

Chapitre 8

Éléments de Théorie des probabilités

*Ce chapitre est un exposé élémentaire
de la Théorie des probabilités.*

La théorie des probabilités constitue un pan entier des mathématiques et fait appel à des notions très abstraites liées à la difficulté du sujet¹. Celle-ci peut être mesurée par le fait que, si la notion de probabilité est finalement assez ancienne, il a fallu attendre les années 1930 pour qu'en soit faite une construction axiomatique² cohérente (Kolmogorov, 1933). On touchera du doigt plus loin les subtilités qui rendent les exposés élémentaires insatisfaisants quand le but est l'édification d'une théorie dénuée d'ambiguïtés et logiquement irréfutable.

On s'en tiendra ici à un exposé très élémentaire, essentiellement guidé par le pragmatisme et par le souci de garder le contact avec les préoccupations usuelles du physicien. Ceci n'interdira pas, de loin en loin, de mentionner des situations un peu exotiques que l'on rencontre parfois en Physique³.

8.1 Notion de variable aléatoire. Notion de probabilité

Il semble indiscutable que la notion de variable aléatoire est inévitablement liée à la notion d'*expérience*, ou de *mesure*⁴. On entend par là qu'un dispositif expérimental ayant été mis en place, le but de l'expérience est d'obtenir les valeurs d'une certaine grandeur d'intérêt. Il est toujours possible d'établir une correspondance (bijective) entre les résultats d'un type d'expérience et des nombres ; par exemple, s'agissant de distinguer quantitativement des boules de couleur différente, des blanches, des rouges et des noires par exemple, on peut décider d'affecter la valeur -1 à la couleur blanc, la valeur 0 à la couleur rouge, la valeur 1 à la couleur noir. De même, pour la pièce de monnaie, on pourra dire que face correspond à 1 , et pile à 0 . Pour le dé cubique (six faces), on peut décider d'attribuer à chaque face le nombre correspondant à sa valeur. Cette association est purement conventionnelle et n'est définie que dans un but opératoire mais, évidemment, une fois choisie il faut s'y tenir. En tout cas, on pourra donc toujours supposer que l'expérience (la mesure) débouche sur une valeur *numérique*.

Le caractère *aléatoire* est révélé par le fait que la répétition de l'expérience, sensée se reproduire à l'identique d'une fois à l'autre, ne produit pas toujours le même résultat : N jets de la pièce produiront au

¹Selon Daniel Dugué "Le calcul des probabilités est certainement l'une des branches les plus récentes des mathématiques, bien qu'il ait en fait trois siècles et demi d'existence" (*Encyclopædia Universalis*)

²D'après Georges Glæser "Ainsi, lorsque A. Kolmogorov a formalisé, en 1933, les fondements du calcul des probabilités, il a permis à cette science de se dégager des notions vagues et subjectives de hasard et de chance ; l'étude des probabilités, restée longtemps stagnante, a pris ainsi un nouvel essor, devenant une partie très vivante des mathématiques.", *ibid.*

³par exemple les mesures singulières continues à propos des systèmes désordonnés.

⁴Le sens de ce terme est précisé dans les lignes qui suivent. Le même terme est employé (pas par hasard !) par les mathématiciens, pour désigner une notion abstraite dûment définie qui, dans les cas élémentaires, se réduit à celle utilisée par l'arpenteur (longueur d'un chemin, surface d'une parcelle, . . .)

total n fois pile et $N - n$ fois face, sans que l'on puisse prédire, lorsque la pièce est en l'air, de quel côté elle va retomber. L'expérience est ici le jet de la pièce, le résultat de chaque expérience est soit pile, soit face ; avec la correspondance définie plus haut, les valeurs numériques obtenues sont donc soit 1 soit 0 respectivement ; avec ce cadre de travail, la variable aléatoire est le résultat d'un lancer de la pièce, et peut prendre les deux valeurs 0 et 1. Une variable aléatoire pouvant prendre un nombre fini de valeurs distinctes, ou un nombre infini dénombrable, est dite *discrète*. Par la suite, on considérera aussi des variables aléatoires *continues*, qui prennent leurs valeurs dans \mathbb{R} ou une partie de \mathbb{R} . L'abréviation consacrée pour *variable aléatoire* est *v.a.* Il convient toujours de soigneusement distinguer d'une part la variable aléatoire abstraite, d'autre part les valeurs obtenues lors d'une expérience – tout comme on doit distinguer une fonction $x \rightarrow f(x)$ des valeurs numériques $f(x_n)$. Le plus souvent, on notera X la v. a. et x_n ses valeurs numériques possibles.

Dans le cas du jet de la pièce, la non-reproductibilité du résultat de chaque mesure tient au fait que l'expérimentateur, de fait, ne contrôle pas tous les paramètres déterminant le mouvement de la pièce. En toute rigueur, si tous ces paramètres étaient contrôlés et connus, un protocole complètement défini devrait donner toujours le même résultat, étant entendu que la pièce est un objet macroscopique obéissant aux équations déterministes de la Mécanique classique⁵. En fait, au-delà des apparences sensibles, de nombreuses petites causes influent le mouvement de la pièce en dehors de tout contrôle, et ce sont elles qui font finalement varier le résultat de chaque lancer. L'ensemble de ces causes est ce que l'on appelle communément le *hasard* et traduit, en quelque sorte, l'ignorance inévitable d'effets certes petits mais essentiels puisqu'ils produisent une dispersion des résultats. Le hasard est ainsi l'ensemble des causes responsables de la variabilité des résultats de mesures réputées identiques, au sens où le protocole expérimental est complètement défini.

Le but de toute théorie probabiliste est non pas de tenter de décrire (de façon déterministe) les causes d'une telle variabilité (par nature le hasard échappe à tout contrôle), mais de fournir un cadre alternatif permettant d'affecter à l'issue de chaque mesure un nombre (sa probabilité) permettant quantifier les résultats d'un *très grand nombre* d'expériences. En quelque sorte, l'approche probabiliste part d'un constat (l'impossibilité de décrire dans le détail les raisons premières qui produisent un résultat ou un autre) et, par pragmatisme, aborde le problème posé suivant une autre méthode⁶. En définitive, confronté à la pertinence incontournable du hasard pour un phénomène donné, on recourt à une méthodologie radicalement différente, dont les ambitions quantitatives restent à un niveau d'exigence élevé et dont les vertus prédictives sont d'une extrême richesse.

Dans le cas des boules de couleurs, le hasard tient au fait que l'on définit un jeu, consistant à tirer *en aveugle* les boules dans un sac. En quelque sorte, le hasard tient à la règle du jeu elle-même, car il n'y a rien d'aléatoire en soi dans la couleur d'une boule donnée : dans le sac, chaque boule a sa couleur, qui ne change pas au cours du temps. Il convient de dire dès à présent que la théorie des jeux a joué un rôle moteur dans le développement historique de la Théorie des probabilités.

En Physique, tout résultat de mesure est une variable aléatoire, puisque toute mesure est entachée d'imprécision : quel que soit le soin pris dans la réalisation d'une expérience et sa conduite à terme, la valeur obtenue n'est pas infiniment précise ; d'ailleurs, la *barre d'erreur* est un ingrédient essentiel pour légitimer la valeur finalement retenue pour la grandeur physique d'intérêt. Ceci est vrai même si, selon la théorie en vigueur, la grandeur que l'on cherche à mesurer n'a en principe qu'une seule valeur possible⁷, par exemple l'énergie d'un système isolé.

⁵Dans le domaine quantique, il en va tout autrement : on doit considérer que les objets quantiques ont une dynamique *foncièrement* aléatoire. D'un autre côté, contrairement à une affirmation fréquente, la Mécanique quantique est elle aussi une théorie déterministe : elle prévoit avec certitude (?) ce que sont les probabilités des événements observables. Par rapport à la Mécanique classique, le déterminisme n'a pas disparu, il a seulement changé de registre.

Le domaine quantique est caractérisé par le fait que le *hasard* y est irréductible, et ne tient en aucune façon à l'imprécision d'une expérience ou à une incapacité de formalisation détaillée – voir la note 6.

⁶C'est la même forme de renoncement qui constitue les prémisses de la Mécanique Statistique : aucun physicien doué de raison n'envisage une seconde de décrire les propriétés d'un gaz en se fixant comme objectif de décrire les trajectoires des atomes du gaz. Cet objectif est irréalisable, ses résultats (inaccessibles !) seraient inexploitable, enfin et surtout cette approche est en contradiction flagrante avec l'expérience : pour décrire (parfaitement les propriétés d'un gaz, quelques paramètres en nombre très réduits suffisent (pression, volume, température, pour un système fermé) : rien à voir avec les $\sim 10^{23}$ degrés de liberté mécaniques du gaz !

⁷Il existe évidemment des grandeurs *intrinsèquement* aléatoires : la longueur d'une règle a des fluctuations liées aux fluctuations thermiques, toutes les grandeurs quantiques,...

D'un autre côté, on peut, paradoxalement à première vue, affirmer que le monde quantique est *infiniment déterminé*, dans la mesure où certaines grandeurs physiques sont quantifiées, c'est-à-dire ne peuvent prendre leurs valeurs que dans un ensemble de mesure nulle au sens du mathématicien. Par exemple, les énergies possibles d'un système lié forment un ensemble fini ou isomorphe à \mathbb{N} , $\{E_n\}_n$. Le résultat d'une mesure donnant la valeur E_{n_0} est de fait *infiniment précis* dès lors que l'incertitude liée aux

En ce qui concerne les grandeurs classiques (*i. e.* non quantiques), l'imprécision a essentiellement trois causes :

1. les appareils utilisés ont une résolution finie Δ , qui permet seulement de définir une grille de points isolés les uns des autres, séparés de Δ . Pour toute valeur mesurée x_0 , on peut au plus dire que la valeur de la grandeur est dans l'intervalle $[x_0 - \frac{\Delta}{2}, x_0 + \frac{\Delta}{2}]$
2. L'état de l'appareil dépend de divers paramètres, qui peuvent varier d'une expérience à l'autre (ce que l'on peut désigner – assez vaguement, il faut bien le dire⁸ – par le vocable de *fluctuations*). Il en résulte que la valeur effectivement mesurée peut varier d'une expérience à l'autre – clairement, la grille de mesures elle-même peut fluctuer
3. la grandeur elle-même peut être aléatoire, faute d'une maîtrise absolue des conditions expérimentales.

Ceci étant, deux situations extrêmes peuvent être envisagées :

1. la résolution Δ est très bonne, c'est-à-dire que Δ est très petit devant les variations mesurées résultant des fluctuations de toute nature. Alors, on pourra tracer un histogramme⁹ dont les bâtons ont pour largeur Δ , qui constituera l'histogramme des valeurs mesurées. Sa moyenne et sa largeur seront caractéristiques des variations statistiques des mesures
2. la résolution est médiocre, ou en tout cas Δ est grand vis-à-vis des fluctuations. Alors, toutes les valeurs mesurées coïncideront les unes avec les autres, l'écart entre les deux valeurs extrêmes n'excédant pas Δ

Finalement, affirmer que le résultat de la mesure d'une grandeur physique (classique, voir la note 7) est une variable aléatoire est d'une extrême banalité. Si X note la grandeur mesurée, les valeurs obtenues seront (dans la bonne unité de X), des nombres x_n définis par la grille discrète de pas Δ . On pourra dire que le résultat de chaque mesure est x_n , si l'on a trouvé une valeur dans l'intervalle $[x_n - \frac{\Delta}{2}, x_n + \frac{\Delta}{2}]$. Dans le premier cas ci-dessus (très bonne résolution), la grandeur X ressort comme une variable aléatoire et est distribuée suivant un certain histogramme. Dans le second cas (fluctuations d'appareil très petites et/ou résolution médiocre), la variable X apparaît comme certaine, mais il serait aventureux d'affirmer, au vu de cette série d'expériences, qu'elle est intrinsèquement une variable certaine, c'est-à-dire ne pouvant prendre qu'une et une seule valeur¹⁰.

La règle générale est donc l'aléatoire, la variable certaine étant en fait le cas-limite idéal où toutes les valeurs sauf une sont impossibles ; cette valeur unique donne à la grandeur la qualité de variable non distribuée. Pour une grandeur physique donnée, le statut aléatoire *vs* certaine n'est pas toujours propre à cette grandeur, mais peut dépendre du cadre d'étude¹¹ et/ou du degré de précision des expériences envisagées. Ici encore, le zéro du physicien n'est pas celui du mathématicien : pour le physicien, une variable X prenant une valeur x_1 avec la probabilité $10^{-10^{23}}$ et une valeur différente x_2 avec la probabilité $1 - 10^{-10^{23}}$ sera considérée comme une variable non aléatoire¹² (certaine) . . .

Pour désigner le résultat d'une expérience, le terme conventionnel est *événement*, noté ω . Plus précisément, on se doit de définir d'abord les événements *élémentaires* ("atomiques"), c'est-à-dire dresser la liste de

appareils de mesure est plus petite que l'écart entre E_{n_0} et ses deux valeurs les plus proches. On peut ainsi comprendre pourquoi des caractéristiques atomiques (la longueur d'onde d'une transition, par exemple) peuvent servir de standard métrologique.

⁸L'un des enjeux est précisément de préciser et de quantifier ces variations imprévisibles.

⁹Le mot est pris ici au sens élémentaire ; un langage plus précis (et propre aux probabilistes) appelle *histogramme* la fonction de répartition définie plus loin (section 8.3).

¹⁰Répetons qu'une distinction est ici toutefois nécessaire. Quand on mesure des grandeurs prenant des valeurs $x \in \mathbb{R}$ continues – c'est toujours le cas pour les grandeurs physiques classiques –, la prise en compte de l'imprécision expérimentale, toujours inévitable, permet de dire que la mesure de toute telle grandeur fournit une variable aléatoire. Ceci n'est plus vrai pour les grandeurs physiques *quantifiées* qu'introduit la Mécanique quantique, dont l'un des postulats affirme que toute mesure ne peut fournir qu'une valeur *discrète* x_n , à prendre dans un ensemble fini ou infini $\{X\}$, mais toujours *dénombrable*. En pareil cas, si Δ est plus petit que $\min_{\{X\}} |x_n - x_{n'}|$, mais toujours forcément fini, la mesure peut bel et bien être qualifiée d'*infinitement* précise, puisqu'elle fournit alors la seule et unique valeur disponible dans l'intervalle de résolution de l'appareil.

¹¹Pour un système isolé (microcanonique), l'énergie est une grandeur certaine ; pour le même système placé en situation canonique, son énergie est une grandeur fluctuante.

¹²D'où l'affirmation contenue dans le Second principe de la Thermodynamique.

tous les résultats possibles de l'expérience. Pour la pièce de monnaie, il y a deux événements élémentaires, $\omega_1 =$ pile (P) ou $\omega_2 =$ face (F). Pour les boules à trois couleurs, les événements (élémentaires) sont $\omega_1 =$ blanc, ou $\omega_2 =$ noir, ou $\omega_3 =$ rouge ; pour le dé, il y a 6 événements élémentaires. L'ensemble des événements élémentaires sera noté Ω et est appelé *espace des épreuves*.

Les événements élémentaires sont des “briques” permettant de définir des événements composites (“molécules”) ; par exemple, on peut décider de lancer N fois la pièce (ou de lancer N pièces supposées identiques) : chaque expérience peut être représentée par le mot de N lettres PPFPPPPFFPF...F (on peut aussi envisager de lancer simultanément plusieurs pièces), c'est un événement (composite), fabriqué avec les briques que constituent les événements élémentaires. Les résultats de toutes ces expériences peuvent être complètement spécifiés à l'aide des deux seuls événements élémentaires. Obtenir pile *ou* face est un événement non-élémentaire¹³, ainsi qu'obtenir pile *et* face¹⁴. Pour les boules dans le sac, obtenir deux blanches et une noire en trois essais est un exemple d'événement non-élémentaire, qui d'ailleurs n'est ni certain ni impossible. Si nécessaire, on désignera par des capitales (A, B, \dots) des événements composites, pour bien les distinguer des événements élémentaires ω_n .

Deux événements sont dits *exclusifs* si l'obtention de l'un interdit celle de l'autre. Deux événements élémentaires sont donc forcément exclusifs : tout jet de la pièce fournit à chaque fois une valeur, excluant l'autre *ipso facto* – on peut même caractériser les événements élémentaires en disant que pour chacun d'entre eux, le résultat de toute expérience est soit cet événement, soit son contraire.

À l'inverse, deux événements composites A et B ne sont pas forcément exclusifs ; par exemple, les deux événements :

$$A = \text{tirer deux boules blanches et une boule noire} , \quad B = \text{tirer trois boules de couleurs différentes} \quad (8.1)$$

ne sont pas exclusifs. Dans ce cas on voit aussi que la réalisation de A entraîne la réalisation de B : on dit que A entraîne B , ou que A est inclus dans B – ce que l'on note $A \subset B$.

L'événement consistant en la réalisation de A *ou* de B est noté $A \cup B$; celui consistant en la réalisation de A *et* de B est noté $A \cap B$ (si $A \subset B$, alors $A \cap B \neq \emptyset$). Cette notation est très utile, et parle à l'esprit : on verra (sans grande surprise, compte tenu des habitudes algébriques de l'esprit) que l'union (\cup) est en relation étroite avec l'*addition* des probabilités¹⁵, que l'intersection (\cap) est associée avec la *multiplication* des probabilités¹⁶, et qu'enfin l'inclusion (\subset) correspond à l'inégalité \leq ou $<$. Pour les événements élémentaires, on a évidemment :

$$\bigcup_n \omega_n = \Omega , \quad \omega_n \cap \omega_{n'} = \emptyset \quad (\forall n \neq n') \quad (8.2)$$

En revanche, pour des événements composites, $A \cap B \neq \emptyset$, sauf si A et B sont incompatibles. Dans la suite, on utilisera l'une ou l'autre des notations, ou les deux (voir par exemple (8.18)).

L'événement *certain* est un événement qui se produit à chaque expérience ; obtenir pile *ou* face est l'événement certain du jeu de pile et face. Pour la complétude, on définit aussi la notion d'événement impossible : c'est un événement qui ne peut pas se produire (tirer une boule de couleur indigo). Il y a autant d'événements impossibles que l'on veut !

Il s'agit maintenant d'aborder la description quantitative formant la base d'une méthodologie de nature probabiliste, en suivant comme annoncé une démarche essentiellement pragmatique. Dans un premier temps, on s'en tient à une approche intuitive où il sera supposé que le nombre d'événements élémentaires est soit fini, soit dénombrable ; ceci veut dire que le cardinal de Ω est un nombre entier fini ou infini : Ω pourra toujours être mis en bijection avec $\{1, 2, \dots, N_0\}$ ou avec \mathbb{N} . En conséquence, l'ensemble des événements composites construits avec les $\omega_n \in \Omega$ est forcément dénombrable.

Énoncer la liste des événements élémentaires possibles est un premier stade, de type qualitatif. Pour donner une véritable dimension quantitative à la description théorique, il convient d'attribuer un nombre mesurant

¹³ c'est dans ce cas l'événement dit *certain* !

¹⁴ c'est l'événement dit *impossible*.

¹⁵ pour des événements *exclusifs* – voir plus loin.

¹⁶ pour des événements *indépendants* – voir plus loin.

que tel ou tel événement se produit plus ou moins souvent. Pour ce faire, l'approche intuitive consiste à se reposer sur une longue série d'expériences réputées identiques¹⁷, et à faire le relevé soigneux du nombre de fois où un événement donné s'est produit ; clairement, il suffit d'effectuer ce travail sur les événements élémentaires. Si N est le nombre total d'expériences, et si l'événement (élémentaire) ω_n est survenu N_n fois, on définit alors sa *fréquence statistique* f_n par le rapport :

$$\text{fréquence d'observation de } \omega_n = f_n = \frac{N_n}{N} . \quad (8.3)$$

L'acte de foi consiste alors à *admettre* que dans la limite d'un nombre infini d'expériences, chaque fréquence a une limite qui est, suivant cette définition pragmatique, la *probabilité* de l'événement ω_n :

$$\forall \omega_n \in \Omega, \exists \lim_{N \rightarrow +\infty} f_n \quad \text{et} \quad p_n = \lim_{N \rightarrow +\infty} f_n \iff \text{Probabilité de } \omega_n = p_n . \quad (8.4)$$

Évidemment, la convergence (supposée) de la suite des f_n vers les p_n est fort subtile : en réalité, chaque série d'expériences a sa propre suite de fréquences. Si on fixe n , en se concentrant sur un événement donné, on peut tracer une trajectoire donnant f_n en fonction du nombre d'expériences N . Cette trajectoire a un aspect "rugueux" et passe sans cesse d'un côté à l'autre de la valeur-limite p_n , en suivant le plus souvent un chemin très surprenant : il peut rester très longtemps d'un même et unique côté, faire une brève excursion de l'autre, avant de refranchir la limite et s'en écarter en donnant l'impression de n'y plus vouloir revenir. Par ailleurs, pour un événement donné, la trajectoire varie d'une série d'expériences à l'autre. Le point essentiel est de postuler que toutes les trajectoires relatives à un événement élémentaire donné ont une limite commune – dans le langage convenu, on peut parler de *bassin d'attraction*, vers lequel convergent toutes les trajectoires (*tous les chemins mènent à Rome*).

Ici, rien n'est donc affirmé sur la loi de convergence des fréquences f_n vers leurs valeurs limites p_n : la seule hypothèse, à ce stade, est l'existence de la limite p_n pour chaque fréquence f_n . La grande difficulté de l'échantillonnage statistique tient au fait qu'il faut en général un *très grand* nombre d'expériences pour que la fréquence f_n soit proche de la valeur théorique p_n .

