



HAL
open science

Les processus stochastiques

Bahram Houchmandzadeh

► **To cite this version:**

Bahram Houchmandzadeh. Les processus stochastiques. Doctoral. Les processus stochastiques, Grenoble, France. 2010. cel-01148909v2

HAL Id: cel-01148909

<https://hal.science/cel-01148909v2>

Submitted on 13 Nov 2019 (v2), last revised 18 Aug 2020 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

BAHRAM HOUCHEMANDZADEH

PROCESSUS
STOCHASTIQUES.

Aucun droit réservé. Chaque partie de ce manuscrit peut être reproduite, modifiée ou transmise sans l'autorisation de l'auteur.
Tout commentaire, critique ou correction sera grandement apprécié.

Remerciements.

Je remercie sincèrement You XIA pour sa lecture attentive du manuscrit et ses très nombreuses corrections et suggestions.

web : [www-liphy.ujf-grenoble.fr/
pagesperso/bahram/Phys_Stat/phystat.html](http://www-liphy.ujf-grenoble.fr/pagesperso/bahram/Phys_Stat/phystat.html)
mél : bahram.houchmandzadeh (à) univ-grenoble-alpes.fr
Première version : Septembre 2008
Version présente : 13 novembre 2019

Table des matières

1	<i>Introduction.</i>	5
2	<i>Concepts de probabilités.</i>	7
	2.1 <i>Les notions fondamentales.</i>	7
	2.2 <i>Approfondissement.</i>	12
	2.3 <i>Exercices.</i>	15
3	<i>Les processus stochastiques.</i>	17
	3.1 <i>Exemple fondamental.</i>	17
	3.2 <i>Définitions et généralisations.</i>	20
4	<i>L'équation Maîtresse à travers des exemples.</i>	25
	4.1 <i>Quelques exemples de processus stochastique.</i>	25
	4.2 <i>Manipulation de l'équation maîtresse.</i>	30
	4.3 <i>Applications aux exemples choisis.</i>	33
	4.4 <i>Problèmes.</i>	40
5	<i>Quelques généralisations de l'équation Maîtresse.</i>	43
	5.1 <i>Système à deux composants.</i>	43
	5.2 <i>Système spatialement étendu.</i>	45
	5.3 <i>Transitions entre états quelconque.</i>	49
	5.4 <i>Les états continus.</i>	54
	5.5 <i>Un rapide aperçu de la théorie des jeux.</i>	55
	5.6 <i>Problèmes.</i>	55

6	<i>Probabilité de fixation et temps de premier passage.</i>	57
6.1	<i>L'équation maîtresse amont (backward).</i>	57
6.2	<i>Probabilité de fixation.</i>	58
6.3	<i>Probabilité de survie & temps de fixation.</i>	61
7	<i>Intégrales de chemins.</i>	65
7.1	<i>Formulation générale.</i>	65
7.2	<i>Physique statistique.</i>	67
7.3	<i>Théorie des grandes déviations.</i>	67
8	<i>Résolution numérique des équations maîtresse.</i>	69
9	<i>L'équation de diffusion (Fokker-Planck).</i>	71
9.1	<i>L'équation aval (forward).</i>	71
9.2	<i>Les systèmes continus.</i>	74
9.3	<i>Exemples d'application de l'équation de Diffusion.</i>	77
9.4	<i>L'équation amont (backward) et le temps de premier passage.</i>	79
9.5	<i>Problèmes.</i>	79
10	<i>Les équations différentielles stochastiques et le calcul d'Ito.</i>	81
10.1	<i>Les équations différentielles stochastiques.</i>	81
10.2	<i>Manipulation de $dW(t)$.</i>	83
10.3	<i>Stratonovitch vs Ito.</i>	84
10.4	<i>Problèmes.</i>	84
11	<i>Les filtres de Kalman.</i>	85
11.1	<i>Prélude.</i>	85
11.2	<i>Observation d'une Chaîne de Markov.</i>	87
A	<i>Le traitement du mouvement Brownien par Einstein, Langevin et Smoluchowski.</i>	89

1

Introduction.

Nous avons tous une idée plus ou moins précise de ce qu'implique le titre de ce document. "Stochastique" est un mot un peu chic pour dire "aléatoire", ce qui évoque bien sûr l'idée des probabilités; "processus" évoque l'idée d'un changement dans le temps. Selon le Larousse,

Processus : Enchaînement ordonné de faits ou de phénomènes, répondant à un certain schéma et aboutissant à quelque chose.

Le lecteur est sûrement déjà familier avec le concept de probabilité : la probabilité d'obtenir "2" avec un dès non biaisé est $1/6$; la probabilité d'obtenir 7 en jetant deux dès est encore $1/6$. Un processus stochastique est plutôt comme un jeu de l'oie : si j'obtiens 2 j'avance de deux cases, si je tombe dans la case serpent, je descend 7 cases et je rejette les dès 2 fois ... Quelle est alors la probabilité d'arriver au bout ou celle de revenir à la case de départ ?

Évidemment, si l'on étudie les processus stochastiques, c'est que nous les voyons apparaître dans toutes les branches de la science. Donnons nous quelques exemples que nous étudierons plus tard plus en détail.

Évolution : Nous avons N individus dans un écosystème (numérotions les de 1 à N). A chaque génération, chaque individu produit un nombre aléatoire de descendant; la plupart des descendants meurent pour que le nombre d'habitant de l'écosystème reste autour de N . En supposant que tous les individus soient équivalents, quelle est la probabilité de voir des descendants de l'individu i au bout de m générations? Et si l'individu i était légèrement meilleur que les autres à se reproduire, comment cette probabilité doit être évalué? Ce problème a été résolu pour la première fois dans les années 1930 par les pères fondateurs de la nouvelle synthèse (le mariage entre l'évolution Darwinienne et la génétique Mendélienne) et constitue le socle mathématiques de notre compréhension de la théorie de l'Évolution.

Physique Statistique : L'hypothèse fondamentale de la physique statistique est que dans un système micro-canonique qui ne peut rien échanger avec l'extérieur (ni énergie, ni volume, ni matière, ...) tous les micro-états dont l'énergie est E ont la même probabilité. Ce théorème était profondément choquant pour les mathé-

maticiens et mécanicien de l'époque. Boltzmann s'en est sorti¹ en introduisant l'hypothèse du chaos moléculaire. Considérons la fonction $p(x, t)$, la (densité de) probabilité d'observer une particule à l'énergie x au temps t . Considérons maintenant un choc entre deux particules. Lors d'un choc, une particule peut transférer une partie de son énergie à l'autre² et Boltzmann a émis l'hypothèse que cette redistribution est aléatoire (d'où le nom chaos moléculaire). Connaissant la forme de la fonction $p(x, 0)$, que peut-on dire sur la fonction $p(x, t)$? Comment évolue la quantité appelée Entropie, $S(t) = \int p \log p dx$? La résolution de ce problème (connu sous le nom du théorème H) est le fondement de la physique statistique.

Mécanique Quantique : La mécanique quantique est bien sûr formulée comme un processus stochastique. Connaissant l'amplitude initiale $\psi(x, 0)$, et connaissant l'hamiltonien H du système (un opérateur que l'on suppose pour l'instant indépendant du temps), que vaut l'amplitude à un temps ultérieur $\psi(x, t)$? La réponse à cette question est la base de notre compréhension du monde microscopique et nous apprenons pendant nos études les mille et une manières de calculer efficacement la fonction $\psi(x, t) = e^{itH}\psi(x, 0)$. Une interprétation alternative a été donnée par Feynman dans les années 1950, mettant encore plus l'accent sur le processus, en calculant l'amplitude des trajectoires comme $e^{iS/\hbar}$, où S est l'action classique.

La Chimie : à petite échelle, quand le nombre de molécule n'est pas abondant, la chimie appartient pleinement au domaine des processus stochastiques : une réaction $A + B \rightarrow C + D$ se traite comme la probabilité que deux molécules A et B donnent lieu à une réaction chimique ; si la molécule D participe elle-même à une autre réaction chimique, nous pouvons avoir des cascades probabilistes. A l'échelle des cellules, ceci provoque la différence entre les bactéries isogénique, comme celle qui digèrent le sucre lactose et celle qui ne le font pas. Depuis les années 2000, c'est cette vision probabiliste des cellules qui émerge, en remplaçant la vision cellule, machine déterministe.

Nous verrons tous ces exemples et bien d'autres à travers ce cours. Ce cours est construit beaucoup plus à travers les exemples choisis comme illustration des méthodes que comme un cours de mathématiques avec des énoncés de théorème et de Lemmes.

La première partie est consacrée au rappel des résultats principaux des Probabilités. Nous allons ensuite introduire les probabilités de transition et aboutir à l'équation Maître³. Nous passerons beaucoup de temps avec l'équation Maître et nous la généraliserons au fur et à mesure. Nous introduirons ensuite une approximation fondamentale de cette équation connue sous le nom de l'équation de diffusion amont et aval, ce qui nous permettra d'utiliser l'arsenal des outils des équations aux dérivées partielles.

Bon, assez parlé, passons aux choses concrètes.

1. Pour beaucoup, cela reste insuffisant

2. Nous avons trois équations de conservation de mouvement, une équation de conservation d'énergie, et 6 inconnues qui sont les vitesses des deux particules. Deux inconnues ne peuvent donc pas être tirées des équations de conservation.

3. Appelé Chapman-Kolmogorov par les mathématiciens et Fokker-Planck par les physiciens.

2

Concepts de probabilités.

2.1 Les notions fondamentales.

2.1.1 Loi de distribution.

Nous parlons de probabilité quand nous manquons d'information sur un système et que l'on ne peut pas prédire son comportement. Connaissant le temps aujourd'hui, la météo peut prédire de façon presque sûr le temps demain, mais aura beaucoup de mal à prédire le temps dans 10 jours. Comment quantifier notre incertitude? Supposons que nous faisons une expérience de jeté de dès. La connaissance des contraintes physiques nous amène à postuler que les dès ne peuvent pas produire n'importe quel résultat, mais seulement un entier entre 1 et 6. Nous avons énormément réduit l'ensemble des résultats possible de la "variable aléatoire jeté de dès"; par contre, nous avons utilisé toute l'information que nous avons et nous ne pouvons pas aller plus loin dans nos prédictions. Dans ce cas, nous postulons que "la probabilité d'obtenir n ($n \in \{1, 2, \dots, 6\}$) est $1/6$ ". Ce postulat vient de notre expérience passé : ayant jeté des dès des centaines de fois, nous avons eu l'impression qu'aucun nombre n'était privilégié par rapport aux autres. Formellement, nous imaginons un robot ayant jeté des dès des milliard de fois et qui a dressé le nombre relatif de chaque résultat obtenu, et constaté que ce nombre relatif n'est pas différent des autres. Ceci est le cœur du concept de probabilité : nous ne pouvons rien dire de très précis, nous ne faisons que résumer le résultat d'un très grand nombre de réalisation passé. Si nous avions grandi avec des dès pipés vers le six, nous aurions postulé probablement une phrase du genre "la probabilité d'obtenir 6 est $1/2$, la probabilité d'obtenir les autres chiffres est $1/10$ ".

Supposons maintenant que quand nous obtenons 6, nous gagnons 10 euros et que pour les autres chiffres, nous devons payer 1 euro. Dans ce cas, nous basant sur nos postulats de dès non pipés, nous pouvons formuler la phrase suivante : "A chaque réalisation, la probabilité d'obtenir 10 euros est $1/6$, la probabilité de perdre 1 euro est $5/6$ ". Ce postulat vient de notre connaissance de la loi des dès non pipés (équiprobabilité des entiers entre 1 et 6) et de la loi de transformation (ou là règle du jeu). Ceci est exactement le tra-

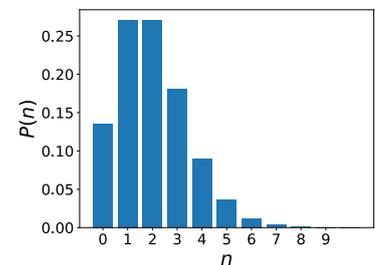


FIGURE 2.1 – Exemple d'une loi de distribution (ici loi de Poisson avec $\lambda = 2$).

vail du scientifique : connaissant la loi de répartition des différents états microscopiques, et comment chacun de ces états influe sur le comportement global du système, prédire ce comportement global.

Précisons tout cela. Nous appelons une variable aléatoire X un phénomène physique que l'on ne peut pas prévoir exactement, comme par exemple la tension exacte au bord d'une résistance ou le nombre exact de molécule à un temps donné. Chaque mesure du phénomène physique est appelé "une réalisation". Supposons pour l'instant que notre variable aléatoire produit des valeurs distincts x_1, x_2, x_3, \dots . Pour caractériser notre variable aléatoire, nous faisons un très grand nombre de réalisations (mesures), disons N . Nous comptons ensuite le nombre de fois N_1 ou nous avons obtenu x_1 , le nombre de fois N_2 ou nous avons obtenu x_2 et ainsi de suite, et nous résumons tout cela dans un tableau ou une graphie ou l'on porte x_i en abscisse et N_i en ordonnées. En faite, on fait même un peu mieux, on porte le nombre relatif $P(x_i) = N_i/N$ en ordonné. En théorie des probabilité, nous supposons toujours que nous avons fait une infinité de mesures, ou que l'on peut imaginer un robot fictif qui aurait fait cette infinité de mesure pour nous.¹ Dans ce cas, la donnée des couples $\{x_i, P(x_i)\}$ caractérise entièrement le phénomène aléatoire en question (ou autrement dit résume toute l'information que nous possédons sur le phénomène). Nous appelons la fonction P la *loi de distribution* du phénomène aléatoire. Pour un nombre donné x_i , nous appelons $P(x_i)$ sa *probabilité*. Par exemple, pour la variable aléatoire "jet de dès", la probabilité d'obtenir 6 est $P(6) = 1/6$.

2.1.2 Composition des probabilités : ET.

Supposons que nous avons deux dès non pipés. Quelle est la probabilité d'obtenir 1 avec le premier et 3 avec le deuxième lors d'un jeté simultané? Si les deux événements sont indépendants², les probabilités se multiplient. La raison est simple : considérons une expérience de N ($N \rightarrow \infty$) jets de deux dès. Dans cet ensemble, une proportion $P(X_1 = 1)$ a obtenu le premier dès égal à 1. Considérons maintenant l'ensemble des résultats où le premier jet est égal à 1. Une proportion $P(X_2 = 3)$ a obtenu le deuxième dès égal à 3. En tout donc, une proportion $P(1)P(3)$ ont satisfait au critère. De façon générale, pour deux événements indépendants E_1 et E_2 combinés par un ET logique et régis par des lois P_1 et P_2 , nous avons

$$P(E_1, E_2) = P_1(E_1)P_2(E_2)$$

2.1.3 Composition des probabilités : OU.

la probabilité pour obtenir une réalisation ou une autre, si ces deux réalisations sont mutuellement exclusives, est la somme des probabilités individuelles³. Par exemple, la probabilité d'obtenir 1 ou 3 est de $1/6 + 1/6 = 1/3$. Considérons à nouveau une expérience de N ($N \rightarrow \infty$) jets de dès. Un nombre N_1 retourne le résultat 1

1. Évidemment, une infinité de mesure n'appartient pas à notre monde physique. Le concept d'infini peut cependant être une très très bonne approximation pour les phénomènes physiques. Surtout, nous savons que dans la plupart des cas, plus nous faisons de mesure, plus nous nous approchons de la probabilité théorique. Le lecteur notera que je présente le concept de probabilité à travers l'œil du physicien. Pour le mathématicien, le concept de probabilité est juste une extension de la théorie mathématique de la mesure et n'a pas besoin de "robot fictif" faisant des expériences répétées.

2. Les événements sont indépendants si le résultat obtenu par l'un n'influe par sur l'autre.

3. En théorie axiomatique des probabilités que nous avons évité d'énoncer ici, nous prenons cette assertion comme un axiome.

et un nombre N_2 le résultat 3. Le nombre total de jet ayant rendu 1 ou 3 est donc $N_1 + N_2$ et la probabilité de cet événement est $(N_1 + N_2)/N = N_1/N + N_2/N = P_1 + P_2$. De façon général, pour deux événement mutuellement exclusive, nous avons

$$P(E_1 \text{ OU } E_2) = P(E_1) + P(E_2)$$

Par exemple, pour la variable aléatoire “somme de deux jets de dès”, $P(7) = 1/6$. Nous avons différentes façon d’obtenir ce dernier résultat. Une façon de voir cela est de dire que pour obtenir 7 pour la somme d’un double jet, nous devons obtenir soit $\{1,6\}$, soit $\{2,5\}$,... soit $\{6,1\}$; chaque événement mentionné a une probabilité de $1/36$, d’où le résultat énoncé.

2.1.4 Variable aléatoire continue.

Dans beaucoup de cas, l’appareil de mesure que nous utilisons est de précision limitée par rapport au phénomène que nous mesurons. Par exemple, notre voltmètre a une précision de millivolt ou nous mesurons les positions d’une bille au nanomètre près. Dans ce cas, nous parlons de “probabilité d’obtenir un résultat entre x et $x + h$, où h est la précision de notre mesure, et nous notons cela $P(x < X < x + h)$. Un problème se pose quand nous mesurons le même phénomène par deux appareils différents de précision différentes. Comment faire le lien entre les deux ensemble de mesures? Nous voyons que la fonction P dépend ici de deux variables x et h . Cependant, il nous semble naturelle de dire en divisant l’intervalle $[x, x + h]$ en deux, que

$$P(x < X < x + h) = P(x < X < x + h/2) + P(x + h/2 < X < x + h)$$

Cela découle directement de la loi des composition OU. En divisant l’intervalle en N morceau au lieu de deux ($N \rightarrow \infty$), on abouti à

$$P(x < X < x + h) = \int_x^{x+h} p(x) dx$$

Où la fonction d’une seule variable $p(x)$ caractérise entièrement la fonction de deux variables $P(x, h)$. Autrement dit

$$P(x < X < x + dx) = p(x) dx$$

où dx est une précision infiniment petite⁴. La fonction $p(x)$ est appelé *une densité de probabilité*. Expérimentalement bien sûr nous ne pouvons pas accéder à la précision absolue et les histogrammes expérimentaux portent toujours la trace de h (du “binning” ou de la “fonction d’appareil”); si cependant h est très petite par rapport à l’intervalle de mesure pertinente, nous estimons que nous avons une bonne approximation de la densité de probabilité.⁵

Notons enfin que si la variable aléatoire continue X à un domaine fini, c’est à dire ses mesures (réalisations) tombent toujours dans l’intervalle $[a, b]$, nous pouvons toujours prendre la densité de

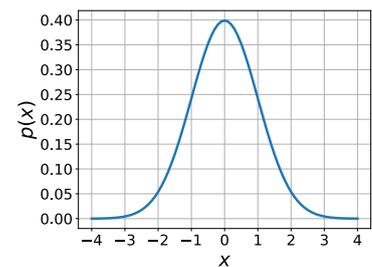


FIGURE 2.2 – Une densité de probabilité continue (ici la gaussienne $(2\pi)^{-1/2} \exp(-x^2/2)$).

4. Les mathématiques classiques n’aiment pas trop les infiniment petits et leurs usages (pourtant très intuitif pour un physicien) a été banni; Les infiniment petits ont été “relégitimisés” dans les années 1960 à travers ce qu’on appelle l’Analyse non standard et les nombres hyper-réels.
5. Parfois, quand nous connaissons bien notre appareil de mesure, nous pouvons tenter de *déconvoluer* le signal et en exclure l’alliassage produit par l’appareil de mesure. Le résultat n’est pas toujours garanti.

probabilité $p(x)$ définie sur $] -\infty, \infty[$ avec $p(x < a) = p(x > b) = 0$. Cela nous permettra de toujours écrire les intégrales que nous avons à faire sur $] -\infty, \infty[$ sans faire des cas particulier pour chaque variable aléatoire.

2.1.5 Loi de distribution de deux variables aléatoires.

Soit deux variables aléatoires discrètes X et Y de loi de distribution $P_1(x)$ et $P_2(y)$. La probabilité d'obtenir, lors d'une mesure combinée, x dans la première mesure et y dans la deuxième est donnée par une nouvelle loi, $P(x, y)$. Comme nous l'avons dit plus haut, si les deux évènements sont indépendants, $P(x, y) = P_1(x).P_2(y)$ mais en général nous n'avons pas à supposer l'indépendance. Pour une variable continue, nous parlerons de la densité de probabilité $p(x, y)$. Il faut le comprendre comme suit : la probabilité pour que le résultat (x, y) d'une mesure combinée tombe dans un domaine \mathcal{D} est (fig.2.3a)

$$\int \int_{\mathcal{D}} p(x, y) dx dy$$

Si les deux évènements sont indépendants, nous pouvons comme avant écrire $p(x, y) dx dy = p_1(x) dx p_2(y) dy$ ou simplement $p(x, y) = p_1(x) p_2(y)$.

Revenons au cas général de deux variables aléatoires, qu'elles soient indépendante ou non. Nous avons vu comment nous pouvons définir, par des expériences répétées, la loi combinée $P_{i,j}$ ou si les distributions sont continues, $p(x, y) dx dy$. Comment calculer la probabilité qu'une mesure de X tombe donne i quelque soit le résultat de Y ? Du point de vue expérimental, cela revient à ignorer où tombe la variable Y à chaque fois que l'on fait une mesure simultanée. D'un autre côté, si nous avons fait N mesures combinées des deux variables et noté soigneusement les nombres N_{ij} d'obtenir i sur la variable X et j sur la variable y , nous savons que le nombre total de fois où nous avons obtenus $X = i$ quelque soit le résultat obtenu sur Y et

$$N_i = \sum_j N_{ij}$$

ou écrit en langage des nombres relatifs (probabilités) en divisant les deux côtés par N :

$$P(X = i) = \sum_j P(X = i, Y = j)$$

qui pour des probabilités continues, se transforme en

$$p(x) = \int_{\mathbb{R}} p(x, y) dy$$

2.1.6 Moyenne, variance, et moments d'ordre supérieurs.

Connaître une loi de distribution nécessite le stockage de beaucoup d'information : l'ensemble des couples $(x_i, P(x_i))$ pour une variable discrète et infiniment plus si la variable est continue⁶. Il

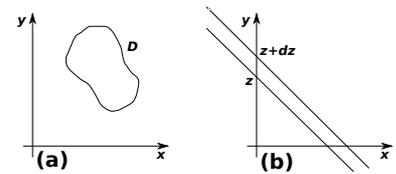


FIGURE 2.3 – (a) La probabilité pour qu'une mesure combinée (x, y) tombe dans un domaine \mathcal{D} est $\int \int_{\mathcal{D}} p(x, y) dx dy$. (b) Pour que la variable aléatoire $Z = X + Y$ tombe dans le domaine $[z, z + dz]$, il faut que le résultat (x, y) d'une mesure combinée tombe dans le domaine délimitée par les deux droites ;

6. Pour un physicien, l'infini est "finiment" limité par la précision des mesures.

se trouve que nous ne sommes pas très bon à imaginer et établir des comparaisons quand nous avons trop de données (c'est un peu comme essayer de tenir plusieurs pastèques dans une main) et nous devons trouver des moyens de réduire de façon intelligente la quantité totale d'information. De plus, il se trouve que dans beaucoup de cas qui intéressent le physique, on peut faire des prédictions en disposant de quelques nombres que l'on tire de la distribution dans son ensemble.

Pour un professeur qui veut comparer le niveau de deux classes⁷ une donnée intéressante est la *moyenne*⁸

$$\mu(X) = \sum_i iP_i$$

où P_i est le nombre relatif d'étudiant (autrement dit, la probabilité) ayant la note i . Si nous manipulons des densités de probabilité continues, $P_i = p(x_i)dx$ et la somme se transforme en intégrale

$$\mu(X) = \int_{-\infty}^{+\infty} xp(x)dx$$

La moyenne est probablement la fonction la plus célèbre en statistique et probabilité, et nous permet d'un seul coup d'œil d'avoir une idée de l'ordre de grandeur du processus aléatoire. Il existe une autre quantité qui est moins sensible aux très grandes fluctuations⁹ qu'on appelle la médiane : la moitié des mesures tombent plus bas que cette valeur, l'autre moitié plus haut. Cette fonction se traite cependant moins bien analytiquement et est moins utilisée.

Considérons maintenant deux classes de 100 étudiants. Dans la première, tous les étudiants ont 10/20. Dans la deuxième, 50 étudiants ont 20/20, l'autre moitié 0/20. Les deux classes, ayant la même moyenne, ont clairement des distributions très différentes. Pour quantifier cette différence, nous pouvons calculer "la moyenne des écarts à la moyenne", par exemple $\sum |x - \mu|p(x)$. Nous n'avons pas trop cette quantité puisqu'elle contient la valeur absolue, qui ne se traite pas bien analytiquement. Nous calculons donc plus souvent la quantité

$$\text{Var}(X) = \int_{-\infty}^{+\infty} (x - \mu(X))^2 p(x)dx$$

ou sa version discrète. La racine carrée de cette quantité donne l'ordre de grandeur des fluctuations autour de la moyenne. On appelle cette quantité *l'écart quadratique moyen* ou *déviations standard* :

$$\sigma(X) = \sqrt{\text{Var}(X)}$$

De façon générale, nous appelons le moment d'ordre n d'une distribution la quantité

$$\mu_n(X) = \int_{-\infty}^{+\infty} x^n p(x)dx$$

et le moment centré d'ordre n la quantité

$$\bar{\mu}_n(X) = \int_{-\infty}^{+\infty} (x - \mu(X))^n p(x)dx$$

7. ou un manager comparant deux lignes de production, ou un physicien comparant les tensions au bord de deux circuits,...

8. On appelle cette quantité en mathématique l'espérance de la variable X qu'on note $E(X)$. Ce mot a été choisi par Blaise Pascal pour justifier pourquoi il faut croire en Dieu. Le lecteur ayant quelques connaissances en probabilité et ayant parcouru *les pensées* (science malheureusement très peu répandue auprès des philosophes) pourrait résoudre l'exercice suivant : pourquoi le calcul de Pascal est faux ?

9. Si Bill Gate, le président de Microsoft, entre dans un bar à Grenoble, la moyenne des revenus des clients augmente brusquement. La médiane change beaucoup moins.

Le moment centré d'ordre 3 donne par exemple la dissymétrie de la distribution autour de la moyenne. Il n'est pas fréquent cependant en pratique d'utiliser plus que les deux premiers moments.

Notons enfin que le moment d'ordre zéro est trivial :

$$\mu_0 = \int_{-\infty}^{+\infty} p(x)dx = 1$$

Revenons à notre exemple de la classe. Supposons que des prix $F(i)$ sont distribués aux étudiants en fonction de leur note i . Quel est le montant moyen de prix accordé à une classe? Un calcul trivial montre que

$$\mu(F(X)) = \sum_i F(i)P(i)$$

ou en version continue

$$\mu(F(X)) = \int_{-\infty}^{+\infty} F(x)p(x)dx$$

Si par exemple nous disposons de N particules dont les vitesses sont distribuées selon la densité de probabilité $p(v)$, l'énergie cinétique moyen de ces particules est $\mu = \int_I (m/2)v^2 p(v)dv$. D'ailleurs, nous voyons que les moments d'ordre n peuvent s'écrire simplement

$$\mu_n(X) = \mu(X^n)$$

2.1.7 La notation $\langle \rangle$.

Les physiciens ont pris l'habitude d'utiliser la notation $\langle X \rangle$ à la place de $\mu(X)$ présenté ci-dessus pour désigner la moyenne d'une distribution. Cette notation nous évite à chaque fois d'écrire des \sum ou des \int et facilite beaucoup l'écriture des formules longues. Voici les règles de manipulation de ce symbole qui se démontre de façon trivial.

1. $\langle X + Y \rangle = \langle X \rangle + \langle Y \rangle$ quelque soit les variables X et Y .
2. $\langle aX \rangle = a \langle X \rangle$ pour un nombre arbitraire $a \in \mathbb{R}$ ou \mathbb{C} .
3. $\langle X.Y \rangle = \langle X \rangle \langle Y \rangle$ si et seulement si les deux variables X et Y sont indépendantes.

2.2 Approfondissement.

En utilisant les résultats somme toute assez intuitifs ci-dessus, nous allons obtenir quelques résultats légèrement plus compliqués et utiles pour la suite de ce cours. Dans la plupart des cas, nous utiliserons une variable aléatoire continue, plus fréquemment rencontrée dans le monde de la physique, mais évidemment, les résultats ne dépendent pas du caractère discret ou continu du processus. Dans la formulation, il faut seulement remplacer des $\sum P(n)$ par $\int p(x)dx$.

