentielles.

2.4.1 La loi Gamma

La simulation d'une file d'attente où les clients arrivent aléatoirement à une fréquence d'exponentielle est très simple à réaliser. En effet, il suffit de se donner une suite de variables aléatoires indépendantes S_n de même loi exponentielle

$$\mathbb{P}(S_n \in ds) = \lambda e^{-\lambda s} 1_{[0, \infty[}(s) ds \quad \text{avec } \lambda > 0$$

Dans la section 1.4.5 consacrée aux horloges exponentielles, nous avons vu que le paramètre λ représente l'intensité d'arrivée des clients. Plus précisément, nous avons $\mathbb{E}(S_n) = 1/\lambda$. La fréquence d'arrivée des clients est donc d'autant plus grande que le paramètre λ est petit.

Chaque variable S_n représente le temps d'attente aléatoire entre l'arrivée du $(n - 1)^{\text{ième}}$ client, et le suivant. Autrement dit, si l'on note T_n la date d'arrivée du $n^{\text{ième}}$ client, nous avons

$$T_n = T_{n-1} + S_n = S_1 + \dots + S_{n-1} + S_n \quad \text{avec } T_0 = 0$$

Deux questions naturelles viennent à l'esprit : Quelles est la loi des instants d'arrivée du groupe de clients, et quelle est la loi d'arrivée du $n^{\text{ième}}$ individu ?

Pour calculer la loi du vecteur aléatoire (T_1, \dots, T_n) , on se donne une fonction f suffisamment régulière de $[0, \infty[^n$ dans \mathbb{R} , et l'on commence par remarquer que l'on a

$$\begin{aligned} \mathbb{E}(f(T_1, \dots, T_n)) &= \int f(t_1, \dots, t_n) p^{(T_1, \dots, T_n)}(t_1, \dots, t_n) dt_1 \dots dt_n \\ &= \mathbb{E}(f(S_1, S_1 + S_2, \dots, S_1 + \dots + S_n)) \\ &= \int_0^\infty \dots \int_0^\infty f(s_1, s_1 + s_2, \dots, s_1 + \dots + s_n) \lambda^n e^{-\lambda (s_1 + \dots + s_n)} ds_1 \dots ds_n \end{aligned}$$

Pour identifier la densité de probabilité $p^{(T_1, \dots, T_n)}$ du vecteur aléatoire (T_1, \dots, T_n) , on effectue le changement de variable

$$\begin{cases} t_1 = s_1 \\ t_2 = s_1 + s_2 \\ t_3 = s_1 + s_2 + s_3 \\ \vdots \\ t_n = s_1 + s_2 + \dots + s_n \end{cases}$$

Par construction le vecteur des temps (t_1, \dots, t_n) est formée de nombres croissants. Plus précisément, nous avons

$$(t_1, \dots, t_n) \in C_n = \{(t_1, \dots, t_n) \in \mathbb{R}_+^n ; t_1 < \dots < t_n\}.$$

Inversement, nous avons

$$\begin{cases} s_1 = t_1 \\ s_2 = t_2 - t_1 \\ s_3 = t_3 - t_2 \\ \vdots \\ s_n = t_n - t_{n-1} \end{cases}$$

En effectuant un simple changement de variable dans chaque intégrale, on obtient les associations infinitésimales

$$ds_1 = dt_1, \quad ds_2 = dt_2, \quad \dots \quad ds_n = dt_n$$

et finalement

$$\mathbb{E}(f(T_1, \dots, T_n)) = \int_{C_n} f(t_1, \dots, t_n) \lambda^n e^{-\lambda t_n} dt_1 \dots dt_n$$

En résumé, la loi du vecteur (T_1, \dots, T_n) est donnée par

$$\mathbb{P}((T_1, \dots, T_n) \in d(t_1, \dots, t_n)) = \lambda^n e^{-\lambda t_n} 1_{C_n}(t_1, \dots, t_n) dt_1 \dots dt_n$$

Pour calculer la loi des arrivées aléatoires T_n du $n^{\text{ième}}$ client, on commence par noter que

$$\begin{aligned} \mathbb{P}(T_n \in dt_n) &= \mathbb{P}(T_n \in dt_n, (T_1, \dots, T_{n-1}) \in C_{n-1}(t_n)) \\ &= \lambda^n e^{-\lambda t_n} dt_n \times \left[\int_{C_{n-1}(t_n)} dt_1 \dots dt_{n-1} \right] \end{aligned} \quad (2.4)$$

avec les ensembles

$$\begin{aligned} C_{n-1}(t_n) &= \{(t_1, \dots, t_{n-1}) \in \mathbb{R}_+^{n-1} ; 0 < t_1 < \dots < t_{n-1} < t_n\} \\ &= \{(t_1, \dots, t_{n-1}) \in C_{n-1} \text{ tels que } t_{n-1} < t_n\} \end{aligned}$$

Il nous reste donc à calculer les intégrales de la forme suivante :

$$\begin{aligned} I_n(t) &= \int 1_{C_n(t)}(t_1, \dots, t_n) dt_1 \dots dt_n \\ &= \int_{0 < t_1 < \dots < t_{n-1} < t_n < t} dt_1 \dots dt_n \\ &= \int_0^t \left[\int_{0 < t_1 < \dots < t_{n-1} < t_n} dt_1 \dots dt_{n-1} \right] dt_n = \int_0^t I_{n-1}(t_n) dt_n \end{aligned}$$

Une simple récurrence nous permet alors de vérifier que l'on a

$$\begin{aligned}
 I_1(t) = \int_0^t ds = t &\Rightarrow I_2(t) = \int_0^t I_1(s) ds = \frac{t^2}{2!} \\
 &\Rightarrow I_3(t) = \int_0^t I_2(s) ds = \frac{t^3}{3!} \\
 &\vdots \\
 &\Rightarrow I_n(t) = \frac{t^n}{n!}
 \end{aligned}$$

En utilisant (2.4), on obtient finalement

$$\begin{aligned}
 \mathbb{P}(T_n \in dt) &= I_{n-1}(t) \times \lambda^n e^{-\lambda t} 1_{[0, \infty[}(t) dt \\
 &= \frac{(t)^{n-1}}{(n-1)!} \times \lambda^n e^{-\lambda t} 1_{[0, \infty[}(t) dt \\
 &= \lambda e^{-\lambda t} \times \frac{(\lambda t)^{n-1}}{(n-1)!} 1_{[0, \infty[}(t) dt
 \end{aligned}$$

Lorsque $\lambda = 1$, les arrivées aléatoires du $(n+1)^{\text{ième}}$ client sont distribuées selon la loi Gamma de paramètre $(n+1)$ définie par la formule suivante

$$\mathbb{P}(T_{n+1} \in dt) = e^{-t} \frac{t^n}{n!} 1_{[0, \infty[}(t) dt \tag{2.5}$$

2.4.2 La loi de Poisson

Les distributions de Poisson sont des lois de comptage du nombre de fois où un évènement aléatoire se réalise sur un intervalle de temps donné. Ces lois sont souvent utilisées pour estimer le nombre de clients en attente dans un réseau de communication, le nombre de pannes d'une machine dans une journée, le nombre de mutations génétiques au cours de l'évolution d'une espèce, etc. En termes mathématiques, on dit qu'une variable aléatoire N à valeurs entières suit une loi de Poisson de paramètre γ lorsque l'on a

$$\mathbb{P}(N = n) = e^{-\gamma} \frac{\gamma^n}{n!}$$

pour chaque entier $n \in \mathbb{N}$. Le paramètre γ correspond à la valeur moyenne de la variable de Poisson N

$$\mathbb{E}(N) = \sum_{n \geq 1} e^{-\gamma} n \frac{\gamma^n}{n!} = \gamma \left[\sum_{n \geq 1} e^{-\gamma} \frac{\gamma^{n-1}}{(n-1)!} \right] = \gamma$$

La forme de la loi Gamma introduite en (2.5) est très proche de la loi de Poisson. L'inversion des rôles des paramètres t et n est basée sur les formules suivantes

$$\frac{\gamma^n}{n!} e^{-\gamma} = \int_{\gamma}^{+\infty} \frac{\partial}{\partial t} \left(-e^{-t} \frac{t^n}{n!} \right) dt$$

et

$$\frac{\partial}{\partial t} \left(-e^{-t} \frac{t^n}{n!} \right) = e^{-t} \frac{t^n}{n!} - e^{-t} \frac{t^{n-1}}{(n-1)!}.$$

Lorsque les variables aléatoires S_n sont distribuées selon une loi exponentielle de paramètre $\lambda = 1$ alors d'après (2.5) on en déduit que

$$\begin{aligned} \frac{\gamma^n}{n!} e^{-\gamma} &= \int_{\gamma}^{+\infty} e^{-t} \frac{t^n}{n!} dt - \int_{\gamma}^{+\infty} e^{-t} \frac{t^{n-1}}{(n-1)!} dt \\ &= \mathbb{P}(T_{n+1} > \gamma) - \mathbb{P}(T_n > \gamma) \\ &= \mathbb{P}(T_{n+1} > \gamma \text{ et } T_n \leq \gamma). \end{aligned}$$

La dernière égalité provient du fait que l'on a

$$A = \{T_n > \gamma\} \subset B = \{T_{n+1} > \gamma\} \quad \text{et} \quad \mathbb{P}(B) = \mathbb{P}(B - A) + \mathbb{P}(A).$$

Ainsi, en posant

$$N = \inf\{n \geq 1 ; T_{n+1} > \gamma\}$$

on a clairement

$$\{N = n\} = \{T_n \leq \gamma \text{ et } T_{n+1} > \gamma\}$$

Cette simple équivalence événementielle nous conduit au procédé de simulation d'une variable de Poisson :

Si S_n sont des variables aléatoires indépendantes de loi exponentielle de paramètre $\lambda = 1$ alors pour tout $\gamma > 0$ la variable aléatoire entière suivante

$$N = \inf\{n \geq 1 ; S_1 + \dots + S_{n+1} > \gamma\}$$

est distribuée sur \mathbb{N} selon une loi de Poisson de paramètre γ .

2.4.3 La statistique d'ordre uniforme

Examinons de plus près la répartition des dates d'arrivées relatives de $(n + 1)$ visiteurs Gamma décrites dans la section 2.4.1

$$(V_1, \dots, V_n) =_{\text{déf.}} \left(\frac{T_1}{T_{n+1}}, \frac{T_2}{T_{n+1}}, \dots, \frac{T_n}{T_{n+1}} \right) \quad (2.6)$$

Ces dates sont rangées dans un ordre croissant dans l'intervalle $[0, 1]$. Le dernier visiteur sert de référence temporelle. Comment se distribuent ces dates aléatoires relatives ? Les premiers visiteurs arrivent-ils généralement plus rapidement que les derniers ? Intuitivement non, toutes ces dates semblent uniformément réparties sur l'intervalle $[0, 1]$. Si cela était le cas, il serait équivalent de simuler une suite indépendantes de dates uniformément réparties sur $[0, 1]$, puis de les ranger ! Cependant, les coûts de calculs pour ordonner cette longue séquence d'instant, uniformes sur $[0, 1]$, seraient

bien plus élevés que ceux nécessaires à la simulation des arrivées exponentielles relatives décrites plus haut. De plus, si ces deux phénomènes aléatoires étaient équivalents, cela impliquerait que les dates d'arrivées relatives (2.6) des visiteurs Gamma seraient indépendantes de l'instant d'arrivée T_{n+1} du dernier visiteur ! Bien que les conséquences de notre hypothèse d'équivalence statistique semblent déroutantes, nous allons vérifier par de simples calculs que notre intuition est bien cohérente.

Pour poursuivre notre étude, il convient de se rappeler que la distribution (2.4) du vecteur (T_1, \dots, T_{n+1}) peut s'écrire sous la forme suivante :

$$\begin{aligned} & \mathbb{P}((T_1, \dots, T_{n+1}) \in d(t_1, \dots, t_{n+1})) \\ &= 1_{C_{n+1}}(t_1, \dots, t_{n+1}) \lambda^{n+1} e^{-\lambda t_{n+1}} dt_1 \dots dt_{n+1} \\ &= (n! 1_{C_n(t_{n+1})}(t_1, \dots, t_n) t_{n+1}^{-n} dt_1 \dots dt_n) \left(\lambda e^{-\lambda t_{n+1}} \frac{(\lambda t_{n+1})^n}{n!} 1_{[0, \infty[}(t_{n+1}) dt_{n+1} \right) \\ &= \mathbb{P}((T_1, \dots, T_n) \in d(t_1, \dots, t_n) \mid T_{n+1} = t_{n+1}) \times \mathbb{P}(T_{n+1} \in dt_{n+1}) \end{aligned}$$

avec les ensembles désormais classiques

$$C_n(t) = \{(t_1, \dots, t_n) \in \mathbb{R}_+^n ; 0 < t_1 < \dots < t_n < t\}.$$

Afin de trouver la loi du vecteur (V_1, \dots, V_n) , on se donne une fonction suffisamment régulière et bornée sur $[0, 1]^n$, et l'on note que

$$\begin{aligned} & \mathbb{E} \left(f \left(\frac{T_1}{T_{n+1}}, \frac{T_2}{T_{n+1}}, \dots, \frac{T_n}{T_{n+1}} \right) \mid T_{n+1} = t_{n+1} \right) \\ &= \int_{C_n(t_{n+1})} f \left(\frac{t_1}{t_{n+1}}, \frac{t_2}{t_{n+1}}, \dots, \frac{t_n}{t_{n+1}} \right) n! t_{n+1}^{-n} dt_1 \dots dt_n \end{aligned}$$

Un simple changement de variable

$$\left(v_1 = \frac{t_1}{t_{n+1}}, \quad v_2 = \frac{t_2}{t_{n+1}}, \quad \dots, \quad v_n = \frac{t_n}{t_{n+1}} \right) \implies (dv_1 \dots dv_n = t_{n+1}^{-n} dt_1 \dots dt_n)$$

permet de vérifier que l'espérance conditionnelle précédente est indépendante de la date d'arrivée du $(n+1)^{\text{ième}}$ visiteur ! Plus précisément, on obtient

$$\begin{aligned} & \mathbb{E} \left(f \left(\frac{T_1}{T_{n+1}}, \frac{T_2}{T_{n+1}}, \dots, \frac{T_n}{T_{n+1}} \right) \mid T_{n+1} = t_{n+1} \right) \\ &= \int_{C_n(1)} f(v_1, \dots, v_n) n! dv_1 \dots dv_n \end{aligned}$$

D'après les calculs d'intégrales effectués à la page 56, on peut observer que l'on a

$$\int_{C_n(1)} dv_1 \dots dv_n = \frac{1}{n!} = \mathbb{P}((U_1, \dots, U_n) \in C_n(1))$$

où (U_1, \dots, U_n) désignent une suite de variables aléatoires indépendantes et indentiquement distribuée selon une loi uniforme sur l'intervalle $[0, 1]$. Le vecteur des dates relatives

$$(V_1, \dots, V_n) =_{\text{déf.}} \left(\frac{T_1}{T_{n+1}}, \frac{T_2}{T_{n+1}}, \dots, \frac{T_n}{T_{n+1}} \right)$$

est donc indépendant de la date T_{n+1} , et il se distribue sur l'intervalle $[0, 1]$ comme une suite ordonnée de nombres aléatoires choisis indépendamment et uniformément sur $[0, 1]$. En résumé, nous avons montré qu'il est statistiquement équivalent **de simuler et ordonner** une suite de variables uniformes et indépendantes sur $[0, 1]$, ou de **simuler une suite d'arrivées exponentielles** et relatives de visiteurs.

Chapitre 3

Chaînes de Markov discrètes

3.1 Introduction

Un processus aléatoire est un phénomène dont une partie de l'évolution temporelle est aléatoire. On rencontre ces processus dans divers domaines de la physique, ou des sciences de l'ingénieur. Par exemple la répartition et l'évolution de la chaleur dans un corps, la turbulence atmosphérique ; c'est aussi le cas des temps d'arrivée d'appels téléphoniques, les erreurs de mesures dues aux perturbations thermiques dans des capteurs électroniques de type radar ou sonar, ou encore l'évolution des cours de marchés boursiers.

La théorie de processus aléatoires est une théorie mathématique très riche, offrant de nombreuses interactions avec diverses branches des mathématiques, telles la théorie des graphes, l'analyse fonctionnelle, la théorie des opérateurs, ainsi que la théorie ergodique des systèmes dynamiques.

Formellement, un processus aléatoire est une succession de variables aléatoires $(X_n)_{n \geq 0}$

$$X_0 \longrightarrow X_1 \longrightarrow \dots \longrightarrow X_n \longrightarrow X_{n+1} \longrightarrow \dots$$

L'indice $n \in \mathbb{N}$ représente la paramètre temporel. Les variables X_n peuvent être à valeurs dans un espace discret E , dans des espaces euclidiens $E = \mathbb{R}^d$, ou dans des espaces plus complexes tels des espaces de chemins ou d'excursions.

Ces modèles trajectoriels sont utiles en biologie dans l'analyse d'évolutions ancestrales de population. On rencontre aussi des processus à valeurs dans des espaces de matrices en physique statistique, ou en analyse d'images. L'étude de tels modèles sort bien entendu du cadre de cet ouvrage.

Les états X_n peuvent évoluer aléatoirement au cours du temps selon des mécanismes plus ou moins complexes. Dans certains cas, les épreuves X_n sont indépendantes ; c'est le cas des séquences de lancers de dés, ou les successions de jets de pièces de monnaies.

Dans d'autres situations, le processus aléatoire est donné par une équation physique récursive de la forme suivante

$$X_n = F_n(X_{n-1}, U_n)$$

La v.a. X_0 désigne la condition initiale du système, F_n des fonctions de dérive déterministes, et enfin U_n des variables aléatoires “indépendantes” des précédents états X_0, \dots, X_{n-1} . De telles séquences sont appelées des *chaînes de Markov*.

En traitement du signal, de tels systèmes peuvent représenter l'évolution temporelle d'un cible dans l'espace. Dans ce contexte, les variables aléatoires U_n ont une double dimension. Elles correspondent à la fois, aux erreurs de modélisation, ainsi qu'aux stratégies inconnues de guidage.

En mathématiques financières, l'évolution des prix d'actifs dans des marchés boursiers sont aussi modélisés par des chaînes de Markov. Dans ce contexte, les variables U_n représentent les fluctuations, et la volatilité stochastique du marché financier.

Nous reviendrons sur des exemples plus précis dans la suite.

3.2 Chaînes de Markov discrètes

Les chaînes de Markov les plus élémentaires sont bien entendu celles dont les états aléatoires ne prennent qu'un nombre fini, ou au plus dénombrable de valeurs. Cette section est consacrée à l'étude de ces chaînes élémentaires. Nous insisterons sur les réalisations “canoniques”, et dynamiques, de ces processus aléatoires. La seconde partie de cette section concerne l'étude des semigroupes d'évolution des ces chaînes de Markov.

Afin de satisfaire la curiosité du lecteur, et souligner les points d'interaction avec d'autres domaines scientifiques, nous soulignerons les interactions entre la théorie des chaînes de Markov, l'analyse fonctionnelle, l'algèbre matricielle, et la théorie des graphes. Enfin, nous illustrerons ces nouvelles notions sur une variété d'exemples précis liés à des phénomènes aléatoires issus de la physique, de la biologie, et des sciences de l'ingénieur.

Dans la suite de cette section, X_n désigne une chaîne de Markov discrète à valeurs dans un espace E , au plus dénombrable. Ses transitions de probabilités seront données par une **collection** $(M_n(x, \cdot))_{x \in E}$ de mesures de probabilités sur E .

Autrement dit, les probabilités de passage d'un état $X_{n-1} = x$ à un nouvel état aléatoire X_n sont données par l'application suivante

$$y \in E \mapsto M_n(x, y) = \mathbb{P}(X_n = y \mid X_{n-1} = x) \in [0, 1]$$

On utilise parfois la notation

$$\mathbb{P}^{X_n|X_{n-1}}(y|x) = \mathbb{P}(X_n = y \mid X_{n-1} = x)$$

On désigne par la suite $\eta_n = \mathbb{P}^{X_n}$, la loi de l'état X_n de la chaîne, à chacun des instants $n \in \mathbb{N}$.

On remarquera que la loi de la trajectoire (X_0, \dots, X_n) de l'origine jusqu'à l'instant n est alors décrite par la formule multiplicative

$$\begin{aligned} \mathbb{P}^{(X_0, X_1, \dots, X_n)}(x_0, x_1, \dots, x_n) &= \mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \\ &= \eta_0(x_0)M_1(x_0, x_1) \dots M_n(x_{n-1}, x_n) \end{aligned}$$

pour toute trajectoire $(x_p)_{0 \leq p \leq n} \in E^{n+1}$.

