Introduction aux probabilités et aux statistiques

ENS Cachan - UPMC

Notes de Cours de Licence *L3-Phytem*

Nicolas Sator
Laboratoire de Physique Théorique de la Matière Condensée
Université Pierre et Marie Curie (Paris 6)

LPTMC - Septembre 2017

Préambule

Dans le champ de l'observation, le hasard ne favorise que les esprits préparés.

(Louis Pasteur, 1854)

De la théorie des jeux à la modélisation financière, en passant par la biologie et bien sûr la physique (en particulier la mécanique quantique et la physique statistique), le hasard est omniprésent.

Par manque d'information, que l'objet étudié soit trop complexe ou que nos connaissances sur son état soient trop imprécises, il n'est pas toujours possible (ou même nécessaire) de prévoir le résultat d'une expérience avec certitude. C'est par exemple le cas d'une pièce lancée au jeu de pile ou face. Bien que le mouvement de la pièce soit déterminé par les lois de la mécanique, le résultat du jeu reste incertain. Mais derrière cette incertitude, se cache une quasi-certitude : en lançant 1000 fois la pièce, on sait que l'on obtiendra environ 500 fois "pile". Un événement *aléatoire* sera donc imprévisible, mais sa probabilité d'occurrence pourra être estimée en répétant l'expérience un grand nombre de fois. [1]

Bien que l'utilisation des probabilités soit quotidienne (météo, sondages, jeux, économie...), l'intuition est souvent trompeuse comme nous le verrons dans la suite de ce cours. Pour éviter les erreurs, un préalable est de définir une expérience idéale qui modélise au mieux la réalité. Par exemple, au jeu de pile ou face, on supposera par défaut que seuls deux résultats sont possibles (la pièce, *parfaite*, ne tombe pas sur sa tranche) et qu'ils ont la même probabilité $p = 1/2$ (n'ayant pas d'autres informations sur l'objet ou son propriétaire, tous les événements ont la même probabilité). Mais la pièce peut être (un peu) biaisée...

1. Le mot *aléatoire* vient du latin *alea* et veut dire "jeu de dés", puis, par extension, hasard (qui vient de l'arabe *az-zahr* qui signifie aussi "dés"). Le hasard était déjà associé au jeu... En revanche, la notion d'ordre ou de régularité ressort de l'étymologie d'un synonyme du mot aléatoire : *stochastique*. En effet, ce mot plus savant venu du grec $\sigma\tau o\chi\alpha\sigma\tau\iota\kappa o\sigma$ n'a rien à voir avec le hasard, mais signifie plutôt "qui tend vers un but", puis "habile à conjecturer". Curieusement, le mot $\sigma\tau o\chi\alpha\sigma\tau\eta\sigma$ signifie le devin...

Table des matières

Chapitre I

Probabilités

1 Définitions

Depuis les travaux de A.N. Kolmogorov [1] (1933), les calculs de probabilités reposent sur la théorie de la mesure. [2] Dans la suite de ce cours, nous nous passerons de ce formalisme mathématique. Plus simplement, le calcul de probabilité est basé sur le trio suivant :

Événement

Un événement est le résultat possible d'une expérience idéale. En général, l'issue d'une expérience peut être décrite par un nombre entier (variable discrète, par exemple obtenue par un lancé de dé) ou par un nombre réel (variable continue, comme la taille d'un individu ou la vitesse d'une particule). Un événement peut être *élémentaire* (on dit aussi *simple*), par exemple faire 3 au dé, ou bien *composé* de plusieurs événements élémentaires (obtenir un nombre pair au dé, c'est-à-dire l'un de ces trois événements élémentaires : 2, 4 ou 6).

Espace des observables

L'*espace des observables* (également appelé *espace des événements*), noté Ω, est l'ensemble de tous les événements élémentaires d'une expérience. [3] Puisque l'espace des observables caractérise une expérience, il est fondamental de bien le définir, c'est-à-dire d'énumérer tous les résultats possibles. En revanche, la nature de ses points n'est pas importante : que l'on distribue des balles dans des boîtes ou des convives à des tables, le traitement est le même. Tout événement étant un ensemble d'événements élémentaires (au moins 1), on peut définir une structure algébrique munie de lois de compositions.

Probabilités

A tout événement élémentaire E_i de l'espace des observables Ω, on associe un nombre, la probabilité $p(E_i)$ d'obtenir l'événement E_i, telle que

- $p(E_i) \geq 0$ pour tout événement E_i,
- $Pr(E_i \textit{ ou } E_j) = p(E_i) + p(E_j)$. La probabilité d'un événement composé est donc la somme des probabilités de tous les événements élémentaires le constituant, [4]
- $\sum_{E_i \in \Omega} p(E_i) = p(\Omega) = 1$ (l'événement certain est de probabilité 1).

On passe du vocabulaire probabiliste à celui de la théorie des ensembles à l'aide des lois de compositions

1. Andreï Nikolaevich Kolmogorov (1903-1987) mathématicien russe qui posa les bases de la théorie des probabilités sous sa forme axiomatique.
2. On définit un espace de probabilités $(\Omega, \mathcal{C}, p)$, où Ω est un ensemble, $\mathcal{C}$ une *tribu* définie sur Ω, et p une mesure sur l'espace mesurable $(\Omega, \mathcal{C})$ telle que $p(\Omega) = 1$.
3. En mécanique statistique, cet ensemble s'appelle l'espace des phases, un événement est alors la donnée de la position $\vec{q}_i$ et de la quantité de mouvement $\vec{p}_i$ de toutes les particules.
4. On notera dans la suite $Pr(A)$, la probabilité de l'événement A décrit par l'argument.

suivantes. Soient A et B deux événements (a priori composés) :

- $A = \Omega$: l'événement A est certain et $p(A) = 1$ (obtenir un entier avec un dé).
- $A = \emptyset$: l'événement A est impossible et $p(A) = 0$ (obtenir 0 avec un dé).
- $A \subset B$: l'événement A implique l'événement B et $p(A) \leq p(B)$ (obtenir 2 implique un entier pair).
- L'événement $A^c = \Omega - A$ est l'événement complémentaire de A et $p(A^c) = 1 - p(A)$ (obtenir un entier pair et obtenir un entier impair).
- Loi de multiplication : $A \bigcap B$ (parfois notée $A.B$) signifie que les événements A *et* B sont réalisés. Les événements A et B sont dits *incompatibles* (ou *exclusifs ou disjoints*) si $A \bigcap B = \emptyset$ et donc $p(A \bigcap B) = 0$ (obtenir 2 et un nombre impair).
  Les événements compatibles A et B sont dits *indépendants* si et seulement si
  $$p(A \bigcap B) = p(A)p(B).$$
  (obtenir un nombre pair *et* obtenir un multiple de 3, c'est-à-dire 6, avec la probabilité $p = 1/2.1/3 = 1/6$).
- Loi d'addition : $A \bigcup B$ signifie que les événements A *ou* B sont réalisés ("ou" non exclusif, au moins l'un des deux) et
  $$p(A \bigcup B) = p(A) + p(B) - p(A \bigcap B).$$
  (obtenir un nombre pair *ou* obtenir un multiple de 3, c'est-à-dire 2, 4, 6 et 3 avec la probabilité $p = 1/2 + 1/3 - 1/6 = 2/3$).
  Cas particulier : si A et B sont *incompatibles* alors
  $$p(A \bigcup B) = p(A) + p(B).$$
  (obtenir 2 ou un nombre impair $p = 1/6 + 1/2 = 2/3$).

En pratique, on utilise une définition en fréquence : si on observe N événements *indépendants* lors d'une expérience dont n_i sont de type E_i, la probabilité de l'événement E_i est estimée par
$$p(E_i) = \lim_{N \to \infty} \frac{n_i}{N}.$$

Un exemple

Considérons l'ensemble des placements possibles de r balles *discernables* (on peut les distinguer les unes des autres) dans n boîtes. Chaque boîte peut recevoir un nombre illimité de balles. Explicitons les n^r événements élémentaires de Ω avec $r = n = 3$.

1.	$\{abc\|-\|-\}$	10.	$\{a\|bc\|-\}$	19.	$\{-\|a\|bc\}$
2.	$\{-\|abc\|-\}$	11.	$\{b\|ac\|-\}$	20.	$\{-\|b\|ac\}$
3.	$\{-\|-\|abc\}$	12.	$\{c\|ab\|-\}$	21.	$\{-\|c\|ab\}$
4.	$\{ab\|c\|-\}$	13.	$\{a\|-\|bc\}$	22.	$\{a\|b\|c\}$
5.	$\{ac\|b\|-\}$	14.	$\{b\|-\|ac\}$	23.	$\{a\|c\|b\}$
6.	$\{bc\|a\|-\}$	15.	$\{c\|-\|ab\}$	24.	$\{b\|a\|c\}$
7.	$\{ab\|-\|c\}$	16.	$\{-\|ab\|c\}$	25.	$\{b\|c\|a\}$
8.	$\{ac\|-\|b\}$	17.	$\{-\|ac\|b\}$	26.	$\{c\|a\|b\}$
9.	$\{bc\|-\|a\}$	18.	$\{-\|bc\|a\}$	27.	$\{c\|b\|a\}$

Table 1. *Les 27 placements de 3 balles numérotées a, b et c dans 3 boîtes.*

Par exemple, l'événement composé A, "une boîte est occupée plusieurs fois", est l'ensemble des événements 1 à 21. Et l'événement B, "la boîte de gauche n'est pas vide", est la réunion des événements 1, 4-15 et 22-27. On peut décider d'attribuer la même probabilité ($p_i = 1/27$) aux événements élémentaires [5] et s'intéresser par exemple à la probabilité de $A \bigcup B$ ou de $A \bigcap B$. Ainsi, $p(A) = 21/27$, $p(B) = 19/27$, $p(A \bigcap B) = 13/27$ et $p(A \bigcup B) = 1$. Les événements A et B ne sont donc pas indépendants.

5. Mais nous aurions pu choisir des balles "indiscernables", il n'y aurait alors plus que 10 éléments dans Ω. C'est la statistique quantique de Bose-Einstein.