2.2.1 Entropie de l'information.

Nous voudrions trouver une fonction S pour caractériser le degré d'imprédictibilité d'un système, et nous allons lui donner le nom d'entropie. Par exemple, quelle est la différence entre un dès à 6 faces et un dès à 10 faces? Le premier est plus déterministe, puisque il possède moins de résultats possible. On peut donc simplement prendre le nombre W de cas possible pour un système comme une mesure de sa nature probabiliste. Allons un peu plus loin, nous voulons que la fonction S soit additive, c'est à dire que si on considère deux systèmes (indépendants) ensemble, leurs entropie s'ajoute. Si W_1 et le nombre de cas possible du premier système et W_2 le nombre de cas du deuxième système, le nombre de cas possible sera $W_1 W_2$; si nous avons choisi le nombre de cas comme la définition de l'entropie, cette fonction serait multiplicative et non additive. Qu'à cela ne tienne, nous connaissons la fonction logarithme qui transforme les multiplications en additions, et donc nous pouvons définir

$$S = \log W \quad (2.1)$$

Cette définition a quelques petites avantages supplémentaire. Si par exemple un système est parfaitement déterministe et nous connaissons toujours à l'avance le résultat d'une mesure, alors le système n'a qu'un cas possible et son entropie est $S = 0$. Boltzmann (vers 1875) était tellement satisfait avec cette définition qu'il l'a fait gravé sur sa tombe.

Nous pouvons un peu arranger cette définition pour l'exprimer en terme des probabilité. Dans le cas d'un système ou tous les événements sont équiprobable, la probabilité pour un événement i est simplement $p_i = 1/W$ et nous pouvons donc réécrire S

$$S = -\log p_i$$

et pour ne pas favoriser un événement i en particulier,

$$S = -(1/W) \sum_i \log p_i$$

où même mieux encore

$$S = -\sum_i p_i \log p_i \quad (2.2)$$

L'effort semble un peu vain, c'est comme si au lieu de parler du chiffre 1, nous parlions du chiffre "232 fois 1 232ème". Mais cela est pour palier à un gros défaut de l'expression (2.1). Considérons un système avec deux résultats possible A et B avec $P(A) = p; P(B) = 1 - p$. Si nous utilisons l'expression (2.1), l'entropie de ce système est toujours $\log 2$. Mais il est évident qu'un système ou $p = 1/2$ est plus aléatoire qu'un système où $p < 1/2$. A la limite où $p \rightarrow 0$, le système devient presque déterministe et nous nous attendrions à ce que $S \rightarrow 0$; pourtant, si nous utilisons la définition (2.1) l'entropie sera $\log 2$ quelque soit p . Une meilleur solution serait de pondérer

la contribution de chaque résultat possible à l'entropie (le facteur $\log p_i$) par sa probabilité p_i , ce que fait exactement l'expression (2.2). Pour notre système à deux états, quand $p \rightarrow 0$, l'entropie définie par (2.2) tend effectivement vers zéro.

Notons que cette définition conserve l'additivité de l'entropie. Considérons deux systèmes indépendants déterminés par les probabilités $\{P_i | i \in A_1\}$ et $\{Q_j | j \in A_2\}$. Le système combiné sera caractérisé par $\{R_{ij} = P_i Q_j | (i, j) \in A_1 \times A_2\}$. L'entropie du système combiné sera

$$\begin{aligned} S_{12} &= \sum_{i,j} R_{ij} \log R_{ij} \\ &= \sum_{i,j} P_i Q_j (\log P_i + \log Q_j) \\ &= \sum_i P_i \log P_i \sum_j Q_j + \sum_j Q_j \log Q_j \sum_i P_i \end{aligned}$$

Or, par définition des probabilités, $\sum_i P_i = \sum_j Q_j = 1$ et nous trouvons donc finalement

$$S_{12} = S_1 + S_2$$

Cette définition de l'entropie est due plutôt à Gibbs (vers 1890) mais n'est pas gravée sur sa tombe. Une cinquantaine d'années plus tard, Shannon a introduit ce concept en théorie de communication qui plus tard est devenue théorie de l'information. C'est en se basant sur ce concept que nous pouvons par exemple compresser des fichiers de données. Nous dédions un chapitre entier à ces concepts plus tard dans le cours de ce manuscrit.

2.2.2 Théorème central limite.

Parmi toutes les fonctions de distribution de probabilité, la "gaussienne" joue un rôle très particulière :

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

le facteur $1/\sqrt{2\pi}\sigma$ assure la normalisation de la fonction : $\int_{\mathbb{R}} f(x) dx = 1$ ¹⁰. Cette fonction étant symétrique, sa moyenne est nulle, et sa variance $\int_{\mathbb{R}} x^2 f(x) dx$ vaut justement σ^2 . La primitive de cette fonction est appelée la fonction d'erreur $\text{erf}(x) = (2/\sqrt{\pi}) \int_0^x \exp(-t^2) dt$. La probabilité pour que la variable aléatoire X ayant cette fonction de distribution tombe dans l'intervalle $[-x, x]$ est donc

$$F(x) = P(-x/\sigma < X < x/\sigma) = \int_{-x}^x f(x) dx = \text{erf}(x/\sqrt{2}\sigma)$$

En particulier, $F(1) = 0.68$, $F(2) = 0.955$ et $F(3) = 0.997$. Cela veut dire en particulier qu'il y a très peu de chance (< 0.003) d'observer la variable au delà du 3σ ; les chances de l'observer au delà de 5σ est de l'ordre de 10^{-7} .

Revenons maintenant au problème de somme Y de N variables aléatoires X_n identiques. On peut imaginer les variables comme

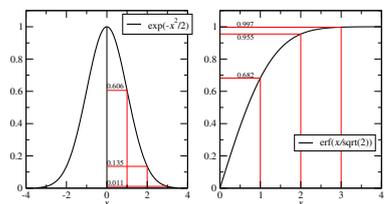


FIGURE 2.4 – Les fonctions $\exp(-x^2/2)$ et $\text{erf}(x/\sqrt{2})$

10. Pour retrouver facilement le facteur de normalisation, il suffit de se souvenir que $\int_{\mathbb{R}} \exp(-x^2) dx = \sqrt{\pi}$. Il suffit ensuite d'effectuer le changement de variable $x \rightarrow x/\sqrt{2}\sigma$ pour retrouver le bon facteur.

“jeté de dès” et leur somme comme somme de N jetés. Nous avons défini cela comme

$$Y_N = \sum_{n=1}^N X_n$$

Appelons μ et σ^2 la moyenne et la variance des X_n : $\mu = \langle X_n \rangle$ et $\sigma^2 = \text{Var}(X_n)$. Nous savons que

$$\langle Y_N \rangle = \sum_{n=1}^N \langle X_n \rangle = N\mu$$

et quand les X_n sont indépendantes,

$$\text{Var}(Y_N) = \sum_{n=1}^N \text{Var}(X_n) = N\sigma^2$$

Nous pouvons dire quelque chose de beaucoup plus précis sur Y si le nombre N est très grand. Il existe un théorème fantastique en mathématique¹¹, appelé théorème “limite centrale” qui énonce que la fonction de distribution de Y_N , $f_N(y)$ converge vers une Gaussienne quand $N \rightarrow \infty$:

$$f_N(y) \approx \frac{1}{\sqrt{2\pi N}\sigma} e^{-(y-N\mu)^2/2N\sigma^2}$$

Peut importe la loi de distribution des X_n , leurs somme, pourvu que N soit assez grand, est une gaussienne. La convergence est extrêmement rapide dans la plupart des cas et la somme d’une centaine de variable aléatoire est très souvent une très bonne gaussienne. Si N est un nombre aussi important qu’une mole, $\sim 10^{23}$, ou même une fraction de ce nombre, on ne peut distinguer $f_N(y)$ d’une gaussienne.

2.3 Exercices.

§ 2.1 Quelques lois célèbres.

Soient les lois suivantes

— binomiale (discret) :

$$P(0) = 1 - p; P(1) = p$$

— Poisson (discret) :

$$P(n) = e^{-\lambda} \lambda^n / n! \quad (n = 0, 1, 2, \dots)$$

— Normale (continue) :

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

— Poisson(continue) :

$$p(t) = \mu e^{-\mu t} \quad (t \in [0, +\infty[$$

— Cauchy (continue) :

$$p(x) = \frac{1}{\pi} \frac{\sigma}{(x-\mu)^2 + \sigma^2} \quad (x \in]-\infty, +\infty[$$

11. Dû à Laplace (~1810) pour certains cas particuliers et démontré dans le cas général par Lyapunov (1901).

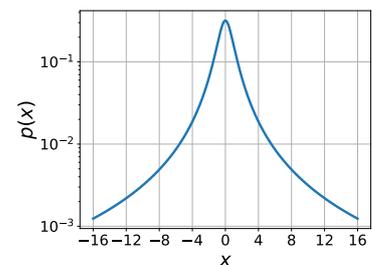


FIGURE 2.5 – La densité de probabilité de Cauchy avec $\mu = 0$ et $\sigma = 1$.

Dans chaque cas, tracer la fonction (soit la probabilité, soit sa densité, selon le cas), calculer la moyenne et la variance de la variable aléatoire.

§ 2.2 La notation $\langle \rangle$

Démontrer

1. $\langle X + Y \rangle = \langle X \rangle + \langle Y \rangle$ quelque soit les variables X et Y .
2. $\langle aX \rangle = a \langle X \rangle$ pour un nombre arbitraire $a \in \mathbb{R}$ ou \mathbb{C} .
3. $\langle X.Y \rangle = \langle X \rangle \langle Y \rangle$ si et seulement si les deux variables X et Y sont indépendantes.
4. $Var(X) = \langle (X - \langle X \rangle)^2 \rangle = \langle X^2 \rangle - \langle X \rangle^2$
5. $Var(X + Y) = Var(X) + Var(Y)$ si et seulement si les deux variables X et Y sont indépendantes.

§ 2.3 Combinaison des probabilités.

Soit un ensemble de lignes horizontales équidistant de a . Soit une barrette de taille a qu'on lance sur le plan. Démontrer que la probabilité que la barrette coupe une ligne est $2/\pi$.

§ 2.4 Combinaison des probabilités.

Soit un sac contenant une proportion p de boules blanches. Nous faisons l'expérience suivante : nous tirons une boule, notons sa couleur et le remplaçons dans le sac ; Nous réitérons cette expérience N fois. Quelle est la probabilité d'obtenir n boules blanches ? Quelle est la limite quand $N \rightarrow \infty, p \rightarrow 0$ mais $Np \rightarrow \lambda > 0$?

§ 2.5 Pour deux variables aléatoires continues indépendantes X et Y , donner la densité de probabilité de la variable $Z = X - Y$, $Z = X/Y$ et $Z = XY$. Donner dans chaque cas les hypothèses sur X et Y nécessaires à l'existence de la variable.

§ 2.6 Fonction d'une variable aléatoire.

Les autres cas (fonction monotone décroissante et fonction oscillante).

§ 2.7 démontrer qu'en général, $\langle f(X) \rangle \neq f(\langle X \rangle)$. A quelle condition l'égalité est vérifiée ?

§ 2.8 Démontrer l'additivité des moments centrés pour des variables indépendantes.

§ 2.9 Démontrer la relation (??) entre les dérivées des cumulants et les moments centrés pour $n = 1, 2, 3$. Démontrer que pour $n = 4$, cette relation n'est plus vérifiée.

§ 2.10 Démontrer que pour une variable normale, seule les deux premiers cumulants sont $\neq 0$.

3

Les processus stochastiques.

3.1 Exemple fondamental.

Nous allons maintenant nous intéresser aux phénomènes où la probabilité d'obtenir un certain résultat dépend de ce que nous avons déjà obtenu. Prenons l'exemple fondamental du mouvement Brownien à une dimension. Supposons que notre monde est divisée en cases, que nous numérotions à l'aide d'un indice n . A chaque pas de temps, notre particule fait un saut vers la case immédiatement à sa gauche ou à sa droite. Nous nous demandons maintenant : quelle est la probabilité $P(n, t)$ de trouver la particule dans la case numéro n au temps t ?

Telle quelle, cette question n'a pas de sens. Il nous faut préciser où était la particule à l'instant initial; la question précise est donc quelle est la probabilité de trouver la particule en n au temps t sachant quelle était dans la case n_0 au temps t_0 . Nous notons cette probabilité par $P(n, t; n_0, t_0)$. Comme cette notation est un peu lourd, nous omettons parfois le n_0, t_0 mais gardons à l'esprit que ce terme est toujours présent.

La figure 3.1 représente 25 réalisations d'un tel processus : à l'instant initial $t = 0$, nous posons la particule en case $n = 0$ et nous le faisons évoluer de case en case pendant 100 tour par notre jeu probabiliste : nous tirons au pile ou face, si c'est face nous déplaçons la particule d'une case vers la gauche, si c'est pile nous déplaçons la particule d'une case vers la droite. Nous refaisons cette expérience 25 fois, chaque *trajectoire* représente une expérience. Évidemment, pour parler probabilité, nous avons besoins de beaucoup plus que 25 trajectoire. La figure 3.2 représente la même chose mais pour 500 trajectoires.

Nous pouvons maintenant nous donner un temps t , par exemple $t = 400$ et nous poser la question : Combien de trajectoires arrivent dans la case $n = 40$ à ce temps là ? Ce nombre relatif est bien sûr la probabilité $P(40, 400; 0, 0)$. Nous pouvons déplacer notre appareil de mesure du "nombre relatif de trajectoire" de case en case et constituer ainsi la loi de distribution au temps $t = 400$, $P(n, 400; 0, 0)$. Nous pouvons d'ailleurs faire plusieurs coupes temporelles à divers temps t_i et de la même manière constituer $P(n, t_i)$. La figure 3.3 montre, à partir de la figure ci-dessus, la construction

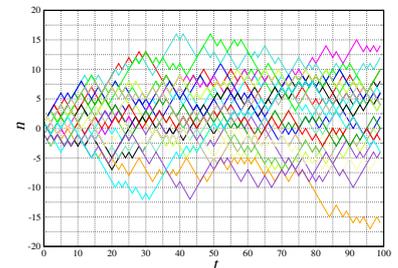


FIGURE 3.1 – 25 réalisations du mouvement brownien discret.

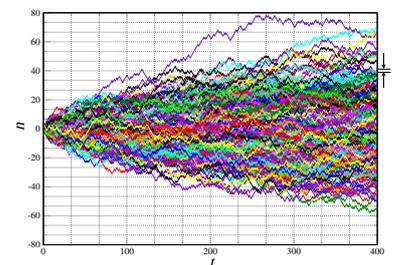


FIGURE 3.2 – 500 réalisations du mouvement brownien discret, pendant 400 tours. La proportion de trajectoires qui aboutissent à $n = 40$ au temps $t = 400$ est la probabilité $P(40, 400; 0, 0)$.

des probabilité pour 4 coupes temporelles.

. Nous verrons plus tard que cette fonction est assez proche d'une gaussienne.

Construire toute la loi de distribution à une coupe temporelle t_i est un processus un peu lourd, mais nous pouvons, à chaque coupe temporelle, extraire les moments statistiques, surtout les deux plus importants qui sont la moyenne et la variance : à une coupe temporelle donnée, nous enregistrons le point d'impact de chaque trajectoire sur la coupe et notons ces nombres n_1, n_2, \dots . Il suffit ensuite de calculer la moyenne et la variance de ces nombres. La figure 3.4 montre la moyenne et la variance en fonction du temps.

Ce que nous pouvons remarquer sur cette figure est que la moyenne reste autour de zéro : si à un instant, on photographie la scène et que l'on calcule la moyenne de la position des particule, ce nombre est proche de zéro. Par contre, nous observons que la variance croit linéairement en fonction du temps¹.

Nous reviendrons sur ces problèmes en détail, mais voyons comment on peut comprendre ces résultats rapidement. Soit D la variable aléatoire de déplacement à chaque tour. D prend la valeur ± 1 avec une probabilité de 0.5 ; sa moyenne est donc $\mu = \langle D \rangle = 0$ et sa variance est $\sigma^2 = \langle (D - \mu)^2 \rangle = 1$. La position de la particule est une variable aléatoire X qui au bout de T tours est la somme de T variables D indépendantes :

$$X(T) = \sum_{i=1}^T D_i$$

Nous savons que pour les moyennes

$$\langle X(T) \rangle = \left\langle \sum_{i=1}^T D_i \right\rangle = \sum_{i=1}^T \langle D_i \rangle = 0$$

Par ailleurs, comme les variables D_i sont indépendantes les variances sont additives :

$$\text{Var}(X(T)) = \text{Var}\left(\sum_{i=1}^T D_i\right) = \sum_{i=1}^T \text{Var}(D_i) = T\sigma^2 = T$$

Nous voyons donc que la moyenne reste nulle, tandis que la variance croit linéairement avec le temps. Nous voyons aussi que les moments statistiques peuvent se calculer facilement, en général beaucoup plus facilement que les probabilités.

Notons que toute sorte de généralisation son possible à partir de là. Par exemple, la variable aléatoire D_i de déplacement à chaque tour pourrait prendre les valeurs $-1, 0, 1$ avec les probabilité $q, 1 - (p + q), p$: la particule a une probabilité de rester sur place avec la probabilité $1 - (p + q)$ et le mouvement vers la gauche et la droite ne sont plus équiprobable. Cette fois, $\mu = p - q$ et $\sigma^2 = p + q - (p - q)^2$; la moyenne se déplace donc également dans le temps $X(T) = T(p - q)$, mais les résultats essentiels ne change pas : la variance croit linéairement avec le temps.

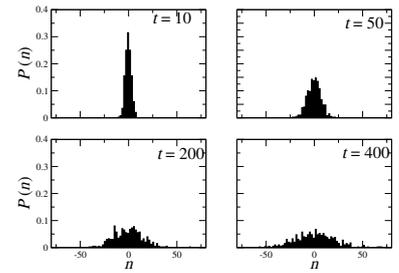


FIGURE 3.3 – Les probabilités $P(n, t_i)$ pour quatre temps $t_i = 10, 50, 200, 400$ obtenu à partir des trajectoire de la figure 3.2.

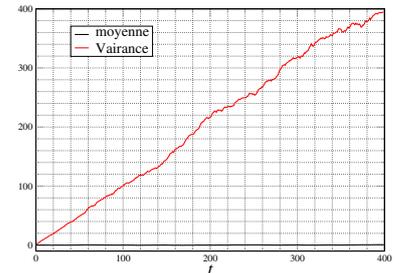


FIGURE 3.4 – Moyenne $\langle n(t) \rangle$ et variance $V(t)$ extraite des coupes temporelles des trajectoires de la figure 3.2.

1. bien sûr, nous n'avons que 400 trajectoires et la croissance est un peu fluctuante autour de la droite ; si vous refaites la même expériences numérique avec 10000 trajectoires, la croissance devient une droite presque parfaite.

§ 3.1 Obtenir la moyenne et la variance du mouvement brownien en fonction du temps si les sauts ne sont pas limités aux cases voisines : la variable aléatoire D rend la valeur m ($m \in \mathbb{Z}$) avec la probabilité $P(m)$.

§ 3.2 Obtenir la moyenne et la variance du mouvement brownien *avec persistance* en fonction du temps : les déplacements D_i ne sont pas indépendants, mais chaque déplacement D_i dépend du déplacement précédent :

$$D_i = (1 - p)D_{i-1} + p\zeta_i$$

où p est un paramètre fixé $\in [0, 1]$ et ζ_i est une variable aléatoire qui rend ± 1 avec la même probabilité.²

3.1.1 Le temps continu.

Dans l'exemple ci-dessus, le temps était traité comme une variable discrète qu'on énumérait avec un indice entier. Dans la plupart des situations physiques, le temps est une variable *continue* : le mouvement d'une particule, la rencontre de deux molécules, la mort d'un individu sont des variables aléatoires qui peuvent occuper n'importe quelle position temporelle. Le temps discret simplifie parfois les calculs, mais plus souvent il les complique. De toute façon, si notre description du modèle est correct, nous devons pouvoir passer du temps discret au temps continu sans modifier les résultats essentiels du problème : il suffit de prendre la limite du pas de temps de discrétisation tendant vers zéro.

Revenons à l'exemple ci-dessus du mouvement brownien essayons de le formuler en temps continu. Faisons l'expérience (par l'esprit) suivant : Donnons nous un temps infinitésimal dt , observons une case particulière n et comptons la proportion de particules qui sautent d'une case vers la gauche à partir de cette case. Il est évident que cette proportion dépend fortement de la durée dt : plus dt est court, moins les particules auront le temps d'opérer un saut. En nous remémorant notre discussion sur les variables aléatoires continues (voir §2.1), nous pouvons écrire la probabilité de ces sauts comme $W^-(n)dt$. $W^-(n)$ est une *densité* de probabilité de saut vers la gauche si la particule se trouve sur la case n . De la même manière, nous avons la densité de probabilité $W^+(n)$ des sauts vers la droite à partir de la case n . Dans la littérature, nous rencontrons ces quantités sous le nom de taux (ou probabilité) de transition. Nous pouvons bien sûr généraliser cela à des sauts de plusieurs cases et considérer par exemple $W(n, m)$, la densité de probabilité de sauter de m cases ($m \in \mathbb{Z}$) à partir de la case n . Nous verrons dans les prochaines sections que le temps continu se traite plus facilement que le temps discret.

3.1.2 Les états continus ?

La même démarche que nous avons effectuée pour le temps continu pourrait être généralisée aux états : au lieu de considérer les états discrets n , nous pouvons considérer les états continus x . La position d'une particule par exemple est une variable continue et

2. Ceci est un problème important par exemple pour les polymères qui peuvent être considérés comme des bâtonnets liés les uns aux autres par leurs extrémités : chaque bâtonnet tend à s'aligner sur le précédent, avec une déviation due à l'agitation thermique (voir par exemple B. Houchmandzadeh & S. Dimitrov, *Elastic measurements show the existence of thin rigid cores inside mitotic chromosomes*, J. Cell Biol. 145 :215-23 (1999)). De même pour le mouvement d'un micro organisme qui tend à conserver sa trajectoire mais est soumis à de petites déviations aléatoires (voir par exemple B. Houchmandzadeh, *Neutral Clustering in a Simple Experimental Ecological Community*, Phys. Rev. Lett. 101, 078103 (2008)).

non discrète. Nous allons cependant déferer cette généralisation au chapitre sur l'équation de diffusion. Il y a plusieurs raison à cela : d'abord, beaucoup de problèmes sont naturellement traité en état discret et non continu : la chimie à faible nombre de molécule, les organismes vivants (nous ne pouvons pas avoir 2.3 individu), ... Les phénomènes discrets occupent une très grande place. Ensuite, beaucoup de concepts fondamentaux sont facile à comprendre en formulation discrète et se généralisent ensuite intuitivement aux problèmes continus. Enfin, la formulation rigoureuse des problèmes continus alourdit extraordinairement L'énoncé des résultats. Il nous faut remarquer que le monde physique (et ses observables) est beaucoup plus simple que le monde mathématique et nous n'avons très souvent pas besoin des complications que les mathématiciens doivent gérer pour énoncer leurs problèmes en des termes les plus généraux³.

3.2 Définitions et généralisations.

Dans l'exemple de la section précédente, nous avons introduit la plupart des concepts des processus stochastiques. En probabilité classique, chaque réalisation (ou mesure, ou expérience) est un point (un état) que nous désignons par n . Nous généralisons ce concept et dorénavant, chaque réalisation est un trajectoire $n(t)$ (voir figure 3.5) . La probabilité $P(n, t; n_0, t_0)$ est⁴ celle d'occuper l'état n au temps t sachant qu'au temps t_0 , l'état occupé était n_0 . Cette quantité est le poids relatif des trajectoires qui passent par le point (t, n) .

Nous allons souvent omettre la condition initiale dans notre écriture pour alléger les notations, mais nous gardons à l'esprit qu'elle est toujours présente. Nous pouvons évaluer les quantités classiques calculées en probabilité en faisant des coupes temporelles. La moyenne au temps t est donné par exemple par

$$\langle n(t) \rangle = \sum_n n P(n, t)$$

Définition qui se généralise naturellement à tous les moments statistiques :

$$\langle f(n(t)) \rangle = \sum_n f(n) P(n, t)$$

Pour évaluer $P(n, t)$, nous avons à notre dispositions les taux de transitions $W(n, m)$, la densité de probabilité d'effectuer un saut de taille m si nous nous trouvons au point n (Figure 3.6). Ainsi, $W(n, m)dt$ est la probabilité de se trouver en $n + m$ au temps $t + dt$ si nous étions en n au temps t .

Comme dans n'importe quelle approche différentielle, ces taux de transitions microscopique nous donnent accès, via une équation différentielle, aux probabilités $P(n, t)$. Pour cela, il suffit d'appliquer les règles de ET et de OU logique. Restreignons nous pour l'instant aux saut de ± 1 , et allégeons la notation en posant $W^\pm(n) = W(n, \pm 1)$. On peut aboutir à l'état n au temps $t + dt$ de

3. A titre d'exemple, considérons le passage des intégrales de Riemann en 1850 aux intégrales de Lebesgue en 1902 : il a fallu 50 ans d'efforts aux mathématiciens pour pouvoir étendre la notion d'intégrabilité aux fonctions étranges comme celle de Riemann (discontinue en une infinité dénombrable de points) ou de Volterra (différentiable partout, à différentielle bornée, mais non intégrable au sens de Riemann). Pourtant, ces fonctions ne se rencontrent pas dans le monde physique et nous avions continuent de voler en nous contentant de l'intégrale de Riemann. Kronecker avait lancé à Cantor, quand celui-ci se battait avec le cardinal de l'ensemble \mathbb{R} et l'hypothèse du continu : "pourquoi vous vous occupez tant des nombres réelles ? Ne perdez pas votre temps, il n'y a que les nombres rationnelles qui existent".
4. plus exactement, c'est la densité de probabilité de se trouver en l'état n au temps t ; la probabilité de se trouver en n entre les temps t et $t + dt$ est cette quantité multiplié par dt .

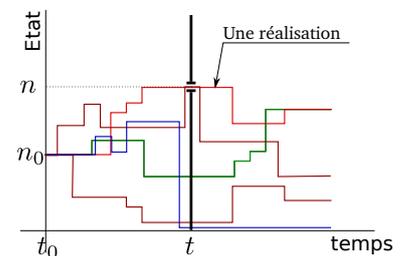


FIGURE 3.5 – Chaque réalisation est un trajectoire ; $P(n, t; n_0, t_0)$ est la proportion des trajectoires passant par (t_0, n_0) qui passent également par (t, n) .

trois façons différentes : OU nous étions en $n \pm 1$ à l'instant t ET nous avons effectué un saut de ∓ 1 ; OU nous étions en n à l'instant t et nous n'avons effectué aucun saut. Comme ces trois possibilités sont mutuellement exclusives, les OU se traduisent par des sommes de probabilité. Les événements à l'intérieur de chaque probabilité étant indépendants, les ET se traduisent pas des produits de probabilité. En résumé, nous avons donc

$$\begin{aligned} P(n, t + dt) &= P(n + 1, t)W^-(n + 1)dt \\ &+ P(n - 1, t)W^+(n - 1)dt \\ &+ P(n, t) (1 - (W^+(n) + W^-(n)) dt) \end{aligned}$$

En développant $P(n, t + dt)$ à l'ordre 1 en dt et en réarrangeant les termes, nous obtenons alors

$$\begin{aligned} \frac{\partial P(n, t)}{\partial t} &= W^+(n - 1)P(n - 1, t) - W^+(n)P(n, t) \\ &+ W^-(n + 1)P(n + 1, t) - W^-(n)P(n, t) \end{aligned} \quad (3.1)$$

L'équation ci-dessus s'appelle une équation Maître⁵. Nous pouvons l'imager simplement comme une équation de balance : la variation de probabilité par unité de temps dans l'état n est égale au flux entrant moins le flux sortant. La première ligne de l'équation ci-dessus est le flux entrant par la gauche et sortant par la droite ; la deuxième ligne désigne le flux dans l'autre sens (voir Figure 3.7).

Il est facile à partir de là de généraliser quand toutes les longueurs de saut m sont permises : en écrivant la même balance entre les flux, nous obtenons une équation légèrement plus compliquée :

$$\frac{\partial P(n, t)}{\partial t} = \sum_m W(n - m, m)P(n - m, t) - W(n, m)P(n, t) \quad (3.2)$$

L'équation ci-dessus mesure la balance entre les flux de sauts de taille m : en sommant sur tous ces flux, (vers et à partir de l'état n), on trouve ce qui reste (ou quitte) l'état n .