Par leur définition (8.3), les fréquences sont des quantités positives ; leur limite p_n (la probabilité de ω_n) est donc elle aussi positive. En outre, par construction, la somme des fréquences est, quel que soit N , égale à 1 :

$$\sum_n f_n = \sum_n \frac{N_n}{N} = \frac{\sum_n N_n}{N} = \frac{N}{N} = 1 . \quad (8.5)$$

Il en va de même des limites p_n des fréquences :

$$\sum_n p_n = 1 . \quad (8.6)$$

Par ailleurs, la fréquence d'avoir un événement ω_n ou un autre événement $\omega_{n'}$ est évidemment la *somme* de leurs fréquences respectives, somme du nombre de cas où ω_n s'est produit et du nombre de cas où $\omega_{n'}$ s'est produit :

$$\text{fréquence d'observation de } (\omega_n \text{ ou } \omega_{n'}) = f_n + f_{n'} ; \quad (8.7)$$

on a donc aussi :

$$\text{Probabilité de } (\omega_n \text{ ou } \omega_{n'}) = p_n + p_{n'} . \quad (8.8)$$

L'événement composite : obtention de ω_n ou de $\omega_{n'}$, noté $\omega_n \cup \omega_{n'}$, conformément à ce qui a été mentionné plus haut, a donc tout naturellement une fréquence égale à la *somme* des fréquences, une propriété qui se reporte sur les probabilités correspondantes (répétons que deux événements élémentaires sont forcément exclusifs). D'un autre côté, la composition *et* est en un sens triviale lorsqu'il s'agit d'événements élémentaires, pour lesquels on a toujours $\omega_n \cap \omega_{n'} = \emptyset$, d'où :

$$\text{Probabilité de } (\omega_n \text{ et } \omega_{n'}) = 0 \quad \text{pour tout couple d'événements élémentaires} . \quad (8.9)$$

¹⁷On appelle souvent *tirage* ou *échantillonnage*, une telle longue suite d'expériences.

À ce stade, on pourrait être tenté de vouloir définir la notion de probabilité sans faire référence à la notion de fréquence statistique. Par exemple, pour le dé cubique parfait, on pourrait dire d'emblée que la probabilité de chaque face est $\frac{1}{6}$, rapport entre le nombre de cas favorables et le nombre de cas possibles. En réalité, cette définition n'en est pas une, car elle suppose (implicitement) que chaque face a... la même probabilité que toutes les autres ! En effet, s'il n'en va pas ainsi (dé pipé), il est bien clair que les probabilités des différentes faces ne sont pas égales entre elles et ne valent pas toutes $\frac{1}{6}$. Il en résulte que la définition ci-dessus devrait être précisée en disant que la probabilité est le rapport du nombre de cas favorables au nombre de cas possibles, *sachant que tous les cas possibles ont la même probabilité*. Alors le caractère spécieux de cette définition saute aux yeux : pour définir la notion de probabilité, on fait appel à... la notion de probabilité. En Logique, une telle définition est souvent dite *circulaire*. Quoi qu'il en soit, cette définition n'est pas satisfaisante¹⁸. On s'en tiendra donc ici à la définition intuitive présentée ci-dessus et reposant fondamentalement sur la notion de fréquence statistique d'observation, une attitude pas vraiment incompatible avec le fait que la Physique est d'abord une science expérimentale !

Pour terminer cette approche intuitive, il convient de rappeler le lien étroit entre la Théorie des probabilités et l'analyse combinatoire, elle-même liée à la Théorie des jeux. Ces deux disciplines ont joué un rôle moteur dans l'édification de la Théorie des probabilités. Fondamentalement, le lien est précisément la définition (circulaire !) énoncée plus haut, la résolution de chaque problème consistant à dénombrer les différents cas pour finalement retenir la probabilité cherchée comme le rapport du nombre de cas favorables au nombre de cas possibles. Dans cet exposé élémentaire, l'aspect combinatoire sera délaissé, au profit de l'introduction d'instruments d'usage courant en Physique, étant admis que les bases des techniques élémentaires de dénombrement sont connues.

8.2 Axiomes. Premières conséquences

Les considérations précédentes permettent de construire un cadre théorique où, sans vouloir formaliser à l'extrême, il est bon de rassembler un nombre minimum de propositions non-contradictaires définissant un cadre cohérent (axiomatisation élémentaire), et débouchant sur des instruments de travail indispensables, dont le plus important est ce que l'on appelle la *fonction de répartition*¹⁹ d'une variable aléatoire, généralement notée F , définie précisément dans la section 8.3.

8.2.1 Axiomes

Il s'agit ici de rationaliser les constatations intuitives précédentes pour énoncer finalement les axiomes sur lesquels est construite la version élémentaire de Théorie des probabilités exposée dans ce chapitre. Les affirmations suivantes sont faites à propos de l'espace des épreuves Ω , formant l'ensemble des événements élémentaires. Elles permettront de trouver la probabilité de tout événement qui est une combinaison des événements élémentaires.

1. toute probabilité est un nombre compris entre 0 et 1 :

$$0 \leq p(\omega_n) \leq 1 \quad \forall \omega_n \in \Omega . \quad (8.10)$$

2. la probabilité de l'événement certain est égale à 1, ce que l'on peut noter²⁰ :

$$p(\Omega) = 1 \quad \iff \quad \sum_{n \in \mathbb{N}} p_n = 1 . \quad (8.11)$$

¹⁸C'est l'une des difficultés historiques de la théorie première des probabilités. L'autre difficulté majeure survient quand le nombre d'événements a la puissance du continu. Dans tous les exemples intuitifs présentés ci-dessus, le nombre d'événements est fini. La plupart des résultats s'étendent sans trop de difficulté au cas où Ω est isomorphe à \mathbb{N} , moyennant la définition convenable de la convergence sous différentes formes. En revanche, si Ω a la puissance du continu, la difficulté est immense ; c'est sa résolution effective qui a motivé les travaux de Kolmogorov, et la formulation axiomatique qu'il a proposée en 1933.

¹⁹On dit aussi *loi de répartition*.

²⁰On suppose toujours Ω fini ou dénombrable. Toutes les sommations peuvent être notées sur \mathbb{N} : le cas où Ω possède un nombre fini N_0 événements élémentaires s'obtient en posant formellement que toute probabilité $p_{m > N_0}$ est nulle.

Cette relation est le pendant de (8.6).

3. la probabilité de l'événement consistant en la réalisation de l'un ou l'autre de deux événements exclusifs²¹ est la somme de leurs probabilités :

$$\text{Prob} [\omega_n \text{ ou } \omega_{n'}] = p_n + p_{n'} . \quad (8.12)$$

Ceci est l'affirmation de principe découlant de (8.7) et (8.8). Dans la limite où l'on considère *tous* les événements ω_n de Ω , on obtient $\text{Prob} [\omega_1 \text{ ou } \omega_2 \text{ ou } \dots \text{ ou } \omega_n \text{ ou } \dots]$, soit la somme de toutes les probabilités p_n , qui vaut 1 d'après l'axiome 2.

8.2.2 Premières conséquences

Des affirmations précédentes résultent immédiatement un certain nombre de règles d'usage constant, énoncées d'abord à propos des événements élémentaires ω_n :

1. si ω_n a la probabilité p_n , le contraire de ω_n a la probabilité $1 - p_n$. En effet, l'événement contraire $\tilde{\omega}_n$ de ω_n consiste à d'obtenir n'importe lequel des autres $\omega_{n'}$, $n \neq n'$; d'après l'axiome 3, la probabilité d'obtenir ω_1 ou ω_2 ou ... ou ω_{n-1} ou ω_{n+1} ou ... est $\sum_{n' \neq n} p_{n'}$. D'après (8.11) :

$$p_n + \sum_{n' \neq n} p_{n'} = 1 \iff \sum_{n' \neq n} p_{n'} = 1 - p_n ; \quad (8.13)$$

d'où la règle importante :

$$\text{Prob} [\tilde{\omega}_n \equiv \text{contraire de } \omega_n] = 1 - \text{Prob} [\omega_n] . \quad (8.14)$$

2. tout événement impossible a une probabilité nulle. En effet, la probabilité d'obtenir ω_n est p_n , celle d'obtenir le contraire est $1 - p_n$. L'événement impossible est d'obtenir le contraire de ω_n ou son contraire $\tilde{\omega}_n$, deux événements qui s'excluent mutuellement ; la probabilité de l'événement impossible est donc $1 - [p_n + (1 - p_n)] = 0$.

En ce qui concerne tout événement quelconque A , pas forcément élémentaire, les affirmations suivantes sont vraies :

1. la probabilité de A , $P(A)$, est un nombre compris entre 0 et 1

$$0 \leq P(A) \leq 1 \quad \forall A . \quad (8.15)$$

2. la probabilité d'un événement certain est égale à 1

3. la probabilité d'un événement impossible est égale à 0. La réalisation de A et B quand A et B sont incompatibles est un événement impossible :

$$A \cap B = \emptyset \implies P(A \cap B) = 0 . \quad (8.16)$$

4. la probabilité de \tilde{A} , contraire de A , est $P(\tilde{A}) = 1 - P(A)$

5. si $A \subset B$, alors $P(A) \leq P(B)$; autrement dit :

$$(A \implies B) \iff P(A) \leq P(B) . \quad (8.17)$$

²¹On dit aussi *incompatibles*, comme on parle des observables incompatibles en Mécanique quantique.

6. la probabilité de deux événements exclusifs A et B est la somme de leurs probabilités :

$$P(A \text{ ou } B) = P(A) + P(B) \iff P(A \cup B) = P(A) + P(B) . \quad (8.18)$$

Plus généralement, si tous les A_n sont deux à deux incompatibles :

$$P\left(\bigcup_n A_n\right) = \sum_n P(A_n) . \quad (8.19)$$

7. Les deux événements (A ou B) et (A et B) sont clairement incompatibles et la somme de leurs probabilités est la probabilité d'avoir ou A , ou B ; il en résulte que, quels que soient A et B :

$$P(A \cup B) + P(A \cap B) = P(A) + P(B) \iff P(A \cup B) = P(A) + P(B) - P(A \cap B) . \quad (8.20)$$

(8.18) est le cas particulier de (8.20) lorsque A et B sont incompatibles, auquel cas $P(A \cap B) = 0$.

8. la probabilité *conditionnelle* est la probabilité d'un événement sachant qu'un autre est réalisé. On note ainsi $P(A|B)$ la probabilité de A sachant que B est réalisé. On peut alors écrire la formule²² dite de Bayes donnant la probabilité pour que A et B soient réalisés :

$$\text{Prob}[A \text{ et } B] \equiv P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) . \quad (8.21)$$

Il est préférable de lire cette formule à l'envers :

$$P(A|B) \stackrel{\text{déf}}{=} \frac{\text{Prob}[A \text{ et } B]}{P(B)} \equiv \frac{P(A \cap B)}{P(B)} , \quad (8.22)$$

la considérant comme la *définition* de la probabilité conditionnelle.

Dans les dernières formules, A et B jouent des rôles strictement symétriques. Ceci permet de définir la notion d'événements *indépendants* ; A et B sont indépendants si $P(A|B)$ ne dépend pas de B ou, de façon équivalente, si $P(B|A)$ ne dépend pas de A :

$$A \text{ et } B \text{ indépendants} \iff P(A|B) \text{ indépendant de } B \iff P(B|A) \text{ indépendant de } A . \quad (8.23)$$

Dans ces conditions, la probabilité conditionnelle $P(A|B)$ ne dépend nullement de B , et n'est autre que la probabilité attachée au seul événement A , $P(A)$, et de même pour B , les rôles étant inversés. Pour des variables indépendantes, on a donc $P(A|B) = P(A)$, $P(B|A) = P(B)$ et (8.20) devient :

$$A \text{ et } B \text{ indépendants, } \text{Prob}[A \text{ et } B] = P(A)P(B) . \quad (8.24)$$

Cette relation joue un rôle de tout premier plan dans les applications. On peut dire que la difficulté d'un problème augmente *infiniment* lorsque les variables ne sont pas indépendantes, un peu comme un système de particules en interaction est en un sens infiniment plus difficile à étudier qu'un ensemble des mêmes particules sans interaction mutuelle.

8.3 Fonction de répartition

La fonction de répartition (et ses produits dérivés²³) est l'instrument primordial pour la description quantitative d'une variable aléatoire. Si dans les cas "ordinaires" F n'est pas à proprement parler l'outil que l'on manipule le plus souvent (quand les densités (au sens large) existent, elles sont plus indiscutablement plus usitées), cette fonction demeure conceptuellement l'objet fondamental de la description exhaustive de la "dispersion" des valeurs inhérente à la notion de variable aléatoire, et à la mesure quantitative de ce que le sens commun perçoit comme leur fréquence d'apparition plus ou moins grande.

²²On dit parfois *axiome* de Bayes.

²³sans jeu de mot !

Dans la suite de cette section, X désigne une variable aléatoire²⁴ prenant une suite de valeurs discrètes x_n , chacune de ces dernières étant associée à la réalisation d'un événement particulier. Par exemple, pour la pièce de monnaie, si on associe la valeur -1 à face, la valeur $+1$ à pile, le résultat du lancer de la pièce constitue la variable aléatoire X , qui peut prendre les deux valeurs²⁵ $x_1 = -1$ et $x_2 = +1$.

L'ensemble des x_n est une image parfaite de l'espace des épreuves Ω , et on peut les confondre dans une seule et même notation :

$$\{x_1, x_2, \dots, x_m, \dots\} = \Omega . \quad (8.25)$$

Pour la commodité des raisonnements, on supposera que l'indexation des valeurs de l'aléatoire X a été choisie une fois pour toutes de sorte que :

$$x_1 < x_2 < \dots < x_m < \dots . \quad (8.26)$$

La fonction de répartition, notée $F(x)$, est par définition la fonction dont la valeur en x est la probabilité pour que X prenne une valeur inférieure ou égale à x . Très précisément, on définit $F(x)$ comme suit²⁶ :

$$F(x) \stackrel{\text{déf}}{=} \text{Prob}[X \leq x] . \quad (8.27)$$

De cette définition, résultent immédiatement les propriétés suivantes :

1. $F(x)$ est une fonction positive ou nulle (c'est une probabilité)
2. La probabilité est nulle que X prenne une valeur inférieure à $-\infty$:

$$F(-\infty) = 0 . \quad (8.28)$$

3. La probabilité que X prenne une valeur inférieure à $+\infty$ est égale à 1 :

$$F(+\infty) = 1 . \quad (8.29)$$

4. $F(x)$ est une fonction non-décroissante. En effet, si $a < b$, alors $X \leq a \implies X < b$; d'après (8.17), il en résulte :

$$F(a) \leq F(b) \quad (a < b) . \quad (8.30)$$

5. Il est facile de donner un sens à la différence des valeurs de F entre deux points. Soit deux nombres a et b tels que $a < b$. La probabilité que $x \leq b$ peut se décomposer en la somme de deux événements exclusifs : (X est inférieur ou égal à a) ou (X est compris entre a et b). En terme d'équation, on a donc :

$$\text{Prob}[X \leq b] = \text{Prob}[X \leq a] + \text{Prob}[a < X \leq b] . \quad (8.31)$$

En vertu de la définition de F , ceci se transcrit comme suit :

$$F(b) = F(a) + \text{Prob}[a < X \leq b] , \quad (8.32)$$

soit :

$$\text{Prob}[a < X \leq b] = F(b) - F(a) . \quad (8.33)$$

6. La fonction $F(x)$ est partout continue à droite, c'est-à-dire que :

$$\forall \varepsilon > 0, \lim_{\varepsilon \rightarrow 0} F(x + \varepsilon) = F(x) \iff F(x + 0) = F(x) \quad \forall x \quad (8.34)$$

Dans le cas de la pièce avec les deux valeurs conventionnelles choisies ± 1 , à quoi ressemble la fonction de répartition ? Tant que $x < -1$, plus petite valeur possible de X , F est nulle. Dès que x franchit la valeur -1 ,

²⁴ X est ici généralement une variable sans dimension dont les valeurs x_n sont des nombres purs. En pratique, X est le rapport entre une grandeur physique L et une échelle de référence l_0 faisant office d'unité.

²⁵ Une telle variable est dite *binnaire* ou *de Bernoulli*. Dans le contexte des phénomènes critiques, on parle de *variable d'Ising*, en référence à la *Drosophile* des experts en la matière.

²⁶ On définit parfois $F(x)$ comme $F(x) = \text{Prob}[X < x]$.

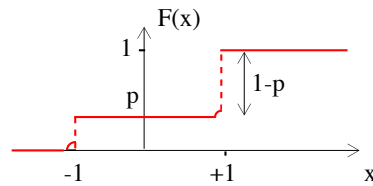


Figure 8.1: Fonction de répartition $F(x)$ pour le jeu de pile ou face ; pour une pièce parfaite $p = \frac{1}{2}$ Le quart de rond à gauche de chaque saut rappelle que $F(x)$ est définie en chaque saut par sa limite à droite (continuité à droite).

F est égale à la probabilité associée à face, soit p (p vaut $\frac{1}{2}$ pour une pièce parfaite), puisque $\text{Prob}[x \leq a]$ avec $-1 < x$ et $a < +1$ (la valeur attachée à pile) est la probabilité d'obtenir face. Tant que x , tout en augmentant, reste inférieur à $+1$, la probabilité ne change pas, F reste donc constante et égale à p (traduisant le fait qu'il n'y a pas d'événement "intermédiaire" entre pile et face). Enfin, quand $x \geq +1$, $F(x)$ est la probabilité d'avoir ou pile ou face, et vaut donc 1. En définitive, F est une fonction en *escalier*, variant par saut à chaque fois que x franchit l'une des valeurs possibles pour l'aléatoire X (c'est cette fonction F que les probabilistes appellent *histogramme*, contrairement au langage du *vulgum pecus*). À chaque valeur, le saut est égal à la probabilité de l'événement correspondant (voir fig. 8.1). On doit d'ores et déjà noter que la *dérivée* de la fonction F est nulle presque partout, puisque F est une fonction constante par morceaux²⁷.

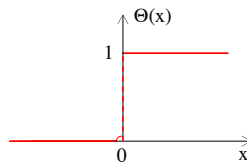


Figure 8.2: Graphe de la fonction $\Theta(x)$ définie en (8.37). Cette fonction est continue à droite.

C'est le bon moment pour examiner précisément F au voisinage d'un saut, en illustration de la propriété 6 ci-dessus. Considérons le premier saut en $x_1 = -1$; par définition de F :

$$x < -1, F(x) = \text{Prob}[X \leq x < -1] = 0, \quad x = -1, F(x) = \text{Prob}[X \leq x = -1] = p, \quad (8.35)$$

d'où $F(-1-0) = 0$, $F(-1) = p$ et $F(x) = p$ si $-1 < x < +1$, d'où $F(-1+0) = F(-1)$, exprimant la continuité à droite. Il en va de même près de $x = +1$. Finalement, pour la pièce de monnaie et le jeu de pile ou face :

$$F(x) = \begin{cases} 0 & x < -1 \\ p & -1 \leq x < +1 \\ 1 & x \geq +1 \end{cases} . \quad (8.36)$$

Introduisons la fonction échelon $\Theta(x)$ précisément définie²⁸ comme :

$$\Theta(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases} ; \quad (8.37)$$

alors la fonction (8.36) s'écrit :

$$F(x) = p\Theta(x+1) + (1-p)\Theta(x-1) . \quad (8.38)$$

²⁷Bien sûr, le physicien écrira $F'(x) = p\delta(x+1) + (1-p)\delta(x-1)$: pour lui, la dérivée de F est un peigne de Dirac à deux dents. On reviendra par la suite sur cette audace, au demeurant fort légitime.

²⁸ Θ diffère d'un détail de la fonction de Heaviside définie antérieurement : Θ est définie en $x = 0$ et vaut 1 (voir fig. 8.2).

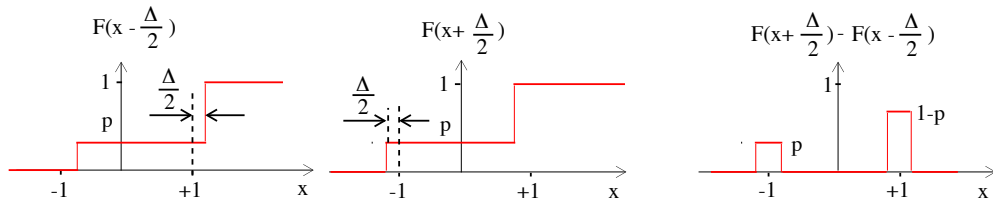


Figure 8.3: Construction (à gauche) de l’“histogramme” (à droite) donnant la probabilité de trouver une variable binaire dans l’intervalle $]x - \frac{\Delta}{2}, x + \frac{\Delta}{2}]$. Chaque bâton a une largeur Δ .

Ces éléments étant acquis, on peut par exemple trouver la probabilité pour que X soit compris dans un intervalle d’amplitude donné autour d’une valeur x quelconque. Plus précisément, compte tenu de (8.33) :

$$\text{Prob} \left[x - \frac{\Delta}{2} < X \leq x + \frac{\Delta}{2} \right] = F \left(x + \frac{\Delta}{2} \right) - F \left(x - \frac{\Delta}{2} \right) \equiv H_{\Delta}(x) . \quad (8.39)$$

Le graphe de $H_{\Delta}(x)$ se déduit aisément de celui de $F(x)$, en décalant chaque tracé de $\frac{\Delta}{2}$ vers la droite, qui donne $F(x - \frac{\Delta}{2})$, vers la gauche, qui donne $F(x + \frac{\Delta}{2})$ et en faisant la différence pour fabriquer $H_{\Delta}(x)$ conformément à (8.39) (voir fig. 8.3). C’est ce type de dessin que l’on appelle “histogramme” dans le langage courant, en contradiction avec la terminologie des experts en Théorie des probabilités. $H_{\Delta}(x)$ est la différence de deux fonctions continues à droite, c’est aussi une fonction continue à droite.