2 Arrangements et dénombrements

Dans l'exemple précédent, le faible nombre de boîtes et de balles a permis de déterminer explicitement l'espace des observables. En général, on a recours à l'analyse combinatoire.

Soit un ensemble de n éléments, par exemple n boules numérotées dans une urne, on appelle *arrangement* de p éléments toute série *ordonnée* d'éléments *distincts* (tirage sans remise dans l'urne). Le nombre A_n^p d'arrangements de p éléments parmi n est :

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

En effet, les tirages sont indépendants et on a n possibilités pour le premier, $(n-1)$ pour le second et ainsi de suite. Un arrangement de n éléments parmi n s'appelle une *permutation* : $A_n^n = n! = n(n-1).(n-2)...2.1$ (par convention $0! = 1$). Remarque : si les éléments ne sont pas forcément distincts (tirage avec remise dans l'urne), le nombre de séries possibles est n^p, puisque l'on a n choix à chaque tirage indépendant.

Une collection *non ordonnée* de p éléments distincts parmi n est appelée *combinaison*. Le nombre $\binom{n}{p}$ de combinaisons de p éléments parmi n est le coefficient binomial (également noté C_n^p) :

$$\binom{n}{p} = C_n^p = \frac{n!}{p!(n-p)!}. \tag{I.1}$$

Ainsi, le nombre de combinaisons possibles de p boules dans une urne qui en contient n (sans remise dans l'urne), indépendamment de l'ordre de sortie, est $\binom{n}{p}$. Prenons l'exemple d'un ensemble de trois éléments $\{a, b, c\}$. Les arrangements possibles de deux éléments sont : ab, ba, ac, ca, bc et cb. Soit $A_3^2 = 6$ arrangements. En revanche il y a $\binom{3}{2} = 3$ combinaisons possibles : ab, ac et bc.

Quelques propriétés importantes :

- Symétrie : $\binom{n}{n-p} = \binom{n}{p}$
- Triangle de Pascal :

$$\binom{n}{p} = \binom{n-1}{p} + \binom{n-1}{p-1}.$$

  On démontre cette relation en isolant un élément parmi les p. Il y a deux cas possibles : Soit cet élément appartient à la sélection, il reste alors à choisir $p-1$ éléments parmi $n-1$. Soit il n'appartient pas à la sélection et il faut choisir p éléments parmi $n-1$.
- Binôme de Newton : soient x et y deux réels,

$$(x+y)^n = \sum_{p=0}^{n} \binom{n}{p} x^p y^{n-p}.$$

  Cette formule se démontre par récurrence en utilisant le triangle de Pascal. [6]

6. Pour $n = 0$ et $n = 1$ la formule est vérifiée. Par hypothèse de récurrence :

$$\begin{aligned}
(x+y)^{n+1} &= (x+y).\sum_{p=0}^{n} \binom{n}{p} x^p y^{n-p} \\
&= x^{n+1} + x.\sum_{p=0}^{n-1} \binom{n}{p} x^p y^{n-p} + y^{n+1} + y.\sum_{p=1}^{n} \binom{n}{p} x^p y^{n-p} \\
&= x^{n+1} + y^{n+1} + \sum_{p=1}^{n} [\binom{n}{p} + \binom{n}{p-1}] x^p y^{n-p+1} \\
&= x^{n+1} + y^{n+1} + \sum_{p=1}^{n} \binom{n+1}{p} x^p y^{n+1-p} = \sum_{p=0}^{n+1} \binom{n+1}{p} x^p y^{n+1-p}.
\end{aligned}$$

Exercice 1 *: les anniversaires*[†]

Quelle est la probabilité que les anniversaires de k personnes n'aient pas lieu le même jour ? En déduire la probabilité qu'au moins deux personnes aient le même jour d'anniversaire.

3 Probabilités conditionnelles

Axiome de Bayes :[7] Soient A et B deux événements compatibles de l'espace des observables Ω, on définit la *probabilité conditionnelle* $p(A|B)$ de réaliser l' événement A quand on a déjà réalisé l'événement B, par

$$p(A|B) = \frac{p(A \bigcap B)}{p(B)}. \tag{I.2}$$

C'est donc une probabilité définie sur le sous-ensemble $B \subset \Omega$. Ce qui donne la formule symétrique (et plus intuitive)

$$p(B|A)p(A) = p(A|B)p(B) = p(A \bigcap B).$$

Si les événements A et B sont *indépendants* alors $p(B|A) = p(B)$ (ou $p(A|B) = p(A)$) et on retrouve $p(A \bigcap B) = p(A)p(B)$.

Par exemple, la probabilité d'obtenir un multiple de 3 (événement A : 3 ou 6) sachant que l'on a obtenu un nombre pair (événement B : 2, 4 ou 6) est $p(A|B) = p(A \bigcap B)/P(B) = (1/6)/(1/2) = 1/3$.

Exercice 2 *: les dés pipés*[†]

Soient 100 dés dont la moitié est pipée de telle sorte que la probabilité de tirer un 6 vaut 1/2*. On choisit un dé au hasard, on le jette et on obtient un 6. Quelle est la probabilité que le dé choisi soit pipé ?*

---

†. **Réponse** : Pour la première personne, il y a 365 possibilités, pour la seconde 364, pour la troisième 363...donc, puisque les événements sont indépendants :

$$p = \frac{364}{365} \cdot \frac{363}{365} ... \frac{365-(k-1)}{365} = A_{365}^k / 365^k.$$

La probabilité pour qu'au moins deux personnes aient le même jour d'anniversaire est donc $1 - p$. Pour k petit, on a :

$$1 - p = 1 - (1 - \frac{1}{365})(1 - \frac{2}{365})...(1 - \frac{k-1}{365}) \simeq 1 - (1 - \frac{1+2+...+k-1}{365}) = k(k-1)/730.$$

Ainsi, pour $k = 20$ on trouve $1 - p \simeq 0.5$, ce qui n'est pas si intuitif.

7. Thomas Bayes (1702-1761), mathématicien et théologien anglais qui proposa une théorie des probabilités.

†. **Réponse** : Soit A l'événement "choisir un dé pipé" ($p(A) = 1/2$) et soit B l'événement "tirer un 6" ($p(B) = 1/2.1/6 + 1/2.1/2 = 1/3$). La probabilité de tirer un 6 alors que le dé est pipé est bien sûr $p(B|A) = 1/2$ et la probabilité, si on a tiré un 6, que le dé soit pipé est $p(A|B) = p(B|A).p(A)/p(B) = 3/4$.

Chapitre II

Variables aléatoires

Un événement aléatoire est un événement dont on ne peut pas déterminer l'issue, mais seulement lui associer une probabilité. Plus formellement, une *variable aléatoire* X est une fonction [1] définie sur l'espace des observables : à chaque point (événement) de cet espace, on associe un nombre, la probabilité que X vaille x, par une règle unique. Dans la suite, on notera en petites lettres les valeurs des variables aléatoires.

1 Distribution et densité de probabilité

Soit X une variable aléatoire *discrète* qui peut prendre une valeur appartenant à un ensemble discret $x_1, x_2, ...$, l'événement tel que $X = x_j$ a une probabilité $P(x_j)$, où P est une fonction définie sur les valeurs de X appelée la *distribution de probabilité* de la variable aléatoire X. C'est le cas par exemple du résultat obtenu par le lancé d'un dé. Les nombres inscrits sur les faces étant des entiers, ou pas.

Soit X une variable aléatoire *continue* qui peut prendre toute valeur réelle dans un certain intervalle de $\mathbf{R}$, on définit la densité de probabilité [2] $P(x)$ par la probabilité que la variable aléatoire X ait une valeur comprise entre x et $x + dx$:

$$P(x)dx = Pr(x \leq X < x + dx).$$

Notons que $P(x)dx$ étant une probabilité, il s'agit d'un nombre sans dimension. C'est le cas par exemple de la distribution en taille des individus. La taille d'une personne est un nombre réel mesuré avec une certaine précision dx.

La distribution/densité de probabilité vérifie les propriétés suivantes :
- $P(x_i) \geq 0$ pour tout x_i, ou $P(x) \geq 0$ pour tout x,
- $\sum_{x_i} P(x_i) = 1$, ou $\int P(x)dx = 1$ (normalisation).

Fonction de répartition

La fonction de répartition d'une variable aléatoire réelle est définie par

$$F(x) = Pr(X \leq x).$$

C'est une fonction croissante qui caractérise une variable aléatoire. Si la densité de probabilité existe on a alors :

$$F(x) = \int_{-\infty}^{x} P(t)dt.$$

Il s'agit en général d'une primitive de la densité de probabilité, sauf quand elle n'est pas dérivable partout (c'est le cas de la fonction de répartition de la loi uniforme ou de la loi exponentielle par exemple).

1. Il vaudrait d'ailleurs mieux parler de fonction aléatoire plutôt que de variable aléatoire.
2. Pour alléger les notations, on utilisera la même notation $P(x)$ pour une distribution et une densité de probabilité.

2 Moments d'une distribution

Toute l'information sur une expérience dont l'issue est une variable aléatoire X est contenue dans la distribution de probabilité $P(x)$. Pour décrire sommairement cette fonction, on utilise sa *valeur moyenne* $\langle X\rangle$ (ou espérance mathématique $E(X)$) et sa *variance* $Var(X)$.

Valeur moyenne

La *valeur moyenne* $\langle X\rangle$ est définie par

$$\langle X\rangle = \sum_j x_j P(x_j) \quad \text{ou} \quad \langle X\rangle = \int xP(x)\,dx.$$

Par rapport à la connaissance de la distribution, la moyenne est une perte d'information : "Un homme s'est noyé en traversant une rivière dont la profondeur moyenne est de 10 cm"...

Le terme "d'espérance mathématique" est très parlant dans l'exemple suivant : Au casino, un joueur mise 100 € sur l'un des 37 numéros de la roulette. Il a donc une chance sur 37 de gagner 35 fois sa mise et 36 "chances" sur 37 de la perdre. Son espérance de gain est donc $\langle X\rangle = \frac{1}{37}(3500) - \frac{36}{37}(100) \simeq -2.7$€...