Nous pouvons aussi réécrire cette équation directement par balance de flux entre deux états n et m quelconque (sans nous soucier de la taille des sauts) :

$$\frac{\partial P(n, t)}{\partial t} = \sum_m W(m \rightarrow n)P(m, t) - W(n \rightarrow m)P(n, t) \quad (3.3)$$

Et en collectant les $P(n, t)$ dans un vecteur colonne (indexé par n) $|P(t)\rangle = (P(0, t), P(1, t), \dots)^T$, de façon encore plus compact nous écrivons

$$\partial_t |P(t)\rangle = \mathcal{L} |P(t)\rangle$$

où les coefficients de la matrice \mathcal{L} sont donnés par

$$\mathcal{L}_m^n = W(m \rightarrow n) ; \quad \mathcal{L}_n^n = - \sum_m W(n \rightarrow m)$$

Dorénavant, nous étudierons essentiellement les problèmes où les sauts sont limités à ± 1 , la généralisation étant souvent triviale.

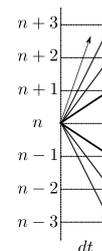


FIGURE 3.6 – La probabilité de transition $W(n \rightarrow n + m)dt$ est la probabilité de se trouver en $n + m$ au temps $t + dt$ sachant que le système était en n au temps t

5. En physique statistique, si $\partial_t P(n, t) = 0$ pour tout n , nous disons que nous avons atteint l'équilibre, ce que par ailleurs nous appelons état stationnaire. En physique, on assume qu'à l'équilibre, nous avons une *balance détaillée*, c'est à dire que le flux entre deux cases quelconques se compensent toujours exactement. On peut cependant facilement construire des états stationnaires qui ne respectent pas la balance détaillée. Voir par exemple, B. Morrisey, *Microscopic reversibility and detailed balance*, Journal of Chemical Education, 52 :296 (1975).

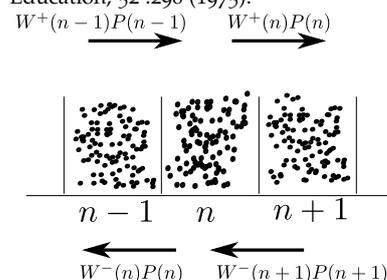


FIGURE 3.7 – L'équation maîtresse vue comme une balance de flux entre les états.

Exemple 3.1 le processus de Poisson.

Les processus de Poisson⁶ est le socle fondamental de notre compréhension des processus stochastiques. Ce sont les processus que l'on rencontre le plus souvent : le nombre de particule radioactives qui se désintègrent pendant un temps T , le nombre de communication reçu par un central pendant une période ou de réaction chimique, ... La loi de poisson se rencontre par tout et il est important de bien le comprendre. Prenons l'exemple des gens qui arrivent dans une queue : les arrivées sont indépendantes les uns des autres et nous excluons les arrivées groupées pour l'instant. Supposons que la probabilité d'occurrence d'un événement (une arrivée, une naissance, une désintégration radioactive,...) pendant un intervalle de temps dt infinitésimal soit αdt ⁷. Quelle est la probabilité d'observer n évènements pendant un temps t ?

Dans notre langage de processus stochastique, nous avons la condition initiale $t_0 = 0, n_0 = 0, W^+(n) = \lambda$ (la probabilité d'occurrence d'un événement ne dépend pas de combien d'évènement nous avons déjà observé), $W^-(n) = 0$ (nous n'avons pas d'anti-évènement), et donc

$$\frac{\partial P(n,t)}{\partial t} = \alpha (P(n-1,t) - P(n,t)) \quad n \geq 1 \quad (3.4)$$

$$\frac{\partial P(0,t)}{\partial t} = -\alpha P(0,t) \quad (3.5)$$

Insistons que ceci constitue une infinité d'équation différentielles couplées, une équation pour chaque état $n = 0, 1, 2, \dots$. Il n'est pas difficile de vérifier que la solution de cette équation est donnée par

$$P(n,t) = e^{-\alpha t} \frac{(\alpha t)^n}{n!} \quad (3.6)$$

Si l'intervalle de temps est fixé, on pose souvent $\lambda = \alpha t$ et la loi nous donne la probabilité d'observer n évènements pendant ce temps sous la forme classique

$$P(n) = e^{-\lambda} \lambda^n / n!$$

Nous pouvons également nous poser la question de la de probabilité du temps entre deux évènements. Dans ce cas, nous posons $n = 0$ et nous avons la loi de Poisson temporelle

$$P(t) = e^{-\alpha t}$$

Ceci est la probabilité qu'aucun événement n'arrive dans l'intervalle $[0, t]$. Nous rencontrons parfois cette loi sous une forme légèrement différente : quelle est la *densité* de probabilité qu'un événement arrive exactement dans l'intervalle de temps $[t, t + dt]$? Ceci est la probabilité pour qu'aucun événement n'arrive pendant $[0, t]$, et qu'un événement arrive entre $[t, t + dt]$, ce qui nous donne $\exp(-\alpha t)\alpha dt$. La densité de cette probabilité est donc

$$p(t) = \alpha e^{-\alpha t}$$

6. Siméon Denis Poisson, auteur prolifique dans deux nombreux disciplines de physique/mathématiques. Cette loi de distribution apparaît dans son livre "Recherches sur la probabilité des jugements en matières criminelles et matière civile" (1838) disponible sur Europeana.

7. Nous voyons ici pourquoi nous écartons la probabilité d'observer deux évènements pendant un temps infinitésimal dt : sa probabilité est de l'ordre de dt^2 , donc négligeable.

Ce sont les deux aspects du même processus stochastique.

Les techniques par lesquelles nous résolvons effectivement les équations maîtresses ou au moins en extrayons des quantités intéressantes sont traités au prochain chapitre. Cependant, les équations (3.4-3.5) sont suffisamment facile pour les résoudre directement. Notons d'abord que la condition initiale est $P(n, 0) = \delta_{n,0}$. La transformée de Laplace des équations ci-dessus ($t \rightarrow s, P(n, t) \rightarrow \tilde{P}(n, s)$) nous donne alors

$$\begin{aligned}(s + \alpha)\tilde{P}(0, s) &= 1 \\(s + \alpha)\tilde{P}(1, s) &= \alpha\tilde{P}(0, s) \\(s + \alpha)\tilde{P}(2, s) &= \alpha\tilde{P}(1, s) \\&\dots\end{aligned}$$

La récurrence se résout sans problème

$$\tilde{P}(n, s) = \frac{\alpha^n}{(s + \alpha)^{n+1}}$$

dont la transformée de Laplace inverse est donnée par l'équation (3.6).

4

L'équation Maîtresse à travers des exemples.

4.1 Quelques exemples de processus stochastique.

Voyons maintenant quelques prototypes de processus stochastiques. Après les avoir énoncer, nous verrons quelques techniques pour extraire des résultats concret à partir de ces équations.

Exemple 4.1 Le mouvement brownien.

Le mouvement Brownien, la danse aléatoire des particules de pollen observée par M. Brown dans les années 1830, est devenu un des thèmes les plus fondamental de la physique-mathématique depuis les années 1900. Au fur et à mesure que les scientifiques ont essayé de décrire les lois de distributions de ce processus stochastique, il est devenu claire qu'on le rencontre partout. Trois articles fondamentaux sont apparus sur le sujet entre 1905 et 1907 par Einstein, Smoluchowski et Langevin donnant différent traitement mathématiques de ce mouvement et établissant le lien profond entre cette *danse* et l'existence des molécules (ce qu'on appelait encore à l'époque l'hypothèse atomique).

De nombreux travaux ont suivi, comme ceux de Perrin et l'observation expérimentales précise du mouvement Brownien en 1909¹ et N. Wiener qui les a utilisé pour formuler rigoureusement la théorie des processus stochastique continu, travaux qui ont servi ensuite à Otto de formuler dans les années 1950 la théorie des équations différentielles stochastiques. De nos jours une médaille Fields est décernée régulièrement à quelqu'un travaillant sur ce sujet : en effet, le mouvement Brownien pose les même problèmes de mesure dans l'espace des fonctions que les mathématiciens affrontaient dans les années 1850 dans l'espace \mathbb{R} . Comme cela a un lien profond avec la mécanique quantique et sa formulation à travers les intégrales de chemins, une formulation rigoureuse intéresse les physiciens au plus haut degrés.

Il y a de nombreuse formulation du mouvement Brownien, toutes identiques quant à leurs résultats. Nous allons ici étudier le mouvement Brownien d'une particule à une dimension où le temps est continu et l'espace discret. Nous verrons par la suite les généralisations qui nous ferons dépasser ces contraintes,

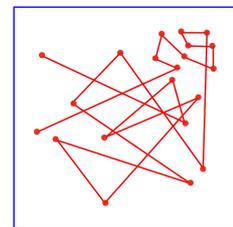


FIGURE 4.1 – Observation d'une particule Brownien au cours du temps.

1. Voir J. Perrin, Mouvement Brownien et réalité moléculaire, Annales de Physique et de Chimie, 1909 (Ces Annales sont disponible librement sur la toile).

mais tous les résultats essentiels seront rencontrés dans le cas le plus simple. Considérons l'espace unidimensionnel discret, divisé en cellules (ou cases) indexé par n ($n \in \mathbb{Z}$). Une particule effectue des saut aléatoires d'une case vers la gauche ou la droite avec la même densité probabilité α . Dans ce problème, les états du système sont les positions n , et $P(n, t)$ est la (densité de) probabilité de trouver la particule à la position n au temps t .

$$W^+(n) = W^-(n) = \alpha$$

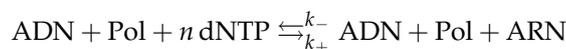
Noter que les taux de transition ne dépendent pas de la position n .

Exemple 4.2 Une équation chimique simple : la transcription de l'ARN.

Les réactions chimiques sont par nature des processus stochastiques. Dans la réaction simple $A + B \rightarrow C$, la rencontre entre les deux molécules A et B , la réaction chimique une fois la collision ayant eu lieu, la décomposition de la molécule C sont tous des phénomènes aléatoires. En chimie macroscopique, le nombre de molécule est très important ($\sim 10^{23}$) et le théorème central limite nous indique que les fluctuations sont faibles; c'est pourquoi en cinétique chimique classique, ces réactions sont traitées par des équations différentielles (et donc déterministes).

Par contre, quand les réactions ont lieu avec un petit nombre de molécule dans un volume réduit, ce qui est typiquement le cas de la chimie dans les cellules vivantes, le vent stochastique prend le dessus. Prenons l'exemple de la bactérie *E. Coli* : sa capacité à digérer le sucre lactose est le résultat d'une série d'équations chimiques enchaînées. Si on fait croître un nombre important de bactérie à partir du même clone et qu'on les place ensuite dans un milieu à une certaine concentration de lactose, on observe qu'une proportion des bactérie est capable de digérer le lactose et une autre proportion non². Pour pouvoir correctement déduire l'état des bactérie, nous devons plutôt obtenir la probabilité $P(c, t)$, la probabilité que l'enzyme responsable de la digestion soit à la concentration c au temps t . Le fait que les cellules vivantes ne sont pas des machines déterministes mais des machines probabilistes, et cela à cause de la chimie à faible nombre de molécule, fait partie intégrante de notre vision du monde vivant depuis les années 2000.

Prenons le cas le plus simple de la transcription de l'ARN à partir de l'ADN. La protéine ARN-Polymérase s'accroche à la partie "promoteur" de l'ADN, parcourt l'ADN et synthétise une molécule d'ARN à partir des dNTP (voir figure 4.3)³. Cette réaction chimique s'écrit



Comme l'ADN et le polymérase se retrouvent des deux côtés de la réaction inchangés et en supposant que la quantité de dNTP

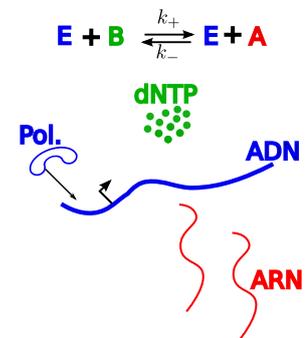
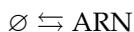


FIGURE 4.2 – Transcription d'ARN. Dans la réaction chimique ci-dessus, E représente l'ADN et le polymérase (en bleu), B les dNTP (en vert) et A l'ARN (rouge).

2. Le phénomène a été décrit par Novick et Wiener dans les années 1950 et a reçu une nouvelle démonstration à l'aide des outils modernes de la biologie moléculaire dans les années 2000. Voir par exemple (i) NOVICK, A. and WEINER, M. *Enzyme induction as an all-or-none phenomenon*. Proc. Natl. Acad. Sci. USA 43 : 553-566 (1957); (ii) Raj A, van Oudenaarden A. *Nature, nurture, or chance: stochastic gene expression and its consequences*, Cell, 135 :216-226 (2008).

3. Voir par exemple I. Golding *et al.*, *Real-Time Kinetics of Gene Activity in Individual Bacteria*, Cell 123 :1025-1036 (2005) où les auteurs ont expérimentalement construit la courbe $P(n, t)$ en observant les molécules d'ARN individuellement.

est tellement saturante que les réactions de transcription ne la changent pas, nous pouvons simplifier cette réaction en l'écrivant comme la synthèse ex-nihilo d'ARN :



qui nous servira de prototype de réaction chimique enzymatique. Plus tard, nous compliquerons ce schéma à souhait. L'état du système est donné par le nombre n de molécule d'ARN présente. Le taux de synthèse de la molécule ne dépend que des rencontre Pol-ADN et nous le prenons constante :

$$W^+(n) = \alpha$$

Les molécules d'ARN sont dégradé à leur tour par d'autre enzymes, et la densité de probabilité pour une molécule d'être dégradé est une constante que nous appellerons μ . Quelle est la probabilité, si on observe n molécules, d'observer une dégradation? Ceci est la probabilité de voir la molécule 1 se dégrader, OU la molécule 2, OU la molécule 3, ... Nous avons des OU logiques d'évènement mutuellement exclusif, leurs probabilités s'ajoutent donc et

$$W^-(n) = \mu n$$

Nous compliquerons se schéma à souhait, mais notons que nous avons là des taux de transition qui dépendent de l'état dans lequel on se trouve.

Exemple 4.3 La dynamique stochastique des populations.

Prédire la dynamique des populations est une tâche primordiale pour l'écologiste qui veut comprendre la nature ou aider la commission européenne à établir les quotas de pêche. Les écologistes ont l'habitude de traiter ces problèmes à l'aide d'équation différentielle, mais cette approche peut s'avérer grossièrement fausse. Dès le milieu du XIXème siècle par exemple, les navigateurs avait remarqué que les planctons⁴ apparaissaient en "tâches" sur l'océan et leur distribution n'était pas homogène. Pourtant, l'eau est un milieu liquide et les plancton devraient pouvoir diffuser et leur concentration se homogénéiser. Ceci est pourtant pas le cas et tout les être vivant ont tendance à s'agglomérer. Ceci est un effet subtil des morts et naissances, qui change le nombre d'individu par pallier discret : ont passe de 2 à trois individu par exemple, mais jamais à 2.3 ou à 1.8.

Oublions la distribution spatiale pour l'instant et considérons un système à zéro dimension : par exemple, une boite de pétri dans laquelle on compte le nombre de bactérie n à chaque instant. L'état du système est donné par le nombre d'individus présents que nous appellerons n . A chaque instant, chaque individu a une densité de probabilité μ de mourir et α de donner naissance à un descendant. Dans ce cas,

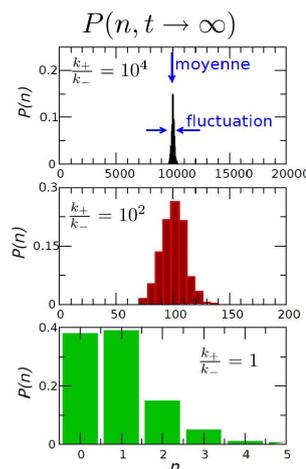


FIGURE 4.3 – Quand le nombre de molécules est important (cas $k_+/k_- = 10^4$), les fluctuations sont faible et la dynamique peut se traiter par des équation déterministe. Quand par contre le nombre moyen de molécule d'ARN est faible ($k_+/k_- \approx 1$), le système doit se traiter par les probabilités.

4. des micro-organisme marins, pour certains à la base de la chaîne alimentaire dans les océan. Leur présence change la coloration de l'eau en vert et cette couleur peut-être utilisée pour estimer leur densité.

$$W^+(n) = \alpha n$$

$$W^-(n) = \mu n$$

Noter que nous avons beaucoup simplifié : en général, les ressources sont limitées et le taux de naissance doit décroître quand la population croît; un individu ne peut pas se multiplier immédiatement après une division. Le taux de mort dépend de l'âge de chaque individu. Nous pouvons bien sûr inclure tous ces phénomènes au fur et à mesure, mais contentons nous pour l'instant de ce premier étape, qui a une grande pertinence en écologie.

Que peut on dire par exemple de ce système si $\alpha = \mu$? On s'attend à priori que les morts et les naissances se compensent et que les probabilités soit distribuées autour d'une moyenne. Ceci est loin d'être le cas, ce système ne peut pas atteindre un régime stationnaire et nous aurons une très forte densité dans certaines boites de pétri, tandis que d'autres resterons vide⁵ (figure 4.4).

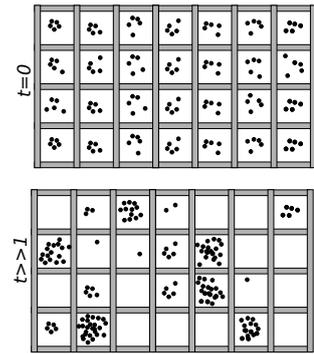


FIGURE 4.4 – Évolution de nombre de bactérie par les morts et naissances aléatoires dans M boites de pétri. A l'instant initial, chaque boîte comporte le même nombre d'individu; au fur et à mesure que le temps s'écoule, certaines boîtes s'enrichissent fortement tandis que d'autres deviennent vide.

Exemple 4.4 La marche vers l'équilibre : le modèle d'Ehrenfest.

Les systèmes physique à grand nombre de particules atteignent toujours l'équilibre thermodynamique : Les réactions chimique, la diffusion des gaz d'une pièce vers l'autre, ... Toute la science de la thermodynamique est dédié à l'étude de cet équilibre. Mais pourquoi atteint on l'équilibre? Les équations microscopique du mouvement des particules sont réversible dans le temps : on ne peut pas prédire dans quel sens coule le temps en regardant simplement un film de mouvement de particules à l'échelle microscopique. Mais regardons la courbe de température d'un gigot qu'on sort du four, et on saura que "chaud" est le passé et "froid" le futur. L'origine de cette irréversibilité a consommé (et consomme toujours) des générations de scientifiques. Nous avons déjà mentionné le calcul de Boltzmann et l'hypothèse du chaos moléculaire. Cependant, ces calculs sont un peu ardu⁶ et on peut avoir l'impression que nous n'avons pas compris le fond des choses : il faut des "toy models" plus simple pour mettre en évidence l'origine de l'irréversibilité. Beaucoup de "toy models" ont été inventé par les physiciens, comme le très célèbre modèle de Pasta Fermi Ulam dans les années 1950 qui a été la première expérience numérique.

Dans les années 1930, Ehrenfest a imaginé un modèle très simple : imaginez N cartes numéroté de 1 à N distribué en deux piles (figure 4.5). De façons aléatoires, nous tirons un numéro et nous déplaçons la carte correspondante de la pile où elle se trouve vers l'autre pile. Nous pouvons décrire l'état du système par le nombre n de cartes qui se trouvent sur la pile de gauche, $N - n$ sera alors le nombre de cartes sur la pile de droite. Nous sommes donc intéresser par la probabilité $P(n, t)$ de trouver n cartes sur la pile de gauche.

5. Voir B. Houchmandzadeh, *Neutral Clustering in a Simple Experimental Ecological Community*, Phys. Rev. Lett. 101, 078103 (2008) pour les expériences sur les écosystèmes modèles, et B. Houchmandzadeh, *Theory of neutral clustering for growing populations*, Phys. Rev. E. 80, 051920 (2009) pour le traitement théorique complet, incluant la diffusion spatiale.

6. Nous les verrons un peu plus loin.

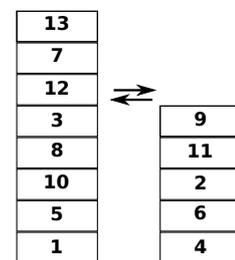


FIGURE 4.5 – L'urne d'Ehrenfest

Si à l'instant initial, nous avons mis toutes les cartes sur une des piles, comment le système évolue ? Y a-t'il un équilibre à atteindre $P(n, t) \lim_{t \rightarrow \infty} P_s(n)$? Comment évolue l'entropie du système, définie comme

$$S(t) = - \sum P(n, t) \log (P(n, t))$$

D'après notre définition du système, les taux de transition pour ce modèle s'écrivent

$$\begin{aligned} W^+(n) &= \alpha(N - n) \\ W^-(n) &= \alpha n \end{aligned}$$

Exemple 4.5 Évolution Darwinienne.

La théorie de l'Évolution a été énoncé dans les années 1850 par Darwin et Wallace. Ses points faibles ont été comblés après sa fusion avec la génétique Mendélienne dans les années 1920 par les pères fondateurs de la synthèse moderne que sont Fisher, Wright, Malécot. Les processus stochastiques y ont été intégré à ce moment et des solutions efficace y ont été amené par Kimura à la fin des années 1950. Dans sa première formulation dogmatique, l'évolution était vu comme une "marche vers le progrès", comme si les scientifiques avaient remplacé le mot Dieu par le mot Évolution, mais avaient conservé les mêmes concepts ; même de nos jours, on voit des traces de ce concept dans une certaine littérature.

Le principe de l'évolution comme processus stochastique est simple : l'écosystème peut supporter N individus ; chaque individu produit plus de descendants que nécessaire, et le surplus par rapport à N est éliminé. Si un individu "découvre" un gène qui donne un léger avantage à ses descendants d'être présent dans la prochaine génération, ce gène a une certaine probabilité de se répandre (le nombre d'individu le portant d'augmenter). Toute la subtilité est là : "a une certaine probabilité de se répandre" et non "se répandra". Comme on va le voir, on peut avoir des gènes neutres ou légèrement maléfique qui auront également une certaine chance de se répandre. Le tout est de pouvoir estimer ces probabilités.

Dans la version originale, Fisher et Wright considéraient des populations sexués où les parents , ainsi que le surplus des descendants mourraient à chaque génération. Le modèle était inutilement compliqué : nous avons vu que nous pouvons toujours remplacer les modèles "par tour" par des modèles "temps continus" ; nous allons de plus pour l'instant laisser tomber la reproduction sexué, l'essentiel du vivant étant constitué d'organisme qui se reproduisent de façon asexué. Le modèle d'évolution qui en résulte a été proposé par Moran dans les années 1960 (figure 4.6) : considérons N individu dans un écosystème qui supporte exactement N individu. Supposons que ces individus portent différent allèle du même gène. Les individus meurent avec un certain taux μ ; quand un individu meurent, il est immédiatement

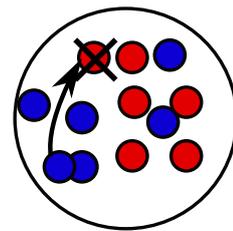


FIGURE 4.6 – Le modèle de Moran de l'évolution : quand un individu meure, il est remplacé par un autre tiré au hasard dans la population.

remplacé par le descendant d'un autre individu choisi au hasard. Nous supposons que tous les allèles sont similaires, excepté un qu'on appellera A et qui confère aux individus qui le portent un léger avantage $(1 + s)$ à se dupliquer quand quelqu'un meurent (s peut être négatif, ce sera alors un désavantage).

L'état du système est caractérisé par le nombre n des A -individus, $N - n$ étant le nombre de tous les autres. La probabilité pour que les A diminuent leur nombre de 1 pendant le temps dt est la probabilité qu'un A meure (μndt) ET que ce soit un non- A qui le remplace : $(N - n)/N$. La probabilité pour que les A augmentent leur nombre de 1 pendant le temps dt est la probabilité qu'un non- A meure (μndt) ET que ce soit un A qui le remplace : $(1 + s)(N - n)/N$. On peut résumer tout cela par

$$\begin{aligned} W^-(n) &= \mu n(N - n)/N \\ W^+(n) &= \mu(1 + s)n(N - n)/N \end{aligned}$$

Nous remarquons que si n tombe à zéro, l'allèle A disparaît définitivement du système; d'un autre côté, si n atteint N , tous les autres allèles sont éliminés et le nombre n restera toujours à N : on dira que A est devenu dominant.

La question que l'on se pose est le suivant : si à l'instant initial, $n = n_0$, quel est la probabilité pour A de disparaître ou de devenir dominant? Quelle est l'échelle de temps sur lequel cela se passe? La réponse à ces questions a rendu un statut de vraie théorie quantitative à l'Évolution.

4.2 Manipulation de l'équation maîtresse.

Dans la section précédente, les exemples précédents étaient choisis pour exhiber des taux de transition $W^\pm(n)$ polynomial en fonction de n , et de degrés 0, 1 ou 2. Une fois que nous savons maîtriser ces exemples, on peut les compliquer à souhait.

Revenons à nos exemples. On peut bien sûr tenter de les résoudre directement, mais ceci est un peu compliqué : une équation maîtresse est en fait une infinité d'équations différentielles couplées, une équation pour chaque $P(n)$. On peut l'écrire de façon matricielle comme

$$\frac{d}{dt} \begin{pmatrix} \dots \\ P(-1, t) \\ P(0, t) \\ P(1, t) \\ \dots \end{pmatrix} = A \begin{pmatrix} \dots \\ P(-1, t) \\ P(0, t) \\ P(1, t) \\ \dots \end{pmatrix}$$

où A est une matrice, contenant les taux de transfert. Les vecteurs et la matrice A sont bien sûr infini; ceci dit, dans tous les problèmes ci-dessus, les transitions ne sont qu'entre états voisins et la matrice A est donc tri-diagonal. Avec un peu d'entraînement, nous apprenons à diagonaliser certaines de ces matrices et résoudre directement notre problème, mais cela reste en général compliqué.

Dans beaucoup de problèmes stochastiques cependant, ce qui nous intéresse de premier abord est l'évolution de la moyenne et de la variance en fonction du temps. Si on connaissait ces quantités, nous saurions énormément de chose sur le système. Or souvent⁷, nous pouvons extraire ces quantités directement de l'équation maîtresse sans même la résoudre.

7. Tant que les taux sont des polynômes de degrés 1 au plus.

4.2.1 La moyenne.

Reprenons notre équation maîtresse générale, qui est, comme nous l'avons dit, une infinité d'équations couplées :

$$\begin{aligned} \frac{\partial P(n,t)}{\partial t} &= W^+(n-1)P(n-1,t) - W^+(n)P(n,t) \\ &+ W^-(n+1)P(n+1,t) - W^-(n)P(n,t) \end{aligned} \quad (4.1)$$

Les quantités qui nous intéressent sont par exemple $\langle n(t) \rangle = \sum_n nP(n,t)$. Multiplions les deux côtés de l'équation ci-dessus par n et sommons sur n . A gauche, nous avons le terme $\sum_n \partial_t nP(n,t)$, et nous acceptons que nous pouvons échanger l'ordre de la dérivation par rapport au temps et sommation sur n . Le terme de gauche devient alors tout simplement

$$\frac{d}{dt} \sum_n nP(n,t) = \frac{d}{dt} \langle n(t) \rangle$$

ou dit autrement, la dérivée temporelle de la moyenne. Considérons maintenant un des termes de droite, par exemple

$$\sum_n nW^+(n-1)P(n-1,t)$$

Si dans cette somme, nous changeons d'indice et que nous effectuons le changement d'indice $n \rightarrow n+1$, ce terme devient

$$\sum_n (n+1)W^+(n)P(n,t)$$

Or, dans l'équation (4.1), nous avons déjà un terme en $-\sum_n nW^+(n)P(n,t)$, le bilan de la première ligne est donc simplement $\sum_n W^+(n)P(n,t)$.

En faisant la même chose pour la deuxième ligne, nous aboutissons à une équation différentielle ordinaire pour la moyenne

$$\frac{d}{dt} \langle n(t) \rangle = \sum_n (W^+(n) - W^-(n)) P(n,t) \quad (4.2)$$

Si les taux de transition sont des polynômes de degrés au plus 1, le terme de droite s'exprime également en terme de moyenne et nous avons une équation différentielle bien définie. Nous pouvons quelque peu alléger les notations et d'après ce que nous avons dit à la section §2.1, noter l'équation ci-dessus

$$\frac{d \langle n(t) \rangle}{dt} = \langle W^+(n) - W^-(n) \rangle \quad (4.3)$$

On traitera l'ensemble des exemples de la section précédente par cette méthode, mais un peu de patience pour l'instant. Nous pouvons dire que nous avons extrait la moyenne en appliquant l'opérateur $\sum_n n$ à l'équation (4.1).

