Amine Asselah– Année 2009–2010.
Université de Paris XII

12 SEMAINES DE PROBABILITÉS. SECTION L2-MATH

Amine Asselah– Année 2009–2010., Université de Paris XII

12 SEMAINES DE PROBABILITÉS. SECTION L2-MATH

Amine Asselah– Année 2009–2010.
Université de Paris XII

TABLE DES MATIÈRES

CHAPITRE 1

ESPACE DES ISSUES, PROBABILITÉS, VARIABLES ALÉATOIRES

La théorie des probabilités a pour objectif de modéliser des jeux ou des expériences où plusieurs issues sont possibles mais où leur réalisation n'est pas déterminée à l'avance : par exemple un lancer de dés. La théorie des probabilités ne va pas permettre de prédire quelle issue va se réaliser, mais quelle chance a chaque issue de se réaliser.

Ainsi, dans un premier temps, on va associer à chaque issue possible un nombre entre 0 et 1 qui traduit notre estimation des chances que cette issue a de se réaliser : on appelle ce nombre *la probabilité* de cette issue. On appelle "évènement" un ensemble d'issues. La probabilité qu'on associe à un évènement est la somme des probabilités de chaque issue qui constitue cet évènement. Typiquement, la question est de déterminer la probabilité d'un évènement, et la difficulté est d'une part de décrire l'expérience de façon commode afin d'énumérer toutes les issues possibles et leur probabilité respective, et d'autre part de décomposer l'évènement considéré en issues.

1.1. Modélisation

1.1.1. Espace des issues. — Avant de calculer les probabilités d'évènements, il faut définir l'espace des issues de façon commode et complète. Cet espace comprendra toutes les issues possibles du jeu, ou de l'expérience aléatoire que l'on considère, même celles qui ne nous intéressent pas, a priori. Dans chaque situation, l'espace des issues sera noté Ω (grand omega), alors que les issues seront notées ω (petit omega).

1.1.2. Espace des évènements. — Un évènement est une collection d'issues. On note $\mathcal{F}$ l'ensemble des évènements. $\mathcal{F}$ est l'ensemble des parties de Ω. Si $\Omega = \{1, 2, 3\}$, alors

$$\mathcal{F} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{1, 3\}, \{1, 2, 3\}\},$$

où $\emptyset$ est l'ensemble vide. On note $A \cup B = \{\omega \in A \text{ ou } \omega \in B\}$, et $A \cap B = \{\omega \in A \text{ et } \omega \in B\}$. On interprète $A \cap B$ comme A **et** B se réalisent, et $A \cup B$ comme A **ou** B se réalisent. On note $A \backslash B = \{\omega \in A \text{ et } \omega \notin B\}$ et $A^c = \Omega \backslash A$. On appelle A^c le complémentaire de A, il correspond à l'évènement 'A ne se réalise pas'.

Deux évènements A et B de $\mathcal{F}$ sont **disjoints** s'ils n'ont aucune issue en commun, c'est à dire que $A \cap B = \emptyset$. Par exemple, A et A^c sont disjoints, ainsi que $\emptyset$ et A.

1.1.3. Probabilités. —

Définition 1. — . Une probabilité P est une fonction de $\mathcal{F}$ dans $[0,1]$ qui doit vérifier les propriétés suivantes :

- $P(\Omega) = 1$ et $P(\emptyset) = 0$.
- Si A et B sont disjoints

$$P(A \cup B) = P(A) + P(B). \tag{1.1.1}$$

On remarque que $P(A) + P(A^c) = 1$, ce qui découle de $P(\Omega) = 1$ et (1.1.1). Lorsque les ensembles A et B ne sont pas disjoints, alors on peut les découper en deux ensembles disjoints : A et $B \backslash (A \cap B)$, tels que

$$A \cup B = A \cup (B \backslash (A \cap B)) .$$

Noter aussi que

$$B = (B \backslash (A \cap B)) \cup (A \cap B).$$

On applique (1.1.1) deux fois :

$$P(B) = P(B \backslash (A \cap B)) + P(A \cap B)$$

et,

$$P(A \cup B) = P\left(A \cup (B \backslash (A \cap B))\right) = P(A) + P(B \backslash (A \cap B)).$$

