

Partie II : Probabilités

La théorie des probabilités n'est, au fond,
que le bon sens réduit au calcul ;
elle fait apprécier avec exactitude ce que
les esprits justes sentent par une sorte d'instinct,
sans qu'ils puissent souvent s'en rendre compte.

Pierre-Simon Laplace

La théorie des probabilités est née au XVI^{ème} siècle de l'étude quantitative des jeux de hasard⁶ (travaux de Fermat, Pascal, Huyghens). Son étude a été poursuivie aux XVIII^e et XIX^e siècles par de grands mathématiciens, les Bernoulli, Laplace, Gauss, Poincaré, jusqu'au début du XX^{ème}, où, sous la double pression des compagnies d'assurance (évaluation des risques) et des sciences exactes, en particulier de la physique statistique, et humaines, les travaux d'É. Borel, P. Lévy, A.N. Kolmogorov et d'autres lui ont donné sa forme moderne. Étroitement associée aux statistiques, elle est aujourd'hui omniprésente dans les sciences "pures" et appliquées, depuis la physique quantique ou statistique à l'économie et à la finance, aux sciences de la terre et de l'environnement, à la biologie et à l'épidémiologie, etc.

Chapitre 6. Évènements et probabilités

Ce chapitre est consacré à une introduction aux concepts de base de la théorie des probabilités – événements indépendants, probabilités conditionnelles, . . . –, ainsi qu'à quelques notions de combinatoire des dénombrements.

1. Premières définitions

La théorie des probabilités s'intéresse à des événements –accidents, jeux de hasard, phénomènes complexes, caractéristiques de grandes populations biologiques ou physiques etc– qui dépendent de causes mal contrôlées ou non identifiées (le "hasard"). On doit alors remplacer une information déterministe –la certitude de tel événement à tel instant– par une notion statistique, sa *probabilité*.

⁶ Le mot "hasard" vient de l'arabe *az-zhar*", dés, et de même *alea* veut dire dé en latin.

Par exemple, quelle est la probabilité qu'un nouveau-né soit de sexe féminin ? quelle est la probabilité d'un "pile" à un lancer de pièce ? Dans ces deux cas, l'expérience, fondée sur de nombreuses observations, nous donne la réponse : une chance sur deux, c'est-à-dire une probabilité $\frac{1}{2}$. Autre exemple, on dispose d'un sac contenant b_1 billes rouges, b_2 bleues et b_3 blanches, $b_1 + b_2 + b_3 = b$. On effectue une *expérience* consistant à tirer k fois une bille au hasard et à la remettre dans le sac après avoir noté sa couleur. La donnée de l'expérience est donc constituée des nombres k_1 , k_2 et k_3 de billes rouges, bleues et blanches tirées dans l'épreuve, avec $k_1 + k_2 + k_3 = k$. Notre intuition nous dit, et les "lois des grands nombres" que nous établirons plus tard confirment, que quand k devient très grand, les rapports k_i/k tendent vers les rapports des "populations" b_i/b de billes dans le sac. On voit que dans ce cas, l'échantillonnage d'un nombre k suffisamment grand de billes permet d'évaluer la composition de la population du sac. Les sondages d'opinion procèdent de façon analogue. . .

La physique moderne fait un appel constant à des concepts de probabilités et de statistiques. D'une part, dans des situations impliquant un nombre gigantesque d'atomes, d'électrons, de photons, etc –gaz, métal, . . . –, le déterminisme de la physique classique perd son sens : il est inconcevable de donner les conditions initiales et de résoudre les équations du mouvement de $N \sim 10^{23}$ particules ! Il faut remplacer ces données inconnues et ingérables par une information de nature statistique : quelle est la probabilité de tel type de particules d'avoir telle vitesse, etc. C'est sur ce principe qu'est bâtie la Mécanique Statistique, qui fournit une relation entre le monde microscopique des constituants et les propriétés macroscopiques observées.

D'autre part, le monde de la physique quantique est intrinsèquement de nature probabiliste. On y renonce à prédire avec certitude les positions, vitesses ou évolutions des objets étudiés et on se contente de calculer et mesurer leurs probabilités de présence, de transition, etc.

Il est donc fondamental de disposer d'un formalisme mathématique précis et opératoire qui permette de discuter ces questions de probabilités. Pour cela on va devoir préciser un peu les choses pour bien définir de quoi on parle.

Un *événement* est le nom donné au résultat d'une expérience ou d'une observation. Dans les cas les plus simples, il est codé par une variable prenant une valeur binaire 0 ou 1 –pile ou face, garçon ou fille, être vivant ou mort, etc–, ou entière –valeur du dé après un lancer, âge de décès, . . .– ou réelle continue –poids des bébés à la naissance, taille d'une plante, angle de diffusion d'une particule, etc. Mais l'événement peut-être plus complexe, impliquer plusieurs variables (plusieurs lancers de dé, poids *et* âge de décès, etc),

ou dépendre d'autres événements, comme on verra. En général on parlera de l'*espace des épreuves* Ω pour un problème donné, une situation donnée. (Un tel "espace" ne doit pas être confondu avec les espaces vectoriels de la première partie de ce cours !) Un événement A est associé à une partie (un sous-ensemble) Ω_A de Ω , qui rassemble toutes les épreuves conduisant à la réalisation (à l'observation) de A .

Exemples :

1. on s'intéresse au résultat d'un lancer de deux dés. L'espace Ω est l'ensemble de tous les résultats possibles de tels lancers. Si A est l'événement consistant en deux valeurs identiques des 2 dés, il est associé au sous-ensemble $\Omega_A = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$;

2. un projectile est émis depuis l'origine O avec une vitesse \vec{v} proche de l'axe Ox mais qui a une certaine indétermination en direction et en module ; ce projectile peut être classique –une balle de fusil– ou microscopique –un photon ou un électron– et l'incertitude sur \vec{v} peut donc être liée à la qualité du fusil ou de la cartouche dans le premier cas, à des effets quantiques dans le second. On s'intéresse au point où il atteindra le plan $x = a$. L'espace Ω est donc l'ensemble des points d'impact dans ce plan, et peut être décrit par les coordonnées y, z dans le plan $x = a$. Un événement A possible est que le projectile atteint le plan $x = a$ à une distance $r_1 \leq r < r_2$ de l'axe Ox .

◦ *Définitions de la probabilité*

Intuitivement, pour définir la probabilité d'un événement A , on imagine qu'on peut répéter un grand nombre N de fois une expérience où l'événement A se produit ou non. S'il se produit n_A fois, on dira que sa probabilité est

$$\mathbf{P}(A) = \lim_{N \rightarrow \infty} \frac{n_A}{N} . \quad (1.1)$$

La probabilité est alors définie comme une fréquence (fréquence au sens commun de taux de répétition, et non au sens d'un phénomène périodique). Évidemment $0 \leq n_A \leq N$ donc $0 \leq \mathbf{P}(A) \leq 1$.

Mais on peut aussi se donner ces nombres $\mathbf{P}(A)$ a priori, ou les calculer, et on dit que l'espace Ω a été probabilisé. C'est en particulier le cas si l'espace des épreuves Ω a un nombre d'éléments (ou *cardinal*) fini, noté $|\Omega|$ (et si tous ces événements ont la même fréquence d'occurrence). Il en est bien sûr alors de même de tous ses sous-ensembles Ω_A et il est naturel de définir la probabilité de A comme le rapport

$$\mathbf{P}(A) = \frac{|\Omega_A|}{|\Omega|} \quad (1.2)$$

du nombre des épreuves où A se réalise au nombre total d'épreuves, ou comme on dit souvent, "le rapport du nombre de cas favorables au nombre total de cas". Le fait que les deux définitions (1.1) et (1.2) fournissent la même réponse est assez intuitif et sera justifié dans la suite. Ainsi dans l'exemple précédent du lancer de deux dés (non pipés), le calcul se ramène au comptage (ou *dénombrement*) des différentes combinaisons de dés : les $N = |\Omega| = 6^2 = 36$ lancers de Ω ont autant de chances de "sortir" (si les dés sont bien équilibrés), le nombre de "cas favorables" de deux nombres identiques est $|\Omega_A| = 6$ et la probabilité de A est le rapport du nombre de cas favorables au nombre total $\mathbf{P}(A) = \frac{6}{36} = \frac{1}{6}$.

Dans le cas plus général où l'espace d'épreuves Ω n'est pas un ensemble fini, comme c'est le cas quand on s'intéresse par exemple à des quantités entières non bornées (nombre de clients dans un grand magasin, nombre de désintégrations d'un matériau radioactif, nombre de photons dans une cavité, ...) ou continues (distributions de masses, de tailles d'une population animale ou physique, de vitesses comme dans un exemple précédent), la définition *a priori* des probabilités des parties de Ω n'a plus rien d'évident. Donner un sens à de telles probabilités en général relève d'un domaine mathématique ardu (lié à la "théorie de la mesure"). Contentons-nous ici d'admettre que la donnée de probabilités sur l'espace Ω est liée à celle d'une *mesure* sur Ω et sur (certaines de) ses parties Ω_A . Notant $\mu(\Omega)$ et $\mu(\Omega_A)$ ces mesures (positives) de Ω et Ω_A , la probabilité de l'événement A est donnée par la mesure relative de l'ensemble Ω_A par rapport à celle de Ω

$$\mathbf{P}(A) = \frac{\mu(\Omega_A)}{\mu(\Omega)}. \quad (1.3)$$

Intuitivement plus la mesure $\mu(\Omega_A)$ est élevée, c'est-à-dire plus le sous-ensemble Ω_A est "gros", plus l'événement A est probable. Dans le cas d'événements discrets en nombre fini, cette mesure $\mu(\Omega_A)$ est simplement fournie par le nombre d'éléments de Ω_A , comme on vient de le voir. La définition (1.3) englobe et généralise donc (1.2). Nous verrons au chapitre suivant comment procéder quand les épreuves et les événements sont repérés par des variables continues.

Dans les deux définitions (1.1) et (1.2) il est manifeste que $\mathbf{P}(A)$ vérifie les inégalités

$$0 \leq \mathbf{P}(A) \leq 1 \quad (1.4)$$

pour tout événement A . Cette inégalité est fondamentale pour la cohérence du concept de probabilité et on supposera (ou on vérifiera) qu'elle est toujours satisfaite dans les cas plus généraux comme (1.3).

Si $\mathbf{P}(A) = 1$, on dit que l'événement est *presque sûr* (ou presque certain). Si $\mathbf{P}(A) = 0$, on dit qu'il est *presqu'impossible* ou ne se produit *presque jamais*.

Exemple : A est un lancer de dé compris entre 1 et 6 : certain ; $A =$ le dé atterrit (et reste en équilibre !) sur un coin : $\mathbf{P}(A) = 0$.

◦ Histogrammes

Supposons que l'on répète un certain nombre N de fois une expérience (une mesure par exemple) dont le résultat est un nombre (somme T des deux points dans un lancer de deux dés, taille d'un individu, ...). Il est commode de construire un diagramme qui rassemble ces données et leur nombre (ou leur fréquence) d'apparitions. Dans le cas où le résultat x de l'expérience est un nombre entier i borné entre a et b , on porte $x = i$ en abscisse, et en ordonnée $y = n_i$ nombre de fois où le résultat $x = i$ est observé. On obtient l'*histogramme* des données. Une variante est de porter en ordonnée non pas le nombre absolu mais sa fréquence $f = n_i/N$. Selon l'argument/définition donné plus haut en (1.1), la fonction $f(x)$ doit dans la limite $N \gg 1$ donner la probabilité d'observer x , plus précisément $p_i = f(i)$.

Exemple : une série de 200 lancers de deux dés m'a fourni le résultat suivant (5, 3), (1, 3), (5, 4), (6, 4), (4, 1), (1, 3), (6, 6), (1, 1), (6, 6), (4, 3), etc, et les sommes des points des deux dés valent donc 8, 4, 9, 10, 5, 4, 12, 2, 12, 7, etc.

On construit alors l'histogramme suivant

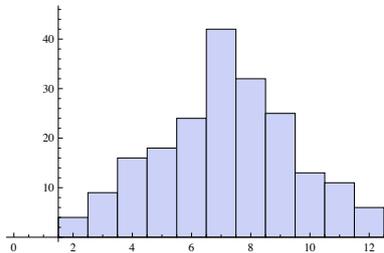


Fig. 16: Histogramme du total de deux dés

qui se lit ainsi : il y a 4 “événements” de total $x = 2$, 9 pour $x = 3$, etc, 6 pour $x = 12$.

On verra au chapitre suivant un exemple d'histogramme pour des quantités continues.

2. Propriétés et axiomes

2.1. Événements composés

Attention ! pour alléger les notations, nous noterons dans la suite par le même symbole A , B , etc, un événement et le sous-ensemble de Ω qui lui est associé (noté Ω_A ... au

paragraphe précédent). Supposons donc qu'à toute partie A de l'espace Ω , on a associé une probabilité $\mathbf{P}(A)$. On peut donc appliquer les notions et notations de la théorie des ensembles aux combinaisons d'événements et parler des événements \emptyset , $A \cup B$ et $A \cap B$ etc.

▷. $A \subset B$ signifie du point de vue de la théorie des ensembles que A est un sous-ensemble de B et du point de vue des événements que A implique B . Exemple : “ A = taille d'une plante inférieure à 50 cm” \subset “ B = taille \leq 100 cm”.

▷. $A = \emptyset$. Il est naturel de postuler que l'événement \emptyset a la probabilité $\mathbf{P}(\emptyset) = 0$.

▷. De même pour $A = \Omega$, (qui représente un événement quelconque de l'ensemble !) $\mathbf{P}(\Omega) = 1$, en accord avec les définitions (1.1), (1.2) ou (1.3). L'événement Ω est certain.

▷. $\bar{A} := \Omega \setminus A$, le complémentaire de A , représente l'événement contraire : \bar{A} se produit ssi A ne se produit pas ; c'est l'événement (non- A).

▷. $A \cap B$ (noté parfois $A \cdot B$) : les événements A et B se produisent, c'est l'événement (A et B).

Une définition très importante en probabilités est celle d'*événements indépendants*.

Définition : Deux événements A et B sont indépendants si et seulement si

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B) . \quad (2.1)$$

▷. A et B sont *mutuellement exclusifs* (ou “ils s'excluent mutuellement”) si $\mathbf{P}(A \cap B) = 0$: l'événement (A et B) est presque impossible. Exemple : A = taille \leq 50 cm, B = taille \geq 60 cm.

▷. $A \cup B$: événement A ou B (ou les deux).

2.2. Espace d'épreuves fini

Dans le cas d'un espace d'épreuves Ω ayant un nombre fini d'éléments, la définition de la probabilité d'un événement (un sous-ensemble) A de Ω se ramène au calcul du nombre de ses éléments, comme on l'a vu dans (1.2).

On peut donc préciser alors les probabilités des événements composés précédents :

▷. $A \subset B$ implique $|A| \leq |B|$ donc aussi $\mathbf{P}(A) \leq \mathbf{P}(B)$.

▷. $A = \emptyset$, $|\emptyset| = 0$ donc $\mathbf{P}(\emptyset) = 0$ est tout naturel.

▷. De même, $\mathbf{P}(\Omega) = |\Omega|/|\Omega| = 1$.

▷. $|\bar{A}| := |\Omega| - |A|$, donc $\mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A)$.

On a alors le

Theorème des probabilités totales : $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$

qui découle de la même propriété pour le nombre d'éléments dans un ensemble fini.

En particulier, si A et B sont mutuellement exclusifs, $\mathbf{P}(A \cap B) = 0$, $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$.

2.3. Exemples

1. On joue à la roulette (36 numéros de 1 à 36 plus le 0) et on parie sur $A = \text{“pair”}$ (avec 0 exclu) et sur $B = \text{“manque”}$ (numéro entre 1 et 18). Quelle est la probabilité de gain ? On a $\mathbf{P}(A) = 18/37$, $\mathbf{P}(B) = 18/37$, $A \cap B = \{2, 4, 6, 8, 10, 12, 14, 16, 18\}$, $\mathbf{P}(A \cap B) = 9/37$, donc $\mathbf{P}(A \cup B) = (18 + 18 - 9)/37 = 27/37$.

2. Dans un tirage de 13 cartes dans un paquet de 52 cartes bien battues, soit $A = \text{présence d’un as}$, $P = \text{présence d’un pique}$, $A \cap P = \text{présence d’un as et d’un pique}$ (qui peut être ou non l’as). Le calcul de $\mathbf{P}(A)$, $\mathbf{P}(P)$, $\mathbf{P}(A \cap P)$, $\mathbf{P}(A \cup P)$ n’est pas si aisé. . .

2.4. Espace probabilisé

Dans le cas où l’espace d’épreuves Ω n’est pas un ensemble fini, on a vu plus haut que le concept de probabilités sur Ω est lié à celui de mesure sur cet ensemble. En général on dira qu’on a probabilisé l’espace Ω si on a pu définir une famille \mathcal{F} de ses parties (sous-ensembles) A et leur attacher un nombre $0 \leq \mathbf{P}(A) \leq 1$ avec les axiomes suivants (“axiomes de Kolmogorov”)

▷. $\emptyset \in \mathcal{F}$, et $\mathbf{P}(\emptyset) = 0$;

▷. $\Omega \in \mathcal{F}$ et $\mathbf{P}(\Omega) = 1$;

▷. Si $A \in \mathcal{F}$, son complémentaire $\bar{A} := \Omega \setminus A \in \mathcal{F}$, et $\mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A)$;

▷. Si $A, B \in \mathcal{F}$, $A \subset B$ implique $\mathbf{P}(A) \leq \mathbf{P}(B)$;

▷. Si $A, B \in \mathcal{F}$, $A \cap B$ et $A \cup B \in \mathcal{F}$ et $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$;

toutes propriétés qui généralisent celles que nous venons de voir pour un ensemble fini. Nous y reviendrons au chapitre suivant.