4.2.2 *La variance.*

Si nous voulons calculer la variance, nous avons d'abord besoin du moment d'ordre 2 $\langle n^2(t) \rangle = \sum_n n^2 P(n, t)$. En suivant exactement la même démarche que ci-dessus, mais en appliquant l'opérateur $\sum_n n^2$ à l'équation (4.1), nous obtenons :

$$\frac{d\langle n^2 \rangle}{dt} = 2\langle n(W^+(n) - W^-(n)) \rangle + \langle W^+(n) + W^-(n) \rangle \quad (4.4)$$

Nous pouvons même pousser un peu plus loin : comme la variance $V(t) = \langle n^2(t) \rangle - \langle n(t) \rangle^2$, nous avons

$$\frac{dV}{dt} = \frac{d\langle n^2 \rangle}{dt} - 2\langle n(t) \rangle \frac{d\langle n(t) \rangle}{dt}$$

et en tirant les deux expressions de $\langle n \rangle$ et $\langle n^2 \rangle$ des relation ci-dessus, aboutir directement à une équation sur la variance :

$$\frac{dV}{dt} = 2\langle (n - \langle n \rangle)(W^+(n) - W^-(n)) \rangle + \langle W^+(n) + W^-(n) \rangle \quad (4.5)$$

4.2.3 *Fonctions caractéristique et génératrice des probabilités.*

Nous avons calculé $\langle n(t) \rangle$ en appliquant $\sum_n n$ et $\langle n^2(t) \rangle$ en appliquant l'opérateur $\sum_n n^2$. Il ne faut pas être grand clerc pour généraliser cela au calcul de la moyenne d'une fonction quelconque $f(n)$ en appliquant l'opérateur $\sum_n f(n)$, et nous obtenons alors

$$\frac{d\langle f(n) \rangle}{dt} = \langle (f(n+1) - f(n))W^+(n) \rangle - \langle (f(n) - f(n-1))W^-(n) \rangle \quad (4.6)$$

Supposons que nous avons choisi $f(n) = e^{ins}$ où s est un paramètre quelconque. Posons

$$\phi(s, t) = \langle f(n) \rangle = \sum_n e^{ins} P(n, t)$$

le terme de droite de l'équation (4.6) est donc $\partial_t \phi(s, t)$. Si nous pouvons arranger les termes de droite pour qu'ils soient également exprimable en fonction de ϕ , nous avons touché le jackpot : en effet, nous voyons que les quantités $P(n)$ ne sont que les coefficients de Fourier de la fonction $\phi(s)$ et par conséquent :

$$P(n, t) = \frac{1}{2\pi} \int_0^{2\pi} \phi(s, t) e^{-ins} ds$$

Si nous connaissons la fonction ϕ , nous avons en réalité résolu complètement notre problème. La fonction ϕ s'appelle la fonction caractéristique.

Enfin, parfois il est plus naturel de choisir $f(n) = z^n$, où z est un paramètre quelconque. Ce choix est judicieux par exemple quand $n \geq 0$. Le calcul est strictement équivalent à la fonction caractéristique, mais la fonction

$$\psi(z, t) = \sum_n z^n P(n, t)$$

qui s'appelle la transformée en z ou la fonction génératrice des probabilités, peut-être plus facilement interprétable. En regardant la relation ci-dessus comme un développement de Taylor, nous obtenons immédiatement l'inversion :

$$P(n, t) = \frac{1}{n!} \left. \frac{\partial^n \psi}{\partial z^n} \right|_{z=0}$$

Si n peut également négatif, il faut alors regarder la définition de ψ comme un développement de Laurent et les formules d'inversion sont alors similaires aux transformées de Fourier.

Note sur les états absorbants.

4.3 Applications aux exemples choisis.

Nous allons maintenant passer les exemples des sections précédentes à la moulinette des techniques vu plus haut.

Exemple 4.6 Le processus de Poisson.

Commençons par notre processus le plus connu, où nous avons vu que les taux de transition vers le haut et le bas sont donnés par

$$W^+(n) = \alpha ; W^-(n) = 0$$

et ne dépendent pas de l'état n . D'après ce que nous avons dit (eq. 4.2),

$$\frac{d \langle n(t) \rangle}{dt} = \langle \alpha \rangle = \alpha$$

et donc la moyenne croît linéairement avec le temps⁸. Comme à l'instant initial, $n = 0$

$$\langle n(t) \rangle = \alpha t$$

Le même calcul donne pour le moment d'ordre 2

$$\frac{d \langle n^2(t) \rangle}{dt} = 2 \langle \alpha n \rangle + \langle \alpha \rangle = 2\alpha \langle n \rangle + \alpha$$

Or, nous avons déjà calculer $\langle n \rangle$, il suffit de remplacer pour obtenir

$$\frac{d \langle n^2(t) \rangle}{dt} = 2\alpha^2 t + \alpha$$

et comme $\langle n^2(0) \rangle = 0$, nous avons

$$\langle n^2(t) \rangle = (\alpha t)^2 + \alpha t$$

Ce qui nous donne le résultat fondamental des processus de poisson sur la variance $V = \langle n^2 \rangle - \langle n \rangle^2$:

$$V(t) = \alpha t = \langle n(t) \rangle$$

Les calculs ci-dessus nous montre un phénomène tout a fait remarquable : pour calculer le moment d'ordre k , nous n'avons besoin que des moments d'ordre inférieur $k - 1, k - 2, \dots$. Ce fait

8. N'oublions pas les règles élémentaires du symbole $\langle \rangle$:

- $\langle 1 \rangle = 1$ puisque $\langle 1 \rangle = \sum_n 1P(n) = \sum_n P(n) = 1$
- $\langle \alpha \rangle = \alpha \langle 1 \rangle = \alpha$ où α est une constante.

est dû à la forme particulière de nos taux de transition (polynôme de degrés < 2 en n). Mais il ne faut pas boudier son plaisir, si nous arrivons à établir une hiérarchie des moments, nous savons que le problème est soluble.

Comme nous l'avons énoncé, nous pouvons aller plus loin et obtenir directement les $P(n, t)$. Posons

$$\phi(s, t) = \sum_n e^{ins} P(n, t) = \langle e^{ins} \rangle$$

Nous savons d'après l'équation 4.6 que

$$\frac{\partial \phi}{\partial t} = \alpha \langle e^{i(n+1)s} - e^{ins} \rangle = \alpha(e^{is} - 1) \langle e^{ins} \rangle = \alpha(e^{is} - 1)\phi$$

Malgré ses allures d'équation aux dérivées partielles, ceci est une équation différentielle ordinaire de premier ordre du genre $y' = ay$ et la solution est simplement

$$\phi(s, t) = A(s) \exp\left(\left(e^{is} - 1\right) \alpha t\right)$$

où $A(s)$ est une constante d'intégration à déterminer à l'aide des conditions initiales; A dépend de s , puisque l'équation sur ϕ n'impliquait que la variable t . A $t = 0$, nous savons de façon sûr que $n = 0$, et donc $P(n, 0) = \delta_{n,0}$, ce qui impose $\phi(s, 0) = 1$ et donc $A(s) = 1$. Finalement, nous avons

$$\phi(s, t) = \exp\left(\left(e^{is} - 1\right) \alpha t\right)$$

et nous obtenons les probabilités en effectuant l'intégrale

$$P(n, t) = \frac{1}{2\pi} \int_0^{2\pi} \exp\left(\left(e^{is} - 1\right) \alpha t\right) e^{-ins} ds$$

Pour le lecteur familier avec la théorie des fonctions complexe, cette intégration ne pose pas de problème; nous allons le cependant faire un peu en détail, cela nous servira par la suite. Nous allons effectuer le changement de variable $z = \exp(is)$: quand s parcourt l'intervalle $[0, 2\pi]$, z parcourt le cercle unité autour de l'origine dans le plan complexe. Il nous suffit de savoir⁹ que

$$\oint z^n dz = 2i\pi \delta_{n,-1}$$

C'est à dire que la seule puissance dont l'intégrale est non nulle est $n = -1$. Nous avons pour notre intégrale,

$$P(n, t) = \frac{-i}{2\pi} e^{-\alpha t} \oint e^{\alpha t z} z^{-n-1} dz$$

En développant dans l'intégrale $\exp(\alpha t z)$ en puissance de z et en retenant le seul terme qui donne une intégrale non-nulle, nous obtenons

$$P(n, t) = e^{-\alpha t} \frac{(\alpha t)^n}{n!}$$

§ 4.1 Obtenir le même résultat en utilisant la fonction génératrice des probabilités $\psi(z, t) = \langle z^n \rangle$.

9. Cela est très facile à démontrer, le faire en exercice.

Exemple 4.7 Le mouvement Brownien.

Nous pouvons répéter facilement les calculs ci-dessus pour le processus poissonien ; les taux de transition sont toujours des constantes, mais cette fois la probabilité de saut vers le bas n'est pas nulle :

$$W^+(n) = W^-(n) = \alpha$$

Ceci nous donne immédiatement la moyenne

$$\frac{d \langle n(t) \rangle}{dt} = \langle W^+(n) - W^-(n) \rangle = 0$$

Et comme nous avons placé la particule en $n = 0$ à l'instant initiale, nous aurons $\langle n(t) \rangle = 0$. Bien, la moyenne ne bouge pas, comme nous nous y attendions. Le calcul pour la variance donne

$$\frac{dV}{dt} = 2\alpha \text{ donc } V(t) = 2\alpha t$$

Puisque $V(0) = 0$. En suivant la même route que précédemment, nous obtenons pour la fonction caractéristique $\phi(s, t) = \langle e^{ins} \rangle$ l'équation différentielle

$$\frac{\partial \phi}{\partial t} = \alpha(e^{is} + e^{-is} - 2)\phi = 2\alpha(\cos(s) - 1)\phi$$

dont la solution, en utilisant la condition initiale, est

$$\phi(s, t) = \exp [2\alpha(\cos(s) - 1)t]$$

Ce qui nous donne les probabilités

$$P(n, t) = e^{-2\alpha t} \frac{1}{2\pi} \int_0^{2\pi} e^{2\alpha t \cos s - ins} ds$$

Nous pouvons à nouveau effectuer l'intégrale, mais la fonction définie par cette intégrale porte le nom de fonction de Bessel I d'ordre n et est une des plus connues fonction spéciale¹⁰(voir figure 4.7); finalement,

$$P(n, t) = e^{-2\alpha t} I_n(2\alpha t)$$

Pour les temps long, c'est à dire $t \gg n^2/8\alpha$, nous avons à notre disposition l'excellent approximation des fonctions de Bessel

$$P(n, t) = \frac{1}{\sqrt{4\pi\alpha t}} e^{-n^2/(4\alpha t)}$$

Jusque là, nous référençons les cases par leur numéro n . Supposons que les cases de l'espace ont une largeur ℓ , et posons $x = n\ell$; de cette façon, nous énumérons les cases par leurs distance à l'origine x : être dans la case n est équivalent d'être entre x et $x + \ell$ de l'origine. Nous pouvons donc à la place de la probabilité $P(n)$ utiliser la densité de probabilité $p(x)$ où $p(x)\ell = P(n, t)$. En effectuant le changement de variable, nous avons alors

$$p(x, t) = \frac{1}{\sqrt{4\alpha\ell^2 t}} e^{-x^2/(4\alpha\ell^2 t)}$$

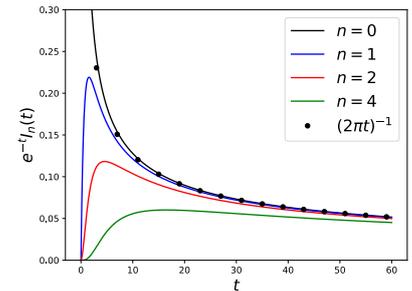


FIGURE 4.7 – Les fonctions $e^{-t} I_n(t)$ et $1/\sqrt{2\pi t}$. La limite asymptotique de ces fonctions reflète leur lente décroissance en fonction de t .

10. Voir La Bible, M. Abramowitz, *Handbook of mathematical functions*, p.374. La version électronique se trouve sur le site du DLMF.

En posant $D = \alpha \ell^2$, nous avons la forme classique de la gaussienne

$$p(x, t) = \frac{1}{\sqrt{4Dt}} e^{-x^2/4Dt}$$

Nous savons par ailleurs¹¹ que cette fonction est la solution de l'équation de diffusion

$$\frac{\partial p}{\partial t} = D \frac{\partial^2 p}{\partial x^2}$$

avec la condition initiale $p(x, 0) = \delta(x)$. Nous avons tout un chapitre dédié à l'équation de diffusion, cela nous en donne un avant goût.

11. Voir par exemple mon cours de mathématiques supérieures disponible sur la toile.

Exemple 4.8 Équation chimique.

Nous avons ici $W^+(n) = \alpha$ et $W^-(n) = \mu n$. La moyenne est donnée par

$$d \langle n \rangle / dt = \alpha - \mu \langle n \rangle$$

C'est une équation différentielle simple, avec la condition $n_0 = 0$ et donc

$$\langle n(t) \rangle = (\alpha/\mu)(1 - e^{-\mu t})$$

Pour les temps long, la moyenne tend vers le rapport des coefficient cinétique sur une échelle de temps de $\tau = 1/\mu$.

§ 4.2 Démontrer que la variance vaut

$$V(t) = (\alpha/\mu)(1 - e^{-\mu t})$$

Nous remarquons qu'à tout instant, et surtout pour les temps long devant $1/\mu$ quand l'état stationnaire est atteint, que la variance et la moyenne sont égales, ce qui est une signature des processus de poisson. Cela veut dire par exemple que si l'on observe nos cellules pendant un temps T et que l'on dresse l'histogramme du nombre d'ARN observés, nous obtenons toujours une distribution de Poisson de paramètre $\lambda = \alpha/\mu$, quelque soit le temps d'observation T , pourvu qu'il soit long devant $1/\mu$.

§ 4.3 Vérifier que pour l'état stationnaire quand $\partial_t P = 0$ (que l'on atteint pour les temps $t \gg 1/\mu$), la solution stationnaire est donnée par une distribution de Poisson de paramètre $\lambda = \alpha/\mu$.

§ 4.4 Obtenez l'équation de la fonction génératrice des probabilités :

$$\frac{\partial \phi}{\partial t} = (z - 1) \left\{ -\frac{\partial \phi}{\partial z} + \alpha \phi \right\}$$

C'est une EDP de première ordre qui se résout assez facilement si l'on pose¹² $\phi = \exp(Nu)$.

12. on peut vérifier que la solution est de la forme

$$u = (1 - e^{-t})(z - 1)$$

Le comptage individuel d'un ARN non régulé dans la bactérie *E. Coli* (voir l'expérience de I. Golding cité plus haut) a effectivement confirmé ces calculs, à ce détail près que la ratio moyenne/variance est bien une constante, mais égale à 4. Ce n'est plus vraiment un processus poissonien ! Autant pour les grandes considérations théoriques. En faite, cette observation a mené les chercheurs à mieux comprendre la transcription de l'ARN, qui arrive par "bouffée" (burst en anglais). Nous verrons plus bas comment en plaçant la

transcription sous le contrôle d'un autre processus aléatoire, nous pourrions obtenir ce phénomène. En résumé, chaque arrivée (aléatoire) de polymérase ne peut pas résulter en un ARN : le gène switch¹³ aléatoire entre deux états ON et OFF, et ce n'est que les polymérase arrivant pendant les phases ON qui peuvent donner lieu à la synthèse d'ARN.

§ 4.5 Croissance exponentielle.

Prenons une population qui ne peut que croître, avec un taux μ . Nous avons donc

$$W^+(n) = \mu n ; W^-(n) = 0$$

Soit le moment factoriel

$$(n)_\alpha = n(n+1)\dots(n+\alpha-1) = \Gamma(n+\alpha)/\Gamma(n)$$

Le symbole $(n)_\alpha$ est appelé le symbole de Pochhammer et généralise les factorielles.

Démontrer alors que

$$\frac{d}{dt} \langle (n)_\alpha \rangle = \alpha \langle (n)_\alpha \rangle$$

en déduire l'évolution de la moyenne et de la variance en fonction du temps.

Exemple 4.9 Dynamique des populations.

Nous considérons ici que le cas $\alpha = \mu$; les autres cas peuvent facilement se déduire à partir de ce traitement. Nous avons alors

$$W^+(n) = W^-(n) = \alpha n$$

La figure 4.8 nous montre une simulation numérique de ce qui se passe dans les boîtes de pétri : dans beaucoup de boîtes, le nombre d'organisme tombe à zéro et y reste par la suite : pour une naissance, il faut un individu déjà présent ; autrement dit, $W^+(0) = 0$ et $n = 0$ est un état absorbant. Dans certaines boîtes au contraire, le nombre d'individu, par pure chance, devient de plus en plus grand. Au fur et à mesure que le temps s'écoule, de plus en plus de boîtes se vident, tandis que certaines continuent à s'enrichir au delà de toute limite. D'après l'équation 4.2,

$$\frac{d \langle n \rangle}{dt} = \langle W^+(n) - W^-(n) \rangle = 0$$

et donc si on avait placé n_0 individu dans chaque boîtes à l'instant zéro (autrement dit $P(n, 0) = \delta_{n, n_0}$)

$$\langle n(t) \rangle = n_0$$

Ce qui est bien ce que l'on observe sur la figure 4.8b. Pour la variance, l'équation 4.5 nous donne

$$\frac{dV}{dt} = \langle W^+(n) - W^-(n) \rangle = \alpha n_0$$

et donc

$$V(t) = \alpha n_0 t$$

(Figure 4.8b). Pour mesurer le taux d'agglomération, nous utilisons souvent le ratio F de variance sur moyenne. Quand $F \sim 1$,

13. "commute" en français.

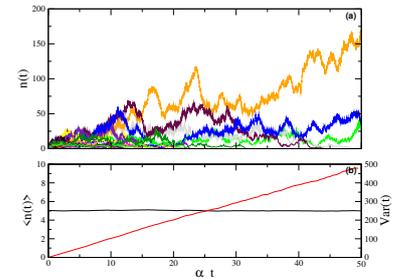


FIGURE 4.8 – (a) : Simulation numérique de 500 évolutions temporelles de nombre d'individu. A l'instant initial, exactement 5 individus étaient placés dans chaque boîte de pétri. Le nombre d'individu varie alors selon la règle neutre suivant : par unité de temps, tous les individus ont la même probabilité de se dupliquer que de mourir. (b) : la moyenne (noir, axe gauche) et la variance (rouge, axe droit) obtenus à partir des coupes temporelles de ces trajectoires.

la distribution dans les boîtes de pétrie est poissonien et nous verrons plus ou moins le même nombre d'individu par boîte. Quand $F \gg 1$, nous avons affaire à une distribution agglomérée, où certaines boîtes ont considérablement plus d'individu que d'autres. Pensez par exemple à la distribution des humains, où les villes condensent l'essentielle de la population.

Nous voyons ici que $F = \alpha t \rightarrow \infty$! Sans parler des humains, nous voyons que le phénomène le plus fondamental de la vie, mort et naissance aléatoire, provoque des agglomérations extrêmement importantes. Nous pouvons ouvrir les barrières entre les boîtes et laisser les organismes *diffuser* des zones de forte concentration vers les zones de basse concentration, mais la diffusion n'est pas suffisante pour empêcher le phénomène d'agglomération. Nous verrons cela plus bas, quand nous étudierons les systèmes spatialement étendus. Notons simplement que sur les plusieurs milliers d'espèces vivantes dont on a mesuré le taux d'agglomération, toutes possédaient des ratios F très supérieurs à l'unité.

§ 4.6 Obtenez une équation à dérivée partielle pour la fonction caractéristique, et vérifiez que la solution de cette équation est donnée par (sans perte de généralité, nous posons $\alpha = 1$)

$$\phi(s, t) = \left[\frac{(t-1)e^{is} - t}{te^{is} - (t+1)} \right]^{n_0}$$

démontrer alors que $P(0, t)$ est donnée par

$$P(0, t) = \frac{1}{2\pi} \int_0^{2\pi} \phi(s, t) ds = \left(\frac{t}{1+t} \right)^{n_0}$$

et que pour $n_0 = 1$,

$$P(n, t) = \frac{t^{n-1}}{(1+t)^{n+1}}$$

Démontrer que $P(n, t)$ atteint son maximum en $t_n^* = (n-1)/2$ et que pour n grand

$$P(n, t_n^*) \approx \frac{4}{e^2} \frac{1}{(n+1)^2}$$

Exemple 4.10 La marche vers l'équilibre.

Nous laissons le traitement de ce problème comme exercice au lecteur intéressé. Il n'est pas très difficile de démontrer que l'équation de la fonction génératrice des probabilités (PGF en anglais) est donnée par

$$\frac{\partial \phi}{\partial t} = (z-1) \left\{ N\phi - (z+1) \frac{\partial \phi}{\partial z} \right\}$$

l'équilibre s'établit quand $\partial_t \phi = 0$ et le calcul de $\phi(z, \infty)$ nous donne accès aux probabilités données par une distribution binomiale

$$P(n, \infty) = 2^{-N} C_N^n$$

En dépensant un peu de temps¹⁴, on peut même obtenir l'expression exacte de l'entropie $S(t)$ et démontrer que c'est une fonction monotone croissante. Remarquons que nous avons la réversibilité *microscopique*, chaque carte prise individuellement a

14. posez $\phi = u^N$ et obtenez une équation pour u . Comme $\phi(z, t)$ est un polynôme de degré N en z , $u(z, t)$ est un polynôme de degré 1 et il est facile d'obtenir ses deux valeurs propres.

autant de chance d'aller de la pile A à la pile B ou vice et versa. Pourtant, *macroscopiquement*, nous finissons par deux piles de taille plus ou moins similaire, quelque soit la configuration de départ. Notons que l'accent est sur le mot macroscopique, qui veut dire N très grand.

Exemple 4.11 L'évolution Darwinienne.

Nous laissons à nouveau le traitement de ce problème au lecteur intéressé. La façon efficace de traiter ce problème est à travers la fonction génératrice des probabilités

$$\psi(z, t) = \langle z^n \rangle$$

§ 4.7 Démontrer que la fonction ψ obéit à l'équation à dérivée partielle

$$\sigma \frac{\partial \psi}{\partial t} = (z-1)(z-\sigma) \frac{\partial}{\partial z} \left[N\psi - z \frac{\partial \psi}{\partial z} \right]$$

[A compléter : un long commentaire sur les mutations neutres, le temps de fixation, compétition entre sélection et dérive génétique]

Exemple 4.12 L'expérience de Luria & Delbruck

Un des résultats le plus fondamentaux en théorie d'Évolution a été obtenu par Luria et Delbruck¹⁵. Le problème a été de déterminer si les mutations sont induites et dirigées, ou si elles sont aléatoires. L'expérience consiste à exposer des bactéries à des virus et de compter le nombre de survivant (bactérie résistante).

Supposons que nous soumettons M cultures de bactéries à un virus. Supposons que les mutations sont induites par la présence de virus, la probabilité pour une bactérie de survivre à un virus étant λ , le nombre de survivant est donné par une distribution de poisson. Cela veut dire que si nous calculons la moyenne $\langle m \rangle$ et la variance V des survivants dans les M cultures, nous devons trouver $V / \langle m \rangle \sim 1$.

Si les mutations ne sont pas induites, mais qu'elles arrivent de façons aléatoires, le calcul donne un résultat différent. Prenons une culture de bactérie en croissance avec un taux α , et soit $N(t)$ le nombre de bactérie *sauvage*¹⁶ au temps t . Nous avons $N(t) = N_0 \exp(\alpha t)$. Soit a le taux de mutation sauvage \rightarrow résistant. Soit m le nombre de bactérie résistantes. Nous avons alors

$$W^+(m) = aN + \alpha m$$

Le nombre de bactérie résistante augmente d'une part par les mutations des sauvages, d'autre part par la croissance de ceux déjà existant, dont nous supposons qu'ils ont le même taux de croissance. Nous pouvons alors déduire la moyenne et la variance. Commençons par la moyenne :

$$\begin{aligned} \frac{d \langle m \rangle}{dt} &= \langle W^+(m) \rangle \\ &= \alpha \langle m \rangle + aN \end{aligned}$$

ce qui nous donne

$$\begin{aligned} \langle m \rangle &= \alpha N_0 e^{\alpha t} \\ &= \alpha N(t) \end{aligned}$$

15. Luria, S. E. ; Delbrück, M. (1943). *Mutations of Bacteria from Virus Sensitivity to Virus Resistance*. Genetics 28 (6) : 491-511.

16. Wild type (WT) en anglais, c'est à dire non mutant

Nous voyons que la croissance des mutants est plus rapide que la croissance des sauvages ; cela cependant est très difficilement mesurable.

Continuons avec le second moment. Nous avons

$$\begin{aligned} \frac{d \langle m^2 \rangle}{dt} &= 2 \langle mW^+(m) \rangle + \langle W^+(m) \rangle \\ &= 2\alpha \langle m^2 \rangle + (2aN + \alpha) \langle m \rangle + aN \end{aligned}$$

ce qui nous donne

$$\langle m^2 \rangle = a^2 t^2 N_0^2 e^{2\alpha t} + 2 \frac{a}{\alpha} N_0 e^{\alpha t} (e^{\alpha t} - 1 - \alpha t / 2)$$

Nous remarquons que le premier terme de droite est juste $\langle m \rangle^2$. Par ailleurs, quand $\alpha t \gg 1$, nous pouvons négliger les termes linéaires et nous avons pour la variance $V = \langle m^2 \rangle - \langle m \rangle^2$

$$V = 2 \frac{a}{\alpha} \frac{N^2}{N_0}$$

Cela nous donne le rapport

$$\frac{V}{\langle m \rangle} = 2 \frac{N}{N_0 \alpha t} = \frac{2(N/N_0)}{\log(N/N_0)}$$

Nous voyons que dans le cas où les mutations sont aléatoire, le rapport $V / \langle m \rangle$ peut être très nettement supérieur à l'unité.

Luria et Delbruck ont donc mis M culture en croissance, en partant d'un petit nombre de bactérie ($N_0 \sim 500$), et ont attendu que le nombre de bactérie dans chaque tube atteigne le maximum (de l'ordre de 10^8). Ils ont ensuite exposés les cultures à un virus particulier, et ont compter le nombre de survivant dans chaque tube. Le rapport $V / \langle m \rangle$ a été très supérieur à 1, et de l'ordre de 1000. Il ont ainsi établit que les mutations sont purement aléatoire. Ces travaux ont été reconnu par le prix Nobel de 1969.

Remarque finale : moment closure.

Que faire quand nous n'avons pas une hiérarchie des moments ?

4.4 Problèmes.

Problème 4.1 Mouvement Brownien borné.

Supposons qu'une particule brownien de position n soit confiné entre les deux bornes N_i et N_f ($N_f > N_i$), et les taux de transitions soient donnés par

$$\begin{aligned} W^+(n) &= \alpha(N_f - n) \\ W^-(n) &= \alpha(n - N_i) \end{aligned}$$

Soit la fonction génératrice des probabilité (PGF)

$$\phi(z, t) = \sum_n z^n P(n, t) = \langle z^n \rangle$$

En utilisant le fait que $\langle nz^n \rangle = z \partial_z \phi$, et que $\phi(1, t) = 1$, démontrer que le PGF obéit à l'équation

$$\frac{\partial \phi}{\partial t} = (z - 1) \left\{ \left(N_f + \frac{N_i}{z} \right) \phi - (z + 1) \frac{\partial \phi}{\partial z} \right\}$$

Montrer alors que la solution stationnaire est donnée par

$$\phi(z) = 2^{N_i - N_f} z^{N_i} (1 + z)^{N_f - N_i}$$

En développant le binôme, montrer alors que pour $N_i \leq n \leq N_f$, la distribution stationnaire est donnée par

$$P(n) = \binom{N_f - N_i}{N_f - n}$$

et pour $N_f - N_i \gg 1$, nous pouvons approximer cette distribution par une distribution normale $N(\mu, \sigma)$ où

$$\begin{aligned} \mu &= (N_f + N_i) / 2 \\ \sigma &= \sqrt{N_f - N_i} \end{aligned}$$

Faire le rapprochement avec un mouvement brownien continu dans un potentiel harmonique.

Problème 4.2 Biologie de la mort.

Considérer une particule qui peut passer de l'état n à l'état $n + 1$ avec un certain taux α . Cependant, il existe une possibilité β que la particule tombe à zéro. En d'autre terme,

$$W(n \rightarrow n + 1) = \alpha ; W(n \rightarrow 0) = \beta$$

On peut voir ce système comme vieillir avec le taux α et mourir avec le taux β . Montrer que dans ce système, la distribution des ages est poissonien. Ceci ne reflète pas du tout la vraie distribution des ages chez les humains : la probabilité d'avoir plus de 140 ans est nulle. Montrer que nous pouvons retrouver cet aspect si nous supposons que la probabilité de mourir augmente avec l'age : $W(n \rightarrow 0) = \beta n$.

Problème 4.3 Mouvement aléatoire et vol de Lévy.