Pour une variable aléatoire X prenant N_0 valeurs distinctes ordonnées conformément à (8.26), et telles que $\text{Prob} [X = x_n] = p_n$, la fonction de répartition $F(x)$ présente un saut à chaque fois que son argument x franchit une nouvelle valeur possible x_n . Chaque saut a une amplitude égale à la probabilité p_n de l’événement où X prend la valeur x_n ; la variation totale de F entre $\pm\infty$ est égale à la somme des sauts, soit $p_1 + p_2 + \dots + p_{N_0}$, qui est égale à 1 comme il se doit. À nouveau, la dérivée de F est nulle presque partout²⁹, et F est l’escalier représenté sur la fig. 8.4. L’expression de F est dans ce cas :

$$F(x) = \sum_{n=1}^{N_0} p_n \Theta(x - x_n) , \quad (8.40)$$

en généralisation immédiate de (8.38).

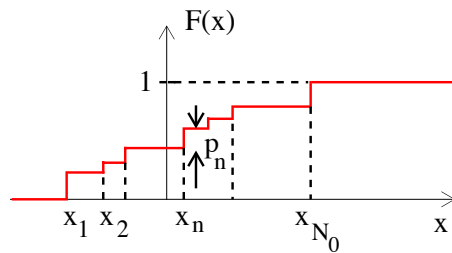


Figure 8.4: Fonction de répartition $F(x)$ pour une aléatoire discrète à N_0 valeurs.

8.4 Variables aléatoires continues

Jusqu’à présent, toutes les variables aléatoires considérées sont *discrètes*, au sens où le nombre d’événements élémentaires est soit fini, soit dénombrable, permettant d’imager Ω par un ensemble $\{x_1, x_2, \dots\}$ de cardinal fini ou ayant la puissance de \mathbb{N} .

²⁹et, à nouveau, le physicien écrira $F'(x) = \sum_n p_n \delta(x - x_n)$.

D'un autre côté, la notion de variable aléatoire peut se généraliser au cas où les valeurs possibles sont dans un intervalle inclus au sens large dans \mathbb{R} (ce peut être \mathbb{R} lui-même). En effet, rien dans la notion de v.a. ne s'oppose à considérer un tel cas, c'est même à la réflexion le cas le plus naturel, puisque (théoriquement) la plupart des grandeurs physiques prennent en principe leurs valeurs dans \mathbb{R} (le cas discret est à la réflexion plutôt exceptionnel, sauf dans le domaine quantique évidemment). Le physicien peut par exemple se poser la question : quelle est la probabilité pour que la vitesse d'un atome de gaz soit inférieure à une valeur v_0 choisie à l'avance ? Ou : quelle est la probabilité pour que l'activité d'une source radioactive soit supérieure à un seuil de tolérance donné ? Dans la vie de tous les jours, les variables réputées aléatoires sont souvent des variables prenant des valeurs continues : quelle est la probabilité pour que la taille d'un individu soit inférieure à une taille donnée ?

La considération de variables aléatoires *continues* pose de grandes difficultés de fond ; il est même possible de mettre en évidence des problèmes apparemment bien posés pouvant recevoir différentes réponses, en raison du *flou* concernant la définition ce que l'on appelle un cas favorable (notamment, *Paradoxe de Bertrand*). On admettra dans la suite que le problème posé n'est pas suffisamment *vicieux* pour conduire à de tels paradoxes.

Quoi qu'il en soit, la fonction F reste l'outil de référence, et se prête immédiatement à la généralisation aux variables continues. De fait, la définition (8.27) peut se recopier à l'identique pour une variable aléatoire X dont les valeurs possibles sont denses :

$$F(x) \stackrel{\text{déf}}{=} \text{Prob}[X \leq x] . \quad (8.41)$$

La définition étant strictement la même que pour une variable discrète, toutes les propriétés 1 à 6 de F énoncées plus haut à propos d'une variable aléatoire discrète restent valides. Le point qualitatif qui d'emblée saute aux yeux est que maintenant F ne varie plus seulement par morceaux : les valeurs possibles de l'aléatoire X étant réparties de façon dense par intervalles disjoints ou non, $F(x)$ augmente graduellement. Schématiquement, F n'est plus réduite à une fonction en escalier constante par morceaux, mais monte graduellement entre deux marches, ou même ne présente aucune marche.

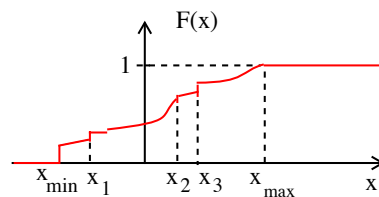


Figure 8.5: Fonction de répartition $F(x)$ pour une aléatoire prenant ses valeurs dans $[x_{\min}, x_{\max}] \subset \mathbb{R}$, dont les probabilités varient par saut pour certaines valeurs discrètes x_n . Le graphe de F_{cont} est la juxtaposition continue des arcs disjoints. Celui de F_{at} est (ici) un escalier à trois marches en x_i , $i = 1, 2, 3$. Souvent, les deux composantes F_{at} et F_{cont} sont définies sur deux intervalles disjoints.

Un résultat majeur dû à Lebesgue décrit les propriétés de la fonction de répartition F pour une variable aléatoire *quelconque*, dont les valeurs possibles sont denses par intervalles, avec en plus la possibilité de valeurs discrètes, isolées les unes des autres, noyées dans les intervalles denses ou à l'extérieur de ceux-ci. Ce résultat peut s'énoncer comme suit : F peut toujours se décomposer au plus en la somme de trois composantes, notées F_{at} , F_{cont} et $F_{\text{sing cont}}$:

$$F = F_{\text{at}} + F_{\text{cont}} + F_{\text{sing cont}} , \quad (8.42)$$

avec les propriétés suivantes :

1. chaque composante est non décroissante
2. F_{at} est la *fonction des sauts*, constante en dehors des points de discontinuité de F formant un ensemble *dénombrable*. Aux points de discontinuité de F , le saut de F_{at} est égal à celui de F . En d'autres termes, F_{at} représente la partie "escalier" de F , quand elle existe. La valeur à droite de F_{at} est égale à la somme des sauts p_n . En Physique, une composante de type F_{at} est souvent appelée *distribution atomique*

3. La différence $F - F_{\text{at}}$ est donc une fonction continue, puisque c'est F_{at} qui contient toute la partie discontinue de F . Elle se décompose à son tour en deux morceaux :

$$F - F_{\text{at}} = F_{\text{cont}} + F_{\text{sing cont}} , \quad (8.43)$$

- (a) F_{cont} est une fonction *absolument continue*, c'est-à-dire qu'elle peut s'écrire sous la forme d'une intégrale :

$$F_{\text{cont}}(x) = \int_{-\infty}^x \rho(x') dx' ; \quad (8.44)$$

en termes plus ordinaires, la dérivée de F_{cont} existe, c'est ρ . Revenant à (8.33) et prenant $a = x$ et $b = x + \Delta$, on en déduit (puisque F' existe par hypothèse) compte tenu de (8.44) :

$$dF_{\text{cont}}(x) = \rho(x)dx \iff \rho(x) = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} [F(x + \Delta) - F(x)] ; \quad (8.45)$$

tout naturellement, la fonction $\rho(x)$ est appelée³⁰ *densité de probabilité* associée à F

- (b) enfin, $F_{\text{sing cont}}$ est la différence $F - F_{\text{at}} - F_{\text{cont}}$; c'est une fonction non décroissante, continue et non différentiable, souvent appelée *composante singulière continue*. En quelque sorte, cette composante un peu exotique est ce qui reste dans F quand on lui a retranché ce à quoi on est plus ou moins habitué. Il n'y a pas de doute que si on trouve $F_{\text{at}}(x > x_{\text{max}}) + F_{\text{cont}}(x > x_{\text{max}}) < 1$, alors une telle composante existe

Il n'est pas aisé de se faire une représentation mentale intuitive d'une fonction continue non dérivable ; on peut partir de l'image d'une ligne rugueuse constituée de tout petits segments de droite ; dans un deuxième temps, on décompose chaque petit segment en un grand nombre d'autre petits segments d'orientation "désordonnée" (mais de sorte que la fonction soit toujours non décroissante), et on recommence. . . Les *escaliers du diable* que l'on rencontre en Physique à propos de certains systèmes désordonnés peuvent fournir un exemple de composante singulière continue (dans la mesure où on peut diviser sans fin). L'expression tient à l'image simplifiée suivante : à l'étape zéro, on part d'un escalier avec un certain nombre de marches ; puis on effectue une transformation d'échelle qui décompose chacune des marches elle-même en un escalier, et l'on itère à l'infini, l'intervalle en abscisse entre la première et la dernière marche étant fixé une fois pour toutes.

Cette situation est typique des systèmes possédant une invariance d'échelle, traduisant le fait qu'ils ont le même aspect quand on fait un premier *zoom*, puis un second, puis un troisième, *etc* : ces systèmes sont communément appelés *fractales*, et offrent à l'œil un spectacle magnifique, étrange, voire parfois inquiétant.

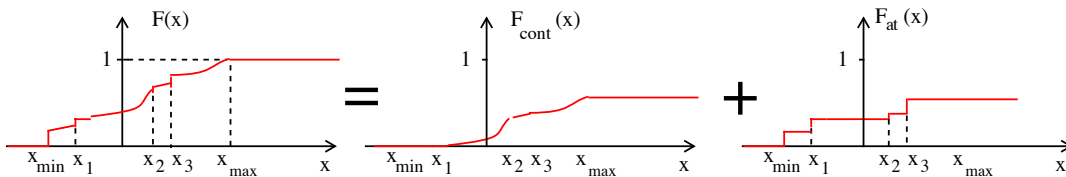


Figure 8.6: Illustration de la décomposition de Lebesgue dans le cas où il n'y a pas de composante singulière continue

Bien évidemment, une ou deux composantes de F dans (8.42) peut(ven) être(s) absente(s), c'est-à-dire être identiquement nulle³¹. Dans toute la suite, on ne considèrera plus la possibilité d'une composante du type $F_{\text{sing cont}}$; pour simplifier encore les choses, on se placera le plus souvent dans deux cas :

³⁰Si X a une dimension physique, $\rho(x)$ a la dimension inverse puisque le produit $\rho(x)dx$ est un nombre pur.

³¹Par exemple, si on considère l'énergie E de l'atome d'hydrogène dans un état Ψ donné comme une variable aléatoire (c'en est une, aux dires de la Mécanique quantique !), sa fonction de répartition a l'allure suivante :

$$F(E) = \underbrace{\sum_{n \in \mathbb{N}} p_n \Theta \left(E + \frac{E_0}{n^2} \right)}_{F_{\text{at}}} + \underbrace{\int_0^E \rho(E') dE'}_{F_{\text{cont}}} \quad (8.46)$$

- F n'a qu'une composante de type F_{at} : c'est en fait le cas d'une variable aléatoire discrète, qui a servi de point de départ pour la définition de la fonction de répartition (section 8.3)
- F n'a qu'une composante de type F_{cont} , c'est-à-dire que d'une part la variable aléatoire est continue, que d'autre part il existe une densité de probabilité. Compte tenu de (8.44), on peut aussi écrire :

$$\text{Prob}[x < X \leq x + dx] = F(x + dx) - F(x) = F'(x)dx = \rho(x)dx . \tag{8.47}$$

En fait, le premier cas peut être considéré comme un cas particulier du second, si l'on accepte de manipuler la fonction de Dirac, en la prenant sans état d'âme comme la *dérivée* de la fonction $\Theta(x)$ définie en (8.37). Alors, dans les deux situations précédentes, F admet une dérivée et tous les cas considérés dans la suite³² peuvent être traités en bloc, en jouant avec la densité de probabilité associée à F . Ceci étant admis, on pourra dès lors toujours supposer que toute fonction de répartition sans composante singulière continue est donnée par :

$$F(x) = \sum_{n \in \mathbb{N}} p_n \Theta(x - x_n) + \int_{-\infty}^x \rho(x') dx' \tag{8.48}$$

et admet la dérivée $F'(x) = p(x)$ donnée par³³ :

$$p(x) = \sum_{n \in \mathbb{N}} p_n \delta(x - x_n) + \rho(x) . \tag{8.49}$$

En toute rigueur, c'est $\delta^{(+)}$ qui figure dans la composante continue. Répétons que l'une ou l'autre de ces deux composantes peut parfaitement être identiquement nulle : si $\rho(x) = 0$, la variable X est purement discrète, elle est purement continue si tous les poids p_n sont nuls. Dans la littérature francophone, $p(x)$ s'appelle *loi de distribution*.

On vérifiera d'ici de là que cette méthode opératoire fonctionne bien. En particulier, en intégrant membre à membre (8.49) entre $\pm\infty$, on trouve³⁴ :

$$\int_{-\infty}^{+\infty} p(x) dx = \sum_{n \in \mathbb{N}} p_n + \int_{-\infty}^{+\infty} \rho(x) dx = 1 , \tag{8.50}$$

après avoir remarqué que le premier membre est $F(+\infty) - F(-\infty)$, une différence égale à $1 - 0 = 1$, par définition de la fonction de répartition. Il en résulte que $p(x)$ est une fonction intégrable, au sens où son intégrale entre $\pm\infty$ est nécessairement finie (et vaut 1). On note que l'existence des deux composantes atomique et continue entraîne :

$$\exists F_{\text{at}} \text{ et } F_{\text{cont}} \implies \sum_{n \in \mathbb{N}} p_n < 1 \text{ et } \int_{-\infty}^{+\infty} \rho(x) dx < 1 . \tag{8.51}$$

Pour faire court, on peut aussi écrire (8.50) sous la forme $\int p = 1$. À ce propos, on doit remarquer que trouver par exemple $\int_{-\infty}^{+\infty} \rho(x) dx < 1$ est le symptôme de l'existence d'une composante atomique.

Notons que rien n'interdit à une densité $\rho(x)$ de diverger en un point, mais elle doit rester localement intégrable. C'est pourquoi toute divergence du genre $|x - x_0|^{-\alpha}$ est acceptable³⁵ pourvu que l'exposant (positif) α soit plus petit que 1.

où Θ est la fonction de Heaviside définie en (8.37). Les p_n et la fonction $\rho(E)$ dépendent de l'état Ψ considéré et satisfont la relation (dite de *normalisation*) $\sum_{n \in \mathbb{N}} p_n + \int_0^{+\infty} \rho(E) dE = 1$; E_0 est l'échelle typique d'énergie de la Physique atomique et vaut à peu près 13.6 eV. La partie discrète (énergies négatives, *spectre discret*) correspond aux états liés : ce sont eux qui représentent l'atome constitué (d'où, sans doute, la terminologie F_{at} pour une fonction de répartition en escalier) ; la partie continue (*spectre continu*) représente les états non-liés, décrivant quantiquement ce qui se passe quand on envoie un électron (projectile) en direction d'un proton fixe (cible). Dans le cas considéré, F contient effectivement les deux composantes, le sens physique de chacune étant parfaitement clair.

³² c'est-à-dire avec l'hypothèse que la troisième composante de Lebesgue $F_{\text{sing cont}}$ est nulle.

³³ La sommation $n \in \mathbb{N}$ ne signifie pas qu'il y a une infinité dénombrable d'événements discrets : cette notation contient le cas où seuls des p_n en nombre fini sont non nuls.

³⁴ en admettant éventuellement toutes les hypothèses de convergence uniforme nécessaires.

³⁵ C'est aussi la raison pour laquelle la fonction d'onde de la Mécanique quantique – en dehors de toute autre considération – ne peut avoir de singularité plus violente que $|x - x_0|^{-\alpha/2}$, $\alpha < 1$.

Un exemple d'escalier du diable pas vraiment diabolique

Imaginons que la fonction de répartition $F(x)$ d'une aléatoire X , dont on sait par ailleurs qu'elle prend ses valeurs sur l'intervalle $[0, 1] \subset \mathbb{R}$, satisfait l'équation suivante :

$$F(x) = pF(\alpha x) \quad (x < 1) , \quad (8.52)$$

où p et α sont réels, avec $0 < p < 1$, et $\alpha > 1$. Avec un peu d'ingéniosité, on trouve que la solution de (8.52) satisfaisant $F(x \leq 0) = 0$ et $F(x \geq 1) = 1$ est la fonction :

$$F(x) = \sum_{k=0}^{+\infty} (1-p)p^k \Theta(x - \alpha^{-k}) , \quad (8.53)$$

montrant que la variable aléatoire considérée X prend les valeurs α^{-k} , $k \in \mathbb{N}$ et que :

$$\text{Prob}[X = \alpha^{-k}] = (1-p)p^k . \quad (8.54)$$

On vérifie en outre que la somme des sauts, $(1-p)(1+p+p^2+\dots)$, est bien égale à 1.

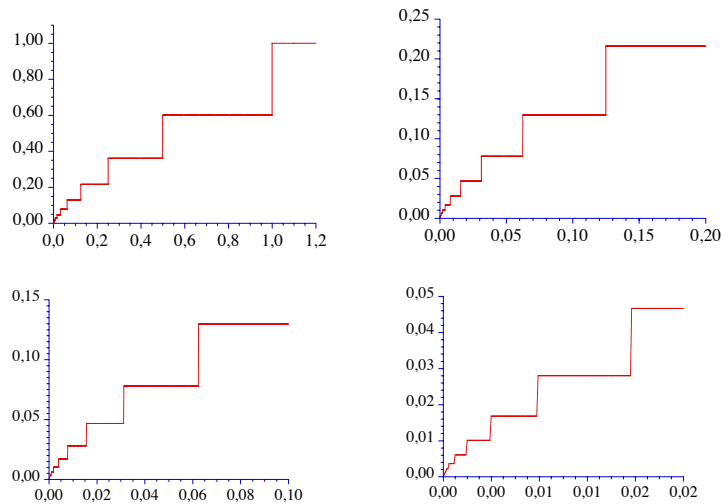


Figure 8.7: Zooms successifs sur le graphe de la fonction (8.53) (avec $q = 0.6$, $\alpha = 2$), révélant son caractère quelque peu diabolique...

Le graphe de la fonction (8.53) révèle toutes les curiosités de cette dernière : quand on fait un *zoom*, on retrouve une figure ayant le même aspect. Toutefois, le *zoom* sur une marche donnée révèle qu'elle ne se décompose pas elle-même en un escalier : l'invariance d'échelle apparente a ses limites. En tout cas, cet escalier du diable n'est pas assez diabolique pour être une composante singulière continue. La fonction de répartition (8.53) est en fait du genre composante atomique, puisque les sauts forment un ensemble *dénombrable* ! De fait, la densité (purement atomique) s'écrit :

$$p(x) = (1-p) \sum_{k=0}^{+\infty} p^k \delta(x - \alpha^{-k}) . \quad (8.55)$$

La densification des points de discontinuité quand on se dirige vers $x = 0$ ne change rien au fait que la dérivée de $F(x)$ est presque partout nulle.

Note

Pour les esprits curieux, quelques indications sur les méthodes permettant de trouver la solution (8.53). Comme on sait ce que vaut F en dehors de $[0, 1]$ (0 à gauche, 1 à droite), il suffit de considérer des valeurs $x \in [0, 1]$.

Avec de la réflexion, la solution se trouve à vue, sans calculs. En effet, si x est choisi de sorte que $\alpha x > 1$, alors l'équation (8.52) s'écrit explicitement $F(x) = p \times 1$, puisque $\alpha x > 1$. On en déduit $F(x) = p \forall x \in [\frac{1}{\alpha}, 1]$.

Examinons maintenant $x < \frac{1}{\alpha}$; si on se restreint en fait à l'intervalle $[\frac{1}{\alpha^2}, \frac{1}{\alpha}]$, l'équation donne $F(x) = p \times p = p^2$, puisque αx étant entre $\frac{1}{\alpha}$ et 1, on sait de ce qui précède que $F(\alpha x)$ vaut alors p . Et ainsi de suite : F est la fonction en escalier montant de p^{k+1} à p^k au point $x = \alpha^{-k}$ et s'écrit donc :

$$F(x) = \sum_{k=0}^{+\infty} (p^k - p^{k+1}) \Theta(x - \alpha^{-k}) \tag{8.56}$$

C'est bien la fonction trouvée en (8.53).

Donnons d'autres façons de s'y prendre, constituant d'ailleurs un petit exercice à propos des transformées de Fourier et Laplace.

Sans pousser le raisonnement jusqu'au bout comme ci-dessus, on remarque que, puisque $F(x \geq 1) = 1$, l'équation permet d'emblée de trouver les valeurs de $F(x)$ aux points de la forme α^{-k} . En effet, l'équation dit que $F(\frac{1}{\alpha}) = p \times F(x = 1) = p$, $F(\frac{1}{\alpha^2}) = pF(\frac{1}{\alpha}) = p^2$, et ainsi de suite. Ceci suggère que la solution doit pouvoir s'écrire $F(x) = x^\lambda$ où toute la question est de trouver l'exposant λ . Si on reporte cette forme dans (8.52), on obtient :

$$1 = p\alpha^\lambda, \tag{8.57}$$

dont on tire naïvement $\lambda = -\frac{\ln p}{\ln \alpha}$; ceci donne bien une sorte d'enveloppe pour $F(x)$, avec les valeurs p^k aux points $x_k = \alpha^{-k}$. En réalité, l'équation (8.57) a une infinité de solutions ! En effet, tous les exposants λ_n de la forme :

$$\lambda_n = \frac{2in\pi - \ln p}{\ln \alpha} \tag{8.58}$$

conviennent tout autant. Comme l'équation est linéaire, toute combinaison linéaire de toutes les solutions correspondantes doit aussi être essayée. À ce stade, on peut affirmer que $F(x)$ est une série de la forme :

$$F(x) = \sum_{n \in \mathbb{Z}} F_n x^{\frac{2in\pi - \ln p}{\ln \alpha}} = x^{-\frac{\ln p}{\ln \alpha}} \sum_{n \in \mathbb{Z}} F_n x^{\frac{2in\pi}{\ln \alpha}}. \tag{8.59}$$

On note que la dernière somme à droite est une série de Fourier par rapport la variable $y = 2\pi \frac{\ln x}{\ln \alpha}$, mais il faut se souvenir que toutes les écritures doivent être comprises dans l'intervalle $[0, 1]$ pour x . Plutôt que de chercher à obtenir directement les coefficients de Fourier F_n , montrons que la fonction obtenue d'emblée en (8.55) peut effectivement se mettre sous la forme d'un tel développement.