On peut clairement étendre la notion de chaîne de Markov précédente à des modèles markoviens X_n prenant leurs valeurs dans des espaces E_n **liés au paramètre temporel!** Dans ce contexte, $M_{n+1}(x_n, x_{n+1})$ désigne la probabilité de passer d'un état $x_n \in E_n$ vers un état $x_{n+1} \in E_{n+1}$. Plus formellement, nous avons à nouveau

$$M_{n+1}(x_n, x_{n+1}) = \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n)$$

et

$$\begin{aligned} \mathbb{P}^{(X_0, X_1, \dots, X_n)}(x_0, x_1, \dots, x_n) &= \mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \\ &= \eta_0(x_0)M_1(x_0, x_1) \dots M_n(x_{n-1}, x_n) \end{aligned}$$

Bien que cette extension ne puisse paraître purement formelle, sans d'autres intérêts que mathématiques, cette notion apparaît naturellement dans la représentation de chaînes trajectorielles.

3.2.1 Semigroupes de transitions

Les probabilités de transitions M_n d'une chaîne de Markov sur un espace E sont associées à deux opérateurs naturels :

1. Le premier agit à droite sur les fonctions bornées sur E . A chacune de ces fonctions f , on associe la fonction bornée $M_n[f]$ définie par la formule suivante

$$M_n[f](x) =_{\text{dét.}} \mathbb{E}(f(X_n) | X_{n-1} = x) = \sum_{y \in E} M_n(x, y) f(y)$$

2. Le second agit à gauche sur les mesures de probabilités sur E . A chacune de ces mesures η , on associe la mesure de probabilité (ηM_n) définie par

$$(\eta M_n)(y) = \sum_{x \in E} \eta(x) M_n(x, y) \in [0, 1]$$

Dans ce contexte, il est aussi très utile de voir une mesure de probabilité η comme un opérateur sur l'ensemble des fonctions f bornées sur l'espace des états E

$$\eta[f] = \sum_{x \in E} \eta(x) f(x) \in \mathbb{R}$$

Avec ce système de notations, pour les fonctions indicatrices $f = 1_B$ d'ensembles $B \subset \mathbb{R}$, on retrouve la mesure $\eta(B)$ des ensembles B

$$\eta[1_B] = \sum_{x \in B} \eta(x) = \eta(B)$$

De même, on observe que

$$M_n[1_B](x) = \mathbb{E}(1_B(X_n) | X_{n-1} = x) = M_n(x, B) = \mathbb{P}(X_n \in B | X_{n-1} = x)$$

Lorsqu'il n'y a pas de confusions, il est coutume de noter $\eta(f)$, et $M_n(f)$ plutôt que $\eta[f]$, et $M_n[f]$. Dans ce système de notations, nous avons

$$\mathbb{E}(f(X_n)) = \sum_{x \in E} f(x) \eta_n(x) = \eta_n(f)$$

et

$$\mathbb{E}(f(X_n) | X_{n-1} = x) = \sum_{y \in E} M_n(x, y) f(y) = M_n(f)(x)$$

En utilisant la formule des conditionnements emboîtés on montre que

$$\begin{aligned}\eta_n(f) &= \mathbb{E}(\mathbb{E}(f(X_n)|X_{n-1})) \\ &= \mathbb{E}(M_n(f)(X_{n-1})) \\ &= \sum_{x \in E} \eta_{n-1}(x) M_n(f)(x) = \eta_{n-1}(M_n(f))\end{aligned}$$

Autrement dit, les lois η_n des différents états de la chaîne de Markov X_n peuvent “se calculer” récursivement. Ces dernières sont solution d’un système dynamique discret à valeurs dans l’espace des probabilités :

$$\eta_n = \eta_{n-1}M_n$$

Notre prochain objectif est de décrire plus précisément le semigroupe d’évolution du flot de mesures $(\eta_n)_n$. On introduit pour cela la probabilité conditionnelle de X_n en X_p

$$M_{p,n}(x, y) = \mathbb{P}^{X_n|X_p}(y|x) = \mathbb{P}(X_n = y | X_p = x)$$

avec $0 \leq p \leq n$, Comme précédemment, on associe à toute fonction f sur E , la fonction $M_{p,n}(f)$ sur E donnée par

$$M_{p,n}(f)(x) = \mathbb{E}(f(X_n)|X_p = x) = \sum_{y \in E} M_{p,n}(x, y) f(y) \quad (3.1)$$

En particulier pour $p = n$, on a $M_{n,n}(f) = f$; et pour $p = (n - 1)$, nous avons

$$M_{n-1,n}(f)(x) = M_n(f)(x) = \mathbb{E}(f(X_n)|X_{n-1} = x)$$

D’après la formule de conditionnements emboîtés, nous avons la formule de récurrence

$$\begin{aligned}M_{p,n}(f)(x) &= \mathbb{E}(\mathbb{E}(f(X_n)|X_{n-1})|X_p = x) \\ &= \mathbb{E}(M_n(f)(X_{n-1})|X_p = x) = M_{p,n-1}(M_n(f))(x)\end{aligned}$$

Par conséquent, l’opérateur de transition $M_{p,n}$ est donné par la formule (3.1) avec

$$\begin{aligned}M_{p,n}(x, y) \\ = \sum_{x_{p+1} \in E, \dots, x_{n-1} \in E} M_{p+1}(x, x_{p+1})M_{p+1}(x_{p+1}, x_{p+2}) \dots M_n(x_{n-1}, y)\end{aligned}$$

On en déduit une formulation des transitions conditionnelles $X_p \rightsquigarrow X_n$, en terme de compositions d’opérateurs

$$M_{p,n}(f) = M_{p+1}M_{p+2} \dots M_n(f)$$

En utilisant la formule

$$\forall p \leq n \quad \mathbb{E}(f(X_n)) = \mathbb{E}(\mathbb{E}(f(X_n)|X_p)) = \mathbb{E}(M_{p,n}(f)(X_p))$$

on en conclut que

$$\begin{aligned} \eta_n(f) &= \eta_p(M_{p,n}f) = \sum_{x \in E} \eta_p(x) \left[\sum_{y \in E} M_{p,n}(x,y) f(y) \right] \\ &= (\eta_p M_{p,n})(f) = \sum_{y \in E} \left(\sum_{x \in E} \eta_p(x) M_{p,n}(x,y) \right) f(y) \end{aligned}$$

L'ordre des sommes étant sans importance, on simplifie les notations, et on écrit tout simplement

$$\eta_n = \eta_p M_{p,n} \quad \text{avec} \quad M_{p,n} = M_{p+1} M_{p+2} \dots M_n \quad (3.2)$$

Nous avons donc montré que les opérateurs $M_{p,n}$, $p \leq n$, correspondent au semigroupe d'évolution du flot de mesures $(\eta_n)_{n \geq 0}$.

Lorsque la chaîne est homogène, ces opérateurs correspondent à la composition du même opérateur de transition M . Dans cette situation, on utilise souvent les notations synthétiques suivantes

$$M_{p,p+n} = M^n \quad \text{et} \quad \eta_n = \eta_0 M^n \quad \text{avec} \quad M^n = M^{n-1} M = M M^{n-1}$$

3.2.2 Processus historique

On considère une chaîne de Markov discrète X'_n à valeurs dans un espace au plus dénombrable E' . On note $M'_n(x', y')$, la probabilité de passage de $X_{n-1} = x'$ vers $X_n = y'$. La séquence de trajectoires aléatoires

$$X_n = (X'_0, \dots, X'_n) \in E_n =_{\text{déf.}} (E')^{n+1}$$

forme à nouveau une chaîne de Markov à valeurs dans les espaces trajectoriels E_n . En effet, le passage de X_n à X_{n+1} s'effectue en deux temps. On conserve tout d'abord le segment de trajectoire $X_n = (X'_0, \dots, X'_n)$, puis on lui adjoint une extension élémentaire $X'_{n+1} = x'$ de loi $M'_{n+1}(X'_n, x')$. Autrement dit, nous avons

$$\begin{aligned} X_n = (X'_0, \dots, X'_n) &\rightsquigarrow X_{n+1} = \left(\underbrace{(X'_0, \dots, X'_n)}_{X_n}, X'_{n+1} \right) \\ &= (X_n, X'_{n+1}) \in E_{n+1} = (E_n \times E') \end{aligned}$$

On notera que pour toute fonction f_{n+1} bornée sur E_{n+1} , et pour tout segment de trajectoire $x_n = (x'_0, \dots, x'_n) \in E_n$, nous avons

$$\begin{aligned} & \mathbb{E}(f_{n+1}(X_{n+1}) \mid X_n = x_n) \\ &= \mathbb{E}(f_{n+1}([X'_0, \dots, X_n], X'_{n+1}) \mid (X'_0, \dots, X'_n) = (x'_0, \dots, x'_n)) \\ &= \sum_{x'_{n+1} \in E'} f_{n+1}([x'_0, \dots, x'_n], x'_{n+1}) M'_{n+1}(x'_n, x'_{n+1}) \end{aligned}$$

Cette équation s'exprime aussi sous la forme trajectorielle suivante

$$\begin{aligned} & \mathbb{E}(f_{n+1}(X_{n+1}) \mid X_n = x_n) \\ &= \sum_{(y'_0, \dots, y'_{n+1}) \in E_{n+1}} f_{n+1}(y'_0, \dots, y'_{n+1}) \mathbf{1}_{(x'_0, \dots, x'_n)}(y'_0, \dots, y'_n) M'_{n+1}(y'_n, y'_{n+1}) \end{aligned}$$

La dernière formule montre que X_n est une chaîne de Markov, de probabilités de transitions M_{n+1} de E_n vers E_{n+1} , données par

$$M_{n+1}((x'_0, \dots, x'_n), (y'_0, \dots, y'_{n+1})) = \mathbf{1}_{(x'_0, \dots, x'_n)}(y'_0, \dots, y'_n) M'_{n+1}(y'_n, y'_{n+1})$$

3.2.3 Interprétation matricielle

Lorsque l'espace d'état E est, soit fini, soit indexé de façon naturelle par \mathbb{N} , ou encore par \mathbb{Z} , les semigroupes de transitions $M_{p,n}$ définis dans la section 3.2.1 correspondent à des compositions "élémentaires" de matrices. Pour préciser cette assertion, commençons par l'exemple le plus simple, où l'espace d'état est donné par un ensemble à deux points $E = \{1, 2\}$. Dans ce cas, la donnée des transitions de la chaîne

$$M_n(x, y) = \mathbb{P}(X_n = y \mid X_{n-1} = x)$$

est équivalente à la donnée des matrices (2×2) suivantes

$$M_n = (M_n(x, y))_{x, y \in E} = \begin{pmatrix} M_n(1, 1) & M_n(1, 2) \\ M_n(2, 1) & M_n(2, 2) \end{pmatrix}$$

Par exemple, la matrice

$$M = \begin{pmatrix} 6/7 & 1/7 \\ 1/2 & 1/2 \end{pmatrix}$$

correspond au mouvement aléatoire entre deux états $E = \{1, 2\}$, synthétisé par le schéma ci-dessous

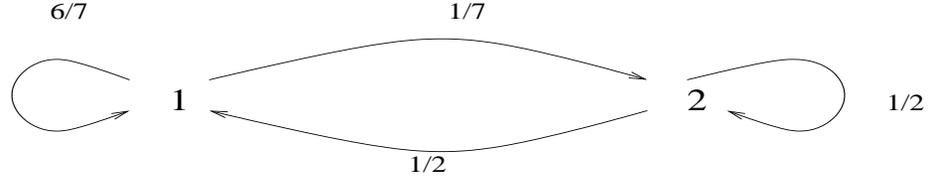


FIG. 3.1 – Chaîne à deux états

Par définition de l'opérateur de transition M_n , nous avons, pour toute fonction f sur $E = \{1, 2\}$, la formule matricielle

$$\begin{aligned}
 \begin{bmatrix} M_n(f)(1) \\ M_n(f)(2) \end{bmatrix} &= \begin{bmatrix} M_n(1,1) f(1) + M_n(1,2) f(2) \\ M_n(2,1) f(1) + M_n(2,2) f(2) \end{bmatrix} \\
 &= \begin{pmatrix} M_n(1,1) & M_n(1,2) \\ M_n(2,1) & M_n(2,2) \end{pmatrix} \begin{bmatrix} f(1) \\ f(2) \end{bmatrix}
 \end{aligned} \tag{3.3}$$

D'autre part, en utilisant la formule des conditionnements emboîtés

$$\mathbb{E}(f(X_n)) = \mathbb{E}(\mathbb{E}(f(X_n)|X_{n-1})) = \mathbb{E}(M_n(f)(X_{n-1}))$$

on obtient une nouvelle formule matricielle

$$\begin{aligned}
 [\eta_n(1), \eta_n(2)] &= \begin{bmatrix} \eta_{n-1}(1) M_n(1,1) & \eta_{n-1}(2) M_n(2,1) \\ \eta_{n-1}(1) M_n(1,2) & \eta_{n-1}(2) M_n(2,2) \end{bmatrix} \\
 &= [\eta_{n-1}(1), \eta_{n-1}(2)] \begin{pmatrix} M_n(1,1) & M_n(1,2) \\ M_n(2,1) & M_n(2,2) \end{pmatrix}
 \end{aligned} \tag{3.4}$$

Par conséquent, si l'on représente une fonction numérique f , et une mesure de probabilité η sur E par les vecteurs colonnes et lignes

$$f = \begin{bmatrix} f(1) \\ f(2) \end{bmatrix} \quad \text{et} \quad \eta = [\eta(1), \eta(2)]$$

alors, les équations (3.3) et (3.4) s'expriment sous la forme d'un **semigroupe matriciel**

$$M(f) = M f \quad \text{et} \quad \eta_n = \eta_{n-1} M_n = \eta_0 M_1 \dots M_n$$

Ces interprétations matricielles restent valables pour des chaînes de Markov à valeurs dans des espaces d'états finis abstraits $E = \{x_1, \dots, x_d\}$. Dans ce contexte, les matrices de transitions sont données par

$$M_n = \begin{pmatrix} M_n(x_1, x_1) & \dots & M_n(x_1, x_d) \\ \vdots & \ddots & \vdots \\ M_n(x_d, x_1) & \dots & M_n(x_d, x_d) \end{pmatrix}$$

Les mesures de probabilités η , et les fonctions f sur E sont associées au vecteurs lignes et colonnes suivants

$$\eta = [\eta(x_1), \dots, \eta(x_d)] \quad \text{et} \quad f = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_d) \end{bmatrix}$$

Exemple 3.2.1 *La figure suivante*

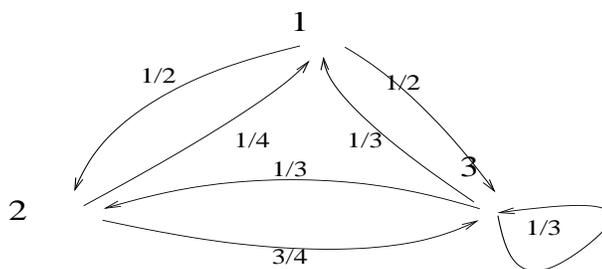


FIG. 3.2 – Chaîne à 3 états

présente un schéma d'évolution de chaîne de Markov sur un espace à trois points $E = \{1, 2, 3\}$, et de matrice de transition homogène donnée par

$$M = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/4 & 0 & 3/4 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$$

Lorsque l'espace d'états est dénombrable, et indexé par les entiers positifs $E = \{x_i, i \in \mathbb{N}\}$, les matrices de transition sont infini-dimensionnelles

$$M_n = (M_n(x, y))_{x, y \in E} = \begin{pmatrix} M_n(x_0, x_0) & M(x_0, x_1) & M(x_0, x_2) & \dots \\ M_n(x_1, x_0) & M(x_1, x_1) & M(x_1, x_2) & \dots \\ M_n(x_2, x_0) & M(x_2, x_1) & M(x_2, x_2) & \dots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Dans ce cas, les probabilités η , et les fonctions f sur E sont associés au vecteurs

$$\eta = [\eta(x_0), \eta(x_1), \eta(x_2), \dots] \quad \text{et} \quad f = \begin{bmatrix} f(x_0) \\ f(x_1) \\ f(x_2) \\ \vdots \end{bmatrix}$$

Exemple 3.2.2 Une marche aléatoire homogène sur \mathbb{N} , absorbée en 0, peut être représentée par le schéma suivant

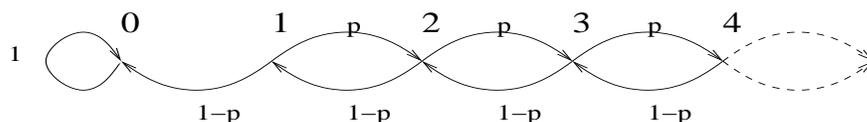


FIG. 3.3 – Chaîne absorbée en 0

La matrice de transition associée à cette chaîne est donnée par

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots \\ (1-p) & 0 & p & 0 & \dots \\ 0 & (1-p) & 0 & p & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Lorsque le point 0 est réfléchissant, la matrice de transition s'écrit sous la forme

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots \\ (1-p) & 0 & p & 0 & \dots \\ 0 & (1-p) & 0 & p & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Cette situation correspond au schéma suivant

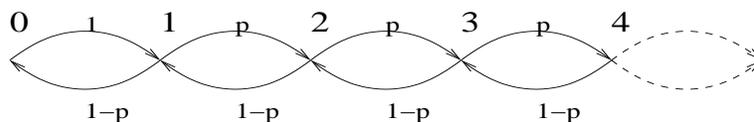


FIG. 3.4 – Chaîne réfléchie en 0

Dans le dernier cas, où l'espace d'états est dénombrable, et indexé par les entiers relatifs $E = \{x_i, i \in \mathbb{Z}\}$, les matrices de transition sont données par

$$M_n = \begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & M(x_{-1}, x_{-1}) & M_n(x_{-1}, x_0) & M(x_{-1}, x_1) & \dots \\ \dots & M(x_0, x_{-1}) & M_n(x_0, x_0) & M(x_0, x_1) & \dots \\ \dots & M(x_1, x_{-1}) & M_n(x_1, x_0) & M(x_1, x_1) & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

et les probabilités η , et les fonctions f sur E sont associés au vecteurs

$$\eta = [\dots, \eta(x_{-1}), \eta(x_0), \eta(x_1), \dots] \quad \text{et} \quad f = \begin{bmatrix} \vdots \\ f(x_{-1}) \\ f(x_0) \\ f(x_1) \\ \vdots \end{bmatrix}$$

Exemple 3.2.3 Une marche aléatoire simple sur \mathbb{Z} peut être représenté par le schéma suivant cette situation correspond au schéma suivant

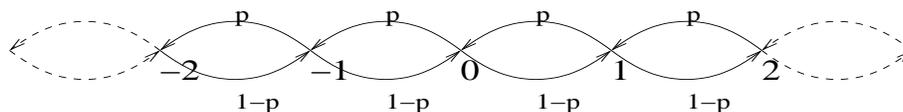


FIG. 3.5 – Marche simple sur \mathbb{Z}

La matrice de transition de cette évolution est donnée par

$$M = \begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & 0 & (1-p) & 0 & \dots \\ \dots & p & 0 & (1-p) & \dots \\ \dots & & p & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

3.3 Quelques Exemples

3.3.1 Files d'attentes

Soit $(U_n)_{n \geq 1}$ une suite de v.a. positives et indépendantes de lois respectives $(\mu_n)_{n \geq 1}$. On considère la chaîne de Markov définie de façon récursive par l'équation suivante

$$\begin{cases} X_{n+1} = (X_n - 1)^+ + U_{n+1} \\ X_0 = 0 \end{cases}$$

Dans la formule précédente $a^+ = \max(a, 0)$ désigne le maximum entre un nombre réel $a \in \mathbb{R}$ et 0.

Ce processus aléatoire peut s'interpréter comme la longueur d'**une file d'attente**, ou encore **le temps d'attente d'un client** arrivant à un guichet, servant une personne par unité de temps. Dans ce contexte, la v.a. U_{n+1} représente le nombre de clients arrivant dans la file d'attente au temps $(n+1)$.

On peut aussi interpréter X_n comme le nombre de paquets (symboles binaires représentant de l'information : voix, vidéo, données,...) en attente dans **la mémoire d'un canal de communication**, transmettant un paquet par unité de temps. Dans cette situation, la v.a. U_{n+1} représente le nombre de paquets arrivant dans le canal à l'instant $(n+1)$.