On conclut que pour deux évènements quelconques A, B

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

1.1.4. Exemples. —

Exemple 1. — . On considère un dé à 6 faces, numérotées de 1 à 6. On suppose que le dé est équilibré, ce qui veut dire que les 6 faces ont la même chance de sortir. On appelle Ω l'ensemble des issues possibles d'un lancer, $\Omega = \{1,2,3,4,5,6\}$, et on associe à chaque issue la probabilité $1/6$. On pense donc à une probabilité comme à une fonction P, sur Ω, telle que

$$P(1) + \cdots + P(6) = 1. \tag{1.1.2}$$

On peut maintenant répondre à la question : quelle est la probabilité qu'un nombre pair sorte ? Il suffit de décomposer cet évènement, disons A, en les différentes issues qu'il contient : $A = \{2,4,6\}$

$$P(A) = P(2) + P(4) + P(6) = 1/2. \tag{1.1.3}$$

Exemple 2. — . On considère le même dé, sauf que sur les faces 2,3,4,5 et 6 a été inscrit le nombre 2. Ainsi, deux seuls nombres peuvent sortir : 1 et 2. L'ensemble des issues est différent $\tilde{\Omega} = \{1,2\}$. Comme le dé est équilibré, les faces 2,3,4,5 et 6 ont chacune la probabilité $1/6$ de sortir, alors que $P(2) = 5/6$, et $P(1) = 1/6$. Voici donc un exemple où deux issues se réalisent avec des probabilités différentes.

Exemple 3. — . Une urne (c'est-à-dire un sac) contient 1 boule blanche et 5 boules noires. On tire une boule au hasard. Quel est le *bon* espace des issues :

$$\Omega_A = \{b, n, n, n, n, n\} \quad \text{ou} \quad \Omega_B = \{b, n\}?$$

(On a utilsé b pour 'boule blanche', et n pour 'boule noire'.

Réponse : Ω_A représente le contenu de l'urne, alors que Ω_B représente le résultat d'un tirage, et c'est ce qui nous intéresse. La probabilité de tirer une boule blanche est 1/6, et la probabilité de tirer une boule noire est 5/6.

Exemple 4. — . On considère le lancer de deux dés. Une personne I voit le résultat des deux dés et communique à une personne II la somme des deux dés : ainsi la personne II n'aura accès qu'à un nombre entre 2 et 12. Quel est l'espace des issues ? Il y a une certaine ambiguité puisque I et II ne voient pas la même chose. La personne I voit les résultats des lancers 1 et 2, et il est plus commode de considérer

$$\Omega_I = \{(1,1), (1,2), (2,1), \ldots, (6,6)\} = \{(i,j) : \ i,j \in \{1,\ldots,6\}\}.$$

Cet espace est commode car chaque couple à la même chance de sortir :

$$\text{Pour tous } i,j \text{ dans } \{1,\ldots,6\}, \quad P_I(i,j) = \frac{1}{|\Omega|} = 1/36.$$

Pour la personne II, les issues possibles sont

$$\Omega_{II} = \{2, 3, 4, \ldots, 11, 12\}.$$

Par exemple si la somme des deux dés est 4, cela englobe les couples $\{(1,3), (2,2), (3,1)\}$ de Ω_I. Ainsi,

$$P_{II}(4) = P_I(1,3) + P_I(2,2) + P_I(3,1) = \frac{3}{36}.$$

Il faut faire la même chose pour trouver la probabilité associée à chaque issue : faites-le, nous aurons besoin dans la suite de P_{II}.

On pourra aussi décider que l'ensemble des issues est toujours Ω_I, et que la personne II a seulement accès à des sous- ensembles de Ω_I : il est important de noter que le choix de l'espace des issues n'est pas unique ! Ce qui doit vous guider est la facilité d'associer une probabilité à une issue.

Exemple 5. — . On dispose de 15 sacs contenant chacun un alphabet de 26 lettres $\mathcal{A} = \{\mathrm{a}, \mathrm{b}, \ldots, \mathrm{z}\}$. On forme un mot de 15 lettres en piochant successivement une lettre dans chaque sac. Pour chaque lettre $\mathrm{x} \in \mathcal{A}$, on note B_{x}, l'évènement selon lequel la lettre x n'apparait pas dans le mot.

- Donner l'espace des issues, et les probabilités associées.
- Que vaut $P(B_{\mathrm{x}})$?

Solution : une issue dans ce jeu est un choix de 15 lettres : $\omega = (a_1, \ldots, a_{15})$, où chaque a_i appartient à l'alphabet. Ainsi, on notera cela

$$\Omega = \{\omega = (a_1, \ldots, a_{15}) : \ a_k \text{ appartient à l'alphabet, pour } k = 1, \ldots, 15\}.$$

Chaque lettre est tirée avec la probabilité 1/26, et donc pour chaque issue ω dans Ω,

$$P(\omega) = \underbrace{\frac{1}{26} \cdots \frac{1}{26}}_{15 \text{ fois}} = (\frac{1}{26})^{15}.$$

Comment écrire B_{x} en fonction des issues de Ω ? Il est facile de décrire B_{x} en français : c'est l'ensemble des issues où les lettres sont toutes différentes de x. Ainsi,

$B_{\mathrm{x}} = \{\omega = (a_1, \ldots, a_{15}) :\ a_k$ appartient à l'alphabet moins la lettre x, pour $k = 1, \ldots, 15\}$.