2.5. Probabilités conditionnelles

Soient deux événements A et B . On définit la *probabilité conditionnelle* $\mathbf{P}(A|B)$ comme la probabilité de réaliser l’événement A quand B est déjà réalisé. On dit que $\mathbf{P}(A|B)$ est la “probabilité de A si B ” (ou la “probabilité de A sachant B ”). On a la formule

Axiome de Bayes

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} .$$

(2.2)

Cette proposition se démontre aisément pour un espace d’épreuves Ω qui est un ensemble fini. En effet par l’hypothèse “ B s’est réalisé”, on doit restreindre notre espace

d'épreuves au sous-ensemble de Ω où B est réalisé. Son cardinal est $\mathbf{P}(B) \cdot |\Omega|$. Celui de l'espace des événements recherchés ("cas favorables") est $\mathbf{P}(A \cap B) \cdot |\Omega|$. Leur rapport, qui est la probabilité conditionnelle, est bien le rapport (2.2). En général, pour un espace Ω quelconque, il faut supposer que la mesure des différentes parties de Ω satisfait bien (2.2), et on préfère parler d'"axiome", c'est-à-dire de principe indémontrable que l'on ajoute aux propriétés du § 2.4.

Noter que $\mathbf{P}(A|B)\mathbf{P}(B) = \mathbf{P}(B|A)\mathbf{P}(A) = \mathbf{P}(A \cap B)$ qui est bien symétrique en A et B . Noter encore que dans le cas particulier où A et B sont indépendants, donc $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$, on a $\mathbf{P}(A|B) = \mathbf{P}(A)$, ce qui correspond bien à l'intuition : l'occurrence de l'événement A est indépendante de l'occurrence ou non de l'événement B .

Exemple. Dans un lancer d'un dé, probabilité d'avoir un nombre multiple de 3, sachant qu'on a un nombre pair ? $B = \{2, 4, 6\}$, $\mathbf{P}(B) = 1/2$, $A = \{3, 6\}$, $A \cap B = \{6\}$, $\mathbf{P}(A \cap B) = 1/6$, $\mathbf{P}(A|B) = 1/3$.

Theorème de Bayes. Si A et B sont deux événements de probabilité non nulle,

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A)\mathbf{P}(B|A)}{\mathbf{P}(B)}. \quad (2.3)$$

Cette formule qui relie la probabilité d'avoir A sachant B à la probabilité réciproque d'avoir B sachant A peut être vue comme une sorte de formule d'inversion des causalités.

3. Un peu de combinatoire

Dans les questions de probabilités portant sur des variables discrètes, on est souvent amené à un problème de dénombrement : combien d'éléments (épreuves) y a-t-il dans l'espace Ω , combien d'éléments dans un de ses sous-ensembles, donc de réalisations de l'événement correspondant ?

Par exemple, avec un lancer de trois dés, combien de configurations des trois dés donnent un total de points donné ?

◦ *Permutations. Arrangements et combinaisons.*

Rappelons d'abord qu'étant donné un ensemble \mathcal{E} de n éléments, éléments qu'on suppose distincts et numérotés de 1 à n , (par exemple, n boules portant les numéros de 1 à n), le nombre de tirages ordonnés de ces n éléments (de ces n boules) est le nombre de *permutations* de ces n objets, soit $n! = n \cdot (n-1) \cdots 2 \cdot 1$. En effet il y a n choix pour le premier, $n-1$ choix pour le second, etc.

Le nombre de tirages ordonnés de p numéros quelconques est, par le même argument

$$A_n^p = n(n-1)\cdots(n-p+1) = \frac{n!}{(n-p)!}. \quad (3.1)$$

(n choix pour le premier, $n-1$ choix pour le second, \dots , $n-p+1$ pour le p -ième). La seconde expression peut aussi se comprendre de la façon suivante : A_n^p est le nombre total $n!$ de tirages de n éléments divisé par le nombre des tirages des $n-p$ éléments restants, qu'on n'effectue pas. Le nombre A_n^p est appelé *nombre d'arrangements* de p objets choisis parmi n .

Supposons maintenant qu'on s'intéresse aux tirages non ordonnés de ces p éléments (p boules). On compte alors les *combinaisons* de p éléments choisis parmi n . Leur nombre est

$$C_n^p = \frac{n(n-1)\cdots(n-p+1)}{p(p-1)\cdots 2 \cdot 1} = \frac{n!}{p!(n-p)!}. \quad (3.2)$$

Cette expression s'obtient en divisant A_n^p par le nombre $p!$ d'ordres différents des p éléments tirés, puisque cet ordre n'importe pas.

Attention, ce nombre C_n^p est noté $\binom{n}{p}$ dans la littérature anglo-saxonne, y compris dans les commandes de certains logiciels... (par exemple `Binomial[n,p]` en Mathematica).

Exemple. On a 12 boules distinctes dans une urne. Le nombre total de tirages de 5 boules où par exemple le tirage $\{1, 3, 8, 7, 4\}$ est distingué de $\{1, 3, 4, 7, 8\}$ est $A_{12}^5 = 12!/7! = 95040$. La probabilité d'un tirage particulier (un ordre donné de 5 boules données) est donc $1/95040$. Si on ne s'intéresse pas à l'ordre, le nombre de tirages différents est $C_{12}^5 = A_{12}^5/5! = 792$. La probabilité d'un tirage donné est maintenant $1/792$, qui est $5! = 120$ fois plus élevée que la précédente.

Attention, ces calculs supposent que les n objets sont distincts et qu'on les retire du choix à chaque tirage. S'ils ne sont pas distincts, ou si on les remet en jeu (on remet les boules dans l'urne), le calcul est différent ! (voir TD)

Exercice : le tiercé. Imaginons la situation totalement irréaliste où une course hippique met en piste $n = 18$ chevaux de force égale (et où la course n'est pas truquée). Quelle est la probabilité de gagner le tiercé (3 chevaux gagnants dans l'ordre) ?, celle de gagner le tiercé dans le désordre ? *Réponse:* Nombre de tiercés possibles $A_{18}^3 = 4896$, $\mathbf{P}(\text{tiercé}) = \frac{1}{A_{18}^3} \approx 2 \cdot 10^{-4}$; $\mathbf{P}(\text{tiercé dans le désordre}) = \frac{1}{C_{18}^3} \approx 1,2 \cdot 10^{-3}$.

◦ *Propriétés des C_n^p*

Les C_n^p sont définis pour $1 \leq p \leq n$ mais on peut les étendre à tout p entier positif ou nul : $C_n^0 = 1$, et $C_n^p = 0$ si $p > n$. Ils satisfont les propriétés suivantes :

Chapitre 7. Variables aléatoires

Le chapitre précédent a été consacré à des situations où l'événement pouvait être décrit par une ou plusieurs variables discrètes : valeur du dé, numéro du cheval gagnant etc. On va maintenant examiner des situations où il faut faire appel à une variable continue X . Par exemple, la taille ou le poids d'enfants à un certain âge, le volume d'eau tombée par m^2 à un certain site pendant un certain temps, etc.

En pratique, une fois qu'on a enregistré des données concernant de tels événements, il est commode de construire leur *histogramme* : on divise l'ensemble des résultats possibles en intervalles de la droite réelle d'une taille choisie à l'avance, $I_n = [a + n\Delta, a + (n + 1)\Delta[$, et on porte verticalement au dessus de chaque intervalle le nombre d'événements qui ont fourni un résultat tombant dans cet intervalle I_n .

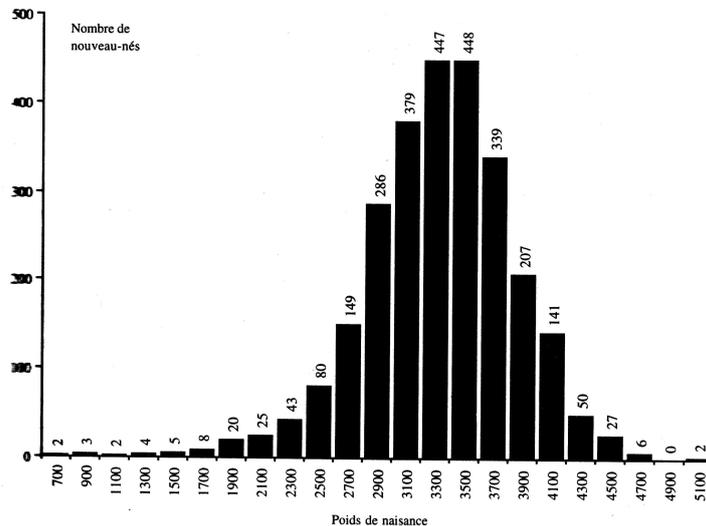


Fig. 17: Histogramme des poids de 2673 nouveaux nés dans une maternité du Bas-Rhin.

Exemples. Poids des 2673 bébés nés à la maternité d'Haguenau (Bas-Rhin) en 1984-85. L'histogramme de la figure 17 regroupe les poids de 200g en 200g : 80 bébés sont nés avec un poids dans l'intervalle de 2400 à 2600 g, etc. Celui de la figure 18 collecte les 90 notes sur 20 du contrôle continu de l'UE LP207 d'avril 2012 en 19 intervalles de point en point...

Pour des variables continues, le choix de l'intervalle (*bin* en français) n'est pas unique : on a un certain arbitraire sur le centre et la largeur de l'intervalle. C'est tout l'art du statisticien de trouver les bonnes valeurs qui donneront un aspect régulier à son histogramme...

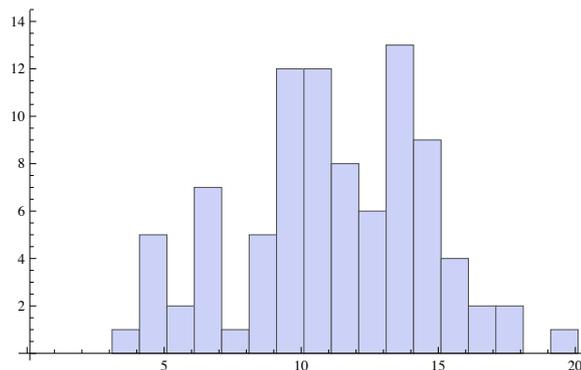


Fig. 18: Histogramme des notes du contrôle continu de l'UE 207, avril 2012

1. Variables aléatoires. Distributions de v.a.

1.1. Définition d'une variable aléatoire

On s'intéresse donc maintenant au cas où l'espace d'épreuves est la droite réelle $\Omega = \mathbb{R}$. Les sous-ensembles de Ω qu'on va pouvoir "probabiliser" sont certains types d'intervalles, typiquement des intervalles de la forme $[a, b[$, ouverts à droite, c'est-à-dire des ensembles des réels $a \leq x < b$. Un "événement" élémentaire est donc un tel intervalle et on suppose qu'on connaît les probabilités $\mathbf{P}([a, b[)$ pour tous les intervalles $[a, b[$, avec éventuellement le point a rejeté à moins l'infini $(-\infty, b[$. On interprète cette probabilité comme décrivant la distribution d'une quantité réelle X (taille, poids, volume d'eau dans les exemples ci-dessus), appelée *variable aléatoire* (v.a.), et la probabilité de l'intervalle $[a, b[$ est la probabilité que cette variable X prenne des valeurs dans cet intervalle

$$0 \leq \mathbf{P}(a \leq X < b) \leq 1 . \quad (1.1)$$

Ces probabilités satisfont tous les axiomes qu'on a détaillés au chapitre précédent (Chap. 6, § 2.4). En particulier

- a) $\mathbf{P}(\mathbb{R}) = \mathbf{P}(-\infty < X < \infty) = 1$ (il est certain que X prend une valeur réelle)
- b) quels que soient $a < b < c$,

$$\mathbf{P}(a \leq X < c) = \mathbf{P}(a \leq X < b) + \mathbf{P}(b \leq X < c) \quad (1.2)$$

puisque ces deux intervalles ont une intersection vide, etc.

L'ensemble de ces $\mathbf{P}([a, b[)$ définit ce qu'on appelle la *loi de probabilité* de la v.a. X .

Remarques

- Noter que les données contenues dans un histogramme comme dans les exemples précédents devraient permettre de définir les probabilités \mathbf{P} pour les intervalles choisis I_n , par $\mathbf{P}(X \in I_n) = \# \text{ événements dans l'intervalle } I_n / \# \text{ total d'événements}$. Cela devrait se justifier dans la limite où ce nombre total d'événements est grand.
- Bien entendu, la v.a. X peut ne prendre ses valeurs que dans un sous-ensemble strict de \mathbb{R} . Ainsi par exemple, les ordinateurs sont pourvus de programmes *générateurs de nombres aléatoires* compris entre 0 et 1. Pour une telle v.a., $\mathbf{P}(X < x) = 0$ pour tout $x < 0$ et $\mathbf{P}(X < x) = 1$ pour tout $x > 1$.
- Il est clair aussi que le concept de v.a. englobe les cas de variables discrètes : ainsi pour une v.a. ne prenant que des valeurs entières, $\mathbf{P}(X = k) = p_k$ au sens où on l'a vu au chapitre précédent.

1.2. Les fonctions importantes attachées à une v.a.

On définit alors la *fonction de répartition*⁷ $F(x)$

$$F(x) = \mathbf{P}(X < x) \quad (1.3)$$

correspondant donc à un intervalle $(-\infty, x[$. Il découle de (1.2) que cette fonction est monotone croissante : si $x < y$, $F(x) \leq F(y)$.

Cette fonction peut être discontinue. C'est en particulier le cas pour une v.a. discrète, comme celles considérées au chapitre précédent. Exemple : la valeur d'un dé étant considérée comme v.a., la fonction de répartition est discontinue à chaque entier.

Si on définit la probabilité avec une inégalité "large" $\mathbf{P}(X \leq a) = \lim_{x \rightarrow a+0} \mathbf{P}(X < x) = \lim_{x \rightarrow a+0} F(x) \equiv F(a+0)$ par une limite, on voit que $\mathbf{P}(X = a) = \mathbf{P}(X \leq a) - \mathbf{P}(X < a) = F(a+0) - F(a)$, la discontinuité de F en a est la probabilité que $X = a$.

Supposons maintenant que la fonction $F(x)$ est non seulement continue mais aussi dérivable. Soit $f(x) = F'(x)$. Inversement $F(x)$ est la primitive de $f(x)$ s'annulant à $x \rightarrow -\infty$, donc

$$F(x) = \int_{-\infty}^x f(x') dx' , \quad (1.4)$$

et en général

$$\mathbf{P}(a \leq X \leq b) = \int_a^b f(x) dx . \quad (1.5)$$

⁷ appelée aussi "fonction de distribution cumulative", le danger étant que "fonction de distribution" est aussi utilisé dans un autre sens, voir plus bas.

La fonction f , appelée *densité de probabilité*⁸, est positive (puisque F est croissante) et telle que $\mathbf{P}(\mathbb{R}) = \int_{-\infty}^{\infty} f(x) dx = 1$, on dit qu'elle est *normalisée*. Noter que $f(x) dx$ représente la probabilité que la v.a. X appartienne à l'intervalle $[x, x + dx]$.

Remarque. Bien comprendre que dans ce cas d'une v.a. continue X à fonction de répartition F dérivable, on ne parle pas de la probabilité que X prenne une *valeur* x , mais seulement qu'elle soit dans un *intervalle*, qu'il soit fini $[a, b[$, infini $(-\infty, a[$ ou infinitésimal $[x, x + dx]$.

Graphes d'une loi f . Il est commode de visualiser une loi de probabilité par le graphe de la fonction f . Noter la relation entre ce graphe et les histogrammes que l'on peut construire à partir d'échantillons de N événements : tout histogramme à N fini donne (à un facteur N près) une approximation du graphe de f . Intuitivement, on s'attend à ce que dans la limite où $N \rightarrow \infty$ et où la largeur des intervalles tend vers 0, l'histogramme approche le graphe, puisque $f(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h}$, donc pour h petit, $f(x)h \approx \mathbf{P}(x \leq X < x + h)$, mais cela mériterait d'être justifié plus précisément ...

Fonction caractéristique.

Un outil très utilisé en probabilités mais qui fait appel à des mathématiques pas encore étudiées en L2 est la *fonction caractéristique* de la v.a., définie comme la *transformée de Fourier* de la densité f , c'est-à-dire

$$\varphi(u) = \int_{-\infty}^{\infty} e^{iux} f(x) dx . \quad (1.6)$$

Cette intégrale converge absolument puisque $|\int e^{iux} f(x) dx| < \int |e^{iux}| f(x) dx = \int f(x) dx = 1$. On démontre qu'inversement la densité f est complètement déterminée par φ grâce à la formule suivante (théorème d'inversion de Fourier)

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} \varphi(u) du . \quad (1.7)$$

L'utilité de cette fonction caractéristique tient au théorème suivant :

La somme de N v.a. indépendantes a pour fonction caractéristique le produit des fonctions caractéristiques,
et en particulier

La somme de N v.a. indépendantes et de même loi f a pour fonction caractéristique $\varphi_N(u) = \varphi(u)^N$.

On peut la définir aussi dans le cas discret : soit X une v.a. discrète prenant la valeur entière k avec la probabilité p_k . On a $\varphi(u) = \sum_k p_k e^{iuk}$ ou, de façon équivalente, $\phi(z = e^{iu}) = \sum_k p_k z^k$. Sous cette dernière forme, comme fonction de z , on l'appelle aussi *fonction génératrice* des probabilités p_k . En effet on retrouve les p_k à partir de cette fonction génératrice par différentiation

$$p_k = \frac{1}{k!} \left(\frac{d}{dz} \right)^k \phi(z) \Big|_{z=0} . \quad (1.8)$$

Exemple: loi binomiale. Une v.a. prend la valeur 0 avec la probabilité p et la valeur 1 avec la probabilité $1 - p$, donc $\phi(z) = p + (1 - p)z$. La somme de N v.a. indépendantes et de même loi a pour fonction caractéristique $\phi_N(z) = \phi(z)^N = \sum_{k=0}^N C_N^k p^{N-k} (1 - p)^k z^k$. Cela donne la réponse à un problème posé au chapitre précédent : si on tire à pile (0) ou face (1), quelle est la probabilité d'un score total k après N tirages? Réponse, si $p = 1 - q = \frac{1}{2}$, $p_k = (\frac{1}{2})^N C_N^k$. On reviendra sur cette loi binomiale au chapitre suivant.

⁸ appelée aussi *probability distribution fonction*, ou PDF, en anglais. . .