Empiriquement, lorsqu'on réalise une expérience N fois, la valeur moyenne est estimée par

$$\langle X\rangle = \frac{1}{N}\sum_{p=1}^{N} x_p,$$

où x_p est la valeur de X obtenue la p^{eme} fois. Nous y reviendrons lorsque nous traiterons la loi des grands nombres au chapitre IV.

Soit X une variable aléatoire de densité de probabilité $P(x)$ et f une fonction, alors $F(X)$ est une nouvelle variable aléatoire qui peut prendre les valeurs $f(x)$. On appellera valeur moyenne de $F(X)$ la quantité

$$\langle F(X)\rangle = \int f(x)P(x)\,dx.$$

La densité de probabilité $\phi(f)$ de la variable aléatoire $F(X)$ est donnée par

$$\phi(f) = P[x(f)]\frac{1}{|f'(x)|},$$

si f est dérivable et monotone.[3]

Variance

Pour caractériser la largeur d'une distribution, ou l'écart à la moyenne,[4] on définit sa *variance* (ou écart quadratique moyen), notée $Var(X)$ ou σ^2, par la valeur moyenne du carré de l'écart par rapport à la moyenne :

$$Var(X) = \sigma^2 \equiv \langle (X - \langle X\rangle)^2\rangle = \langle X^2\rangle - \langle X\rangle^2 \geq 0.$$

La racine carrée de la variance, $\sigma = \sqrt{Var(X)}$, est appelée *écart type*. Par exemple, les températures moyennes annuelles à San Francisco et Tokyo sont très proches (environ $13.5C^o$), mais l'écart type est de $1.4C^o$ dans la première et de $2.8C^o$ dans la seconde.

Soient a et b deux réels, alors

$$\langle aX + b\rangle = a\langle X\rangle + b \quad \text{et} \quad Var(aX + b) = a^2 Var(X).$$

3. En effet, $\langle F(X)\rangle = \int \frac{f(x)}{|f'(x)|}P(x)df = \int f\phi(f)df$.
4. On a bien sûr : $\langle X - \langle X\rangle\rangle = 0$.

La translation ne modifie pas la variance d'une distribution.

On définit la variable *réduite* (normalisée) X^* associée à la variable aléatoire X par

$$X^* = (X - \langle X \rangle)/\sigma.$$

Ainsi $\langle X^* \rangle = 0$ et $Var(X^*) = 1$.

Moments

Plus généralement, une distribution $P(x)$ est caractérisée par son *moment d'ordre* n, où n est un entier positif ou nul, défini par :

$$\langle X^n \rangle = \sum_{x_j} x_j^n P(x_j) \quad \text{ou} \int x^n P(x)\, dx,$$

respectivement dans les cas de variable aléatoire discrète et continue. Le premier moment est donc la valeur moyenne et le second est lié à la variance. S'ils existent tous, leur connaissance permet en théorie de retrouver $P(x)$ (si P est analytique). La somme ou l'intégrale définissant le moment peut ne pas converger, dans ce cas, le moment n'existe pas. C'est le cas de la variance des lois larges que nous verrons plus loin. On remarquera que si le moment d'ordre n existe, alors tous les moments d'ordre inférieurs existent également.

3 Ensemble de variables aléatoires

On considère deux variables aléatoires continues X et Y dont les densités de probabilité sont respectivement $P(x)$ et $Q(y)$. La densité de probabilité *jointe*, $H(x, y)$ (par unité de X et de Y) est définie par

$$H(x, y)dxdy = Pr\left(x < X < x + dx \; et \; y < Y < y + dy\right).$$

Comme toute distribution de probabilité, $H(x, y)$ est normée. Si X et Y sont deux variables aléatoires indépendantes, alors :

$$H(x, y) = P(x)Q(y).$$

On définit la probabilité *conditionnelle* de l'événement $Y = y$ sachant que $X = x$ par

$$H(y|x) = \frac{H(x, y)}{\int H(x, y)\, dy}.$$

Remarquons que la fonction $H(y|x)$ ne dépend que de y pour un x donné, mais est différente de $Q(y)$.

La distribution de probabilité $P(x)$ est la distribution *marginale* de la variable X :

$$P(x) = \int_{y_{min}}^{y_{max}} H(x, y)dy.$$

Si X et Y sont deux variables aléatoires *indépendantes*, $S = X + Y$ est une nouvelle variable aléatoire dont la densité de probabilité est donnée par le produit de convolution [5]

$$H(s) = P * Q = \int P(t)Q(s - t)dt, \tag{II.1}$$

5. À cause de l'indépendance des variables aléatoires X et Y, la probabilité que $X \in [x, x + dx]$ et $Y \in [y, y + dy]$ vaut $P(x)Q(y)\, dx\, dy$. La probabilité $H(s)\, ds$ est donc

$$H(s)\, ds = \int\int_{\mathcal{D}} P(x)Q(y)\, dxdy = \int_{-\infty}^{\infty} P(x)\, dx \int_{s-x}^{s+ds-x} Q(y)\, dy,$$

où le domaine $\mathcal{D}$ est défini par $\{x + y \in [s, s + ds]\}$. L'intégrale sur y se calcule par le théorème de la moyenne et vaut $Q(s - x)\, ds$.

où $P(x)$ et $Q(y)$ sont respectivement les densités de probabilité de X et Y.

Exercice 3 *: Dérivation de la loi de Maxwell*[6] *de distribution des vitesses.*[†]

On considère un gaz à l'équilibre. Soit $P(v_x, v_y, v_z)\,dv_x\,dv_y\,dv_z$ la probabilité pour que dans l'espace des vitesses, l'extrémité du vecteur vitesse $\vec{v}$ soit dans le domaine $[v_x, v_x + dv_x]$, $[v_y, v_y + dv_y]$, $[v_z, v_z + dv_z]$.

1) Déterminer $P(v_x, v_y, v_z)$ à une constante près à l'aide des seules considérations de symétrie (homogénéité, isotropie, indépendance des 3 coordonnées). On introduira $F(v_x)dv_x$ la probabilité que $v_x \in [v_x, v_x + dv_x]$.

2) Montrer que la probabilité pour que le module v de la vitesse soit compris entre v et $v + dv$ est (distribution maxwellienne)

$$W(v)dv = \frac{4}{\sqrt{\pi}} b^{3/2} v^2 e^{-bv^2} dv.$$

Calculer v_0, la valeur la plus probable de v, sa valeur moyenne $\langle v \rangle$, puis sa valeur quadratique moyenne $\langle v^2 \rangle$. Dessiner $W(v)$.

4 Corrélations et indépendance

Soient X et Y deux variables aléatoires continues de densité de probabilité $P(x)$ et $Q(y)$ respectivement. Posons

$$\Delta X = X - \langle X \rangle \quad \text{et} \quad \Delta Y = Y - \langle Y \rangle.$$

On définit la *covariance* des variables X et Y par

$$cov(X, Y) = \langle \Delta X \Delta Y \rangle = \langle XY \rangle - \langle X \rangle \langle Y \rangle.$$

Ainsi, si $Y = aX$, $cov(X, Y) = a\sigma_X^2 = \sigma_X \sigma_Y$, la corrélation est maximale. En revanche, si les variables X et Y sont *indépendantes*, alors $cov(X, Y) = 0$ et donc

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

Un exemple qui intéresse beaucoup l'industrie textile est la corrélation entre la taille d'un individu et la longueur de sa jambe (une étude statistique a montré que $cov(X, Y) \simeq 0.8\sigma_X \sigma_Y$).

Soient $X_1, ..., X_N$ des variables aléatoires de variances *finies* $\sigma_1^2, ...\sigma_N^2$, alors $\sum_{i=1}^N X_i$ est une variable aléatoire de variance

$$\boxed{Var(\sum_{i=1}^N X_i) = \sum_{i=1}^N \sigma_i^2 + 2 \sum_{i<j} cov(X_i, X_j).} \tag{II.2}$$

Ainsi, la variance de la somme de variables aléatoires indépendantes est la somme de leurs variances.

6. James Clerk Maxwell (1831-1879) physicien anglais qui fonda la théorie de l'électromagnétisme et la théorie cinétique des gaz.

†. **Réponse** :

1) Puisque le système est isotrope, les distributions des composantes des vitesses selon les trois directions de l'espace sont les mêmes. En supposant l'indépendance des probabilités (ce qui est critiquable !) on a :

$$P(v_x, v_y, v_z)\,dv_x\,dv_y\,dv_z = F(v_x)F(v_y)F(v_z)\,dv_x\,dv_y\,dv_z.$$

La distribution P ne peut dépendre que de la norme de la vitesse (isotropie) donc

$$P(v_x, v_y, v_z) = F(v_x)F(v_y)F(v_z) = f(v_x^2 + v_y^2 + v_z^2).$$

Il est facile de vérifier que seule l'exponentielle satisfait cette équation. On obtient donc

$$P(v_x, v_y, v_z) = Ae^{-b(v_x^2+v_y^2+v_z^2)},$$

où A et b sont > 0. La normalisation entraîne $A = (\frac{b}{\pi})^{3/2}$. Il ne reste plus qu'un paramètre, b. On montre en physique statistique que $b = m/2k_bT$, où T est la température, k_b la constante de Boltzmann et m la masse d'une molécule.

2) Le vecteur vitesse pointe alors dans le volume $4\pi v^2 dv$ décrit par la sphère de rayon v et d'épaisseur dv. On trouve alors à l'aide de l'annexe 1 $v_0 = 1/\sqrt{b} = \sqrt{2k_bT/m}$, $\langle v \rangle = \sqrt{8k_bT/\pi m}$ et $\sqrt{\langle v^2 \rangle} = \sqrt{3k_bT/m}$.

L'indépendance de deux variables est donc une *condition suffisante, mais pas nécessaire,* pour qu'elles ne soient pas corrélées : l'absence de corrélation n'implique pas l'indépendance. Ainsi, des variables dépendantes peuvent ne pas être corrélées.