Ainsi, à chaque tirage, on a 25 choix possibles sur 26 :

$$P(B_{\mathrm{x}}) = \underbrace{\frac{25}{26} \cdots \frac{25}{26}}_{15 \text{ fois}} = (\frac{25}{26})^{15}.$$

Exemple 6. — . L'urne A contient 1 boule blanche et 5 boules noires. L'urne B contient 5 boules blanches et 1 boule noire. On s'intéresse à modéliser le jeu suivant : on jette d'abord un pièce de monnaie bien equilibrée.

- Si elle tombe sur pile, on tire une boule de l'urne A.
- Si elle tombe sur face, on tire une boule de l'urne B.

Décrire l'espace des issues de ce jeu et donner les probabilités associées à chaque issue.

Solution : Le jeu consiste à lancer une pièce, puis à tirer une boule : la pièce peut-être pile ou face, et la boule peut-être blanche ou noire. Ainsi le plus simple est de décrire le résultat du lancer et du tirage par un couple (l, t) où l est dans {pile,face}, et t est dans {blanche,noire}. Ainsi

$$\Omega = \{(\text{pile,blanche}), (\text{pile,noire}), (\text{face,blanche})(\text{face,noire})\}.$$

On obtient pile 1 fois sur 2, et lorsque c'est le cas, on a 1 fois sur 6 une boule blanche, car on tire de l'urne A. Donc

$$P((\text{pile,blanche})) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12}.$$

Faites de même pour les autres issues. Vérifier que la somme des probabilités sur Ω donne bien 1.

Exemple 7. — . On a 3 cartes coloriées sur chaque face : rouge-rouge (RR), blanc-blanc (BB) et rouge-blanc (RB). Une personne tire une carte au hasard et vous en montre une face, elle aussi tirée au hasard. Si la face montrée est R, que pariez-vous pour la couleur de la face cachée ?

Intuitivement, il y a trois cas possibles qui se produisent avec la même probabilité : deux cas où la face montrée est R lorsque la carte est RR, un cas où la face montrée est R lorsque la carte est RB. Donc, deux cas où la face cachée sera R, et un cas où elle sera B. On a donc deux fois plus de chance de gagner en pariant R.

Plus rigoureusement, l'expérience qu'on veut modéliser consiste en 2 étapes : on choisit une carte parmi {RR,RB,BB}, puis on choisit une des deux faces {recto, verso}. On va supposer que dans la notation XY des cartes, X correspond au recto et Y au verso. On traduit que ces choix se font au hasard en disant que toutes les issues possibles ont la même probabilité. On va décrire séparement les deux expériences : le choix d'une carte est une issue de

$$\Omega_1 = \{RR, RB, BB\}$$

et on appelle P_1 la probabilité sur Ω_1, avec

$$P_1(RR) = P_1(RB) = P_1(BB) = 1/3.$$

Puis, le choix d'une face correspond à une issue de $\Omega_2 = \{\text{recto, verso}\}$ et $P_2(\text{recto}) = P_2(\text{verso}) = 1/2$. Pour les deux expériences, l'ensemble des issues est ici l'ensemble des couples (carte, face). Donc, si l'on utilise r pour recto et v pour verso

$$\Omega = \{(RR, r), (RR, v), (RB, r), (RB, v), (BB, r), (BB, v)\} = \Omega_1 \times \Omega_2. \tag{1.1.4}$$

La probabilité associée aux issues de Ω est P

$$P(\text{carte, face}) = P_1(\text{carte})P_2(\text{face}) = 1/6.$$

Le fait que P se factorise traduit *l'indépendance* des deux expériences.

On décompose l'évènement A "la face montrée est R et la face cachée est R" en issues de Ω : $A = \{(RR, r), (RR, v)\}$. De même, D est "la face montrée est R et la face cachée est B" $= \{(RB, r)\}$. On voit donc que $P(A) = 2/6$ alors que $P(D) = 1/6$.