Considérer un mouvement aléatoire sur réseau donné par les taux

$$\begin{aligned} W^+(n) &= n^2 - \alpha n + \beta \\ W^-(n) &= n^2 + \alpha n + \beta \end{aligned}$$

Montrer que quelque soit α , nous pouvons trouver β tel que $W^\pm(n) > 0$. Analyser ce système au comportement un peu ... pathologique.

Problème 4.4 Les réactions chimiques comme système à trois état.

[Bien dégager les temps de réaction vs l'équilibre et faire le lien avec la thermodynamique].

Problème 4.5 La polymérisation des microtubule.

Problème 4.6 Oscillateur harmonique en équilibre avec un champs électromagnétique.

5

Quelques généralisations de l'équation Maître.

Le chapitre précédent nous a habitué à manipuler les équations maîtresses simple. Nous pouvons maintenant essayer de compléter le tableau et enlever les contraintes au fur et à mesure. Les mathématiques deviennent un peu plus compliqué, mais à partir du moment où l'on arrive à formuler correctement un problème, nous avons presque gagné : le reste peut être estimé à l'aide de simulation ou d'approximations diverses et variées.

5.1 Système à deux composants.

Jusque là, nous avons traité des systèmes où les états pouvaient être énumérés à l'aide d'un seul indice. Nous avons souvent affaire à des systèmes à plusieurs indices. Par exemple, le mouvement Brownien à d dimension a besoin de d indices. Nous avons également affaire à des cas où une des variables aléatoire et sous le contrôle d'une autre. Ceci est le cas des réactions chimiques enchaînées par exemple. Contentons nous pour l'instant à des systèmes à deux indices seulement, que nous appellerons n et m . Pour pouvoir formuler l'équation maîtresse, en nous contentant seulement des saut entre état immédiatement voisin, nous avons besoin de 4 taux

$$W(n \rightarrow n \pm 1, m), W(n, m \rightarrow m \pm 1)$$

Nous noterons ces taux W_1^\pm et W_2^\pm selon que nous avons affaire au premier ou au deuxième indice. En suivant la même démarche de balance des flux entrant en et sortant de l'état (n, m) , nous obtenons

$$\begin{aligned} \frac{\partial P(n, m, t)}{\partial t} &= W_1^+(n-1, m)P(n-1, m, t) - W_1^+(n, m)P(n, m, t) \\ &+ W_1^-(n+1, m)P(n+1, m, t) - W_1^-(n, m)P(n, m, t) \\ &+ W_2^+(n, m-1)P(n, m-1, t) - W_2^+(n, m)P(n, m, t) \\ &+ W_2^-(n, m+1)P(n, m+1, t) - W_2^-(n, m)P(n, m, t) \end{aligned}$$

Comme précédemment, nous pouvons extraire la moyenne de l'état n ou m en appliquant l'opérateur $\sum_{n,m} n$ ou $\sum_{n,m} m$ à l'équation ci-dessus. Nous laissons au lecteur de vérifier que

$$\frac{d \langle n \rangle}{dt} = \langle W_1^+(n, m) - W_1^-(n, m) \rangle$$

en gardant à l'esprit que le symbole $\langle \rangle$ veut cette fois dire $\langle f(n, m) \rangle = \sum_{n,m} f(n, m)P(n, m, t)$. Il est surtout très instructif de voir que pour obtenir la moyenne sur l'état n , nous n'avons qu'à tenir compte des taux W_1^\pm .

En continuant le jeu d'indice, nous nous rendons compte que pour obtenir l'expression de $\langle n^2 \rangle$ ou $\langle m^2 \rangle$, nous avons besoin d'une nouvelle quantité, $\langle nm \rangle$, qui reflète la corrélation entre les deux indices. En appliquant l'opérateur $\sum_{n,m} nm$ à l'équation 5.1, nous obtenons l'équation pour cette quantité :

$$\frac{d \langle nm \rangle}{dt} = \langle m (W_1^+(n, m) - W_1^-(n, m)) \rangle + \langle n (W_2^+(n, m) - W_2^-(n, m)) \rangle$$

Exemple 5.1 le mouvement Brownien.

Pour le mouvement Brownien, les quatre taux de transition sont des constantes : les états n et m sont indépendantes et nous pouvons les traiter comme deux variables aléatoires séparées. Nous laissons quand même le soin au lecteur de démontrer que comme avant,

$$\langle n(t) \rangle = \langle m(t) \rangle = 0$$

Comme les deux variables sont indépendantes,

$$\langle nm \rangle = \langle n \rangle \langle m \rangle = 0$$

et l'expression pour les variances sont strictement similaires à ce que nous avons déjà obtenu pour le mouvement Brownien 1d.

Exemple 5.2 Production de protéine.

Nous avons déjà rencontré la production d'ARN, dont nous notons le nombre par r . Dans la cellule vivante, les ARN sont utilisés par les ribosomes pour produire des protéine, dont on note le nombre par p . La production d'ARN des gènes non régulés n'est pas influencé par la quantité de protéines¹ donc nous avons comme précédemment

$$\begin{aligned} W_1^+(r, p) &= \lambda_r \\ W_1^-(r, p) &= \mu_r r \end{aligned}$$

La production des protéines par contre nécessite la présence d'ARN et au premier ordre, proportionnelle à cette quantité. La cellule possède par ailleurs un système de dégradation de protéines² et donc

$$\begin{aligned} W_2^+(r, p) &= \lambda_p r \\ W_2^-(r, p) &= \mu_p p \end{aligned}$$

Nous pouvons alors obtenir les équations de nos différents moments³ :

$$\begin{aligned} \frac{d \langle r \rangle}{dt} &= \lambda_r - \mu_r \langle r \rangle \\ \frac{d \langle p \rangle}{dt} &= \lambda_p \langle r \rangle - \mu_p \langle p \rangle \end{aligned}$$

1. pour certains gènes, la protéine produite peut à son tour influencer la production de son propre ARN, ce qui est par exemple le cas de l'opéron Lac responsable de la digestion du lactose. Cela donne la possibilité de réaliser des *circuits génétiques* dont l'étude dépasse le cadre de ce cours texte. Le lecteur intéressé peut se rapporter à mon petit cours sur la biologie moléculaire.

2. appelé ubiquitine

3. Le lecteur est bien sûr fortement encouragé à effectuer ces calculs

$$\begin{aligned}\frac{d\langle r^2 \rangle}{dt} &= 2\lambda_r \langle r \rangle - 2\mu_r \langle r^2 \rangle + \lambda_r + \mu_r \langle r \rangle \\ \frac{d\langle p^2 \rangle}{dt} &= 2\lambda_p \langle rp \rangle - 2\mu_p \langle p^2 \rangle + \lambda_p \langle r \rangle + \mu_p \langle p \rangle \\ \frac{d\langle rp \rangle}{dt} &= -(\mu_r + \mu_p) \langle rp \rangle + \lambda_r \langle p \rangle + \lambda_p \langle r^2 \rangle\end{aligned}$$

Tout ceci n'est qu'un système d'équations différentielles couplées de premier ordre que nous pouvons résoudre trivialement. La quantité intéressante est cependant l'état stationnaire, et la variance de protéines que l'on peut effectivement mesurer⁴ :

$$\text{Var}(p) / \langle p \rangle = 1 + b$$

Nous laissons au lecteur d'obtenir ce résultat, ou b devrait être quelque chose comme nombre de protéine produit par un ARN.

Nous pouvons compliquer encore un peu plus ces calculs en ajoutant un étage : l'ADN lui même commute entre deux états ON et OFF que nous énumérons par la variable n ($n = 1$ ou 0); nous pouvons écrire les taux pour cette variable; le taux de production d'ARN doit alors être modifié par

$$W_1^+(n, r, p) = \lambda_r n$$

Nous ne continuons pas plus loin ces considérations, le but de cette section était simplement d'illustrer l'étude des systèmes à n composants.

[parler de la distribution Gamma quelques part et citer les calculs du PNAS 2005 que je doit retrouver].

Exemple 5.3 La statistique des syntons dans le génome bactérien modèle d'agrégation (Rivoire 2013).

5.2 Système spatialement étendu.

5.2.1 Définitions et notations.

Les systèmes spatialement étendu sont une généralisation des systèmes à deux composants; on peut les considérer comme des systèmes à une infinité de composants. Cependant, le même jeu avec les indices reste valable et nous pouvons tirer beaucoup d'information de l'équation sans toujours la résoudre.

La première chose à faire est de définir l'équation Maître correctement, et nous ne pouvons l'écrire que si nous utilisons une notation rigoureuse et simple. Prenons un système uni-dimensionnel et appelons n_k le nombre d'individu dans la case k . L'état du système est alors donné par $\mathbf{n} = (\dots, n_{-1}, n_0, n_1, \dots)$. Nous définissons comme $P(\mathbf{n}, t)$ la probabilité d'observer le système dans l'état \mathbf{n} (c'est à dire voir n_1 particules dans la case 1, n_2 dans la case 2 ...) à l'instant t . Pour nos besoins, nous devons définir deux nouveaux vecteurs relié au vecteur \mathbf{n} :

$$\begin{aligned}a_i \mathbf{n} &= (\dots, n_{i-1}, n_i - 1, n_{i+1}, \dots) \\ a_i^\dagger \mathbf{n} &= (\dots, n_{i-1}, n_i + 1, n_{i+1}, \dots)\end{aligned}$$

4. Les protéines peuvent être modifiées pour porter une partie fluorescente, le signal lumineux est alors utilisé pour mesurer leurs quantités

le vecteur $a\mathbf{n}$ représente donc le même vecteur que \mathbf{n} , à ceci près qu'il a une particule de moins dans la case i . On appelle d'ailleurs les deux opérateurs a et a^\dagger les opérateurs d'annihilation et de création, bien que je trouve ces deux termes un peu trop mystificateur. Nous pouvons maintenant formuler notre équation Maître exacte comme avant, en faisant le bilan des flux entrant et sortant vers et de l'état \mathbf{n} . La seule chose que nous devons faire est de déterminer quel sont les états accessible à partir d'un état \mathbf{n} .

Prenons le cas de la diffusion de particule. Supposons que nous avons des particules brownien réparti dans l'espace, n_k particules dans la case k . Chaque particule a la même probabilité par unité de temps de sauter à gauche ou à droite α . Quand une particule saute de la case k à la case $k + 1$ par exemple, le nombre de particule dans la case k est diminué de 1, tandis que le nombre de particule dans la case $k + 1$ est augmenté d'une unité. Un tel événement élémentaire transforme donc l'état \mathbf{n} en l'état $a_k a_{k+1}^\dagger \mathbf{n}$ avec un taux de transition $W = mn_k$, où m et la densité de probabilité de saut par particule. Notre équation maître doit donc s'écrire

Nous avons alors

$$\begin{aligned} \frac{\partial P(\mathbf{n}, t)}{\partial t} &= \sum_i W(a_{i+1}^\dagger a_i \mathbf{n} \rightarrow \mathbf{n}) P(a_{i+1}^\dagger a_i \mathbf{n}, t) \\ &\quad - W(\mathbf{n} \rightarrow a_{i+1}^\dagger a_i \mathbf{n}) P(\mathbf{n}, t) \end{aligned} \quad (5.2)$$

$$+ W(a_{i-1}^\dagger a_i \mathbf{n} \rightarrow \mathbf{n}) P(a_{i-1}^\dagger a_i \mathbf{n}, t) \quad (5.3)$$

$$- W(\mathbf{n} \rightarrow a_{i-1}^\dagger a_i \mathbf{n}) P(\mathbf{n}, t) \quad (5.4)$$

où la sommation est sur les cases de l'espace.

Si en plus de la diffusion, nous avons des naissances (taux α) et des morts (taux μ), nous devons également prendre en compte des événements locaux

$$W(\mathbf{n} \rightarrow a_i^\dagger \mathbf{n}) = \alpha n_i$$

$$W(\mathbf{n} \rightarrow a_i \mathbf{n}) = \mu n_i$$

et ajouter les termes correspondants à l'équation maître.

Exemple 5.4 diffusion+naissance et mort.

Si maintenant, en plus de la diffusion, les particules sont capable de se dupliquer et de mourir, avec des taux β et μ , alors

$$W_i^+(\mathbf{n}) = \alpha(n_{i+1} + n_{i-1}) + \beta n_i$$

$$W_i^-(\mathbf{n}) = 2\alpha n_i - \mu n_i$$

5.2.2 Extraire les moyennes.

Nous pouvons maintenant jouer au même jeu d'indice pour obtenir la moyenne, la corrélation, la variance, ...

Appliquons d'abord l'opérateur

$$\sum_{\mathbf{n}} = \dots \sum_{n_{-1}} \sum_{n_0} \sum_{n_1} \dots$$

aux deux côtés de l'équation (5.3). Considérons pour l'instant le terme

$$I = \sum_{\mathbf{n}} \sum_i W(a_{i+1}^\dagger a_i \mathbf{n} \rightarrow \mathbf{n}) P(a_{i+1}^\dagger a_i \mathbf{n}, t)$$

Ceci est un petit exercice qui nécessite de prendre un crayon et d'écrire explicitement quelques expressions. Nous pouvons échanger l'ordre de sommation sur les états \mathbf{n} et sur l'espace i ; ensuite, pour un i fixé, nous pouvons changer l'indice de sommation de \mathbf{n} à $a_{i+1}^\dagger a_i^\dagger \mathbf{n}$ (comme on faisait à 0 dimension pour changer de variable de n à $m = n + 1$). Nous obtenons, alors

$$\begin{aligned} I &= \sum_i \sum_{\mathbf{n}} W(a_{i+1}^\dagger a_i \mathbf{n} \rightarrow \mathbf{n}) P(a_{i+1}^\dagger a_i \mathbf{n}, t) \\ &= \sum_i \sum_{\mathbf{n}} W(\mathbf{n} \rightarrow a_{i+1}^\dagger a_i^\dagger \mathbf{n}) P(\mathbf{n}, t) \\ &= \sum_{\mathbf{n}} \sum_i W(\mathbf{n} \rightarrow a_{i+1}^\dagger a_i^\dagger \mathbf{n}) P(\mathbf{n}, t) \end{aligned}$$

Et nous voyons que ce terme s'annule exactement avec le flux sortant correspondant. Cela nous donne donc, comme il se doit,

$$\frac{\partial \sum_{\mathbf{n}} P(\mathbf{n}, t)}{\partial t} = 0$$

puisque de toute façon, la somme des probabilités est toujours 1 et donc sa variation nulle. Il ne faut pas mystifier le changement de variable de \mathbf{n} à $a^\dagger \mathbf{n}$, c'est juste la généralisation de la formule

$$\sum_n f(n) = \sum_n f(n+1)$$

quand les limites de sommation sont $\pm\infty$. Ici, nous sommes entrain de faire une multitude de somme en parallèle (c'est le sens de la notation $\sum_{\mathbf{n}}$), mais à part cela, c'est exactement la même chose.

La moyenne de particules dans la case k est

$$\langle n_k \rangle (t) = \sum_{\mathbf{n}} n_k P(\mathbf{n}, t)$$

Appliquons l'opérateur $\sum_{\mathbf{n}} n_k$ aux deux côtés de l'équation (5.3); concentrons nous à nouveau sur le premier terme de droite

$$J = \sum_{\mathbf{n}} \sum_i n_k W(a_{i+1}^\dagger a_i \mathbf{n} \rightarrow \mathbf{n}) P(a_{i+1}^\dagger a_i \mathbf{n}, t)$$

qui vaut donc :

$$\begin{aligned} J &= \sum_i \sum_{\mathbf{n}} n_k W(a_{i+1}^\dagger a_i \mathbf{n} \rightarrow \mathbf{n}) P(a_{i+1}^\dagger a_i \mathbf{n}, t) \\ &= \sum_i \sum_{\mathbf{n}} (n_k + \delta_{k,i} - \delta_{k,i+1}) W(\mathbf{n} \rightarrow a_{i+1}^\dagger a_i^\dagger \mathbf{n}) P(\mathbf{n}, t) \\ &= \sum_{\mathbf{n}} \sum_i n_k W(\mathbf{n} \rightarrow a_{i+1}^\dagger a_i^\dagger \mathbf{n}) P(\mathbf{n}, t) P(\mathbf{n}, t) \\ &\quad + \sum_{\mathbf{n}} W(\mathbf{n} \rightarrow a_{k+1}^\dagger a_k^\dagger \mathbf{n}) P(\mathbf{n}, t) \\ &\quad - \sum_{\mathbf{n}} W(\mathbf{n} \rightarrow a_k a_{k-1}^\dagger \mathbf{n}) P(\mathbf{n}, t) \end{aligned}$$

Comme précédemment, le premier terme de droite s'annule avec le flux sortant; en faisant la même chose sur le reste de l'équation, nous trouvons

$$\frac{d \langle n_k \rangle}{dt} = \langle W(\mathbf{n} \rightarrow a_{k+1} a_k^\dagger \mathbf{n}) - W(\mathbf{n} \rightarrow a_k a_{k-1}^\dagger \mathbf{n}) \rangle$$

En faisant la même opération pour les termes restants (les sauts vers la gauche), nous trouvons en définitif

$$\begin{aligned} \frac{d \langle n_k \rangle}{dt} &= \langle W(\mathbf{n} \rightarrow a_{k+1} a_k^\dagger \mathbf{n}) - W(\mathbf{n} \rightarrow a_k a_{k-1}^\dagger \mathbf{n}) \rangle \\ &+ \langle W(\mathbf{n} \rightarrow a_{k-1} a_k^\dagger \mathbf{n}) - W(\mathbf{n} \rightarrow a_k a_{k+1}^\dagger \mathbf{n}) \rangle \end{aligned}$$

Ce qui, pour la diffusion simple, nous donne

$$\frac{d \langle n_k \rangle}{dt} = m (\langle n_{k+1} \rangle + \langle n_{k-1} \rangle - 2 \langle n_k \rangle)$$

Il est plus usuel de mesurer l'espace avec la variable continue $x = k/\ell$ où ℓ est la taille d'une case et noter $c(x, t) = \langle n_k \rangle(t)/\ell$ qui porte le nom de concentration. Notant $D = m\ell^2$, nous voyons alors que

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2} \quad (5.5)$$

qui est l'équation usuelle de diffusion. Si nous ajoutons les termes de naissances et de mort, cette équation devient

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2} + (\alpha - \mu)c \quad (5.6)$$

Il est crucial de noter cependant que les équations de type (5.5,5.6) n'ont de sens que si nous avons une estimation des fluctuations et des corrélations. Une description par la moyenne est valide seulement si les fluctuations sont petites par rapport à ces derniers. Nous voyons par exemple que les deux équations sur les moyennes sont identiques si le taux de mort et de naissance sont égaux $\beta = \mu$. Cependant, le comportement des deux systèmes est radicalement différent.

5.2.3 Extraire les fluctuations.

Nous laissons au lecteur le soin de déduire les équations pour les moments d'ordre 2

$$\begin{aligned} \frac{d \langle n_k^2 \rangle}{dt} &= \dots \\ \frac{d \langle n_k n_{k+j} \rangle}{dt} &= \dots \end{aligned}$$

En appliquant l'opérateur $\sum_{\mathbf{n}} n_k^2$ ou $\sum_{\mathbf{n}} n_k n_{k+j}$ ($j \neq 0$) à l'équation maîtresse (5.3), comme maintenant nous en avons l'habitude.

5.3 Transitions entre états quelconque.

Revenons maintenant à nos système à une seule composante, énumérée par un seul indice n . Jusque là, nous avons considéré que des probabilités de saut entre état voisin, nous allons maintenant relâcher cette contrainte. Soit $W(n, k)$ la densité de probabilité d'effectuer un saut de taille k quand on se trouve dans l'état n . En général, la taille des sauts dépend explicitement de l'état n . Si par exemple la particule se trouve dans un puits de potentiel, la probabilité des sauts change en fonction de la distance de la particule au centre du puits.

Comme $W(n, k)$ est une densité de probabilité, nous pouvons définir les quantités statistiques usuelles : quel est le saut moyen sachant que l'on se trouve en n ? Nous noterons cette quantité par $\langle k|n \rangle$

$$\langle k|n \rangle = \sum_k kW(n, k) \quad (5.7)$$

la sommation étant sur k , il est évident que cette quantité dépend de n . Tous les autres moments se définissent de la même façon. Par exemple

$$\langle k^2|n \rangle = \sum_k k^2 W(n, k) \quad (5.8)$$

et la variance

$$V|n = \langle k^2|n \rangle - \langle k|n \rangle^2$$

Ce sont là des quantités locales : un observateur peut se positionner dans un état n et dresser l'histogramme des sauts à partir de cet état. Si par ailleurs, nous connaissons la probabilité $P(n)$ de se trouver dans l'état n , nous pouvons déduire la moyenne des sauts (en laissant tomber "sachant n ") :

$$\langle k \rangle = \sum_n \langle k|n \rangle P(n) \quad (5.9)$$

N'oublions pas que si $P(n)$ dépend du temps, $\langle k \rangle$ en dépend également et l'on devrait plutôt l'écrire $\langle k(t) \rangle$. Insistons cependant que nous ne connaissons pas à priori cette quantité, qui nécessite la connaissance des $P(n)$.

Avant de continuer, notons que notre notation $\langle \rangle$ devient un peu confus. Dans les équations 5.7, 5.8, $\langle \rangle$ désigne une sommation sur les probabilités de saut $W(k)$, tandis que dans l'équation 5.9, $\langle \rangle$ désigne une sommation sur les probabilités de présence dans l'état n , $P(n)$. Ce sont deux choses complètement différentes et pour les distinguer, nous distinguerons par un indice S les moyennes sur les sauts et les noterons $\langle \rangle_S$; nous conserverons la notation $\langle \rangle$ sans indice pour les moyennes sur les états.

Considérons maintenant l'équation maîtresse, qui n'est qu'une comptabilité des particules entrant et sortant dans un état s'écrit de façon générale

$$\frac{\partial P(n, t)}{\partial t} = \sum_k W(n-k, k)P(n-k, t) - W(n, k)P(n, t)$$

Comme précédemment, nous sommes intéressés par l'évolution de la moyenne et de la variance. Appliquons l'opérateur $\sum_n n$ à l'équation ci-dessus et échangeons l'ordre de sommation⁵ sur les indices k et n :

$$\begin{aligned} \frac{d \langle n(t) \rangle}{dt} &= \sum_k \sum_n n W(n-k, k) P(n-k, t) \\ &- \sum_k \sum_n n W(n, k) P(n, t) \end{aligned}$$

Effectuons le changement de variable $n \rightarrow n+k$ dans la première somme ; une fois retranchée la deuxième sommation, nous arrivons à

$$\frac{d \langle n(t) \rangle}{dt} = \sum_k \sum_n k W(n, k) P(n, t)$$

et un dernier changement d'ordre de sommation nous donne

$$\frac{d \langle n(t) \rangle}{dt} = \sum_n \langle k|n \rangle_S P(n, t) \quad (5.10)$$

$$= \langle \langle k|n \rangle_S \rangle \quad (5.11)$$

L'équation 5.11 est très parlante : la vitesse de changement de la moyenne n est égale à la moyenne des sauts.

Nous pouvons répéter ces calculs pour le moment d'ordre 2 pour aboutir à

$$\begin{aligned} \frac{d \langle n^2(t) \rangle}{dt} &= \sum_n \left(\sum_k k^2 W(n, k) + \sum_k 2nk W(n, k) \right) P(n, t) \\ &= \sum_n \left(\langle k^2|n \rangle_S + 2n \langle k|n \rangle_S \right) P(n, t) \\ &= \langle \langle k^2|n \rangle_S \rangle + \langle 2n \langle k|n \rangle_S \rangle \end{aligned} \quad (5.12)$$

Exemple 5.5 les sauts entre état voisin.

Au chapitre précédent, nous considérions que des sauts entre états voisins. Nous notons $W^\pm(n) = W(n, \pm 1)$ et posons $W(n, |k| > 1) = 0$. La taille moyenne des sauts est

$$\begin{aligned} \langle k|n \rangle_S &= (+1)W^+(n) + (-1)W^-(n) \\ &= W^+(n) - W^-(n) \end{aligned}$$

L'équation 5.11 s'écrit donc

$$\frac{d \langle n(t) \rangle}{dt} = \langle W^+(n) - W^-(n) \rangle$$

Qui n'est rien d'autre bien sûr que l'équation 4.3 que nous avons utilisé tout le long du chapitre 4.

Le second moment des sauts s'écrit

$$\begin{aligned} \langle k^2|n \rangle_S &= (+1)^2 W^+(n) + (-1)^2 W^-(n) \\ &= W^+(n) + W^-(n) \end{aligned}$$

L'équation 5.12 s'écrit alors

$$\frac{d \langle n^2(t) \rangle}{dt} = \langle 2n (W^+(n) - W^-(n)) \rangle + \langle W^+(n) + W^-(n) \rangle$$

Et on retrouve à nouveau nos vieux amours.

5. Sans rentrer dans le détail, nous supposons que nos processus sont "gentils", ce qui est le cas de tous les processus physique, et nous pouvons sans crainte échanger les deux sommations. Cela serait beaucoup plus problématique si nous avions affaire par exemple à des "vols de Lévy" où certains des moments statistiques n'existent pas.

Exemple 5.6 Transcription d'ARN et les bouffées.

Au chapitre précédent, nous avons traité la transcriptions d'ARN comme un processus à deux composante : le gène commute aléatoirement entre deux états ON et OFF; quand le gène est dans l'état ON, le processus aléatoire de transcription peut prendre place. Il se trouve que ces deux phénomènes se déroulent sur des échelles de temps très différentes, la commutation de gène étant beaucoup plus rapide (\sim milliseconde) que la transcription et la dégradation d'ARN (\sim minute). Pour l'observateur qui doit faire des pauses d'une seconde avec sa caméra, les ARN arrivent par bouffée (burst) et se dégradent ensuite lentement (voir figure 5.1)

. Il est plus judicieux de se concentrer sur la quantité d'ARN qui est un observable et laisser tomber la commutation de gène qui ne l'est pas.

Résumons : les ARN arrivent par bouffée, la probabilité d'avoir k ARN arriver en même temps est $W(n, k > 0) = f(k)$ indépendant de n : en effet, la synthèse d'ARN pour un gène non régulé ne dépend pas de la quantité d'ARN déjà présent. Par ailleurs, le phénomène de dégradation étant lent, nous continuons de le traiter comme avant :

$$W(n, -1) = \mu n ; W(n, k < -1) = 0$$

Cela veut dire que pendant le temps dt , la probabilité d'observer deux ou plus de dégradation en même temps est négligeable.

Posons

$$E = \sum_{k \geq 0} k f(k)$$

$$F = \sum_{k \geq 0} k^2 f(k)$$

Nous pouvons alors écrire pour la taille moyenne des sauts dans la quantité d'ARN (ce qui inclut les dégradations)

$$\langle k | n \rangle_S = E - \mu n$$

$$\langle k^2 | n \rangle_S = F + \mu n$$

et ceci nous donne l'équation d'évolution des moments

$$\frac{d \langle n(t) \rangle}{dt} = \langle E - \mu n \rangle = -\mu \langle n \rangle + E$$

$$\frac{d \langle n^2(t) \rangle}{dt} = \langle (F + \mu n) \rangle + \langle 2n (E - \mu n) \rangle$$

$$= -2\mu \langle n^2 \rangle + (2E + \mu) \langle n \rangle + F$$

Nous sommes en présence de deux équations de premier ordre couplées. Nous sommes surtout intéressés par l'état stationnaire ($t \rightarrow \infty$), qui est

$$\langle n \rangle = E / \mu$$

$$\langle n^2 \rangle = \frac{E^2}{\mu^2} + \frac{E}{2\mu} + \frac{F}{2\mu}$$

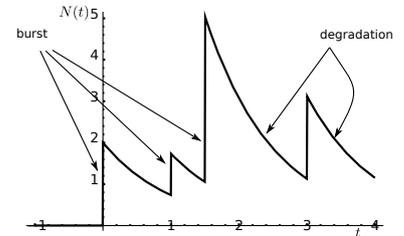


FIGURE 5.1 – les *bursts* d'ARN

Et un peu de calcul nous montre que le ratio variance sur moyenne vaut

$$R = \frac{\langle n^2 \rangle - \langle n \rangle^2}{\langle n \rangle} = \frac{1 + E}{2} + \frac{F - E^2}{2E}$$

Qu'on peut réécrire comme

$$R = 1 + b$$

Si les ARN arrivent par paquet de exactement 1, $E = 1$ et $F - E^2 = 0$; Dans ce cas, $R = 1$ est la distribution des ARN est effectivement poissonien. Si la taille des paquets d'ARN devient variable, le ratio R dépasse 1. Nous avons mentionné l'expérience de Golding qui a mesuré R pour un gène dans *E. Coli* et a trouvé la valeur de 4.