Exemple : Soit une distribution de probabilité $P(x)$ symétrique en $x = 0$ ($\langle X \rangle = 0$). Soit $Y = X^2$ une variable aléatoire, clairement dépendante de X ($\langle Y \rangle = \sigma^2$) alors $cov(X, Y) = \langle X^3 \rangle - \langle X \rangle \sigma^2 = 0$ (car $\langle X^3 \rangle = \langle X \rangle = 0$), X et Y sont donc dépendantes mais pas corrélées.

En résumé, l'indépendance implique l'absence de corrélation, mais la dépendance n'implique pas les corrélations (exemple ci-dessus). Et plus généralement, corrélation n'implique pas causalité...

5 Fonction caractéristique

Nous verrons par la suite que la fonction caractéristique est un outil puissant pour étudier une distribution de probabilité. La fonction caractéristique d'une variable aléatoire X est définie par $\phi(u) = \langle e^{iuX} \rangle$. Si la densité de probabilité $P(x)$ de la variable aléatoire existe,[7] $\phi(u)$ est la transformée de Fourier de $P(x)$:

$$\phi(u) = \langle e^{iuX} \rangle = \sum_j e^{iux_j} P(x_j) \quad \text{ou} \quad \phi(u) = \int e^{iux} P(x) dx.$$

Une distribution est donc entièrement déterminée par sa fonction caractéristique. Donnons quelques exemples d'utilisation de la fonction caractéristique.

- Calcul des moments de la distribution. En effet,
$$\langle X^n \rangle = \frac{1}{i^n} \frac{d^n \phi(u)}{du^n}\bigg|_{u=0}. \tag{II.3}$$
Une autre méthode pour calculer les moments est de développer $\phi(u)$ en puissance de u.
- Produit de convolution. Soient deux variables aléatoires X et Y, de densité de probabilité $P(x)$ et $Q(y)$. Exprimons la densité de probabilité $H(s)ds$ de la variable aléatoire $S = X + Y$.
La transformée de Fourier de $H(s)$ est :
$$\hat{H}(u) = \langle e^{iu(X+Y)} \rangle = \langle e^{iuX}.e^{iuY} \rangle.$$
Par ailleurs, si X et Y sont indépendantes
$$\langle e^{iuX}.e^{iuY} \rangle = \langle e^{iuX} \rangle \langle e^{iuY} \rangle = \hat{P}(u)\hat{Q}(u).$$
En appliquant la transformée de Fourier inverse et le théorème sur la transformée de Fourier du produit de convolution,[8] on retrouve l'équation (II.1) :
$$H(s) = P * Q = \int P(s-t)Q(t)dt.$$
- Somme de variables aléatoires indépendantes. Soient N variables aléatoires X_j, où $j = 1, 2, ...N$, *indépendantes* de distribution $P_j(x)$ dont la fonction caractéristique est $\phi_j(u)$. En utilisant à nouveau le produit de convolution, on montre que la fonction caractéristique de la variable $Y = \sum_{j=1}^{N} X_j$ est :
$$\phi_Y(u) = \langle e^{iuY} \rangle = \prod_{j=1}^{N} \phi_j(u).$$
Dans le cas particulier où les N variables ont la même distribution, on a donc
$$\phi_Y(u) = \phi_j^N(u).$$

7. Un contre-exemple : L'escalier de Cantor (une fonction continue "en escalier", dérivable presque partout et dont la dérivée est nulle presque partout) est la fonction de répartition d'une variable aléatoire qui n'a pas de densité de probabilité.

8. La transformée de Fourier d'un produit de convolution est le produit des transformées de Fourier :
$$\hat{H}(u) = \int e^{ius} H(s)ds = \int e^{ius} ds \int P(s-t)Q(t)dt = \int dt\, e^{iut} Q(t) \int ds\, e^{iu(s-t)} P(s-t) = \hat{P}(u)\hat{Q}(u).$$

Chapitre III

Quelques distributions de probabilité

Dans la pratique, seules quelques lois de probabilité interviennent fréquemment. Il s'agit en particulier de la loi binomiale, de la loi de Poisson et de la gaussienne (ou loi normale), dont les relations asymptotiques sont résumées sur la figure III.1.

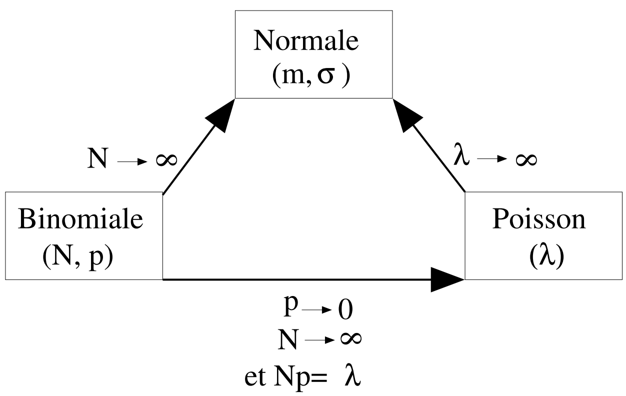

FIGURE III.1 – *Relations asymptotiques entre les distributions gaussienne, binomiale et de Poisson.*

distribution	variable	paramètres	fonction	valeur moyenne	variance
Uniforme	$a \leq x \leq b$	a,b réels	$P(x) = \frac{1}{b-a}$	$m = (b+a)/2$	$(b-a)^2/12$
Gaussienne	x réel	m réel, σ réel positif	$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$	m	σ^2
Binomiale	$k = 0, 1, ..., N$ entier	N entier, $0 \leq p \leq 1$	$P(k) = C_N^k p^k (1-p)^{N-k}$	Np	$Np(1-p)$
Poisson	$k = 0, 1, ...$ entier	λ réel positif	$P(k) = e^{-\lambda} \frac{\lambda^k}{k!}$	λ	λ

1 Distribution uniforme

C'est la distribution qui résulte d'une expérience dans laquelle chaque évé--nement a la même probabilité. On dit que les événements sont *équiprobables.* [1] Par exemple, lorsqu'on lance un dé parfait, chaque face a une probabilité $1/6$ d'apparaître.

Soit une variable aléatoire réelle [2] X dont les valeurs sont comprises entre a et b. La distribution de probabilité uniforme est donnée par

$$P(x) = \frac{1}{b-a} \quad si \quad a \leq x \leq b, \quad 0 \quad sinon. \tag{III.1}$$

Sa valeur moyenne est $\langle X \rangle = (b+a)/2$ et sa variance $Var(X) = (b-a)^2/12$. Dans le cas d'une variable entière, ces formules sont fausses en général (voir l'exercice 4).

Exercice 4 *: Lancé de dé* [†]

Le chiffre N_1 obtenu par un lancé de dé est distribué uniformément avec une probabilité $1/6$. Calculer la valeur moyenne et la variance de N_1.

Exercice 5 *: Somme de deux variables aléatoires uniformes* [‡]

Le lancé de deux dés fournit deux nombres aléatoires N_1 et N_2. Quelle est la distribution de la somme $N = N_1 + N_2$? Calculer la valeur moyenne et la variance de N.

2 Loi gaussienne ou normale

Nous verrons au chapitre suivant pourquoi cette loi de probabilité est si fréquente. Soit une variable aléatoire X continue qui varie sur $\mathcal{R}$ tout entier. La distribution gaussienne [3] (ou normale) dépend de deux paramètres réels m et $\sigma > 0$:

$$\boxed{P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}} \tag{III.2}$$

1. En physique statistique, c'est la distribution de probabilité des états microscopiques d'un système isolé (ensemble *microcanonique*).
2. La variable peut également être discrète comme dans le cas du lancé de dé.

†. **Réponse** : On trouve facilement en examinant les 6 événements possibles : $\langle N_1 \rangle = 3.5$ et $Var(N_1) = 35/12$. Attention, la formule de la variance d'une variable réelle aurait donné un résultat faux (25/12).

‡. **Réponse** : La probabilité $p(n)$ d'obtenir une somme égale à $n = 2, ...12$ est

$$p(n) = (1/6)^2 \sum_{n_1=1}^{6} \sum_{n_2=1}^{6} \delta(n - n_1 - n_2).$$

En énumérant tous les cas possibles, on trouve : $p(n) = (n-1)/36$ si $2 \leq n \leq 7$ et $p(n) = (13-n)/36$ si $8 \leq n \leq 12$. La somme de deux variables distribuées uniformément n'est donc pas distribuée uniformément : on dit que cette distribution n'est pas *stable* (voir chapitre IV). Puisque N_1 et N_2 sont indépendantes, on retrouve facilement $\langle N \rangle = 7$ et $Var(N) = 35/6$.

3. Carl Friedrich Gauss (1777-1855) mathématicien et physicien allemand.

Cette loi est bien normalisée ($\int_{-\infty}^{+\infty} P(x)\,dx = 1$) et sa valeur moyenne et sa variance sont données [4] par $\langle X \rangle = m$ et $Var(X) = \sigma^2$. La valeur moyenne est donc la valeur la plus probable. La fonction caractéristique de la loi normale est $\phi_n(u) = e^{ium - u^2\sigma^2/2}$. Ce qui permet, par exemple, de calculer les moments de la distribution en utilisant la relation (II.3).

Ainsi, on sait depuis le XIX^{eme} siècle que la taille des individus est approximativement distribuée sur une loi gaussienne décrite uniquement par sa moyenne m et sa variance σ^2. Par exemple, pour les japonais, $m = 165.5$ cm et $\sigma = 5.8$ cm, et pour les américains, $m = 175.5$ cm et $\sigma = 7.1$ cm.

Il est intéressant de garder en tête les valeurs des probabilités d'écart à la moyenne :

$$Pr(m - \sigma < X < m + \sigma) \simeq 68\%$$

et

$$Pr(m - 2\sigma < X < m + 2\sigma) \simeq 96\%.$$

Ces probabilités se "lisent" clairement sur la figure III.2 comme l'aire sous la courbe avec une base de largeur 2σ et 4σ respectivement. On voit que contrairement à une loi large (voir section 6) une variable gaussienne s'écarte peu de sa moyenne.