1.2. Variable Aléatoire

1.2.1. Partitions. —

Définition 2. — . On appelle **partition** de Ω, une suite d'ensembles $A_1, \ldots, A_n$ de Ω, disjoints deux à deux, et tels que leur reunion soit Ω. En d'autres termes,

$$\text{Pour tout } i \text{ différent de } j,\ A_i \cap A_j = \emptyset \quad \text{et} \quad A_1 \cup A_2 \cup \cdots \cup A_n = \Omega.$$

1.2.2. Fonction Indicatrice. — La *fonction indicatrice*, 1_A, pour un ensemble $A \subset \Omega$, est une fonction de Ω dans $\{0, 1\}$ avec

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

Nous énonçons sans démonstration 4 propriétés simples et utiles des fonctions indicatrices :

1. $1_{A\cap B} = 1_A.1_B$.
2. Si $A \cap B = \emptyset$, alors $1_{A\cup B} = 1_A + 1_B$.
3. $1_{A^c} = 1 - 1_A$.
4. Si $\{A_i,\ i = 1, \ldots, n\}$ forment une partition, alors $1 = 1_{A_1} + \cdots + 1_{A_n}$.

Il est important de vérifier seul ces 4 points.

Exemple 8. — . Nous continuons l'exemple 5. On pose $A_{\mathrm{x}} = B_{\mathrm{x}}^c$.
– Dire en français ce que représente A_{x} ?

– Que représente
$$N(\omega) = \sum_{\mathrm{x}\in\mathcal{A}} 1_{A_{\mathrm{x}}}(\omega)?$$
Solution : A_{x} est l'ensemble des mots où la lettre x apparait au moins une fois. $N(\omega)$ est le nombre de lettres distinctes qui apparaissent dans le mot ω.

1.2.3. Variable aléatoire. —

Définition 3. — . Une fonction de Ω dans les nombres réels est appelée variable aléatoire.

Ainsi, une variable aléatoire n'est rien d'autre qu'une fonction tout à fait classique. Pourquoi avoir donc un nom de plus ? En analyse, il est fondamental de savoir quel est le domaine de la fonction : comme $\sin(x)$, ou $\exp(x)$: on doit savoir où se trouve x : sur $[0,\pi[$, ou sur $[0,\infty[$, etc... En probabilité, ce qui est important est la valeur de la fonction, le domaine par contre est toujours Ω, l'espace des issues qui est un peu artificiel, puisqu'il dépend de notre façon de décrire le jeu ou l'expérience que l'on modélise ; ce qui est par contre très important est la probabilité que l'on définit sur Ω.

On note en général les variables aléatoires par $X, Y, \dots$. Intuitivement, on pense à la valeur d'une variable aléatoire $X(\omega)$ comme à l'argent qu'on gagnerait si l'issue ω se réalisait.

Définition 4. — . Soit $\{x_1,\dots,x_n\}$ l'ensemble des valeurs distinctes d'une variable X. On associe à X une partition de Ω :
$$A_1 = \{\omega : X(\omega) = x_1\}, \quad \dots, \quad A_n = \{\omega : X(\omega) = x_n\}.$$
En effet, on vérifie que si $i \neq j$ alors $A_i \cap A_j = \emptyset$, et pour tout $\omega \in \Omega$, il y a un indice i tel que $\omega \in A_i$. On note $\mathcal{A}(X)$ la partition $\{A_1,\dots,A_n\}$, que l'on appelle aussi la partition engendrée par X.

Ainsi, les éléments de la partition engendrée par X correspondent aux régions où X est constante. Inversement, à toute partition de Ω, on peut associer une variable aléatoire qui sera une fonction constante sur les éléments de cette partition. Il est très utile de conserver cette représentation en tête.
$$X(\omega) = x_1 1_{A_1}(\omega) + \cdots + x_n 1_{A_n}(\omega), \quad \text{avec} \quad A_i = \{\omega : X(\omega) = x_i\}.$$

Définition 5. — . Soient $\mathcal{A} = \{A_1,\dots,A_n\}$ et $\mathcal{B} = \{B_1,\dots,B_m\}$ deux partitions de Ω. On dira que la partition $\mathcal{A}$ est plus fine que la partition $\mathcal{B}$ si pour tout $i = 1,\dots,n$ il existe un $j = 1,\dots,m$ tel que $A_i \subset B_j$. Ainsi, lorsque $\mathcal{A}$ est plus fine que $\mathcal{B}$, si $I_j = \{k : A_k \subset B_j\}$, alors
$$B_j = \bigcup_{k\in I_j} A_k.$$

Proposition 1. — . *Soient $X, Y : \Omega \to \mathbb{R}$. Les deux propriétés suivantes sont équivalentes :*

1. *$\mathcal{A}(X)$ est plus fine que $\mathcal{A}(Y)$.*
2. *Si $\omega, \omega' \in \Omega$, et $X(\omega) = X(\omega')$ alors $Y(\omega) = Y(\omega')$.*

Démonstration. — (1) implique (2). Si $X(\omega) = X(\omega')$, alors ω et ω' appartiennent au même élément $A \in \mathcal{A}(X)$. Or, il existe $B \in \mathcal{A}(Y)$ tel que $A \subset B$, et donc ω et ω' appartiennent à B, ce qui signifie que $Y(\omega) = Y(\omega')$.