1.3. Plusieurs variables aléatoires

Ce qui vient d'être dit pour une v.a. continue peut être répété pour deux v.a. : l'espace $\Omega = \mathbb{R}^2$ peut être visualisé comme le plan muni de deux coordonnées x et y , les deux v.a. X, Y prenant des valeurs x et y avec une certaine loi décrivent un point aléatoire dans le plan.

La fonction de répartition $F(x, y)$ est alors définie comme

$$F(x, y) = \mathbf{P}(X < x, Y < y) ; \quad (1.9)$$

quand F est continue et dérivable, on lui associe la densité

$$f(x, y) dx dy = \mathbf{P}(x < X \leq x + dx, y < Y \leq y + dy) , \quad (1.10)$$

et $F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(x, y) dx dy$.

Si f est factorisée, $f(x, y) = f_1(x)f_2(y)$, les deux v.a. X et Y sont indépendantes, puisque

$$\mathbf{P}(a \leq X < b, c \leq Y < d) = \int_a^b dx \int_c^d dy f(x, y) = \int_a^b f_1(x) dx \int_c^d f_2(y) dy = \mathbf{P}(a \leq X < b) \mathbf{P}(c \leq Y < d)$$

qui est bien la définition de l'indépendance, voir chap. 6, (2.1).

On passe sans difficulté de 2 à n v.a. X_1, \dots, X_n . On rencontrera souvent dans la suite le cas de n v.a. indépendantes et de même loi, ou "également distribuées"⁹, donc dotées d'une densité factorisée $f_n(x_1, \dots, x_n) = \prod_{i=1}^n f(x_i)$.

2. Moyenne, variance, écart-type, moments

2.1. Moyenne

Pour une v.a. discrète X prenant la valeur x_i avec la probabilité p_i , il est naturel et d'usage commun de définir sa *moyenne* par

$$\langle X \rangle = \bar{X} = EX = \sum_i p_i x_i . \quad (2.1)$$

⁹ Les probabilistes anglo-saxons parlent de variables *i.i.d*, *independent, identically distributed*.

On a introduit ici les trois principales notations en usage : les crochets $\langle \cdot \rangle$, très prisés des physiciens, la barre, utilisée quand les crochets prêtent à ambiguïté, et le E (pour espérance mathématique, ou *expectation*) très en vogue chez les mathématiciens. . . Plus généralement pour toute fonction $G(x)$, la moyenne ou espérance de cette fonction pour la v.a. X est

$$\langle G(X) \rangle = \overline{G(X)} = EG(X) = \sum_i p_i G(x_i) . \quad (2.2)$$

Pour une v.a. X continue dotée d'une densité f , on définit de même

$$\begin{aligned} \langle X \rangle &= \overline{X} = EX = \int_{-\infty}^{\infty} x f(x) dx \\ \langle G(X) \rangle &= \overline{G(X)} = EG(X) = \int_{-\infty}^{\infty} G(x) f(x) dx . \end{aligned} \quad (2.3)$$

Exercice. Soient X et Y deux v.a. indépendantes. Montrer que $\langle X.Y \rangle = \langle X \rangle \langle Y \rangle$. Inversement si $\langle X.Y \rangle \neq \langle X \rangle \langle Y \rangle$, cela implique que X et Y ne peuvent être indépendantes.

Remarques : 1. En Mécanique, on introduit la notion de *barycentre* (ou centre de gravité) de N points massifs de coordonnées \vec{x}_i et de masse m_i par $\vec{x}_B = \frac{1}{M} \sum_i m_i \vec{x}_i$, $M = \sum_i m_i$, ou dans le cas d'une distribution continue de masse de densité massique $\rho(\vec{x})$,

$$\vec{x}_B = \frac{\int d^3 \vec{x} \rho(\vec{x}) \vec{x}}{\int d^3 \vec{x} \rho(\vec{x})} . \quad (2.4)$$

Bien noter l'analogie entre les moyennes des formules (2.1)-(2.3) et (2.4). Le rôle des probabilités p_i ou de la densité $f(x)$ est joué ici par les rapports m_i/M ou $\rho(\vec{x})/\int d^3 \vec{x} \rho(\vec{x})$.

2. Soit X une v.a., on définit la nouvelle v.a. $X' = X - \langle X \rangle$: il est clair que $\langle X' \rangle = 0$, et on dit que la v.a. X' est *centrée*.

2.2. Variance et écart-type

La *variance* d'une v.a. X est définie comme la moyenne de $(X - \langle X \rangle)^2$, soit

$$\text{var } X = \sigma_X^2 = \langle (X - \langle X \rangle)^2 \rangle . \quad (2.5)$$

À nouveau, les deux notations en vigueur ont été données. L'*écart-type* (ou déviation standard) est $\sigma_X = \sqrt{\text{var } X}$.

Intuitivement l'écart-type est une mesure de l'étalement des valeurs de la v.a. autour de sa moyenne. Une variance (donc un écart-type) faible signale des valeurs de la v.a. resserrées autour de la moyenne $\langle X \rangle$, avec donc un histogramme pointu. À l'inverse, une variance élevée signifie que X peut prendre (avec une probabilité non négligeable) des valeurs éloignées de $\langle X \rangle$.

Démontrons une utile formule pour la variance. Développons $(X - \langle X \rangle)^2 = X^2 - 2X\langle X \rangle + \langle X \rangle^2$. La moyenne étant une opération linéaire, son calcul peut se faire terme à terme, donc $\langle (X - \langle X \rangle)^2 \rangle = \langle X^2 \rangle - 2\langle X \rangle^2 + \langle X \rangle^2 = \langle X^2 \rangle - \langle X \rangle^2$, et finalement

$$\boxed{\text{var } X = \sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2 .} \quad (2.6)$$

Remarques.

- La variance est insensible à une translation de la v.a. X par une constante c . En effet, si $X' = X + c$, $\langle X' \rangle = \langle X \rangle + c$ et $X - \langle X \rangle = X' - \langle X' \rangle$, donc les variances de X et X' sont les mêmes. Par une dilatation par un facteur λ , la moyenne et l'écart-type sont dilatés par ce même facteur $X \mapsto X' = \lambda X$, $\langle X' \rangle = \lambda \langle X \rangle$, $\text{var } X' = \lambda^2 \text{var } X$, donc $\sigma_X / \langle X \rangle$ est invariant.
- Les définitions de $\langle X \rangle$ et de $\text{var } X$ qui impliquent des intégrales “impropres” sur $(-\infty, \infty)$ n'ont de sens que si ces intégrales convergent. La convergence à l'infini de la moyenne de X^2 est évidemment plus contraignante que celle de X : pour certaines distributions de v.a., la moyenne peut exister sans que la variance existe, voir exemples au chapitre suivant.

2.3. Moments

La définition des moments généralise celles de la moyenne et de la variance. En général, pour tout k entier positif, on définit le k -ième moment d'une v.a. comme

$$m_k = \langle X^k \rangle = \int_{-\infty}^{\infty} x^k f(x) dx \quad (2.7)$$

si cette intégrale existe (converge).

Pour une v.a. discrète, on écrit simplement, avec les notations ci-dessus

$$m_k = \sum_i x_i^k p_i .$$

On verra au chapitre suivant des exemples explicites de “lois” et des calculs détaillés de leurs moments.

◦ *Inégalité de Tchebychev*

Supposons que l'intégrale du k -ième moment converge absolument, $\langle |X|^k \rangle < \infty$. On peut borner supérieurement $\mathbf{P}(|X| > a)$ par

$$\mathbf{P}(|X| > a) = \int_{|x|>a} f(x) dx \leq \int_{|x|>a} \left| \frac{x}{a} \right|^k f(x) dx \leq \int_{-\infty}^{\infty} \left| \frac{x}{a} \right|^k f(x) dx = \left\langle \left| \frac{X}{a} \right|^k \right\rangle \quad (2.8)$$

L'inégalité $\mathbf{P}(|X| > a) \leq \langle |\frac{X}{a}|^k \rangle$ est appelée inégalité de Tchebychev. Elle nous dit que quand a croît, la probabilité que $|X| > a$ décroît au moins comme a^{-k} , pour tout k pour lequel le k -ième moment existe. Cette inégalité nous sera utile au Chapitre 9.

3. Fonctions de corrélation de plusieurs v.a.

Quand on dispose de plusieurs v.a., comme X et Y au § 1.3 ci-dessus, on s'intéresse à leur *fonction de corrélation* définie par

$$\langle XY \rangle = \int f(x, y) xy dx dy . \quad (3.1)$$

Sous l'hypothèse de factorisation de la densité $f(x, y) = f_1(x)f_2(y)$, donc d'indépendance des v.a. X et Y , on a

$$\langle XY \rangle = \langle X \rangle \langle Y \rangle .$$

Le calcul de

$$\langle XY \rangle_c = \langle XY \rangle - \langle X \rangle \langle Y \rangle \quad (3.2)$$

fournit donc une mesure de la non-indépendance, ou de la corrélation, des v.a. X et Y . Plus généralement, on peut bien sûr définir des fonctions de corrélation plus compliquées, telles $\langle P(X)Q(Y) \rangle$ où P et Q sont par exemple des polynômes, etc.

Exemple physique : dans un barreau de fer, l'aimantation est due aux moments magnétiques \vec{m}_i portés par les \mathcal{N} atomes, plus précisément à l'effet coopératif de ces moments microscopiques. Selon les lois de la Mécanique Statistique, on peut décrire un tel système par une approche probabiliste. À chaque configuration de moments magnétiques $\{\vec{m}_i\}$, on associe son énergie $\mathcal{E}(\{\vec{m}_i\})$ qui dépend de tous ces moments et de leurs interactions, et aussi d'éventuels champs magnétiques appliqués, etc. Le principe fondamental de la Mécanique Statistique est qu'à l'équilibre, à une température T (température absolue, mesurée en degrés Kelvin), la probabilité d'une telle configuration est

$$\mathbf{P}(\{\vec{m}_i\}) = \frac{1}{Z} e^{-\beta \mathcal{E}(\{\vec{m}_i\})} \quad (3.3)$$

où $\beta := \frac{1}{k_B T}$, avec k_B la constante de Boltzmann, $k_B = 1,380 \cdot 10^{-23} \text{ J K}^{-1}$. Le facteur exponentiel $e^{-\beta \mathcal{E}(\{\vec{m}_i\})}$ est appelé *poids de Boltzmann* de la configuration. Le facteur Z , appelé *fonction de partition* du système, est défini par

$$Z = \sum_{\text{config.}} e^{-\beta \mathcal{E}(\{\vec{m}_i\})} \quad (3.4)$$

où on somme sur toutes les configurations des moments magnétiques, de telle façon que la probabilité (3.3) est bien normalisée : $\sum_{\text{config.}} \mathbf{P}(\{\vec{m}_i\}) = 1$. Le résultat de l'analyse théorique, en accord avec l'observation, est que dans la *limite thermodynamique* d'un grand nombre \mathcal{N} de moments et à haute température les moments microscopiques sont désordonnés, "pointant" dans toutes les orientations, l'aimantation totale $\vec{M} = \langle \sum_i \vec{m}_i \rangle$ est nulle et le corps est dans sa *phase paramagnétique* ; à température inférieure à la *température de Curie*, en revanche, les moments magnétiques ont tendance à s'orienter parallèlement les uns aux autres, créant ainsi une aimantation macroscopique \vec{M} non nulle : on est dans la *phase ferromagnétique*. La valeur moyenne $\langle \vec{m}_i \rangle$ du moment de l'atome i est nulle dans la phase paramagnétique, et non nulle et égale à $\frac{1}{\mathcal{N}} \vec{M}$ (par définition) dans la phase ferromagnétique. La fonction de corrélation $\langle \vec{m}_i \vec{m}_j \rangle$ des moments de deux atomes i et j distants s'annule avec leur séparation dans la phase paramagnétique, elle tend vers le carré de $\langle \vec{m} \rangle$ dans la phase ferromagnétique.

Remarque : $\langle XY \rangle_c = 0$ (défini en (3.2)) est une condition *nécessaire* pour que les v.a. X et Y soient indépendantes, mais elle n'est pas *suffisante* : elle n'implique pas que X et Y sont indépendantes. Ainsi soit φ une v.a. angulaire prenant des valeurs sur $[0, 2\pi[$ avec une densité uniforme $f(\varphi) = \frac{1}{2\pi}$. Soient les v.a. $X = \cos \varphi$ et $Y = \sin \varphi$. Montrer que $\langle XY \rangle_c = 0$. Il est pourtant clair que X et Y ne sont pas indépendantes, étant toutes deux liées à la v.a. φ . Exercice : montrer dans cet exemple que $\langle X^2 Y^2 \rangle \neq \langle X^2 \rangle \langle Y^2 \rangle$, ce qui vérifie la non-indépendance de X et Y .

4. Changement de variable aléatoire

On considère un phénomène aléatoire dont les événements sont décrits par une v.a. X réelle, par exemple une coordonnée d'un point M . Pour une raison ou une autre, on est intéressé à changer la description du phénomène en passant de la variable X à une autre, par exemple $Y = aX + b$, $a > 0$ ou toute autre fonction (monotone croissante)¹⁰ $Y = \phi(X)$.

¹⁰ afin que la fonction $\phi : x \mapsto y$ soit inversible, $x = \phi^{-1}(y)$

Comment la loi de X se modifie-t-elle par ce changement de variable (ce changement de coordonnée) ? Autrement dit, quelle est la loi de Y ? Appelons $f(x)$ et $\tilde{f}(y)$ les densités de probabilité qui décrivent le phénomène dans les deux variables. La quantité $f(x)dx$ qui fournit la probabilité que $x \leq X < x + dx$ doit être indépendante du choix de variable, en ce sens que

$$\mathbf{P}(M \in I) = \mathbf{P}(x < X < x + dx) = \mathbf{P}(y < Y < y + dy)$$

donc

$$\begin{aligned} f(x) dx &= \tilde{f}(y) dy & \text{avec } y &= \phi(x) \text{ et } dy = \phi'(x) dx \\ \text{donc } f(x) &= \tilde{f}(y) \frac{dy}{dx} = \tilde{f}(y) \phi'(x) & \iff \tilde{f}(y) &= \frac{f(x)}{\phi'(x)}. \end{aligned} \quad (4.1)$$

Noter que l'on a supposé explicitement que y est une fonction croissante de x , sinon que faudrait-il faire ?

Exemples

1. Un point aléatoire sur le segment $[-1, 1]$ est décrit par une coordonnée X ou par un angle T , $-\frac{1}{2}\pi \leq T \leq \frac{1}{2}\pi$, avec $X = \sin T$. Les densités sont reliées par $\tilde{f}(t) = f(x) \frac{dx}{dt} = f(\sin t) \cos t$ ou $f(x) = \frac{\tilde{f}(t)}{\sqrt{1-x^2}}$.
2. Passage en coordonnées radiales. Supposons qu'on se donne une distribution aléatoire de points dans le plan, décrits par une densité (de probabilité) $f(x, y)$. En coordonnées polaires, $(x, y) \mapsto (r, \theta)$, on va écrire

$$f(x, y) dx dy = \tilde{f}(r, \theta) dr d\theta$$

mais on sait bien que l'élément de surface infinitésimal $dx dy = r dr d\theta$, donc $\tilde{f}(r, \theta) = r f(x, y)$. Si la distribution en θ est uniforme, on peut se réduire à la seule coordonnée radiale r et considérer $\hat{f}(r) = \int_0^{2\pi} d\theta \tilde{f}(r, \theta) = 2\pi r f(x, y)$. Exemple : des points sont distribués dans le plan avec la densité $f(x, y) = A \exp -\alpha(x^2 + y^2)$ (loi gaussienne, voir chapitre suivant). Cette densité est évidemment invariante par rotation dans le plan (elle ne dépend que de la distance de l'origine au point M de coordonnées (x, y)), il est naturel d'utiliser la coordonnée radiale $r = \sqrt{x^2 + y^2}$. La densité de probabilité de la v.a. R correspondante est $\hat{f}(r) = 2\pi A r \exp -\alpha r^2$. De la même façon, une distribution dans l'espace \mathbb{R}^3 peut se récrire en coordonnées sphériques (r, θ, ψ) , avec $0 \leq \theta \leq \pi$, $0 \leq \psi \leq 2\pi$. On a alors $\tilde{f}(r, \theta, \psi) = r^2 \sin \theta f(x, y, z)$, ou s'il y a invariance par rotation (indépendance en θ, ψ), $\hat{f}(r) = 4\pi r^2 f(x, y, z)$. On verra en TD et TP des applications de ces considérations à différents problèmes, dont la distribution de Boltzmann des vitesses des molécules d'un gaz.

Chapitre 8. Distributions classiques

Dans ce chapitre nous définissons et étudions quelques distributions usuelles, ou *lois*, de variables aléatoires discrètes ou continues.

1. Distribution uniforme

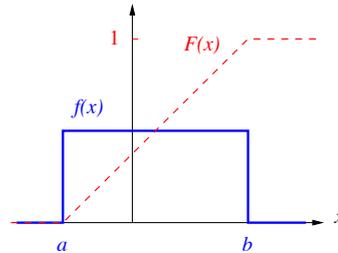


Fig. 1: Densité et fonction de répartition d'une distribution uniforme sur l'intervalle $[a, b]$.

C'est la plus simple des distributions de v.a. : la densité est constante et non nulle dans un intervalle $[a, b]$. La valeur de cette constante p est fixée par la normalisation $\int_a^b f(x) dx = (b - a)p = 1$ soit $p = \frac{1}{b-a}$, voir Fig. 1. La fonction de répartition $F(x)$ est nulle pour $x \leq a$, croît linéairement de a à b et vaut 1 pour $x \geq b$.

Dans sa version discrète, où la v.a. prend un nombre fini de valeurs x_i avec des probabilités p_i , les p_i sont égales, les valeurs x_i sont *équiprobables*. C'est le cas d'une pièce (au jeu de pile ou face) ou d'un dé bien équilibrés, avec des probabilités respectives de $p_p = p_f = \frac{1}{2}$ ou de $p_i = \frac{1}{6}$. Dans sa version continue, c'est la distribution des générateurs de nombres aléatoires usuels, qui engendrent des nombres x_i de l'intervalle $[0, 1]$ avec une densité de probabilité constante. À partir de cette densité sur $[0, 1]$, on fabrique une distribution uniforme sur tout autre intervalle $[a, b]$ en translatant et dilatant la variable en $y = x(b - a) + a$ et en prenant, selon la formule (4.1) du Chap. 7, $\tilde{f}(y) = \frac{1}{b-a} f(x) = \frac{1}{b-a} f\left(\frac{y-a}{b-a}\right) = \frac{1}{b-a}$ pour $y \in [a, b]$.