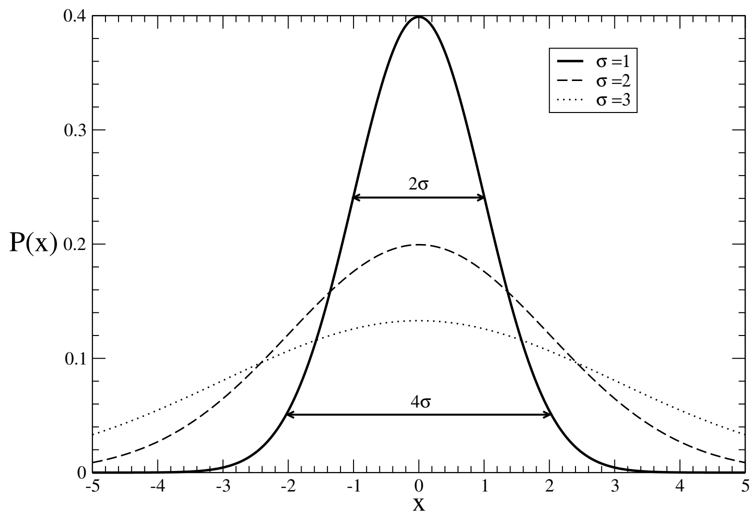

FIGURE III.2 – *Distributions normales centrées pour $\sigma = 1, 2$ et 3.*

Quand une variable aléatoire a une densité de probabilité gaussienne, on dit qu'elle est *normale*, si elle a une moyenne nulle elle est *centrée* et si sa variance vaut 1, elle est *réduite.* [5]

4. En effet, Il faut calculer

$$\langle X \rangle = \int x P(x)\,dx = \frac{1}{\sqrt{\pi}} \left[\int_{-\infty}^{\infty} \sigma\sqrt{2} t e^{-t^2}\,dt + m \int_{-\infty}^{\infty} e^{-t^2}\,dt \right],$$

où l'on a effectué le changement de variable $t = \frac{x-m}{\sigma\sqrt{2}}$. La première intégrale est nulle par parité et la seconde vaut $\sqrt{\pi}$ (voir l'annexe 1). Calculons maintenant

$$\langle (X - \langle X \rangle)^2 \rangle = \int (x-m)^2 P(x)\,dx.$$

Posons

$$J(\sigma) = \int e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx = \sqrt{2\pi}\sigma$$

et dérivons par rapport à σ

$$\frac{dJ}{d\sigma} = \frac{1}{\sigma^3} \int (x-m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx,$$

d'où on tire immédiatement

$$\langle (X-m)^2 \rangle = \sigma^2.$$

5. Voir la définition d'une variable réduite section 2.

3 Loi binomiale

Cette loi intervient lorsque l'on répète *indépendamment* un événement qui n'a que deux issues possibles de probabilité p et $q = 1 - p$ respectivement ("Bernouilli trials" en anglais). Par exemple, le jeu de "pile ou face" avec une pièce parfaite implique $p = q = 1/2$. La loi binomiale est la probabilité de réaliser k fois exactement un événement de probabilité p en effectuant N essais *indépendants* (la variable aléatoire entière K varie donc de 0 à N) :

$$\boxed{P(k) = C_N^k p^k (1-p)^{N-k}} \tag{III.3}$$

où C_N^k est le nombre de façons de choisir k objets parmi N, donné par l'équation (I.1). On montre facilement † que $\sum_{k=0}^{N} P(k) = 1$ (normalisation), $\langle K \rangle = Np$ et $Var(K) = Np(1-p)$. La fonction caractéristique de la loi binomiale est $\phi_b(u) = (pe^{iu} + 1 - p)^N$.

Quand $N \to \infty$, la distribution binomiale (III.3) tend vers la loi gaussienne (pour p pas trop proche de 0 ou 1) § (III.2), où $k \in [0, N]$ devient une variable continue $x \in [0, +\infty]$, avec $\langle X \rangle = pN$ et $\sigma^2 = Np(1-p)$. Cette propriété de la loi binomiale est illustrée par la figure III.3.

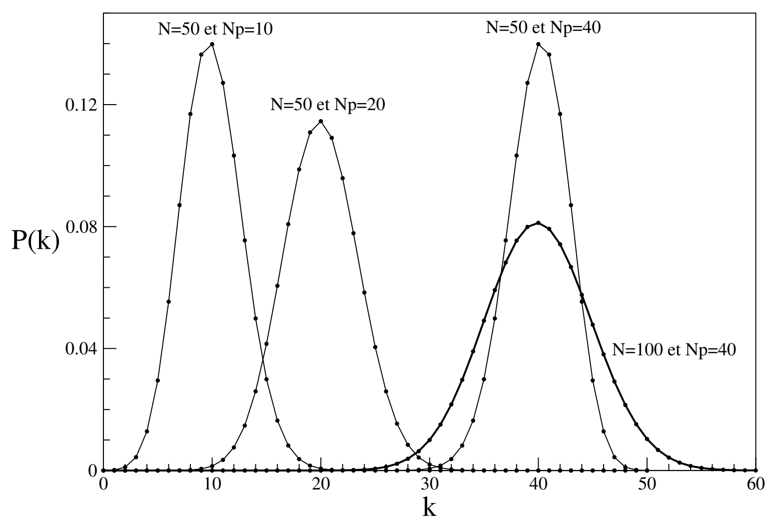

FIGURE III.3 – *Distributions binomiales pour différentes valeurs de N et de Np. Contrairement aux trois courbes calculées pour $N = 50$, la distribution obtenue pour $N = 100$ et $Np = 40$ est indiscernable d'une loi gaussienne (avec $m = 40$ et $\sigma^2 = 24$).*

†. A l'aide de la formule du binôme de Newton, $(p+q)^N = \sum_{k=0}^{N} \binom{N}{k} p^k q^{N-k}$, que l'on peut dériver par rapport à p, ou en utilisant la fonction caractéristique $\phi_b(u)$ et la relation II.3.

§. Pour le montrer, on peut utiliser la fonction caractéristique de la distribution binomiale $\phi_b(u)$ et montrer qu'elle tend vers la fonction caractéristique de la loi normale $\phi_n(u)$ lorsque $u \to 0$ (car $k \to \infty$ pour p pas trop proche de 0 ou 1). En effet,

$$\begin{aligned}
\ln \phi_b(u) &= N \ln(1 + p(e^{iu} - 1)) = N[p(e^{iu} - 1) - \frac{p^2}{2}(e^{iu} - 1)^2 + ...] \\
&= N[p(iu - \frac{u^2}{2} + ...) - \frac{p^2}{2}(iu - \frac{u^2}{2} + ...)^2 + ...] \\
&\simeq iuNp - \frac{u^2}{2} Np(1-p) \simeq \ln \phi_n(u).
\end{aligned}$$

On reconnaît la fonction caractéristique de la gaussienne au deuxième ordre en u.

Exercice 6 *: Lapins verts*†

Un lapin sur 100 est vert (si, si...). On prend aléatoirement 100 lapins dans un clapier, quelle est la probabilité qu'il y en ait un (exactement) vert ?

4 Loi multinomiale

C'est une généralisation de la loi binomiale. Une variable aléatoire K peut prendre r valeurs $(k_1, k_2, \cdots k_r)$ avec les probabilités respectives $p_1, p_2, \cdots p_r$ telles que $\sum_{i=1}^r p_i = 1$. La probabilité d'obtenir k_i tirages (indépendants) de $K = k_i$ en N essais ($\sum_{i=1}^r k_i = N$) est donnée par

$$P(k_1, k_2 \cdots k_r) = \frac{N!}{k_1!k_2! \cdots k_r!} p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}.$$

La démonstration est évidente : $p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$ est la probabilité d'obtenir le résultat dans un ordre donné et la dégénérescence vaut

$$\frac{N!}{k_1!k_2! \cdots k_r!} = \binom{N}{k_1}\binom{N-k_1}{k_2} \cdots \binom{N-k_1-k_2 \cdots - k_{r-1}}{k_r}.$$

C'est la distribution de probabilité obtenue lors de tirages indépendants de boules de r couleurs différentes contenues dans une urne. Les valeurs moyennes, les variances et les covariances sont données respectivement par $\langle K_i \rangle = np_i$, $Var(K_i) = Np_i(1-p_i)$ et $cov(K_i, K_j) = -Np_ip_j$.

5 Loi de Poisson

Lorsque des événements *indépendants* se produisent aléatoirement avec un taux constant (nombre d'événements par unité de temps ou d'espace), la loi de Poisson [6] est la probabilité d'en trouver $k = 0, 1, 2, ...$ alors que l'on en attend λ en moyenne :

$$\boxed{P(k) = e^{-\lambda}\lambda^k/k!} \tag{III.4}$$

On montre facilement que la loi de Poisson est normalisée et que la valeur moyenne de K et sa variance sont égales : $\langle K \rangle = Var(K) = \lambda$. La fonction caractéristique de la loi de Poisson est $\phi_p(u) = e^{\lambda(e^{iu}-1)}$.

Considérons par exemple N appels téléphoniques, [7] passés au hasard et indépendamment pendant un temps T. On s'attend donc à recevoir $\lambda = N/T$ appels *en moyenne* par unité de temps. La probabilité de recevoir k appels pendant une durée $t \leq T$ est donnée par la distribution de Poisson de paramètre λt.

La loi de Poisson est la limite de la loi binomiale (III.3) quand $N \to \infty$ (nombreux essais) et $p \to 0$

†. **Réponse** : La probabilité pour que le *premier* lapin choisi soit vert et qu'aucun des 99 suivants ne le soit est $1/100$ multiplié par $(99/100)^{99}$ (les probabilités indépendantes se multiplient). Maintenant on ne demande pas qu'un lapin déterminé (le premier) soit vert. Ce peut être le second, le troisième ou ...le dernier. Il y a 100 telles possibilités. Comme elles s'excluent mutuellement, la probabilité qu'un lapin *quelconque* soit vert est donc 100 fois la probabilité précédente, soit $(99/100)^{99} \simeq 1/e \simeq 0.37$. On aurait pu aussi partir directement de la loi binomiale (III.3) et écrire que

$$P(1) = \binom{100}{1}\frac{1}{100}\left(1 - \frac{1}{100}\right)^{99}.$$

"Un lapin sur 100 est vert" ne signifie donc pas que sur 100 lapins tirés au hasard il y en a nécessairement un qui le soit. Par contre, si on effectue une série de 100 prélèvements dans le clapier, on aura bien *en moyenne* $Np = 100/100 = 1$ lapin sur 100 vert.