On notera que les transitions de cette chaîne sont données pour tout $i \geq 1$, et pour tout $j \geq 0$, par la formule suivante

$$\mathbb{P}(X_{n+1} = (i-1) + j \mid X_n = i) = \mu_{n+1}(j) = \mathbb{P}(X_{n+1} = j \mid X_n = 0)$$

3.3.2 Modèle d'urne

On considère une urne contenant initialement B_0 boules blanches, et N_0 boules noires. A chaque instant n , on choisit au hasard une boule, puis on remet cette boule dans l'urne accompagnée d'une nouvelle boule de la même couleur. On note (B_n, N_n) le nombre de boules blanches, et noires, dans l'urne au temps n . Par construction, le couple $X_n = (B_n, N_n)$ est une chaîne de Markov à valeurs dans $E = \mathbb{N}^2$, et de probabilités de transitions

$$\mathbb{P}^{X_n | X_{n-1}}(\cdot | (b, m)) = \frac{b}{m+b} \delta_{(b+1, m)} + \frac{m}{b+m} \delta_{(b, m+1)}$$

3.3.3 Marche aléatoire sur \mathbb{Z}

La marche aléatoire simple sur \mathbb{Z} correspond à un mouvement aléatoire d'une particule sur les entiers relatifs, se déplaçant soit d'un pas vers la droite, avec une probabilité p , soit d'un pas vers la gauche, avec une probabilité $(1-p)$, avec $p \in (0, 1)$. Ce mouvement peut être représenté schématiquement par la figure suivante :

Plus formellement, ce mouvement aléatoire est défini par une chaîne de Markov $X = (X_n)_{n \geq 0}$ définie sur un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$, d'origine $X_0 = 0$ et de probabilités de transitions homogènes

$$\begin{aligned} M(x, y) &=_{\text{déf.}} \mathbb{P}(X_n = y \mid X_{n-1} = x) = \mathbb{P}^{X_n | X_{n-1}}(y|x) \\ &= p \mathbf{1}_{x+1}(y) + (1-p) \mathbf{1}_{x-1}(y) \end{aligned}$$

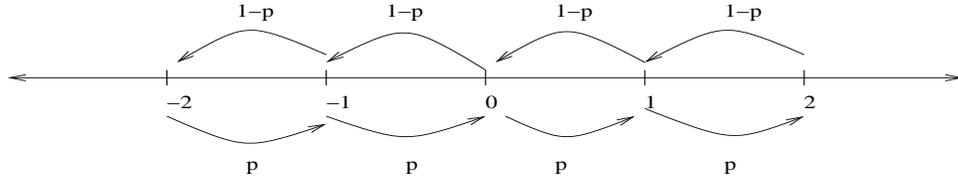


FIG. 3.6 – Marche aléatoire simple sur \mathbb{Z}

Pour décrire une réalisation dynamique de cette chaîne, on se donne une suite de v.a. indépendantes $U = (U_n)_{n \geq 1}$ distribuées sur $\{-1, +1\}$ selon la même loi de Bernoulli

$$\mathbb{P}(U_n = +1) = 1 - \mathbb{P}(U_n = -1) = p$$

On suppose, comme d’habitude, que cette suite est définie sur un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$. On associe à U , le système dynamique aléatoire donné par

$$\begin{cases} X_n = X_{n-1} + U_n \\ X_0 = 0 \end{cases} \quad (3.5)$$

Cette interprétation dynamique offre un certain nombre d’avantages. Par exemple, elle permet une représentation “explicite” de l’état X_n en terme de v.a. indépendantes, et simplifie l’analyse des transitions de probabilités.

Exercice 3.3.1 *Cet exercice a pour objectif d’analyser plus en profondeur la marche aléatoire sur \mathbb{Z} décrite dans l’exemple 3.3.3.*

1. Vérifier que l’interprétation dynamique introduite en (3.5) correspond bien à la donnée d’une marche aléatoire simple sur \mathbb{Z} .
2. Montrer que la position moyenne de la particule au temps n est donnée par la formule $\mathbb{E}(X_n) = n \times (2p - 1)$. En conclure que

$$\begin{cases} \lim_{n \rightarrow \infty} \mathbb{E}(X_n) = -\infty & \text{si } p \in [0, 1/2) \\ \mathbb{E}(X_n) = 0 & \text{si } p = 1/2 \\ \lim_{n \rightarrow \infty} \mathbb{E}(X_n) = +\infty & \text{si } p \in (1/2, 1] \end{cases}$$

3. Vérifier que les transitions de la chaîne entre deux instants, m et $(m + n)$, sont données par la formule

$$\mathbb{P}(X_{m+n} = x + [k - (n - k)] \mid X_m = x) = C_n^k p^k (1 - p)^{n-k}$$

pour tous les $k \in \{0, \dots, n\}$, et

$$\mathbb{P}(X_{m+n} \notin \{2k - n : k = 0, \dots, n\} \mid X_n = x) = 0$$

4. En déduire que

$$\mathbb{P}(X_{m+2k} = 0 \mid X_m = 0) = \frac{(2k)!}{k!k!} (p(1-p))^k$$

En utilisant la formule de Stirling ($k! \simeq \sqrt{2\pi k} k^k e^{-k}$), montrer que

$$\mathbb{P}(X_{m+2k} = 0 \mid X_m = 0) \simeq \frac{(4p(1-p))^k}{\sqrt{\pi k}} \quad (= 1/\sqrt{\pi k} \text{ si } p = 1/2)$$

3.3.4 Marche aléatoire sur \mathbb{Z}^d

On note $|\cdot|$ la distance l^1 sur $E = \mathbb{Z}^d$ définies par

$$|x| = \sum_{i=1}^d |x^i|$$

pour tout $x = (x^i)_{1 \leq i \leq d} \in \mathbb{Z}^d$. On associe à une mesure de probabilité p sur l'ensemble des $2d$ vecteurs unitaires directionnels

$$\mathcal{U} = \{u \in \mathbb{Z}^d : |u| = 1\}$$

la transition homogène

$$M(x, y) = \sum_{u \in \mathcal{U}} p(u) 1_{x+u}(y)$$

L'évolution aléatoire de la chaîne X_n associée à M est claire. A chaque étape n , la particule choisit aléatoirement un vecteur $u \in \mathcal{U}$ avec la probabilité $p(u)$, et se déplace dans cette direction. Autrement dit, si $(U_n)_{n \geq 1}$ désigne une suite de v.a. indépendantes de même loi p sur \mathcal{U} , on a une représentation dynamique de l'évolution

$$X_n = X_{n-1} + U_n = X_0 + \sum_{i=1}^n U_i$$

La figure suivante présente une réalisation d'une trajectoire aléatoire de la chaîne X_n sur \mathbb{Z}^2 , d'origine $X_0 = 0 \in \mathbb{Z}^2$.

3.3.5 Marche aléatoire arrêtée

Supposons qu'une particule évolue sur \mathbb{Z}^d selon les principes de transitions élémentaires décrits dans l'exemple 3.3.4, mais cette dernière ne peut se mouvoir que s'il elle se trouve dans une région spécifique de l'espace $B \subset \mathbb{Z}^d$. Autrement dit, lorsque la chaîne X_n sort de B , elle s'immobilise. Ce modèle physique peu à nouveau être

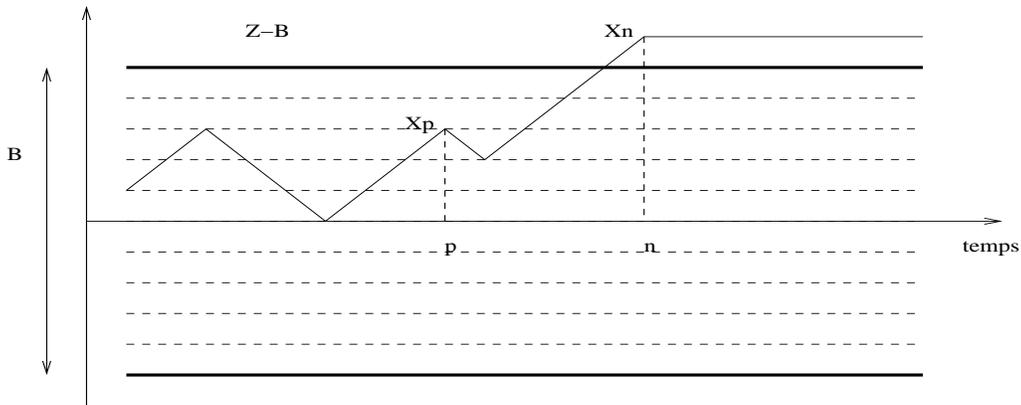


FIG. 3.8 – Chaîne stoppée

Lorsque les v.a. de branchement $(N_n^i)_{i \geq 1, n \geq 1}$ sont des copies indépendantes d'une même v.a. entière N , on notera que

$$\mathbb{E}(X_{n+1}|X_n) = \sum_{i=1}^{X_n} \mathbb{E}(N_n^i) = X_n \times \mathbb{E}(N) \Rightarrow \mathbb{E}(X_{n+1}) = \mathbb{E}(X_0) \mathbb{E}(N)^{n+1}$$

Par conséquent, en supposant que $\mathbb{E}(X_0) \neq 0$, la population moyenne s'éteindra lorsque $\mathbb{E}(N) < 1$, et elle explosera lorsque $\mathbb{E}(N) > 1$. La figure ci-dessous présente une réalisation d'un arbre de descendance d'un individu ($X_0 = 1$).

Examinons la situation où chaque individu se dédouble avec une probabilité p , ou disparaît avec la probabilité $(1 - p)$. Ce modèle correspond au choix d'une v.a. N de Bernoulli loi

$$\mathbb{P}^N = p \delta_2 + (1 - p) \delta_0$$

Dans ce contexte, on notera que $\mathbb{E}(N) = 2p$. De plus, on montre les transitions de cette chaîne sont données par la formule binomiale

$$\mathbb{P}(X_{n+1} = 2y \mid X_n = x) = C_x^y p^y (1 - p)^{x-y}$$

pour tout $y \in \{0, \dots, x\}$, et $\mathbb{P}(X_{n+1} \in 2\mathbb{N} + 1 \mid X_n = x) = 0$, pour tout $x \in \mathbb{N}$.

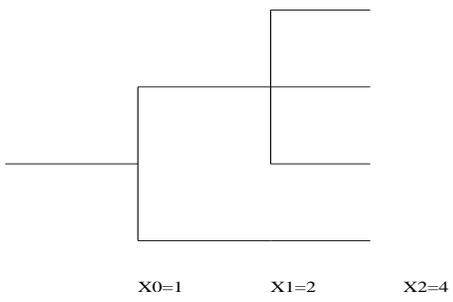


FIG. 3.9 – Processus de branchement

Chapitre 4

Chaînes de Markov abstraites

4.1 Description des modèles

Comme nous l'avons souligné dans l'introduction, la notion de chaîne de Markov, est loin d'être restreinte à des phénomènes aléatoires à valeurs dans des espaces discrets. Les fluctuations de température d'un liquide par unité de temps correspondent à des chaînes de Markov à valeurs réelles. L'évolution d'une cible spatiale peut être modélisée par une chaîne de Markov à valeurs dans l'espace euclidien \mathbb{R}^3 , ou bien dans \mathbb{R}^9 si l'on tient compte des coordonnées de position, vitesse et accélération. Enfin, si l'on tient compte des évolutions de cette même cible entre certains paliers, on obtient une chaîne de Markov à valeurs dans des espaces de chemins ou d'excursions.

Tous ces modèles physiques, s'inscrivent dans la théorie abstraite des chaînes de Markov à valeurs dans des espaces mesurables. Leur analyse s'exprime de façon naturelle dans la théorie de l'intégration de Lebesgue. Afin d'éviter un langage trop technique, j'ai volontairement choisi de ne pas inscrire cette introduction à l'ingénierie stochastique dans ce cadre trop mathématique. Nous ignorerons donc les notions d'ensembles et de fonctions mesurables, les théorèmes de Fubini, et d'autres propriétés fondamentales de la théorie de l'intégration de Lebesgue. La nature est suffisamment stable pour polir les erreurs et les confusions dues à de tels écarts. J'ai plutôt essayé de coller au plus près à la réalité scientifique et technique, tout en aiguisant la curiosité mathématique du lecteur.

Dans la suite, on utilisera donc la terminologie "chaîne de Markov abstraite" pour désigner une chaîne de Markov X_n à valeurs dans des ensembles suffisamment réguliers E_n , de transitions de probabilités

$$M_n(x_{n-1}, dx_n) =_{\text{déf.}} \mathbb{P}^{X_n|X_{n-1}}(dx_n|x_{n-1}) = \mathbb{P}(X_n \in dx_n \mid X_{n-1} = x_{n-1})$$

et de loi initiale $\eta_0(dx_0) = \mathbb{P}^{X_0}(dx_0)$ sur E_0 . On note alternativement

$$\eta_n(dx_n) = \mathbb{P}^{X_n}(dx_n) = \mathbb{P}(X_n \in dx_n)$$

la loi de la v.a. X_n sur E_n , donnée pour toute fonction f_n bornée sur E_n par la formule

$$\mathbb{E}(f_n(X_n)) = \int f_n(x_n) \eta_n(dx_n) \stackrel{\text{d\'ef.}}{=} \eta_n(f_n)$$

Dans ce système de notations abusives, une mesure donnée sur E_n s'interprète, soit comme un opérateur intégral, soit comme une fonction ensembliste. Ainsi, pour tout sous ensemble suffisamment régulier $B_n \subset E_n$, nous avons les représentations équivalentes

$$\begin{aligned} \eta_n(B_n) &= \mathbb{P}^{X_n}(B_n) = \eta_n[1_{B_n}] \\ M_n(x_{n-1}, B_n) &= \mathbb{P}^{X_n|X_{n-1}}(B_n|x_{n-1}) \\ &= \mathbb{P}(X_n \in B_n \mid X_{n-1} = x_{n-1}) = M_n[1_{B_n}](x_{n-1}) \end{aligned}$$

On dit qu'une chaîne de Markov **réelle** $X = (X_n)_{n \geq 0}$ est **absolument continue**, lorsque la v.a. initiale X_0 est absolument continue de densité p_0 sur \mathbb{R} , et lorsque ses transitions de probabilités sont données en terme d'une famille de densités de probabilités $\{p_n(x, \cdot), x \in \mathbb{R}\}$ sur \mathbb{R}

$$M_n(x_{n-1}, dx_n) = \mathbb{P}^{X_n|X_{n-1}}(dx_n|x_{n-1}) = p_n(x_{n-1}, x_n) dx_n \quad (4.1)$$

Dans la formule précédente dx_n désigne la mesure de Lebesgue sur \mathbb{R} . A titre illustratif, on pourra considérer les familles de densités gaussiennes, ou exponentielles données par la formule

$$p_n(x, y) = \frac{e^{-\frac{1}{2}(y-x)^2}}{\sqrt{2\pi}} \quad \text{ou} \quad p_n(x, y) = 1_{[0, \infty)}(y) (|x| + 1) e^{-(|x|+1)y}$$

Dans ce contexte, la trajectoire (X_0, \dots, X_n) de l'origine jusqu'à l'instant n est une v.a. absolument continue, de loi donnée par la formule

$$\mathbb{P}^{(X_0, \dots, X_n)}(d(x_0, \dots, x_n)) = p_0(x_0)p_1(x_0, x_1) \dots p_n(x_{n-1}, x_n) dx_0 dx_1 \dots dx_n$$

4.1.1 Semigroupe des transitions

Comme dans le cas des chaînes discrètes, il existe un système de notations naturel permettant de décrire les semigroupes d'évolution des lois des états X_n .

Les probabilités de transitions $M_n(x_{n-1}, dx_n)$ permettent de définir deux opérateurs intégraux naturels :

1. Le premier agit à droite sur les fonctions bornées. A chacune de ces fonctions f , on associe la fonction bornée $M_n[f]$ définie par la formule suivante

$$M_n[f] : x \in \mathbb{R} \mapsto M_n[f](x) =_{\text{déf.}} \mathbb{E}(f(X_n) | X_{n-1} = x) \in \mathbb{R}$$

2. Le second agit à gauche sur les mesures de probabilités sur \mathbb{R} . A chacune de ces mesures η , on associe la mesure de probabilité (ηM_n) définie par

$$(\eta M_n) : B \subset \mathbb{R} \mapsto (\eta M_n)(B) = \int \eta(dx) M_n(x, B) \in [0, 1]$$

Dans ce contexte, une mesure de probabilité η correspond à un opérateur intégral sur l'ensemble des fonctions f mesurables et bornées

$$\eta[f] = \int \eta(dx) f(x) \in \mathbb{R}$$

Avec ce système de notations, pour les fonctions indicatrices $f = 1_B$ d'ensembles $B \subset \mathbb{R}$, nous avons $\eta(B) = \eta[1_B]$. Lorsqu'il n'y a pas de confusions, il est coutume de noter $\eta(f)$ et $M_n(f)$ plutôt que $\eta[f]$ et $M_n[f]$. Cet abus de notation évident est parfois poussé à l'extrême. Certains auteurs notent parfois tout simplement ηf , pour insister sur le fait que cette opération intégrale n'est autre qu'une extension naturelle du produit matriciel.

À la différence des chaînes discrètes, les lois $\eta_n(dx)$ des états X_n de chaînes abstraites sont données par des équations d'évolution intégrales. Il est bien entendu hors de question d'être tenté de calculer, ou d'estimer ces formules de transport. Il convient néanmoins de souligner qu'il existe des stratégies numériques et probabilistes pour le faire! Ces techniques sont connues sous le nom de *méthodes de Monte-Carlo*, en référence au fait qu'elles sont basées sur des simulations concrètes de trajectoires aléatoires.

En utilisant la propriété de Markov, on notera que pour toute fonction f_{n+1} , mesurable et bornée sur E_{n+1} , on a

$$\begin{aligned}
\mathbb{E}(f_{n+1}(X_{n+1})|X_{n-1}) &= \mathbb{E}(\mathbb{E}(f_{n+1}(X_{n+1})|X_{n-1}, X_n) |X_{n-1}) \\
&= \mathbb{E}(\mathbb{E}(f_{n+1}(X_{n+1})|X_n) |X_{n-1}) \\
&= \mathbb{E}(M_{n+1}(f_{n+1})(X_n) |X_{n-1}) \\
&= M_n[M_{n+1}(f_{n+1})](X_{n-1})
\end{aligned}$$

On note $(M_n M_{n+1})(x_{n-1}, dx_{n+1})$ la collection de mesures sur E_{n+1} , indexée par les $x_{n-1} \in E_{n-1}$ et définies pour tout $B_{n+1} \subset E_{n+1}$ par la formule de composition intégrale

$$(M_n M_{n+1})(x_{n-1}, B_{n+1}) = \int_{E_n} M_n(x_{n-1}, dx_n) M_{n+1}(x_n, B_{n+1})$$

En terme d'indicatrices cette équation s'exprime sous la forme suivante

$$(M_n M_{n+1})(1_{B_{n+1}})(x_{n-1}) = \int_{E_n} M_n(x_{n-1}, dx_n) M_{n+1}(1_{B_{n+1}})(x_n)$$

Plus généralement, on a la formule de conditionnement

$$\mathbb{E}(f_{n+p}(X_{n+p})|X_{n-1}) = M_n M_{n+1} \dots M_{n+p}(f_{n+p})(X_{n-1})$$

pour toute fonction f_{n+p} bornée sur E_{n+p} , et pour tout décalage d'indice temporel $p \geq 1$. Les opérateurs intégraux

$$M_{n,n+p} \stackrel{\text{aéf.}}{=} M_n M_{n+1} \dots M_{n+p}$$

avec $n \geq 1$ et $p \geq 1$, forment *un semigroupe d'opérateurs intégraux*, en ce sens où

$$\forall n \leq m \leq n+p \quad M_{n,n+p} = M_{n,m} M_{m,n+p}$$

4.1.2 Équations de Chapman-Kolmogorov

D'après la formule des conditionnements emboîtés, nous avons

$$\begin{aligned}
\eta_n(f_n) &= \mathbb{E}(\mathbb{E}(f_n(X_n)|X_{n-1})) = \mathbb{E}(M_n(f_n)(X_{n-1})) \\
&= \int_{E_{n-1}} \eta_{n-1}(dx_{n-1}) M_n(f_n)(x_{n-1}) = \eta_{n-1}(M_n(f_n)) \quad (4.2)
\end{aligned}$$

Soit $(\eta_{n-1} M_n)$ la mesure sur E_n définie, pour tout $B_n \subset E_n$, par la formule

$$(\eta_{n-1} M_n)(1_{B_n}) = \int_{E_{n-1}} \eta_{n-1}(dx_{n-1}) M_n(1_{B_n})(x_{n-1})$$

Par construction, nous avons les représentations équivalentes suivantes

$$\begin{aligned}
\eta_n(f_n) &= \eta_{n-1}(M_n(f_n)) \\
&= \int_{E_{n-1}} \eta_{n-1}(dx_{n-1}) \left[\int_{E_n} M_n(x_{n-1}, dx_n) f_n(x_n) \right] \\
&= \int_{E_n} \left[\int_{E_{n-1}} \eta_{n-1}(dx_{n-1}) M_n(x_{n-1}, dx_n) \right] f_n(x_n) \\
&= (\eta_{n-1}M_n)(f_n) \stackrel{\text{déf.}}{=} \eta_{n-1}M_n(f_n)
\end{aligned}$$

En utilisant (4.2), nous obtenons la formule de transport intégral des lois des états de la chaîne

$$\eta_n = \eta_{n-1}M_n = \eta_0M_1M_2 \dots M_n \quad (4.3)$$

Cette équation intégrale, appelée **la formule de Chapman-Kolmogorov**, permet de voir les lois η_n comme solution d'un système dynamique (déterministe) intégral (et donc linéaire) sur les espaces de mesures de probabilités.