L'expression (8.55) peut s'écrire :

$$F(x) = p^{1+E[-\frac{\ln x}{\ln \alpha}]}, \tag{8.60}$$

où $E[x]$ est la fonction partie entière. Maintenant, la fonction $E(y) - y$ est une fonction 1-périodique, donc toute quantité de la forme $r^{E(y)-y}$ ($0 < r < 1$) est une fonction de même période, dont la série de Fourier se trouve facilement :

$$r^{E(y)-y} = -\frac{1-r}{r} \sum_{k \in \mathbb{Z}} \frac{e^{2ik\pi y}}{\ln r + 2ik\pi}. \tag{8.61}$$

Ce dernier résultat peut d'ailleurs s'obtenir directement en calculant d'abord la transformée de Laplace de $r^{E(y)}$:

$$\mathcal{L} [r^{E(y)}] (z) = \sum_{k=0}^{+\infty} \int_k^{k+1} r^k e^{-zy} dy = \frac{1}{z} \frac{1 - e^{-z}}{1 - re^{-z}} \tag{8.62}$$

puis en effectuant la transformation de Laplace inverse ; par résidus³⁶ on obtient :

$$r^{E(y)} = \sum_{k=-\infty}^{+\infty} \frac{1 - e^{-z_k}}{z_k \times 1} e^{z_k t} = \sum_{k=-\infty}^{+\infty} \frac{1 - \frac{1}{r}}{\ln r + 2ik\pi} e^{(\ln r + 2ik\pi)y} = \frac{r-1}{r} r^y \sum_{k=-\infty}^{+\infty} \frac{e^{2ik\pi y}}{\ln r + 2ik\pi} \tag{8.63}$$

³⁶ z_k est l'un des zéros de $1 - re^{-z}$

d'où :

$$r^{E(y)-y} = \frac{r-1}{r} \sum_{k=-\infty}^{+\infty} \frac{e^{2ik\pi y}}{\ln r + 2ik\pi} . \quad (8.64)$$

En particulier ($r \rightarrow p$, $y \rightarrow -\frac{\ln x}{\ln \alpha}$) :

$$F(x) = p^{1+E[-\frac{\ln x}{\ln \alpha}]} = p p^{-\frac{\ln x}{\ln \alpha}} \frac{p-1}{p} \sum_{k=-\infty}^{+\infty} \frac{e^{2ik\pi - \frac{\ln x}{\ln \alpha}}}{\ln p + 2ik\pi} = (p-1)p^{-\frac{\ln x}{\ln \alpha}} \sum_{k=-\infty}^{+\infty} \frac{e^{-2ik\pi \frac{\ln x}{\ln \alpha}}}{\ln p + 2ik\pi} . \quad (8.65)$$

Comme $p^{-\frac{\ln x}{\ln \alpha}} = e^{-\frac{\ln x}{\ln \alpha} \ln p} = e^{-\frac{\ln p}{\ln \alpha} \ln x} = x^{-\frac{\ln p}{\ln \alpha}}$, (8.65) s'écrit aussi :

$$F(x) = x^{-\frac{\ln p}{\ln \alpha}} \sum_{k=-\infty}^{+\infty} \frac{(p-1)}{\ln p + 2ik\pi} e^{-2ik\pi \frac{\ln x}{\ln \alpha}} = x^{-\frac{\ln p}{\ln \alpha}} \sum_{n \in \mathbb{N}} \frac{(p-1)}{\ln p - 2in\pi} x^{\frac{2in\pi}{\ln \alpha}} . \quad (8.66)$$

Cette expression est bien de la forme (8.59), et on lit les coefficients F_n par identification.

8.5 Espérances mathématiques (moyennes)

Dès qu'une description probabiliste est nécessaire (parce que les variables d'intérêt sont aléatoires), la caractérisation maximale possible est la connaissance d'une loi de répartition F ou, avec les hypothèses désormais admises, de sa dérivée $F'(x) = p(x)$ au sens large précisé ci-dessus. L'une ou l'autre de ces fonctions étant connues, il est possible de calculer *toutes* les valeurs moyennes relatives à la variable aléatoire X .

La quantité la plus naturelle pour quantifier une variable aléatoire, et aussi la plus simple, est la valeur moyenne de l'aléatoire X elle-même. Il est facile de la construire intuitivement en se disant qu'une valeur possible x_n est d'autant plus importante que sa probabilité p_n est élevée. La valeur moyenne de X , notée³⁷ $\langle X \rangle$ émerge donc naturellement, pour une variable aléatoire discrète, comme la somme pondérée :

$$\langle X \rangle = \sum_n p_n x_n , \quad (8.67)$$

qui se lit aussi :

$$\langle X \rangle = \sum_n x_n \times \text{Prob}[x = x_n] . \quad (8.68)$$

Finalement, se référant à un grand nombre N d'expériences et aux fréquences statistiques f_n , on voit que la moyenne n'est rien d'autre que la moyenne arithmétique des valeurs trouvées ; si on trouve N_n fois la même valeur x_n , elle apparaît N_n fois quand on effectue la moyenne arithmétique ; notant x_λ la valeur obtenue lors de l'expérience λ (elle coïncide avec l'une des valeurs possibles x_n) :

$$\text{moyenne arithmétique des valeurs trouvées} = \frac{1}{N} \sum_{\lambda=1}^N x_\lambda = \frac{1}{N} \sum_n N_n x_n = \sum_n \frac{N_n}{N} x_n \equiv \sum_n f_n x_n . \quad (8.69)$$

La première somme court sur les résultats des N expériences, les deux suivantes s'effectuent en sommant sur les valeurs différentes de la v.a. Dans la limite $N \rightarrow \infty$, les f_n tendent vers les probabilités et on retrouve (8.67).

Pour une aléatoire purement continue de densité $\rho(x)$, la moyenne se construit suivant le même argument ; on écrit :

$$\langle X \rangle = \int_{-\infty}^{+\infty} x \times \text{Prob}[x < X \leq x + dx] , \quad (8.70)$$

soit, d'après (8.47) :

$$\langle X \rangle = \int_{-\infty}^{+\infty} x \rho(x) dx . \quad (8.71)$$

³⁷Parfois notée $E(X)$.

C'est à nouveau une somme pondérée : on affecte à chaque valeur x un poids égal à la probabilité élémentaire de trouver X entre x et $x + dx$, puis on fait la somme sur toutes les valeurs possibles.

À la réflexion, on voit que la moyenne dans le cas discret, (8.67), peut aussi s'écrire sous forme intégrale ; en effet, en raison de la règle opérationnelle de la fonction de Dirac :

$$\langle X \rangle = \int_{-\infty}^{+\infty} \sum_{n \in \mathbb{N}} p_n \delta(x - x_n) dx = \sum_{n \in \mathbb{N}} p_n x_n . \quad (8.72)$$

En définitive, pour une aléatoire quelconque décrite par une densité du type (8.49), la valeur moyenne est donnée par :

$$\langle X \rangle = \int_{-\infty}^{+\infty} x p(x) dx . \quad (8.73)$$

La moyenne $\langle X \rangle$ s'appelle aussi *espérance mathématique* ; bien évidemment, X et $\langle X \rangle$ ont la même dimension physique.

Remarque

La moyenne peut aussi se calculer avec la fonction de répartition F . À partir de (8.73), utilisant le fait que $p(x) = F'(x)$ et intégrant par parties, il vient :

$$\langle X \rangle = \lim_{A, B \rightarrow +\infty} \left[[xF(x)]_{-B}^A - \int_{-B}^A F(x) dx \right] . \quad (8.74)$$

Le terme tout intégré est nul en $-\infty$ si $F(x)$ se rapproche de 0 plus vite que $\frac{1}{|x|}$ quand $x \rightarrow -\infty$, c'est-à-dire quand $\rho(x)$ tend vers zéro plus vite $\frac{1}{x^2}$. Quoi qu'il en soit, on devine que, sauf cas exceptionnel, la formule (8.73) est plus efficace que (8.74).

Ayant identifié le mode opératoire pour calculer la valeur moyenne de X , on peut maintenant trouver la valeur moyenne de n'importe quelle fonction de X , $f(X)$, définie comme la fonction prenant les valeurs $f(x)$; pour une v.a. discrète :

$$\langle f(X) \rangle = \sum_{n \in \mathbb{N}} p_n f(x_n) . \quad (8.75)$$

Un exemple très important est celui où $x \rightarrow f(x) = x^2$; la moyenne du carré de X dans le cas discret est :

$$\langle X^2 \rangle = \sum_{n \in \mathbb{N}} p_n x_n^2 . \quad (8.76)$$

À nouveau, dans le cas général d'une densité du type (8.49), la moyenne s'obtient comme :

$$\langle X^2 \rangle = \int_{-\infty}^{+\infty} x^2 p(x) dx . \quad (8.77)$$

Il faut noter dès maintenant que $\langle X^2 \rangle \geq \langle X \rangle^2$. En effet, calculons la valeur moyenne de la combinaison³⁸ $(X - \langle X \rangle)^2$. On a :

$$\langle (X - \langle X \rangle)^2 \rangle = \langle X^2 \rangle - 2\langle X \rangle \langle X \rangle + \langle X \rangle^2 = \langle X^2 \rangle - \langle X \rangle^2 . \quad (8.78)$$

Le premier membre est une quantité positive ou nulle : c'est l'intégrale de la fonction positive $(x - \langle X \rangle)^2 p(x)$. D'où :

$$\langle X^2 \rangle - \langle X \rangle^2 \geq 0 \iff \langle X^2 \rangle \geq \langle X \rangle^2 . \quad (8.79)$$

En fait, l'égalité n'est obtenue que si la fonction $(x - \langle X \rangle)^2 p(x)$ est identiquement nulle ; ceci ne survient que dans un cas, celui où X ne prend qu'une seule valeur (et alors c'est forcément sa valeur *moyenne* !), soit quand X est en fait une variable certaine.

³⁸La variable $X - \langle X \rangle$ est appelée variable *centrée*, en raison du fait que sa valeur moyenne est visiblement nulle.

Une variable aléatoire X , par définition, présente une dispersion des valeurs qui “sortent” successivement d’une série d’expériences. Il est très important de pouvoir décrire quantitativement la dispersion de ces valeurs qui, lorsque le nombre d’expériences devient assez grand, se déplacent de façon erratique de part et d’autre de la valeur moyenne et représentent les *fluctuations* de X autour de la moyenne. D’après les observations précédentes, la mesure la plus naturelle de ces fluctuations est l’*écart-type* ΔX , définie comme :

$$\Delta X = \sqrt{\langle X^2 \rangle - \langle X \rangle^2} ; \quad (8.80)$$

la quantité sous le radical est appelée *variance*, ou *écart quadratique moyen* ou simplement *écart quadratique* ; d’après (8.79), c’est bien une quantité positive, nulle dans le cas-limite d’une variable certaine. Des exemples simples montrent ce que l’on peut pressentir : la valeur de l’écart-type est d’autant plus grande que les valeurs de X sont très dispersées, c’est à dire que la distribution densité $p(x)$ est large³⁹ ; *a contrario*, l’écart-type ΔX est nul dans le cas où la variable X est certaine, c’est-à-dire ne fluctue pas du tout. L’écart-type est donc une bonne mesure *globale* des fluctuations ; toutefois, en tant que grandeur intégrale calculée avec $p(x)$, il “efface” forcément certains détails de la densité de probabilité. On retiendra en tout cas l’équivalence :

$$\Delta X = 0 \iff X \text{ est une variable certaine} . \quad (8.81)$$

En terme de distribution de probabilité, le cas certain correspond à $p_n = \delta_{n n_0}$ pour une distribution atomique, à $\delta(x - x_0)$ pour une distribution absolument continue. En toute généralité, la fonction de répartition d’une variable certaine (prenant la valeur x_0 avec probabilité 1) est $F(x) = \Theta(x - x_0)$.

D’une façon générale, on définit les *moments* M_k d’une variable aléatoire⁴⁰ ; ce sont les valeurs moyennes des X^k (k n’est pas forcément entier, ni forcément positif) :

$$M_k \stackrel{\text{déf}}{=} \langle X^k \rangle = \int_{-\infty}^{+\infty} x^k p(x) dx = \sum_n x_n^k p_n + \int_{-\infty}^{+\infty} x^k \rho(x) dx . \quad (8.82)$$

Il est bien clair que rien ne garantit d’avance que les différents moments existent et sont finis ; par exemple, lorsqu’il n’y a pas de composante atomique :

$$M_k = \int_{-\infty}^{+\infty} x^k \rho(x) dx . \quad (8.83)$$

Si ρ est à support borné ne contenant pas l’origine, l’intégrale existe $\forall k$; si l’origine est dans le support de ρ , il faut $k \geq -1$. Quand le support s’étend jusqu’à l’infini, tout dépend de la nature de la décroissance de ρ à l’infini : pour que le moment M_k existe, il faut et il suffit que ρ décroisse plus vite que $|x|^{-(k+1)}$. Notons que le moment d’ordre zéro existe toujours, et vaut 1 : ce n’est rien d’autre que l’intégrale et/ou la somme de la densité $p(x)$: $\int p = 1 \forall p$.

Il est bien clair que la connaissance de $p(x)$ (ou de $F(x)$) constitue l’information maximale escomptable dans un cadre probabiliste. En pratique, on est souvent bien démuni, ne connaissant d’expérience qu’un petit nombre de moments – sauf bien sûr si on est capable d’échafauder un modèle permettant de trouver théoriquement les fonctions $F(x)$ ou $p(x)$. Dans le premier cas, se pose alors parfois le problème difficile de reconstruire la fonction inconnue à partir des moments obtenus d’une façon ou d’une autre. Il existe un théorème important, le théorème de Marcinkiewicz, que l’on énoncera plus précisément par la suite (section 8.7), et qui érige quelques garde-fous évitant les canulars (par exemple des distributions contenant des probabilités... négatives).

8.6 Lois de distribution courantes

8.6.1 Loi binomiale

La loi binomiale est une loi discrète qui survient quand on considère une population dont les individus, *agissant indépendamment les uns des autres*, peuvent être chacun dans deux états possibles⁴¹, avec les probabilités p et

³⁹ On dit aussi *plate*, mais c’est un abus de langage : $p(x)$ doit être intégrable et donc décroît à l’infini plus vite que $\frac{1}{x}$.

⁴⁰ Avec cette définition, l’écart quadratique est $\Delta X^2 = M_2 - M_1^2$.

⁴¹ Une variable aléatoire qui peut prendre deux valeurs (comme pile et face, 0 et 1, etc.) est aussi appelée *variable de Bernoulli*.

$q = 1 - p$. Par exemple, il peut s'agir d'une source de N_0 noyaux radioactifs au départ qui, à un certain instant t (fixé pour le moment), peuvent être soit excités ("vivants", "non désintégrés"), soit à l'état fondamental ("morts", "désintégrés"). La variable aléatoire est ici N , nombre de noyaux survivants à l'instant considéré, chaque noyau vivant sa vie indépendamment de tous les autres (autre exemple : les boules blanches et noires que l'on tire dans un sac, les différentes suites possibles de pile et face quand on jette N_0 fois la pièce).

La question centrale est ici de trouver p_n :

$$\text{Probabilité d'avoir } n \text{ survivants} \equiv \text{Prob}[N = n] \stackrel{\text{déf}}{=} p_n . \quad (8.84)$$

Le nombre de façons distinctes d'avoir n individus vivants est $C_{N_0}^n$. En effet, numérotons les individus de 1 à N_0 et disposons-les sur une ligne en mettant les n vivants à gauche, les $N_0 - n$ morts à droite. Il est clair que l'on peut effectuer n'importe quelle permutation parmi les vivants (il y a $n!$ permutations), ou parmi les morts ($(N_0 - n)!$ permutations), sans modifier la nature de l'événement " n vivants et $N - n$ morts" ; par ailleurs, la numérotation des individus est arbitraire et il y a $N_0!$ numérotations distinctes. Au total, le nombre de configurations contenant n morts et $N_0 - n$ vivants est égal à $\frac{N_0!}{n!(N_0 - n)!} = C_{N_0}^n$.

Chaque répartition exclut toute autre ; une répartition donnée a la probabilité $p^n(1 - p)^{N_0 - n}$. La probabilité p_n d'avoir n noyaux vivants est donc la somme des $C_{N_0}^n$ probabilités $p^n(1 - p)^{N_0 - n}$:

$$p_n = C_{N_0}^n p^n (1 - p)^{N_0 - n} \quad (0 \leq n \leq N_0) . \quad (8.85)$$

On vérifie sans peine que la somme des p_n vaut 1 puisque $[p + (1 - p)]^{N_0} = 1$. La distribution n'est pas symétrique dans la transformation $n \rightarrow N_0 - n$, sauf si $p = \frac{1}{2}$, évidemment.

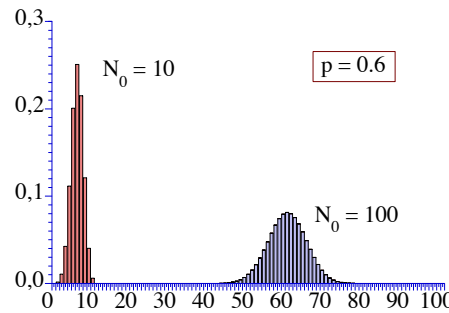


Figure 8.8: Distribution des probabilités binomiales (8.85) p_n en fonction de n , pour $p = 0.6$ et $N_0 = 10, 100$. Les probabilités sont strictement nulles si n est en dehors de $[0, N_0]$.

À p donné, la valeur moyenne du nombre de survivants est :

$$\langle N \rangle = \sum_{n=0}^{N_0} n p^n (1 - p)^{N_0 - n} C_{N_0}^n = \left[\frac{d}{dx} [x^n p^n (1 - p)^{N_0 - n} C_{N_0}^n] \right]_{x=1} = \left[\frac{d}{dx} [xp + (1 - p)]^{N_0} \right]_{x=1} = N_0 p . \quad (8.86)$$

Par définition, la moyenne de $\langle N^2 \rangle$ est :

$$\langle N^2 \rangle = \sum_{n=0}^{N_0} n^2 p^n (1 - p)^{N_0 - n} C_{N_0}^n \quad (8.87)$$

et se calcule facilement ; au total, on trouve que l'écart quadratique est :

$$\Delta N^2 = \langle N^2 \rangle - \langle N \rangle^2 = N_0 p (1 - p) ; \quad (8.88)$$

Ceci permet d'écrire l'expression des fluctuations relatives :

$$\frac{\Delta N}{\langle N \rangle} = \frac{1}{\sqrt{N_0}} \sqrt{\frac{1}{p} - p} . \quad (8.89)$$

Il faut noter l'occurrence du facteur $N_0^{-1/2}$, qui reviendra souvent par la suite dans la considérations des fluctuations.

S'agissant de la source radioactive, la probabilité pour un noyau d'être encore vivant à l'instant t est $p(t) = e^{-\lambda t}$, où λ^{-1} est la durée de vie moyenne d'un noyau à l'état actif ; cette probabilité décroît toujours (plus on attend, plus on a de chances de trouver qu'un noyau donné est mort), et tend évidemment vers zéro. La dépendance en t de p se reporte sur toutes les probabilités, $p_n(t)$, donc sur toutes les espérances mathématiques. Ainsi, la moyenne de N devient une fonction du temps et (8.86) donne $\langle N \rangle(t) = N_0 e^{-\lambda t}$, expression usuelle de la loi de décroissance radioactive (qui porte en fait sur une *moyenne*, comme on omet trop souvent de le mentionner⁴²). Aux premiers instants ($t \ll \frac{1}{\lambda}$, $p \sim 1$ et $q \equiv 1 - p \sim 0$), la distribution est ramassée vers N_0 (d'ailleurs, au départ ($t = 0$) N est une variable certaine : on se donne comme information que la source contient N_0 noyaux actifs). À l'instant⁴³ T où $p = q$, soit $e^{-\lambda T} = \frac{1}{2}$, ou encore $T = \frac{\ln 2}{\lambda} \simeq \frac{0.693}{\lambda}$, la distribution est symétrique autour de $\frac{N_0}{2}$, il y a – en termes de probabilités – autant de morts que de vivants. Enfin, pour $t \gg \lambda^{-1}$, $p \simeq 0$ et $q \simeq 1$, toutes les probabilités sont exponentiellement petites, sauf $p_0 \lesssim 1$. Ceci ne veut pas dire qu'une source radioactive devenue vieille est inoffensive : selon (8.89), on a alors $\frac{\Delta N(t)}{\langle N \rangle(t)} \simeq \frac{1}{\sqrt{N_0}} e^{+\frac{\lambda t}{2}}$, ce qui signifie que les fluctuations relatives $\frac{\Delta N}{\langle N \rangle} \dots$ divergent exponentiellement quand $t \rightarrow +\infty$! Ainsi, même une vieille source – dont l'activité *moyenne* est devenue faible en raison de son grand âge – est capable d'émettre des *bouffées* dangereuses ; cette observation élémentaire est à la base de l'éternel débat : peut-on tolérer des faibles doses (moyennes) ou au contraire la seule dose acceptable est-elle la dose *zéro* ? – qui n'existe nulle part d'ailleurs.