(2) implique (1). A faire!

∎

Soit f une fonction de $\mathbb{R}$ dans $\mathbb{R}$. Alors, $f(X) : \Omega \to \mathbb{R}$ pourra se décomposer sur $\mathcal{A}(X)$ de la façon suivante

$$f(X(\omega)) = f(x_1)1_{A_1}(\omega) + \cdots + f(x_n)1_{A_n}(\omega). \tag{1.2.1}$$

Ainsi, pour tout $i = 1, \ldots, n$, la variable $f(X)$ vaut $f(x_i)$ sur A_i, ce qui veut dire que $B_i := \{\omega : f(X(\omega)) = f(x_i)\}$ contient A_i, et donc $\mathcal{A}(X)$ est plus fine que $\mathcal{A}(f(X))$. Notons qu'il se peut que les valeurs $\{f(x_i), i = 1, \ldots, n\}$ ne soient pas toutes distinctes, auquel cas $\mathcal{A}(f(X))$ est différente de $\mathcal{A}(X)$.

On pourra montrer que si f est injective alors $\mathcal{A}(X) = \mathcal{A}(f(X))$.

Remarque 1. — . Noter qu'à partir de deux partitions $\mathcal{A}, \mathcal{B}$ de Ω, on peut construire une partition $\mathcal{C}$ plus fine que $\mathcal{A}$ et que $\mathcal{B}$.

$$\mathcal{C} := \{C \subset \Omega : \exists A \in \mathcal{A}, \exists B \in \mathcal{B},\ C = A \cap B\}.$$

On notera $\mathcal{C} = \mathcal{A} \bullet \mathcal{B}$.

Remarque 2. — . Soit $\mathcal{A} = \mathcal{A}(X)$ et $\mathcal{B} = \mathcal{A}(Y)$. Alors, $\mathcal{A} \bullet \mathcal{B}$ est plus fine que $\mathcal{A}(X+Y)$. Il faut montrer pour cela que $X+Y$ est constante sur les parts de $\mathcal{A} \bullet \mathcal{B}$. Soient $\omega, \omega' \in A \in \mathcal{A} \bullet \mathcal{B}$. Alors, $X(\omega) = X(\omega')$ et $Y(\omega) = Y(\omega')$. Ainsi $X(\omega) + Y(\omega) = X(\omega') + Y(\omega')$.

Définition 6. — . Soit $\{x_1, \ldots, x_n\}$ l'ensemble des valeurs distinctes d'une variable X. On appelle loi de X la collection des couples $(x_i, P(\{\omega :\ X(\omega) = x_i\})$ pour $i = 1, \ldots, n$.

CHAPITRE 2

INDÉPENDANCE, PROBABILITÉS CONDITIONNELLES

2.1. Indépendance

La notion intuitive d'indépendance s'illustre avec l'exemple suivant : on lance deux dés, et l'on demande la probabilité que le premier donne 3, et le second 5 ? La réponse immédiate est qu'on a 1 chance sur 6 que le premier dé donne 3, et 1 chance sur 6 que le second donne 5 : donc au total 1/6 fois 1/6, soit 1/36 d'obtenir le couple $(3,5)$.

Soyons plus abstrait : soit un ensemble d'issues Ω et soit P une probabilité sur Ω. On supposera que Ω a un nombre fini d'issues et que chaque issue a une probabilité strictement positive. On pose la définition suivante.

Définition 7. — . Deux évènements A et B sont indépendants si

$$P(A \cap B) = P(A) \times P(B) \tag{2.1.1}$$

Définition 8. — . Deux partitions de Ω, $\mathcal{A} := \{A_1, \ldots, A_n\}$ et $\mathcal{B} := \{B_1, \ldots, B_m\}$ sont indépendantes si pour tous $i \in \{1, \ldots, n\}$ et tous $j \in \{1, \ldots, m\}$ on a

$$P(A_i \cap B_j) = P(A_i) \times P(B_j)$$

Proposition 2. — . *Soient $\mathcal{A}, \mathcal{A}'$ et $\mathcal{B}, \mathcal{B}'$ quatre partitions de Ω et $\mathcal{A}$ et $\mathcal{B}$ sont indépendantes. Si $\mathcal{A}$ est plus fine que $\mathcal{A}'$ et $\mathcal{B}$ plus fine que $\mathcal{B}'$, alors $\mathcal{A}'$ et $\mathcal{B}'$ sont indépendantes.*