4.1.3 Processus historique

Comme nous l'avons vu pour les chaînes discrètes dans la section 3.2.2, le cadre non homogène est utile pour représenter des modèles trajectoriels, tel le processus historique associé à une chaîne de Markov élémentaire. La construction abstraite de ces modèles trajectoriels est analogue à celle présentée à l'exemple 3.2.2. Ainsi, si X'_n est une chaîne de Markov de transitions M'_n sur des ensembles E'_n , les séquences de trajectoires

$$X_n = (X'_0, \dots, X'_n)$$

forment un processus de Markov sur les espaces produits

$$E_n = (E'_0 \times \dots \times E'_n)$$

Le passage de X_n à X_{n+1} s'effectue en deux temps. On conserve tout d'abord le segment de trajectoire $X_n = (X'_0, \dots, X'_n)$, puis on lui adjoint une extension élémentaire aléatoire X'_{n+1} de loi $M'_{n+1}(X'_n, dx')$. Plus formellement, nous avons

$$\begin{aligned}
X_n = (X'_0, \dots, X'_n) &\rightsquigarrow X_{n+1} = \underbrace{((X'_0, \dots, X'_n))}_{X_n}, X'_{n+1} \\
&= (X_n, X'_{n+1}) \in E_{n+1} = (E_n \times E')
\end{aligned}$$

On notera que pour toute fonction f_{n+1} bornée sur E_{n+1} , et pour tout segment de trajectoire $x_n = (x'_0, \dots, x'_n) \in E_n$, nous avons

$$\begin{aligned} & \mathbb{E}(f_{n+1}(X_{n+1}) \mid X_n = x_n) \\ &= \mathbb{E}(f_{n+1}([X'_0, \dots, X_n], X'_{n+1}) \mid (X'_0, \dots, X'_n) = (x'_0, \dots, x'_n)) \\ &= \int_{x'_{n+1} \in E'} f_{n+1}([x'_0, \dots, x'_n], x'_{n+1}) M'_{n+1}(x'_n, dx'_{n+1}) \end{aligned}$$

Cette équation s'exprime aussi sous la forme trajectorielle suivante

$$\begin{aligned} & \mathbb{E}(f_{n+1}(X_{n+1}) \mid X_n = x_n) \\ &= \int_{(y'_0, \dots, y'_{n+1}) \in E_{n+1}} f_{n+1}(y'_0, \dots, y'_{n+1}) \delta_{(x'_0, \dots, x'_n)}(d(y'_0, \dots, y'_n)) M'_{n+1}(y'_n, dy'_{n+1}) \end{aligned}$$

La dernière formule montre que X_n est une chaîne de Markov, de probabilités de transitions M_{n+1} de E_n vers E_{n+1} , données par la formule

$$\begin{aligned} & M_{n+1}((x'_0, \dots, x'_n), d(y'_0, \dots, y'_{n+1})) \\ &= \delta_{(x'_0, \dots, x'_n)}(d(y'_0, \dots, y'_n)) M'_{n+1}(y'_n, dy'_{n+1}) \end{aligned}$$

Ces processus historiques interviennent de façon naturelle dans divers problèmes issus de la physique, ou de la biologie. Ils offrent un cadre markovien naturel pour modéliser et analyser des évolutions aléatoires complexes, liées le plus souvent à des effets de dépendance trajectorielles. Ainsi, dans la section 7.3, ces processus historiques nous permettront de définir des modèles d'arbres généalogiques en terme d'algorithmes génétiques trajectoriels. Dans la section 7.4, nous utiliserons à nouveau ces modèles pour représenter des explorations évolutives basées sur des mécanismes de mémoire renforçant les probabilités de retours vers des sites qui ont déjà été visités.

4.2 Chaînes linéaires et gaussiennes

La distribution gaussienne joue un rôle essentiel dans la théorie des probabilités. L'importance des variables gaussiennes est en grande partie due au théorème central de la limite. Ce dernier nous informe que toute accumulation de petites fluctuations indépendantes, et de nature quelconque, se traduit asymptotiquement et irrémédiablement par une variable gaussienne. Tout phénomène résultant d'une addition d'effets aléatoires indépendants est donc nécessairement de nature gaussienne. Dans de nombreux problèmes pratiques, il est donc naturel de considérer comme gaussiennes les erreurs de mesures, et les erreurs de modélisation.

4.2.1 Formulation canonique

Les systèmes linéaires et gaussiens $X = (X_n)_{n \geq 0}$ sont définis par la donnée d'une suite de v.a. à valeurs réelles de lois marginales

$$\mathbb{P}^{(X_0, \dots, X_n)}(d(x_0, \dots, x_n)) = \left[\prod_{p=1}^n \frac{e^{-\frac{(x_p - a_p x_{p-1})^2}{2\sigma_p^2}}}{\sqrt{2\pi\sigma_p^2}} dx_p \right] \frac{e^{-\frac{x_0^2}{2\sigma_0^2}}}{\sqrt{2\pi\sigma_0^2}} dx_0 \quad (4.4)$$

Dans le formule ci-dessus $(a_n)_{n \geq 1}$ désigne une suite de nombres réels, et $(\sigma_n)_{n \geq 0}$ une suite de nombres strictement positifs.

En intégrant la formule (4.4) en la coordonnée x_n , on montre facilement que

$$\begin{aligned} & \mathbb{P}^{X_n | (X_0, \dots, X_{n-1})}(dx_n | x_0, \dots, x_{n-1}) \\ &= \mathbb{P}^{X_n | X_{n-1}}(dx_n | x_{n-1}) \\ &= M_n(x_{n-1}, dx_n) =_{\text{déf.}} \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left\{-\frac{1}{2\sigma_n^2} (x_n - a_n x_{n-1})^2\right\} dx_n \end{aligned}$$

pour tout $n \geq 1$, et $x_0, \dots, x_{n-1} \in \mathbb{R}$. De même, on montre que la loi de la condition initiale X_0 est donnée par

$$\mathbb{P}^{X_0}(dx_0) = \eta_0(dx_0) =_{\text{déf.}} \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{x_0^2}{2\sigma_0^2}\right\} dx_0$$

La formule multiplicative (4.4) doit donc se lire comme une formule de Bayes séquentielle

$$\begin{aligned} \mathbb{P}^{(X_0, \dots, X_n)}(d(x_0, \dots, x_n)) &= \left[\prod_{p=1}^n \mathbb{P}^{X_p | X_{p-1}}(dx_p | x_{p-1}) \right] \mathbb{P}^{X_0}(dx_0) \\ &= \eta_0(dx_0) M_1(x_0, dx_1) \dots M_n(x_{n-1}, dx_n) \end{aligned}$$

4.2.2 Formulation dynamique

Notons $W = (W_n)_{n \geq 0}$ une suite de v.a. indépendantes et gaussiennes, et telles que

$$\mathbb{P}^{(W_0, \dots, W_n)}(d(w_0, \dots, w_n)) = \prod_{p=0}^n \mathbb{P}^{W_p}(dw_p) = \prod_{p=0}^n \left(\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{w_p^2}{2}\right\} dw_p \right)$$

On associe à cette séquence de v.a. gaussiennes, le système dynamique

$$\begin{cases} X_n &= a_n X_{n-1} + \sigma_n W_n \\ X_0 &= \sigma_0 W_0 \end{cases} \quad (4.5)$$

Par construction, nous avons pour toute fonction bornée f

$$\begin{aligned}
& \mathbb{E}(f(X_n)|X_{n-1} = x_{n-1}) \\
&= \mathbb{E}(f(x_{n-1} + \sigma_n W_n)|X_{n-1} = x_{n-1}) \\
&= \int_{-\infty}^{+\infty} f(x_{n-1} + \sigma_n w_n) \frac{e^{-\frac{w_p^2}{2}}}{\sqrt{2\pi}} dw_p \\
&= \int_{-\infty}^{+\infty} f(x_n) \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{1}{2\sigma_n^2} (x_n - a_n - x_{n-1})^2} dx_n = \int M_n(x_{n-1}, dx_n) f(x_n)
\end{aligned}$$

Il est donc équivalent de définir la chaîne X_n soit de façon dynamique, soit directement par la donnée de ses transitions de probabilités.

4.3 Processus de Poisson

Le processus de Poisson est souvent associé à des phénomènes de comptages dans le temps : arrivées de clients dans une file d'attente, nombres de transactions journalières autour d'une action boursière, arrivées et départs d'avions dans des aéroports, nombres d'appels dans un central téléphonique, etc. Nous renvoyons le lecteur à l'ouvrage de F.A. Haight [?] consacré aux différentes applications de ce processus.

D'un point de vue mathématique, le processus de Poisson est défini en terme d'une suite de v.a. $(T_n)_{n \geq 0}$, à valeurs positives, et de lois marginales données par la formule

$$\begin{aligned}
& \mathbb{P}^{(T_0, \dots, T_n)}(d(t_0, \dots, t_n)) \tag{4.6} \\
&= [1_{[0, \infty)}(t_0) \lambda e^{-\lambda t_0} dt_0] [1_{[t_0, \infty)}(t_1) \lambda e^{-\lambda(t_1 - t_0)} dt_1] \\
&\quad \dots \times [1_{[t_{n-1}, \infty)}(t_n) \lambda e^{-\lambda(t_n - t_{n-1})} dt_n]
\end{aligned}$$

Dans la formule ci-dessus, $\lambda > 0$ désigne un paramètre fixé. En interprétant (4.6) comme une formule de Bayes séquentielle, on obtient

$$\mathbb{P}^{(T_0, \dots, T_n)}(d(t_0, \dots, t_n)) = \left[\prod_{p=1}^n \mathbb{P}^{T_p|T_{p-1}}(dt_p|t_{p-1}) \right] \mathbb{P}^{T_0}(dt_0)$$

avec

$$\begin{aligned}
\mathbb{P}^{T_0}(dt_0) &= 1_{[0, \infty)}(t_0) \lambda e^{-\lambda t_0} dt_0 \\
\mathbb{P}^{T_n|T_{n-1}}(dt_n|t_{n-1}) &= 1_{[t_{n-1}, \infty)}(t_n) \lambda e^{-\lambda(t_n - t_{n-1})} dt_n
\end{aligned}$$

Par un simple changement de variable, on montre aussi facilement que

$$U_0 =_{\text{déf.}} T_0 \quad \text{et} \quad U_n =_{\text{déf.}} (T_n - T_{n-1})$$

forment une suite de v.a. exponentielles et indépendantes de paramètre λ . Le processus continu de comptage des sauts

$$t \in \mathbb{R}_+ \mapsto N(t) = \sum_{n \geq 0} 1_{[T_n, \infty)}(t) \in \mathbb{N}$$

est appelé le Processus de Poisson d'intensité λ . Une réalisation de ce processus est donnée dans la figure suivante.

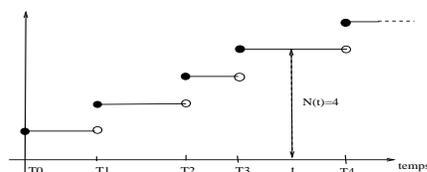


FIG. 4.1 – Processus de Poisson

4.4 Évolutions dans des milieux absorbants

En termes physiques, la chaîne de Markov suivante représente l'évolution d'une particule physique dans un puit de potentiel absorbant. Ce modèle peut aussi s'interpréter comme une désintégration, ou encore comme une absorption de la radioactivité dans des containers de stockages de déchets nucléaires.

Très brièvement, une particule évolue aléatoirement dans un environnement absorbant avec un taux de survie $G(x) \in [0, 1]$, en chaque site x .

Les obstacles sont d'autant plus absorbants que les valeurs de G sont proches de 0. Inversement, la particule évolue librement dans des régions où G vaut 1. Les régions où le potentiel est strictement inférieur à 1 représentent des obstacles dans lesquels la particules peut être piégée.

Ces trappes peuvent représenter des **niveaux de sécurité** dans tout type d'environnements, tels des aéroports, des milieux carcéraux, des chaînes de production, des réseaux de télécommunications, des containers de stockage de radioactivité, des tissus cellulaires, etc.

En pharmacologie, ces modèles d'absorption sont aussi utilisés pour modéliser l'évolution des taux de leucocytes dans des traitements du cancer. Dans ce contexte, les trappes reflètent des chutes de niveaux de leucocytes dans lesquels un individu risque de décéder.

En biologie et en recherche médicale, ces modèles d'absorption sont aussi utilisés pour modéliser l'évolution de photons émis par un laser sur un tissu cellulaire. Les trappes représentent des régions cellulaires, telles que des tumeurs plus sombres absorbant les photons. L'analyse de ces modèles permet de localiser et d'analyser ces régions de gonflement pathologiques des tissus.

Enfin, ces modèles peuvent aussi s'interpréter comme des **processus économiques** tels que des évolutions de portefeuilles, ou tout autre indicateur économique, dans des milieux financiers ou géopolitiques-politiques.

Dans tous ces domaines d'applications, l'un des problèmes majeurs est de calculer les probabilités de défaillances, autrement dit les probabilités pour que l'absorption de la particule ne soit pas effective à des instants donnés.

Pour fixer les idées, nous conviendrons par la suite que l'espace des états est donné par le réseau \mathbb{Z}^d . Plus formellement, on adjoint à l'espace des sites \mathbb{Z}^d , un point cimetière " c ". On considère alors une particule évoluant sur l'espace augmenté $E = \mathbb{Z}^d \cup \{c\}$, selon une transition $Y_n \rightsquigarrow Y_{n+1}$ se décomposant en deux étapes

$$Y_n \xrightarrow{\text{absorption}} \widehat{Y}_n \xrightarrow{\text{évolution}} Y_{n+1}$$

Pour construire le mécanisme d'absorption, on considère à chaque instant n , une collection de v.a. indépendantes $(\epsilon_n(x))_{x \in \mathbb{Z}^d}$, de Bernoulli à valeurs dans $\{0, 1\}$, de paramètre $G(x)$. Sur l'évènement $\epsilon_n(x) = 0$, le site x devient une pure trappe hautement absorbante. Inversement, si $\epsilon_n(x) = 1$, le site x reste viable, et la particule peut le traverser sans encombre.

Le mécanisme d'évolution libre est analogue à celui de la marche aléatoire sur \mathbb{Z}^d présenté dans l'exemple 3.3.4. Pour le décrire, on se donne une mesure de probabilité p , sur l'ensemble des $2d$ vecteurs unitaires directionnels

$$\mathcal{U} = \{u \in \mathbb{Z}^d : |u| = 1\}$$

ainsi qu'une suite de vecteurs indépendants $(U_n)_{n \geq 1}$ de même loi p sur \mathcal{U} . On suppose que les suites de v.a. d'exploration U_n , et de v.a. d'absorption $\epsilon_n(x)$, sont indépendantes. On convient enfin que $\epsilon_n(c) = 0 = g(c)$.

Les transitions élémentaires de la chaîne sont définies récursivement de la façon suivante. Supposons que la particule se trouve à l'instant n sur un site $Y_n = x \in \mathbb{Z}^d$ (si $Y_n = c$, on pose $Y_{n+1} = c$).

1. Avec une probabilité $(1 - G(x))$, la particule est tuée, puis placée dans l'état cimetièrre. Dans ce cas, on pose $\widehat{Y}_n = c$. Dans le cas contraire, la particule reste active et l'on pose $\widehat{Y}_n = x$. Plus formellement, on a

$$\widehat{Y}_n = \epsilon_n(Y_n) Y_n + (1 - \epsilon_n(Y_n)) c$$

2. Comme il n'y a semble-t-il pas de vie après la mort, lorsque la particule a été tuée, elle reste inactive à l'instant suivant. Dans ce cas, on pose $Y_{n+1} = c$. Dans le cas contraire, la particule est encore active, et effectue un mouvement exploratoire analogue à celui d'une marche aléatoire sur \mathbb{Z}^d . Plus formellement, on pose dans ce dernier cas

$$Y_{n+1} = \widehat{Y}_n + U_{n+1}$$

Par construction, la suite

$$Y_0 \rightarrow \widehat{Y}_0 \rightarrow \dots \rightarrow Y_n \rightarrow \widehat{Y}_n \rightarrow Y_{n+1} \rightarrow \widehat{Y}_{n+1} \rightarrow \dots$$

forme une chaîne de Markov de transitions

$$\begin{aligned} \mathbb{P}(\widehat{Y}_n = y \mid Y_n = x) &= G(x) 1_x(y) + (1 - G(x)) 1_c(y) \\ \mathbb{P}(Y_{n+1} = z \mid \widehat{Y}_n = y) &= 1_c(y) 1_c(z) + 1_{\mathbb{Z}^d}(y) K(y, z) \end{aligned}$$

avec la probabilité de transition

$$K(y, z) = \sum_{u \in \mathcal{U}} p(u) 1_{y+u}(z)$$

La figure suivante montre deux réalisations de trajectoires absorbées dans des "poches" d'obstacles associés à des régions où le potentiel $G(x) < 1$. Lorsque $G(x) = 1$, la particule ne subit pas le mécanisme d'absorption, et évolue librement comme une marche aléatoire sur \mathbb{Z}^2 . Les traits pointillés témoignent du fait qu'une particule visitant une poche d'obstacles s'essouffle. Sa durée de vie diminue à chaque instant, en "subissant" les v.a. trappes de Bernoulli.

Exercice 4.4.1 On note T l'instant d'absorption de la particule

$$T = \inf \{n \geq 0 : \widehat{Y}_n = c\} = \inf \{n \geq 0 : \epsilon_n(Y_n)_n = c\}$$

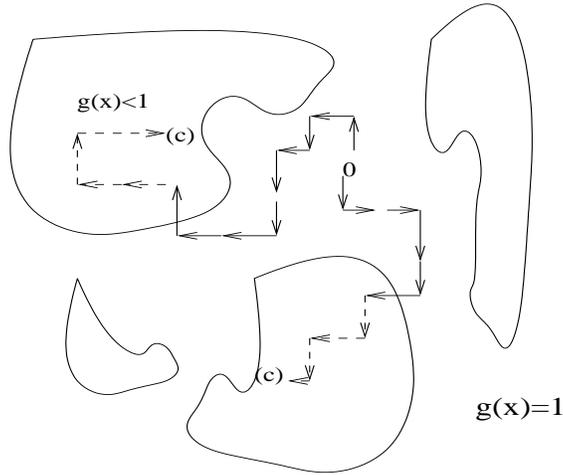


FIG. 4.2 – Particule absorbée dans \mathbb{Z}^2

1. Montrer que

$$\mathbb{P}(T > n) = \mathbb{E} \left(\prod_{p=0}^n G_p(X_p) \right)$$

où X_n désigne une chaîne de Markov sur \mathbb{Z}^d de transitions de probabilités K , et de même loi que Y_0 sur \mathbb{Z}^d . Lorsque le potentiel est uniformément majoré $G(x) \leq e^{-\lambda}$, avec $\lambda > 0$, montrer que

$$\mathbb{P}(T > n) \leq e^{-\lambda(n+1)}$$

Interpréter ce résultat.

2. Vérifier que la loi d'une particule non absorbée est donnée pour toute fonction bornée f sur \mathbb{Z}^d par la formule renormalisée de Feynman-Kac

$$\mathbb{E}(f(Y_n) \mid T > n) = \frac{\mathbb{E} \left(f(X_n) \prod_{p=0}^n G_p(X_p) \right)}{\mathbb{E} \left(\prod_{p=0}^n G_p(X_p) \right)}$$

Chapitre 5

Chaînes de Markov non linéaires

5.1 Introduction

Les modèles de chaînes de Markov non linéaires présentés dans cette section sont une extension naturelle des modèles markoviens étudiés dans la première partie de ce chapitre.

Les premiers modèles de ce type sont semble-t-il apparus dans la littérature du traitement du signal, et plus particulièrement en filtrage non linéaire ([?, ?]). L'idée de départ est la suivante. Les lois conditionnelles des états X_n du signal par rapports aux observations reçues $(Y_p)_{0 \leq p < n}$ sont données par un flot de mesures

$$\eta_0 = \text{Loi}(X_0) \longrightarrow \eta_1 = \text{Loi}(X_1 | Y_0) \longrightarrow \dots \longrightarrow \eta_n = \text{Loi}(X_n | Y_0, \dots, Y_{n-1}) \longrightarrow \dots$$

Ces mesures sont appelées les prédicteurs optimaux. Elles peuvent être calculées de façon récursives suivant une équation de la forme

$$\eta_n = \Phi_n(\eta_{n-1})$$

où Φ_n désigne une transformation plus ou moins complexe sur l'espace des distributions sur l'espace d'état du signal. Bien évidemment, ces équations sont en général impossible à résoudre explicitement. Pour utiliser des méthodes de simulation de type Monte Carlo, l'idée naturelle est d'interpréter ces mesures η_n comme les lois d'une chaîne de Markov \bar{X}_n sur l'espace d'états du signal :

$$\eta_n = \text{Loi}(\bar{X}_n)$$

Ces chaîne de Markov \bar{X}_n ne sont pas uniques. On peut choisir par exemple des suite de variables indépendentes \bar{X}_n de lois

$$\eta_n = \Phi_n(\eta_{n-1}) \quad \text{avec} \quad \eta_{n-1} = \text{Loi}(\bar{X}_{n-1})$$

On peut aussi chercher à écrire les transformations Φ_n comme un transport de mesures markovien

$$\eta_n = \Phi_n(\eta_{n-1}) = \eta_{n-1}K_{n,\eta_{n-1}}$$

Dans la formule précédente, $K_{n,\eta}(x_{n-1}, dx_n)$ désigne une famille de probabilités de transitions de E_{n-1} vers un espace E_n , indexées par le paramètre temporel et par les mesures η sur E_{n-1} . Dans ce contexte, nous avons à nouveau

$$\eta_n = \text{Loi}(\bar{X}_n)$$

où $(\bar{X}_n)_{n \geq 0}$ désigne une chaîne de Markov de loi initiale $\eta_0 = \text{Loi}(\bar{X}_0)$ sur E_0 et de transitions de probabilités élémentaires d'un espace E_{n-1} vers un espace E_n décrites par :

$$\mathbb{P}(\bar{X}_n \in dx_n \mid \bar{X}_{n-1} = x_{n-1}) = K_{n,\eta_{n-1}}(x_{n-1}, dx_n) \quad \text{avec} \quad \eta_{n-1} = \text{Loi}(\bar{X}_{n-1})$$

Nous étudierons ces modèles mathématiques avec plus de détails dans la section 5.2. Il me semble est important de souligner que ces modèles discrets sont très proches des équations à temps continu issues de la mécanique des fluides, telles les équation diffusives de type McKean-Vlasov et/ou les modèles de gaz de type Boltzmann.