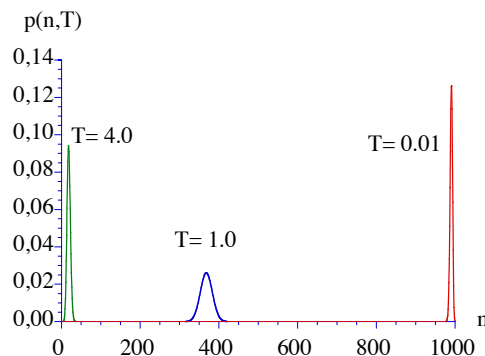


Figure 8.9: Distribution des noyaux survivants à trois âges d'une source contenant initialement $N_0 = 1000$ noyaux actifs. Chaque distribution est repérée par la valeur du paramètre $T = \lambda t$ (à droite : source jeune ; à gauche : source vieille).

Dynamique des probabilités pour le déclin radioactif

La présentation précédente peut être reprise d'un autre point de vue, permettant au passage d'introduire la notion de *processus stochastique*. Sans vouloir donner une définition générale, on peut dire qu'un processus stochastique est la succession des valeurs d'une variable aléatoire lorsqu'un certain paramètre varie, le plus souvent il s'agit du temps ou d'une coordonnée spatiale⁴⁴. Ici, la v.a. est N , nombre de

⁴²Ce fait est en pratique sans importance : pour une source contenant un grand nombre de noyaux, et en comptant les désintégrations pendant des intervalles de temps grands par rapport à $\frac{1}{N\lambda}$, les fluctuations sont très faibles (le *bruit* est très petit). Il n'empêche que, conceptuellement, la différence est de taille. Pouvoir confondre (au moins dans un premier temps !) valeur moyenne et variable aléatoire résulte ici, comme pour tout système macroscopique, de l'énormité du nombre d'Avogadro.

⁴³ T est appelé (improprement) *période*. En ordre de grandeur, $T \sim \lambda^{-1}$.

⁴⁴mais ce peut être aussi l'énergie, quand un système se promène entre des états caractérisés par leur énergie.

survivants à l'instant t – qui prend les valeurs entières $0 \leq n \leq N_0$ –, dont les probabilités sont les p_n données en (8.85), et qui deviennent des fonctions du temps par l'injection de la loi exponentielle pour p : $p = e^{-\lambda t}$.

On peut obtenir directement les $p_n(t)$ sans effectuer le petit dénombrement qui a conduit à (8.85), en écrivant, puis en résolvant, l'équation d'évolution des $p_n(t)$. Celle-ci s'obtient par un raisonnement de *bilan*, en dressant l'inventaire des événements possibles entre deux instants t et $t + \delta t$, compte tenu du seul ingrédient physique : chaque noyau a une probabilité $\lambda \delta t$ de se désexciter entre t et $t + \delta t$.

Soit donc la probabilité $p_n(t)$ de trouver n noyaux actifs à l'instant t , sachant qu'il y en avait N_0 à l'instant initial ($t = 0$). En appliquant les règles usuelles :

- événements indépendants \leftrightarrow multiplication des probabilités $\iff \text{Prob}[A \text{ et } B] = P(A)P(B)$
- événements exclusifs \leftrightarrow addition des probabilités $\iff \text{Prob}[A \text{ ou } B] = P(A) + P(B)$,

et en faisant l'inventaire des processus élémentaires pouvant survenir entre t et $t + \delta t$ ⁴⁵, on obtient le système d'équations de bilan :

$$p_n(t + \delta t) = (1 - \lambda \delta t)^n p_n(t) + (n + 1) \lambda \delta t p_{n+1}(t) + \mathcal{O}((\lambda \delta t)^k \geq 2) \quad (n = 0, 1, \dots, N_0) . \quad (8.90)$$

Le premier terme provient de la possibilité : aucun des n noyaux présents à l'instant t n'a changé d'état ; le second terme signifie qu'il y avait $n + 1$ noyaux à l'instant t et que l'un de ces $n + 1$ noyaux a basculé. Le symbole noté \mathcal{O} rassemble des termes d'ordre supérieur en δt (par exemple : plusieurs noyaux se sont désexcités durant δt). Au même ordre en δt , le système (8.90) est :

$$p_n(t + \delta t) = (1 - n \lambda \delta t) p_n(t) + (n + 1) \lambda \delta t p_{n+1}(t) + \mathcal{O}((\lambda \delta t)^k \geq 2) \quad (n = 0, 1, \dots, N_0) . \quad (8.91)$$

Après division par δt et passage à la limite $\delta t \rightarrow 0$, on obtient le système⁴⁶ différentiel suivant, de dimension $N_0 + 1$:

$$\frac{dp_n}{dt} = (n + 1) \lambda p_{n+1}(t) - n \lambda p_n(t) \quad (n = 0, 1, 2, \dots, N_0; p_{N_0+1}(t) \equiv 0 \quad \forall t) . \quad (8.92)$$

Avec la condition initiale $p_n(t = 0) = \delta_{n N_0}$, on vérifie sans peine que la solution est de fait la distribution donnée en (8.85).

‡ Appendice

Une façon commode de résoudre le système (8.92) est la suivante. On définit la fonction :

$$f(t; s) = \sum_{n=0}^{N_0} s^n p_n(t) , \quad (8.93)$$

que l'on peut appeler⁴⁷ *fonction génératrice* des probabilités⁴⁸ puisque, au facteur $n!$ près, les dérivées successives de f en $s = 0$ sont précisément ces probabilités ($p_n = \frac{1}{n!} \left[\frac{\partial^n}{\partial s^n} f(t; s) \right]_{s=0}$). À partir de (8.92), on voit de suite que f satisfait l'équation aux dérivées partielles :

$$\frac{\partial f}{\partial t} = \lambda(1 - s) \frac{\partial f}{\partial s} , \quad (8.94)$$

avec la condition aux limites $f(t = 0; s) = s^{N_0}$ puisque $p_n(t = 0) = \delta_{n N_0}$. L'inspection⁴⁹ de cette équation montre que sa solution est de la forme $F[(1 - s)a(t)]$, où F est une fonction pour l'instant quelconque, et à condition que $a(t)$ satisfasse l'équation $\dot{a} = -\lambda a \iff a(t) = e^{-\lambda t}$, d'où résulte :

$$f(t; s) = F[(1 - s)e^{-\lambda t}] . \quad (8.95)$$

⁴⁵Par exemple, la probabilité qu'aucun des n noyaux vivants à t ne bascule entre t et $t + \delta t$ est $(1 - \lambda \delta t)^n$.

⁴⁶Il est facile de vérifier que la somme des premiers membres est bien égale à zéro.

⁴⁷En posant $s = e^{i\omega}$, on voit que f n'est autre que la fonction caractéristique introduite plus loin (voir section 8.7) ; c'est la transformée de Fourier discrète des p_n .

⁴⁸Ici, le nombre de probabilités p_n est fini (il est égal à $N_0 + 1$), de sorte que f est un polynôme.

⁴⁹Une procédure moins intuitive et plus systématique consiste à effectuer une transformation de Laplace de l'équation (8.94) ; en posant $F(z; s) = \int_0^{+\infty} e^{-zt} f(t; s) dt$, on trouve que $F(z; s)$ satisfait une équation différentielle ordinaire. L'intégration de celle-ci, suivie de la transformation de Laplace inverse, redonne l'expression (8.96).

La condition aux limites impose $F(1-s) = s^{N_0} \iff F(S) = (1-S)^{N_0}$ et finalement :

$$f(t; s) = [1 - (1-s)e^{-\lambda t}]^{N_0} = [(1 - e^{-\lambda t}) + se^{-\lambda t}]^{N_0} . \quad (8.96)$$

Le développement de cette expression en puissances de s (suivant la formule du binôme), et l'identification avec le développement (8.93), reproduit l'expression (8.85) des probabilités, comme il se doit.

Notons que si on affecte la valeur 1 à l'état vivant, et zéro à l'état mort, N est égal à la somme de ces N_0 variables de Bernoulli. On peut donc dire généralement que la loi binomiale est de fait la loi de probabilité d'une somme de variables de Bernoulli (indépendantes). C'est donc aussi la distribution du gain d'un joueur au jeu de pile et face (pour une pièce parfaite, $p = \frac{1}{2}$).

Conformément à (8.40), la fonction de répartition binomiale est :

$$F(x) = \sum_{n=0}^{N_0} C_{N_0}^n p^n (1-p)^{N_0-n} \Theta(x-n) . \quad (8.97)$$

8.6.2 Loi de Poisson

La loi de Poisson est relative à une variable aléatoire N prenant ses valeurs dans \mathbb{N} et s'écrit par définition⁵⁰ :

$$\text{Prob}[N = n] \stackrel{\text{déf}}{=} p_n = e^{-\lambda} \frac{\lambda^n}{n!} \quad (n \in \mathbb{N}, \lambda \in \mathbb{R}_+) . \quad (8.98)$$

Comme on va le voir, cette loi, qui ne dépend que du seul paramètre λ , décrit notamment, dans une certaine limite qui va être précisée, ce que l'on appelle les *phénomènes d'attente*. Afin de préciser ceci, on se place dans la situation concrète suivante.

Soit un grand nombre d'individus (clients), \mathcal{N} , souhaitant se rendre en un lieu donné (un guichet) en un temps total donné, T . Chaque individu agit pour son propre compte sans se soucier des autres. La question que l'on se pose est la suivante : quelle est la probabilité p_n pour que n individus arrivent pendant le même intervalle de temps fixe, donné, noté Δt ?

Une donnée fondamentale est la probabilité p pour qu'un client arrive pendant Δt . L'hypothèse la plus naturelle (en tout cas la plus simple) consiste à poser :

$$p = \frac{\Delta t}{T} . \quad (8.99)$$

C'est une façon de traduire le *hasard* complet, au sens où aucun instant n'est privilégié entre 0 et T , que tous les temps se valent.

Pendant l'intervalle de temps Δt , un individu arrive au guichet avec une certaine probabilité p , ou n'y arrive pas, avec la probabilité $1-p$. Pour une population de \mathcal{N} clients potentiels agissant indépendamment les uns des autres, la probabilité qu'il arrive n personnes est donc la même probabilité que celle d'avoir n fois pile (par exemple) avec \mathcal{N} lancers de la pièce : c'est donc une probabilité binomiale, donnée par (8.85) et qui ici s'écrit ($N_0 \rightarrow \mathcal{N}$) :

$$\text{Prob}[N = n] \stackrel{\text{déf}}{=} p_n = C_{\mathcal{N}}^n p^n (1-p)^{\mathcal{N}-n} . \quad (8.100)$$

Ceci fournit réponse à la question posée, mais il est intéressant de l'analyser dans une certaine limite, où à la fois \mathcal{N} est très grand et où l'intervalle de temps de comptage Δt est très petit devant la durée totale d'observation T , ce qui constitue typiquement une analyse à grande échelle du problème, dans un modèle de quasi-*continuum*.

Il est clair que les deux limites $\mathcal{N} \gg 1$ et $\Delta t \ll T$ ne doivent pas être prises indépendamment si l'on souhaite une réponse non triviale : par exemple, si on fait tendre formellement T vers $+\infty$ en gardant \mathcal{N} fixe,

⁵⁰On vérifie immédiatement que $\sum_{n \in \mathbb{N}} p_n = 1$.

la réponse immédiate : il ne se passe rien, $p = 0$ par (8.99), et p_n est nulle, sauf si $n = 0$; c'est d'ailleurs bien ce que donne l'expression (8.100) si on y fait brutalement $p = 0$. À l'inverse, si T étant fixé, \mathcal{N} augmente indéfiniment, toutes les probabilités sont nulles, sauf $p_{\mathcal{N}}$, qui vaut 1, ce que donne aussi (8.100) si on y pose $p = 1$.

La façon la plus simple de lier les deux limites est de dire que le rapport $\frac{\mathcal{N}}{T}$ garde une valeur fixe :

$$\mathcal{N} \rightarrow +\infty, \quad T \rightarrow +\infty, \quad \frac{\mathcal{N}}{T} = C^{\text{ste}} = \gamma. \quad (8.101)$$

En fait, c'est aussi la façon la plus naturelle de poser le problème en termes physiques ; la grandeur constante γ , inverse d'un temps, n'est pas un nouveau paramètre du problème et a un sens physique très clair : en tant que limite du rapport $\frac{\mathcal{N}}{T}$, c'est le nombre moyen d'individus arrivés par unité de temps dans la limite d'une population infinie et pour un phénomène durant une éternité. γ représente le taux d'arrivée des clients ; c'est une grandeur susceptible d'une mesure expérimentale révélant, pour le phénomène considéré, si elle est bien définie, et si oui combien elle vaut. Un tel test statistique permet de savoir dans quelle mesure un phénomène *réel* relève ou non de la loi de Poisson.

Avec ces hypothèses, reportant $T = \mathcal{N}\gamma^{-1}$ dans (8.99) :

$$p = \gamma \frac{\Delta t}{\mathcal{N}} \equiv \frac{a}{\mathcal{N}}, \quad a = \gamma \Delta t. \quad (8.102)$$

La prescription (8.101), définissant sans ambiguïté le processus de limite, est une nécessité formelle imposée par le désir d'avoir une réponse non triviale dans le cas $\mathcal{N} \gg 1$. On va voir que, de fait, les probabilités binomiales p_n (8.100) admettent alors une limite, et que cette limite est la loi de Poisson donnée par (8.98).

Ceci étant, il reste un peu d'algèbre à effectuer pour arriver à (8.98). Écrivons l'expression explicite de p_n à partir de (8.100) :

$$p_n = C_{\mathcal{N}}^n p^n (1-p)^{\mathcal{N}-n} = \frac{\mathcal{N}(\mathcal{N}-1)\dots(\mathcal{N}-n+1)}{n!} \left(\frac{a}{\mathcal{N}}\right)^n \left(1 - \frac{a}{\mathcal{N}}\right)^{\mathcal{N}-n}, \quad (8.103)$$

qui se réécrit comme suit :

$$p_n = \frac{1}{n!} \frac{\mathcal{N}(\mathcal{N}-1)\dots(\mathcal{N}-n+1)}{\mathcal{N}^n} a^n \left(1 - \frac{a}{\mathcal{N}}\right)^{\mathcal{N}-n}. \quad (8.104)$$

À n fixé, la grande fraction a pour limite 1 dans la limite $\mathcal{N} \rightarrow \infty$; toujours à n fixé, on a :

$$\lim_{\mathcal{N} \rightarrow +\infty} \left(1 - \frac{a}{\mathcal{N}}\right)^{\mathcal{N}-n} = \lim_{\mathcal{N} \rightarrow +\infty} e^{(\mathcal{N}-n)\ln(1-\frac{a}{\mathcal{N}})} = e^{-a}. \quad (8.105)$$

D'où finalement, dans la limite considérée :

$$\lim p_n = \frac{a^n}{n!} e^{-a}; \quad (8.106)$$

c'est bien la loi de Poisson annoncée, (8.98), avec $\lambda = a$. En résumé, si N dénote toujours la variable aléatoire "nombre d'individus arrivés pendant l'intervalle de référence Δt ", on a :

$$\text{Prob}[N = n] = \frac{(\gamma \Delta t)^n}{n!} e^{-\gamma \Delta t}, \quad (8.107)$$

dans la limite d'une population \mathcal{N} infinie et d'une durée T infinie, le rapport $\frac{\mathcal{N}}{T}$ étant fixe et égal à γ . La discussion détaillée ci-dessus montre que c'est le jeu $T \rightarrow +\infty$, $\mathcal{N} \rightarrow +\infty$, $\mathcal{N}/T = C^{\text{ste}}$ qui donne une limite non-triviale, bien que $p \rightarrow 0$, conformément à (8.102).

La loi de Poisson se rencontre également à propos d'un gaz de particules (supposées classiques) confiné dans un récipient de volume V , quand on veut calculer la probabilité qu'un petit volume δV contienne un nombre donné de particules. Pour une seule particule, son appartenance à δV est une variable binaire, 0 si

la particule est dehors de V , 1 si elle est dedans par exemple, avec les probabilités $1 - p$ et p : il s'agit donc d'une variable de Bernoulli ; avec une hypothèse de type ergodique, on a $p = \frac{\delta V}{V}$. Pour des particules sans interactions, la probabilité pour une particule d'être dans δV reste égale à p en présence des autres puisqu'elles ne se voient pas les unes les autres. Les particules étant indépendantes, leur présence simultanée dans δV est un événement composite formé avec des événements indépendants ; la probabilité correspondante, p_n , est donc le produit des probabilités, étant entendu qu'il y a $C_{\mathcal{N}}^n$ façons de répartir n particules parmi \mathcal{N} dans δV : p_n est donc à nouveau $C_{\mathcal{N}}^n p^n (1 - p)^{\mathcal{N} - n}$.

Le récipient étant fermé, le nombre de particules \mathcal{N} est donné une fois pour toutes. Si $\delta V \ll V$, $p \ll 1$, mais avec δV "infinitésimal", on peut aussi assurer $\mathcal{N}p \lesssim 1$ bien que $\mathcal{N} \gg 1$: on est ainsi dans le cas où la limite ci-dessus s'applique, $\mathcal{N}p$ étant un nombre d'ordre 1 ($=a$) ; on retrouve ainsi la loi de Poisson :

$$\lim p_n = \frac{\bar{\nu}^n}{n!} e^{-\bar{\nu}}, \quad (8.108)$$

où $\bar{\nu} = \mathcal{N}p$ s'interprète ici comme le nombre moyen de particules dans le volume δV .

La loi de Poisson est aussi celle qui régit des phénomènes de remplissage ; soit à remplir un segment de longueur L avec des points jetés au hasard. Il suffit, dans ce qui précède, de remplacer *individu* par *point*, T par L , le rapport noté γ est alors égal à $\frac{\mathcal{N}}{L}$: c'est la densité de points supposée constante quand on fait tendre le nombre de points \mathcal{N} et la longueur L du segment vers l'infini.

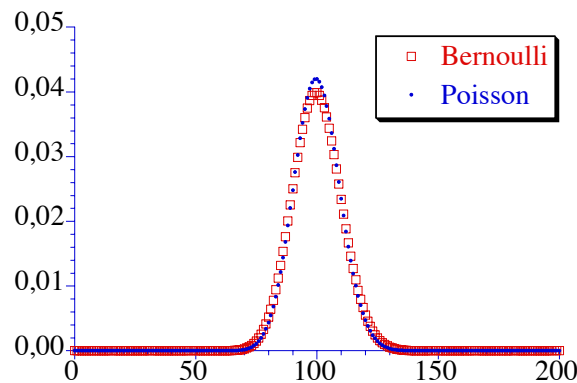


Figure 8.10: Comparaison des distributions de Bernoulli et de Poisson. La loi de Bernoulli est tracée avec $p = 0.1$ et $\mathcal{N} = 1000$ (noté N_0 dans (8.85)). Conformément à l'analyse conduisant à (8.109), la loi de Poisson correspondante a un paramètre a égal à $p\mathcal{N}$ soit $a = 100$.

La loi de Poisson porte aussi le nom de *Loi des petites probabilités* puisque, selon (8.102), la probabilité p tend vers zéro par le processus de limite, précisément comme $\frac{1}{\mathcal{N}}$. On retiendra :

$$\lim_{p \rightarrow 0, \mathcal{N} \rightarrow +\infty, \mathcal{N}p = \text{Cste} = a} p_n^{(\text{binomiale})}(\mathcal{N}, p) = p_n^{(\text{Poisson})}(a = \mathcal{N}p) \quad (8.109)$$

où $p_n^{(\text{Poisson})}(a)$ désigne la loi de Poisson de paramère a (noté λ dans la présentation générale, voir (8.98)). La fig.8.10 fournit la comparaison entre les deux lois ; si la proximité des deux lois est évidente à l'œil, l'examen attentif révèle que, loin dans les ailes, les valeurs numériques (certes très petites) des deux lois sont en fait très différentes : ce phénomène est fréquent et se retrouve également lorsque l'on compare une loi et son approximation par sa loi attractive (voir la section 8.8). On désigne souvent par *région d'échelle* la zone centrale où la coïncidence est très satisfaisante.

La distribution de Poisson (8.98) présente des aspects très différents selon la valeur du paramètre $a = \gamma\Delta t$ (voir fig. 8.11, en remplaçant λ par a). Lorsque $a \ll 1$, p_0 est très voisine de 1, cependant que les autres probabilités décroissent très vite avec n . C'est le cas lorsque l'intervalle de comptage des clients, Δt est très petit devant γ^{-1} , intervalle moyen séparant deux arrivées. Pour $a \lesssim 1$, les toutes premières probabilités sont voisines de 1, les autres décroissant assez vite. Pour $a = 1$, on a $p_0 = e^{-1} = p_1$, $p_n = \frac{e^{-1}}{n!}$. Quoi qu'il en soit,

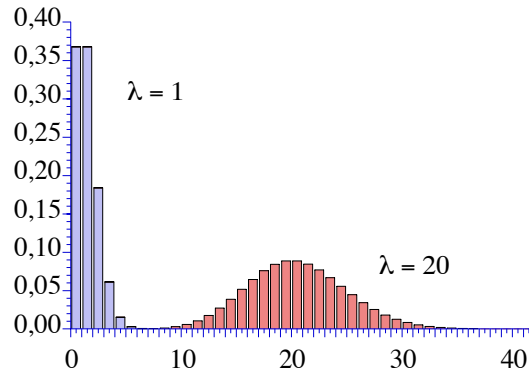


Figure 8.11: Distribution des probabilités de Poisson (8.98) pour $\lambda = 1$ et $\lambda = 20$. Les probabilités sont toutes non-nulles (le support est \mathbb{N}), mais deviennent très vite très petites.

pour $a \sim 1$, ce sont les probabilités avec un petit n qui sont les plus importantes, le maximum de p_n survenant pour $n_{\max} \sim 1$:

$$n_{\max} \sim 1, \quad p_{n_{\max}} \lesssim 1 \quad (8.110)$$

Au contraire, si $a \gg 1$, les probabilités avec un petit n sont très petites, l'exponentielle e^{-a} l'emportant sur a^n tant que $n \lesssim \frac{a}{\ln a}$; par ailleurs, pour $n \rightarrow \infty$, $p_n \rightarrow 0 \forall a$. Il en résulte que p_n passe par un maximum pour un certain n_{\max} qui est grand⁵¹ devant 1. Ce maximum est petit ; en effet, en appliquant la formule de Stirling, on voit que :

$$a \gg 1 \Rightarrow n_{\max} \gg 1, \quad p_{n_{\max}} \simeq \frac{1}{\sqrt{2\pi n_{\max}}} \ll 1 \quad (8.111)$$

Autrement dit, quand $a \gg 1$, la distribution des p_n est relativement “plate”, et présente un faible maximum d'ordre $(n_{\max})^{-\frac{1}{2}}$ (pour $n_{\max} = 100$, $p_{n_{\max}} \simeq 0.040$). Ceci est bien normal : supposer $a \gg 1$, c'est se placer dans la situation où l'intervalle de comptage est long devant γ^{-1} , auquel cas beaucoup d'événements (arrivée de n clients) ont une probabilité sensiblement notablement différente de zéro.