◦ *Moyenne, variance, moments*

Le calcul de la moyenne, de la variance et des moments est aisé

$$\langle X \rangle = \frac{1}{b-a} \int_a^b dx x = \frac{a+b}{2} \quad (1.1a)$$

$$m_k = \langle X^k \rangle = \frac{1}{b-a} \int_a^b dx x^k = \frac{b^{k+1} - a^{k+1}}{(k+1)(b-a)} = \frac{1}{k+1} (a^k + a^{k-1}b + \dots + b^k) \quad (1.1b)$$

$$\text{var } X = \langle X^2 \rangle - \langle X \rangle^2 = \frac{1}{12} (a-b)^2 \quad (1.1c)$$

Une situation où on rencontre cette loi est dans l'étude de l'*erreur d'arrondi*. Si un nombre réel x est approximé par sa partie entière $E(x)$ (le plus petit entier inférieur ou égal à x), $E(x) - x$ est un nombre compris entre 0 et 1. Considérant un nombre réel quelconque X comme une v.a., $X - E(X)$ est uniformément distribuée entre 0 et 1, et son écart type vaut $\frac{1}{\sqrt{12}}$ selon (1.1b). De la même façon, arrondir x à deux chiffres décimaux par valeur inférieure revient à l'approximer par $E(100x)/100$, et l'écart-type de l'erreur d'arrondi est donc alors de $0,01/\sqrt{12}$.

Que se passe-t-il dans la limite $b \rightarrow a$? La densité de probabilité est de plus en plus concentrée au voisinage de a (on dit "piquée" en a) ; montrer que dans cette limite, les moments $m_k = \langle X^k \rangle \rightarrow a^k$.

2. Distribution binomiale

C'est la distribution rencontrée au chapitre 7 (§1.2) dans la discussion des sommes des valeurs de pile ou face : si on a tiré à pile ou face n fois, quelle est la probabilité de k "faces" ?

Il est instructif de faire ce calcul de plusieurs façons différentes :

- a) en se rappelant la définition de la probabilité comme fréquence d'apparition des configurations (cf. chap 6) et en dénombrant ces configurations : dans un tirage de n fois, il y a 2^n configurations possibles et il y a C_n^k façons d'avoir k faces. La probabilité cherchée est donc

$$\mathbf{P}(X = k) = \left(\frac{1}{2}\right)^n C_n^k ;$$

- b) en utilisant les propriétés générales des probabilités. Généralisons un peu le problème en supposant la pièce mal équilibrée : elle a une probabilité p de tomber sur face, $0 < p < 1$ et $q = 1 - p$ de tomber sur pile (la somme de p et q vaut 1, pourquoi ?). Dans une série de n lancers, les résultats successifs sont des variables indépendantes (la pièce ne garde pas le souvenir si elle est tombée sur pile ou face les fois précédentes !). Par conséquent la probabilité d'une suite donnée de k "faces" et donc de $n - k$ "piles" est indépendante de l'ordre de ces faces ou piles et elle est factorisée de la forme $p^k q^{n-k}$. Comme deux suites de k "faces" différant par leur ordre sont des événements A et B mutuellement exclusifs, la probabilité de $A \cup B$ est la somme des probabilités de A et de B , chacune égale à $p^k q^{n-k}$, et plus généralement la probabilité totale cherchée $p_k^{(n)}$ est $p^k q^{n-k}$ fois le nombre total de suites avec k faces, soit C_n^k , d'où

$$\mathcal{B}(n, p) : \quad \mathbf{P}(X = k) = p_k^{(n)} := C_n^k p^k q^{n-k} , \quad (2.1)$$

avec bien sûr $\mathbf{P}(X = k) = 0$ si $k < 0$ ou $> n$. La loi est bien normalisée : $\sum p_k^{(n)} = \sum_{k=0}^n C_n^k p^k q^{n-k} = (p + q)^n = 1$ par l'identité (3.4) du chapitre 6. Pour $p = q = \frac{1}{2}$ on retrouve bien sûr le résultat précédent du a) ci-dessus.

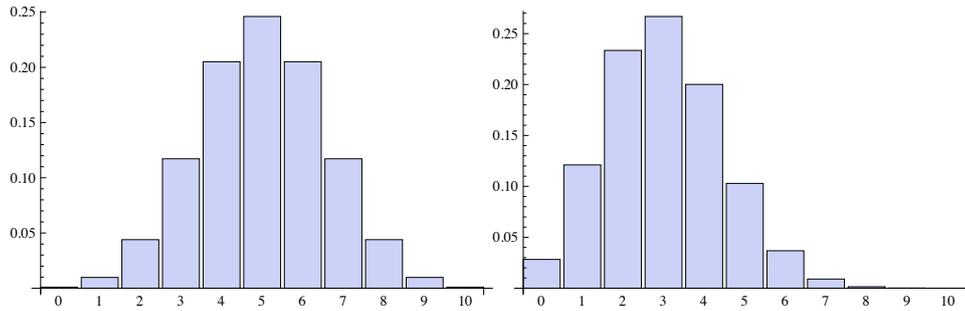


Fig. 2: Distribution binomiale pour $p = 0,5$ et pour $p = 0,3$ et $n = 10$. Noter que $\mathbf{P}(X = k; n, p) = \mathbf{P}(X = n - k; n, 1 - p)$. Le graphe pour $p = 0,7$ s’obtiendrait donc en prenant le réfléchi de celui pour $p = 0,3$ par rapport à un axe vertical passant par $n/2$.

La distribution (2.1) est appelée *loi binomiale*. Elle se rencontre chaque fois qu’on s’intéresse à n occurrences d’un événement se produisant ou non avec une probabilité p ou $1 - p$. Par exemple, on considère n noyaux radioactifs dont chacun a la probabilité p de se désintégrer pendant un intervalle Δt . Quelle est la probabilité que k noyaux se soient désintégrés au bout de ce temps ? (on suppose que les désintégrations sont des phénomènes aléatoires indépendants, c’est-à-dire qu’elles ne s’influencent pas mutuellement.) Nous reviendrons en détail sur cet exemple au § 6.3.

◦ *Moyenne, variance, moments*

À nouveau, le calcul de la moyenne et de la variance est un bon exercice

$$\begin{aligned} \langle X \rangle &= \sum_k k p_k^{(n)} = np \\ \langle X^2 \rangle &= \sum_k k^2 p_k^{(n)} = n(n-1)p^2 + np \\ \text{var } X &= \langle X^2 \rangle - \langle X \rangle^2 = n(p - p^2) = npq \end{aligned} \tag{2.2}$$

dans lequel on pourra utiliser de façon répétée l’identité $kC_n^k = nC_{n-1}^{k-1}$ (la vérifier).

◦ *Fonction caractéristique*

Au vu de ce qui a été dit plus haut, au § 1.2 du chap. 7, il est également utile de construire la fonction caractéristique. Rappelons qu’elle est définie comme $\varphi(u) = \sum_k p_k e^{iuk}$, donc pour la distribution binomiale

$$\varphi_{\text{Bin}(n)}(u) = \sum_k p_k^{(n)} e^{iuk} = \sum_{k=0}^n C_n^k e^{iuk} p^k q^{n-k} = (1 - p + pe^{iu})^n . \tag{2.3}$$

Nous ferons usage de cette expression dans la suite.

3. Distribution normale

Rappelons d'abord que

$$I = \int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi} . \quad (3.1)$$

C'est la très fameuse intégrale gaussienne.

Cette formule très utile se démontre aisément de la façon suivante : calculons le *carré* de l'intégrale, considéré comme une intégrale dans le plan et appliquons lui le changement de variables en coordonnées polaires. Explicitement $I^2 = \int_{-\infty}^{\infty} dx e^{-x^2} \int_{-\infty}^{\infty} dy e^{-y^2} = \int_{\mathbb{R}^2} dx dy e^{-x^2-y^2} = \int_{\mathbb{R}^2} r dr d\theta e^{-r^2} = 2\pi \int_0^{\infty} \frac{1}{2} d(r^2) e^{-r^2} = \pi \int_0^{\infty} du e^{-u} = \pi$, cqfd.

Par un simple changement de variable, on en déduit que pour tout $\alpha > 0$

$$I(\alpha) = \int_{-\infty}^{\infty} e^{-\alpha x^2} dx = \sqrt{\frac{\pi}{\alpha}} \quad (3.2)$$

et en particulier

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx = 1 . \quad (3.3)$$

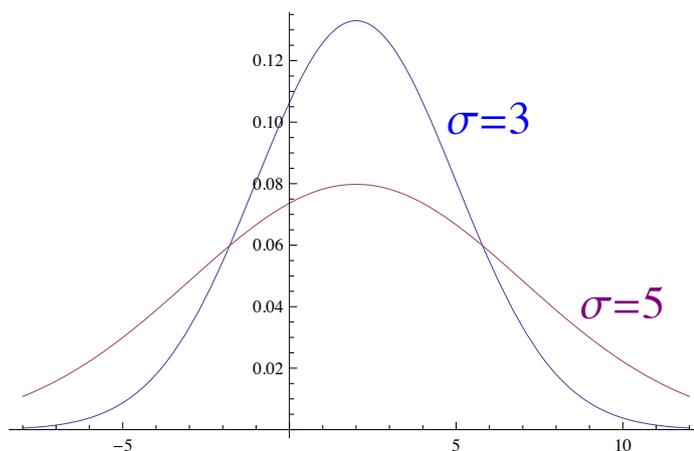


Fig. 3: Densité d'une distribution normale, $\mu = 2$, $\sigma = 3$ et $\sigma = 5$. Le maximum est atteint en $x = \mu$, la largeur du pic (par exemple à la hauteur e^{-1} fois le maximum) est proportionnelle à σ .

On définit alors une v.a. gaussienne x par sa densité

$$\mathcal{N}(\mu, \sigma) : \quad f_{\text{Gauss}}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} . \quad (3.4)$$

On parle aussi de *loi normale*. Elle dépend de deux paramètres μ et σ^2 que nous allons maintenant interpréter. Calculons pour cela la moyenne et la variance de cette v.a.

$$\langle X \rangle = \int_{-\infty}^{\infty} dx x f_{\text{Gauss}}(x) = \int_{-\infty}^{\infty} dx \left((x - \mu) + \mu \right) f_{\text{Gauss}}(x) = 0 + \mu :$$

le premier terme est nul car dans le calcul de $\langle x - \mu \rangle$ on peut effectuer le changement de variable d'intégration $x' = x - \mu$ et on obtient l'intégrale d'une fonction impaire $x' e^{-ax'^2}$ sur $(-\infty, \infty)$ qui est nulle ; le deuxième terme de la moyenne est égal à μ car l'intégrale est bien normalisée.

Pour le calcul de la variance, remarquons d'abord qu'en différentiant les deux membres de $I(\alpha)$ dans (3.2)¹¹ par rapport à α on obtient $-\frac{\partial}{\partial \alpha} I(\alpha) = \int dx x^2 e^{-\alpha x^2} = \frac{1}{2\alpha} I(\alpha)$. Prenant $\alpha = (2\sigma^2)^{-1}$, on est alors en mesure de calculer

$$\text{var} X = \int_{-\infty}^{\infty} dx (x - \mu)^2 f_{\text{Gauss}}(x) = \int \frac{dx}{\sqrt{2\pi\sigma^2}} (x - \mu)^2 e^{-\frac{(x - \mu)^2}{2\sigma^2}} = \int \frac{dx'}{\sqrt{2\pi\sigma^2}} x'^2 e^{-\frac{x'^2}{2\sigma^2}} = \sigma^2 .$$

Par conséquent les deux paramètres μ et σ^2 dont dépend la loi normale (3.4) s'interprètent respectivement comme sa moyenne et sa variance.

Le graphe de $f(x)$ a une allure caractéristique de courbe en cloche, voir Fig. 3. Elle est "centrée" en $x = \mu$, et est d'autant plus étroite que l'écart-type σ (ou la variance σ^2) est plus faible. Cette variance peut être mesurée par la "largeur du pic", c'est-à-dire la distance $2\sqrt{2}\sigma$ entre les points $x = \mu \pm \sqrt{2}\sigma$ où f est réduit d'un facteur e par rapport à son maximum:

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

Définition : on dit qu'une v.a. X a une *loi normale centrée normalisée* si elle est normale avec $\mu = 0$ et $\sigma^2 = 1$.

◦ *Moments.*

Les calculs qui précèdent ont préparé le terrain au calcul des moments d'ordre arbitraire. Par changement de variable on ramène tout calcul de moment de X à celui de $X' = X - \mu$. Pour la variable X' qui est centrée, tous les moments impairs sont nuls, par l'argument de parité donné pour la moyenne. Ses moments pairs s'obtiennent à nouveau par différentiation répétée par rapport à α dans (3.2).

¹¹ Cette procédure de "dériver sous le signe somme", c'est-à-dire d'invertir la dérivation par rapport à α et l'intégration est justifiée par la (très rapide) convergence de l'intégrale gaussienne.

Exercice. Selon ce principe, vérifier les formules suivantes

$$\langle X^2 \rangle = \mu^2 + \sigma^2 \quad ; \quad \langle X^3 \rangle = \mu^3 + 3\mu\sigma^2 \quad ; \quad \langle X^4 \rangle = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 .$$

On verra en exercice de TD une autre manière de calculer ces moments.

La conclusion importante est que *pour la distribution normale, tous les moments existent et sont calculables explicitement.*

Au vu de l'inégalité de Tchebychev (chapitre 7, équ. (2.8)), l'existence de moments d'ordre quelconque implique que la probabilité $\mathbf{P}(|X| > a)$ s'annule très vite quand a croît, plus vite que toute puissance inverse a^{-k} ; ceci est en accord avec notre intuition : dans la distribution normale les événements très éloignés de la moyenne sont extrêmement rares.

Exercice : exprimer cette probabilité $\mathbf{P}(|X| > a)$ pour une variable normale centrée normalisée en termes de la "fonction d'erreur complémentaire" $\operatorname{erfc}(x) := \frac{2}{\sqrt{\pi}} \int_0^x dy e^{-y^2}$, qu'on retrouvera au TP3 et au Chap. 9, § 3.4.

◦ *Fonction caractéristique*

Ici aussi, il est utile de calculer la fonction caractéristique de la loi normale :

$$\begin{aligned} \varphi_{\text{Gauss}}(u) &= \int_{-\infty}^{\infty} dx e^{iux} f(x) = \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi\sigma^2}} e^{iux} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\ &= e^{iu\mu} \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi\sigma^2}} e^{iu(x-\mu)} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = e^{iu\mu} \int_{-\infty}^{\infty} \frac{dx'}{\sqrt{2\pi\sigma^2}} e^{-\frac{x'^2}{2\sigma^2} + iux'} \\ &= e^{iu\mu} \int_{-\infty}^{\infty} \frac{dx'}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x' - iu\sigma^2)^2 + u^2\sigma^4}{2\sigma^2}} = e^{iu\mu} e^{-\frac{1}{2}u^2\sigma^2} . \end{aligned} \quad (3.5)$$

où nous avons d'abord effectué le changement de variable familier $x' = x - \mu$, suivi d'un deuxième : $x'' = x' - iu\sigma^2$, translation par une quantité complexe que l'on peut justifier, et qui nous ramène une fois encore à une intégrale gaussienne. Au final

$$\varphi_{\text{Gauss}}(u) = e^{iu\mu} e^{-\frac{1}{2}u^2\sigma^2} \quad (3.6)$$

une expression qui permet aussi de calculer aisément tous les moments de la distribution normale, voir TD 8.

◦ *Importance des variables normales*

Pour des raisons que l'on va voir, cette loi normale joue un rôle particulièrement important. La distribution normale joue un rôle central aussi bien en probabilités et statistiques qu'en physique. En probabilités et statistiques, cela tient en particulier au théorème limite central, que nous étudierons au chapitre suivant, selon lequel cette distribution apparaît comme limite de nombreux phénomènes aléatoires. En physique, elle est au cœur de très nombreuses considérations ; en physique statistique, en relation avec les poids de Boltzmann : pour un système de particules sans interactions, tel un gaz parfait, l'énergie cinétique est $E = \sum \frac{1}{2}mv_i^2$, et les vitesses considérées comme des v.a. obéissent à une loi normale dictée par le poids de Boltzmann $e^{-\frac{E}{k_B T}} = \prod_i e^{-\frac{mv_i^2}{2k_B T}}$, (voir ci-dessous § 6.2) ; c'est le cas encore avec l'étude d'ensembles d'oscillateurs harmoniques, que l'on rencontre en physique des solides et en théorie quantique des champs, etc etc.

4. Distribution de Poisson

Soit λ un paramètre réel positif, et k une v.a. discrète à valeurs entières ≥ 0 . La *distribution de Poisson* est définie par les probabilités

$$\mathcal{P}(\lambda) : \quad \mathbf{P}(X = k) = p_k^{\text{Poisson}} := \frac{\lambda^k}{k!} e^{-\lambda} \quad (4.1)$$

Il convient d'abord de vérifier que cette loi est bien normalisée, ce qui est immédiat :

$$\sum_{k=0}^{\infty} p_k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = 1.$$

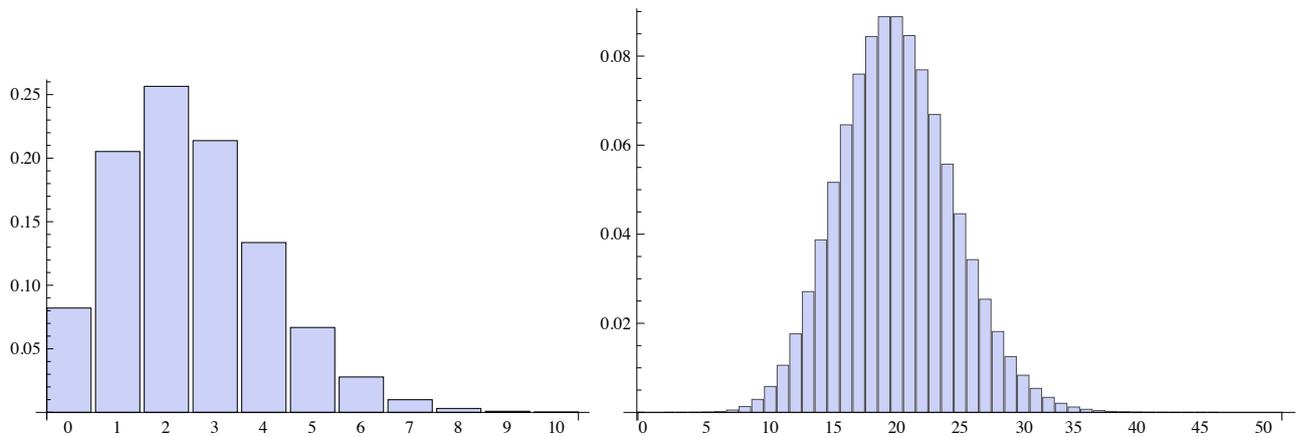


Fig. 4: Distribution de Poisson pour $\lambda = 2, 5$ et $\lambda = 20$.