5.2 Description des modèles

La plus simple façon de les définir une chaîne de Markov non linéaires est de se donner un chaîne de Markov $(\bar{X}_n)_{n \geq 0}$ de loi initiale $\eta_0 = \text{Loi}(\bar{X}_0)$ sur E_0 et de transitions de probabilités élémentaires d'un espace E_{n-1} vers un espace E_n décrites par une formule de la forme suivante :

$$\mathbb{P}(\bar{X}_n \in dx_n \mid \bar{X}_{n-1} = x_{n-1}) = K_{n,\eta_{n-1}}(x_{n-1}, dx_n) \quad \text{avec} \quad \eta_{n-1} = \text{Loi}(\bar{X}_{n-1})$$

Dans la définition précédente, $K_{n,\eta}(x_{n-1}, dx_n)$ désigne une famille de probabilités de transitions de E_{n-1} vers un espace E_n , indexées par le paramètre temporel et par les mesures η sur E_{n-1} .

Il est assez aisé de vérifier que l'on a

$$\mathbb{P}((\bar{X}_0, \dots, \bar{X}_n) \in d(x_0, \dots, x_n)) = \eta_0(dx_0)K_{1,\eta_0}(x_0, dx_1) \dots K_{n,\eta_{n-1}}(x_{n-1}, dx_n)$$

Ces mesures sont appelés les mesures de McKean associés aux transitions de probabilités $(K_{n,\eta})_{n,\eta}$. Une simple intégration par rapports aux coordonnées temporelles permet de vérifier que

$$\text{Loi}(\bar{X}_n) = \eta_0 K_{1,\eta_0} \dots K_{n,\eta_{n-1}} = \eta_{n-1} K_{n,\eta_{n-1}}$$

Une simple récurrence permet de s'assurer que

$$\eta_n = \text{Law}(\bar{X}_n) = \eta_{n-1} K_{n,\eta_{n-1}}$$

5.3 Interprétations particulières en champ moyen

Supposons dans un premier temps que les transitions de la chaîne

$$\bar{X}_{n-1} \rightsquigarrow \bar{X}_n$$

sont faciles à simuler. Dans ces conditions, pour approcher les lois η_n de la chaîne à chaque instant, il suffit de simuler N copies indépendantes $(\bar{X}_n^i)_{1 \leq i \leq N}$ de la chaîne \bar{X}_n . Ce schéma de simulation numérique est illustré par la figure suivante

$$\begin{array}{ccc} \bar{X}_{n-1}^1 & \xrightarrow{K_{n,\eta_{n-1}}} & \bar{X}_n^1 \\ \vdots & & \vdots \\ \bar{X}_{n-1}^i & \xrightarrow{K_{n,\eta_{n-1}}} & \bar{X}_n^i \\ \vdots & & \vdots \\ \bar{X}_{n-1}^N & \xrightarrow{K_{n,\eta_{n-1}}} & \bar{X}_n^N \end{array}$$

Une simple application de la loi des grands nombres nous donne l'approximation suivante :

$$\forall n \geq 0 \quad \eta_n^N := \frac{1}{N} \sum_{i=1}^N \delta_{X_n^i} \simeq_{N \uparrow \infty} \eta_n$$

Malheureusement, dans la plupart des cas les lois de ces chaînes non linéaires η_n n'ont aucune expression analytique explicite et/ou ne peuvent être simulées de façon exacte par aucun algorithme de simulation dans un temps raisonnable.

L'idée des méthodes particulières de type champ moyen est d'utiliser la population courante pour approcher ces lois complexes. Sous certaines conditions de régularités sur les transitions $K_{n,\eta}$, nous avons

$$\eta_{n-1}^N \simeq_{N \uparrow \infty} \eta_{n-1} \longrightarrow K_{n,\eta_{n-1}^N} \simeq_{N \uparrow \infty} K_{n,\eta_{n-1}}$$

En utilisant ces formules d'approximation, on définit récursivement une chaîne de Markov

$$(\xi_n^{(1,N)}, \xi_n^{(2,N)}, \dots, \xi_n^{(N,N)})_{n \geq 0}$$

sur les espaces produits E_n^N . L'état initial $(\xi_0^{(1,N)}, \xi_0^{(2,N)}, \dots, \xi_0^{(N,N)})$ est donné par N variables aléatoires indépendantes et identiquement distribuées de loi η_0 . Les transitions élémentaires de cette chaîne sont données par le schéma suivant

$$\begin{array}{ccc}
\xi_{n-1}^{(1,N)} & \xrightarrow{K_{n,\eta_{n-1}^N}} & \xi_n^{(1,N)} \\
\vdots & & \vdots \\
\xi_{n-1}^{(i,N)} & \xrightarrow{K_{n,\eta_{n-1}^N}} & \xi_n^{(i,N)} \\
\vdots & & \vdots \\
\xi_{n-1}^{(N,N)} & \xrightarrow{K_{n,\eta_{n-1}^N}} & \xi_n^{(N,N)}
\end{array}$$

Plus formellement, nous avons

$$\mathbb{P} \left(\xi_{n+1}^{(N)} \in d(x^1, \dots, x^N) \mid \xi_n^{(N)} \right) = \prod_{i=1}^N K_{n+1,\eta_n^N}(\xi_n^{(N,i)}, dx^i) \quad (5.1)$$

avec les mesures d'occupation du système

$$\eta_n^N := \frac{1}{N} \sum_{j=1}^N \delta_{\xi_n^{(N,j)}}$$

Autrement dit, connaissant une réalisation du système au temps n , la population des individus au temps $(n+1)$

$$\left(\xi_{n+1}^{(N,1)}, \xi_{n+1}^{(N,2)}, \dots, \xi_{n+1}^{(N,N)} \right)$$

est formée de N variables aléatoires indépendantes de lois respectives

$$K_{n+1,\eta_n^N}(\xi_n^{(N,1)}, dx^1), \quad K_{n+1,\eta_n^N}(\xi_n^{(N,2)}, dx^2), \quad \dots \quad K_{n+1,\eta_n^N}(\xi_n^{(N,N)}, dx^N)$$

Sous certaines hypothèses de régularité, on peut montrer que l'on a en un certain sens

$$\eta_n^N := \frac{1}{N} \sum_{j=1}^N \delta_{\xi_n^{(N,j)}} \simeq_{N \uparrow \infty} \eta_n$$

5.4 Champs moyens de type gaussien

On considère l'équation d'évolution suivante sur l'ensemble $\mathcal{P}(\mathbb{R}^d)$ des mesures de probabilités sur l'espace d -dimensionnel \mathbb{R}^d :

$$\eta_{n+1}(dy) = (\eta_n K_{n+1,\eta_n}) (dy) := \int_{\mathbb{R}^d} \eta_n(dx) K_{n+1,\eta_n}(x, dy) \quad (5.2)$$

Dans la définition précédente, les probabilités de transitions $K_{n,\eta}$ sur $E_n = \mathbb{R}^d$ définis par les mesures gaussiennes :

$$K_{n,\eta}(x, dy) = \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{|Q_n|}} \exp -\frac{1}{2} [(y - a_n(x, \eta))' Q_n^{-1} (y - a_n(x, \eta))] dy$$

où

- Q_n désigne une matrice ($d \times d$) symétrique et définie positive, $|Q_n| := \det(Q_n)$.
- dy désigne la mesure de Lebesgue sur \mathbb{R}^d , $y = (y^1, \dots, y^d)$ et $x \in \mathbb{R}^d$.
- $a_n : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ est une fonctions suffisamment régulière et bornée.

Cet exemple est assez intéressant car les mesures de McKean associées

$$\eta_0(x_0)K_{1,\eta_0}(x_0, dx_1) \dots K_{n,\eta_{n-1}}(x_{n-1}, dx_n)$$

peuvent s'interpréter comme les lois des trajectoires $(\bar{X}_0, \dots, \bar{X}_n)$ d'une chaîne de Markov donnée par les équations cinétiques suivantes :

$$\bar{X}_n = a_n(\bar{X}_{n-1}, \eta_{n-1}) + W_n, \quad n \geq 1$$

avec

- W_n , $n \geq 1$, une suite de variables aléatoires indépendantes, à valeurs dans \mathbb{R}^d , et de lois gaussienne

$$\lambda_n(dy) = \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{|Q_n|}} \exp -\frac{1}{2} [y' Q_n^{-1} y] dy$$

- \bar{X}_0 une variable aléatoire de loi η_0 , et indépendante de la suite W_n .
- pour chaque $n \geq 0$, η_n est la loi de \bar{X}_n .

L'interprétation particulière de type champ moyen (5.1) associée au modèle d'évolution (5.2) est la chaîne de Markov

$$\xi_n^{(N)} = \left(\xi_n^{(N,1)}, \xi_n^{(N,2)}, \dots, \xi_n^{(N,N)} \right) \in (\mathbb{R}^d)^N$$

définie par les équations cinétiques suivantes :

$$\forall 1 \leq i \leq N \quad \xi_n^{(N,i)} = a_n \left(\xi_{n-1}^{(N,i)}, \eta_{n-1}^N \right) + W_n^i \quad \text{avec} \quad \eta_{n-1}^N := \frac{1}{N} \sum_{j=1}^N \delta_{\xi_{n-1}^{(N,j)}}$$

Dans cette formulation, la suite $(W_n^i)_{1 \leq i \leq N}$ est formée de N copies indépendantes de W_n . L'états initial $(\xi_0^{(N,i)})_{1 \leq i \leq N}$ est encore formé de N copies indépendantes de \bar{X}_0 .

5.5 Modèles simplifiés de gaz de McKean

Dans cette section, nous présentons le modèle simplifié de gaz à deux vitesses de McKean étudié dans l'article [?]. L'analyse qui suit peut s'étendre à des modèles plus sophistiqués avec des vitesses multiples. Le modèle le plus élémentaire est défini sur l'espace $E = \{-1, +1\}$ en associant à chaque mesure η dans l'ensemble $\mathcal{P}(\{-1, +1\})$ des mesures de probabilités sur $E = \{-1, +1\}$ la transition suivante :

$$K_{n,\eta}(x, dy) = \eta(+1) \delta_x(dy) + \eta(-1) \delta_{-x}(dy). \quad (5.3)$$

On associe à ces transitions le processus à valeurs mesure suivant :

$$\begin{aligned} \eta_{n+1}(+1) &= \eta_n(+1) K_{n+1,\eta_n}(+1, +1) + \eta_n(-1) K_{n+1,\eta_n}(-1, +1) \\ &= \eta_n(+1)^2 + \eta_n(-1)^2 = \eta_n(+1)^2 + (1 - \eta_n(+1))^2 \end{aligned}$$

On pourra remarquer que l'on a

$$\begin{aligned} \eta_{n+1}(-1) &= \eta_n(+1) K_{n+1,\eta_n}(+1, -1) + \eta_n(-1) K_{n+1,\eta_n}(-1, -1) \\ &= 2 \eta_n(+1)\eta_n(-1) \end{aligned}$$

L'interprétation particulière (5.1) associée à ces équations est la chaîne de Markov

$$\xi_n^{(N)} = \left(\xi_n^{(N,1)}, \xi_n^{(N,2)}, \dots, \xi_n^{(N,N)} \right) \in \{-1, +1\}^N$$

définie par la formule

$$\forall 1 \leq i \leq N \quad \xi_n^{(N,i)} = \epsilon_n^{(N,i)} \xi_{n-1}^{(N,i)}$$

Dans la définition précédente $(\epsilon_n^{(N,i)})_{1 \leq i \leq N}$ représente une suite de variables aléatoires conditionnellement indépendantes sur $\{-1, +1\}$ de loi :

$$\mathbb{P} \left(\epsilon_n^{(N,i)} = \epsilon \mid \xi_{n-1}^{(N)} \right) = \eta_{n-1}^N(+1) 1_1(\epsilon) + \eta_{n-1}^N(-1) 1_{-1}(\epsilon) \quad \text{avec} \quad \eta_{n-1}^N := \frac{1}{N} \sum_{j=1}^N \delta_{\xi_{n-1}^{(N,j)}}$$

5.6 Flots de mesures de Feynman-Kac

5.6.1 Description des modèles

On se donne une suite de fonctions positives et bornées G_n sur E_n . On note $\mathcal{P}_{G_n}(E_n)$ l'ensemble des mesures de probabilités μ sur E_n telles que $\mu(G_n) \neq 0$. On associe à chaque fonction potentiel G_n une transformation de Boltzmann-Gibbs Ψ_{G_n} définie par les formules suivantes

$$\Psi_{G_n} : \mu \in \mathcal{P}_{G_n}(E) \mapsto \Psi_{G_n}(\mu) \in \mathcal{P}_{G_n}(E)$$

avec la mesure de probabilité $\Psi_{G_n}(\mu)$ donnée par

$$\Psi_{G_n}(\mu)(dx) := \frac{1}{\mu(G_n)} G_n(x) \mu(dx)$$

On se donne $M_n(x_{n-1}, dx_n)$ une suite de probabilités de transitions de E_{n-1} dans E_n , avec $n \geq 1$. On considère enfin une mesure de probabilité η_0 sur E_0 telle que $\eta_0(G_0) > 0$. On notera par la suite X_n une chaîne de Markov à valeurs dans E_n de loi initiale η_0 et de probabilités de transitions

$$M_n(x_{n-1}, dx_n) := \mathbb{P}(X_n \in dx_n \mid X_{n-1} = x_{n-1})$$

On convient que

$$\forall \hat{\eta}_n \in \mathcal{P}_{G_n}(E_n) \quad \hat{\eta}_n M_{n+1} \in \mathcal{P}_{G_{n+1}}(E_{n+1})$$

autrement dit

$$\hat{\eta}_n(G_n) > 0 \implies \hat{\eta}_n M_{n+1}(G_n) > 0$$

Lorsque les fonctions potentiels sont strictement positives, les espaces $\mathcal{P}_{G_n}(E_n)$ coïncident avec l'espace de toutes les mesures de probabilités sur E_n . Dans ce cas, les conditions précédentes sont trivialement satisfaites.

On considère $(\eta_n)_{n \geq 0}$ le flot de mesures de probabilités défini par les équations :

$$\eta_n = \Phi_n(\eta_{n-1}) := \Psi_{G_{n-1}}(\eta_{n-1})M_n \quad (5.4)$$

Ces mesures η_n et leurs mesures mises à jour ou corrigées

$$\hat{\eta}_n := \Psi_{G_n}(\eta_n)$$

peuvent s'exprimer sous forme d'intégrales de chemins pondérés connus sous le nom de formules de Feynman-Kac. Ces représentations fonctionnelles sont données sur des fonctions tests f_n par les formules suivantes

$$\eta_n(f_n) = \gamma_n(f_n)/\gamma_n(1) \quad \text{et} \quad \hat{\eta}_n(f_n) = \hat{\gamma}_n(f_n)/\hat{\gamma}_n(1) \quad (5.5)$$

avec des mesures non normalisées γ_n et $\hat{\gamma}_n$ décrites ci-dessous :

$$\gamma_n(f_n) = \mathbb{E}[f_n(X_n) \prod_{0 \leq k < n} G_k(X_k)] \quad \text{et} \quad \hat{\gamma}_n(f_n) = \gamma_n(f_n G_n) \quad (5.6)$$

Pour vérifier cette formule, on utilise tout d'abord la propriété de Markov pour vérifier que

$$\begin{aligned} \gamma_n(f_n) &= \mathbb{E} \left(\mathbb{E}(f_n(X_n) \mid (X_p)_{0 \leq p < n}) \prod_{0 \leq p < n} G_p(X_p) \right) \\ &= \mathbb{E} \left(\mathbb{E}(f_n(X_n) \mid X_{n-1}) \prod_{0 \leq p < n} G_p(X_p) \right) \end{aligned}$$

Ceci entraîne que

$$\gamma_n(f_n) = \mathbb{E} \left(M_n(f_n)(X_{n-1}) \prod_{0 \leq p < n} G_p(X_p) \right)$$

avec

$$M_n(f_n)(x_{n-1}) := \mathbb{E}(f_n(X_n) \mid X_{n-1} = x_{n-1})$$

Dans un second temps, on constate que l'on a :

$$\gamma_n(f_n) = \mathbb{E} \left(G_{n-1}(X_{n-1}) M_n(f_n)(X_{n-1}) \prod_{0 \leq p < (n-1)} G_p(X_p) \right)$$

Par définition de la mesure γ_{n-1} , nous avons

$$\gamma_n(f_n) = \gamma_{n-1}(G_{n-1}M_n(f_n)) \quad \text{and} \quad \gamma_n(1) = \gamma_{n-1}(G_{n-1})$$

Ceci entraîne que

$$\eta_n(f_n) = \frac{\gamma_n(f_n)}{\gamma_n(1)} = \frac{\gamma_{n-1}(G_{n-1}M_n(f_n))}{\gamma_{n-1}(G_{n-1})}$$

et par conséquent

$$\eta_n(f_n) = \frac{\gamma_{n-1}(G_{n-1}M_n(f_n))/\gamma_{n-1}(1)}{\gamma_{n-1}(G_{n-1})/\gamma_{n-1}(1)} = \frac{\eta_{n-1}(G_{n-1}M_n(f_n))}{\eta_{n-1}(G_{n-1})}$$

On en conclut que

$$\forall f_n \in \mathcal{B}_b(E_n) \quad \eta_n(f_n) = \Psi_{G_{n-1}}(\eta_{n-1})(M_n(f_n))$$

\Downarrow

$$\eta_n = \Psi_{G_{n-1}}(\eta_{n-1})M_n$$

Les mêmes arguments permettent d'analyser les mesures $(\hat{\gamma}_n, \hat{\eta}_n)$.

Ces mesures de probabilités permettent de modéliser une variété considérable de problèmes issus de la physique, ou de la biologie : traitement du signal non linéaire, description de macro-polymères et de chaînes auto-évitant, analyse d'évènements rares, représentation de valeurs propres et d'états fondamentaux d'opérateurs de Schrödinger,... Nous examinerons un certain nombre de ces questions dans le chapitre ??, pour plus de détails nous renvoyons le lecteur aux ouvrages [?], et [?]. Dans la section ??, nous présentons un algorithme de simulation universel de ces mesures de Feynman-Kac. Ces modèles particuliers sont fondés sur l'évolution d'individus en interaction explorant l'espace selon des mécanismes de mutation et sélection de type génétique.