Toutes ces propriétés se confirment précisément en remarquant simplement que, selon (8.98), on a :

$$\frac{p_{n+1}}{p_n} = \frac{a}{n+1} ; \quad (8.112)$$

cette relation montre que $p_{n+1} > p_n$ tant que $n+1 < a$. Si $a \lesssim 1$, la distribution est toujours décroissante ; si $a \gg 1$, les probabilités croissent jusqu'à $n \simeq a \gg 1$, puis décroissent.

Ces constatations se retrouvent également sur la moyenne de N et son écart-type. On a :

$$\langle N \rangle = \sum_{n=0}^{+\infty} n e^{-a} \frac{a^n}{n!} = \sum_{n=1}^{+\infty} e^{-a} \frac{a^n}{(n-1)!} = a ; \quad (8.113)$$

la moyenne de N^2 est :

$$\langle N^2 \rangle = \sum_{n=0}^{+\infty} n^2 e^{-a} \frac{a^n}{n!} = \sum_{n=1}^{+\infty} [n(n-1) + n] e^{-a} \frac{a^n}{n!} = a^2 + a \equiv \langle N \rangle^2 + a , \quad (8.114)$$

d'où l'écart-type :

$$\Delta N = \sqrt{\langle N^2 \rangle - \langle N \rangle^2} = \sqrt{a} \quad (8.115)$$

⁵¹On va voir que $n_{\max} \simeq a$ (voir (8.112)).

et les fluctuations relatives :

$$\frac{\Delta N}{\langle N \rangle} = \frac{1}{\sqrt{a}} . \quad (8.116)$$

Si $a \lesssim 1$, la moyenne de N est près de $n = 0$, et les p_n sont très vite très petites : la distribution est ramassée près de $n \sim 1$; au contraire, si $a \gg 1$, la loi est en gros centrée en un point $n_{\max} \sim a \gg 1$, loin de l'origine, et sa largeur est égale à \sqrt{a} .

Dynamique conduisant à une distribution de Poisson

Tout comme pour le déclin radioactif, on peut exhiber un processus stochastique produisant tout naturellement des probabilités de Poisson. Historiquement, ce problème a été traité par Schottky, dans le but d'expliquer le bruit électronique dans ces circuits, en conséquence de l'arrivée aléatoire de charges (bruit de *grenaille*, *shot noise* en anglais). L'hypothèse de base est que la probabilité d'arrivée d'une charge supplémentaire entre t et $t + \delta t$ est égale à $\lambda \delta t$, où λ est donc la probabilité d'arrivée d'une charge par unité de temps⁵², supposée indépendante du temps (hypothèse de *stationnarité*).

Soit $p_n(t)$ la probabilité que n charges soient arrivées entre les instants 0 et t . Le bilan des variations possibles des probabilités entre t et $t + \delta t$ donne :

$$p_n(t + \delta t) = (1 - \lambda \delta t) p_n(t) + \lambda \delta t p_{n-1}(t) + \mathcal{O}((\lambda \delta t)^k \geq 2) . \quad (8.117)$$

Après division par δt et passage à la limite $\delta t \rightarrow 0$, il vient :

$$\frac{dp_n}{dt} = -\lambda p_n(t) + \lambda p_{n-1}(t) \quad (n \in \mathbb{N}) . \quad (8.118)$$

Ce système d'équations se résout en introduisant à nouveau la fonction génératrice comme en (8.93) :

$$f(t; s) = \sum_{n=0}^{+\infty} s^n p_n(t) , \quad (8.119)$$

qui est ici une *série* (rien ne limite le nombre de particules arrivées), et dont on voit immédiatement à partir de (8.118) qu'elle satisfait l'équation aux dérivées partielles⁵³ :

$$\frac{\partial f}{\partial t} = \lambda(s-1)f(t; s) \quad (8.120)$$

(comparer avec (8.94)). En supposant la population nulle à $t = 0$ ($p_n(0) = \delta_{n0}$, soit $f(t = 0; s) = 1$), on a immédiatement :

$$f(t; s) = e^{\lambda(s-1)t} ; \quad (8.121)$$

développant alors le second membre en puissances de s , on en déduit $p_n(t)$ par identification :

$$p_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} . \quad (8.122)$$

Il s'agit bien d'une loi de Poisson de paramètre $a = \lambda t$. La moyenne du nombre de charges à l'instant t est λt , et l'écart-type est $\sqrt{\lambda t}$ (voir (8.113) et (8.114)). L'allure de l'évolution temporelle de la distribution se déduit de la fig. 8.11, en remplaçant λ par λt . Aux temps courts devant λ^{-1} , la distribution est ramassée près de $n = 0$. Aux grands temps, la moyenne est grande, et la fluctuation relative est $\frac{1}{\sqrt{\lambda t}}$ (voir (8.116)) : elle tend donc vers zéro, mais très lentement.

Après avoir cité quelques lois courantes pour une v.a. discrète, on considère maintenant des variables aléatoires continues.

⁵²Le même formalisme décrit aussi une population d'individus augmentant spontanément avec l'arrivée des cigognes apportant chacune un nouveau-né, et en l'absence de décès, c'est-à-dire sur échelle de temps courte devant la durée de vie moyenne d'un individu.

⁵³Comme il se doit, la dérivée $\frac{\partial f}{\partial t}$ est nulle $\forall t$ si $s = 1$.

8.6.3 Loi uniforme

C'est la plus simple des densités pour une variable aléatoire continue X : on suppose que la probabilité de trouver X au voisinage d'un point quelconque d'un intervalle $[a, b]$ est la même pour tous les points de cet intervalle :

$$\rho(x) = \begin{cases} C^{\text{ste}} & \text{si } a < x < b \\ 0 & \text{autrement} \end{cases} . \quad (8.123)$$

Comme l'intégrale de ρ doit être normalisée à l'unité, la C^{ste} vaut $\frac{1}{b-a}$. Par exemple, si Φ est un angle que l'on tire uniformément entre $-\pi$ et $+\pi$, $C^{\text{ste}} = \frac{1}{2\pi}$ et :

$$\langle \Phi \rangle = \int_{-\pi}^{+\pi} \frac{1}{2\pi} \phi \, d\phi = 0 , \quad \langle \Phi^2 \rangle = \int_{-\pi}^{+\pi} \frac{1}{2\pi} \phi^2 \, d\phi = \frac{\pi^2}{3} . \quad (8.124)$$

L'écart-type d'un angle tiré uniformément est donc $\frac{\pi}{\sqrt{3}} \simeq 1.81$ radians. D'une façon générale, pour une variable uniforme X , le rapport entre l'écart-type et la longueur de l'intervalle des valeurs possibles est égal à $\frac{1}{2\sqrt{3}} \simeq 29\%$.

La loi de répartition $F(x)$ associée à la densité uniforme $\rho(x)$ est nulle si $x < a$, égale à 1 si $x \geq b$ et augmente linéairement entre a et b (c'est l'intégrale d'une constante) :

$$F(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } a < x < b \\ 1 & \text{si } x \geq b \end{cases} . \quad (8.125)$$

C'est bien une fonction continue.

8.6.4 Loi de Gauss

La loi de Gauss est la plus connue des lois de distribution⁵⁴ pour une variable aléatoire continue X . La densité correspondante $\rho_{\text{Gauss}}(x)$ est :

$$\rho_{\text{Gauss}}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-x_0)^2}{2\sigma^2}} . \quad (8.126)$$

où x_0 et σ sont deux paramètres indépendants réels positifs. Le facteur $\frac{1}{\sqrt{2\pi}\sigma}$ assure la normalisation à 1 de l'intégrale de ρ : $\int_{-\infty}^{+\infty} \rho_{\text{Gauss}}(x) \, dx = 1$. On vérifie immédiatement que :

$$\langle X \rangle = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} dx \, x e^{-\frac{(x-x_0)^2}{2\sigma^2}} = x_0 , \quad \langle X^2 \rangle = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} dx \, x^2 e^{-\frac{(x-x_0)^2}{2\sigma^2}} = x_0^2 + \sigma^2 . \quad (8.127)$$

Ceci donne l'interprétation des paramètres apparaissant dans l'expression (8.126) de la gaussienne ; x_0 n'est autre que la valeur moyenne et σ est l'écart-type :

$$\Delta X = \sqrt{\langle X^2 \rangle - \langle X \rangle^2} = \sigma ; \quad (8.128)$$

La variance est donc égale à σ^2 . Le graphe de $\rho_{\text{Gauss}}(x)$ est une courbe en cloche, symétrique autour de x_0 , $\rho_{\text{Gauss}}(x - x_0) = \rho_{\text{Gauss}}(-x + x_0)$; le maximum est d'autant plus grand et la largeur d'autant plus fine que σ est petit⁵⁵, et inversement.

La loi de Gauss présente la particularité suivante : tous les moments M_k , $k \geq 3$, s'expriment en fonction des moments d'ordre 1 et 2, M_1 et M_2 ; ceci est une caractéristique de l'intégrale gaussienne $\int_{-\infty}^{+\infty} x^k e^{-ax^2} \, dx$, qui joue un rôle fondamental dans les applications⁵⁶. La loi de Gauss possède bien d'autres propriétés remarquables, qui seront données au fur et à mesure dans la suite.

⁵⁴sans doute parce que c'est la plus universelle, pour les raisons qui seront expliquées dans la section 8.8.

⁵⁵La fonction de Gauss est un précurseur de la fonction de Dirac $\delta(x - x_0)$, vers laquelle elle tend dans la limite $\sigma \rightarrow 0$.

⁵⁶Notamment pour la description des processus stochastiques gaussiens. C'est aussi grâce à cette propriété que l'on peut "démontrer" l'équation de la diffusion et, plus généralement, établir l'équation fondamentale de Fokker - Planck.

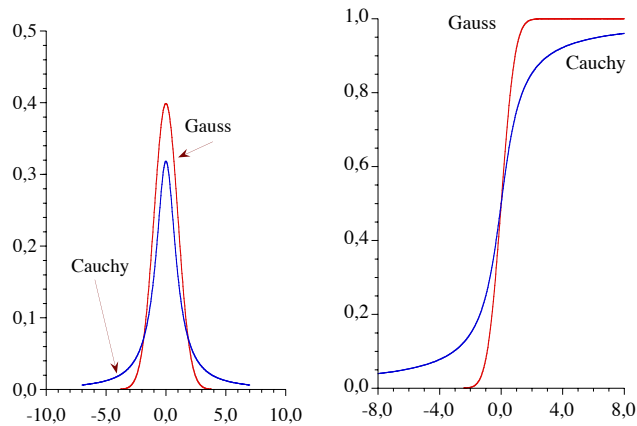


Figure 8.12: Comparaison des lois de Gauss et de Cauchy de moyennes nulles, de même largeur (prise égale à 1), cette dernière étant σ (écart-type) pour la gaussienne, a (voir (8.132)) pour la lorentzienne (dont l'écart-type est infini). À gauche : densités ; la lorentzienne est vaguement triangulaire, la gaussienne est beaucoup plus ronde et plus ramassée. À droite : fonctions de répartition ; la gaussienne est nettement plus raide que la lorentzienne et sature très vite.

La fonction de répartition gaussienne est par définition :

$$F_{\text{Gauss}}(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x dx' e^{-\frac{(x'-x_0)^2}{2\sigma^2}} . \quad (8.129)$$

$F_{\text{Gauss}}(x)$ a un point d'inflexion en x_0 et s'exprime commodément à l'aide de la fonction⁵⁷ $\Phi(x)$ définie comme :

$$\Phi(x) = \frac{2}{\sqrt{\pi}} \int_0^x dy e^{-y^2} , \quad \Phi(\pm\infty) = \pm 1 ; \quad (8.130)$$

on trouve :

$$F_{\text{Gauss}}(x) = \frac{1}{2} \left[1 + \Phi\left(\frac{x-x_0}{\sqrt{2}\sigma}\right) \right] . \quad (8.131)$$

La fonction $1 - \Phi(x)$ est aussi notée $\text{erfc}(x)$.

On verra par la suite (section 8.8) en quoi la loi de Gauss (dite *normale*) a un caractère universel et se rencontre très fréquemment dans la pratique... pas toujours d'ailleurs en vertu du Théorème central limite : le profil (en fréquence) d'une raie atomique émise par une vapeur en équilibre thermodynamique est essentiellement gaussienne – trouver pourquoi –, une raison qui n'a strictement rien à voir avec le fait que la gaussienne est une loi attractive !

8.6.5 Loi de Cauchy

Appelée *lorentzienne* par les Physiciens, cette loi a pour densité :

$$\rho_{\text{Cauchy}}(x) = \frac{1}{\pi} \frac{a}{(x-x_0)^2 + a^2} \quad (a \in \mathbb{R}_+) . \quad (8.132)$$

À nouveau, $\rho_{\text{Cauchy}}(x)$ est symétrique de part et d'autre de x_0 – de toute évidence (?!), la moyenne de X est x_0 . Cette loi a une caractéristique remarquable : la variance est *infinie* ! En effet :

$$\Delta X^2 \equiv \langle X^2 \rangle - \langle X \rangle^2 = \langle (X - \langle X \rangle)^2 \rangle = \frac{a}{\pi} \int_{-\infty}^{+\infty} \frac{(x-x_0)^2}{(x-x_0)^2 + a^2} dx = +\infty , \quad (8.133)$$

⁵⁷souvent appelée *fonction erreur*, notée aussi $\text{Erf}(x)$, mais dont la définition précise peut varier d'un auteur à l'autre.

en conséquence du fait que $\rho_{\text{Cauchy}}(x)$ décroît trop lentement à l'infini⁵⁸, comme x^{-2} . De surcroît, quoique l'écart-type ΔX soit infini, la largeur (géométrique) est visiblement finie : en la définissant conventionnellement comme la largeur à mi-hauteur, elle vaut $2a$ puisque $\rho_{\text{Cauchy}}(x_0 \pm a) = \frac{1}{2}\rho_{\text{Cauchy}}(x_0)$.

La loi de Cauchy est l'exemple le plus simple de loi dite *large*, au sens où les valeurs très éloignées de la moyenne (ici égale à x_0) ont une probabilité qui n'est pas si petite que cela : d'où la *pertinence des événements rares*, qui joue un rôle important en Physique, notamment pour la relaxation de certains systèmes complexes, ou la dynamique au sein des milieux désordonnés.

La fonction de répartition de Cauchy est :

$$F_{\text{Cauchy}}(x) = \int_{-\infty}^x dx' \frac{a}{\pi} \frac{1}{(x' - x_0)^2 + a^2} = \frac{1}{2} + \frac{1}{\pi} \text{Arctan} \frac{x - x_0}{a} . \quad (8.135)$$

Elle se rapproche de ses valeurs limites avec un écart variant comme $\frac{1}{|x|}$.

Exemple où l'on trouve précisément une densité de Cauchy⁵⁹ : les valeurs de l'aléatoire $Y = \tan X$ quand X est uniformément répartie dans l'intervalle $]-\frac{\pi}{2}, +\frac{\pi}{2}[$. En effet, si F_X et F_Y sont les fonctions de répartition correspondantes, on a par définition⁶⁰ :

$$\text{Prob}[Y \leq y] = \text{Prob}[X \leq x] \quad \text{sachant que } y = \tan x \iff x = \text{Arctg} y , \quad (8.136)$$

soit :

$$F_Y(y) = F_X(x = \text{Arctg} y) . \quad (8.137)$$

Si $F_X(x)$ est absolument continue et si la densité correspondante est $\rho_X(x)$, alors la probabilité $\text{Prob}[X \leq x]$ l'intégrale de $\rho_X(x)$ de $-\frac{\pi}{2}$ jusqu'à y . Au total :

$$F_Y(y) = \text{Prob}[X \leq x = \text{Arctg} y] = \int_{-\frac{\pi}{2}}^{\text{Arctg} y} \rho_X(x) dx . \quad (8.138)$$

Si x est répartie *uniformément* entre $\pm\frac{\pi}{2}$, alors $\rho_X(x)$ est une constante et vaut $\frac{1}{\pi}$. En définitive :

$$F_Y(y) = \int_{-\frac{\pi}{2}}^{\text{Arctg} y} \frac{1}{\pi} dx = \frac{1}{\pi} \left(\frac{\pi}{2} + \text{Arctg} y \right) . \quad (8.139)$$

Il en résulte que Y admet aussi exclusivement une fonction de répartition absolument continue, dont la densité est :

$$\rho_Y(y) = \frac{dF_Y}{dy} = \frac{1}{\pi} \frac{1}{y^2 + 1} , \quad (8.140)$$

qui est une loi de Cauchy centrée de largeur $a = 1$.

Signalons une propriété géométrique intéressante du graphe de la loi de Cauchy, illustrée sur la fig. 8.13 : pour tracer semi-quantitativement cette courbe, on part d'un triangle isocèle de base $4a$ et de hauteur $\frac{1}{\pi a} \simeq \frac{1}{3a}$ et, partant du sommet, on fait "tangenter" la courbe aux milieux des deux autres côtés du triangle. Ceci permet de distinguer à l'œil une loi de Cauchy d'une loi de Gauss : la lorentzienne a une allure triangulaire alors que la gaussienne est plutôt ronde. Ces deux distributions se démarquent aussi fortement l'une de l'autre par l'aspect des ailes : la lorentzienne traîne en longueur alors que la gaussienne s'annule très vite.

Pour terminer cette section, mentionnons d'autres lois assez fréquentes⁶¹ :

⁵⁸En fait, aucun moment n'existe à proprement parler, sauf les moments impairs de la variable centrée $X - x_0$ et à condition de les régulariser en écrivant :

$$\tilde{M}_{2k+1} \stackrel{\text{d'éf}}{=} \langle (X - x_0)^{2k+1} \rangle = \lim_{A \rightarrow \infty} \frac{a}{\pi} \int_{-A}^{+A} \frac{y^{2k+1}}{y^2 + a^2} dy = 0 ; \quad (8.134)$$

c'est à ce prix qu'il est légitime de dire notamment que la moyenne de X avec la loi (8.132) est égale à x_0 ...

⁵⁹La lorentzienne est aussi le profil en fréquence d'une raie atomique quand seule la largeur naturelle est en jeu – pourquoi ?

⁶⁰Noter que cette relation simple résulte du fait que la fonction $x \rightarrow \tan x = y$ est *monotone non-décroissante* dans l'intervalle considéré pour x .

⁶¹ C désigne à chaque fois la constante de normalisation telle que l'intégrale de $\rho(x)$ soit égale à l'unité.

- loi Γ : $\rho(x) = C x^\alpha e^{-kx}$, $x \geq 0$, $\alpha > -1$,
- loi de Maxwell : $\rho(x) = C x^2 e^{-\lambda x^2}$, ainsi appelée pour des raisons évidentes
- loi Beta⁶² : $\rho(x) = C x^\alpha (a-x)^\beta$, $0 \leq x \leq a$, $\alpha, \beta > -1$
- loi de Laplace : $\rho(x) = C e^{-k|x|}$ (qui est essentiellement la transformée de Fourier d'une lorentzienne).

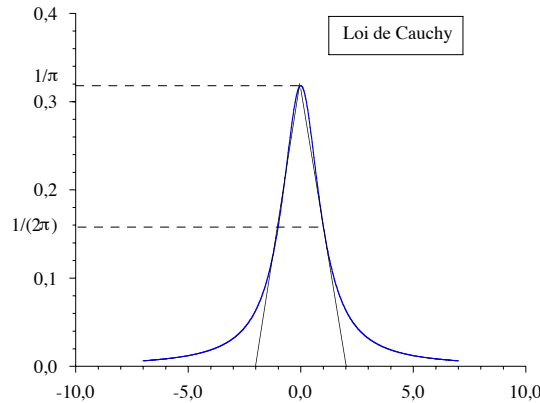


Figure 8.13: Illustration de la propriété géométrique de la densité de Cauchy ($a = 1$) : les deux tangentes à mi-hauteur (en $x = \pm a$) se coupent au sommet, et traversent l'axe des abscisses en $\pm 2a$ (le dessin est tracé avec $a = 1$).