◦ *Moyenne, variance, moments*

Une fois encore, on calcule aisément les moyenne et variance de la distribution :

$$\begin{aligned} \langle X \rangle &= \sum_{k=0}^{\infty} k \lambda^k \frac{e^{-\lambda}}{k!} = \lambda \\ \langle X^2 \rangle &= \sum_{k=0}^{\infty} k^2 \lambda^k \frac{e^{-\lambda}}{k!} = \lambda^2 + \lambda \\ \text{var } X &= \langle X^2 \rangle - \langle X \rangle^2 = \lambda \end{aligned} \quad (4.2)$$

Exercice : vérifier ces formules.

◦ *Importance de la loi de Poisson*

La loi de Poisson apparaît dans de nombreux phénomènes naturels ou de la vie pratique, chaque fois que des événements aléatoires indépendants se produisent à un taux moyen

constant λ (dans l'espace et/ou le temps) : la loi p_k^{Poisson} donne alors la probabilité d'avoir k occurrences de l'événement. Ainsi supposons que N objets ont été distribués selon une loi de Poisson $\mathcal{P}(\lambda)$ pendant un temps T donné (cela peut s'appliquer à des appels téléphoniques, à des impacts de particules émises par une source, à des gouttes de pluie ou autres projectiles supposés répartis à un rythme constant, etc). On s'attend donc à en recevoir en moyenne $\lambda = N/T$ par unité de temps. On démontre alors (voir un exemple dans l'exercice 2 du TD 8) que la probabilité d'en recevoir k pendant un temps $t \leq T$ est donnée par la loi de Poisson de paramètre λt . Exemple : je reçois en moyenne 80 e-mails par jour, répartis *grosso modo* de façon uniforme dans le temps. Quelle est la probabilité que j'en reçoive k un certain jour ? Réponse : p_k^{Poisson} avec $\lambda = 80$. Que j'en reçoive 19 en une demi-journée ? Réponse : p_{19}^{Poisson} avec $\lambda' = Nt/T = 40$.

Cette apparition de la loi de Poisson tient au fait qu'elle décrit une certaine limite de la loi binomiale, comme on va le voir au § suivant.

5. Limites de la loi binomiale

◦ *Limite normale.*

Comme le suggère le graphe de la distribution binomiale (voir Fig. 2), avec sa forme en cloche très proche de celle de la gaussienne, la distribution normale est une bonne approximation de la loi binomiale $p_k^{(n)}$ dans la limite des n grands, et pour $k \sim np$, avec p et $q = 1 - p$ finis, donc $n \gg 1$, $k \gg 1$, $n - k \gg 1$. C'est donc la partie centrale de la distribution que l'on explore là. On peut en effet démontrer après un calcul non trivial que

$$p_k^{(n)} \simeq \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(k - np)^2}{2\sigma^2}\right) (1 + O(n^{-1})), \quad (5.1)$$

c'est-à-dire une loi normale de moyenne np et de variance $\sigma^2 = np(1 - p)$.

La démonstration repose sur l'étude dans la limite considérée des fonctions caractéristiques calculées plus haut, cf (2.3) et (3.6). Calculons le logarithme de $\varphi_{\text{Bin}(n)}(u)$ pour u petit (qui seul nous intéresse si $k \gg 1$)

$$\begin{aligned} \ln \varphi_{\text{Bin}(n)}(u) &= n \ln(1 + p(e^{iu} - 1)) \simeq n \left(p(e^{iu} - 1) - \frac{1}{2}p^2(e^{iu} - 1)^2 + \dots \right) \\ &= n \left(p(iu - \frac{u^2}{2} + \dots) + \frac{1}{2}p^2u^2 + O(u^3) \right) \\ &= n \left(iup - \frac{u^2}{2}p(1 - p) + \dots \right) \end{aligned}$$

et comparons-le à celui de $\varphi_{\text{Gauss}}(u)$, cf (3.6)

$$\ln \varphi_{\text{Gauss}}(u) = iu\mu - \frac{1}{2}u^2\sigma^2$$

Pour $k \gg 1$, donc $u \ll 1$, on voit que les deux fonctions caractéristiques coïncident pourvu que d'une part les moyennes $\langle k \rangle = np$ et μ , de l'autre les variances $\text{var } X = np(1-p)$ et σ^2 coïncident.

En pratique, pour comparer la distribution (discrète) binomiale et la distribution (continue) normale, il faut calculer numériquement à l'aide de cette dernière la probabilité que la variable x soit dans un intervalle de longueur 1 centré sur l'entier k

$$p'_k = p(k - \frac{1}{2} < x < k + \frac{1}{2}) = \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} dx f_{\text{Gauss}}(n) = F(k + \frac{1}{2}) - F(k - \frac{1}{2})$$

et la comparer à $p_k^{(n)}$ calculé avec la loi binomiale. Par exemple pour $n = 50$ et $p = 0,2$, on trouve que p'_k diffère de $p_k^{(n)}$ d'au plus 8% dans l'intervalle $6 \leq k \leq 16$, voir la figure suivante.

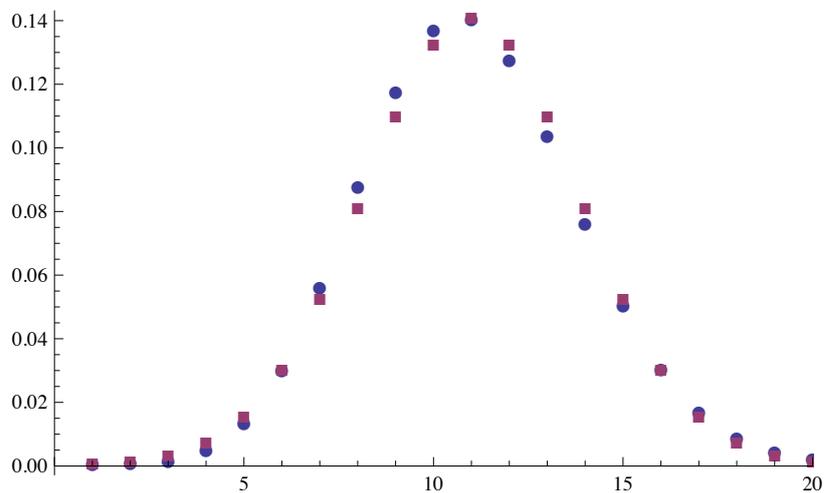


Fig. 5: Comparaison de la distribution binomiale ($n = 50$, $p = 0,2$: petits disques) et de la normale ($\mu = 10$, $\sigma = \sqrt{np(1-p)}$: petits carrés) intégrée dans chaque intervalle de longueur 1.

o *Limite poissonnienne*

Il peut être aussi intéressant d'étudier la limite de la loi binomiale quand p est très petit et np (la valeur moyenne de k) est finie, tandis que $n \gg 1$. Il faut bien comprendre que l'on explore maintenant le *bord* de la distribution binomiale, loin de son maximum. Poisson a montré que l'on trouve alors la distribution qui porte son nom.

$$p_k^{(n)} \Big|_{n \gg 1} \simeq p_k^{\text{Poisson}} = \frac{\lambda^k}{k!} e^{-\lambda} \tag{5.2}$$

avec $\lambda = np$ supposée finie tandis que $n \rightarrow \infty$.

Démonstration de cette limite : partout où apparaît p dans $p_k^{(n)}$, on substitue $p = \lambda/n$ puis on développe à n grand

$$\begin{aligned}
 p_k^{(n)} &= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!} \frac{(np)^k}{n^k} \left(1 - \frac{np}{n}\right)^{n-k} \\
 &= \frac{n(n-1)\cdots(n-(k-1))}{k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
 &= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \times \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \left(1 - \frac{\lambda}{n}\right)^{-k} \\
 &\simeq \frac{\lambda^k}{k!} e^{-\lambda} (1 + O(n^{-1}))
 \end{aligned}
 \tag{5.3}$$

en se rappelant que $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$.

A nouveau une petite expérience numérique est utile. Pour $n = 50$, $p = 0,02$, on voit que la distribution de Poisson pour $\lambda = np = 1$ approxime à moins de 2% la loi binomiale pour $0 \leq k \leq 3$, comme le montre cette table

k	:	0	1	2	3	4	5
Binomiale $p_k^{(50)}$:	0,36417	0,37160	0,18580	0,06067	0,01454	0,002731
Poisson $p_k^{\text{Poisson}} _{\lambda=1}$:	0,36788	0,36788	0,18394	0,06131	0,01533	0,003066

Récapitulons. Pour n très grand et p finie, la loi binomiale est bien approximée par une normale pour les valeurs de k les plus probables, $k \approx pn \sim n$: les sommets des deux cloches binomiale et gaussienne se superposent bien. Pour p très petit et $\langle k \rangle = pn$ fini, les événements à $k \ll n$ (le bord gauche de la cloche binomiale) sont bien décrits par la loi de Poisson.

6. Quelques exemples concrets

6.1. L'aiguille de Buffon

Voici le problème posé par Buffon : on laisse tomber une aiguille sur un parquet, quelle est la probabilité que l'aiguille chevauche deux lames du parquet ?

De façon plus précise, soit un ensemble de droites parallèles à l'axe des x , espacées d'une distance ℓ . Quelle est la probabilité pour un segment de longueur $b < \ell$ uniformément distribué en position et en orientation d'être sécant à une de ces droites ?

Dans la bande (la lame du parquet) où se trouve le milieu O du segment, soit y la distance de ce milieu au bord inférieur. Soit θ l'angle qu'il fait avec le bord, $0 \leq \theta < \pi$.

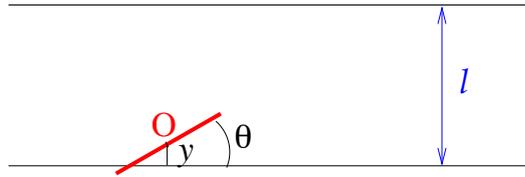


Fig. 6: Problème de Buffon

L'hypothèse de distribution uniforme en position et angle signifie que la densité de probabilité autour du point (y, θ) est $f(y, \theta) = \frac{1}{\ell} \frac{1}{\pi}$ qui est bien normalisée puisque $\int dy d\theta f(y, \theta) = \int_0^\ell \frac{dy}{\ell} \int_0^\pi \frac{d\theta}{\pi} = 1$.

La condition que le segment recoupe un bord est $y < \frac{1}{2}b \sin \theta$ (bord inférieur) ou $\ell - y < \frac{1}{2}b \sin \theta$ (bord supérieur). Les deux événements sont exclusifs puisque $b < \ell$. La probabilité pour le segment de *ne pas être sécant* est donc

$$\begin{aligned} P_{NC} &= \int_0^\pi \frac{d\theta}{\pi} \int_{\frac{1}{2}b \sin \theta}^{\ell - \frac{1}{2}b \sin \theta} \frac{dy}{\ell} \\ &= \int_0^\pi \frac{d\theta}{\pi} \left(1 - \frac{b}{\ell} \sin \theta\right) \\ &= 1 - \frac{2b}{\pi \ell}. \end{aligned}$$

La probabilité de croisement est donc

$$P_C = \frac{2b}{\pi \ell}.$$

C'est un bon exercice de reprendre le calcul dans l'hypothèse où $b > \ell$. Montrer que la la probabilité d'intersection vaut alors $P_C = 1 - \frac{2}{\pi} \text{Arc sin } \frac{\ell}{b} + 2 \frac{b}{\pi \ell} \left(1 - \sqrt{1 - \frac{\ell^2}{b^2}}\right)$. Vérifier que les deux limites $b \rightarrow \ell \pm 0$ coïncident.

6.2. Distribution de Maxwell des vitesses dans un gaz

Dans un gaz (parfait) à l'équilibre, soit $f(v_x, v_y, v_z)$ la densité de probabilité des vitesses d'une molécule. Montrons que des considérations de symétrie et d'indépendance conduisent à l'expression de f . Supposons en effet que les v.a. v_x, v_y et v_z sont indépendantes, une hypothèse qu'il faudrait justifier... La fonction f est donc factorisée, cf chap 7, § 1.3, et le gaz étant isotrope, les trois directions d'espace jouent le même rôle et les trois fonctions sont identiques

$$f(v_x, v_y, v_z) = g(v_x)g(v_y)g(v_z).$$

Par ailleurs, par invariance par rotation, la fonction $f(v_x, v_y, v_z)$ doit n'être fonction que de $\vec{v}^2 = v_x^2 + v_y^2 + v_z^2$. On cherche donc une fonction ayant la propriété que

$$f(v_x, v_y, v_z) = h(v_x^2 + v_y^2 + v_z^2) = g(v_x)g(v_y)g(v_z) .$$

En prenant la dérivée logarithmique de cette expression par rapport à v_x ou v_y , on a $\frac{1}{2v_x} \frac{g'(v_x)}{g(v_x)} = \frac{1}{2v_y} \frac{g'(v_y)}{g(v_y)} = \frac{h'(v_x^2+v_y^2+v_z^2)}{h(v_x^2+v_y^2+v_z^2)} = \alpha$, une constante indépendante de x et y , ce qui s'intègre en $g(v_x) = a \exp -\alpha v_x^2$. Donc seule la fonction exponentielle a cette propriété:

$$f(v_x, v_y, v_z) = A e^{-\alpha(v_x^2+v_y^2+v_z^2)}$$

et c'est donc une distribution normale ! La normalisation de f fixe $A = \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}}$ et le coefficient α est identifié avec $\frac{1}{2}m\beta$, $\beta = (k_B T)^{-1}$, de telle sorte qu'apparaît dans l'exponentielle $\beta \times \frac{1}{2}m\vec{v}^2 = \frac{1}{k_B T} \times$ énergie cinétique de la molécule, cf fin du chap 7

$$f(v_x, v_y, v_z) = \left(\frac{m}{2\pi k_B T}\right)^{\frac{3}{2}} e^{-\frac{m(v_x^2+v_y^2+v_z^2)}{2k_B T}} . \quad (6.1)$$

Enfin il peut être utile d'exprimer cette loi en termes de la seule variable $v = \sqrt{\vec{v}^2}$. Pour cela on effectue un changement de variable des coordonnées $(v_x, v_y, v_z) \mapsto (v, \theta, \phi)$, cf fin du chap 7, on intègre sur les angles θ et ϕ dont la densité f ne dépend pas et on trouve finalement

$$f(v) = 4\pi \left(\frac{m}{2\pi k_B T}\right)^{\frac{3}{2}} v^2 e^{-\frac{mv^2}{2k_B T}} , \quad (6.2)$$

c'est la distribution de Maxwell-Boltzmann des vitesses d'un gaz, que l'on étudiera au TP 3. Exercice : utiliser le "truc" de la dérivation sous le signe somme pour calculer $\langle v^2 \rangle$ et montrer que $\langle v^2 \rangle = 3\langle v_x^2 \rangle$ où la moyenne de v_x^2 est calculée avec la distribution (6.1). Cela n'était-il pas attendu ?

6.3. Désintégrations radioactives

On s'intéresse ici à la désintégration de noyaux d'un isotope $X \rightarrow X'$. On décrit en général le phénomène par une loi "empirique" : le nombre de désintégrations par unité de temps est proportionnel au nombre $N(t)$ de noyaux X présents à l'instant t , avec un taux de désintégration constant λ , caractéristique du noyau étudié X et de son mode de désintégration $X \rightarrow X'$. Pendant le temps dt , le nombre de noyaux se désintégrant selon ce mode est

$$|dN| = \lambda N dt = -dN . \quad (6.3)$$

Autrement dit, la fonction $N(t)$ satisfait l'équation différentielle $\dot{N}(t) = -\lambda N(t)$. En l'intégrant, on trouve

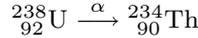
$$N(t) = N(0)e^{-\lambda t} , \quad (6.4)$$

la quantité de l'isotope considéré décroît exponentiellement. On définit la *demi-vie*, notée τ , par le temps au bout duquel le nombre N a décréu par un facteur 2. On a donc $N(\tau) = N(0)/2$, ce qui conduit à

$$\tau = \frac{\ln 2}{\lambda} \quad N(t) = N(0) e^{-t \ln 2 / \tau} = N(0) 2^{-t/\tau}. \quad (6.5)$$

Le nombre de noyaux de l'isotope aura décréu d'un facteur 1000 au bout d'un temps $\frac{\ln 1000}{\ln 2} \tau \approx 10 \tau$ (se rappeler que $2^{10} \approx 10^3$).

Par exemple, pour l'isotope naturel ^{238}U de l'Uranium, qui se désintègre en Thorium selon



en émettant un noyau d'Helium (particule α), la demi-vie est $\tau = 4,5$ Gans, (1 "Gan" = 1 milliard d'années !) donc le nombre N aura décréu d'un facteur 1000 au bout d'un temps $t = 45$ Gans, soit trois fois l'âge actuel de l'Univers... (Pour mémoire, d'autres noyaux radioactifs peuvent avoir des demi-vies beaucoup plus courtes, tels ces deux isotopes du Cesium et de l'Iode dont a beaucoup parlé au moment de l'accident de Fukushima, $^{137}_{55}\text{Cs}$ et $^{131}_{53}\text{I}$, de demi-vies respectives 30 ans et 8 jours.)