6. Siméon Denis Poisson (1781-1840) mathématicien français.

7. Mais nous aurions pu étudier le nombre de particules α émises par un matériau radioactif pendant un intervalle de temps donné ou la distribution spatiale des impacts de bombes tombées sur Londres pendant la seconde guerre mondiale.

(événement rare) de telle sorte que $Np = \lambda$ soit fini.[¶] On peut montrer ce résultat de la façon suivante :

$$\begin{aligned}
P(k) &= \frac{N!}{k!(N-k)!}p^k(1-p)^{N-k} \\
&= \frac{p^k}{k!}\underbrace{N(N-1)...(N-k+1)}_{\sim N^k} e^{\overbrace{(N-k)}^{\sim N}\overbrace{\ln(1-p)}^{\sim -p}} \\
&\simeq \frac{p^k}{k!}N^k e^{-pN} \\
&\simeq \frac{\lambda^k}{k!}e^{-\lambda},
\end{aligned}$$

où $\lambda = NP$. Par ailleurs, la loi de Poisson (III.4) tend, lorsque $\lambda \to \infty$, vers la loi gaussienne[‡] (III.2). Ces propriétés sont illustrées par la figure III.4.

Exercice 7 *: Anniversaire*[§]

Quelle est la probabilité pour que dans un amphi de $N = 500$ étudiants, exactement k personnes aient leur anniversaire le 16 janvier ?

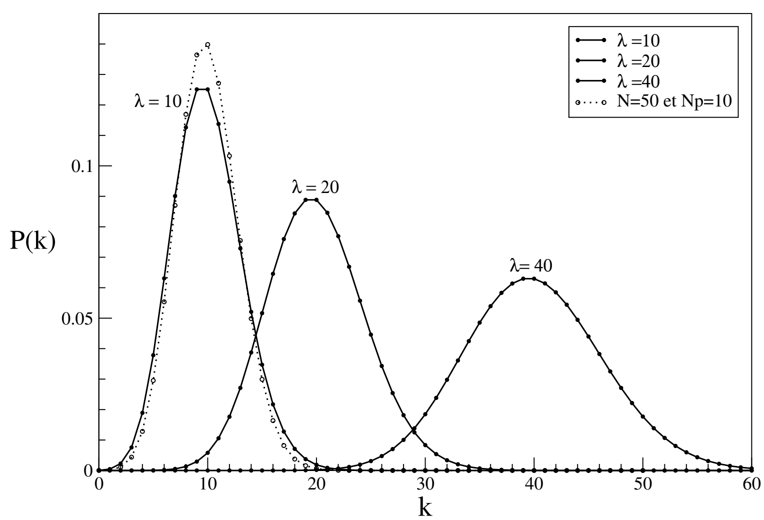

FIGURE III.4 – *Distributions de Poisson pour différentes valeurs de λ. La distribution binomiale ($N = 50$ et $Np = 10$) est représentée en pointillés à titre de comparaison. La loi de Poisson ($\lambda = 10$) est indiscernable de la loi binomiale ($N = 1000$ et $Np = 10$). De même la loi de Poisson ($\lambda = 40$) est indiscernable d'une distribution gaussienne ($m = \sigma^2 = 40$).*

¶. A l'aide des fonctions caractéristiques $\phi_b(u)$ et $\phi_p(u)$ des distributions binomiale et de Poisson quand $N \to \infty$ et $p \to 0$:

$$\begin{aligned}
\ln\phi_b(u) &= N\ln(1+p(e^{iu}-1)) = (Np)(e^{iu}-1) - \frac{(Np)^2}{2N}(e^{iu}-1)^2 + ...] \\
&\simeq Np(e^{iu}-1) = \ln\phi_p(u).
\end{aligned}$$

‡. Une fois de plus :

$$\ln\phi_p(u) = \lambda(e^{iu}-1) = \lambda[iu - \frac{u^2}{2} + ...] \simeq \ln\phi_n(u).$$

§. **Réponse** : La probabilité qu'une personne ait son anniversaire à une date donnée est $p = 1/365$. La probabilité recherchée est la distribution de Poisson en prenant $\lambda = 500/365$. Comparer ce résultat pour $k = 0, 1, 2, 3$ avec la probabilité obtenue avec la distribution binomiale.

6 Lois larges et lois de puissance

Une distribution est dite *large* lorsqu'elle n'a pas de second moment fini (variance non définie).

Loi de Cauchy

C'est le cas, par exemple, de la loi de Cauchy [8] (également appelée loi de Lorentz) représentée sur la figure III.5 :

$$P(x) = \frac{a}{\pi} \frac{1}{a^2 + (x - x_0)^2}. \tag{III.5}$$

Plus le paramètre $a > 0$ est grand, plus cette distribution, centrée en x_0, est étalée. La loi de Cauchy est bien normalisée, n'admet ni de valeur moyenne, ni de moment d'ordre plus élevé (loi large) et sa fonction caractéristique
est [9] $\phi(u) = e^{ix_0 u - a|u|}$.

Les deux distributions représentées sur la figure III.5 semblent très proches, mais le comportement asymptotique en loi de puissance ($\sim 1/x^2$ pour $x \to \infty$) de la loi de Cauchy permet des valeurs élevées de la variable aléatoire qui conduisent à des comportement très différents d'une loi gaussienne.

.

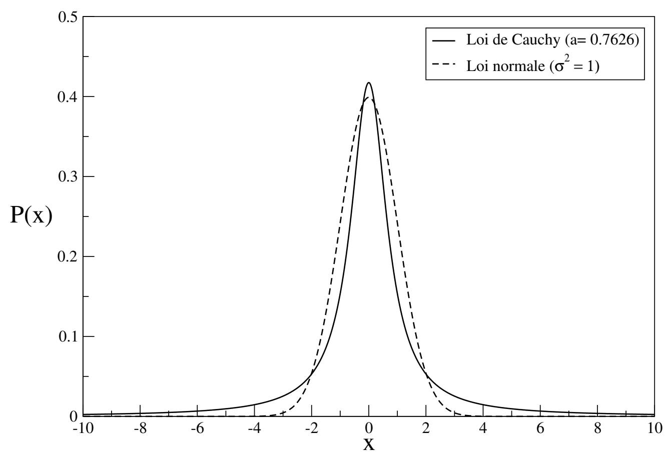

FIGURE III.5 – *Loi de Cauchy pour* $a = 0.7626$ *et loi gaussienne centrée et réduite.*

Lois de puissance

Les distributions en loi de puissance sont de la forme suivante :

$$P(x) = cx^{-\tau},$$

où c est une constante. Si l'exposant $\tau \leq 3$, ces lois sont larges (et si $\tau \leq 2$, non seulement la variance mais également la valeur moyenne divergent). Ce sont les seules lois qui n'ont pas de grandeur

8. Augustin Louis Cauchy (1789-1857) mathématicien français.
9. On a bien

$$\frac{a}{\pi} \int_{-\infty}^{+\infty} \frac{1}{a^2 + (x - x_0)^2} dx = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{1}{1 + x^2} dx = \frac{1}{\pi} [\arctan(\infty) - \arctan(-\infty)] = 1.$$

La moyenne, $\langle X \rangle = \frac{a}{\pi} \int \frac{x}{a^2 + (x - x_0)^2} dx$, n'existe pas puisque l'intégrand se comporte en $1/x$ pour $x \to \infty$. De même pour $\langle X^2 \rangle$. La fonction caractéristique est :

$$\phi(u) = \frac{a}{\pi} \int_{-\infty}^{\infty} e^{iux} \frac{1}{a^2 + (x - x_0)^2} dx = e^{iux_0} e^{-a|u|}.$$

caractéristique, c'est-à-dire que leur allure ne dépend pas de *l'échelle* à laquelle on les étudie. Les lois de puissances interviennent dans les phénomènes physiques où le système présente le même aspect quelle que soit l'échelle d'observation. C'est le cas en particulier des transitions de phase continues (point critique thermodynamique, seuil de percolation...).

Par ailleurs, de nombreux phénomènes donnent lieu à des distributions dont le comportement asymptotique (pour x grand) est en loi de puissance.[10] Les exemples suivants sont illustrés par la figure III.6 :

- La loi de Zipf : la variable X est alors le rang d'un mot dans un texte (en anglais, les mots les plus courants sont "the", "of" ...) et $P(x)$ sa probabilité d'occurrence. Empiriquement, $\tau \simeq 2.2$.
- La distribution des tremblements de terre en fonction de leur magnitude : $\tau \simeq 3.04$.
- La loi de Pareto[11] (1896 ou 1909 ? ? ?) donne le nombre d'individus $P(X)$ possédant une certaine "richesse" X. Cette loi semble bien vérifiée dans des situations très diverses pour les revenus les plus élevés. L'exposant τ caractérise alors la violence de l'inégalité sociale. Dans l'exemple de la figure III.6, $\tau \simeq 2.09$ et la richesse la plus élevée aux USA, 46 10^9\$, est celle d'un certain William H. Gates III. En Inde, pour les années 2002-2004, on trouve $\tau \simeq 1.9$.

.