8.7 Fonction caractéristique

8.7.1 Définitions et propriétés

Par définition, et en adoptant les notations traditionnelles en Théorie des probabilités, la *fonction caractéristique* $\phi(t)$ d'une loi de probabilité⁶³ $p(x)$ est la valeur moyenne de la quantité e^{itX} :

$$\phi(t) = \mathbb{E}(e^{itX}) \equiv \langle e^{itX} \rangle . \quad (8.141)$$

En l'absence de composante singulière continue (une hypothèse acceptable pour la plupart des besoins), il est équivalent de dire que $\phi(t)$ est la transformée de Fourier de $p(x)$ définie en (8.49) :

$$\phi(t) = \int_{-\infty}^{+\infty} p(x) e^{itx} dx \iff p \xrightarrow{\mathcal{F}} \phi = \mathcal{F}[p] . \quad (8.142)$$

⁶²Pourquoi doit-on avoir $\alpha, \beta > -1$?

⁶³Il est aussi parfois très utile de considérer le logarithme de $\phi(t)$, $\psi(t) = \ln \phi(t)$, soit $e^{\psi(t)} = \langle e^{itX} \rangle$ (de ce point de vue, *mutatis mutandis*, ψ ressemble à une énergie libre, cependant que $\langle e^{itX} \rangle$ joue le rôle d'une fonction de partition). La fonction ψ permet de calculer commodément les *cumulants* C_k , qui sont certaines fonctions des moments M_k ($C_1 = M_1$, $C_2 = M_2 - M_1^2$, $C_3 = M_3 - 3M_1M_2 + 2M_1^3$, ...). Dans certains problèmes, les cumulants sont les *bonnes* quantités décrivant globalement la dispersion des valeurs d'une variable aléatoire. Pour une distribution gaussienne centrée, tous les cumulants sont nuls au-delà du second ($C_2 = \sigma^2$, $C_{k>2} = 0$) ; pour une variable certaine, tous les cumulants sont nuls (sauf le premier, égal à la valeur certaine). Si on considère le nombre de cumulants non-nuls comme une mesure de l'aléatoire perçu qualitativement, la gaussienne apparaît à cette aune comme une caractéristique de l'aléatoire *minimum*, du plus *faible* hasard, du plus *petit* désordre : c'est la distribution de probabilité la moins "hasardeuse".

Le théorème de Marcinkiewicz affirme que $\psi(t)$ est soit un polynôme du second degré (pour la gaussienne), soit une *série*. Tronquer cette série dans un cas donné – à titre d'approximation dans un problème compliqué – expose à de sérieux déboires, comme par exemple trouver des probabilités négatives (par Fourier inverse de la fonction caractéristique *approchée* qui en résulte).

Notons que ϕ existe toujours puisque $p(x)$ est une fonction positive dont l'intégrale est finie :

$$\forall t \in \mathbb{R} : |\phi(t)| \leq \int_{-\infty}^{+\infty} |e^{itx} p(x)| dx = \int_{-\infty}^{+\infty} p(x) dx = 1 . \quad (8.143)$$

D'une façon ou d'une autre, l'expression explicite de $\phi(t)$ est :

$$\phi(t) = \sum_n p_n e^{itx_n} + \int_{-\infty}^{+\infty} \rho(x) e^{itx} dx . \quad (8.144)$$

Notons que compte tenu de la normalisation de $p(x)$, on a toujours :

$$\phi(0) = 1 . \quad (8.145)$$

À titre d'exemple, les fonctions caractéristiques des lois présentées ci-dessus sont :

$$\phi(t) = \begin{cases} 1 + p(e^{it} - 1) & \text{(loi de Bernoulli)} \\ [1 + p(e^{it} - 1)]^{N_0} & \text{(loi binomiale)} \\ e^{a(e^{it}-1)} & \text{(loi de Poisson)} \\ e^{itx_0 - \frac{1}{2}t^2\sigma^2} & \text{(loi de Gauss)} \\ e^{-|t|a} & \text{(loi de Cauchy)} \end{cases} , \quad (8.146)$$

Dans le cas purement continu, et parce qu'elles forment un couple de Fourier, les fonctions $\rho(x)$ et $\phi(t)$ ont des propriétés analytiques duales l'une de l'autre. Par exemple, si $\rho(x)$ possède des moments M_k de tous les ordres, alors $\phi(t)$ possède un développement en série entière (série de Taylor) centré en $t = 0$; en développant e^{itx} sous l'intégrale dans (8.142), on voit apparaître les moments $M_k = \langle X^k \rangle$:

$$\phi(t) = \sum_{k=0}^{+\infty} \frac{(it)^k}{k!} \int_{-\infty}^{+\infty} x^k p(x) dx \equiv \sum_{k=0}^{+\infty} \frac{(it)^k}{k!} M_k . \quad (8.147)$$

C'est le cas de la gaussienne, pour laquelle il est facile maintenant d'exhiber la propriété remarquable annoncée plus haut. Prenons une variable gaussienne centrée (de moyenne nulle) ; le calcul direct de l'intégrale (8.142) avec $p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$ donne $\phi(t) = e^{-\frac{1}{2}t^2\sigma^2}$, d'où :

$$e^{-\frac{1}{2}t^2\sigma^2} = \sum_{k=0}^{+\infty} \frac{(it)^k}{k!} M_k ; \quad (8.148)$$

en développant en série le premier membre, l'identification donne :

$$\sum_{k=0}^{+\infty} \frac{1}{k!} \left(-\frac{1}{2}t^2\sigma^2\right)^k = \sum_{k=0}^{+\infty} \frac{(it)^k}{k!} M_k \iff M_{2k+1} = 0 , \quad M_{2k} = \frac{(2k)!}{2^k k!} \sigma^{2k} . \quad (8.149)$$

Ainsi, pour une variable gaussienne centrée, tous les moments pairs s'expriment simplement à l'aide de $\sigma^2 \equiv M_2$. Le point important est que M_{2k} est simplement proportionnel à la $k^{\text{ème}}$ puissance de σ^2 : cette propriété joue un rôle majeur dans l'étude des processus stochastiques dits à *diffusion lente*⁶⁴. Pour la référence ultérieure⁶⁵, retenons que pour une v.a. gaussienne centrée :

$$M_2 = \sigma^2 , \quad M_4 = 3\sigma^4 , \quad M_6 = 15\sigma^6 . \quad (8.150)$$

À l'inverse, la loi de Cauchy n'a aucun moment (sauf les moments impairs, et à condition de les régulariser comme indiqué dans la note 58) ; de fait, sa fonction caractéristique $e^{-|t|a}$ n'est pas analytique en $t = 0$.

⁶⁴C'est cette propriété qui conduit à l'équation de Fokker - Planck, une des équations fondamentales permettant de décrire la dynamique hors d'équilibre d'un système macroscopique.

⁶⁵voir sous-section 8.8.2 et notamment la fig. 8.14.

La fonction caractéristique est très commode pour obtenir d'emblée tous les moments, quand ils existent. En effet, de (8.147) on déduit :

$$M_k = i^{-k} \left[\frac{d^k}{dt^k} \phi(t) \right]_{t=0} . \quad (8.151)$$

C'est pourquoi $\phi(t)$ est parfois appelée fonction génératrice des moments, par abus de langage ; en effet, si $\phi(t)$ existe toujours, ce n'est pas le cas des M_k : pour la lorentzienne, $\phi = e^{-a|t|}$ alors qu'aucun moment n'existe. À l'inverse, s'agissant de déterminer la distribution de probabilités d'un phénomène donné, une procédure standard consiste à mesurer les différents moments M_k afin de bâtir peu à peu la fonction caractéristique $\phi(t)$. Cette *reconstruction* à partir des moments est une opération délicate.

Enfin, la fonction $\phi(t)$, en tant que transformée de Fourier d'une fonction *positive* possède des propriétés remarquables, qu'il convient de respecter dans les traitements approximatifs. Ce sont ces propriétés qu'il faut faut s'efforcer de respecter lors de la reconstruction à partir des moments M_k , afin de s'éviter des déboires.

8.7.2 Somme de variables aléatoires

Il s'agit ici d'examiner les propriétés élémentaires de la somme de plusieurs variables aléatoires. On va voir que le théorème de convolution permet de montrer simplement que la loi de probabilité d'une somme de variables *indépendantes* a pour fonction caractéristique le *produit* des fonctions caractéristiques. En effet, soit X_1 et X_2 deux telles v.a. dont les fonctions de répartition sont F_1 et F_2 , et soit $X = X_1 + X_2$ leur somme. La question est de trouver la fonction de répartition de X , $F(x)$.

La probabilité que X soit inférieure ou égale à x est égale à la somme des probabilités des événements exclusifs où à la fois $X_1 \leq x - x'$ et X_2 est compris entre x' et $x' + dx'$:

$$\text{Prob}[X \leq x] = \text{Somme sur } x' \{ \text{Prob}[X_1 \leq x - x' \text{ et } x' < X_2 \leq x' + dx'] \} ; \quad (8.152)$$

cette dernière probabilité s'obtient par un simple produit de probabilités puisque les deux variables sont indépendantes⁶⁶ par hypothèse ; de la sorte :

$$\text{Prob}[X \leq x] = \text{Somme sur } x' \{ \text{Prob}[X_1 \leq x - x'] \times \text{Prob}[x' < X_2 \leq x' + dx'] \} , \quad (8.153)$$

soit :

$$F(x) = \int F_1(x - x')[F_2(x' + dx') - F_2(x')] = \int F_1(x - x')F_2'(x')dx' . \quad (8.154)$$

Il est clair que dans le raisonnement conduisant à (8.153), les rôles de X_1 et de X_2 peuvent être échangés : on peut tout autant dire que la probabilité pour que X soit inférieure ou égale à x est égale à la somme des probabilités des événements où à la fois $X_2 \leq x - x'$ et X_1 est compris entre x' et $x' + dx'$. Au total :

$$F(x) = \int F_1(x - x')F_2'(x') dx' = \int F_2(x - x')F_1'(x') dx' . \quad (8.155)$$

Ainsi, F est la convolution de l'une des F_j avec la dérivée de l'autre.

Une fois trouvée la fonction de répartition, on en déduit sa dérivée au sens général, $p(x)$ en dérivant par rapport à x :

$$p(x) = \lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} \iff p(x) = \int p_1(x')p_2(x - x') dx' \equiv (p_1 * p_2)(x) . \quad (8.156)$$

Ainsi, en raison de l'indépendance des variables, $p(x)$ ressort comme la convolution des p_j ; par le théorème de convolution, sa fonction caractéristique (qui est sa transformée de Fourier) est donc le simple produit des deux fonctions caractéristiques :

$$p_j \xrightarrow{\mathcal{F}} \phi_j , \quad p \xrightarrow{\mathcal{F}} \phi_1 \phi_2 . \quad (8.157)$$

⁶⁶Cette hypothèse est visiblement cruciale !

Il s'agit d'un résultat majeur, mais répétons que l'hypothèse que les X_j sont indépendantes est absolument essentielle. Ce dernier résultat se généralise immédiatement à la somme d'un nombre fini de variables aléatoires indépendantes :

$$p_j \xrightarrow{\mathcal{F}} \phi_j, \quad p(x) = \text{Prob} [x < X = \sum_{j=1}^N x_j \leq x + dx] \xrightarrow{\mathcal{F}} \phi_1(t)\phi_2(t) \dots \phi_N(t). \quad (8.158)$$

Pour des variables continues, cette propriété se transcrit à l'identique pour les densités ρ_1 et ρ_2 . Pour des variables purement discrètes ayant les "densités" :

$$p_1(x) = \sum_n p_{1,n} \delta(x - x_{1,n}), \quad p_2(x) = \sum_m p_{2,m} \delta(x - x_{2,m}), \quad (8.159)$$

la convolution $p_1 * p_2$ est :

$$(p_1 * p_2)(x) = \sum_{n,m} p_{1,n} p_{2,m} \int dx' \delta(x - x' - x_{1,n}) \delta(x' - x_{2,m}) = \sum_{n,m} p_{1,n} p_{2,m} \delta(x - x_{1,n} - x_{2,m}). \quad (8.160)$$

La fonction caractéristique est la transformée Fourier de cette convolution, et a l'expression :

$$\phi(t) = \sum_{n,m} p_{1,n} p_{2,m} \int dx \delta(x - x_{1,n} - x_{2,m}) e^{itx} = \sum_{n,m} p_{1,n} p_{2,m} e^{it(x_{1,n} + x_{2,m})} \equiv \phi_1(t) \phi_2(t) \quad (8.161)$$

où $\phi_j(t) = \sum_n p_{j,n} e^{itx_{j,n}}$.

Notons que le résultat (8.158) peut aussi s'obtenir comme suit. Par définition :

$$\phi(t) = \langle e^{it(X_1 + X_2 + \dots + X_N)} \rangle; \quad (8.162)$$

par hypothèse, les variables X_j sont indépendantes, donc les moyennes se factorisent :

$$\phi(t) = \langle e^{itX_1} \rangle \langle e^{itX_2} \rangle \dots \langle e^{itX_N} \rangle \equiv \phi_1(t) \phi_2(t) \dots \phi_N(t). \quad (8.163)$$

En particulier, si toutes les variables indépendantes sont distribuées suivant la même loi $p_0(x)$, de fonction caractéristique $\phi_0(t)$, alors, suivant (8.158), la somme est distribuée suivant la loi ayant la fonction caractéristique $\phi(t)$ telle que :

$$\phi(t) = [\phi_0(t)]^N. \quad (8.164)$$

C'est pour cette raison que, dans (8.146), la fonction caractéristique de la loi binomiale est la puissance $N_0^{\text{ème}}$ de la fonction caractéristique de la loi de Bernoulli : on sait qu'une variable binomiale prenant les valeurs 1, 2, ..., N_0 est la somme de N_0 variables de Bernoulli indépendantes prenant les valeurs 0 et 1.

Pour terminer, signalons une façon expéditive d'établir l'expression de la densité $\rho(x)$ de la somme de deux variables aléatoires indépendantes. Il suffit de filtrer toutes les possibilités indépendantes de x_1 et de x_2 par la fonction de Dirac ne retenant que les termes correspondants à une valeur donnée x pour la somme, et d'appliquer la règle opérationnelle de $\delta(x)$:

$$\rho(x) = \int dx_1 \int dx_2 p_1(x_1) p_2(x_2) \delta(x - (x_1 + x_2)) = \int dx_1 p_1(x_1) p_2(x - x_1). \quad (8.165)$$

8.7.3 Stabilité d'une loi par l'addition

Le résultat exprimé par (8.158) est d'une extrême importance (mais ne jamais oublier qu'il ne tient que pour des variables indépendantes !). Examinons ce qu'il implique pour les lois présentées ci-dessus, en se référant aux expressions données en (8.146). Prenons d'abord deux variables de Poisson X_1 et X_2 , de paramètres respectifs a_1 et a_2 . La fonction caractéristique de leur somme $X = X_1 + X_2$ est :

$$\phi(t) = e^{a_1(e^{it}-1)} e^{a_2(e^{it}-1)} = e^{(a_1+a_2)(e^{it}-1)}. \quad (8.166)$$

C'est encore la fonction caractéristique d'une loi de Poisson. Il en résulte que la somme suit une loi de Poisson, de paramètre $a = a_1 + a_2$ – une fois établi le fait que la somme suit une loi de Poisson, on a forcément $a = a_1 + a_2$ puisque a est la valeur moyenne.

Pour deux variables gaussiennes X_1 et X_2 de moyennes respectives x_{01} et x_{02} , et d'écart-types respectifs σ_1 et σ_2 , (8.146) dit que la fonction caractéristique de la somme $X = X_1 + X_2$ est $e^{itx_{01} - \frac{1}{2}t^2\sigma_1^2} e^{itx_{02} - \frac{1}{2}t^2\sigma_2^2} = e^{it(x_{01} + x_{02}) - \frac{1}{2}t^2(\sigma_1^2 + \sigma_2^2)}$: c'est encore une fonction caractéristique de gaussienne, de moyenne $x_{01} + x_{02}$ (ce qui n'est pas surprenant), et de variance égale à la somme des variances $\sigma_1^2 + \sigma_2^2$ (ce qui ne l'est pas davantage)⁶⁷. Enfin⁶⁸, pour deux variables de Cauchy, la conclusion est encore la même : leur somme obéit à une loi de Cauchy de paramètre a égal à la somme des paramètres relatifs à X_1 et X_2 .

Le fait remarquable est que, pour ces lois, la variable somme $X = X_1 + X_2$ obéit à une loi de *même nature* que chacune de ses “composantes” X_1 et X_2 . Cette propriété extraordinaire n'est pas vraie en général : elle ne tient visiblement pas pour deux variables de Bernoulli, ou pour deux variables binomiales. Les lois qui la possèdent sont dites *stables par l'addition* ; elles jouent un rôle de première importance en pratique, et sont par ailleurs des curiosités mathématiques justifiant l'intérêt qui leur est porté à la suite des travaux de Paul Lévy dans les années 1920. Inversement, elles suggèrent de se poser des questions intéressantes, comme par exemple : dans quelle mesure une variable aléatoire X peut-elle être décomposée en la somme de deux autres variables aléatoires, ce que l'on formule usuellement en disant : dans quelle mesure une loi est-elle *divisible* ? Le champ de recherche en la matière porte le nom d'*arithmétique des lois de probabilités*, et est encore à l'heure actuelle un domaine en pleine activité.

Il convient de savoir que la réciproque n'est vraie que pour les lois de Poisson et de Gauss : si $X_1 + X_2$ est une variable de Gauss (*resp.* de Poisson), et si X_1 et X_2 sont indépendantes, alors X_1 et X_2 sont des variables de Gauss (*resp.* de Poisson). À l'inverse, il existe des cas où $\phi_1(t)\phi_2(t) = e^{-a|t|}$ sans que X_1 et X_2 soient des variables de Cauchy.

8.8 Lois-limites. Théorème central limite

8.8.1 Problématique

Les remarques qui viennent d'être faites attirent l'attention sur des lois remarquables par le fait qu'elles sont stables par l'addition de variables aléatoires indépendantes. En termes plus précis, ceci signifie qu'étant donné des variables aléatoires indépendantes X_j obéissant à un même type de loi (de Gauss, de Poisson, *etc.*), si on définit une suite de variables S_N (elles aussi évidemment aléatoires) :

$$S_1 = X_1, \quad S_2 = X_1 + X_2, \quad S_3 = X_1 + X_2 + X_3, \quad \iff \quad S_{N+1} = S_N + X_{N+1}, \quad (8.167)$$

alors la loi de probabilité de S_N est de même nature – mais bien sûr a ses propres paramètres dépendant de N : si les X_j sont toutes gaussiennes (moyennes x_{0j} et variances σ_j^2), la gaussienne pour S_N a pour moyenne $\sum_{j=1}^N x_{0j}$ et pour variance $\sum_{j=1}^N \sigma_{X_j}^2$. En quelque sorte, ces lois remarquables sont des *points-fixes*⁶⁹ dans un certain espace de classes de fonctions, quand l'itération élémentaire est l'addition des v.a.

La propriété de *point fixe* est une propriété en soi ; une autre propriété pour un “point” est d'être la *limite* x_∞ d'une suite x_n de nombres proprement définie ; il peut aussi s'agir d'une suite de fonctions f_n , dont

⁶⁷La moyenne d'une somme de variables aléatoires est toujours la somme des valeurs moyennes, que ces variables soient indépendantes ou non. En revanche, la variance n'est égale à la somme des variances que si les variables sont indépendantes.

⁶⁸En sous-produit, on pourra retenir que la convolution de deux gaussiennes est une gaussienne, la convolution de deux poissonniennes est une poissonnienne, la convolution de deux lois de Cauchy (lorentziennes) est une loi de Cauchy.

⁶⁹Un point fixe est un point invariant dans une certaine loi de récursion (itération). Par exemple, soit une suite de nombres x_n définis par $x_{n+1} = f(x_n)$, x_0 donné, où f est une fonction également donnée ; un point fixe, noté x^* , est un point qui satisfait $x^* = f(x^*)$: c'est donc un point invariant par l'application f . On connaît des itérations simplissimes aux propriétés stupéfiantes, lorsque la fonction f est *non-linéaire* (ce qui est une banalité en soi). Par exemple, l'itération dite *logistique*, définie par $f(x) = \lambda x(1-x)$, est sans doute l'exemple le plus simple de route vers le chaos quand le paramètre de contrôle λ augmente (cascade de Feigenbaum).

la limite est notée f_∞ . Un tel point peut être dit *attracteur* au sens où, partant d'une certaine situation, la suite des opérations conduit peu à peu inexorablement vers la limite⁷⁰ considérée.

Ainsi, s'agissant cette fois de la somme S_N de N variables indépendantes X_i distribuées suivant des lois $p_{X_i}(x)$ *a priori quelconques*, on peut se poser la question suivante : dans quelle mesure, et à quelle(s) condition(s), la loi de répartition $F_N(x)$ de S_N peut-elle converger vers une certaine loi quand N tend vers l'infini ? Et si oui, quelle est cette loi ? Il est bien clair qu'une telle loi-limite, si elle existe, aura un caractère *universel* qui la fera apparaître spontanément dans une multitude de phénomènes où la somme de variables aléatoires joue un rôle de premier plan.

La réponse est oui. Une telle loi existe (dans des conditions qui seront précisées), et c'est la loi de Gauss. L'affirmation qui précise les choses porte le nom de *Théorème central limite*, ou *Théorème de la limite centrale*, selon les auteurs. Notons que s'agissant de comprendre ce qui se passe lorsque l'on considère une suite de variables aléatoires, il conviendrait au préalable de redéfinir complètement la notion même de convergence, c'est-à-dire de plonger dans la *topologie aléatoire* – un monde en soi. On se contentera ici d'affirmations élémentaires, dont certaines dissimuleront des difficultés de fond, ou des subtilités le plus souvent inessentiels pour le physicien. Pour tout dire, des affirmations dans la suite ne seront pas seulement *élémentaires*, elles seront aussi quelque peu (ou très ?) vagues aux yeux des puristes. En tout cas, les résultats obtenus de façon parfois peu rigoureuse permettront de comprendre des faits essentiels, en particulier l'omniprésence de la loi de Gauss dans la Nature – un fait expérimental indiscutable, même si évidemment les phénomènes aléatoires n'obéissent tous pas à une telle loi.