Le problème de désintégration de noyaux atomiques peut et doit en fait être traité par une approche probabiliste, puisque c'est bien là le fond de la question : un noyau (comme tout système individuel de nature quantique) a une *probabilité de transition* de tel ou tel état vers tel autre. Le traitement qui a précédé s'est appliqué implicitement à une vaste population de noyaux (dont le nombre a été traité comme une variable continue, et non comme un entier discret), et a concerné en fait, comme on va le voir plus bas, le nombre moyen (la valeur moyenne ou espérance au sens probabiliste) de la v.a. nombre total de noyaux de l'isotope X. L'équation (6.3) se redit en termes probabilistes comme suit : la probabilité d'un noyau donné de subir la désintégration étudiée pendant le temps dt est $\mathbf{P}(X \rightarrow X'; dt) = \lambda dt$. Notons $\mathbf{P}_N(t)$ la probabilité d'avoir N noyaux dans l'état initial (radioactif) X au temps t . Les désintégrations des différents noyaux étant supposées des événements indépendants, on a

$$\mathbf{P}_N(t + dt) = (1 - \lambda dt)^N (\mathbf{P}_N(t) + \lambda(N + 1)\mathbf{P}_{N+1}(t)dt) + O(dt^2),$$

somme des probabilités des événements exclusifs suivants : aucune désintégration des N noyaux (probabilité $(1 - \lambda dt)$ pour chacun, factorisation pour l'ensemble puisque événements indépendants) + une désintégration d'un des $N + 1$ noyaux avec la probabilité λdt fois la probabilité que les N autres restent inchangés + des désintégrations multiples d'ordre plus élevé en dt . Écrivant $\mathbf{P}_N(t + dt) = \mathbf{P}_N(t) + dt \frac{d}{dt} \mathbf{P}_N(t) + O(dt^2)$ et ne gardant que les termes d'ordre dt , on a donc

$$\frac{d}{dt} \mathbf{P}_N(t) = \lambda((N + 1)\mathbf{P}_{N+1}(t) - N\mathbf{P}_N(t)) \quad (6.6)$$

On se donne (avec probabilité 1) la valeur initiale N_0 au temps 0, $\mathbf{P}_N(t = 0) = \delta_{N N_0}$. On va résoudre l'équation (6.6) en considérant la fonction génératrice des $\mathbf{P}_N(t)$ (voir chap. 7, § 1.2) : soit

$$\phi(x, t) = \sum_{N=0}^{N_0} x^N \mathbf{P}_N(t). \quad (6.7)$$

Calculons les dérivées partielles de $\phi(x, t)$

$$\begin{aligned} \frac{\partial}{\partial t} \phi(x, t) &= \sum_{N=0}^{N_0} x^N \frac{d}{dt} \mathbf{P}_N(t), \\ \frac{\partial}{\partial x} \phi(x, t) &= \sum_{N=1}^{N_0} N x^{N-1} \mathbf{P}_N(t) = \sum_{N=0}^{N_0-1} (N + 1) x^N \mathbf{P}_{N+1}(t) = \sum_{N=0}^{N_0} (N + 1) x^N \mathbf{P}_{N+1}(t) \\ &\quad \uparrow \text{car } p_{N_0+1}(t) = 0! \\ x \frac{\partial}{\partial x} \phi(x, t) &= \sum_{N=0}^{N_0} N x^N \mathbf{P}_N(t). \end{aligned} \quad (6.8)$$

On a donc fait apparaître aux membres de droite de (6.8) les trois termes de l'équation (6.6), ce qui conduit donc à une *équation aux dérivées partielles* pour ϕ

$$\frac{\partial}{\partial t}\phi(x, t) = \lambda(1-x)\frac{\partial}{\partial x}\phi(x, t) . \quad (6.9)$$

La méthode systématique de résolution fait appel à la “transformation de Laplace”, qui sera étudiée en L3. Contentons-nous d'observer que toute fonction de la forme $\phi(x, t) = G((1-x)\psi(t))$, avec G une fonction arbitraire, est solution de (6.9) à condition que $\dot{\psi}(t) = -\lambda\psi(t)$, soit $\psi(t) = ae^{-\lambda t}$, avec a une constante quelconque. La fonction jusque là arbitraire G est fixée par la condition initiale $\phi(x, 0) = x^{N_0} = G(a(1-x))$, soit $G(ax) = (1-x)^{N_0}$, et finalement la solution unique de (6.9) complétée par cette condition initiale est

$$\begin{aligned} \phi(x, t) &= \left(1 - (1-x)e^{-\lambda t}\right)^{N_0} = \left((1 - e^{-\lambda t}) + xe^{-\lambda t}\right)^{N_0} \\ &= \sum_{N=0}^{N_0} C_{N_0}^N x^N e^{-N\lambda t} (1 - e^{-\lambda t})^{N_0-N} \end{aligned} \quad (6.10)$$

qui permet d'identifier

$$\mathbf{P}_N(t) = C_{N_0}^N e^{-N\lambda t} (1 - e^{-\lambda t})^{N_0-N} . \quad (6.11)$$

On voit apparaître la loi binomiale pour une probabilité $p(t) = e^{-\lambda t}$, comme anticipé au § 2 : en effet $p(t) = e^{-\lambda t}$ n'est autre que la probabilité qu'un noyau donné ne se soit pas désintégré au temps t .

On peut alors calculer l'espérance de la variable aléatoire N

$$\langle N \rangle(t) = \sum_{N=0}^{N_0} N \mathbf{P}_N(t) = \left. \frac{\partial \phi(x, t)}{\partial x} \right|_{x=1} = N_0 e^{-\lambda t} .$$

On retrouve bien le résultat obtenu plus haut à partir de l'équation différentielle (6.3), qui s'appliquait donc en fait à $\langle N \rangle$.

Noter encore que si N_0 est très grand, N petit par rapport à N_0 et $e^{-\lambda t}$ petit, on peut remplacer la loi binomiale (6.11) par sa limite poissonnienne $\mathbf{P}(\lambda = N_0 e^{-\lambda t})$, dont l'espérance est encore $\langle N \rangle = \lambda = N_0 e^{-\lambda t}$.

On pourrait aussi étudier maintenant ce que sont les fluctuations de la v.a. N autour de cette valeur moyenne.

Chapitre 9. Théorèmes asymptotiques

Dans ce chapitre nous allons énoncer, démontrer et illustrer deux théorèmes asymptotiques, la *loi des grands nombres* et le *théorème limite central*.

1. Préambule

1.1. Convergence en probabilité

Considérons d'abord une suite infinie X_n de v.a. réelles, $n = 1, 2, \dots$. On dira que la suite converge en probabilité vers une limite ℓ si

$$\forall h > 0 \quad \mathbf{P}(|X_n - \ell| > h) \rightarrow 0 \quad \text{quand } n \rightarrow \infty. \quad (1.1)$$

Autrement dit, (cf le cours LP206), $\forall h > 0, \forall \epsilon > 0, \exists N : \forall n > N \quad \mathbf{P}(|X_n - \ell| > h) < \epsilon$.

Plus généralement on dira que la suite de v.a. réelles X_n converge en probabilité vers une v.a. réelle Y si

$$\forall h > 0 \quad \mathbf{P}(|X_n - Y| > h) \rightarrow 0 \quad \text{quand } n \rightarrow \infty. \quad (1.2)$$

1.2. Moyenne arithmétique de N v.a.

Soient X_1, X_2, \dots, X_N N v.a. réelles indépendantes et distribuées selon la même loi de probabilité. (Elles ont donc en particulier même valeur moyenne $\langle X_i \rangle = m$ et même variance $\text{var } X_i = \sigma^2$, si cette moyenne et cette variance existent). Ce peut être par exemple le résultat de N lancers successifs d'un même dé (qu'il soit "honnête" ou pipé), la taille de N individus choisis au hasard dans une population homogène, N mesures répétées d'une quantité physique entachées d'erreurs, etc. On définit alors une nouvelle v.a., leur *moyenne arithmétique*,

$$Y = \frac{1}{N}(X_1 + X_2 + \dots + X_N). \quad (1.3)$$

Bien distinguer cette moyenne arithmétique Y des X_i , qui est une v.a., de leur valeur moyenne $\langle X_i \rangle$, qui est un nombre réel.

On se propose d'étudier la loi que suit Y dans la limite où N est grand.

De façon qualitative (nous allons préciser les hypothèses et énoncer précisément les théorèmes ci-dessous)

- la loi des grands nombres nous dit que la v.a. Y converge en probabilité quand $N \rightarrow \infty$ vers la valeur commune des $\langle X_i \rangle$;
- le théorème central limite affirme que la v.a. Y converge en probabilité vers une v.a. normale (gaussienne) de moyenne $\langle X \rangle$ et de variance $\frac{1}{N}\text{var}(X_i)$ dans la limite $N \rightarrow \infty$.

2. Loi des grands nombres

2.1. Le théorème et sa preuve

Théorème : Soient X_1, X_2, \dots, X_N une suite de v.a. indépendantes et de même loi. On suppose que la valeur moyenne m et la variance σ^2 de cette loi sont bien définies¹². Alors la moyenne arithmétique Y de (1.3) converge en probabilité vers sa valeur moyenne m quand $N \rightarrow \infty$.

Preuve : Par l'hypothèse que les X_i sont indépendantes et ont même loi, donc même valeur moyenne (ou espérance) $\langle X \rangle$, on a d'abord

$$\langle Y \rangle = \frac{1}{N} \sum_{i=1}^N \langle X_i \rangle = \langle X \rangle = m .$$

Calculons alors la variance de Y

$$\begin{aligned} \text{var } Y &= \langle (Y - m)^2 \rangle = \left\langle \left(\frac{1}{N} \sum_{i=1}^N (X_i - m) \right)^2 \right\rangle \\ &= \frac{1}{N^2} \left\langle \sum_i (X_i - m)^2 \right\rangle + \frac{2}{N^2} \sum_{i < j} \langle (X_i - m)(X_j - m) \rangle \\ &= \frac{1}{N^2} \left\langle \sum_i (X_i - m)^2 \right\rangle + \frac{2}{N^2} \sum_{i < j} \langle (X_i - m) \rangle \langle (X_j - m) \rangle \quad (\text{indépendance}) \\ &= \frac{1}{N^2} \left\langle \sum_i (X_i - m)^2 \right\rangle = \frac{1}{N^2} N \sigma^2 \quad \text{puisque } \langle (X_i - m) \rangle = 0 \\ &= \frac{1}{N} \sigma^2 = \frac{1}{N} \text{var } X \end{aligned} \tag{2.1}$$

On est maintenant en mesure d'appliquer l'inégalité de Tchebychev (equ. (2.8) du chap 7, avec $k = 2$) à la v.a. $Y - m$

$$\mathbf{P}(|Y - m| > h) \leq \frac{\langle (Y - m)^2 \rangle}{h^2} = \frac{\sigma^2}{Nh^2} . \tag{2.2}$$

Quel que soit $h > 0$, la probabilité $\mathbf{P}(|Y - m| > h)$ tend vers zéro quand $N \rightarrow \infty$, CQFD.

¹² On peut en fait se passer de l'hypothèse d'existence de la variance, mais la preuve est alors plus compliquée.

2.2. Deux illustrations

○ *Évaluation empirique de la probabilité ou de la moyenne $\langle X \rangle$*

La loi des grands nombres nous permet de justifier *a posteriori* la définition “intuitive” d’une probabilité d’un événement comme une fréquence d’occurrence de cet événement. Soit A un événement aléatoire se réalisant avec la probabilité p , associons lui une v.a. X prenant la valeur $X = 1$ si A se réalise, $X = 0$ sinon. Donc $\mathbf{P}(X = 1) = p$, $\mathbf{P}(X = 0) = 1 - p$, $\langle X \rangle = 1 \cdot p + 0 \cdot (1 - p) = p$. Effectuons un grand nombre d’épreuves X_i , $i = 1, \dots, N$ et considérons leur moyenne arithmétique $Y = \frac{1}{N}(X_1 + X_2 + \dots + X_N)$. Cette v.a. est la fréquence de réalisation de A dans nos N épreuves. La loi des grands nombres nous dit que Y tend en probabilité vers $\langle X \rangle = p$ et fournit donc une estimation de cette probabilité.

C’est bien de cette façon que nous nous assurons qu’un dé est “honnête” : chacun des 6 nombres vient avec une fréquence $\approx \frac{1}{6}$ dans une série d’un grand nombre de lancers. La même méthode permet d’estimer la (valeur) moyenne d’une v.a. X , en “mesurant” la moyenne (arithmétique) d’un échantillon de taille N suffisamment grande.

Les sondages d’opinion procèdent de manière analogue. Pour savoir quel pourcentage de la population aime les savonnettes Truc, ou a une opinion favorable de M. X, etc, c’est-à-dire quelle est la probabilité qu’un individu tiré au hasard ait telle ou telle opinion, on prend un échantillon de taille N et on mesure l’opinion moyenne de cet échantillon. Noter que (2.1) nous enseigne que l’écart-type de Y (c’est-à-dire la largeur de la distribution des Y) décroît comme $N^{-\frac{1}{2}}$: la précision d’un sondage est donc d’ordre $N^{-\frac{1}{2}}$, c’est-à-dire que pour gagner un facteur 10 sur l’incertitude, il faut multiplier la taille de l’échantillon par 100 !

Grâce à l’inégalité de Tchebychev (2.2), on peut même estimer la taille de l’échantillon nécessaire pour obtenir une précision donnée. Quelle doit être cette taille N (ou le nombre d’épreuves ou de mesures dans les exemples précédents) pour que la probabilité que Y fournisse la valeur de $m = \langle X \rangle$ avec une erreur *relative* de moins de 1% soit de 95%, autrement dit pour que $\mathbf{P}(|Y - m| > 0,01m) < 0,05$? Il suffit que $\frac{\sigma^2}{Nh^2} = \frac{10^4}{N} \left(\frac{\sigma}{m}\right)^2 < 0,05$, soit $N > 200\,000 \left(\frac{\sigma}{m}\right)^2$. On a déjà noté que la quantité “sans dimension” $\frac{\sigma}{m}$ est la bonne façon de mesurer l’étalement de la distribution d’une v.a. On voit que c’est elle encore qui donne la taille de l’échantillon requise pour atteindre une précision donnée.

◦ *Évaluation Monte-Carlo d'une intégrale*

Supposons les v.a. X_1, \dots, X_N uniformément distribuées entre 0 et 1, (donc avec une densité de probabilité $f(x) = 1$). Soit $g(x)$ une fonction (de carré intégrable), $g(X_i)$ est une nouvelle v.a. dont la valeur moyenne et la variance se calculent aisément

$$m_g = \langle g(X_i) \rangle = \int_0^1 g(x) dx$$

$$\sigma_g^2 = \int_0^1 (g(x) - m_g)^2 dx .$$

La loi des grands nombres nous dit que la moyenne arithmétique des $g(X_i)$

$$\frac{1}{N}(g(X_1) + \dots + g(X_N))$$

converge en probabilité vers la valeur moyenne $m_g = \langle g(X) \rangle = \int_0^1 g(x) dx$ et que l'écart-type de la différence est $\sigma_g = \sigma/N^{\frac{1}{2}}$. Cela signifie qu'on a une approximation

$$\frac{1}{N}(g(x_1) + \dots + g(x_N)) \approx \int_0^1 g(x) dx \quad (2.3)$$

dans laquelle les x_i sont des valeurs tirées de tables de nombres aléatoires uniformément distribués entre 0 et 1 ; la précision de cette approximation se comporte en $N^{-\frac{1}{2}}$.

Cette méthode de calcul est appelée *méthode de Monte-Carlo*. La convergence en $N^{-\frac{1}{2}}$ n'est pas très rapide et nécessite beaucoup de points pour obtenir une précision satisfaisante. La méthode prend toute son importance pour des intégrales multiples et est très utilisée. On peut lui apporter toutes sortes d'améliorations et de raffinements...

3. Théorème limite central

On a vu au chapitre précédent que la limite quand $N \gg 1$ d'une loi binomiale au voisinage de son maximum est une loi normale. Or cette loi binomiale décrit la distribution des événements quand on répète N fois une épreuve dont la probabilité de succès est p . Le théorème limite central, appelé aussi (incorrectement) "théorème de la limite centrale", généralise ce résultat.

3.1. Énoncé et remarques

Théorème : Soient N v.a. X_1, \dots, X_N indépendantes et de même loi, dont la valeur moyenne m et la variance σ^2 existent (c'est-à-dire sont finies). Leur moyenne arithmétique Y , définie en (1.3), est une v.a. dont la loi \tilde{f} tend quand $N \rightarrow \infty$ vers une loi normale de moyenne m et de variance σ^2/N

$$\tilde{f}(y) \approx f_G^{(N)}(y) = \frac{1}{\sqrt{2\pi} \frac{\sigma}{\sqrt{N}}} e^{-\frac{(y-m)^2}{2\left(\frac{\sigma}{\sqrt{N}}\right)^2}}. \quad (3.1)$$

Les hypothèses excluent donc une loi lorentzienne $f(x) = \frac{a}{\pi(x^2+a^2)}$ dont la valeur moyenne et la variance n'existent pas (cf TP 3).

Il faut d'abord noter que la fonction gaussienne $f_G^{(N)}$ qui apparaît au second membre de (3.1) est de plus en plus "piquée" quand $N \rightarrow \infty$: en effet elle est bien normalisée, $\int_{-\infty}^{\infty} dy f_G^{(N)}(y) = 1$, sa variance est de plus en plus petite $\sigma_N^2 = \frac{\sigma^2}{N}$ et son maximum de plus en plus grand $f_G^{(N)}(m) \propto \sqrt{N}$. Le graphe de $f_G^{(N)}$ est donc un pic de plus en plus haut et de plus en plus étroit au dessus de la valeur $y = m$, mais toujours d'aire 1. À la limite $N \rightarrow \infty$ la fonction $f_G^{(N)}$ est infinie en $y = m$, nulle partout ailleurs, mais on voit que son intégrale avec toute fonction régulière g de y donne $g(m)$

$$\int_{-\infty}^{\infty} dy f_G^{(N)}(y) g(y) \simeq g(m) \int_{-\infty}^{\infty} dy f_G^{(N)}(y) = g(m), \quad (3.2)$$

l'approximation devenant exacte dans la limite $N \rightarrow \infty$. Il ne s'agit plus d'une fonction, mais d'une *distribution*, et on appelle *distribution delta de Dirac* $\delta(x - m)$ (ou "fonction delta", par abus de langage) cette limite. On rencontrera ces distributions et en particulier la "fonction" δ dans le cours de physique quantique de L3.