5.6.2 Chaînes de Markov non linéaires

Commençons par remarquer qu'une transformation de Boltzmann-Gibbs

$$\Psi_G(\mu)(dx) := \frac{1}{\mu(G)} G(x) \mu(dx)$$

associée à une fonction potentiel G sur un espace d'état E , peut s'interpréter comme un transport de mesure markovien. Pour être plus précis, on note $\epsilon(\mu)$ une famille de paramètre pouvant dépendre de G et μ , et telle que

$$\epsilon(\mu) G(x) \leq 1 \quad \text{pour } \mu\text{-presque tous les } x \in E.$$

Avec ce système de notations, nous avons

$$\Psi_G(\mu) = \mu S_\mu \quad \left(\iff \Psi_G(\mu)(dy) = \int_E \mu(dx) S_\mu(x, dy) \right)$$

avec la famille de probabilités de transition sur E données par la formule suivante :

$$S_\mu(x, dy) = \epsilon(\mu) G(x) \delta_x(dy) + (1 - \epsilon(\mu) G(x)) \Psi_G(\mu)(dy).$$

Pour vérifier cette assertion, on se donne une fonction test bornée f sur E et l'on observe que

$$S_\mu(f)(x) = \epsilon(\mu) G(x) f(x) + (1 - \epsilon(\mu) G(x)) \Psi_G(\mu)(f)$$

On en déduit les formules suivantes :

$$\begin{aligned} \mu(S_\mu(f)) &= \epsilon(\mu) \mu(Gf) + (1 - \epsilon(\mu) \mu(G)) \Psi_G(\mu)(f) \\ &= \epsilon(\mu) \mu(Gf) + \Psi_G(\mu)(f) - \epsilon(\mu) \mu(G) \frac{\mu(Gf)}{\mu(G)} = \Psi_G(\mu)(f). \end{aligned}$$

Reprenons les équations non linéaires (5.4) vérifiées par les flots de mesures de Feynman-Kac $(\eta_n)_{n \geq 0}$ introduites en (5.5)

$$\eta_{n+1} = \Phi_{n+1}(\eta_n) := \Psi_{G_n}(\eta_n) M_{n+1}$$

D'après les calculs précédents, nous avons l'équation de transport

$$\Psi_{G_n}(\eta_n) = \eta_n S_{n, \eta_n}$$

avec les transitions $S_{n, \eta_n}(x_n, dy_n)$ sur E_n définies par les formules suivantes

$$S_{n, \eta_n}(x_n, dy_n) = \epsilon_n(\eta_n) G_n(x_n) \delta_{x_n}(dy_n) + (1 - \epsilon_n(\eta_n) G_n(x_n)) \Psi_n(\eta_n)(dy_n)$$

avec des paramètres $\epsilon_n(\eta_n) \geq 0$ tels que $\epsilon_n(\eta_n)G_n(x_n) \leq 1$, pour tous les $x_n \in E_n$. On en conclut que

$$\eta_{n+1} = \eta_n K_{n+1, \eta_n}$$

avec la famille de transitions de probabilités composées

$$\begin{aligned} K_{n+1, \eta_n}(x_{n-1}, dx_n) &= S_{n, \eta_n} M_{n+1}(x_{n-1}, dx_n) \\ &= \int_{E_n} S_{n, \eta_n}(x_n, dy_n) M_{n+1}(y_n, x_{n+1}) dx_{n+1} \end{aligned} \quad (5.7)$$

5.6.3 Champs moyens de type évolutionnaire

L'interprétation particulière (5.1) associée aux équations (5.2) est la chaîne de Markov

$$\xi_n^{(N)} = \left(\xi_n^{(N,1)}, \xi_n^{(N,2)}, \dots, \xi_n^{(N,N)} \right) \in (E_n)^N$$

de transitions élémentaires données par la formule suivante :

$$\mathbb{P} \left(\xi_{n+1}^{(N)} \in dx_{n+1} \mid \xi_n^{(N)} \right) = \prod_{i=1}^N \left(S_{n, \eta_n^N} M_{n+1} \right) \left(\xi_n^{(N,i)}, dx_{n+1}^i \right) \quad (5.8)$$

avec les mesures d'occupation η_n^N données par

$$\eta_n^N := \frac{1}{N} \sum_{j=1}^N \delta_{\xi_n^{(N,j)}}$$

On notera que l'on a

$$\mathbb{P} \left(\xi_{n+1}^{(N)} \in dx_{n+1} \mid \xi_n^{(N)} \right) = \int_{E_n^N} \mathcal{S}_n(\xi_n^{(N)}, dx_n) \mathcal{M}_{n+1}(x_n, dx_{n+1})$$

avec les transitions de Boltzmann-Gibbs \mathcal{S}_n de E_n^N dans lui même et les transitions d'exploration \mathcal{M}_{n+1} de E_n^N dans E_{n+1}^N définies par

$$\begin{aligned} \mathcal{S}_n(\xi_n^{(N)}, dx_n) &= \prod_{i=1}^N S_{n, \eta_n^N}(\xi_n^{(N,i)}, dx_n^i) \\ \mathcal{M}_{n+1}(x_n, dx_{n+1}) &= \prod_{i=1}^N M_{n+1}(x_n^i, dx_{n+1}^i) \end{aligned}$$

Ces décompositions soulignent le fait que la chaîne de Markov suit les deux mêmes étapes de mise à jour/correction et d'exploration/mutation que le flot des mesures de

Feynman-Kac limites. Plus formellement, les deux étapes de correction/prediction dans l'espace des distributions :

$$\eta_n \in \mathcal{P}(E_n) \xrightarrow{S_{n,\eta_n}} \hat{\eta}_n = \eta_n S_{n,\eta_n} = \Psi_{G_n}(\eta_n) \in \mathcal{P}(E_n) \xrightarrow{M_{n+1}} \eta_{n+1} = \hat{\eta}_n M_{n+1} \quad (5.9)$$

sont approchées par deux étapes évolutives de type sélection/mutation dans l'espace des mesures empiriques :

$$\xi_n^{(N)} \in E_n^N \xrightarrow{\text{selection}} \hat{\xi}_n^{(N)} \in E_n^N \xrightarrow{\text{mutation}} \xi_{n+1}^{(N)} \in E_{n+1}^N \quad (5.10)$$

L'équation d'évolution complète du système peut se résumer par le diagramme synthétique suivant :

$$\left[\begin{array}{c} \xi_n^{(N,1)} \\ \vdots \\ \xi_n^{(N,i)} \\ \vdots \\ \xi_n^{(N,N)} \end{array} \right] \xrightarrow{S_{n,\eta_n^N}} \left[\begin{array}{ccc} \hat{\xi}_n^{(N,1)} & \xrightarrow{M_{n+1}} & \xi_{n+1}^{(N,1)} \\ \vdots & & \vdots \\ \hat{\xi}_n^{(N,i)} & \longrightarrow & \xi_{n+1}^{(N,i)} \\ \vdots & & \vdots \\ \hat{\xi}_n^{(N,N)} & \longrightarrow & \xi_{n+1}^{(N,N)} \end{array} \right]$$

avec les probabilités d'acceptation-rejets :

$$\begin{aligned} & S_{n,\eta_n^N}(\xi_n^{(N,i)}, dx) \\ & := \epsilon_n(\eta_n^N) G_n(\xi_n^{(N,i)}) \delta_{\xi_n^{(N,i)}}(dx) + \\ & \quad \left(1 - \epsilon_n(\eta_n^N) G_n(\xi_n^{(N,i)})\right) \sum_{j=1}^N \frac{G_n(\xi_n^{(N,j)})}{\sum_{k=1}^N G_n(\xi_n^{(N,k)})} \delta_{\xi_n^{(N,j)}}(dx) \end{aligned}$$

Le choix du paramètre d'acceptation n'est pas unique. L'algorithme génétique simple de type mutation/sélection correspond au choix $\epsilon_n(\eta_n^N) = 0$. Si on pose $\epsilon_n(\eta_n^N) = 1/\sup_i G_n(\xi_n^{(N,i)})$, alors les élites de la population qui maximisent le potentiel sont toujours acceptées. Lorsque $\|G_n\|$ est explicitement connu, on peut aussi poser $\epsilon_n(\eta_n^N) = 1/\|G_n\|$.

Pour plus de détails concernant ces modèles avec notamment la description des arbres généalogiques associées nous renvoyons le lecteur au chapitre 7, ainsi qu'à l'ouvrage [?].

Chapitre 6

L'équation de la chaleur

6.1 Les fluctuations browniennes

Les physiciens se sont longtemps penché sur la définition de la chaleur. Depuis la nuit des temps, ou presque, nous savons que la chaleur peut être générée par frottements vigoureux entre objets. En est-il de même au niveau atomique? La répartition de la chaleur dans un corps serait elle dictée par les excitations et les évolutions chaotiques des différents atomes constituant la matière? Est il envisageable simuler l'agitation thermique induite par des collisions de molécules monoatomiques dans des gaz, ou dans des liquides? Peut on tenir compte des différents degrés de de frottement induits par la matière?

Pour comprendre l'origine de ces questions, il faut remonter au début du 19^{ième} siècle, et plus précisément en 1827. Cette année là, le biologiste Robert Brown s'amuse à observer les mouvements erratiques de différents grains de pollen dans une goutte d'eau. Ces grains de matière subissent des chocs incessants et imprévisibles avec les molécules du liquide en agitation permanente. Leur trajectoires aléatoires totalement erratiques et chaotiques, semblent néanmoins dictées par des lois physiques bien précises.

D'après la théorie stochastique moderne, la répartition de la chaleur est elle aussi l'expression de la répartition spatiale de "grains de chaleur" animés par ces mouvements browniens. Si l'on chauffe une barre de fer rectiligne en un point donné, les molécules de matière en ce point prennent de l'énergie, et se mettent à vibrer, et à se cogner en elles. Ces successions de collisions de chaque cotés confèrent aux particules des mouvements imprévisibles et chaotiques, allant vers la droite ou vers la gauche. Lorsque le temps s'écoule, les grains de chaleur se trouvent distribués sur la barre selon une distribution gaussienne centrée en la source de chaleur, et de plus en plus étalée sur les bords.

Le mouvement brownien uni-dimensionnel peut être peut être construit à partir d'une simple marche aléatoire sur la droite des réels \mathbb{R} . Un marcheur virtuel partant de l'origine choisit à chaque instant d'évoluer aléatoirement vers la droite ou vers la gauche. Ses déplacements vers la gauche $-\Delta X$ ou vers la droite $+\Delta X$, par unité de

temps Δt sont extrêmement amples et erratiques, en ce sens où

$$[\Delta X]^2 = \Delta t$$

Autrement dit, le marcheur effectue des amplitudes de déplacement du type $\pm\sqrt{\Delta t}$, par unité de temps Δt . Par exemple, pour des fréquences temporelles très rapides de $\Delta t = 10^{-10}$ secondes, les amplitudes spatiales sont données par, disons $\Delta X = 10^{-5}$ mètres. La vitesse de ce marcheur est extrêmement rapide. Il court tout simplement à $\Delta X/\Delta t = 10^5$ mètres par secondes !

Supposons que notre marcheur évolue dans des échelles sub-atomiques. Pour un observateur macroscopique cela revient à s'éloigner de plus en plus du mouvement décrit plus haut en faisant tendre les deux pas de la grille spatio-temporelle ΔX et Δt vers 0. Pour cet observateur, ces particules browniennes ont tout simplement une amplitude de vitesse infinie

$$\frac{\Delta X}{\Delta t} = \frac{\pm 1}{\sqrt{\Delta t}} \longrightarrow \pm\infty \quad \text{lorsque} \quad \Delta X \text{ et } \Delta t \downarrow 0$$

La construction probabiliste du mouvement brownien développée en 1923 par l'américain Norbert Wiener est assez sophistiquée. Dans ce qui suit, nous allons essayer d'en donner les grandes lignes. Par soucis de clarté, nous conviendrons que l'intervalle de temps est simplement donné par le segment unité $[0, 1]$. On choisit une séquence d'instantanés t_i équi-distribués

$$t_0 = 0 < t_1 = 1/n < \dots < t_i = i/n < \dots < t_{n-1} = (n-1)/n < t_n = 1$$

avec

$$t_i - t_{i-1} = \Delta t = 1/n$$

On construit sur cette subdivision, une marche aléatoire simple $(\beta_{t_i}^n)_{i=0, \dots, n}$ centrée en l'origine :

$$\beta_0^n = 0 \quad \text{et} \quad \Delta\beta_{t_i}^n = \beta_{t_i}^n - \beta_{t_{i-1}}^n = \epsilon_{t_i} \sqrt{\Delta t}$$

Les variables aléatoires ϵ_{t_i} représentent les mouvements aléatoires indépendants, vers la gauche ou bien vers la droite, de notre marcheur indécis

$$\mathbb{P}(\epsilon_{t_i} = +1) = \mathbb{P}(\epsilon_{t_i} = -1) = \frac{1}{2}$$

Par construction, cette marche aléatoire possède les deux propriétés essentielles suivantes

$$\begin{aligned} \Delta\beta_{t_i}^n \times \Delta\beta_{t_i}^n &= \Delta t \\ \mathbb{E}(\Delta\beta_{t_i}^n \mid \beta_{t_0}^n, \dots, \beta_{t_{i-1}}^n) &= 0 \end{aligned}$$

Pour vérifier la seconde assertion, on pourra noter que l'on a

$$\mathbb{E}(\Delta\beta_{t_i}^n \mid \beta_{t_0}^n, \dots, \beta_{t_{i-1}}^n) = \mathbb{E}(\epsilon_{t_i}) \sqrt{\Delta t} = \left(\frac{1}{2} - \frac{1}{2}\right) \sqrt{\Delta t} = 0$$

La première propriété concerne la vitesse du marcheur, la seconde souligne le fait que ses déplacements aléatoire et locaux sont nuls, en moyenne.

Entre chaque instant t_{i-1} et t_i , on convient que le marcheur évolue de façon rectiligne, sur la pente entre $\beta_{t_{i-1}}^n$ et $\beta_{t_i}^n$. Autrement dit, sa position aux temps $t \in [t_{i-1}, t_i[$ est donnée par l'interpolation linéaire des deux extrémités $\beta_{t_{i-1}}^n$ et $\beta_{t_i}^n$. Plus formellement, nous avons

$$\beta_t^n = \beta_{t_{i-1}}^n + \frac{\beta_{t_i}^n - \beta_{t_{i-1}}^n}{t_i - t_{i-1}} (t - t_{i-1})$$

Par définition de la séquence d'instants t_i , nous avons $t_i - t_{i-1} = 1/n$, et donc

$$\beta_t^n = \beta_{t_{i-1}}^n + (nt - nt_{i-1}) \Delta\beta_{t_i}^n = \beta_{t_{i-1}}^n + \sqrt{n} (t - [nt]/n) \epsilon_{t_i}$$

Après avoir noter que

$$t \in [t_{i-1}, t_i[= \left[\frac{i-1}{n}, \frac{i}{n} \right[\iff nt \in [i-1, i[\iff \frac{[nt]}{n} = \frac{(i-1)}{n} = t_{i-1}$$

on obtient pour chaque $t \in [t_{i-1}, t_i[$, la décomposition suivante

$$\begin{aligned} \beta_t^n &= \beta_{t_{i-1}}^n + (\beta_t^n - \beta_{t_{i-1}}^n) \\ &= \sum_{j=1}^{i-1} \Delta\beta_{t_j}^n + \sqrt{n} (t - [nt]/n) \epsilon_{t_{([nt]+1)/n}} \\ &= \sum_{j=1}^{[nt]} \Delta\beta_{t_j}^n + \sqrt{n} (t - [nt]/n) \epsilon_{t_{([nt]+1)/n}} \end{aligned}$$

Pour poursuivre notre discussion, il convient de rappeler que les accroissements de la marche aléatoire $\Delta\beta_{t_j}^n$ forment une suite de variables aléatoires donnés par la formule

$$\Delta\beta_{t_j}^n = \epsilon_{t_i} \sqrt{\Delta t} = \frac{1}{\sqrt{n}} \epsilon_{t_i}$$

On a de plus les majorations

$$0 \leq (t - [nt]/n) \leq 1/n \downarrow 0 \quad \text{et} \quad [\beta_t^n - \beta_{t_{i-1}}^n]^2 = (\sqrt{n} (t - [nt]/n))^2 \leq 1/n \downarrow 0$$

On en conclut que

$$\beta_t^n = \frac{1}{\sqrt{n}} \times \sum_{i=1}^{[nt]} \epsilon_{t_i} + o(1) = \sqrt{\frac{[nt]}{n}} \left(\frac{1}{\sqrt{[nt]}} \times \sum_{i=1}^{[nt]} \epsilon_{t_i} \right) + o(1)$$

où $o(1)$ désigne une fonction aléatoire du paramètre n qui tend presque sûrement vers 0, lorsque n tend vers l'infini.

Le terme somme $\sum_{i=1}^{[nt]} \epsilon_{t_i}$ est formé de $[nt]$ variables aléatoires $(\epsilon_{t_i})_{i=1, \dots, [nt]}$ indépendantes et de même loi. Le théorème central de la limite nous assure que la suite

$$\frac{1}{\sqrt{[nt]}} \times \sum_{i=1}^{[nt]} \epsilon_{t_i}$$

converge faiblement, lorsque n tend vers l'infini, vers une variable aléatoire gaussienne, de moyenne nulle et de variance unité.

Il reste à noter que le facteur déterministe $\sqrt{\frac{[nt]}{n}}$ converge vers \sqrt{t} . Par des arguments probabilistes élémentaires (le lemme de Slutsky), on en conclut que

$$\beta_t^n \longrightarrow \beta_t$$

où β_t désigne une variable aléatoire gaussienne de moyenne nulle et de variance \sqrt{t} . Autrement dit, la probabilité pour que notre marcheur β_t se trouve au temps t dans un intervalle $[a, b] \subset \mathbb{R}$, est donnée par

$$\mathbb{P}(a \leq \beta_t \leq b) = \int_a^b \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} dx$$

Si l'on traduit ces probabilités de présence, par des répartitions de chaleur sur une barre rectiligne chauffée en l'origine, il s'en suit que la chaleur est plus élevée autour de la source, et de plus en plus faible lorsque l'on s'en éloigne.

Par des raisonnements analogues, il est possible vérifier que pour tout couple d'instantanés $s, t \in [0, 1]$, avec $s < t$, on a la convergence faible

$$(\beta_t^n - \beta_s^n) \longrightarrow (\beta_t - \beta_s)$$

où $(\beta_t - \beta_s)$ désigne une variable aléatoire gaussienne de moyenne nulle et de variance $\sqrt{t - s}$. On peut approfondir cette étude, et montrer que le processus gaussien $(\beta_t)_{t \in [0, 1]}$ que nous venons de construire est un processus à trajectoires continues, nulle part dérivables, à accroissements indépendants et de nature gaussienne.

6.2 La loi des grands nombres

Dans la section précédente, nous avons interprété la répartition de chaleur sur une barre rectiligne et chauffée en l'origine, en terme de la probabilité de présence d'un mouvement brownien sur la droite réelle centré en l'origine. Avec cette interprétation, la répartition de la chaleur sur un segment $[a, b]$ est donnée par la probabilité pour que β_t appartienne à cet intervalle. On peut écrire cette quantité en terme de moyenne

sur les trajectoires aléatoires conduisant le mouvement brownien à l'instant t dans l'intervalle $[a, b]$

$$\mathbb{P}(\beta_t \in [a, b]) = \mathbb{E}(1_{[a,b]}(\beta_t))$$

La loi des grands nombres nous permet d'approcher ces moyennes par des moyennes empiriques fondées sur la simulation d'un grand nombre N de copies indépendantes β_t^i de β_t . Plus précisément, la probabilité $\mathbb{P}(\beta_t \in [a, b])$ est approximativement égale à la proportion de trajectoires simulées ayant réussi à atteindre cet intervalle au temps t

$$\mathbb{E}(1_{[a,b]}(\beta_t)) \simeq \frac{1}{N} \sum_{i=1}^N 1_{[a,b]}(\beta_t^i)$$

Dans la théorie de probabilités, ces estimateurs empiriques fondés sur la simulation de variables aléatoires font partie des algorithmes d'estimation dits de Monte-Carlo. Leur convergence, lorsque le nombre de simulation augmente, est une conséquence de la loi des grands nombres. Il existe une variété de théorèmes limites, et d'estimations de probabilités d'erreurs.