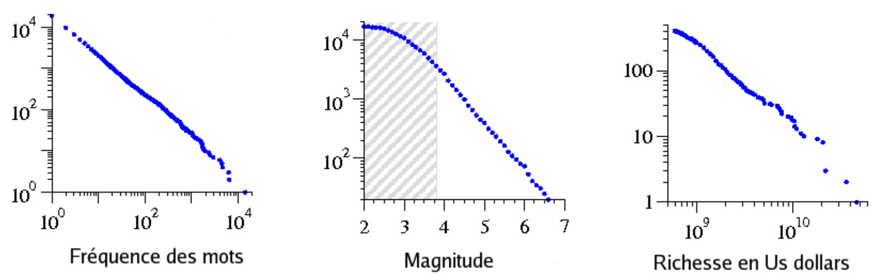

FIGURE III.6 – *Exemples de distributions cumulées ($\int_x^\infty P(x')dx'$, l'exposant vaut donc $\tau - 1$) en loi de puissance (d'après M.E.J. Newman, Contemporary Physics 46, 323-351 (2005)). De gauche à droite : Nombres d'occurrences des mots du roman "Moby Dick" de Hermann Melville ; Magnitudes des tremblements de terre en Californie entre janvier 1910 et mai 1992 ; Revenus en dollar des américains les plus riches en octobre 2003 (voir texte).*

10. On trouvera divers exemples dans : M.E.J. Newman, "Power laws, Pareto distribution and Zipf's law" , Contemporary Physics 46, 323-351 (2005). Disponible sur http ://fr.arxiv.org/abs/cond-mat/0412004.

11. Economiste libéral qui étudia la distribution des richesses en 1909 pour différents pays et différentes époques (il trouva un exposant de 1,5 environ pour décrire les hauts revenus).

Chapitre IV

Loi des grands nombres et théorème central limite

Ces théorèmes jouent un rôle fondamental en statistiques, mais ils sont parfois appliqués (à tort) en dehors de leur domaine de validité.

Soient $X_1,\ X_2 \ldots\ X_N$ des variables aléatoires mutuellement *indépendantes* ayant la même [1] distribution de probabilité, c'est-à-dire la même moyenne *finie* $\langle X_i \rangle = m$ et la même variance, *pas nécessairement finie*, σ^2. Leur somme divisée par N, qui s'écrit

$$M_N = \frac{S_N}{N} = \frac{1}{N}\sum_{i=1}^{N} X_i,$$

est aussi une variable aléatoire. On a évidemment, *en moyenne*, $\langle M_N \rangle = m$ pour tout N : la moyenne de la somme est la somme des moyennes.

1 La loi des grands nombres

La loi des grands nombres stipule que pour $N \to \infty$, $M_N \to m$ en probabilité :

$$\lim_{N\to\infty} Pr\left(|M_N - m| > a\right) = 0 \quad \forall a > 0.$$

Autrement dit, la probabilité que la variable aléatoire M_N s'écarte de m d'une valeur a arbitraire tend vers 0. Ce résultat intuitif a été montré par Khintchine en 1929.

D'après la relation (II.2), et puisque les variables sont indépendantes, on a [2]

$$Var(M_N) = \frac{\sigma^2}{N},$$

soit un écart type qui tend vers 0 en $1/\sqrt{N}$. La somme M_N des N variables aléatoires X_i/N a donc $\langle M_N \rangle = m$ comme valeur moyenne et $Var(M_N) = \frac{\sigma^2}{N}$ comme variance.

Plus formellement on peut commencer par démontrer l'inégalité de Bienaymé-Chebychev. Soit X une variable aléatoire de densité de probabilité $P(x)$. On suppose d'abord que $\langle X \rangle = 0$ et que $\langle X^2 \rangle$ existe. On a la suite d'inégalités :

$$Pr(|X| > a) = \int_{|x|>a} P(x)\,dx < \int_{|x|>a} \left|\frac{x}{a}\right|^2 P(x)\,dx < \int_{-\infty}^{\infty} \left|\frac{x}{a}\right|^2 P(x)\,dx = \frac{\langle X^2 \rangle}{a^2}.$$

1. Plus généralement, il n'est pas nécessaire que les N variables aléatoires aient la même moyenne ou la même variance, elles doivent par contre être bornées.
2. Si la variance σ^2 n'est pas définie, la démonstration est plus délicate (voir par exemple le livre de Feller, chapitre X).

On suppose maintenant que $\langle X \rangle = m$ est différent de zéro. On change X en $X-m$ et l'inégalité précédente devient

$$Pr(|X-m| > a) < \frac{\sigma^2}{a^2}.$$

On comprend mieux comment σ caractérise la dispersion de la mesure. Posons maintenant $a = k\sigma$, il vient

$$Pr(|X-m| > k\sigma) < \frac{1}{k^2}.$$

C'est l'inégalité de Bienaymé-Chebichef. [3] Elle ne nous apprend rien pour $k \leq 1$. On remarquera que cette inégalité a été démontrée en toute généralité, sans connaître la forme de la distribution $P(x)$ (pourvu qu'elle admette un deuxième moment). C'est la majoration la plus grossière possible.

Si on applique maintenant l'inégalité de Bienaymé-Chebichef, on conclut que

$$Pr\left(|M_N - m| > a\right) < \frac{\sigma^2}{a^2 N},$$

ce qui achève la démonstration.

L'évaluation empirique d'une probabilité est un cas particulier de la loi des grands nombres. En effet, considérons une expérience dont l'issue vaut aléatoirement $X = 1$ avec la probabilité p ou $X = 0$ avec la probabilité $1-p$. La moyenne arithmétique réalisée sur une série de N observations, $\frac{1}{N}\sum_{i=1}^{N} X_i$, est une estimation de la probabilité p qui est d'autant plus précise que N est grand.

2 Le théorème central limite

Ce théorème, démontré [4] par Lindeberg en 1922, va au delà de la connaissance de la valeur moyenne et de la variance de la variable aléatoire M_N. Il en donne une loi de probabilité approchée lorsque $N \to \infty$.

Soit $X_1,\ X_2 \ldots\ X_N$ des variables aléatoires mutuellement *indépendantes* ayant la même distribution de probabilité, [a] c'est-à-dire la même *moyenne finie* m et la même *variance finie* σ^2. Leur somme divisée par N, qui s'écrit

$$M_N = \frac{1}{N}\sum_{i=1}^{N} X_i,$$

est une variable aléatoire distribuée selon une loi de probabilité [b] qui tend vers la loi normale de moyenne m et de variance σ^2/N lorsque $N \to \infty$:

$$P(m_N) \simeq \frac{1}{\sqrt{2\pi}(\frac{\sigma}{\sqrt{N}})} e^{-\frac{(m_N-m)^2}{2(\frac{\sigma}{\sqrt{N}})^2}}, \tag{IV.1}$$

a. Plus généralement, il n'est pas indispensable que ces variables aient la même distribution. Il suffit qu'elles aient la même valeur moyenne et la même variance et qu'elles soient finies.

b. C'est en fait une densité de probabilité : en passant à la limite $N \to \infty$, M_N devient une variable continue variant de $-\infty$ à $+\infty$.

Insistons : Contrairement à la loi des grands nombres, le théorème central limite ne s'applique qu'à des variables aléatoires dont la variance est finie.

Nous avons déjà rencontré des applications du théorème central limite lorsque nous avons étudié le comportement asymptotique de la distribution binomiale et de la loi de Poisson au chapitre III.

3. Irénée-Jules Bienaymé (1796-1878) mathématicien français et Pafnuty Lvovich Chebichev (1821-1894) mathématicien russe.
4. Laplace en donne une première version en 1812.

On peut vérifier que †

$$\lim_{N\to\infty} P(m_N) = \delta(m_N - m).$$

Autrement dit, quand $N \to \infty$, $\sigma/\sqrt{N} \to 0$ et donc la fonction $P(m_N)$ centrée en m devient de plus en plus étroite : son maximum tend vers l'infini, mais son intégrale reste égale à 1. Elle n'a pas de limite comme fonction, mais tend vers la distribution de Dirac, δ, placée en m. La figure IV.1 illustre le théorème central limite à l'aide de la somme de N variables aléatoires uniformes.

.

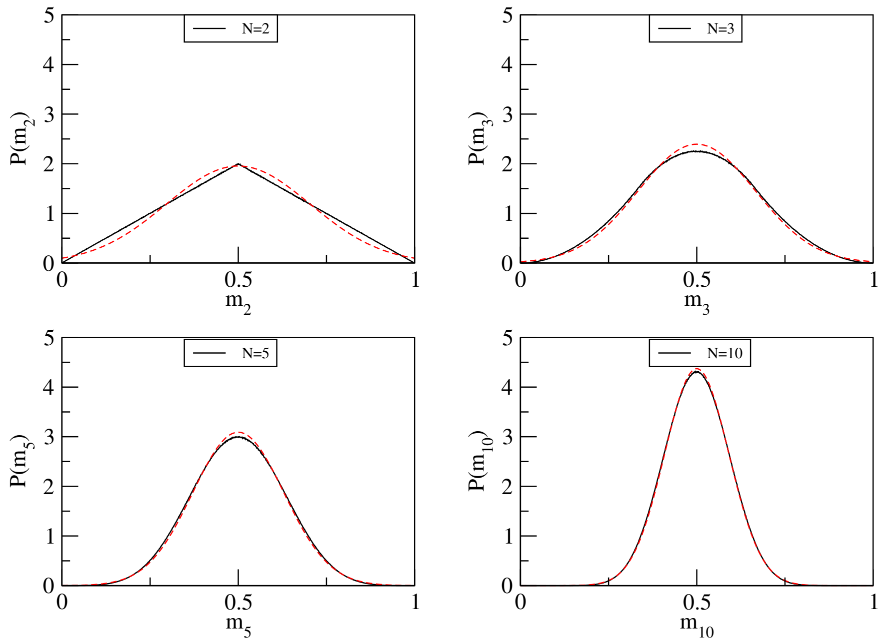

FIGURE IV.1 – *Distribution $P(m_N)$ de la somme de $N = 2$ (en haut à gauche), $N = 3$ (en haut à droite), $N = 5$ (en bas à gauche) et $N = 10$ (en bas à droite) variables aléatoires uniformes (distribuées sur $P(x_i) = 1$ pour $0 < x < 1$, $Var(X_i) = 1/12$).La courbe en trait rouge interrompu est une gaussienne de même variance que la distribution $P(m_N)$: $Var(M_N) = Var(X_i)/N$.*