5.3.1 Formulation (légèrement) alternative de l'équation Maître.

L'équation maître telle que nous l'avons annoncé est implique une sommation sur les sauts k :

$$\frac{\partial P(n, t)}{\partial t} = \sum_k W(n - k, k)P(n - k, t) - W(n, k)P(n, t) \quad (5.13)$$

Nous calculons la balance entre le flux entrant et sortant du à la même taille (algébrique) de saut, et nous sommes ensuite sur tous les sauts (5.2). Nous pouvons changer ce point de vu et considérer la balance entre états : considérons les états $n - k$ et n . Le flux du premier vers le deuxième est $W(n - k, k)P(n - k, t)$; le flux du deuxième vers le premier est $W(n, -k)P(n, t)$. La différence entre ces deux est le flux net entrant dans l'état n du à l'état $n - k$; il suffit ensuite de sommer sur tous les k pour avoir le flux net entrant en n

$$\frac{\partial P(n, t)}{\partial t} = \sum_k W(n - k, k)P(n - k, t) - W(n, -k)P(n, t)$$

Nous aurions pu aboutir à ce résultat en changeant k en $-k$ dans la deuxième sommation de l'équation 5.13. D'ailleurs, si on formule le problème à l'aide des échanges entre états, autant introduire directement les transitions entre états T_{nm} , la probabilité de transiter de l'état n à l'état m ⁶, et l'équation maître devient alors

$$\frac{\partial P(n, t)}{\partial t} = \sum_m T_{mn}P(m, t) - T_{nm}P(n, t) \quad (5.14)$$

C'est essentiellement cette notation que l'on rencontre par exemple en mécanique quantique.

5.3.2 Réversibilité microscopique et théorème H.

La réversibilité veut dire qu'un événement et sa symétrique par renversement de temps ont la même probabilité d'occurrence. Dans le langage des transitions entre état, cela veut dire $T_{nm} = T_{mn}$.

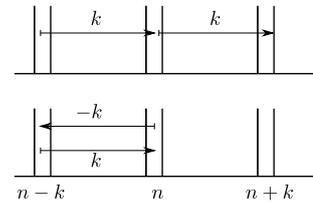


FIGURE 5.2 – Différentes façon de calculer la balance des flux.

6. Le lien entre les deux taux est $T_{nm} = W(n, m - n)$

Considérons un système réversible⁷ et définissons l'entropie du système par

$$S(t) = - \sum_n P(n, t) \log (P(n, t))$$

Le théorème H énonce que pour S doit être une fonction croissante du temps. En dérivant les deux côtés de la relation ci-dessus par rapport au temps, nous avons

$$\begin{aligned} \frac{dS}{dt} &= - \sum_n \frac{\partial P(n, t)}{\partial t} \log (P(n, t)) - \sum_n \frac{\partial P(n, t)}{\partial t} \\ &= - \sum_n \frac{\partial P(n, t)}{\partial t} \log (P(n, t)) \end{aligned}$$

puisque $\sum_n P(n, t) = 1$ et sa dérivée donc est nulle. En utilisant l'expression 5.14, et en utilisant la symétrie des taux $T_{nm} = T_{mn}$, nous obtenons

$$\frac{dS}{dt} = - \sum_{n,m} T_{nm} (P(m, t) - P(n, t)) \log (P(n, t))$$

Or, le système étant symétrique, nous pouvons échanger le rôle des deux indices et écrire de façon équivalente

$$\frac{dS}{dt} = - \sum_{n,m} T_{nm} (P(n, t) - P(m, t)) \log (P(m, t))$$

Nous pouvons même ajouter ces deux expressions pour avoir une expression entièrement symétrique

$$\frac{dS}{dt} = \frac{1}{2} \sum_{n,m} T_{nm} (P(n, t) - P(m, t)) (\log (P(n, t)) - \log (P(m, t))) \quad (5.15)$$

La fonction logarithme est une fonction monotone, ce qui veut dire que si $x > y$, alors $\log x > \log y$. Nous en déduisons que les deux parenthèses de la sommation sont de même signe et que leur produit est donc positif. Donc

$$\frac{dS}{dt} \geq 0$$

Noter également que si tout les états deviennent équiprobable, l'entropie cesse de croître et $dS/dt = 0$. Évidemment, si tous les états sont équiprobable, le calcul de l'entropie devient trivial

$$S = \log \Omega$$

où Ω est le nombre total d'état. C'est souvent comme cela que l'on démontre le théorème H en mécanique quantique, où la symétrie des taux de transition entre états est assurés par la règle de Fermi

$$T_{nm} = \left| \langle n | V | m \rangle \right|^2$$

où V est le potentiel d'interaction qui fait transiter de l'état n vers l'état m .

7. A l'échelle microscopique, tous les phénomènes physiques sont réversible

5.4 Les états continus.

La formulation précédente sur l'équation maîtresse quand les états sont discret se généralise facilement aux états continus. Soit dx la distance entre deux états n et $n + 1$. Nous introduisons la variable continue $ndx = x$ et la densité de probabilité $p(x, t)$. La probabilité $P(n, t)$ n'est autre que $p(x, t)dx$. De même, la probabilité $W(n, k)$ égale $w(x, y)dy$, $w(x, y)$ étant la densité de probabilité d'effectuer un saut de taille y . L'équation maîtresse se transforme alors en

$$\frac{\partial p(x, t)}{\partial t} = \int_{y=-\infty}^{+\infty} (w(x - y, y)p(x - y, t) - w(x, y)p(x, t)) dy$$

Ceci est une équation intégrale qui n'est parfois pas plus compliqué. Cependant, si la densité de probabilité $w(x, y)$ est bien centrée sur x , nous avons une méthode plus efficace de traiter l'équation maîtresse qui est celle de l'équation de diffusion que l'on verra au prochain chapitre. Nous nous contentons ici de donner un exemple de cette formulation.

5.4.1 Le théorème H en mécanique classique.

Dans les années 1870, Boltzmann a énoncé le théorème H dans sa forme classique, en tentant de modéliser la dynamique d'un grand nombre de molécule enfermés dans un réservoir isolé du monde. Nous donnons ici une version simplifiée de sa formulation, uniquement basée sur l'énergie. Notons x l'énergie et posons $p(x, t)$ la densité de probabilité pour une particule d'avoir l'énergie x . Les particules lors de leurs mouvements subissent des chocs les une avec les autres. Supposons que pour une particule, la probabilité des chocs par unité de temps est une constante α .

Lors d'un choc entre deux particule, l'énergie totale se conserve. Si x_1^a et x_2^a sont l'énergie des deux particules avant le choc et x_1^b x_2^b la même chose après le choc, nous devons avoir $x_1^a + x_2^a = x_1^b + x_2^b$; cependant, nous n'avons pas suffisamment d'équation de conservation pour déterminer la répartition exacte de l'énergie entre les deux particules. Boltzmann suppose alors que cette distribution est totalement aléatoire⁸, c'est à dire que

$$\begin{aligned} x_1^b &= q(x_1^a + x_2^a) \\ x_2^b &= (1 - q)(x_1^a + x_2^a) \end{aligned}$$

où q est une variable aléatoire uniformément distribuée entre $[0, 1]$.

Le calcul que nous allons effectuer est légèrement plus compliqué que ce que nous avons vu précédemment puisqu'il implique les interactions à deux corps, mais le principe est le même : calculer le flux entrant et sortant dans l'état x . Pour que l'état x s'enrichissent, il faut que deux particules d'énergie x_1 et x_2 se rencontre ET que l'énergie d'une des parties à la sortie soit égale à x . Le flux entrant est donc :

$$\alpha \int_{x_1=0}^{\infty} \int_{x_2=0}^{\infty} \int_{q=0}^1 p(x_1)p(x_2)\delta(q(x_1 + x_2) - x) dq dx_2 dx_1$$

8. C'est cela qu'on appelle l'hypothèse du chaos moléculaire.

Par contre, une particule d'énergie x qui subit un choc avec une particule de n'importe quelle autre énergie quitte obligatoirement cet état ; le flux sortant est donc

$$\alpha \int_{x_1=0}^{\infty} p(x)p(x_1)dx_1 = \alpha p(x)$$

L'équation Maîtresse s'écrit comme la différence de ces deux termes. De plus, en manipulant un peu le symbole δ , on simplifie l'intégrale triple pour aboutir à

$$\frac{\partial p(x,t)}{\partial t} = \alpha \left(\int_{q=0}^1 (1/q) \int_{x_1=0}^{x/p} p(x_1,t)p(x/q-x_1,t)dx_1dq \right) - \alpha p(x,t)$$

Si nous effectuons une transformée de Laplace par rapport à x , où s est la variable conjuguée à x , les choses deviennent encore plus simple⁹

$$\begin{aligned} \frac{\partial \hat{p}(s,t)}{\partial t} &= \alpha \int_{q=0}^1 \hat{p}(qs,t)^2 dq - \alpha \hat{p}(s,t) \\ &= \frac{\alpha}{s} \int_{u=0}^s \hat{p}(u,t)^2 du - \alpha \hat{p}(s,t) \end{aligned}$$

Il n'est pas difficile de vérifier que l'état stationnaire ($\partial_t \hat{p} = 0$) est donnée par $\hat{p}(s) = \beta / (\beta + s)$ et donc

$$p(x) = \beta e^{-\beta x}$$

En physique statistique, $1/\beta$ est appelé température T . De plus, un petit calcul variationnelle nous montre que la fonction ci-dessus est effectivement celle qui optimise la fonctionnelle

$$S = \int_0^{\infty} p(x) \log(p(x)) dx$$

qui est bien sûr l'entropie. Il y a encore beaucoup de chose à faire ici ; nous n'allons pas continuer ce problème, le lecteur intéressé se tournera vers la littérature de la physique statistique.

[Extension : linéariser autour de la solution stationnaire et traiter le problème temporel.]

5.5 Un rapide aperçu de la théorie des jeux.

5.6 Problèmes.

Exemple 5.7 Biologie de la mort.

[référence au dernier papier de Tadei sur le vieillissement bactérien].

Exemple 5.8 Nombre de descendants.

Calculer le nombre de descendant d'un individu dans le cadre du modèle de Moran de l'évolution vu plus haut : Un individu possède deux états vivant ($n = 1$) ou mort ($n = 0$) et sa densité de probabilité de mort est une constante α : $W(n \rightarrow n-1) = \alpha n$. Tant qu'il est vivant, il produit des descendant à taux fixe : $W(k \rightarrow k+1) = \alpha n$. Calculer $P(k,t)$ à la limite $t \rightarrow \infty$.

9. Une TL transforme les produits de convolution en produit simple, et ici, $p(x < 0) = 0$ puisque l'énergie des particules est seulement cinétique.

Exemple 5.9 Particules browniens couplés.

Soit deux particules browniens à 1 dimensions. La particule 1 est à gauche de la particule 2 et ne peut pas le dépasser. Les deux particules interagissent pour minimiser leurs distance. La position des deux particules étant repérées par les deux variables (n_1, n_2) , un modèle simple pour les taux de transitions seraient

$$W(n_1, n_2 \rightarrow n_1 + 1, n_2) = \alpha f(n_2 - n_1)$$

$$W(n_1, n_2 \rightarrow n_1 - 1, n_2) = \beta$$

$$W(n_1, n_2 \rightarrow n_1, n_2 + 1) = \beta$$

$$W(n_1, n_2 \rightarrow n_1, n_2 - 1) = \alpha f(n_2 - n_1)$$

où $f(p)$ est une fonction quelconque positive et $f(0) = 0$. Soit $x = (n_1 + n_2)/2$ et $w = (n_2 - n_1)$ la position moyenne et la distance entre les deux particules. Démontrer alors que

$$d \langle x \rangle / dt = 0$$

$$d \langle w \rangle / dt = 2\beta - 2\alpha \langle f(n_2 - n_1) \rangle$$

$$d \langle x^2 \rangle / dt = (\beta + \alpha \langle f(n_2 - n_1) \rangle) / 2$$

Résoudre le système pour $f(p) = p$. Déduire également $\text{Var}(w)$ dans ce cas. Que peut on supposer de la distribution des w ?

6

Probabilité de fixation et temps de premier passage.

6.1 L'équation maîtresse amont (backward).

Pour dériver l'équation maîtresse, nous nous sommes posés la question : quelle est la probabilité d'être dans l'état $P(n, t + dt)$. D'une part, nous avons fait un développement de Taylor de cette expression, d'autre part nous avons relié cette probabilité à celle d'être dans l'état $n - k$ au temps t . Une question que nous avons un peu éludé est celle de la condition initiale. Nous avons toujours supposé qu'au temps $t = 0$, le système était dans l'état n_0 et nous prenons cette information en compte au moment de la résolution effective des équations, en posant $P(n, t = 0) = \delta_{n, n_0}$.

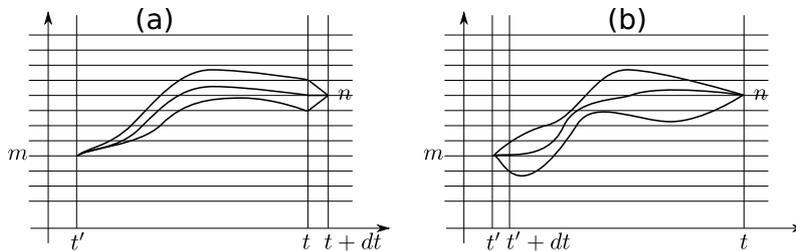


FIGURE 6.1 – dérivation de l'équation (a) aval (forward); (b) amont (backward)

Nous pouvons alourdir nos formules et intégrer explicitement la condition initiale dans les probabilités. Ainsi, nous écrivons $P(n, t | n_0, 0)$ la probabilité d'être dans l'état n au temps t sachant que nous étions dans l'état n_0 au temps 0. D'ailleurs, pour ne pas discriminer inutilement, nous pouvons écrire de façon très symétrique $P(n, t | m, t')$.

L'équation maîtresse que nous avons étudié s'appelle l'équation aval (forward), puisque nous déduisons $P(n, t + dt | m, t')$ (figure 6.1a). Nous aurions pu faire la même démarche, mais sur le point de départ au lieu du point d'arrivé (figure 6.1b) :

$$P(n, t | m, t') = \sum_k P(n, t | m - k, t' + dt) \bar{W}(m \rightarrow m - k) \quad (6.1)$$

A partir de là, nous avons juste à faire des développements limités. Notons que si $k \neq 0$, les facteur \bar{W} contiennent l'élément dt , c'est à dire

$$\bar{W}(m \rightarrow m - k) = W(m \rightarrow m - k) dt$$

où $W(m, m - k)$ est la densité de probabilité de saut de taille k (par unité de temps) que nous connaissons depuis le début ¹. Nous pouvons donc réécrire l'équation (6.1) sous la forme

$$P(n, t|m, t') = P(n, t|m, t' + dt) \left(1 - dt \sum_{k \neq 0} W(m \rightarrow m - k) \right) + dt \sum_{k \neq 0} P(n, t|m - k, t' + dt) W(m \rightarrow m - k)$$

Il nous suffit maintenant développer à l'ordre 1 en dt pour trouver que

$$-\frac{\partial P(n, t|m, t')}{\partial t'} = \sum_k (P(n, t|m - k, t') - P(n, t|m, t')) W(m \rightarrow m - k) \tag{6.2}$$

Nous appelons cette équation l'équation de maîtresse *amont*. Notez deux différences essentielles : d'abord, nous dérivons par rapport au temps t' de départ. Ensuite, le facteur $W(m \rightarrow m - k)$ multiplie la différence des probabilités, contrairement à ce que nous avons pour l'équation aval.

L'équation amont joue un rôle essentielle pour le calcul des probabilités d'absorption et les temps de premier passage que nous étudierons plus tard.

6.2 Probabilité de fixation.

Prenons un jeu simple de casino : vous misez une somme au départ; vous jetez les dès, si vous obtenez 4 2 1, vous doublez la mise, sinon, vous perdez 10€. Si il ne reste plus d'argent devant vous, vous avez perdu et le jeu s'arrête. Sinon, le jeu continue. Quelle est la probabilité p_0 pour que vous quittiez la table en ayant tout perdu? Le casino étant infiniment riche dans ce cas, la réponse est ... 1. Compliquons la chose : Le casino n'est pas infiniment riche, et vous pouvez quitter la table si vous gagnez une somme N . Le calcul de p_0 est cette fois moins trivial, comme nous allons le voir ci-dessous.

Cet exemple, illustré par la figure 6.2, est le prototype de nombreux problèmes apparaissant en science. Soit un processus stochastique $X(t)$ avec la condition initiale $X(t = 0) = x_0$. Quelle est la proportion $p_a(x_0)$ des trajectoires qui touchent un seuil a ? En physique, cela est par exemple le cas de particules méta-stables piégées dans un puits de potentiel. Si par fluctuation, une particule gagne suffisamment d'énergie pour atteindre la barrière, elle bascule dans un autre état de façon irréversible. Beaucoup d'équations chimiques et catalytique peuvent être traité dans ce cadre. La condition de rester dans un état a une fois qu'on l'a atteint est la nullité des taux de transition de cet état vers les autres : $W(a, a + k) = 0$.

L'équation amont/backward (6.2) que nous avons déduite est la clef du calcul des probabilités de fixation. Rappelons l'équation :

$$-\frac{\partial P(n, t|m, t')}{\partial t'} = \sum_k (P(n, t|m - k, t') - P(n, t|m, t')) W(m, m - k)$$

1. et que nous notons plus souvent $W(m \rightarrow m - k)$

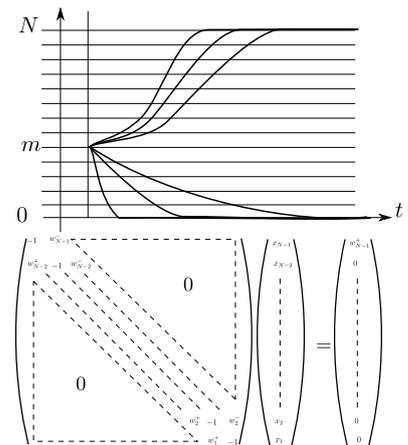


FIGURE 6.2 – Haut : États absorbants. Dans cet exemple, les deux états 0 et N sont absorbant, si un trajectoire stochastique touche une des limites, il y reste. $\pi_N(m)$, la probabilité d'absorption à l'état N en partant de l'état m , est le nombre relatif de tous les trajectoires qui touchent cet état au bout d'un temps infini. . Bas : le système linéaire $N - 1 \times N - 1$ (6.5), où $x_n = \pi_N(n)$.

La probabilité de fixation est celle d'être dans l'état N au bout d'un temps infini, puisque si on a atteint cet état à un temps quelconque, on y reste ! Si nous appelons $\pi_N(m, t')$ la probabilité de fixation en partant de l'état initial m au temps t' , $\pi_N(m, t') = P(N, \infty | m, t')$; d'après l'équation amont ci-dessus, nous avons donc

$$-\frac{\partial \pi_N(m, t')}{\partial t'} = \sum_k (\pi_N(m-k, t') - \pi_N(m, t')) W(m, m-k)$$

Or, si les probabilité de saut ne dépendent pas du temps, la probabilité de fixation ne peut pas dépendre du temps t' de départ non plus et nous devons avoir $\partial_{t'} \pi_N(m, t') = 0$. Nous aboutissons donc à un système d'équations linéaires qui relie les probabilités de fixation en partant d'un point donné les uns aux autres :

$$\sum_k \pi_N(m-k) W(m, m-k) = \pi_N(m) \sum_k W(m, m-k) \quad (6.3)$$

nous pouvons simplifier encore cette expression en normalisant les probabilité de saut et en posant

$$w(m, m+k) = W(m, m+k) / \sum_k W(m, m+k)$$

Le sens des $w(m, m+k)$ est le suivant² : étant dans l'état m , une fois qu'un événement de transition arrive, $w(m, m+k)$ est la probabilité que la transition soit vers l'état $m+k$.

L'équation (6.3) devient :

$$\pi_N(m) = \sum_k \pi_N(m-k) w(m, m-k) \quad (6.4)$$

Ceci est juste un système k -diagonal d'équations linéaires. Dans le cas où nous n'avons que des sauts entre premiers voisins, ceci se réduit à un système tri-diagonal :

$$w^-(m) \pi_N(m-1) - \pi_N(m) + w^+(m) \pi_N(m+1) = 0 \quad (6.5)$$

Ce système se résout directement par récurrence dans le cas où nous avons deux états absorbant 0 et N . La solution se construit comme suit : posons

$$\rho_0 = 1, \quad \rho_k = \prod_{j=1}^k \left(\frac{w^-(m)}{w^+(m)} \right)$$

et nous avons alors

$$\pi_N(m) = \sum_{k=0}^{m-1} \rho_k / \sum_{k=0}^{N-1} \rho_k$$

Cette formule s'obtient assez facilement. Pour simplifier les notation, posons $\pi_N(m) = x_m$. Nous avons $x_0 = 0$ et $x_N = 1$. En utilisant le fait que, par définition, $w^+(m) + w^-(m) = 1$, nous pouvons réécrire la récurrence (6.5) comme

$$w_m^-(x_m - x_{m-1}) = w_m^+(x_{m+1} - x_m)$$

Nous voyons donc que les différences $y_m = x_m - x_{m-1}$ suivent une progression géométrique $y_{m+1} = (w_m^- / w_m^+) y_m$, ce qui nous amène directement à la solution donné plus haut.

2. Voir le chapitre sur l'algorithme de Gillespie

Exemple 6.1 Mouvement Brownien.

Prenons une particule brownien biaisé avec deux états absorbant 0 et N . les probabilités de sauts vers la gauche et la droite sont

$$\begin{aligned}W^+(n) &= \beta(1+s) \\W^-(n) &= \beta(1-s)\end{aligned}$$

où s est le biais (vers la droite si $s > 0$). Les taux normalisés sont

$$w^\pm(n) = (1 \pm s)/2$$

et $\rho_k = [(1-s)/(1+s)]^k$. Posons $c = (1-s)/(1+s)$, nous voyons que la somme d'une série géométrique nous donne immédiatement

$$\pi_N(m) = \frac{1-c^m}{1-c^N}$$

Exemple 6.2 Probabilité de fixation en Évolution Darwinienne.

Reprenons l'exemple de l'évolution Darwinienne que nous avons traité en §4.1. Rappelons les taux de transitions

$$\begin{aligned}W^-(n) &= \mu n(N-n)/N \\W^+(n) &= \mu(1+s)n(N-n)/N\end{aligned}$$

Les taux normalisés sont

$$w^-(n)/w^+(n) = 1/(1+s)$$

et nous voyons que si on pose $c = 1/(1+s)$, nous avons de nouveau

$$\pi_N(m) = \frac{1-c^m}{1-c^N}$$

Nous voyons ici que bien que les probabilités de transition soient très différentes dans les deux cas traités, les probabilité de fixation, qui ne font intervenir que les rapports des probabilité de saut, sont les mêmes. Les probabilités de saut nous donnent le temps (aléatoire) qu'un prochain événement arrive ; une fois que un événement arrive, sa nature est donné par les rapports de probabilité de transitions. Dans le calcul des probabilités de fixations, le temps n'est pas important et c'est pourquoi deux processus apparemment différent donne les même π_N . La distinction entre le temps du prochain événement et sa nature et au cœur de l'algorithme de Gillespie que nous traitons au prochain chapitre.

Exemple 6.3 Apparition de la coopération en Évolution Darwinienne : "kin selection".

Le développement des comportements de coopération posaient problème à certains évolutionnistes : pourquoi aider les autres à ses propres dépens pourrait être sélectionné ? Une première réponse mathématique a été donné par Hamilton dans les

années 1960. Prenons un mutant qui aide ses congénères également mutants, en abaissant ainsi son taux de croissance. Les taux de transition sont

$$\begin{aligned} W^-(n) &= c\mu n(N-n)/N \\ W^+(n) &= \mu n(N-n)(1+ns)/N \end{aligned}$$

Plus il y a de ce type de mutant, plus ils s'aident et favorisent ainsi la coopération.

6.3 Probabilité de survie & temps de fixation.

Nous avons calculé la probabilité de fixation $\pi_N(m)$, c'est à dire la probabilité d'atteindre l'état absorbant N en partant de l'état m . Le temps ici n'avait pas d'importance : nous lâchons notre particule en m au temps $t = 0$, attendons jusqu'au temps $t = \infty$, et comptons le nombre relatif de trajectoire qui atteignent l'état N .

Mais les différentes trajectoires qui atteignent l'état N le font à des durées différentes, et nous pouvons nous demander qu'elle est la durée moyenne du processus d'absorptions. Ceci est d'une importance cruciale si par exemple nous étudions les équations chimique : la probabilité d'atteindre l'état stable est 1, mais si cette réaction prend un temps infini, cette information est de peu d'utilité pour nous³. Nous allons voir que le calcul du temps moyen d'absorption $\bar{t}(m)$ est en tout point similaire au calcul des probabilités de fixation.

3. Nous allons voir plus bas que dans ce cas précis, le temps de fixation est effectivement donné par un processus d'Arhenius.

6.3.1 Probabilité de survie.

Reprenons la quantité $P(N, t|m, t_0)$, la probabilité d'observer la particule dans l'état absorbé N au temps t , sachant que nous l'avons lâché en m au temps t_0 . Remarquons tout de suite que comme nos taux de transitions W^\pm ne dépendent pas du temps, les temps absolus t et t_0 n'ont aucune importance, seul leur différence $T = t - t_0$ compte et donc

$$P(N, 0|m, -T) = P(N, T|m, 0)$$

Soit maintenant $Q(m, T)$ la probabilité d'observer la particule "vivante" (pas absorbée) au bout d'une durée T , l'ayant lâché dans l'état m . Nous avons

$$Q(m, T) = \sum_{n \neq N} P(n, T|m) = 1 - P(N, T|m)$$

et il se trouve que Q obéit à une équation très similaire à celle sur les P :

$$\begin{aligned} \frac{\partial Q(m, T)}{\partial T} &= \sum_{n \neq N} \frac{\partial}{\partial T} P(n|m, -T) \\ &= \sum_{n \neq N} \sum_k [P(n|m-k, -T) - P(n|m)] W(m \rightarrow m-k) \end{aligned}$$

et si maintenant nous changeons l'ordre de sommation, nous avons

$$\frac{\partial Q(m, T)}{\partial T} = \sum_k [Q(m-k, T) - Q(m, T)] W(m \rightarrow m-k)$$

Notons que par définition, $Q(N, T) = 0$: si nous sommes dans l'état N , nous sommes mort!

6.3.2 Temps de fixation (temps de premier passage).

$Q(m, T)$ est la probabilité que la particule soit encore "vivante" au bout d'une durée T , $-\partial_T Q(m, T)$ est donc la densité de probabilité que la particule ne soit pas absorbée dans l'intervalle $[T, T + dT[$. Le temps moyen de survie est donc

$$\begin{aligned} \bar{T}(m) &= - \int_0^\infty T \partial_T Q(m, T) dT \\ &= - [TQ(m, T)]_0^\infty + \int_0^\infty Q(m, T) dT \\ &= \int_0^\infty Q(m, T) dT \end{aligned} \quad (6.6)$$

$$= \int_0^\infty (1 - P(N, T|m)) dt \quad (6.7)$$

Il suffit maintenant de former la quantité

$$I = \sum_k W(m \rightarrow m-k) (\bar{T}(m) - \bar{T}(m-k))$$

et utiliser la relation (6.6) pour obtenir un système d'équations linéaires pour les $\bar{T}(m)$:

$$\sum_k W(m \rightarrow m-k) (\bar{T}(m-k) - \bar{T}(m)) = -1 \quad (6.8)$$

Pour cela, nous avons évidemment utilisé les relations $Q(m, \infty) = 0$, $Q(m, 0) = 1$.

Approche intuitive.

Le résultat (6.8) peut facilement s'expliquer. Prenons le cas particulier où nous pouvons seulement avoir des saut de taille 1, la généralisation aux autres cas étant immédiate. Le temps moyen d'absorption à partir de l'état m égale au temps moyen pour *sortir* de l'état m plus la somme pondérée des temps moyen d'être absorbée à partir de l'état $m \pm 1$ que l'on a atteint en sortant de l'état m .

- Le temps moyen pour sortir de l'état m : $1 / (W^+(m) + W^-(m))$
- Probabilité d'atteindre l'état $m \pm 1$ à partir de l'état m , sachant qu'un saut a eu lieu :

$$W^\pm(m) / (W^+(m) + W^-(m))$$

Nous avons donc :

$$\begin{aligned}\bar{T}(m) &= \frac{1}{(W^+(m) + W^-(m))} + \\ &\quad \frac{W^+(m)}{(W^+(m) + W^-(m))} \bar{T}(m+1) + \\ &\quad \frac{W^-(m)}{(W^+(m) + W^-(m))} \bar{T}(m-1)\end{aligned}$$

En multipliant par $W^+(m) + W^-(m)$ et en regroupant les termes, nous trouvons alors

$$\begin{aligned}W^+(m) (\bar{T}(m+1) - \bar{T}(m)) &+ \\ W^-(m) (\bar{T}(m-1) - \bar{T}(m)) &= -1\end{aligned}$$

Exemple 6.4 le mouvement brownien.