On voit alors que le théorème limite central implique la loi des grands nombres : la loi de la moyenne Y est de plus en plus concentrée en $Y = m$, autrement dit la v.a. Y converge en probabilité vers la valeur moyenne m des v.a. X_i .

3.2. Éléments de preuve du théorème

La méthode de preuve est analogue à celle suivie pour la limite de la loi binomiale au § 5.1 du chap. 8. Elle repose sur la comparaison des fonctions caractéristiques de la v.a. Y et de la loi normale, quand $N \rightarrow \infty$. Puisque Y est la somme de N v.a. indépendantes $\frac{1}{N}X_i$ de même loi, sa fonction caractéristique vaut

$$\varphi_Y(u) = (\varphi_{X/N}(u))^N.$$

On va effectuer des développements limités (D.L.) des fonctions, au deuxième ordre en $1/N$. On sait que

$$\begin{aligned}\varphi_{X/N}(u) &= \left\langle e^{\frac{i u X}{N}} \right\rangle = 1 + \frac{i u}{N} \langle X \rangle - \frac{u^2}{2N^2} \langle X^2 \rangle + O\left(\frac{u^3}{N^3}\right) && \text{(D.L. de } e^{\frac{i u X}{N}}) \\ &= 1 + i m \frac{u}{N} - \frac{u^2}{2N^2} (\sigma^2 + m^2) + O\left(\frac{u^3}{N^3}\right)\end{aligned}$$

et donc pour le logarithme de la fonction caractéristique de Y

$$\begin{aligned}\ln \varphi_Y(u) &= N \ln \varphi_{X/N}(u) = N \ln \left(1 + i m \frac{u}{N} - \frac{u^2}{2N^2} (\sigma^2 + m^2) + O\left(\frac{u^3}{N^3}\right) \right) \\ &= N \left(i m \frac{u}{N} - \frac{u^2}{2N^2} \sigma^2 \right) + O\left(\frac{u^3}{N^3}\right) && \text{(D.L. de } \ln(1 + \dots)) \\ &= i m u - \frac{u^2}{2N} \sigma^2 + O\left(\frac{u^3}{N^2}\right) \\ &= \ln \varphi_{\text{Gauss}}^{(N)}(u) + O\left(\frac{u^3}{N^2}\right)\end{aligned}$$

où $\varphi_{\text{Gauss}}^{(N)}(u) = e^{i u m} e^{-\frac{1}{2} u^2 \sigma^2 / N}$ est la fonction caractéristique de la loi normale de moyenne m et de variance σ^2/N , cf équ. (3.5) du Chap. 8. Ce calcul montre que dans la limite $N \rightarrow \infty$, la fonction caractéristique de la moyenne arithmétique Y des N v.a. X_i tend vers la fonction caractéristique de la loi normale attendue. Nous admettrons que cela suffit à démontrer la convergence de la loi de Y vers la loi normale, c'est-à-dire le théorème limite central.

3.3. Illustrations

Le résultat du théorème limite central est remarquable : quelle que soit la loi des v.a. X_1, \dots, X_N , la loi limite de leur moyenne (1.3) est une loi normale ! La loi normale a ainsi un caractère d'*universalité*. Testons numériquement ce résultat sur des lois simples : les figures qui suivent présentent la densité de probabilité de la v.a. Y pour N v.a. X obéissant à une loi simple f , soit la loi uniforme, soit la loi "triangle" (cf. TD 7)

$$f(x) = \begin{cases} 4x & \text{si } 0 \leq x \leq \frac{1}{2} \\ 4(1-x) & \text{si } \frac{1}{2} \leq x \leq 1 \\ 0 & \text{sinon} \end{cases} \quad (3.3)$$

dont la moyenne est $\frac{1}{2}$ et la variance $\frac{1}{24}$. Pour chacune de ces lois, nous calculons numériquement la densité de Y donnée par l'expression que nous admettrons

$$\tilde{f}(y) = N \int dx_1 \cdots dx_{N-1} f(x_1) f(x_2) \cdots f(x_{N-1}) f(Ny - x_1 - \cdots - x_{N-1})$$

À titre d'exercice, effectuons ce calcul de \tilde{f} pour $N = 2$ v.a. obéissant à une loi uniforme entre 0 et 1

$$\tilde{f}(y) = 2 \int_0^1 dx_1 f(x_1) f(2y - x_1).$$

La première fonction $f(x_1)$ est non nulle (et égale à 1) ssi $0 \leq x_1 \leq 1$, la deuxième $f(2y - x_1)$ ssi $0 \leq 2y - x_1 \leq 1$. Il faut prendre l'intersection des deux domaines $0 \leq x_1 \leq 1$ et $2y - 1 \leq x_1 \leq 2y$, ce qui diffère selon que $y < \frac{1}{2}$ ou $y > \frac{1}{2}$. Si $0 \leq y \leq \frac{1}{2}$, $2y - 1 \leq 0$ mais $2y \leq 1$, donc $\tilde{f}(y) = 2 \int_0^{2y} dx = 4y$, tandis que si $\frac{1}{2} \leq y \leq 1$, $2y \geq 1$ mais $2y - 1 \geq 0$ donc $\tilde{f}(y) = 2 \int_{2y-1}^1 dx = 4(1-y)$. On constate que la loi \tilde{f} obtenue n'est autre que la "loi triangle" de (3.3), cf figure de gauche de la Fig. 1.

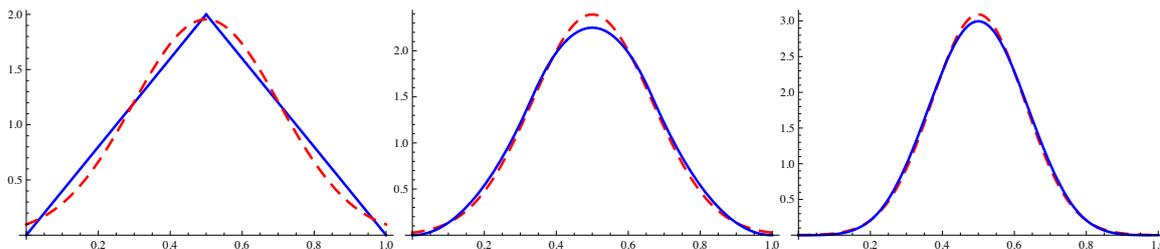


Fig. 1: Distribution $\tilde{f}(y)$ de la moyenne de N v.a. de loi uniforme, et comparaison avec la loi normale de même moyenne $m = \frac{1}{2}$ et de variance $\frac{1}{12N}$ (en ligne brisée). Successivement $N = 2, 3, 5$.

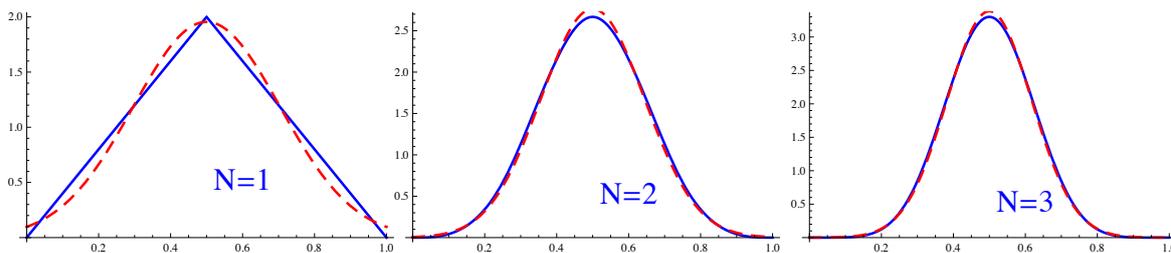


Fig. 2: Distribution $\tilde{f}(y)$ de la moyenne de N v.a. de loi "triangle" (3.3), et comparaison avec la loi normale de même moyenne $m = \frac{1}{2}$ et de variance $\frac{1}{24N}$ (en ligne brisée). Successivement $N = 1, 2, 3$.

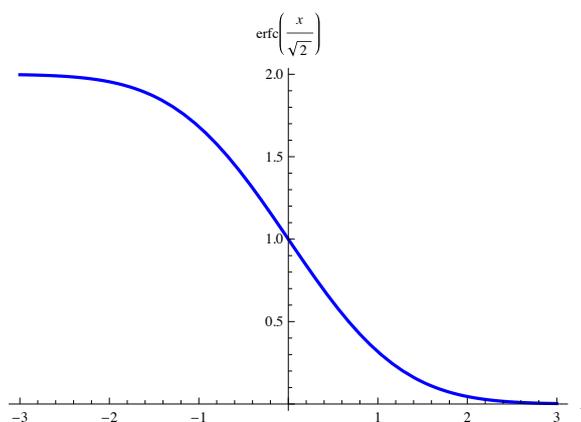


Fig. 3: Graphe de la fonction $\text{erfc}(x/\sqrt{2})$. On y voit que $\text{erfc}(1,96/\sqrt{2}) \simeq 0,05$ (plus exactement=0,0499958 !)

3.4. Évaluations de l'erreur

Réexprimons le théorème en termes de la v.a.

$$Z = \frac{Y - m}{\frac{\sigma}{N^{\frac{1}{2}}}} = \frac{\sum X_i - Nm}{\sigma N^{\frac{1}{2}}}$$

Elle est centrée $\langle Z \rangle = 0$ et le théorème nous dit qu'elle obéit asymptotiquement (quand $N \rightarrow \infty$) à une "loi normale centrée normalisée" (de variance 1). On en déduit que

$$\mathbf{P}(|Z| > \zeta) \rightarrow \frac{2}{\sqrt{2\pi}} \int_{\zeta}^{\infty} e^{-\frac{s^2}{2}} ds \quad (3.4)$$

ou de façon équivalente, dans les variables de départ,

$$\mathbf{P}(|Y - m| > \xi) = \mathbf{P}\left(\left|\frac{1}{N} \sum (X_i - m)\right| > \xi\right) \rightarrow \frac{2}{\sqrt{2\pi}} \int_{\xi N^{\frac{1}{2}}/\sigma}^{\infty} e^{-\frac{s^2}{2}} ds \quad (3.5)$$

En raison de leur importance en probabilités et en statistiques, on a donné un nom aux intégrales gaussiennes incomplètes

$$\begin{aligned} \operatorname{erf}(x) &= \frac{2}{\sqrt{\pi}} \int_0^x dt e^{-t^2} \\ \operatorname{erfc}(x) &= \frac{2}{\sqrt{\pi}} \int_x^{\infty} dt e^{-t^2} = 1 - \operatorname{erf}(x). \end{aligned} \quad (3.6)$$

Pour $x > 0$, on a

$$\operatorname{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{|t| < x} dt e^{-t^2}, \quad \operatorname{erfc}(x) = \frac{1}{\sqrt{\pi}} \int_{|t| > x} dt e^{-t^2}.$$

Ce sont respectivement la *fonction d'erreur* erf et la *fonction d'erreur complémentaire* erfc. Le graphe de la seconde est représenté sur la figure 3 et leurs valeurs sont tabulées dans la Table 1. On peut donc récrire (3.4) et (3.5)

$$\mathbf{P}(|Z| > \zeta) \rightarrow \operatorname{erfc}\left(\frac{\zeta}{\sqrt{2}}\right) \quad \text{ou} \quad \mathbf{P}(|Y - m| > \xi) \rightarrow \operatorname{erfc}\left(\frac{\xi N^{\frac{1}{2}}/\sigma}{\sqrt{2}}\right). \quad (3.7)$$

Avec (3.4) et (3.5) on a retrouvé la loi des grands nombres : la somme $\sum (X_i - m)$ est approximativement normale et centrée avec un écart-type $N^{-\frac{1}{2}}\sigma$. Mais les expressions (3.4)-(3.5) sont plus précises et permettent d'affiner l'estimation faite au paragraphe précédent de la taille d'un échantillon nécessaire pour atteindre une précision donnée. Reprenons l'exemple du § 2.2 : quel N permet d'obtenir $\mathbf{P}(|Y - m| > 0,01m) < 0,05$? D'après (3.7) il faut chercher N tel que $\operatorname{erfc}\left(\frac{N^{\frac{1}{2}}m}{100\sqrt{2}\sigma}\right) < 0,05$. On lit sur la Table 1 que $\operatorname{erfc}(x) = 0,05$ pour $x \approx \sqrt{2}$, (en fait $x = 1,386$), donc $N^{\frac{1}{2}} > N_0^{\frac{1}{2}} \approx 200 \frac{\sigma}{m}$ soit $N_0 \approx 40\,000 \left(\frac{\sigma}{m}\right)^2$. On a gagné un facteur 5 par rapport à l'estimation plus grossière venant de l'inégalité de Tchebychev !

x	erf(x)	erfc(x)	x	erf(x)	erfc(x)
0.	0	1.000000	1.3	0.9340079	0.06599206
0.05	0.05637198	0.9436280	1.4	0.9522851	0.04771488
0.1	0.1124629	0.8875371	1.5	0.9661051	0.03389485
0.15	0.1679960	0.8320040	1.6	0.9763484	0.02365162
0.2	0.2227026	0.7772974	1.7	0.9837905	0.01620954
0.25	0.2763264	0.7236736	1.8	0.9890905	0.01090950
0.3	0.3286268	0.6713732	1.9	0.9927904	0.00720957
0.35	0.3793821	0.6206179	2.	0.9953223	0.00467773
0.4	0.4283924	0.5716076	2.1	0.9970205	0.00297947
0.45	0.4754817	0.5245183	2.2	0.9981372	0.00186285
0.5	0.5204999	0.4795001	2.3	0.9988568	0.00114318
0.55	0.5633234	0.4366766	2.4	0.9993115	0.00068851
0.6	0.6038561	0.3961439	2.5	0.9995930	0.00040695
0.65	0.6420293	0.3579707	2.6	0.9997640	0.00023603
0.7	0.6778012	0.3221988	2.7	0.9998657	0.00013433
0.75	0.7111556	0.2888444	2.8	0.9999250	0.00007501
0.8	0.7421010	0.2578990	2.9	0.9999589	0.00004110
0.85	0.7706681	0.2293319	3.	0.9999779	0.00002209
0.9	0.7969082	0.2030918	3.1	0.9999884	0.00001165
0.95	0.8208908	0.1791092	3.2	0.9999940	6.025 10 ⁻⁶
1.	0.8427008	0.1572992	3.3	0.9999969	3.057 10 ⁻⁶
1.1	0.8802051	0.1197949	3.4	0.9999985	1.522 10 ⁻⁶
1.2	0.9103140	0.08968602	3.5	0.9999993	7.43 10 ⁻⁷

Table 1. Table des valeurs des fonctions erf(x) et erfc(x)*Remarques finales*

1. Comme noté plus haut, le théorème limite central ne s'applique pas à des "lois larges" comme la loi lorentzienne, dont la moyenne ou la variance n'existent pas.
2. Le fait que des lois normales apparaissent empiriquement (et approximativement) dans de très nombreux phénomènes naturels, taille ou poids d'individus dans une certaine population, erreurs dans les mesures successives d'une même grandeur, etc, pourrait trouver son

origine dans le fait que le phénomène en question résulte d'un grand nombre de variables aléatoires indépendantes *et de même loi*. Mais ceci explique-t-il vraiment cela ? ...

4. Marche aléatoire.

Une large classe de phénomènes physiques est modélisée par le problème de la marche au hasard, où les pas successifs d'un marcheur sont des v.a. indépendantes, tirées selon une certaine loi. On s'intéresse au comportement asymptotique de telles marches, dans la limite d'un grand nombre de pas élémentaires. C'est par exemple le cas du célèbre *mouvement brownien*¹³ qu'accomplissent des particules en suspension dans un fluide, sous l'effet des chocs avec les molécules de ce fluide : voir le site de Wikipedia "Brownian motion" pour de très jolies simulations du phénomène ou pour un fac-similé de l'observation du phénomène par le physicien Jean Perrin.

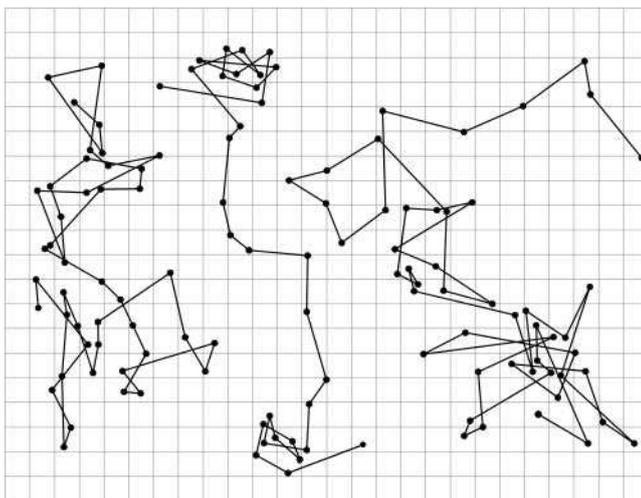


Fig. 4: Observation du mouvement brownien par le physicien Jean Perrin (1908) : la position de trois particules colloïdales (de diamètre $0,53 \mu\text{m}$), est observée sous le microscope et notée toutes les 30 secondes. Le pas de la grille est de $3,2 \mu\text{m}$. (Source : Jean Perrin, *Les Atomes*)

On va d'abord s'intéresser à une version simple du problème, où la marche se fait sur une grille, un *réseau* régulier.