Démontrons maintenant le théorème central limite en utilisant la fonction caractéristique. Puisque M_N est la somme de N variables indépendantes X_i/N, sa fonction caractéristique s'écrit :

$$\phi_{M_N}(u) = \phi^N_{X_i/N}(u), \tag{IV.2}$$

où $\phi_{X_i/N}(u)$ est la fonction caractéristique de la variable X_i/N. De la distribution de X_i/N, on ne connaît que ses deux premiers moments : $\langle X_i\rangle = m$ et $\langle X_i^2\rangle = \sigma^2 + m^2$. En développant $\phi_{X_i/N}(u)$ en série de u

†. En effet, soit une fonction $f(x)$, il faut montrer que

$$\lim_{N\to\infty} \int f(x)P(x)dx = f(m).$$

On fait le changement de variable $u = (x - m)/(\sqrt{2}\frac{\sigma}{\sqrt{N}})$ et il vient

$$\lim_{N\to\infty} \int f(x)P(x)dx = \frac{1}{\sqrt{\pi}} \int f\left(\frac{\sqrt{2}\sigma}{\sqrt{N}}u + m\right) e^{-u^2} du.$$

Le théorème de la convergence dominée permet de passer à la limite $N \to \infty$, il reste

$$\lim_{N\to\infty} \int f(x)P(x)dx = \frac{1}{\sqrt{\pi}} f(m) \int e^{-u^2} du = f(m).$$

on obtient :

$$\begin{aligned}\phi_{X_i/N}(u) &= \langle e^{iu\frac{X_i}{N}}\rangle = 1 + i\frac{u}{N}\langle X_i\rangle - \frac{u^2}{2N^2}\langle X_i^2\rangle + \mathcal{O}(\frac{u^3}{N^3}) \\ &= 1 + im\frac{u}{N} - \frac{u^2}{2N^2}(\sigma^2 + m^2) + \mathcal{O}(\frac{u^3}{N^3}).\end{aligned}$$

D'après l'équation (IV.2), on obtient :

$$\begin{aligned}\ln\phi_{M_N}(u) &= N\ln\left[1 + im\frac{u}{N} - \frac{u^2}{2N^2}(\sigma^2+m^2) + \mathcal{O}(\frac{u^3}{N^3})\right] \\ &= N\left[im\frac{u}{N} - \frac{u^2}{2N^2}(\sigma^2+m^2) - \frac{1}{2}\big(-(\frac{mu}{N})^2\big) + \mathcal{O}(\frac{u^3}{N^3})\right] \\ &= imu - \frac{u^2\sigma^2}{2N} + \mathcal{O}(\frac{u^3}{N^2}) \\ &= \ln\phi_n(u) + \mathcal{O}(\frac{u^3}{N^2}),\end{aligned}$$

où $\phi_n(u)$ est la fonction caractéristique de la loi normale de moyenne m et de variance σ^2/N. A la limite $N \to \infty$, la fonction caractéristique de M_N tend vers la fonction caractéristique d'une loi normale, ce qui démontre le théorème central limite.

En résumé, le théorème central limite peut s'énoncer de la manière suivante :

La moyenne arithmétique d'un échantillon d'une population quelconque de variables aléatoires possédant une distribution de probabilité quelconque mais de moyenne et de variance finies, est distribuée suivant une loi normale quand la taille de l'échantillon est suffisamment grande.

Commentaires et conséquences

- Le fait que les distributions de variables de natures très différentes (répétition de la mesure d'une même grandeur physique, taille des individus, variation de la valeur d'une action sur une courte durée, volume de lait produit par une vache...) soient souvent (mais pas toujours !) correctement reproduites par une gaussienne est une donnée empirique. Elle indique seulement que ces variables peuvent résulter d'une somme de variables aléatoires *indépendantes* de même loi. Mais le mécanisme (complexe) produisant ces variables aléatoires est généralement inconnu.
- Le théorème central limite ne fait aucune hypothèse sur la forme de la distribution de probabilité des variables X_i, il a donc une portée très générale. En revanche, on ne peut pas l'appliquer lorsque ses deux premiers moments ne sont pas définis. Ainsi, la distribution de la moyenne arithmétique de variables distribuées sur une loi large ne tend pas vers la loi normale.
- Lois stables : Lorsque la distribution de la moyenne arithmétique M_N a la même dépendance fonctionnelle que celle des variables aléatoires X_i pour N *quelconque* ($N = 2, 3, ...$), on dit que la distribution est *stable*. Notons que la stabilité d'une distribution n'a rien à voir avec le théorème central limite qui concerne le passage à la limite $N \to \infty$. Cette propriété, qui se démontre facilement à l'aide de la fonction caractéristique, est partagée par les lois normale, de Cauchy, de Poisson et binomiale (pour la même valeur de p) ... En revanche, comme le montre la figure IV.1, la loi uniforme n'est pas stable.

Chapitre V

Annexe Mathématique

1 Fonction $\Gamma(x)$

Soit l'intégrale gaussienne

$$I(n) = \int_0^\infty x^n e^{-\alpha x^2} dx,$$

où $\alpha > 0$.

Calculons $J = \int_{-\infty}^{\infty} e^{-\alpha x^2} dx = 2I(0)$, car $I(0)$ est paire. Pour commencer exprimons J^2 sous la forme d'une intégrale double en coordonnées cartésiennes :

$$J^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\alpha(x^2+y^2)} dxdy.$$

En passant en coordonnées polaires : $x^2 + y^2 = r^2$ et $dxdy = rdrd\theta$. Donc

$$\begin{aligned} J^2 &= \int_0^\infty \int_0^{2\pi} e^{-\alpha r^2} rdrd\theta \\ &= 2\pi \Big[-\frac{e^{-\alpha r^2}}{2\alpha} \Big]_0^\infty \\ &= \frac{\pi}{\alpha}. \end{aligned}$$

Donc :

$$I(0) = \int_0^\infty e^{-\alpha x^2} dx = \frac{1}{2}\sqrt{\frac{\pi}{\alpha}}.$$

De même, on peut calculer directement $I(1)$:

$$\begin{aligned} I(1) &= \int_0^\infty x\, e^{-\alpha x^2} dx \\ &= \Big[-\frac{e^{-\alpha x^2}}{2\alpha} \Big]_0^\infty \\ &= \frac{1}{2\alpha}. \end{aligned}$$

Les intégrales $I(n)$ pour $n > 1$ s'expriment en fonction de $I(0)$ et $I(1)$ de la façon suivante :

$$I(n) = -\frac{\partial}{\partial\alpha}\left(\int_0^\infty x^{n-2} e^{-\alpha x^2} dx \right) = -\frac{\partial I(n-2)}{\partial \alpha}$$

Par exemple, $I(2) = -\frac{\partial I(0)}{\partial\alpha}$ et ainsi de suite. On trouve $I(0) = \frac{1}{2}\sqrt{\frac{\pi}{\alpha}}$, $I(1) = \frac{1}{2\alpha}$, $I(2) = \frac{\sqrt{\pi}}{4\alpha^{\frac{3}{2}}}$, $I(3) = \frac{1}{2\alpha^2}$ et $I(4) = \frac{3\sqrt{\pi}}{8\alpha^{\frac{5}{2}}}$

On définit la fonction Gamma pour $x > 0$:

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt = 2 \int_0^\infty e^{-s^2} s^{2x-1} ds.$$

Clairement

$$\Gamma(1) = 1 \quad \text{et} \quad \Gamma(1/2) = \sqrt{\pi}.$$

En intégrant par parties on montre que

$$\Gamma(x+1) = x\Gamma(x).$$

Donc pour tout n entier positif :

$$n! = \Gamma(n+1). \tag{V.1}$$

2 Formule de Stirling

C'est une approximation très utile de $n!$ d'autant meilleure que n est grand :

$$n! \simeq n^n e^{-n} \sqrt{2\pi n} \ . \tag{V.2}$$

Dans un premier temps, voyons une approximation grossière de $n!$. Remarquons que

$$\ln n! = \ln n(n-1)...1 = \sum_{p=1}^{n} \ln p.$$

Lorsque n est grand par rapport à 1, cette somme peut être approximée par une intégrale ($\ln p$ varie peu lorsque p augmente d'une unité pour $p >> 1$). Ainsi,

$$\ln n! \simeq \int_1^n \ln(x) dx = \Big[x \ln x - x \Big]_1^n \simeq n \ln n - n.$$

Démontrons à présent la formule de Stirling par la méthode du col. D'après l'équation V.1 :

$$n! = \Gamma(n+1) = \int_0^\infty e^{-t} t^n dt = \int_0^\infty e^{-f(t)} dt.$$

où $f(t) = t - n \ln t$. La fonction e^{-t} décroît rapidement avec t, alors que t^n croit rapidement (pour n grand). Le produit $e^{-t}t^n$ présente donc un maximum très marqué en $t_0 = n$ et vaut pratiquement zéro en dehors de cette région. En prenant le logarithme, on obtient la fonction $f(t)$ qui elle présente un minimum en t_0. Puisque l'intégrand ne prend des valeurs notables qu'autour du minimum t_0 ($f'(t_0) = 0$), on va remplacer la fonction $f(t)$ par son développement de Taylor au deuxième ordre :

$$f(t) = f(t_0) + \frac{1}{2}(t - t_0)^2 f''(t_0) + ... \simeq -n \ln n + n + \frac{1}{2n}(t - t_0)^2,$$

puisque $f(t_0) = n - n \ln n$ et $f''(t_0) = 1/n$. On en déduit :

$$n! \simeq n^n e^{-n} \int_0^\infty e^{-\frac{(t-t_0)^2}{2n}} dt.$$

En effectuant le changement de variable $u = t - n$:

$$n! \simeq n^n e^{-n} \int_{-n}^\infty e^{-\frac{u^2}{2n}} du,$$

et puisque $n >> 1$, on reconnaît une intégrale gaussienne et finalement :

$$n! \simeq n^n e^{-n} \sqrt{2\pi n}.$$