Soit une particule brownien dans la partie $n > 0$ de l'axe, avec des taux de transition vers la gauche et la droite de $W^\pm(n) = \alpha$. Nous avons alors

$$\alpha [\bar{T}(m+1) + \bar{T}(m-1) - 2\bar{T}(m)] = -1 \quad m > 0$$

Nous pouvons bien sûr utiliser le même procédé pour évaluer les moments supérieurs du temps de fixation.

6.3.3 Complément : résolution des équations au récurrence.

Les méthodes exposées ci-dessus nécessite souvent de résoudre des équations au récurrence de type

$$a_{n+1} = f_n a_n + g_n \quad (6.9)$$

Nous sommes plus habitués aux équations différentielles, qui apparaissent comme des approximations des équations ci-dessus et que nous traiterons en détail au chapitre sur l'équation de diffusion. Cependant, la résolution de ce type d'équation n'est guère plus compliquée.

Supposons dans un premier temps que $g_n = 0$. La solution de l'équation est alors immédiate :

$$a_{n+1} = \left(\prod_{k=0}^n f_k \right) a_0$$

en posant

$$\mu_{n+1} = \prod_{k=0}^n f_k$$

cela s'écrit facilement $a_n = \mu_n a_0$, où a_0 est donné par la condition initiale que nous supposons connu. Pour garder la cohérence, nous posons de plus $\mu_0 = 1$.

Revenons à l'équation non-homogène (6.9) et posons $a_n = \mu_n A_n$. Nous avons alors

$$\mu_{n+1} A_{n+1} - f_n \mu_n A_n = g_n$$

Or, $f_n \mu_n = \mu_{n+1}$ par définition, et l'équation se transforme en

$$A_{n+1} - A_n = \frac{g_n}{\mu_{n+1}}$$

et en sommant toutes les équations, nous trouvons

$$A_{n+1} = A_0 + \sum_{k=0}^n \frac{g_k}{\mu_{k+1}}$$

où encore

$$a_{n+1} = \mu_{n+1} a_0 + \mu_{n+1} \sum_{k=0}^n \frac{g_k}{\mu_{k+1}}$$

De plus, il est inutile de faire des multiplications inutiles. En posant

$$\gamma_k^{n+1} = \frac{\mu_{n+1}}{\mu_{k+1}} = \prod_{i=k+1}^n f_i$$

nous avons

$$a_{n+1} = \mu_{n+1} a_0 + \sum_{k=0}^n \gamma_k^{n+1} g_k$$

Quand nous étudions les équations différentielles, nous appelons cette méthode de résolution "la variation des constante".

7

Intégrales de chemins.

7.1 Formulation générale.

Dans le précédent chapitre, nous avons fait appelle, sans insister beaucoup, aux *chemins* aléatoires : l'équation amont est un exemple où l'on prend en compte le nombre (infini) de chemin menant d'un point à un autre.

Nous pouvons aller plus loin dans cette approche. Soit un processus de Markov discrète suivant un opérateur \mathcal{L} (la matrice qui collecte les taux de transitions $W(n \rightarrow m)$) dont l'équation Maîtresse (écrite sous forme compact) est

$$\partial_t |P(t)\rangle = \mathcal{L} |P(t)\rangle \quad (7.1)$$

Le vecteur $|P(t)\rangle = (P(0,t), P(1,t), \dots)^T$ collecte les probabilités individuelles $P(n, t|n_0, 0)$; les éléments de la matrice \mathcal{L} sont donnés par

$$\mathcal{L}_m^n = W(m \rightarrow n) \ ; \ \mathcal{L}_n^n = - \sum_m W(n \rightarrow m)$$

Supposons maintenant qu'à chaque *chemin* aléatoire $n(t)$, $t \in [0, T]$ nous associons une valeur S

$$S[T] = \int_0^T c(t) dt$$

où la fonction $c(u)$ est une fonction quelconque (et suffisamment lisse). Cette valeur peut-être par exemple l'énergie d'un polymère, l'action en mécanique, ou simplement la moyenne de la variable le long du chemin. Peut-on calculer la moyenne, sur *tous les chemins*, de cette quantité ?

Pour faire ce travail par la pensée (ou numériquement, voir chapitre 8), nous générons un chemin, calculons S_1 sur ce chemin et enregistrons sa valeur. Nous effectuons cette opération M fois ($M \rightarrow \infty$) et obtenons S_1, S_2, \dots, S_M . Nous calculons ensuite la moyenne de ces valeurs. Nous notons cette opération $\langle S[n(t)] \rangle$, sachant que l'opérateur $\langle \rangle$ désigne la moyenne sur *les chemins*.

Le nombre de chemins aléatoire est un très grand infini : il se trouve cependant que mathématiquement, cette moyenne existe la plupart du temps, en tout cas tant que \mathcal{L} est suffisamment "sage" et que $c(t)$ est réelle (pour prendre en compte les $c(u)$ complexe, il

faut prendre un peu plus de précaution). Non seulement elle existe, mais on peut le calculer.

Considérons la forme

$$G(n, T, 0) = \left\langle e^{\int_0^T c[n(t)]dt} \right\rangle \quad (7.2)$$

L'exponentielle apparaît souvent dans le calcul de l'énergie libre par exemple, où de l'action en mécanique quantique. De façon générale, G désigne la fonction génératrice des moments que nous avons souvent rencontré. Cette moyenne $\langle \rangle$ s'effectue sur tous les chemins partant au temps 0 de l'état initial n et arrivant quelque part au temps T .

Comme d'habitude, nous allons calculer $G(n, T + dT)$ et le relier à $G(n, T)$ à travers une équation différentielle, en utilisant le fait que

$$\int_0^{T+dT} c[n(t)]dt = \int_0^{dT} + \int_{dT}^{T+dT} = c[n]dT + \int_{dT}^{T+dT} c[n(t)]dt$$

Nous avons donc pour G :

$$G(n, T + dT, 0) = e^{c(n)dT} \left\langle e^{\int_{dT}^{T+dT} c[n(t)]dt} \right\rangle \quad (7.3)$$

le terme $\langle \rangle$ de droite n'a pas pour l'instant la forme de G , puisque les chemins ne démarrent pas d'un état m donné. Nous pouvons partitionner tous ces chemins par leur point de départ m et écrire

$$\left\langle e^{\int_{dT}^{T+dT} c[n(t)]dt} \right\rangle = \sum_m P(m) G(m, T + dT, dT)$$

où $P(m)$ est le nombre relatif de chemin démarrant à m au temps dT , sachant qu'ils ont démarré à n au temps 0.

Comme nous considérons des processus qui ne dépendent pas explicitement en temps, nous avons une symétrie de translation dans le temps :

$$G(m, T + dT, dT) = G(m, T, 0)$$

par ailleurs, pour notre processus stochastique, sachant qu'on part de l'état n avec une probabilité de 1 :

$$\begin{aligned} P(m, dT) &= W(n \rightarrow m)dT \text{ si } m \neq n \\ &= 1 - \sum_{m \neq n} W(n \rightarrow m)dT \text{ sinon} \end{aligned}$$

Finalement, en écrivant $e^{c[n]dT} = 1 + c(n)dT + O(dT^2)$, en remplaçant tout dans la relation (7.3) et en ne gardant que le terme d'ordre 1 en dT , et en écrivant $G(n, T)$ au lieu de $G(n, T, 0)$, nous avons

$$\begin{aligned} G(n, T + dT) &= (1 + c(n)dT) \times \\ &\quad \left\{ \sum_{m \neq n} W(n \rightarrow m)dT G(m, T) + \right. \\ &\quad \left. \left(1 - \sum_{m \neq n} W(n \rightarrow m)dT \right) G(n, T) \right\} \end{aligned}$$

ou encore

$$\frac{\partial G(n, T)}{\partial T} = \sum_{m \neq n} \{W(n \rightarrow m)G(m, T) - W(n \rightarrow m)G(n, T)\} + c(n)G(n, T) \quad (7.4)$$

L'équation ci-dessus est connu sous le nom de Feynman-Kac. On peut l'écrire sous forme compact comme

$$\partial_t \langle G(T) | = \langle G(T) | (\mathcal{L}^\dagger + \mathcal{D}) \quad (7.5)$$

où le vecteur ligne $\langle G(T) |$ collecte les $G(n, T)$, la matrice \mathcal{L}^\dagger est la transposé de la matrice \mathcal{L} et \mathcal{D} est une matrice diagonale

$$\mathcal{D}_n^n = c(n)$$

7.2 *Physique statistique.*

7.3 *Théorie des grandes déviations.*

Résolution numérique des équations maîtresse.

L'outil numérique est indispensable pour comprendre les grands traits d'un système stochastique un peu compliqué; de plus, cet outil nous permet d'avoir une compréhension beaucoup plus concrète des équations stochastiques.

La simulation des équations stochastiques discrètes a été proposée dans les années 1970 par Gillespie¹ et l'algorithme porte son nom. Considérons le cas simple d'une seule variable aléatoire, avec les taux de transition $W^+(n)$ et $W^-(n)$. Bien sûr, nous pouvons directement résoudre les N équations différentielles couplées de l'équation maîtresse gouvernant les $P(n, t)$; cela a ses propres complications, mais peut être très efficace.

La méthode de Gillespie consiste au contraire de générer des trajectoires stochastiques, est extraire les probabilités directement de ces trajectoires : on commence par le système à l'état n_0 au temps $t_0 = 0$, on détermine le prochain évènement qui fait passer le système dans l'état $n + k$ au temps t_1 , et on boucle pour générer une trajectoire. On recommence ensuite à partir de l'état initiale pour générer une nouvelle trajectoire, et ainsi de suite. Une fois que nous avons généré suffisamment de trajectoires, nous pouvons faire des statistiques.

Revenons à notre système à une variable avec les taux $W^\pm(n)$: si à un instant t , le système se trouve dans un état n , il a une certaine chance de passer vers l'état $n \pm 1$ selon les taux $W^\pm(n)$. La première question que nous devons poser est : au bout de quel temps cette transition est effectuée? Évidemment, ce temps est une variable aléatoire, et il est le résultat de deux processus (création ou annihilation) en parallèle. L'évènement "création" est un processus poissonien avec un taux $W^+(n)$. Cela veut dire que la densité de probabilité du temps d'arrivée d'un évènement de création est

$$p^+(t) = W^+(n) \exp[-W^+(n)t]$$

la distribution des temps de l'évènement annihilation est également poissonien avec un taux $W^-(n)$. Nous pouvons maintenant nous demander : quel est la densité de probabilité des temps qu'un évènement, quelque soit sa nature (création OU annihilation) arrive? Les deux processus sont parallèle; la somme de deux processus poissonien est un processus poissonien dont le taux est la somme

1. Daniel T. Gillespie (1977). "Exact Stochastic Simulation of Coupled Chemical Reactions". *The Journal of Physical Chemistry* 81 (25) : 2340–2361.

des deux premiers. La réponse à la question précédente est donc

$$p(t) = W(n) \exp [W(n)]$$

où

$$W(n) = W^+(n) + W^-(n)$$

maintenant que nous savons le prochain temps d'un évènement, la question est quelle est sa nature? Cela est directement proportionnel aux taux. La probabilité que se soit un évènement de création ou d'annihilation est

$$p^{up} = \frac{W^+(n)}{W(n)} ; p^{down} = \frac{W^-(n)}{W(n)}$$

L'algorithme de Gillespie consiste donc aux étapes suivantes (voir figure 8.1) :

1. Le système étant dans l'état n , calculer $W^+(n)$, $W^-(n)$ et $W(n)$.
2. Générer un temps poissonien τ avec $W(n)$: cela revient à générer une variable aléatoire uniforme $x \in [0, 1]$ et poser $\tau = -\ln(x)/W(n)$. Incrémenter le temps : $t \leftarrow t + \tau$
3. Générer une variable aléatoire uniforme $y \in [0, 1]$. Si $y \in [0, W^+(n)/W(n)]$ incrémenter n : $n \leftarrow n + 1$. Sinon, décrémenter n : $n \leftarrow n - 1$.
4. Ajouter le couple (t, n) à un tableau et recommencer à l'étape 1 tant que $t < T$, où T est le temps maximum alloué.

Une fois le temps écoulé, l'ensemble des couple (t_i, n_i) constituent une trajectoire stochastique. Il suffit de recommencer pour générer autant de trajectoire que la précision souhaité l'exige.

La grande souplesse du programme réside dans sa *généralisabilité*. Supposons que nous avons M variables aléatoires X_i et que les taux dépendent de l'abondance de chaque variable

$$W_i^\pm = W_i^\pm(n_1, n_2, \dots, n_M)$$

cela ne change rien à l'algorithme : à chaque étape, $W = \sum W_i^\pm$, le temps généré est une variable poissonien avec ce taux. Ensuite, il suffit de partitionner l'intervalle $[0, 1]$ en fonction de l'importance de chaque taux, générer la variable uniforme y et déterminer dans quel sous intervalle elle tombe. Et Voilà!

Rien ne vaut un petit TP numérique pour comprendre les étapes de cet algorithme.

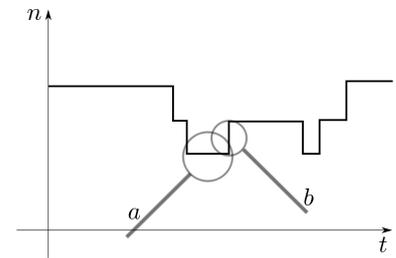


FIGURE 8.1 – Algorithme de Gillespie. (a) le temps jusqu'à l'occurrence du prochain évènement est une variable poissonienne avec la densité de probabilité $p(t) = S \exp(-St)$ où $S = W^+(n) + W^-(n)$; (b) la nature du prochain évènement (incrémentatation ou décrémentatation) est une "pile ou face" biaisée où $p^{up} = W^+(n)/S$ et $p^{down} = W^-(n)/S$.

9

L'équation de diffusion (Fokker-Planck).

L'équation de diffusion est une équation aux dérivées partielles qui remplace l'équation maîtresse dans deux cas : (i) comme une approximation au système discret ; (ii) comme une formulation exacte quand les processus sont naturellement continue, comme par exemple le mouvement Brownien. Dans ce dernier cas cependant, les passages à la limite pour le justifier sont délicats, et il faut imposer des conditions restrictives aux densités de probabilité de saut et on appelle alors cela des processus de Wiener. L'équation de diffusion est également connue dans la littérature physicienne sous le nom de l'équation Fokker-Planck tandis que les mathématiciens y réfèrent sous le nom de Kolmogorov.

9.1 L'équation aval (forward).

Oublions pour un instant nos processus aléatoire et posons nous un problème de mathématiques appliquées : nous devons résoudre une équation à dérivée partielle, mais nous manquons de solution exacte et souhaitons la résoudre numériquement. Comment faire ? Prenons par exemple une équation de la chaleur pour la fonction $u(x, t)$ défini sur l'intervalle $[0, L]$

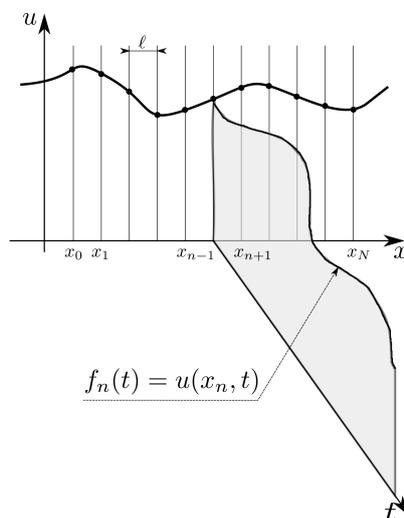


FIGURE 9.1 – Schéma de différences finies : la fonction $u(x, t)$ est échantillonnée aux points $x_n = n\ell$ de l'espace ; Si les fonctions $f_n(t) = u(x_n, t)$ peuvent être calculées, nous disposons alors d'une bonne approximation de la fonction $u(x, t)$.

$$\frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial x^2} \quad (9.1)$$

La façon classique est de précéder par la technique des différences finies¹ : au lieu de chercher la fonction $u(x, t)$, échantillonnons l'espace en choisissant des points $x_n = n\ell$ (voir figure 9.1). Si nous connaissons la valeur de la fonction u seulement en ses points, nous pouvons estimer le problème résolu pourvu que l'échantillonnage ℓ soit suffisamment serré. Nous cherchons donc $N (= L/\ell)$ fonctions inconnues de la seule variable t $f_n(t)$, où $f_n(t) = u(x_n, t)$. Un développement de Taylor nous indique que

$$\left. \frac{\partial^2 u}{\partial x^2} \right|_{x_i} = \frac{u(x_{i-1}, t) + u(x_{i+1}, t) - 2u(x_i, t)}{\ell^2} + O(\ell)$$

En négligeant² les termes en $O(\ell)$ et en remplaçant dans l'équation 9.1, nous trouvons alors

$$\frac{df_n(t)}{dt} = \frac{D}{\ell^2} (f_{n-1}(t) - 2f_n(t) + f_{n+1}(t)) \quad (9.2)$$

C'est à dire que nous avons transformé une équation à dérivée partielle en N équations différentielles simple en la *discrétisant*. Le lecteur aura d'ailleurs reconnu ci-dessus l'équation maîtresse du mouvement Brownien avec $\alpha = D/\ell^2$ que nous avons traité en détail dans les chapitres précédents.

Nous pouvons inverser la technique utilisée ci-dessus : ayant N équations différentielles couplées, nous pouvons obtenir une équation à dérivée partielle dont la solution soit une bonne approximation des équation de départ. Appliquées aux processus stochastiques et à l'équation maîtresse, cette équation à dérivée partielle s'appelle l'équation de diffusion.

Voyons cela de plus près. Comme précédemment (cf section §5.3) notons $W(n, k)$ la densité de probabilité d'effectuer un saut de taille k quand on se trouve dans l'état n . Notre équation maîtresse s'écrit

$$\frac{\partial P(n, t)}{\partial t} = \sum_k [W(n-k, k)P(n-k) - W(n, k)P(n)] \quad (9.3)$$

Supposons que les états sont séparés par une distance ℓ et introduisons la variable continue $x = n\ell$. Introduisons également la densité de probabilité $p(x) : P(n)$, la probabilité d'être dans l'état n est donc la probabilité d'être entre les positions x et $x + \ell$ et vaut donc

$$P(n, t) = p(x, t)\ell$$

De la même façon, la probabilité de saut $W(n, k) = w(x, y)\ell$ où $y = k\ell$. Si $\ell \rightarrow 0$, la somme se transforme en une intégrale (la coutume veut alors qu'on note dy à la place de ℓ) :

$$\frac{\partial p(x, t)}{\partial t} = \int_{-\infty}^{+\infty} \{w(x-y, y)p(x-y) - w(x, y)p(x)\} dy$$

1. C'est d'ailleurs comme cela que les pères fondateurs (d'Euler jusqu'à Poisson) imaginaient les équations différentielles.

2. Négliger ces termes nécessite quelques précautions : en général, il faut choisir $\ell < C$ où C est une constante qui dépend des paramètres du problème et de la précision exigée, pour assurer que pendant l'intervalle de résolution T , la solution numérique reste proche de la solution exacte à ϵ près. Pour nous ici ceci n'est pas un problème, nous pouvons prendre ℓ aussi petit que l'on veut pour illustrer le principe.

Nous pouvons développer à l'ordre 2 le terme entre $\{\}$ dans l'intégrale

$$\{\} = \frac{\partial (w(x,y)p(x))}{\partial x}(-y) + \frac{\partial^2 (w(x,y)p(x))}{\partial x^2} \frac{y^2}{2} + O(y^3)$$

Par ailleurs, la dérivée est sur la variable x (la position), tandis que l'intégrale est sur la variable y (les sauts). Nous pouvons échanger l'ordre de ces deux opérations. Si nous définissons

$$\begin{aligned} a(x) &= \int_{-\infty}^{+\infty} yw(x,y)dy \\ b(x) &= \int_{-\infty}^{+\infty} y^2w(x,y)dy \end{aligned}$$

et négligeons les termes d'ordres supérieures, nous aboutissons à l'équation de diffusion

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial (a(x)p(x,t))}{\partial x} + \frac{1}{2} \frac{\partial^2 (b(x)p(x,t))}{\partial x^2} \quad (9.4)$$

$a(x)$ est le saut moyen à partir de la position x , et est souvent appelé le terme de *drift*. $b(x)$ est le deuxième moment des sauts à partir de la position x et est souvent appelé le coefficient de diffusion³. Notons enfin que $a(x)$ a la dimension d'une vitesse si x désigne l'espace et t le temps.

Nous pouvons effectuer toutes les manipulations que nous faisons avant avec les équations maîtresse à partir de l'équation de diffusion. Remarquons d'abord que $p(x)$ est une densité de probabilité, et donc $\int_I p(x)dx = 1$ où $I =]-\infty, +\infty[$. Ceci implique que $p(x) \rightarrow 0$ quand $x \rightarrow \infty$ et cela plus rapidement que $1/x$. Dans les processus *physique*, $p(x)$ converge beaucoup plus rapidement que cela et tombe à zéro en dehors d'un intervalle fini. Cela veut dire que nous ne considérons que les processus où *tous* les moments existent : $\int_I x^n p(x) < \infty$. $p(x)$ converge vers zéro plus rapidement que x^{-n} ($\forall n > 0$) quand $x \rightarrow \infty$; cela est valable également pour ses dérivées.

Soit maintenant $\langle x(t) \rangle = \int_I xp(x,t)dx$ la position moyenne au temps t . Multiplions les deux côtés de l'équation 9.4 par x et intégrons sur x ; en échangeant à gauche l'intégration sur x et la dérivation sur le temps et en effectuant une intégration par partie à droite, nous obtenons :

$$\begin{aligned} \frac{d \langle x(t) \rangle}{dt} &= [-xa(x)p(x) + x\partial_x (b(x)p(x,t))]_{-\infty}^{+\infty} \\ &+ \int_I a(x)p(x,t)dx + \int_I \partial_x (b(x)p(x,t)) dx \end{aligned}$$

Le premier terme est nul tant que les coefficients $a(x)$ et $b(x)$ sont polynomiaux en x . Le troisième terme est l'intégrale d'une dérivée est s'annule pour les mêmes raisons, ce qui nous donne

$$\begin{aligned} \frac{d \langle x(t) \rangle}{dt} &= \int_I a(x)p(x,t)dx \\ &= \langle a(x) \rangle \end{aligned}$$

3. Cette dérivation est connu sous le nom d'expansion Kramers-Moyal, qui l'ont proposé dans les années 1940.

Ce qui est la version continue des résultats 5.11 ou 4.3. Un résultat analogue peut être obtenu pour le moment d'ordre 2 :

$$\frac{d \langle x^2(t) \rangle}{dt} = 2 \langle xa(x) \rangle + \langle b(x) \rangle$$

9.1.1 *La précision de l'approximation.*

Nous avons négligé, dans l'équation de diffusion, les termes d'ordre supérieur à deux en les traitant de négligeable. Combien cela est-il justifié? La réponse est dans les probabilités de saut $w(x, y)$. Si le système que nous étudions est naturellement discret, comme c'est le cas pour le nombre de particule ou de molécules chimiques, nous avons naturellement une longueur ℓ à notre disposition; si λ est la racine carré de la variance des sauts, l'approximation est de l'ordre de $O((\lambda/\ell)^3)$ et l'erreur relative de l'ordre de λ/ℓ . C'est la valeur de cette quantité qui nous justifie ou non dans le choix de l'équation de diffusion comme approximation. Le mieux est d'étudier un exemple concret, l'exemple de l'évolution Darwinienne et le processus de Moran, où rappelons le

$$\begin{aligned} W^+(n) &= (1+s)n(N-n)/N \\ W^-(n) &= n(N-n)/N \end{aligned}$$

N est le nombre total d'individu que l'écosystème étudié peut supporter. Introduisons la variable continue $x = n/N$ et le pas de discrétisation naturelle $dx = 1/N$; ainsi, un saut nous fait passer de l'état x à l'état $x \pm dx$. En notant $P(n, t) = p(x, t)dx$ et $W^\pm(n) = w^\pm(x)$ et en développant l'équation maîtresse, nous trouvons

$$\frac{\partial p(x, t)}{\partial t} = -Ns \frac{\partial (x(1-x)p(x, t))}{\partial x} + (1 + \frac{s}{2}) \frac{\partial^2 (x(1-x)p(x, t))}{\partial x^2} + O(1/N)$$

Pour les grandes populations, le terme négligé est effectivement faible devant les autres termes. Nous voyons d'ailleurs que l'équation de diffusion devient exacte si $N \rightarrow \infty$, et que Ns reste fini. Ceci est ce que l'on appelle la limite continue, la "distance" dx entre deux états devenant nulle. Nous allons préciser cela à la prochaine section.

L'équation ci-dessus est connu sous le nom de l'équation de Kimura, qui l'a proposé la première fois dans les années 1950⁴ et constitue une des façon les plus efficace de traiter les problèmes d'évolution.

4. M. Kimura, *Solution of a process of random genetic drift with a continuous model*, Proc. Nat. Ac. Sci. (USA), 41 :144 (1954).

9.2 *Les systèmes continus.*

Est-il possible d'envisager des processus réellement continues? Le concept de continu a donné beaucoup de fil à retordre au mathématiciens. Nous pouvons l'imager comme suit ici : Nous avons à notre disposition un microscope à zoom infini. Un processus est

continu si en regardant la trajectoire de la variable stochastique, quelque soit le facteur zoom choisi, la trajectoire paraît continu.⁵

Précisons un peu ce que l'on veut dire par continu. Prenons le cas d'un observateur du mouvement brownien. Il enregistre, sous microscope, la position de sa bille de polystyrène avec sa caméra. Or, son instrument de mesure lui impose ses échelles spatiales ℓ et temporelle τ : la position de la bille est mesuré au pixel près, sa position temporelle dépend du nombre d'image par seconde que la caméra peut enregistrer. Il décrit son phénomène en écrivant la probabilité de saut entre pixels voisins durant le temps τ par la quantité

$$W(n, m; \ell, \tau)$$

Dans notre traitement du mouvement Brownien, nous avons utilisé les probabilité de saut

$$W(n, \pm 1; \ell, \tau) = \alpha(\ell)\tau$$

où α est la probabilité de saut par unité de temps, qui doit évidemment dépendre de la discrétisation de l'espace ℓ . Supposons maintenant que l'observateur change de grossissement de microscope et les pixels représentent maintenant ℓ' micron (voir figure 9.2). Notre observateur continue de modéliser son système par des sauts entre pixels

$$W'(n, \pm 1; \ell', \tau) = \alpha'(\ell')\tau$$

Le phénomène physique n'a pas changé en changeant d'appareil de mesure, les deux descriptions doivent donc être cohérentes entre elle. Il est évident que en réduisant la taille des cases, on doit augmenter la fréquence des sauts, et nous avons vu précédemment, que pour rester cohérent, nous devons avoir

$$\alpha(\ell) = (\ell'/\ell)^2 \alpha(\ell')$$

ou autrement, la quantité $D = \alpha\ell^2$ doit rester constante. La taille moyenne des sauts pendant le temps τ est

$$\lambda = \sqrt{\alpha\ell^2\tau}$$

et donc quand $\tau \rightarrow 0$, la taille moyenne des saut tend également vers zéro, mais seulement comme $\sqrt{\tau}$.

Nous devons maintenant généraliser cette approche à un processus stochastique quelconque en faisant tendre τ et ℓ vers zéro (ce qui multiplie à l'échelle microscopique les "petits" saut) et donner une expression cohérente de $w(x, y, \tau)$ la probabilité d'effectuer, pendant le temps τ , à partir de la position x un saut de taille $y \pm dy$. Ce travail a été fait par N. Wiener⁶ dans les années 1920-30 qui a défini précisément les conditions sur w pour que les trajectoires stochastiques soient continues. Voyons cela de plus près

Soit donc la fonction

$$w(x, y, \tau)$$

5. Pour la plupart des physiciens, la question ne se pose pas : leurs observations sont limités par la précision des mesures et ils ne voient jamais autre chose que des sauts entre états. En dehors de la limitation des mesures, il existe un débat sur la possibilité d'existence physique des processus de Markov continus. Dans les processus physiques, il existe toujours une corrélation entre deux états qui se succèdent dans le temps, et ce temps de corrélation introduit l'échelle de temps naturelle de discrétisation. Un pur processus de Markov suppose que ce temps est nul. Le lecteur intéressé se reportera à la littérature avancée sur les chaînes de Markov.

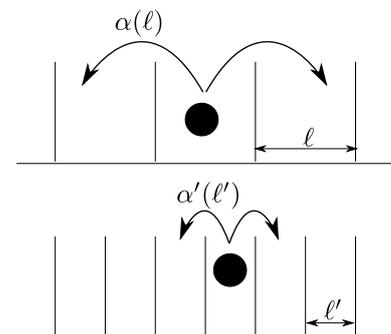


FIGURE 9.2 – Dépendance des probabilités de saut de l'échelle spatiale des discrétisation.