¹³ du nom du botaniste Robert Brown (1773–1858) qui a le premier observé le phénomène sous le microscope

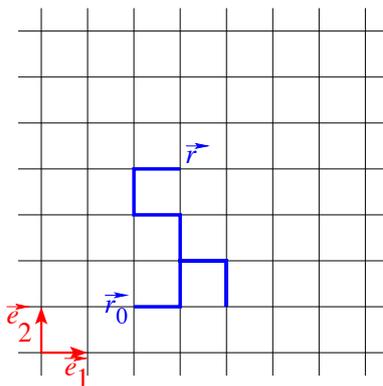


Fig. 5: Marche aléatoire sur un réseau carré

4.1. Marche au hasard sur un réseau

On considère un réseau carré de maille a , de vecteurs de base \vec{e}_1 et \vec{e}_2 (de norme a), voir figure 5¹⁴. Chaque site du réseau a 4 voisins. On considère un mobile (“marcheur”) qui à chaque instant multiple de τ , $t_n = n\tau$, saute du site où il se trouve vers un des 4 sites voisins. On suppose ce saut aléatoire et équidistribué : la probabilité de sauter sur n’importe lequel des sites adjacents est égale à $1/4$. Les sauts successifs sont supposés indépendants.

Soit $\mathbf{P}(\vec{r}, t_n | \vec{r}_0, t_0)$ la probabilité conditionnelle que le marcheur se trouve au point \vec{r} du réseau à l’instant t_n , sachant qu’il était en \vec{r}_0 à l’instant $t_0 = 0$. On va s’intéresser à la détermination de cette probabilité et à ses propriétés, en particulier dans la limite $n \gg 1$.

On observe d’abord que cette probabilité satisfait une condition initiale

$$\mathbf{P}(\vec{r}, t_0 | \vec{r}_0, t_0) = \delta_{\vec{r}, \vec{r}_0} \quad (4.1)$$

où $\delta_{\vec{r}, \vec{r}_0} = 1$ si $\vec{r} = \vec{r}_0$, et $= 0$ sinon, c’est un “symbole de Kronecker” (cf chap. 1) à deux dimensions : $\delta_{\vec{r}, \vec{r}_0} = \delta_{x, x_0} \delta_{y, y_0}$ en termes des composantes x et y des vecteurs \vec{r} et \vec{r}_0 . La probabilité \mathbf{P} satisfait aussi une condition de normalisation

$$\sum_{\vec{r}} \mathbf{P}(\vec{r}, t_n | \vec{r}_0, t_0) = 1 \quad (4.2)$$

où l’on somme sur tous les sites du réseau, condition qui exprime qu’à l’instant t_n , le marcheur se trouve bien quelque part !

¹⁴ Toute la discussion qui suit pourrait être menée en dimension arbitraire, sur un réseau “hypercubique” de dimension d . Nous nous restreignons à $d = 2$ dimensions pour la commodité de l’exposition et des dessins.

Moins trivial est le fait que \mathbf{P} satisfait aussi une relation de récurrence entre les temps t_n et t_{n+1} , qui exprime que le marcheur ne peut être en \vec{r} au temps t_{n+1} que s'il était au temps t_n en un des points voisins sur le réseau, $\vec{r} \pm \vec{e}_j$, $j = 1, 2$. Comme les événements correspondants (être en $\vec{r} \pm \vec{e}_j$ à l'instant t_n) sont exclusifs, leurs probabilités s'ajoutent et on a donc

$$\mathbf{P}(\vec{r}, t_{n+1} | \vec{r}_0, t_0) = \frac{1}{4} \sum_{\vec{r}' = \vec{r} \pm \vec{e}_j} \mathbf{P}(\vec{r}', t_n | \vec{r}_0, t_0) \quad (4.3)$$

(Cette relation rappelle et généralise la relation de récurrence du triangle de Pascal, qui correspondrait à un réseau à une dimension.) Cette relation de récurrence, complétée par la condition initiale, suffit en principe à déterminer complètement la probabilité \mathbf{P} .

Notons encore que le problème étant invariant par translation d'espace et de temps, la fonction $\mathbf{P}(\vec{r}, t_n | \vec{r}_0, t_0)$ ne dépend que des différences $\vec{r} - \vec{r}_0$ et $t_n - t_0 = n\tau$, et que pour des raisons dimensionnelles, elle ne peut dépendre que des rapports $\frac{1}{a}(\vec{r} - \vec{r}_0)$ et $\frac{t_n - t_0}{\tau} = n$

$$\mathbf{P}(\vec{r}, t_n | \vec{r}_0, t_0) = F\left(\frac{\vec{r} - \vec{r}_0}{a}, \frac{t_n - t_0}{\tau}\right). \quad (4.4)$$

4.2. Processus stochastiques, propriété de Markov

La marche au hasard sur un réseau que nous venons de définir est un exemple de *processus aléatoire* ou *processus stochastique*, c'est-à-dire d'un système dont l'évolution temporelle est probabiliste, au contraire d'un système *déterministe*. La marche au hasard a de plus la *propriété de Markov*, ce qui signifie que l'évolution au temps t ne dépend pas de l'histoire passée du système. (Ce ne serait pas le cas si on imposait que le marcheur ne recroise jamais sa trajectoire passée.)

Dans notre problème de marche au hasard sur le réseau, on a aussi la propriété que si $t' > t > t_0$,

$$\mathbf{P}((\vec{r}', t' | \vec{r}_0, t_0) = \sum_{\vec{r}} \mathbf{P}((\vec{r}', t' | \vec{r}, t) \mathbf{P}((\vec{r}, t | \vec{r}_0, t_0) \quad (4.5)$$

qui exprime qu'entre les instants t_0 et t' le marcheur passe en un temps intermédiaire t par un point \vec{r} , que les événements correspondant à des \vec{r} différents sont exclusifs et ont donc des probabilités qui s'ajoutent, et que les évolutions entre les temps t_0 et t et entre t et t' étant indépendantes (processus de Markov), leurs probabilités se multiplient.

4.3. Limite continue et équation de la chaleur

Si on soustrait aux deux membres de l'équation (4.3) la quantité $\mathbf{P}(\vec{r}, t_n | \vec{r}_0, t_0)$, elle peut être réécrite sous la forme suivante :

$$\begin{aligned} \mathbf{P}(\vec{r}, t_{n+1} | \vec{r}_0, t_0) - \mathbf{P}(\vec{r}, t_n | \vec{r}_0, t_0) &= \frac{1}{4} \sum_{\vec{r}' = \vec{r} \pm \vec{e}_j} (\mathbf{P}(\vec{r}', t_n | \vec{r}_0, t_0) - \mathbf{P}(\vec{r}, t_n | \vec{r}_0, t_0)) \\ &= \frac{1}{4} \sum_{j=1,2} (\mathbf{P}(\vec{r} + \vec{e}_j, t_n | \vec{r}_0, t_0) + \mathbf{P}(\vec{r} - \vec{e}_j, t_n | \vec{r}_0, t_0) - 2\mathbf{P}(\vec{r}, t_n | \vec{r}_0, t_0)). \end{aligned} \quad (4.6)$$

Supposons maintenant qu'on laisse les incréments de temps et d'espace τ et a tendre vers zéro. Dans cette limite, les coordonnées de temps et d'espace, qui étaient discrétisées, tendent vers des variables continues. On regarde donc la *limite continue* du problème initial. (De façon équivalente au vu de (4.4), cette limite décrit aussi la situation où les séparations d'espace $\vec{r} - \vec{r}_0$ ou de temps $t_n - t_0 = n\tau$ sont très grandes par rapport à a et τ .) Dans la limite $\tau \rightarrow 0$ et $a \rightarrow 0$, le membre de gauche de (4.6) est approximé par $\tau \frac{\partial}{\partial t} \mathbf{P}(\vec{r}, t_n | \vec{r}_0, t_0)$, celui de droite par $\frac{1}{4} a^2 \Delta \mathbf{P}(\vec{r}, t_n | \vec{r}_0, t_0)$ où Δ est le laplacien à 2 dimensions¹⁵

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \quad (4.7)$$

agissant sur les coordonnées x et y de \vec{r} . Dans cette limite l'équation (4.3) devient

$$\tau \frac{\partial}{\partial t} \mathbf{P} = \frac{1}{4} a^2 \Delta \mathbf{P} \quad (4.8)$$

qui est une équation fondamentale de la physique, décrivant les phénomènes de diffusion, (diffusion de particules dans un milieu, de chaleur etc). Selon le contexte, cette équation est appelée équation de Fick, *équation de la chaleur*, ...

La solution de l'équation (4.8) complétée par une condition initiale qui est la version continue de (4.1), voir ci-dessous, est bien connue :

$$F\left(\frac{\vec{r} - \vec{r}_0}{a}, \frac{t - t_0}{\tau}\right) = \mathbf{P}(\vec{r}, t | \vec{r}_0, t_0) = \frac{A\tau}{(t - t_0)} e^{-\frac{(\vec{r} - \vec{r}_0)^2}{a^2(t - t_0)/\tau}} \quad (4.9)$$

où on suppose $t - t_0 > 0$ et où la constante A va être fixée plus bas. *Exercice* : vérifier que cette fonction est bien solution de (4.8).

Dans la limite qui nous occupe où \vec{r} est une variable (aléatoire) continue, il faut remplacer la probabilité du marcheur d'être en un point donné \vec{r} par une densité de probabilité,

$$p(\vec{r}, t_n | \vec{r}_0, t_0) = \frac{1}{a^2} \mathbf{P}(\vec{r}, t_n | \vec{r}_0, t_0) \quad (4.10)$$

(qui a bien la dimension d'une *densité* à deux dimensions). La condition de normalisation (4.2), $\sum_{\vec{r}} a^2 \frac{\mathbf{P}(\vec{r}, t_n | \vec{r}_0, t_0)}{a^2} = \sum_{\vec{r}} a^2 p(\vec{r}, t_n | \vec{r}_0, t_0) = 1$ peut être considérée comme une somme de Riemann de l'intégrale de p et donne donc dans la limite $a \rightarrow 0$

$$\int d^2 \vec{r} p(\vec{r}, t_n | \vec{r}_0, t_0) = 1, \quad (4.11)$$

¹⁵ Cela généralise à deux dimensions la propriété simple que pour une fonction f deux fois continûment dérivable d'une variable réelle x et a infinitésimal, on a $f(x+a) + f(x-a) - 2f(x) = a^2 f''(x) + O(a^3)$ par le développement de Taylor au 2ème ordre.

qui est bien la condition de normalisation d'une densité de probabilité, comme nous l'avons vu, mais ici pour une v.a. à deux dimensions \vec{r} . Cette condition fixe la constante $A = \frac{1}{\pi}$, (cf les intégrales gaussiennes du chapitre 8), donc

$$p(\vec{r}, t_n | \vec{r}_0, t_0) = \frac{\tau}{\pi a^2 (t - t_0)} e^{-\frac{(\vec{r} - \vec{r}_0)^2}{a^2 (t - t_0) / \tau}}.$$

Le calcul jusqu'ici a supposé que les limites $a \rightarrow 0$ et $\tau \rightarrow 0$ étaient indépendantes. Afin de se débarrasser complètement de ces échelles "microscopiques", on voit qu'il convient de prendre $\tau \propto a^2$. On choisira l'échelle de temps de telle sorte que

$$\tau = \frac{1}{4} a^2 \quad (4.12)$$

(et plus généralement, en dimension d'espace d , on prendrait $\tau = \frac{1}{2d} a^2$). L'expression finale de la densité de probabilité p est donc

$$p(\vec{r}, t_n | \vec{r}_0, t_0) = \frac{1}{4\pi(t - t_0)} e^{-\frac{(\vec{r} - \vec{r}_0)^2}{4(t - t_0)}}, \quad (\text{pour } t > t_0) \quad (4.13)$$

(et plus généralement, en dimension d'espace d , le préfacteur serait $(4\pi(t - t_0))^{-d/2}$).

La condition initiale satisfaite par cette solution est obtenue en prenant la limite $t_n - t_0 \rightarrow 0$. Selon un argument donné plus haut, la fonction gaussienne est de plus en plus piquée au voisinage de $\vec{r} = \vec{r}_0$ mais toujours bien normalisée, et la limite est une "distribution delta de Dirac", cf § 3.1. On écrit donc

$$\lim_{t_n - t_0 \rightarrow 0} p(\vec{r}, t_n | \vec{r}_0, t_0) = \delta(\vec{r} - \vec{r}_0) = \delta(x - x_0) \delta(y - y_0) \quad (4.14)$$

qui est la version continue de la condition initiale (4.1).

Notons enfin que si $t' > t > t_0$, la propriété (4.5) se réécrit pour p selon

$$\int d^2 \vec{r} p(\vec{r}', t' | \vec{r}, t) p(\vec{r}, t | \vec{r}_0, t_0) = p(\vec{r}', t' | \vec{r}_0, t_0)$$

comme on le vérifiera en calculant cette *convolution* d'intégrales gaussiennes, un calcul qui après un changement de variables adéquat se ramène à nouveau à une intégrale d'une gaussienne.

L'apparition d'une gaussienne, c'est-à-dire d'une loi normale pour p , comme solution de notre marche au hasard ne doit pas surprendre : elle découle du théorème limite central, la trajectoire de \vec{r}_0 à \vec{r} résultant de l'addition d'un grand nombre $n = (t - t_0) / \tau$ de pas élémentaires (qui sont des v.a. indépendantes et de même loi).

4.4. Lois d'échelle dans la marche au hasard

Une propriété très remarquable de la solution (4.13) est son invariance par le changement simultané

$$\begin{aligned} \vec{r} - \vec{r}_0 &\rightarrow \lambda(\vec{r} - \vec{r}_0) ; t - t_0 \rightarrow \lambda^2(t - t_0) \\ p(\vec{r}, t_n | \vec{r}_0, t_0) &\rightarrow \lambda^{-2} p(\vec{r}, t_n | \vec{r}_0, t_0) \\ p(\vec{r}, t_n | \vec{r}_0, t_0) d^2\vec{r} &\rightarrow p(\vec{r}, t_n | \vec{r}_0, t_0) d^2\vec{r} \end{aligned} \quad (4.15)$$

où nous avons souligné que par cette dilatation, la *densité* de probabilité p est multipliée par un facteur λ^{-2} , tandis que la *probabilité* $p d^2\vec{r}$ que \vec{r} soit dans un petit domaine au voisinage de \vec{r} est invariante. Cette propriété qui relie les dilatations de l'espace et du temps apparaissait déjà dans la relation (4.12) entre les échelles microscopiques a et τ .

Une conséquence de cette invariance est que l'écart-type de la loi (4.13), $\langle (\vec{r} - \vec{r}_0)^2 \rangle^{\frac{1}{2}}$, qui décrit l'étalement de la distribution, c'est-à-dire le rayon typique R de la région couverte par le marcheur dans sa marche au hasard, croît comme

$$R = \langle (\vec{r} - \vec{r}_0)^2 \rangle^{\frac{1}{2}} \sim |t - t_0|^{\frac{1}{2}} . \quad (4.16)$$

Inversement la relation $|t - t_0| \sim R^2$ est interprétée en disant que la marche au hasard a une *dimension fractale*, (plus précisément une *dimension de Hausdorff*), égale à 2 : si le marcheur déroule une bobine de fil (fil d'Ariane ?) le long de sa trajectoire, à un taux constant par unité de temps, il aura consommé une quantité $\propto R^2$ de fil quand il se sera éloigné de R de son point de départ...

Cette invariance de dilatation (4.15) est un exemple de ce que l'on appelle une *loi d'échelle* et le nombre $\nu = \frac{1}{2}$ qui apparaît dans la puissance de $(t - t_0)$ dans (4.16) est un exemple d'*exposant critique*. Il est tout à fait remarquable que cet exposant soit indépendant du détail de la marche au hasard : pas discrets égaux à \pm les vecteurs de base du réseau, ou mouvement brownien avec des pas de longueur arbitraire, etc. Il ne dépend pas non plus de la dimension de l'espace où s'effectue cette marche. La raison étant une fois encore le théorème limite central et l'universalité de la loi normale qui en découle. En revanche, cet exposant serait modifié si on imposait que le marcheur ne recoupe jamais sa trajectoire passée...

L'étude de systèmes ayant un comportement d'échelle et des exposants critiques est un thème majeur de la physique contemporaine, qui s'applique par exemple au comportement d'un ferromagnétique au point de Curie, mais cela est une autre histoire !

★

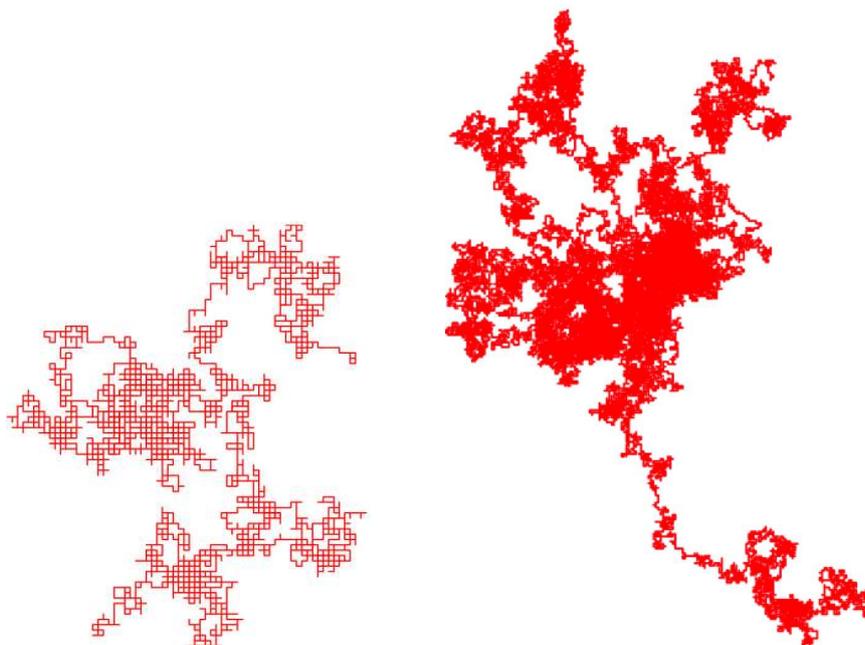


Fig. 6: Deux marches aléatoires sur réseau, vues à des échelles différentes. Sur la première, où on distingue bien les pas sur le réseau discret, on constate que le mouvement ne s'étale pas beaucoup, le marcheur revenant souvent sur ses pas. Cela est encore plus visible sur la seconde, où apparaît aussi une propriété remarquable du mouvement brownien continu, son invariance d'échelle (propriétés fractales) : le grossissement d'une petite partie de la "courbe" est de même nature que la courbe originale.