Sans rentrer en profondeur dans l'étude des convergences de ces schémas, on peut se convaincre de leur qualité numérique par des arguments probabilistes très élémentaires. Commençons par noter que pour toute fonction f suffisamment régulière et bornée, les variances d'erreurs entre l'estimateur empirique $\frac{1}{N} \sum_{i=1}^N f(\beta_t^i)$, et sa moyenne $\mathbb{E}(f(\beta_t))$, sont données par la formule suivante

$$\begin{aligned} & \mathbb{E} \left(\left[\frac{1}{N} \sum_{i=1}^N f(\beta_t^i) - \mathbb{E}(f(\beta_t)) \right]^2 \right) \\ &= \mathbb{E} \left(\left[\frac{1}{N} \sum_{i=1}^N [f(\beta_t^i) - \mathbb{E}(f(\beta_t))] \right]^2 \right) \\ &= \frac{1}{N^2} \sum_{i=1}^N \mathbb{E} ([f(\beta_t^i) - \mathbb{E}(f(\beta_t))]^2) + \frac{1}{N^2} \sum_{i \neq j} \mathbb{E} ([f(\beta_t^i) - \mathbb{E}(f(\beta_t))] [f(\beta_t^j) - \mathbb{E}(f(\beta_t))]) \\ &= \frac{1}{N} \mathbb{E} ([f(\beta_t) - \mathbb{E}(f(\beta_t))]^2) \longrightarrow 0 \quad \text{lorsque } N \uparrow \infty \end{aligned}$$

La dernière assertion provient du fait que

$$\mathbb{E} ([f(\beta_t^i) - \mathbb{E}(f(\beta_t))]^2) = \mathbb{E} ([f(\beta_t) - \mathbb{E}(f(\beta_t))]^2)$$

et

$$\begin{aligned} \mathbb{E} ([f(\beta_t^i) - \mathbb{E}(f(\beta_t))] [f(\beta_t^j) - \mathbb{E}(f(\beta_t))]) &= \mathbb{E} ([f(\beta_t^i) - \mathbb{E}(f(\beta_t))]) \mathbb{E} ([f(\beta_t^j) - \mathbb{E}(f(\beta_t))]) \\ &= 0 \end{aligned}$$

Pour traduire ces estimations de moyennes en terme de probabilités de défauts ou d'erreurs, on utilise l'inégalité de Markov. Cette inégalité affirme que pour toute variable aléatoire positive X , et pour tout nombre $\epsilon > 0$, on a

$$\mathbb{P}(X > \epsilon) \leq \epsilon^{-1} \mathbb{E}(X) \quad (6.1)$$

Ce résultat résulte simplement des majorations élémentaires suivantes

$$\mathbb{E}(X) = \mathbb{E}(X \mathbf{1}_{X \leq \epsilon}) + \mathbb{E}(X \mathbf{1}_{X > \epsilon}) \geq \mathbb{E}(X \mathbf{1}_{X > \epsilon}) \geq \epsilon \mathbb{E}(\mathbf{1}_{X > \epsilon}) = \epsilon \mathbb{P}(X > \epsilon)$$

Notons que l'on a

$$\mathbb{P}(X > \epsilon) = \mathbb{P}(X^2 > \epsilon^2) \leq \epsilon^{-2} \mathbb{E}(X^2)$$

Ainsi, si l'on pose

$$X = \left| \frac{1}{N} \sum_{i=1}^N f(\beta_t^i) - \mathbb{E}(f(\beta_t)) \right|$$

dans l'inégalité précédente, on obtient une estimation de la probabilité de faire une erreur supérieure à ϵ

$$\mathbb{P} \left(\left| \frac{1}{N} \sum_{i=1}^N f(\beta_t^i) - \mathbb{E}(f(\beta_t)) \right| > \epsilon \right) \leq \frac{1}{N \epsilon^2} \mathbb{E}([f(\beta_t) - \mathbb{E}(f(\beta_t))]^2)$$

Notre situation correspond à la fonction indicatrice $f = 1_{[a,b]}$. On obtient dans ce cas

$$\begin{aligned} \mathbb{E}([f(\beta_t) - \mathbb{E}(f(\beta_t))]^2) &= \mathbb{E}(f(\beta_t)^2) - \mathbb{E}(f(\beta_t))^2 \\ &= \mathbb{E}(1_{[a,b]}(\beta_t)) - \mathbb{E}(1_{[a,b]}(\beta_t))^2 \\ &= \mathbb{P}(\beta_t \in [a, b]) (1 - \mathbb{P}(\beta_t \in [a, b])) \leq \frac{1}{4} \end{aligned}$$

En injectant cette majoration dans l'estimation précédente, on en conclut que

$$\mathbb{P} \left(\left| \frac{1}{N} \sum_{i=1}^N 1_{[a,b]}(\beta_t^i) - \mathbb{P}(\beta_t \in [a, b]) \right| > \epsilon \right) \leq \frac{1}{4N\epsilon^2}$$

6.3 Marches aléatoires

D'après les constructions probabilistes précédentes, on peut simuler des mouvements browniens de deux façons distinctes. La première revient à considérer la marche aléatoire simple

$$(\beta_{t_0}^n, \beta_{t_1}^n, \dots, \beta_{t_n}^n)$$

décrite au début de la section 6.1, sur une subdivision suffisamment fine $(t_i)_{i=0, \dots, n}$ de l'intervalle $[0, 1]$.

La seconde, est fondé sur les théorèmes limites gaussiens décrits à la fin de la section 6.1. Supposons que l'on souhaite simuler un mouvement brownien, centré en l'origine $\beta_{t_0} = 0$, sur une suite de pas de temps $(t_i)_{i=1, \dots, n}$

$$(\beta_{t_1}, \dots, \beta_{t_n})$$

En terme d'accroissements, cette séquence peut s'exprimer sous la forme suivante :

$$\begin{aligned} (\beta_{t_1}, \dots, \beta_{t_n}) &= ([\beta_{t_1} - \beta_{t_0}], \beta_{t_1} + [\beta_{t_2} - \beta_{t_1}], \dots, \beta_{t_{n-1}} + [\beta_{t_n} - \beta_{t_{n-1}}]) \\ &= \left([\beta_{t_1} - \beta_{t_0}], \sum_{i=1}^2 [\beta_{t_i} - \beta_{t_{i-1}}], \dots, \sum_{i=1}^n [\beta_{t_i} - \beta_{t_{i-1}}] \right) \end{aligned}$$

On sait de plus que les accroissements

$$([\beta_{t_1} - \beta_{t_0}], \dots, [\beta_{t_n} - \beta_{t_{n-1}}])$$

forment une suite de n variables aléatoires gaussiennes de moyenne nulle, et de variances respectives données par les composantes du vecteur

$$(\sqrt{t_1 - t_0}, \dots, \sqrt{t_n - t_{n-1}})$$

Ces variables indépendantes peuvent être générées par un algorithme de simulation de gaussiennes, tel l'algorithme de Box-Muller décrit à la page 46.

Sur un nombre pair de pas de temps, $n = 2m$, l'algorithme de Box-Muller est fondé sur la simulation de $2m$ variables $(U_k, V_k)_{k=1, \dots, m}$ uniformes, et indépendantes sur $[0, 1]$. Pour simuler les $2m$ accroissements gaussiens, ils reste à poser pour indice $k = 1, \dots, m$

$$\begin{aligned} (\beta_{t_{2k-1}} - \beta_{t_{2k-2}}) &= \sqrt{-2(t_{2k-1} - t_{2k-2}) \log U_k} \cos(2\pi V_k) \\ (\beta_{t_{2k}} - \beta_{t_{2k-1}}) &= \sqrt{-2(t_{2k} - t_{2k-1}) \log U_k} \sin(2\pi V_k) \end{aligned}$$

6.4 L'équation de la chaleur

Dans les précédentes questions, nous avons montré que la répartition de la chaleur sur une barre rectiligne était donnée par la distribution gaussienne d'un mouvement brownien β_t sur \mathbb{R} . Plus formellement, nous avons pour toute fonction régulière et bornée

$$\mathbb{E}(f(\beta_t)) = \int_{\mathbb{R}} f(x) \underbrace{p_t(x) dx}_{\mathbb{P}(\beta_t \in dx)}$$

avec la densité gaussienne

$$p_t(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$$

Cette fonction spatio-temporelle $p : (t, x) \in]0, \infty[\times \mathbb{R} \mapsto p_t(x) \in]0, \infty[$ est la solution d'une équation aux dérivées partielles, appelée l'équation de la chaleur

$$\frac{\partial p_t}{\partial t} = \frac{1}{2} \frac{\partial^2 p_t}{\partial x^2}$$

Pour vérifier cette formule, il suffit simplement de dériver la fonction par rapport aux différentes variables. La dérivée par rapport au paramètre temporel est clairement donnée par

$$\frac{\partial p_t}{\partial t}(x) = \frac{1}{\sqrt{2\pi t}} \left(-\frac{1}{2t} + \frac{x^2}{2t^2} \right) e^{-\frac{x^2}{2t}} = \frac{1}{2} \left(\frac{x^2}{t^2} - \frac{1}{t} \right) p_t(x)$$

La dérivée seconde par rapport à la coordonnée spatiale est aussi donnée par la même formule :

$$\frac{1}{2} \frac{\partial^2 p_t}{\partial x^2}(x) = -\frac{1}{2} \frac{\partial}{\partial x} \left(\frac{x}{t} p_t(x) \right) = \frac{1}{2} \left(\frac{x^2}{t^2} - \frac{1}{t} \right) p_t(x)$$

6.5 Une formulation faible

La dérivation de l'équation de la chaleur décrite ci-dessus est simplement basée sur un jeu de dérivations de fonctions, sans aucun fondement physique. Revenons quelques secondes à notre interprétation de la répartition de la chaleur en terme de probabilité de présence d'une particule brownienne. Dans cette interprétation physique, les variations temporelles des trajectoires moyennes

$$t \in]0, \infty[\mapsto \mathbb{E}(f(\beta_t))$$

sont données par les variations temporelles de la densité de probabilité gaussienne

$$\frac{\partial}{\partial t} \mathbb{E}(f(\beta_t)) = \frac{\partial}{\partial t} \int_{\mathbb{R}} f(x) p_t(x) dx = \int_{\mathbb{R}} f(x) \frac{\partial p_t}{\partial t}(x) dx$$

Essayons de montrer que pour toute fonction continue et bornée f , deux fois dérivable, à dérivés continus et bornés, nous avons

$$\int_{\mathbb{R}} f(x) \frac{\partial p_t}{\partial t}(x) dx = \int_{\mathbb{R}} f(x) \frac{\partial^2 p_t}{\partial x^2}(x) dx \quad (6.2)$$

Cette classe de fonctions est suffisamment vaste pour remplir en un certain sens l'ensemble des fonctions continues et bornées, dual des mesures signées sur \mathbb{R} . Cette propriété topologique, nous permet d'identifier deux mesures signées dès que leurs intégrales sur ces fonctions test f continues et bornées coïncident. La propriété recherchée (6.2) nous permettra donc de conclure à l'égalité au sens faible des densités

$$\frac{\partial p_t}{\partial t}(x) = \frac{\partial^2 p_t}{\partial x^2}(x)$$

On parle dans ce cas de solution faible de l'équation de la chaleur.

Après ce léger aparté topologique, examinons de plus près les variations temporelles des moyennes $m(t) = \mathbb{E}(f(\beta_t))$. Par définition de la dérivée temporelle d'une fonction, et avec quelques abus de notations, nous avons

$$\frac{\partial}{\partial t} \mathbb{E}(f(\beta_t)) \simeq \frac{m(t + \Delta t) - m(t)}{\Delta t} = \frac{\mathbb{E}[f(\beta_t + \Delta\beta_t) - f(\beta_t)]}{\Delta t}$$

avec

$$\Delta\beta_t = (\beta_{t+\Delta t} - \beta_t) \quad \text{et un pas de temps } \Delta t \simeq 0$$

Le développement de Taylor au second ordre de la fonction f , autour du point β_t nous conduit à la formule

$$f(\beta_t + \Delta\beta_t) - f(\beta_t) = \frac{\partial f}{\partial x}(\beta_t) \Delta\beta_t + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}(\beta_t) [\Delta\beta_t]^2 + O((\Delta\beta_t)^3)$$

où $O((\Delta\beta_t)^3)$ désigne une fonction aléatoire telle que

$$\mathbb{E}(O((\Delta\beta_t)^3)) \leq \text{Cte} \mathbb{E}((\sqrt{\Delta\beta_t})^3) \leq \text{Cte} \Delta t \sqrt{\Delta t}$$

pour une certaine constante finie $\text{Cte} < \infty$ ne dépendant que de la fonction f . Après avoir observé que

$$\mathbb{E}\left(\frac{\partial f}{\partial x}(\beta_t) \Delta\beta_t \mid \beta_t\right) = 0 \quad \text{et} \quad \mathbb{E}\left(\frac{\partial^2 f}{\partial x^2}(\beta_t) [\Delta\beta_t]^2 \mid \beta_t\right) = \frac{\partial^2 f}{\partial x^2}(\beta_t) \times \Delta t$$

on en conclut que

$$\frac{1}{\Delta t} \mathbb{E}[f(\beta_t + \Delta t) - f(\beta_t)] = \mathbb{E}\left(\frac{1}{2} \frac{\partial^2 f}{\partial x^2}(\beta_t)\right) + o(1)$$

où $o(1)$ désigne une fonction déterministe convergeant vers 0, lorsque Δt tend vers 0. En faisant tendre le pas de temps vers 0, nous avons montré l'identité suivante

$$\frac{\partial}{\partial t} \mathbb{E}(f(\beta_t)) = \int_{\mathbb{R}} f(x) \frac{\partial p_t}{\partial t}(x) dx = \frac{1}{2} \int_{\mathbb{R}} \frac{\partial^2 f}{\partial x^2}(x) p_t(x) dx$$

Il nous reste donc à vérifier que l'on a

$$\int_{\mathbb{R}} \frac{\partial^2 f}{\partial x^2}(x) p_t(x) dx = \int_{\mathbb{R}} f(x) \frac{\partial^2 p_t}{\partial x^2}(x) dx$$

On démontre cette ultime propriété à l'aide de deux intégrations par parties successives

$$\int_{\mathbb{R}} \frac{\partial^2 f}{\partial x^2}(x) p_t(x) dx = \left[p_t \frac{\partial f}{\partial x} \right]_{-\infty}^{\infty} - \int_{\mathbb{R}} \frac{\partial f}{\partial x}(x) \frac{\partial p_t}{\partial x}(x) dx = - \int_{\mathbb{R}} \frac{\partial f}{\partial x}(x) \frac{\partial p_t}{\partial x}(x) dx$$

et

$$\int_{\mathbb{R}} \frac{\partial f}{\partial x}(x) \frac{\partial p_t}{\partial x}(x) dx = \left[\frac{\partial p_t}{\partial x} f \right]_{-\infty}^{\infty} - \int_{\mathbb{R}} f(x) \frac{\partial^2 p_t}{\partial x^2}(x) dx = - \int_{\mathbb{R}} f(x) \frac{\partial^2 p_t}{\partial x^2}(x) dx$$

Chapitre 7

Dynamiques de population avec branchements

7.1 Processus de branchements spatio-temporels

Dans la section 3.3.6 nous avons étudié des processus de branchement élémentaires dans lesquels chaque individu donne naissance à un certain nombre d'enfants. Ces modèles simplifiés sont assez éloignés de la réalité. Tout d'abord ils ne tiennent pas compte des explorations des individus dans l'espace. De plus, les taux de branchement sont totalement indépendants des régions plus ou moins accueillantes dans lesquelles se trouvent les populations.

Dans ce qui suit, nous allons essayer de raffiner ces modèles stochastiques pour rendre compte de ces deux paramètres, et essayer de coller au mieux à la réalité scientifique. Nous conviendrons que les individus évoluent à chaque instant n dans un espace d'état E_n . Les populations d'individus seront représentées par des vecteurs de E_n^p . Le paramètre entier $p \geq 0$ correspond à la taille des populations. Lorsque $p = 0$, on conviendra que l'espace d'état se réduit à un état cerceuil, ou cimetière $E_n^0 = \{c\}$. L'espace d'état du système est donc donné par l'ensemble

$$S_n = \cup_{p \geq 0} E_n^p$$

La dynamique d'exploration de chaque individu est associée à des transitions markoviennes M_n de E_{n-1} vers E_n . Les mécanismes de branchements dépendent de fonctions potentiel $G_n : E_N \rightarrow [0, \infty)$ représentant les différents degrés de "fertilité" du milieu. Plus précisément, un individu sur un site $x_n \in E_n$ donnera naissance à un nombre aléatoire d'enfants $g_n(x_n)$ avec

$$\mathbb{E}(g_n(x_n)) = G_n(x_n)$$

Pour illustrer ce modèle, on peut supposer que les populations évoluent dans le réseau du plan $E_n = \mathbb{Z}^2$ selon une marche aléatoire simple, et les variables aléatoires de

branchement sont de Bernoulli

$$\begin{aligned}\mathbb{P}(g_n(x_n) = [G_n(x)] + 1) &= G_n(x_n) - [G_n(x_n)] \\ &= 1 - \mathbb{P}(g_n(x_n) = [G_n(x)])\end{aligned}$$

où $[a]$ désigne la partie entière d'un nombre $a \in [0, \infty)$. On peut aussi choisir des branchements poissonniens

$$\forall m \in \mathbb{N} \quad \mathbb{P}(g_n(x_n) = m) = \frac{G_n(x_n)^m}{m!} \exp(-G_n(x_n))$$

Pour définir plus formellement notre dynamique de population, il convient d'introduire une suite de variable $(g_n^i(x_n))_{i \geq 1}$ indépendantes, et de même loi que $g_n(x_n)$. On supposera de plus que les variables sur des sites distincts sont indépendantes.

Nous sommes enfin en mesure de construire récursivement le modèle. Initialement, la population \mathcal{X}_0 est formé d'un seul individu \mathcal{X}_0^1 dans l'état $x_0 \in E_0$

$$\mathcal{X}_0 = \mathcal{X}_0^1 = x_0 \in E_0^{p_0} = E_0^1$$

Cet individu donne naissance à

$$\widehat{p}_0 = g_0^1(\mathcal{X}_0^1)$$

enfants que l'on note

$$\widehat{\mathcal{X}}_0 = (\widehat{\mathcal{X}}_0^1, \dots, \widehat{\mathcal{X}}_0^{\widehat{p}_0}) \in E_0^{\widehat{p}_0}$$

Chacun d'entre eux explore aléatoirement l'espace $\widehat{\mathcal{X}}_0^i \rightsquigarrow \mathcal{X}_1^i$ selon la transition M_1 . Cette transition revient à simuler \widehat{p}_0 variables aléatoires \mathcal{X}_1^i de loi $M_1(\widehat{\mathcal{X}}_0^i, dx_1)$. Lorsque cette étape d'exploration est terminée, nous avons une population formée de

$$p_1 = \widehat{p}_0$$

individus. Cette transition peut s'exprimer de façon synthétique par la formule suivante :

$$\widehat{\mathcal{X}}_0 = (\widehat{\mathcal{X}}_0^1, \dots, \widehat{\mathcal{X}}_0^{\widehat{p}_0}) \rightsquigarrow \mathcal{X}_1 = (\mathcal{X}_1^1, \dots, \mathcal{X}_1^{p_1})$$

Durant l'étape de branchement suivante chaque individu \mathcal{X}_1^i donne naissance à $g_1^i(\mathcal{X}_1^i)$ enfants. À la fin de ce processus nous avons une population formée de \widehat{p}_1 individus

$$\widehat{\mathcal{X}}_1 = (\widehat{\mathcal{X}}_1^1, \dots, \widehat{\mathcal{X}}_1^{\widehat{p}_1}) \in E_1^{\widehat{p}_1} \quad \text{avec} \quad \widehat{p}_1 = \sum_{i=1}^{p_1} g_1^i(\mathcal{X}_1^i)$$

Chacun de ces individus explore aléatoirement l'espace E_2 selon la transition M_2 , etc.

$$\mathcal{X}_n = (\mathcal{X}_n^i)_{1 \leq i \leq p_n} \xrightarrow{\text{branchement}} \widehat{\mathcal{X}}_n = (\widehat{\mathcal{X}}_n^i)_{1 \leq i \leq \widehat{p}_n} \xrightarrow{\text{exploration}} \mathcal{X}_{n+1}$$

Si le système meurt au bout d'un certain temps n , nous avons $\widehat{p}_n = 0$. Dans ce cas, on pose

$$\widehat{\mathcal{X}}_p = \mathcal{X}_{p+1} = 0$$

pour tout les instants suivants $p \geq n$.

Exercice 7.1.1 1. Pour toutes fonctions bornées f_n sur E_n , exprimer les variables aléatoires $\sum_{i=1}^{\hat{p}_n} f_n(\hat{\mathcal{X}}_n^i)$ en fonction de p_n , \mathcal{X}_n , et des variables g_n^i (on utilisera la convention $\sum_{\emptyset} = 0$, lorsque la population est éteinte).

2. On considère les mesures empiriques aléatoires

$$s(x_n) =_{\text{def}} \sum_{i=1}^p \delta_{x_n^i} \quad \text{pour chaque } x_n = (x_n^i)_{1 \leq i \leq d} \in E_n^p$$

Pour toute fonction bornée f_n sur E_n , calculer les moyennes conditionnelles

$$\mathbb{E}(s(\hat{\mathcal{X}}_n)(f_n) \mid \mathcal{X}_n) \quad \text{et} \quad \mathbb{E}(s(\mathcal{X}_{n+1})(f_{n+1}) \mid \hat{\mathcal{X}}_n)$$

En déduire que

$$\mathbb{E}(s(\mathcal{X}_{n+1})(f_{n+1}) \mid \mathcal{X}_n) = s(\mathcal{X}_n)(G_n M_{n+1}(f_{n+1}))$$

3. Vérifier que les premiers moments des dynamiques de population sont donnés par la formule suivante

$$\mathbb{E}(s(\mathcal{X}_{n+1})(f_{n+1})) = \mathbb{E}_{x_0}(f_{n+1}(X_{n+1})) \prod_{k=0}^n G_k(X_k)$$

7.2 Algorithme génétique

En termes biologiques, les algorithmes génétiques représentent les dynamiques de population à nombre constant d'individus. Ces individus explorent des régions plus ou moins accueillantes, selon des mécanismes de mutation et de sélection. Les individus meurent ou donnent naissance à des enfants suivant la qualité de leur milieu, mesurée en terme d'une fonction potentiel.

Dans la section ??, nous montrerons que ces modèles génétiques sont une simple expression "microscopique" des modèles d'évolution-absorption étudiés dans la section 4.4.

Leurs interprétations sont à l'image des différents domaines d'applications décrits à la page 87.