6. Wiener est également connu comme le père de la cybernétique. Personnellement, je ne connais pas la signification de ce mot qui me paraît un concept attrape tout englobant les ordinateurs, les robots, le traitement automatisé des informations, ... Le mot a eu beaucoup de résonance auprès des auteurs de science-fiction, qui lui ont donné un statut de mythe.

la densité de probabilité du passage de la position x à la position $x + y$ pendant le temps τ . Cette fonction est parfois désignée par le "propagateur". Pour la mesurer expérimentalement, nous pouvons collecter N trajectoires passant par x au temps t et aboutissant entre $[y, y + dy[$ au temps $t + \tau$. w étant une densité de probabilité, nous devons avoir

$$\int_{\mathbb{R}} w(x, y, \tau) dy = 1$$

ce qui veut dire que nous sommes sûr de finir *quelque part* au bout d'un temps τ .

La densité de probabilité "d'être" en x au temps $t + \tau$ est évidemment

$$p(x, t + \tau) = \int_{\mathbb{R}} w(x - y, y, \tau) p(x - y, t) dy \quad (9.5)$$

Nous exigeons maintenant trois conditions de $w(x, y, \tau)$ liant la taille des sauts y et le temps de saut τ :

Condition 1. Pour une distance $\epsilon > 0$ donnée,

$$\lim_{\tau \rightarrow 0} \frac{1}{\tau} \int_{|y| > \epsilon} w(x, y, \tau) dy = 0$$

Cela veut dire que la probabilité de "diffuser" en dehors d'une zone de taille donnée tend vers zéro quand le temps de saut tend vers zéro, et cela *plus rapidement* que τ . Prenons le cas de notre bille de polystyrène de taille micro-métrique que nous observons sous microscope. Après une seconde, elle a une chance non négligeable d'être à plus de 1 micron de sa position originale. Après seulement 10^{-3} seconde, cette chance est très amoindri, et après seulement 10^{-9} seconde, cette chance est nulle. La condition 1 établit une relation entre la distance ϵ et le temps de vol τ . Pour le mouvement brownien continu, le propagateur est donné par

$$w(x, y, \tau) = \frac{1}{\sqrt{2\pi D\tau}} \exp\left(-\frac{(y - c\tau)^2}{2D\tau}\right) \quad (9.6)$$

et remplit parfaitement la condition 1.

Condition 2. pour les temps courts, la distance moyenne de diffusion λ depuis la position x est de l'ordre du temps de vol $\lambda \sim \tau$:

$$\lim_{\tau \rightarrow 0} \frac{1}{\tau} \int_{\mathbb{R}} y w(x, y, \tau) dy = a(x)$$

La distance moyenne peu dépendre de la position x ; si le processus est symétrique, elle peut être nulle. Mais l'important dans la condition ci-dessus est que la distance moyenne est au plus de l'ordre du τ .

Condition 3. pour les temps courts, la variance de la distance parcourue est de l'ordre du temps de vol $\lambda \sim \tau$:

$$\lim_{\tau \rightarrow 0} \frac{1}{\tau} \int_{\mathbb{R}} y^2 w(x, y, \tau) dy = b(x)$$

Remarquons que ces trois conditions ont une conséquence profonde.

Conséquence. Tous les moments supérieurs à 2 du propagateur décroissent plus rapidement que le temps de vol τ :

$$\lim_{\tau \rightarrow 0} \frac{1}{\tau} \int_{\mathbb{R}} y^n w(x, y, \tau) dy = 0 \quad \text{si } n > 2$$

Nous laissons au lecteur le soin de démontrer cette conséquence⁷. Mais cela nous permet de développer seulement à l'ordre 2 la relation de consistance (9.5), et en échangeant l'ordre d'intégration sur y et dérivation sur x , nous obtenons trivialement la relation de Fokker-Planck qui est cette fois une relation exacte :

$$\frac{\partial p(x, t)}{\partial t} = - \frac{\partial (a(x)p(x, t))}{\partial x} + \frac{1}{2} \frac{\partial^2 (b(x)p(x, t))}{\partial x^2}$$

§ 9.1 Vérifier que le processus de Wiener (9.6) remplit bien les conditions 2 et 3 et la conséquence 1.

§ 9.2 En utilisant la fonction porte $\Pi(x)$

$$\begin{aligned} \Pi(x) &= 0 \quad \text{si } |x| > 1/2 \\ &= 1 \quad \text{sinon} \end{aligned}$$

construire une fonction $w(x, y, \tau)$ compatible avec les conditions de Wiener.

9.3 Exemples d'application de l'équation de Diffusion.

9.3.1 Mouvement brownien dans un champ de force.

Le mouvement brownien que nous avons déjà longtemps traité correspond à $a(x) = 0$ et $b(x) = D$. Nous savons également que la solution de l'équation de diffusion, avec la condition initiale $p(x, 0) = \delta(x)$ est une gaussienne

$$p(x, t) = \frac{1}{\sqrt{2\pi Dt}} e^{-x^2/2Dt}$$

Notons trivialement que la position moyenne de la particule est une constante :

$$\frac{d \langle x(t) \rangle}{dt} = \langle a(x) \rangle = 0$$

C'est à dire que les forces aléatoires des chocs thermiques sont symétriques et ne peuvent pas provoquer un déplacement net de la moyenne.

Une particule Brownien peut également être soumise à un champ de force $F(x)$: dans ce cas, son mouvement est dû d'une part à l'action des forces aléatoires décrit par le coefficient de diffusion D , et d'autre part à l'action des forces déterministes $F(x)$. Pour une particule dans un milieu dense, la vitesse est proportionnelle à la force appliquée⁸. Comme seule la force $F(x)$ peut provoquer un déplacement net,

$$\frac{d \langle x(t) \rangle}{dt} = \langle F(x) \rangle$$

et donc $a(x) = F(x)$. L'équation de Fokker-Planck s'écrit donc

$$\frac{\partial p(x, t)}{\partial t} = - \frac{\partial (Fp)}{\partial x} + \frac{D}{2} \frac{\partial^2 p(x, t)}{\partial x^2}$$

7. C'est assez facile : se donner un ϵ assez petit, pour les $|y| < \epsilon$, utiliser la majoration $|y^3 w(x, y, \tau)| < \epsilon y^2 w(x, y, \tau)$. Pour $|y| > \epsilon$, utiliser la condition 1.

8. Cela nous ramène à un monde Aristotélicien. En effet, d'un point de vue microscopique,

$$m\ddot{x} + \rho\dot{x} = F(x)$$

mais si la masse est négligeable devant le coefficient de frottement ρ , nous avons la proportionnalité entre la force et la vitesse. Dit autrement, le temps de relaxation m/ρ est très faible devant les autres temps caractéristiques du système. Les particules Browniennes que nous étudions sont toujours dans cette limite.

Exemple 9.1 Force constante.

Prenons une particule Brownienne chutant dans le champs de Gravitation où $F(x) = c = Cte$. L'équation FP s'écrit simplement

$$\frac{\partial p(x, t)}{\partial t} = -F \frac{\partial p}{\partial x} + \frac{D}{2} \frac{\partial^2 p(x, t)}{\partial x^2}$$

Si l'on pose $p(x, t) = u(x - ct, t)$, nous aboutissons à une équation de diffusion simple pour la fonction u ; p s'écrit donc simplement

$$p(x, t) = \frac{1}{\sqrt{2\pi Dt}} e^{-(x-ct)^2/2Dt}$$

c'est à dire une gaussienne dont le centre de gravité se déplace à la vitesse c .

Exemple 9.2 Potentielle harmonique.

Soit $F(x) = -kx$, c'est à dire que la particule se trouve dans un potentiel harmonique $U(x) = (k/2)x^2$. L'équation FP est

$$\frac{\partial p(x, t)}{\partial t} = k \frac{\partial (xp)}{\partial x} + \frac{D}{2} \frac{\partial^2 p(x, t)}{\partial x^2}$$

La solution stationnaire $p(x)$ obéit donc à l'équation

$$p'(x) + \frac{2k}{D} xp(x) = A$$

où A est une constante d'intégration. Comme $p(x)$ doit être une fonction paire, $A = 0$ et (figure 9.3)

$$p(x) = Ce^{-(k/D)x^2}$$

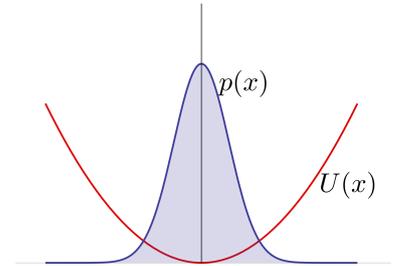


FIGURE 9.3 – La solution stationnaire pour $U(x) = (1/2)kx^2$

§ 9.3 Démontrer que la fonction caractéristique (c'est à dire la transformée de Fourier)

$$\phi(s, t) = \int_I e^{isx} p(x, t) dx$$

obéit à l'équation

$$\partial_t \phi + ks \partial_s \phi = -(D/2)s^2 \phi$$

Ceci est une équation à dérivée partielle de premier ordre que nous savons résoudre exactement. Vérifier que sa solution, avec la condition initiale $p(x, 0) = \delta(x_0)$ est donnée par

$$\phi(s, t) = \exp \left[-\frac{D(1 - e^{-2kt})}{4k} s^2 + ix_0 e^{-kt} s \right]$$

la fonction ϕ est une gaussienne; en prenant la TF inverse, démontrer que $p(x, t)$ est une gaussienne

$$p(x, t) = \frac{1}{\sqrt{2\pi Dt}} e^{-(x-\mu(t))^2/2\sigma^2(t)}$$

où

$$\mu(t) = x_0 e^{-kt}; \sigma^2(t) = \frac{D}{2k} (1 - e^{-2kt})$$

Exemple 9.3 Processus poissonien.

Traiter le poissonien de base [modèle de production d'ARN] quand $\lambda \gg 1$; traiter le bi-poissonien (quand un poissonien pilote un autre dans le même régime et obtenir la distribution gamma).

9.4 L'équation amont (backward) et le temps de premier passage.

[Bien dégager la notion de trajectoire et le poids des trajectoires; peut-être donner un aperçu de la dérivation de l'équation de Schrödinger à partir des intégrales de chemin comme un exemple avancé.]

9.5 Problèmes.

Problème 9.1 Processus de Wiener.

Soit la densité de probabilité de saut

$$f(x, \lambda) = \frac{1}{\lambda} \Pi(x/\lambda)$$

où la fonction porte $\Pi(x) = 1/2$ si $|x| < 1$ et 0 sinon. Démontrer que

$$w(y, \tau) = f(y, \sqrt{D'\tau})$$

est un processus de Wiener et dériver son équation FP. Quelle est la relation entre D' ici et le coefficient D du mouvement Brownien? Soit de façon générale la fonction à support borné $f(x) = 0$ si $|x| > 1$. Construisez un processus de Wiener à partir de cette fonction. Même question pour la fonction $f(x) = \exp(-|x|)$. Même question pour la fonction $f(x, \lambda) = \delta(x - \lambda) + \delta(x + \lambda)$

Problème 9.2 Réaction chimique à double puits de potentiels.

[État Stationnaire, faire le lien avec l'exercice en discret].

Problème 9.3 Le Self focusing par le bruit..

[Faire référence au papier de Johan Paulson].

Problème 9.4 Conformation des polymères.

Problème 9.5 L'équation de Schrodinger et les intégrales de chemins.

Problème 9.6 Thermal ratchets.

Les équations différentielles stochastiques et le calcul d'Ito.

En 1908, Langevin a publié un article¹ qui a donné naissance aux équations différentielles stochastiques. Il a fallu 40 ans avant qu'Ito donne une certaine rigueur mathématiques à cette discipline qui reste toujours assez fragile. Ceci étant, il est très naturel de formuler des problèmes dans ce langage. Par exemple, une particule brownien chargé dans un champs électrique peu se concevoir comme une superposition d'une interaction déterministe et des fluctuations aléatoires. Une fois le problème formulé, on peut trouver son équivalent Fokker-Planck et retomber sur nos pattes. Même si la théorie rigoureuse reste un peu sale, les concepts de bases sont très élégants et facile à entrevoir.

1. C. R. Acad. Sci. (Paris) 146 : 530-533 (1908).

10.1 Les équations différentielles stochastiques.

Commençons par ce que nous comprenons des équations différentielles ordinaires. Supposons que nous souhaitons résoudre l'équation

$$\frac{dx}{dt} = a(x, t) \quad (10.1)$$

et essayons de la résoudre numériquement. Donnons nous un pas de temps dt et considérons les temps discrets $t_i = t_0 + idt$; nous cherchons à trouver les quantités $x_i = x(t_i)$. La relation ci-dessus s'écrit, dans le cadre de notre approximation

$$x_{i+1} = x_i + a(x_i, t_i) dt \quad (10.2)$$

Nous voyons immédiatement que cela nous procure un algorithme : soit x_0 la valeur de x au temps t_0 . Nous calculons la quantité $a(x_0, t_0)$ et la multiplions par dt . Nous incrémentons x_0 par cette valeur, et nous appelons la nouvelle valeur x_1 . Il suffit ensuite de recommencer et construire les x_i de proche en proche. Cet algorithme porte le nom d'Euler qui l'a utilisé dans ces termes pour la première fois.

L'erreur que nous commettons dans notre évaluation est proportionnelle à notre échantillonnage dt . Si nous souhaitons une

meilleure précision, nous pouvons utiliser par exemple l'échantillonnage $dt' = dt/2$; dans ce cas, l'incrément à chaque étape est divisé (en gros) par deux.

La relation de récurrence (10.2) peut se mettre, en remplaçant de proche en proche, sous la forme

$$x_{n+1} = x_0 + \sum_{i=0}^n a(x_i, t_i) dt$$

qui, quand nous faisons tendre dt vers zéro, s'écrit par le signe intégrale :

$$x(t) = x_0 + \int_0^t a(x(s), s) ds \quad (10.3)$$

Nous aurions également pu utiliser pour l'équation différentielle (10.1) la notation

$$dx = a(x, t) dt \quad (10.4)$$

auquel cas la solution intégrale (10.3) aurait paru encore plus naturel.

Comment résoudre une équation différentielle stochastique ?

Supposons maintenant que nous voulons simuler numériquement un mouvement brownien continu. A nouveau, nous considérons un intervalle de temps τ ; à chaque pas de temps, nous "tirons" un saut au hasard (la fonction `rand()` de notre ordinateur par exemple) $\ell = w(x_n, \tau)$ et nous incrémentons la valeur de x_n :

$$x_{n+1} = x_n + \ell$$

Si nous souhaitons réduire les erreurs, il suffit de prendre des intervalles de temps τ de plus en plus petit. Justement! nous touchons là au cœur des processus stochastiques : si nous divisons l'intervalle de temps par 2, la taille des sauts ℓ n'est pas réduit d'un facteur 2, mais d'un facteur $\sqrt{2}$: nous avons vu que dans les processus de Wiener, $\ell \sim \sqrt{\tau}$. Cela dit, si nous respectons cette dépendance, la simulation numérique ne pose aucun problème. Par analogie avec l'équation déterministe (10.4), nous pouvons écrire, quand $\tau \rightarrow 0$

$$dx = b(x, t) dW(t)$$

où $dW(t)$ est un processus aléatoire qui génère des sauts de l'ordre de \sqrt{dt} , et $b(x, t)$ modifie cette amplitude localement au temps t et à la position x .

En combinant la partie déterministe et stochastique, nous pouvons écrire l'équation différentielle générale par

$$dx = a(x, t) dt + b(x, t) dW(t) \quad (10.5)$$

dont la solution est donné par

$$x(s) = x_0 + \int_0^s a(x(s), s) ds + \int_0^s b(x, s) dW(s)$$

Il faut maintenant donner un sens précis à cette intégrale un peu bizarre qui contient un élément de l'ordre de \sqrt{dt} . Pour cela, nous devons définir des règles précises de manipulation du symbole $dW(t)$ ce qui nous permettra de résoudre les équations différentielles stochastiques un peu comme des équations différentielles ordinaires. La chose la plus importante à retenir est que nous allons considérer des temps dt infinitésimal; le détail du processus de Wiener, la forme exacte de la densité de probabilité n'a pas alors d'importance. Nous allons donc choisir un processus de Wiener en particulier qui convient bien à donner un sens exact à nos calculs : à chaque "tirage", le processus de Wiener produit

$$dW(t) = \pm\sqrt{dt}$$

où seulement le signe \pm est aléatoire. C'est à dire (voir exercice 9.5.1)

$$w(x, \tau) = (\delta(x - \sqrt{\tau}) + \delta(x + \sqrt{\tau})/2$$

10.2 Manipulation de $dW(t)$.

Les deux mots clefs de $dW(t)$ sont : (i) de l'ordre de \sqrt{dt} ; (ii) aléatoire. Si nous avons un terme déterministe de l'ordre de \sqrt{dt} , le terme en dt aurait été négligeable. Le fait qu'il soit aléatoire provoque des compensations et ramène le résultat final de l'ordre de dt . Nous exigeons donc, pour que les compensations puissent avoir lieu, que les sauts aléatoires soit symétrique :

$$\langle dW(t) \rangle = 0$$

Visuellement, cela veut dire que nous laissons notre particule brownien évoluer jusqu'au temps t et à la position x et nous tirons un saut aléatoire $\ell_1 = dW(t)$. Si nous refaisons cette expériences des milliards de fois et que nous faisons la moyenne de ces sauts microscopiques, nous avons zéro comme résultat. Il n'est pas difficile de voir que le résultat est, encore plus généralement,

$$\langle b(x, t)dW(t) \rangle = 0$$

Par contre, la moyenne des carrés est non-nulle :

$$[dW(t)]^2 = dt$$

Nous négligeons toutes les puissances supérieures, qui sont négligeable devant dt :

$$[dW(t)]^n = 0$$

Le raisonnement est tout à fait analogue à notre discussion des processus de Wiener et les conditions (??,??).

La fonction stochastique $W(t)$ est la somme des sauts aléatoires microscopiques :

$$W(t) = \int_0^t dW(s)$$

et représente un trajectoire Brownien continu. Nous pouvons définir $f[W(t)]$ où f est une fonction quelconque, comme par exemple $\exp(W(t))$. Ce procédé nous permet de définir par extension des trajectoires aléatoires non Brownien. On peut même définir le différentielle stochastique de cette quantité :

$$\begin{aligned} d \exp(W(t)) &= \exp(W(t) + dW(t)) - \exp(W(t)) \\ &= \exp(W(t)) \exp(dW(t)) - \exp(W(t)) \\ &= \exp(W(t)) \left(dW(t) + \frac{1}{2} dt \right) \end{aligned}$$

10.3 Stratonovitch vs Ito.

10.4 Problèmes.

Problème 10.1 Écologie neutre.

Résoudre

$$dx = \sqrt{x} dW(t)$$

et comparer la solution à mort-naissance traité en discret.

Problème 10.2 Dérive génétique.

Résoudre

$$dx = \sqrt{x(1-x)} dW(t)$$

et comparer à la dérive génétique traité en discret. Compléter en introduisant un terme déterministe de sélection en $sx(1-x)dt$.

Les filtres de Kalman.

Supposons que nous mesurons la position d'une particule par un instrument d'une certaine précision σ . Par exemple, le GPS habituel a une erreur de l'ordre de 10m. Est-il possible d'améliorer la précision de nos mesures en incorporant des connaissances que nous avons sur la particule ? Par exemple, connaissant la position et la vitesse de la particule à l'instant t , peut-on corriger la mesure de sa position à l'instant $t + \Delta t$? Soulignons que la position et la vitesse de la particule au temps t sont les résultats d'une mesure par le même instrument de résolution σ .

La réponse à la question est oui. Cette réponse a émergé à la fin des années 1950 suite aux travaux de nombreux mathématiciens dont Kalman. La méthode joue un rôle crucial dans tous les instruments de guidage.¹

11.1 Prélude.

Considérons d'abord un cas très simple. Une particule dans un monde à une dimension commence son mouvement à $t = 0$ à la position $x = 0$. Après un temps t_1 , la particule se trouve à la position x_1 . Supposons que la particule est soumise, entre autre, à des forces aléatoires et que sa position x_1 est une variable aléatoire de moyenne μ et de largeur σ . Par exemple, si la particule est purement Brownien et diffuse avec un coefficient de diffusion D , nous savons que $\mu = 0$ et $\sigma = 2Dt_1$.

Au temps t_1 , nous mesurons la position de la particule avec un instrument de mesure de précision s , qui nous affiche un résultat y_1 . Connaissant la nature de la particule (en l'occurrence μ et σ), peut-on avoir une meilleure estimation de sa position mesurée que y_1 ?

Pour mettre les choses en langage plus précis, soit la variable aléatoire X (position réelle de la particule) de moyenne μ et variance σ^2 . Soit la variable aléatoire Y (position mesurée de la particule)

$$Y = X + E$$

où E est notre instrument de mesure, une variable aléatoire de moyenne 0 et de variance s^2 . Si nous utilisons Y comme un esti-

1. La rédaction de ce chapitre a été motivée par un excellent séminaire donné à notre groupe en mars 2017 par E. Cinquemani sur le filtre de Kalman, qu'il utilisait entre autre pour une mesure plus précise de l'activité des réseaux génétiques.

mateur de X , l'erreur que l'on commet est $Y - X = E$ qui est la position de notre instrument.

Supposons pour plus de simplicité que $\mu = 0$ pour l'instant. Prenons un autre estimateur de X , la variable aléatoire

$$Z = aY$$

où a est une constante à déterminer. Est-il possible que l'erreur de Z soit moindre que l'erreur de Y ? En d'autres termes, que la variance de $Z - X$ soit inférieure à la variance de $Y - X$?

Le calcul est simple à faire :

$$\begin{aligned} V_K = \text{Var}(Z - X) &= \text{Var}((a - 1)X + aE) \\ &= (a - 1)^2\sigma^2 + a^2s^2 \end{aligned} \quad (11.1)$$

Nous voyons que pour que V_K soit minimum, il faut choisir

$$a = \frac{\sigma^2}{\sigma^2 + s^2}$$

et pour cette valeur de a ,

$$V_K = \frac{\sigma^2}{\sigma^2 + s^2}s^2$$

Et le rapport des variances entre l'estimateur trivial Y et l'estimateur de Kalman Z est

$$\frac{V_K}{s^2} = \frac{\sigma^2}{\sigma^2 + s^2} < 1$$

Est l'estimation améliorée d'autant plus que le mouvement de la particule est prédictible.

En clair, voilà ce que veut dire l'estimateur Z : Nous mesurons la position d'une particule au point y ; mais sachant que la particule vient de l'origine, nous estimons que sa position réelle est plus proche de l'origine que y : nous commettons moins d'erreur en estimant qu'elle se trouve en réalité en $z = ay$. Pour se convaincre de la justesse de la chose, il suffit de faire une petite expérience numérique, comme dans l'exemple ...

Bien sûr, nous n'avons pas à nous restreindre à $\mu = 0$. Si $\mu \neq 0$, l'estimateur de la position devient

$$Z = a(Y - \mu) + \mu$$

où $Y = X + E$ est la mesure de la position par l'instrument comme précédemment. Comme $\langle Y \rangle = \langle X \rangle = \mu$, nous avons $\langle Z \rangle = \mu$. Par ailleurs, $\text{Var}(Z - X)$ est donnée comme précédemment par l'expression (11.1).

Nous n'avons pas trop précisé d'où venait le bruit de position de la particule : cela peut être du fait que la particule est soumise en partie aux forces aléatoires ou au fait que nous avons peu de précision sur la position originale de la particule. Cela n'a pas d'importance du moment que l'on capture ce bruit dans le coefficient σ .

Algorithme 1 Un script en langage Julia pour calculer la variance de $Z - X$

```

M = 10000          #Nombre d'échantillon
sigma = 1
s = 0.05
dx=Normal(0, sigma)
de=Normal(0, s)
x=rand(dx,M)      # génère M réalisation de X
e=rand(de,M)      # génère M réalisation de E
a = sigma^2/(sigma^2+s^2)
r=a*(x+e)-x       # Estime l'écart entre Z et X
sk=std(r)         # Calcul la deviation standard de r

```

11.2 Observation d'une Chaîne de Markov.

Considérons maintenant une suite de position (x_0, x_1, x_2, \dots) que notre instrument mesure par (y_0, y_1, y_2, \dots) où chaque mesure y_i comporte une erreur e_i . Comme précédemment, nous supposons que la variable E est de moyenne nulle et de variance s^2 .

Nous supposons que nous connaissons quelque peu la dynamique interne de la variable position :

$$x_{t+1} = f_t x_t + \eta_t$$

où η_t est la partie aléatoire du mouvement ($\langle \eta_t \rangle = 0$, $\langle \eta_t^2 \rangle = r^2$) et f_t capture la partie déterministe de celui-ci.

Supposons maintenant que X_t est une variable aléatoire de moyenne μ_t et de variance σ_t^2 . Dans ce cas, pour la V.A. X_{t+1} nous avons

$$\begin{aligned} \langle X_{t+1} \rangle &= f_t \mu_t \\ \text{Var}(X_{t+1}) &= (f_t \sigma_t)^2 + r^2 \end{aligned}$$

Nous pouvons maintenant appliquer nos calculs de la précédente section : au lieu d'utiliser Y_{t+1} pour estimer la position X_{t+1} , nous utilisons

$$Z_{t+1} = a_{t+1} (Y_{t+1} - \mu_{t+1}) + \mu_{t+1} \quad (11.2)$$

où

$$a_{t+1} = \frac{(f_t \sigma_t)^2 + r^2}{(f_t \sigma_t)^2 + r^2 + s^2} \quad (11.3)$$

et comme précédemment, la variance de l'erreur $\Delta_{t+1} = Z_{t+1} - X_{t+1}$ sur l'estimation de la position est

$$V_{t+1} = \frac{(f_t \sigma_t)^2 + r^2}{(f_t \sigma_t)^2 + r^2 + s^2} s^2 < s^2$$

Bien. Nous avons comme avant un estimateur meilleur que notre observation brute. Sauf que ... Nous ne connaissons ni μ_t ni σ_t . La seule chose qui nous est donnée sont la loi de la dynamique interne de la particule, c'est à dire f_t et η_t .

Le filtrage de Kalman consiste à faire le maximum que l'on peut avec ce que nous avons, de façon itérative, *en deux étapes* : une phase de prédiction, une phase de correction.

Nous commençons par notre première observation y_0 et posons $\mu_0 = y_0$, $\sigma_0 = s$. Ceci est le mieux que l'on puisse faire au temps initial, où l'on ne dispose d'aucune autre information.

Notre prédiction de la moyenne pour $t = 1$ est $\mu_1^- = f_0\mu_0$. Nous pouvons maintenant confronter cette prédiction à l'observation y_1 : Nous calculons a_1 par la relation (11.1) et

$$z_1 = a_1(y_1 - \mu_1^-) + \mu_1^-$$

Pour la phase de correction, nous utilisons z_1 comme la vraie position de x_1 et utilisons ces valeurs pour avancer. En particulier, nous supposons que

$$\begin{aligned}\mu_1 &= z_1 \\ \sigma_1^2 &= a_1 s^2\end{aligned}$$

Nous pouvons maintenant utiliser ces valeurs μ_1, σ_1 pour prédire la prochaine μ_2^- , a_2 , z_2 . Cette prédiction est utilisée pour estimer μ_2 et σ_2 et ainsi de suite. Ce que nous pouvons supposer est que la chaîne (z_0, z_1, z_2, \dots) est une meilleure estimation des positions que les observations (y_0, y_1, y_2, \dots) .

Insistons sur la différence avec le préluce. Dans le préluce, nous connaissons μ et σ pour construire z . Ici, dans le filtre de Kalman, μ et σ sont eux même des estimations à chaque étape. Nous n'avons donc pas garanti a priori que le filtre de Kalman soit une meilleure estimation. Cependant, pour un large classe de problème, ceci est effectivement le cas.

Résumons. L'algorithme marche comme suit :

1. $\mu_{t+1}^- = f_t \mu_t$
2. $a_{t+1} = ((f_t \sigma_t)^2 + r^2) / ((f_t \sigma_t)^2 + r^2 + s^2)$
3. $z_{t+1} = a_{t+1}(y_{t+1} - \mu_{t+1}^-) + \mu_{t+1}^-$
4. $\mu_{t+1} = z_{t+1}$
5. $\sigma_{t+1}^2 = a_{t+1} s^2$

Les étapes 1-3 sont la phase de prédiction, les étapes 4-5 la phase de correction. Cet algorithme se programme en quelques lignes.

A

*Le traitement du mouvement Brownien par Einstein,
Langevin et Smoluchowski.*