Démonstration. — Nous gardons la notation $\mathcal{A} := \{A_1, \ldots, A_n\}$ et $\mathcal{B} := \{B_1, \ldots, B_m\}$. Rappelons d'abord que si $C_{i,j} = A_i \cap B_j$ alors $\mathcal{C} = \{C_{i,j}, i = 1, \ldots, n,\ j = 1, \ldots, m\}$ est une partition de Ω plus fine que $\mathcal{A}$ et que $\mathcal{B}$ (nous avons noté $\mathcal{C} = \mathcal{A} \bullet \mathcal{B}$ dans le cours précédent). Maintenant, pour tout $A \in \mathcal{A}'$ et $B \in \mathcal{B}'$ notons $I = \{j :\ A_j \subset A\}$, et $J = \{j :\ B_j \subset B\}$. Alors

$$A = \cup_{j \in I} A_j, \quad \text{et} \quad B = \cup_{j \in J} B_j.$$

Noter que $A \cap B = \cup_{I \times J} A_i \cap B_j$, et que les $C_{i,j}$ sont disjoints, on a

$$P(A \cap B) \quad = \sum_{i \in I, j \in J} P(A_i \cap B_j) = \sum_{i \in I, j \in J} P(A_i) \times P(B_j)$$

$$(2.1.2) \qquad = \left(\sum_{i\in I} P(A_i)\right)\left(\sum_{j\in J} P(B_j)\right) = P(A)\times P(B).$$

∎

Définition 9. — . Deux variables aléatoires $X, Y : \Omega \mapsto \mathbb{R}$ sont indépendantes si les partitions qu'elles engendrent le sont.

Proposition 3. — . *Soient deux variables aléatoires indépendantes $X, Y : \Omega \mapsto \mathbb{R}$. Si f et $g : \mathbb{R} \to \mathbb{R}$ sont deux fonctions quelconques, alors $f(X)$ et $g(Y)$ sont indépendantes.*

Démonstration. — Nous avons vu que $\mathcal{A}(X)$ est plus fine que $\mathcal{A}(f(X))$ et pareillement $\mathcal{A}(Y)$ est plus fine que $\mathcal{A}(g(Y))$. D'après la Proposition 2, on a que $f(X)$ et $g(Y)$ sont indépendantes. ∎

Exemple 9. — . Que peut-on dire si X et $f(X)$ sont indépendants ? D'après la Proposition 3, cela entraine que $f(X)$ est indépendante d'elle-même. Ainsi si $A \in \mathcal{A}(f(X))$ on a

$$P(A\cap A) = P(A)\times P(A) \Longrightarrow P(A)(1-P(A)) = 0.$$

Cela veut dire que $P(A)$ vaut 0 ou 1. Ainsi, si notre espace des issues est fini, et que chaque issue a une probabilité positive, alors $\mathcal{A}(f(X)) = \{\Omega\}$. Cette dernière égalité signifie que $f(X)$ est constante.

On peut généraliser à un nombre quelconque de partitions la notion d'indépendance.

Définition 10. — . Les partitions $\mathcal{A}_1$,. . .,$\mathcal{A}_n$ d'un même ensemble Ω sont indépendantes si pour tous $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$, ... et $A_n \in \mathcal{A}_n$, on a

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = \prod_{i=1}^{n} P(A_i).$$

Aussi, les variables $X_1, \ldots, X_n$ sont indépendantes, si les partitions associées $\mathcal{A}(X_1), \ldots, \mathcal{A}(X_n)$ sont indépendantes.

Proposition 4. — . *Si $X_1, X_2, \ldots, X_n : \Omega \to \mathbb{R}$ sont n variables indépendantes. Alors les variables de tout sous groupe de ces n-variables sont indépendantes.*

Démonstration. — Nous montrons que $X_2, \ldots, X_n$ sont indépendantes. Il est alors facile de conclure par récurrence. Soient donc $A_2 \in \mathcal{A}_2$, ... et $A_n \in \mathcal{A}_n$, et pour $A \in \mathcal{A}_1$

$$P(A\cap A_2 \cap \cdots \cap A_n) = P(A)\prod_{i=2}^{n} P(A_i).$$

Notons $B = A_2 \cap \cdots \cap A_n$, et notez que si $A, A' \in \mathcal{A}_1$ distincts alors $(B\cap A)\cap(B\cap A') = \emptyset$. Aussi,

$$B = \bigcup_{A\in\mathcal{A}_1} (B\cap A), \quad \text{et} \quad P(B) = \sum_{A\in\mathcal{A}_1} P(B\cap A).$$