8.8.2 Théorème central limite

La question est de montrer que, dans certaines conditions qui vont apparaître d'elles-mêmes, la somme de N variables *indépendantes* :

$$S_N = \sum_{j=1}^N X_j \quad (8.168)$$

tend vers une variable gaussienne lorsque N augmente indéfiniment. *A priori*, chaque variable X_j a sa propre loi de probabilité $p_{X_j}(x)$, de moyenne $\langle X_j \rangle$ et d'écart-type *finis*⁷¹ σ_{X_j} :

$$\langle X_j \rangle = \int x p_{X_j}(x) dx, \quad |\langle X_j \rangle| < +\infty, \quad \langle X_j^2 \rangle = \int x^2 p_{X_j}(x) dx < +\infty, \quad (8.169)$$

$$\sigma_{X_j}^2 = \langle X_j^2 \rangle - \langle X_j \rangle^2 < +\infty. \quad (8.170)$$

Bien sûr, l'information contenue dans la connaissance de tous les $p_{X_j}(x)$ permet de trouver p_{S_N} : il suffit de convoluer toutes les lois p_{X_j} . Formellement, la loi cherchée a pour expression :

$$p_{S_N}(x) = \int dx_1 \int dx_2 \dots \int dx_{N-1} p_{X_1}(x_1) p_{X_2}(x_2) \dots p_{X_{N-1}}(x_{N-1}) p_{X_N}(x - x_1 - x_2 - \dots - x_{N-1}), \quad (8.171)$$

et il n'y a qu'à calculer l'intégrale multiple pour obtenir p_{S_N} . En général (notamment pour les petites valeurs de N), la loi p_{S_N} est ce qu'elle est, et ne présente aucun caractère d'universalité⁷².

En réalité, le problème se présente souvent autrement : pour un enquêteur d'institut de sondage, N est de l'ordre de 1000 ; quant aux physiciens, ils sont plutôt habitués à manipuler des nombres *énormes*, dont le prototype est le nombre d'Avogadro $\mathcal{N} \sim 10^{23}$. En pareil cas, calculer l'intégrale (même avec un ordinateur puissant) serait une tâche impossible⁷³, et en fait totalement inutile dans bien des cas, comme on va le voir.

⁷⁰L'ensemble des points de départ où il en va ainsi est appelé *bassin attracteur*.

⁷¹Comme on le verra, l'existence de chaque σ_{X_j} est une hypothèse cruciale.

⁷²Encore que... Il est frappant de constater, à l'aide d'une simple calculette, que la somme d'un *petit* nombre (cinq, six) de variables indépendantes distribuées uniformément (sur $[0, 1]$ par exemple) est une variable gaussienne à une très bonne approximation (voir fig. 8.14).

⁷³Si la machine met 10^{-7} seconde pour calculer une intégrale, le temps de calcul total avec \mathcal{N} est $\sim 10^{16}$ secondes, soit environ trois milliards d'années...

Par ailleurs, on ne connaît pas toujours les lois individuelles $p_{X_j}(x)$, une situation dans laquelle on ne sait strictement rien faire en l'état.

L'intrusion de *grands nombres*, voire de *très grands nombres*, est une hypothèse qualitative essentielle pour toute la suite. Elle implique que d'autres conditions doivent être réunies, notamment que *toutes* (ou presque toutes) les variables X_j doivent être réellement fluctuantes : à la limite où seul un petit nombre n d'entre elles fluctuent (les autres étant quasi certaines), on se retrouve de fait dans une situation avec un petit nombre de variables fluctuantes, auquel cas les affirmations énoncées ci-dessous perdent leur validité.

Dire que toutes les variables fluctuent de façon comparable, c'est affirmer que tous les écarts-types σ_{X_j} sont du *même ordre de grandeur* : si certains étaient gigantesques et les autres minuscules, on retrouverait le cas où, relativement, seules certaines variables fluctuent de fait. À partir du moment où tous les σ_{X_j} sont du même ordre de grandeur, on peut tous les supposer voisins de l'unité, au prix d'un recalibrage trivial commun à toutes les variables X_j . Dans toute la suite, on suppose donc :

$$\sigma_{X_j} \sim 1 \quad \forall j . \quad (8.172)$$

Comme la variance d'une somme de variables indépendantes est la somme des variances, cependant que la moyenne de la somme est *toujours* la somme des moyennes⁷⁴, on peut dès maintenant affirmer que S_N est une variable aléatoire de moyenne $\sum_j \langle X_j \rangle$ et de variance égale à la somme de N nombres d'ordre 1, c'est-à-dire est d'ordre N :

$$\langle S_N \rangle = \sum_j \langle X_j \rangle , \quad \sigma_{S_N} = \sqrt{\sum_j \sigma_{X_j}^2} \sim \sqrt{N} . \quad (8.173)$$

L'analyse précédente suggère d'introduire une nouvelle variable qui est centrée et dont les fluctuations sont d'ordre unité ; on définit ainsi :

$$Y_N = \frac{1}{\sqrt{N}}(S_N - \sum_j \langle X_j \rangle) = \frac{1}{\sqrt{N}}(S_N - \langle S_N \rangle) . \quad (8.174)$$

Il s'agit maintenant de préciser la loi de distribution de Y_N , notée $p_{Y_N}(y)$ ou, de façon équivalente, de trouver une approximation de sa fonction caractéristique $\phi_{Y_N}(t)$. On a par définition :

$$\phi_{Y_N}(t) = \langle e^{itY_N} \rangle = \langle e^{it \frac{1}{\sqrt{N}}(S_N - \sum_j \langle X_j \rangle)} \rangle = e^{-i \frac{t}{\sqrt{N}} \langle S_N \rangle} \langle e^{it \frac{1}{\sqrt{N}} S_N} \rangle \equiv e^{-i \frac{t}{\sqrt{N}} \langle S_N \rangle} Z_N(t) . \quad (8.175)$$

C'est la quantité $Z_N(t)$ qu'il convient d'analyser dans la limite $N \gg 1$. On a :

$$Z_N(t) = \langle e^{i \frac{t}{\sqrt{N}}(X_1 + X_2 + \dots + X_N)} \rangle = \langle \prod_j e^{i \frac{t}{\sqrt{N}} X_j} \rangle . \quad (8.176)$$

Comme les variables X_j sont indépendantes, la moyenne du produit est égal au produit des moyennes :

$$Z_N(t) = \langle \prod_j e^{i \frac{t}{\sqrt{N}} X_j} \rangle = \prod_j \langle e^{i \frac{t}{\sqrt{N}} X_j} \rangle = \prod_j \phi_j \left(\frac{t}{\sqrt{N}} \right) , \quad (8.177)$$

où l'on retrouve le produit des fonctions caractéristiques des aléatoires X_j . Compte tenu des hypothèses (existence des moyennes et des variances individuelles), on a :

$$\phi_j(t) = 1 + it \langle X_j \rangle - \frac{t^2}{2} \langle X_j^2 \rangle + \dots \quad \forall j , \quad (8.178)$$

de sorte que le logarithme de $Z_N(t)$ est :

$$\ln Z_N(t) = \sum_j \ln \left[1 + i \frac{t}{\sqrt{N}} \langle X_j \rangle - \frac{t^2}{2N} \langle X_j^2 \rangle + \mathcal{O}(N^{-\frac{3}{2}} t^3) \right] . \quad (8.179)$$

⁷⁴Cette affirmation est un truisme lorsque le nombre N de variables additionnées est *fini*.

On développe maintenant le logarithme suivant $\ln(1+z) = z - \frac{z^2}{2} + \mathcal{O}(z^3)$:

$$\ln Z_N(t) = \sum_j \left[i \frac{t}{\sqrt{N}} \langle X_j \rangle - \frac{t^2}{2N} \langle X_j^2 \rangle + \frac{t^2}{2N} \langle X_j \rangle^2 + \mathcal{O}(N^{-\frac{3}{2}} t^3) \right], \quad (8.180)$$

soit :

$$\ln Z_N(t) = i \frac{t}{\sqrt{N}} \langle S_N \rangle - \frac{t^2}{2N} \sum_j (\langle X_j^2 \rangle - \langle X_j \rangle^2) + \mathcal{O}(N^{-\frac{3}{2}} t^3) \rightarrow i \frac{t}{\sqrt{N}} \langle S_N \rangle - \frac{t^2}{2N} \sum_j \sigma_{X_j}^2. \quad (8.181)$$

En définitive, dans la limite des grands N , on a⁷⁵ (compte tenu de (8.173)) :

$$Z_N(t) \simeq e^{i \frac{t}{\sqrt{N}} \langle S_N \rangle - \frac{t^2}{2N} \sigma_{S_N}^2}, \quad (8.182)$$

d'où, revenant à (8.175) :

$$\phi_{Y_N}(t) \simeq e^{-i \frac{t}{\sqrt{N}} \langle S_N \rangle} e^{i \frac{t}{\sqrt{N}} \langle S_N \rangle - \frac{t^2}{2N} \sigma_{S_N}^2} = e^{-\frac{t^2}{2} \frac{\sigma_{S_N}^2}{N}}. \quad (8.183)$$

On reconnaît à nouveau la fonction caractéristique d'une gaussienne. Ainsi, la variable Y_N est asymptotiquement distribuée suivant la loi normale de moyenne nulle et de variance $\frac{1}{N} \sigma_{S_N}^2$:

$$p_{Y_N}(y) \simeq \frac{1}{\sqrt{2\pi} \frac{1}{\sqrt{N}} \sigma_{S_N}} e^{-\frac{y^2}{2 \frac{1}{N} \sigma_{S_N}^2}} \quad (N \gg 1). \quad (8.184)$$

Entre les deux variables Y_N et S_N , prenant respectivement des valeurs notées x et y , il y a un facteur $\frac{1}{\sqrt{N}}$ (voir la définition (8.174) de Y_N), qui se retrouve au niveau des densités :

$$p_{Y_N}(y) dy = p_{S_N}(x) dx \iff p_{S_N}(x) = p_{Y_N} \left(\frac{x - \langle X \rangle}{\sqrt{N}} \right) \frac{dy}{dx}. \quad (8.185)$$

Comme $\frac{dy}{dx} = \frac{1}{\sqrt{N}}$, on trouve finalement :

$$p_{S_N}(x) \simeq \frac{1}{\sqrt{2\pi} \sigma_{S_N}} e^{-\frac{(x - \langle S_N \rangle)^2}{2 \sigma_{S_N}^2}} \quad (N \gg 1), \quad (8.186)$$

où conformément à (8.173), $\langle S_N \rangle = \sum_j \langle X_j \rangle$, $\sigma_{S_N} = \sqrt{\sum_j \sigma_j^2}$. Ainsi, la somme S_N de N variables aléatoires indépendantes X_j , de moyennes $\langle X_j \rangle$ et de variances $\sigma_{X_j}^2$, mais de lois p_{X_j} quelconques par ailleurs est, pour les grands N , une variable à peu près gaussienne de moyenne et de variance déterminées.

C'est ce résultat qui est appelé *Théorème central limite*, dont la qualification est largement méritée. À la réflexion, ce résultat est assez extraordinaire : il ne repose finalement que sur un petit nombre d'hypothèses, essentiellement l'indépendance des variables et l'existence d'une variance pour chacune d'entre elles – en revanche rien d'autre n'est dit sur les lois individuelles p_{X_j} des différentes variables X_j . S'il s'agit de deux hypothèses très fortes, elles restent néanmoins assez peu contraignantes et, très souvent réalisées de fait en pratique, expliquent fondamentalement l'apparition si fréquente de la loi gaussienne dans la Nature.

Ce théorème prend une forme encore plus simple si les variables X_j sont identiquement distribuées, suivant une seule et même loi p_X ; alors, toutes les X_j ont la même moyenne $\langle X \rangle$ et la même variance σ^2 , de sorte que $\langle S_N \rangle = N \langle X \rangle$ et $\sigma_{S_N}^2 = N \sigma^2$. Dans ces conditions, la loi gaussienne asymptotique est :

$$p_{S_N}(x) \simeq \frac{1}{\sqrt{2\pi} \sqrt{N} \sigma} e^{-\frac{(x - N \langle X \rangle)^2}{2 N \sigma^2}}. \quad (8.187)$$

⁷⁵ Il serait plus juste, à défaut d'analyse plus élaborée, de remplacer \simeq par \sim .

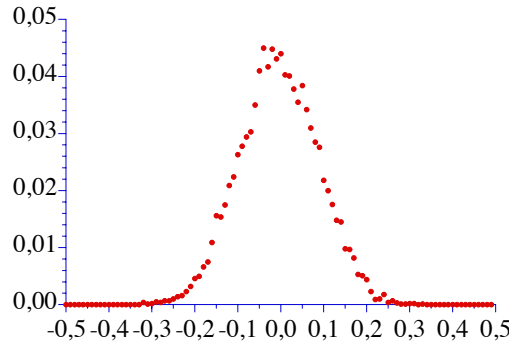


Figure 8.14: Distribution de la somme de $N = 10$ variables aléatoires indépendantes, chacun étant tirée uniformément dans $[-0.5, +0.5]$. L'aspect gaussien reconnaissable à vue se confirme en calculant les premiers moments M_k et en comparant avec les relations spécifiques des moments d'une gaussienne (voir (8.149) et (8.150)). Pour cette expérience, on a trouvé $\frac{M_4}{3M_2^2} = 0.962$, $\frac{M_6}{15M_2^3} = 0.904$, deux nombres assez voisins de 1 assurant que la somme est gaussienne à une bonne approximation. Que dire quand $N \sim 10^{23}$...

Un fait doit être souligné. La fluctuation *relative* de l'aléatoire S_N est bien mesurée par le rapport entre sa fluctuation et sa valeur moyenne⁷⁶. D'après ci-dessus, on a :

$$\frac{\sigma_{S_N}}{\langle S_N \rangle} = \frac{\sqrt{N}\sigma}{N\langle X \rangle} = \frac{1}{\sqrt{N}} \frac{\sigma}{\langle X \rangle} \propto N^{-\frac{1}{2}} . \quad (8.188)$$

Le point essentiel à retenir est que la fluctuation relative décroît comme $\frac{1}{\sqrt{N}}$: elle tend certes vers zéro avec N , mais très lentement. Ce "détail" est sans importance pour le physicien qui pratique la Mécanique statistique ($N \sim 10^{24}$, ce qui donne une fluctuation relative de l'ordre de 10^{-12}), mais est sans doute un souci permanent pour les instituts de sondage ($\frac{1}{\sqrt{1000}} \simeq \frac{1}{32} \simeq 3\%$...).

Donnons un exemple classique d'application de ce théorème à la *marche de l'ivrogne* : une marche au hasard unidimensionnelle où, tous les δt , l'ivrogne fait un pas de longueur a soit dans un sens, soit dans l'autre, avec des probabilités respectives p (un pas vers la droite) et $1 - p$ (un pas vers la gauche). On veut connaître la position typique de l'ivrogne au temps $t_N = N\delta t$.

Le problème peut se résoudre exactement par dénombrement, en supposant qu'à chaque nouveau pas l'ivrogne a complètement oublié les pas précédents (les différents mouvements élémentaires sont alors associés à des v.a. indépendantes). La probabilité $p_n(t_N)$ que l'ivrogne soit en na après N petits pas est $C_N^n p^n (1-p)^{N-n}$, d'où l'on déduit immédiatement toutes les valeurs moyennes souhaitées.

Si l'on s'en tient à la seule position (et à son écart-type), et en se contentant de la dynamique asymptotique (grands temps, soit $N \gg 1$), il est beaucoup plus rapide d'utiliser le Théorème central limite. La position X_N est une variable aléatoire qui est la somme des petits pas élémentaires δX_n ; chacun de ces derniers prend l'une des deux valeurs $\varepsilon_n a$, où $\varepsilon_n = +1$ avec probabilité p , -1 avec probabilité $1 - p$; si les $x_n(N)$ notent les valeurs possibles de la v.a. X_N , on a :

$$x_n(N) = \sum_{n=1}^N \delta x_n , \quad \delta x_n = \varepsilon_n a . \quad (8.189)$$

Compte tenu de l'état de l'ivrogne (et notamment de son amnésie), les ε_n sont des v.a. indépendantes, distribuées suivant la même loi de Bernoulli ; la moyenne de ε_n est $p \times (+1) + (1 - p) \times (-1) = 2p - 1$, celle de ε_n^2 est $p \times (+1)^2 + (1 - p) \times (-1)^2 = 1$. L'écart-type σ est donc $\sqrt{4p(1-p)}$.

⁷⁶On peut dire qu'une variable certaine est le cas limite d'une variable aléatoire pour laquelle un tel rapport tend vers zéro. Par ailleurs, si la dispersion de X résulte d'une variabilité d'une série de mesures physiques, ce rapport joue le rôle d'une incertitude relative.

La moyenne de X_N est la somme des moyennes :

$$\langle X_N \rangle = \sum_{n=1}^N \langle \varepsilon_n \rangle a = N(2p-1)a, \quad (8.190)$$

et comme les δX_n sont des v.a. indépendantes, la variance de la somme est la somme des variances :

$$\sigma_{X_N}^2 \equiv \langle X_N^2 \rangle - \langle X_N \rangle^2 = 4p(1-p)Na^2. \quad (8.191)$$

Ces deux résultats sont indépendants de N (petit ou grand). Maintenant, le Théorème central limite permet d'affirmer que si $N \gg 1$, alors X_N est asymptotiquement une variable gaussienne. Les valeurs de X_N sont de la forme na , où n est entier et $-N \leq n \leq N$; pour $N \gg 1$, la loi de distribution de la position X_N est très bien approximée par :

$$P_{X_N}(n) = \frac{1}{\sqrt{2\pi}\sigma_{X_N}} e^{-\frac{(na - \langle X_N \rangle)^2}{2\sigma_{X_N}^2}}. \quad (8.192)$$

Bien sûr, la vraie distribution de X_N est à support borné, compris entre $\pm Na$, alors que la gaussienne est non-nulle entre $\pm\infty$; cette distorsion dans les ailes est pratiquement sans incidence quand N est très grand.

Avec $t = N\delta t$, le résultat important est :

$$\sigma_{X_N} = \sqrt{4p(1-p)Na} = \sqrt{4p(1-p)\frac{t}{\delta t}a} \iff \sigma_{X_N} \propto \sqrt{t}; \quad (8.193)$$

ce résultat – caractéristique d'un phénomène de *diffusion* – signifie que la taille typique de la région visitée par l'ivrogne croît comme la racine carrée du temps, c'est-à-dire beaucoup plus lentement que si l'ivrogne, débouffé, se déplaçait toujours dans le (bon) sens à vitesse constante (alors la distance parcourue serait $\propto t$). Il est traditionnel d'introduire la constante de diffusion, D définie (à une dimension d'espace⁷⁷) comme :

$$\Delta X_N^2 \equiv 2Dt; \quad (8.194)$$

on obtient ici :

$$D = 2p(1-p)\frac{a^2}{\delta t}. \quad (8.195)$$

La constante de diffusion est maximale pour $p = \frac{1}{2}$ (c'est le hasard le plus complet). À l'inverse, si p tend vers 1, la marche devient de moins en moins au hasard ; à la limite $p = 1$, la constante de diffusion s'annule, et $P_{X_N}(n)$ tend vers une fonction infiniment piquée, centrée au point d'abscisse Na : il s'agit alors d'un mouvement déterministe (abscisse = $Na = \frac{a}{\delta t}t$), l'homme à jeûn se déplaçant à la vitesse constante $v = \frac{a}{\delta t}$, et explorant l'espace beaucoup plus vite que l'ivrogne qui va tantôt dans un sens, tantôt dans l'autre.

Il convient toujours garder à l'esprit le fait que l'approximation par une gaussienne dans la limite des grands N n'est assurément très bonne que dans la région centrale de la distribution (souvent appelée *région d'échelle*). Autrement dit, loin dans les ailes, la vraie distribution de S_N peut différer notablement de son approximation gaussienne ; les moments d'ordre élevés sont sensibles à la formes des ailes et sont parfois mal rendus lorsque l'on s'en tient à l'approximation fournie par le Théorème central limite. En particulier, si la (vraie) loi de S_N a des moments M_k infinis pour k supérieur à un certain k_0 , alors l'approximation gaussienne (qui donne des moments finis $\forall k$) peut conduire à des bêtises, même si elle reproduit bien le centre de la vraie distribution. À nouveau, les événements rares peuvent en fait jouer un rôle déterminant, alors que les événements *typiques* (les plus fréquents) semblent *a priori* plaider en faveur d'un comportement standard.

Ces remarques, nécessaires, n'altèrent la grande importance pratique du Théorème central limite que pour en délimiter l'usage raisonné, mais ne remettent pas fondamentalement en cause l'universalité qu'il sous-tend, et permet d'accepter des hypothèses plausibles comme celle consistant à affirmer que la somme d'un grand nombre de perturbations aléatoires indépendantes est une perturbation gaussienne ; en raison des propriétés remarquables de la distribution de Gauss, une telle hypothèse permet de construire des modèles tractables, dont, à défaut, on ne saurait rien faire. De surcroît, comme Paul Lévy l'a montré, il est susceptible de généralisations diverses (voir par exemple l'ouvrage de Gnedenko et Kolmogorov⁷⁸).

⁷⁷ Dans \mathbb{R}^d , on pose $\Delta r^2 = 2dDt$, si la diffusion est isotrope. Dans (8.194), on a introduit plus naturellement $\Delta X_N \equiv \sigma_{X_N}$.

⁷⁸ B. V. GNEDENKO et A. N. KOLMOGOROV, *Limit distributions for sums of independent random variables* (Addison-Wesley, Cambridge, 1954).