Lorsque la fonction potentiel représente des **niveaux de sécurité**, tels des risques de collisions dans des aéroports, l'étape de sélection permet par exemple de choisir les configurations les moins sécurisées. L'étape de mutation consiste alors à explorer plus en profondeur les risques de collisions suivantes. Dans ce contexte, l'agorithme génétique peut s'interpréter comme des séquences d'évolutions probables conduisant à des collisions d'avion.

En pharmacologie, la fonction potentiel peut représenter les différents niveaux de leukocytes dans un organisme vivant. Dans ce contexte, l'étape de mutation représente les différentes possibilités d'évolution de ces niveaux. La sélection permet d'évaluer les chutes possibles, et les entrées dans des niveaux mortels. L'algorithme génétique correspondant permet de décrire les histoires possibles conduisant aux décès d'un organisme.

Enfin, dans **des modèles économiques**, les potentiels peuvent représenter des valeurs de portefeuilles. L'étape de mutation consiste à décrire les évolutions envisageables de ces quantités partir de niveaux donnés. L'étape de sélection permet de choisir ces propositions suivant leurs tendances à la hausse ou à la baisse. Le modèles génétiques correspondent alors à des évolutions de portefeuilles vers la ruine ou la fortune dans des milieux financiers.

Enfin, dans **les modèles d'apprentissage** développés en intelligence artificielle, les potentiels représentent la qualité d'une proposition émise. L'étape de mutation consiste tout d'abord à élaborer des séquences de propositions envisageables, plus ou moins bonnes. L'étape de sélection permet ensuite d'affiner le raisonnement, en choisissant les propositions les mieux adaptées pour compléter une action donnée, ou pour reconstruire une information partiellement observée.

En termes mathématiques, un algorithme génétique est une chaîne de Markov $X_n = (X_n^i)_{1 \leq i \leq N}$ sur un espace produit E^N . Pour fixer les idées on pourra supposer que l'espace d'état est donné par $E = \mathbb{Z}^d$ ou $E = \mathbb{R}^d$, avec $d \in \mathbb{N} - \{0\}$. Chacune des composantes X_n^i représente la position de l'individu de label i , avec $i = 1, \dots, N$.

L'évolution de cette chaîne se décompose en deux mécanismes bien distincts. Le premier correspond à une sélection des individus, selon un certain critère de qualité. Le second est une exploration pure de l'espace des états. En biologie, cette étape est souvent appelé mutation, par référence au fait que les codes génétiques des individus changent au cours du temps.

$$X_n = (X_n^i)_{1 \leq i \leq N} \xrightarrow{\text{sélection}} \hat{X}_n = (\hat{X}_n^i)_{1 \leq i \leq N} \xrightarrow{\text{mutation}} X_{n+1} = (X_{n+1}^i)_{1 \leq i \leq N}$$

7.2.1 Sélection/Adaptation

Pour décrire l'étape de sélection, on se fixe une fonction potentiel, strictement positive, G sur l'espace E . Une fois connue la configuration de la chaîne $X_n = (X_n^i)_{1 \leq i \leq N}$ au temps n , l'étape de sélection consiste à simuler N v.a. $(\widehat{X}_n^i)_{1 \leq i \leq N}$ indépendantes de même loi

$$\sum_{i=1}^N \frac{G(X_n^i)}{\sum_{j=1}^N G(X_n^j)} \delta_{X_n^i}$$

Autrement dit chaque v.a. \widehat{X}_n^i choisit l'une des valeurs X_n^i , avec la probabilité $\frac{G(X_n^i)}{\sum_{j=1}^N G(X_n^j)}$. On remarquera que ce procédé de sélection peut aussi s'interpréter comme **un mécanisme de naissances et morts**. Dans cette interprétation, les individus X_n^i disparaissent, ou donnent naissance à un certain nombre de copies.

Il existe divers variantes pour sélectionner les individus les mieux adaptés au potentiel G . Dans le cas où le potentiel G est à valeurs dans $[0, 1]$, il est bien plus naturel "d'accepter" chaque individu X_n^i avec une probabilité $G(X_n^i)$, et de le remplacer (avec une probabilité $[1 - G(X_n^i)]$) par un individu choisi avec la loi discrète

$$\sum_{i=1}^N \frac{G(X_n^i)}{\sum_{j=1}^N G(X_n^j)} \delta_{X_n^i}$$

Plus formellement, ce mécanisme de sélection est équivalent à poser pour chaque $i = 1, \dots, N$

$$\widehat{X}_n^i = \begin{cases} X_n^i & \text{avec probabilité } G(X_n^i) \\ \tilde{X}_n^i & \text{avec probabilité } 1 - G(X_n^i) \end{cases}$$

où \tilde{X}_n^i désigne une v.a. de loi $\sum_{j=1}^N \frac{G(X_n^j)}{\sum_{k=1}^N G(X_n^k)} \delta_{X_n^j}$.

Pour des fonctions potentiel pouvant s'annuler sur certaines régions de l'espace, il est possible que tous les individus aient des potentiels nul. Dans cette situation, l'algorithme est stoppé.

7.2.2 Mutation/Exploration

Durant la phase de mutation, les individus sélectionnés explorent l'espace indépendamment les uns des autres, selon des transitions de probabilités élémentaires $M(x, y)$. Autrement dit, nous avons

$$\widehat{X}_n^i \rightsquigarrow X_{n+1}^i$$

où X_{n+1}^i désigne une v.a. de loi $M(\widehat{X}_n^i, \cdot)$.

Plus formellement, nous avons

$$\mathbb{P}(X_{n+1}^1 \in dx^1, \dots, X_{n+1}^N \in dx^N \mid \widehat{X}_n^1, \dots, \widehat{X}_n^N) = M(\widehat{X}_n^1, dx^1) \dots M(\widehat{X}_n^N, dx^N)$$

À titre d'exemple, si $M(x, y)$ désigne la matrice de transition d'une marche aléatoire sur \mathbb{Z}^d , on peut réaliser dynamiquement ces explorations locales en posant

$$X_{n+1}^i = \widehat{X}_{n-1}^i + U_{n+1}^i$$

où U_{n+1}^i désigne une suite de v.a. indépendantes de même loi p , sur l'ensemble des vecteurs unitaires directionnels $\mathcal{U} = \{u \in \mathbb{Z}^d : |u| = 1\}$.

Un exemple schématique d'évolution de $N = 4$ individus est représenté dans la figure suivante. Les nombres entiers entre parenthèse correspondent au nombre d'individus sur le site en question, après l'étape de sélection.

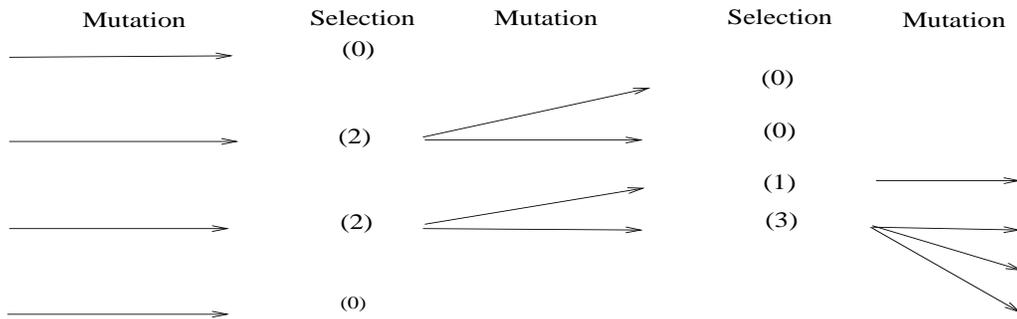


FIG. 7.1 – Algorithme génétique ($N = 4$)

Exercice 7.2.1 Décrire mathématiquement, et schématiquement, l'algorithme génétique sur \mathbb{Z} associé à la fonction de potentiel indicatrice $G(x) = 1_{[-L,L]}$, avec $L \geq 1$. On conviendra que les mutations sont données par les transitions d'une marche aléatoire sur \mathbb{Z} , et l'on initialisera les individus en l'origine.

Exercice 7.2.2 Décrire l'algorithme génétique sur \mathbb{R} associé à des mutations gaussiennes

$$M(x, dy) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(y-x)^2 \right\} dy$$

et un potentiel quadratique centré autour d'un point $a \in \mathbb{R}$

$$G(x) = \exp \left\{ -\frac{1}{2}(x - a)^2 \right\}$$

7.3 Modèles d'arbres généalogiques

7.3.1 Modèles non homogènes

L'algorithme génétique décrit dans la section précédente peut être étendu de façon naturelle à des espaces d'états E_n dépendants du paramètre temporel $n \in \mathbb{N}$. Dans ce contexte, les individus X_n^i vivent à chaque instant n dans l'espace E_n . Les sélections s'effectuent dans ces mêmes espaces, tandis que les mutations s'expriment comme des passages aléatoires d'un état de E_n vers un nouvel état dans E_{n+1} .

Plus formellement, les populations d'individus sont données par des N -uplets

$$X_n = (X_n^i)_{1 \leq i \leq N} \in E_n^N \quad \text{et} \quad \hat{X}_n = (\hat{X}_n^i)_{1 \leq i \leq N} \in E_n^N$$

où E_n^N désigne l'espace produit $(E_n \times \dots \times E_n)$, avec N termes.

Dans ce contexte, les transitions de sélection/mutations s'expriment entre des états non homogènes :

$$X_n \in E_n^N \xrightarrow{\text{sélection}} \hat{X}_n \in E_n^N \xrightarrow{\text{mutation}} X_{n+1} \in E_{n+1}^N$$

Supposons que les sélections à chaque instant n soient aussi dictées par des potentiels non homogènes

$$G_n : x_n \in E_n \mapsto G_n(x_n) \in [0, 1]$$

Dans ce cas, le mécanisme de sélection dans E_n , s'exprime sous la forme suivante :

$$\hat{X}_n^i = \begin{cases} X_n^i & \text{avec proba } G_n(X_n^i) \\ \tilde{X}_n^i & \text{avec proba } 1 - G_n(X_n^i) \end{cases}$$

où \tilde{X}_n^i désigne une v.a. de loi

$$\sum_{j=1}^N \frac{G_n(X_n^j)}{\sum_{k=1}^N G_n(X_n^k)} \delta_{X_n^j}$$

Les individus ainsi sélectionnés \widehat{X}_n^i vivent dans l'espace E_n .

Durant la mutation, ces individus passent de l'état E_n vers un nouvel état E_{n+1} , selon des transitions de probabilités $M_{n+1}(x_n, dx_{n+1})$ de E_n vers E_{n+1} . Autrement dit, nous avons

$$\widehat{X}_n^i (\in E_n) \rightsquigarrow X_{n+1}^i (\in E_{n+1})$$

où X_{n+1}^i désigne une v.a. de loi $M_{n+1}(\widehat{X}_{n-1}^i, \cdot)$. Ces nouveaux individus sont alors sélectionnés en fonction d'un potentiel G_{n+1} sur E_{n+1} , puis ils mutent de E_{n+1} vers E_{n+2} selon une transition markovienne de E_{n+1} vers E_{n+2} , etc.

7.3.2 Modèles trajectoriels

Dans ce qui précède, nous n'avons pas précisé la nature des espaces E_n , ni a fortiori celle des transitions $M_n(x_{n-1}, dx_n)$. Tout ceci semble donc bien abstrait! Revenons donc sur terre en supposant que les espaces E_n sont donnés par des espaces produits

$$E_n = \underbrace{E' \times \dots \times E'}_{(n+1)\text{-fois}}$$

où E' désigne un ensemble quelconque, suffisamment réguliers. Pour fixer les idées, on pourra supposer que $E = \mathbb{Z}^d$, ou $E = \mathbb{R}^d$. On conviendra que les points de E_n sont donnés par des $(n+1)$ -uplets représentant des trajectoires de longueur n dans l'espace E'

$$x_n = (x'_0, \dots, x'_n) \in E_n = (E')^{n+1}$$

L'algorithme génétique précédent est, à chaque étape $n \in \mathbb{N}$, formé de N variables aléatoires trajectorielles, à valeurs dans E_n , que l'on notera

$$X_n^i = (X_{0,n}^i, X_{1,n}^i, \dots, X_{n,n}^i) \in E_n = (E')^{n+1}, \quad 1 \leq i \leq N.$$

et

$$\widehat{X}_n^i = (\widehat{X}_{0,n}^i, \widehat{X}_{1,n}^i, \dots, \widehat{X}_{n,n}^i) \in E_n = (E')^{n+1}, \quad 1 \leq i \leq N.$$

Chacun des individus X_n^i , et \widehat{X}_n^i représente un chemin dans E' de l'origine jusqu'au temps n .

Il est important de souligner que la population initiale

$$X_0^i = X_0^i \in E_0 = E' \quad \text{avec} \quad 1 \leq i \leq N$$

est tout simplement formée de N variables aléatoires à valeurs dans E' .

Sélection trajectorielle :

Dans notre cadre trajectorien, le mécanisme de sélection, peut s'interpréter comme une sélection de trajectoires, en fonction des différents potentiels

$$G_n(X_n^i) = G_n(X_{0,n}^i, X_{1,n}^i, \dots, X_{n,n}^i)$$

L'expression de cette transition reste inchangée. Nous avons à nouveau

$$\widehat{X}_n^i = \begin{cases} X_n^i & \text{avec proba } G_n(X_n^i) \\ \widetilde{X}_n^i & \text{avec proba } 1 - G_n(X_n^i) \end{cases}$$

où $\widetilde{X}_n^i = (\widehat{X}_{0,n}^i, \widehat{X}_{1,n}^i, \dots, \widehat{X}_{n,n}^i)$ désigne une v.a. de loi

$$\sum_{j=1}^N \frac{G_n(X_n^j)}{\sum_{k=1}^N G_n(X_n^k)} \delta_{X_n^j} = \sum_{j=1}^N \frac{G_n(X_{0,n}^j, \dots, X_{n,n}^j)}{\sum_{k=1}^N G_n(X_{0,n}^k, \dots, X_{n,n}^k)} \delta_{(X_{0,n}^j, \dots, X_{n,n}^j)}$$

Les trajectoires ainsi sélectionnés \widehat{X}_n^i vivent désormais dans l'espace de chemins $E_n = (E')^{n+1}$.

Mutation trajectorielle :

L'étape de mutation dépend uniquement de la nature des transitions $M_{n+1}(x_n, dx_{n+1})$ de E_n dans E_{n+1} . Supposons que ces dernières correspondent aux transitions de probabilités d'un processus historique associé à une évolution markovienne sur E' .

Dans cette situation, on rappelle que ces transitions $M_{n+1}(x_n, dy_{n+1})$ s'expriment sous la forme suivante :

$$M_{n+1}((x'_0, \dots, x'_n), d(y'_0, \dots, y'_{n+1})) = \delta_{(x'_0, \dots, x'_n)}(d(y'_0, \dots, y'_n)) M'_{n+1}(y'_n, dy'_{n+1})$$

M'_{n+1} est une transition de Markov de E' vers lui même. Autrement dit, simuler une variable aléatoire trajectorielle de loi

$$M_{n+1}((x'_0, \dots, x'_n), d(y'_0, \dots, y'_{n+1}))$$

revient à conserver tout d'abord le segment de trajectoire

$$x_n = (x'_0, \dots, x'_n)$$

On lui adjoint ensuite une extension élémentaire aléatoire de loi $M'_{n+1}(x'_n, dy'_{n+1})$

$$x'_n \rightsquigarrow y'_{n+1}$$

Durant l'étape de mutation correspondante

$$\widehat{X}_n^i = \underbrace{\left(\widehat{X}_{0,n}^i, \dots, \widehat{X}_{n,n}^i \right)}_{\downarrow} \in E_n^N$$

$$X_{n+1}^i = \left(\underbrace{\left(X_{0,n+1}^i, \dots, X_{n,n+1}^i \right)}_{\left(\widehat{X}_n^i \right)}, X_{n+1,n+1}^i \right) \in E_{n+1}^N$$

chaque chemin s'étend selon un déplacement élémentaire de loi M'_{n+1} , c'est à dire

$$X_{n+1}^i = \left(\underbrace{\left[X_{0,n+1}^i, \dots, X_{n,n+1}^i \right]}_{\parallel}, X_{n+1,n+1}^i \right) \in E_{n+1} = (E')^{n+2}$$

$$= \left(\underbrace{\left[\widehat{X}_{0,n}^i, \dots, \widehat{X}_{n,n}^i \right]}_{\left(\widehat{X}_n^i \right)}, X_{n+1,n+1}^i \right) = \left(\widehat{X}_n^i, X_{n+1,n+1}^i \right)$$

avec une variable aléatoire $X_{n+1,n+1}^i$ de loi $M'_{n+1}(\widehat{X}_{n,n}^i, \cdot)$.

À titre d'exemple, dans le cas où M'_{n+1} désigne la matrice de transition d'une marche aléatoire sur $E' = \mathbb{Z}^d$, on peut réaliser dynamiquement ces explorations locales en posant

$$X_{n+1,n+1}^i = \widehat{X}_{n,n}^i + U_{n+1}^i$$

où U_{n+1}^i désigne une suite de v.a. indépendantes de même loi p , sur l'ensemble des vecteurs unitaires directionnels

$$\mathcal{U} = \{u \in \mathbb{Z}^d : |u| = 1\}$$

Lignes ancestrales :

L'algorithme génétique trajectorien décrit ci-dessus correspond bien à une évolution d'arbres généalogiques.

Dans cette interprétation, chaque trajectoire

$$X_n^i = (X_{0,n}^i, X_{1,n}^i, \dots, X_{n,n}^i) \in E_n = (E')^{n+1}$$

représente la ligne ancestrale de l'individu courant $X_{n,n}^i$ à la date n . Les coordonnées $X_{p,n}^i$, avec $0 \leq p \leq n$, correspondent aux différents ancêtres de cet individu, à chaque niveau temporel $0 \leq p \leq n$.

Dans ce contexte, $X_{0,n}^i$ représente l'ancêtre initial, $X_{1,n}^i$ sa première descendance, $X_{2,n}^i$ sa seconde, etc. Dans le modèle général que nous avons développé, les disparitions et les sélections de lignes ancestrales dépendent de la qualité de la trajectoire de descendance complète. Les différents degrés d'adaptation d'une lignée complète

$$x_n = (x'_0, \dots, x'_n) \in E_n$$

sont mesurés par une fonction potentiel $G_n(x'_0, \dots, x'_n)$. Lorsque ces fonctions ne dépendent que des composantes terminales, c'est à dire lorsque l'on a

$$G_n(x'_0, \dots, x'_n) = G'_n(x'_n),$$

la sélection des lignes ancestrale ne dépend que de la qualité d'adaptation du **dernier descendant**. Dans cette situation, nous laissons le soin au lecteur de se convaincre que ce modèle trajectorien correspond à l'arbre généalogique d'une population d'individus explorant l'espace E' selon M'_n , et s'adaptant en fonction des potentiels de sélection G'_n .

7.4 Chaînes renforcées

Ces modèles de renforcement sont associés à des évolutions aléatoires où chaque transition dépend de la mesure d'occupation des sites visités dans le passé. Ces processus permettent de représenter des stratégies humaines d'exploration de sites comme des rues, des magasins, des restaurants, ou tout autre endroits d'une ville. Dans ce contexte, un site donné est d'autant plus attractif s'il répond à un certain critère de qualité, ou lorsqu'il a été déjà visité de nombreuses fois par le passé!

Le critère de qualité est représenté par une fonction potentiel $G : E \rightarrow [0, 1]$, sur un espace d'états E . Les valeurs de G en un site x sont d'autant plus grandes que le site est attrayant. Ainsi, notre explorateur se trouvant en $X_n = x$, au temps n , choisit d'y rester avec une probabilité $G(x)$, soit préfère retourner vers l'un des sites précédemment visité X_0, \dots, X_{n-1} . Dans cette situation, il choisit un nouveau site avec une probabilité

$$\sum_{p=0}^{n-1} \frac{G(X_p)}{\sum_{q=0}^{n-1} G(X_q)} \delta_{X_p}$$

Cette sélection aléatoire peut être vue comme une transition élémentaire

$$X_n \rightsquigarrow \widehat{X}_n$$

de probabilités de transitions

$$G(X_n) \delta_{X_n} + (1 - G(X_n)) \sum_{p=0}^{n-1} \frac{G(X_p)}{\sum_{q=0}^{n-1} G(X_q)} \delta_{X_p}$$

Partant du site sélectionné \widehat{X}_n , l'individu effectue une nouvelle exploration de la région $\widehat{X}_n \rightsquigarrow X_{n+1}$, selon une transition de probabilité M_{n+1} . Autrement dit, nous avons

$$\mathbb{P}(X_{n+1} \in dy \mid \widehat{X}_n = x) = M_{n+1}(x, dy)$$

Cette transition M_n peut, par exemple, représenter une exploration uniforme des sites voisins à \widehat{X}_n .

Le lecteur aura certainement noté que X_n n'est pas une chaîne de Markov, mais le processus historique possède toujours la propriété markovienne.

Exercice 7.4.1 Examiner la situation où le potentiel est constant $G(x) = \epsilon$, avec $\epsilon \in [0, 1]$. Établir une analogie avec la loi des grands nombres.