Donc

$$P(B) = \sum_{A \in \mathcal{A}_1} P(A) \prod_{i=2}^{n} P(A_i) = \prod_{i=2}^{n} P(A_i) \left(\sum_{A \in \mathcal{A}_1} P(A) \right) = \prod_{i=2}^{n} P(A_i).$$

C'est ce qu'il nous fallait demontrer pour avoir $X_2, \dots, X_n$ indépendantes. ∎

Proposition 5. — . *Supposons que $X_1, \dots, X_n$ soient indépendantes. Alors X_1 est indépendante de $X_2 + \cdots + X_n$.*

Démonstration. — Notons $\mathcal{B} = \mathcal{A}(X_2) \bullet \cdots \bullet \mathcal{A}(X_n)$. Par hypothèse, $\mathcal{A}(X_1)$ et $\mathcal{B}$ sont indépendantes. Notons, comme dans la remarque 2, que si $\omega, \omega' \in B \in \mathcal{B}$, alors

$$X_2(\omega) + \cdots + X_n(\omega) = X_2(\omega') + \cdots + X_n(\omega').$$

Ce qui veut dire que $\mathcal{B}$ est plus fine que $\mathcal{B}' = \mathcal{A}(X_2 + \cdots + X_n)$, et il suffit d'invoquer la proposition 2. ∎

2.2. Lancers de pièces de monnaie indépendantes

2.2.1. Cas d'une pièce. — Les issues possibles sont pile ou face ; ce qu'on représente par

$$\Omega_1 = \{\text{pile,face}\}.$$

Si la pièce est équilibrée, pile et face ont la même chance de sortir. On associe la probabilité P_1

$$P_1(\text{pile}) = P_1(\text{face}) = 1/2$$

En général, on va supposer qu'une des faces est favorisée : il existe $\alpha \in [0, 1]$ tel que

$$P_1(\text{pile}) = \alpha \quad \text{et,} \quad P_1(\text{face}) = 1 - \alpha.$$

On définit la variable aléatoire $X : \Omega_1 \to \{0, 1\}$

$$X(\text{pile}) = 1 \quad \text{et,} \quad X(\text{face}) = 0.$$

X est appelée variable de Bernoulli de paramètre α. Par la notation $\{\omega : X(\omega) = 1\}$, on désigne l'ensemble des issues $\omega \in \Omega$ telles que $X(\omega) = 1$, c'est-à-dire ici $\omega = $ pile. X peut modéliser n'importe quelle situation où deux issues sont possibles.

2.2.2. Cas de deux pièces. — On suppose que pour la seconde pièce, il existe β qui a priori n'est pas lié à α tel que

$$P_2(p) = \beta \quad \text{et} \quad P_2(f) = 1 - \beta.$$

On interprète l'issue $\omega = (\omega_1, \omega_2)$ comme "la première pièce donne ω_1 et la seconde ω_2". On dit que la seconde pièce est semblable à la première si $\alpha = \beta$.

Définition 11. — . On dit que les deux pièces de monnaie sont indépendantes si

$$P((\omega_1, \omega_2)) = P_1(\omega_1).P_2(\omega_2). \tag{2.2.1}$$

Considérons maintenant l'évènement $\{\omega \in \Omega_2 : \ \omega_2 = p\}$ comme un ensemble d'issues de Ω_2 et calculons sa probabilité. D'abord

$$\{\omega \in \Omega_2 : \ \omega_2 = p\} = \{(p,p),(f,p)\}.$$

Et donc,

$$P(\{\omega \in \Omega_2 : \ \omega_2 = p\}) = P\left((p,p)\right) + P\left((f,p)\right) = \alpha\beta + (1-\alpha)\beta = \beta = P_2(p).$$

2.2.3. Cas de n pièces semblables et indépendantes. — Une issue consiste à spécifier l'état, pile ou face, de chacune des n pièces jetées. Donc,

$$\Omega_n = \{(\omega_1,\dots,\omega_n) \text{ ou } \omega_i \in \{p,f\}\}. \tag{2.2.2}$$

On note que Ω_n a 2^n éléments.

On définit, sur Ω_n, n variables de Bernoulli comme suit

$$\forall i = 1,\dots,n \qquad \forall \omega \in \Omega_n \quad X_i(\omega) = X(\omega_i).$$

où X est la variable de Bernoulli définie sur $\{p,f\}$.

Définition 12. — . Les n pièces sont semblables et indépendantes si

$$\forall \omega = (\omega_1,\dots,\omega_n) \in \Omega_n, \quad P\left(\omega_1,\dots,\omega_n\right) = P_1(\omega_1)P_1(\omega_2)\dots P_1(\omega_n).$$

Ceci signifie que $X_1,\dots,X_n$ sont indépendantes, et que pour tout $i = 1,\dots,n$

$$P(\omega : \ X_i(\omega) = p) = P_1(p), \quad P(\omega : \ X_i(\omega) = f) = P_1(f) = 1 - P_1(p).$$

Lorsque $\alpha = 1/2$, on note que chaque issue a la même probabilité $1/2^n$. Cela assure que lorsque l'on somme cette probabilité sur les 2^n issues de Ω_n on trouve bien 1. Comme toutes les issues ont la même probabilité $1/2^n$, lorqu'on demandera la probabilité d'un évènement quelconque, disons A, il suffira de compter combien il contient d'issues, disons K, pour obtenir $P(A) = K/2^n$. Pour n pièces semblables et indépendantes de paramètre $\alpha \neq 1/2$, la probabilité va dépendre du nombre de piles mais pas de l'ordre de sortie. Pour une issue $\omega = (\omega_1,\dots,\omega_n)$, notons $k(\omega)$ le nombre de piles parmi $\omega_1,\dots,\omega_n$. Alors

$$P(\omega) = \alpha^{k(\omega)}(1-\alpha)^{n-k(\omega)}.$$

L'utilité de cette suite $X_1,\dots,X_n$ à valeurs dans $\{0,1\}$ est que

$$S_n(\omega) = X_1(\omega) + X_2(\omega) + \cdots + X_n(\omega)$$

correspond au nombre de piles dans l'issue ω. Typiquement, la question est de calculer la probabilité de l'évènement $\{\omega \in \Omega_n : \ S_n(\omega) = k\}$ que l'on note souvent $\{S_n = k\}$.

2.2.4. Calcul de $P(\{\omega \in \Omega_n : \ S_n(\omega) = k\})$. — L'évènement $\{\omega \in \Omega_n : \ S_n(\omega) = k\}$ consiste en l'ensemble des lancers de n pièces qui ont donné k piles et $n-k$ faces. Combien y-a-t-il de telles issues ? En d'autres termes, on a n lancers, et on doit choisir k d'entres-eux où l'on mettra pile, alors que dans le reste des lancers c'est face qui sortira. Combien y-a-t'il de façon de choisir k objets (les piles) parmi n. Choississons une numérotation arbitraire des k piles : il y aura le premier pile, le second, etc... le premier 'pile' peut se choisir n'importe où parmi les n lancers ; le second 'pile' doit correspondre à une position différente du premier, et donc $n-1$ choix de positions restent possibles... et ainsi de suite jusqu'au k-ème 'pile' qui peut être choisie parmi $n-k+1$ lancers restants. Cependant, avec cet algorithme, un même choix de k positions apparaitra $k!$ fois en raison de la numérotation arbitraire que nous avons choisie au début. (On rappelle que factoriel k noté $k!$ est égal à $1.2.3\dots(k-1).k$, aussi $k!$ correspond au nombre de façons de numéroter k objets) On a donc $n(n-1)(n-2)\dots(n-k+1)/k!$ choix possibles. Ce nombre est appelé C_n^k

$$C_n^k = n(n-1)(n-2)\dots(n-k+1)/k! = \frac{n!}{(n-k)!k!},$$

et correspond donc au nombre de choix distincts de k éléments parmi n. On remarquera que $C_n^0 = 1$, $C_n^1 = n$ et $C_n^n = 1$. Les coéfficients $\{C_n^k, k = 0, \dots, n\}$ jouent un rôle important en mathématiques. On les retrouve dans le développement du Binôme dit de Newton, mais déjà démontré au Xème siècle à Bagdad par al-Karaji :

$$(a+b)^n = a^n + C_n^1 a^{n-1} b + C_n^2 a^{n-2} b^2 + \dots + C_n^{n-1} a b^{n-1} + b^n.$$

Revenons à $P(\{\omega \in \Omega_n : \ S_n(\omega) = k\})$. Notons que chaque issue de $\{\omega \in \Omega_n : S_n(\omega) = k\}$ a la même probabilité $\alpha^k (1-\alpha)^{n-k}$ et donc

$$P(\{\omega \in \Omega_n : \ S_n(\omega) = k\}) = \frac{n(n-1)(n-2)\dots(n-k)}{k!} \ \alpha^k (1-\alpha)^{n-k}.$$

On appelle S_n une variable Binomiale de paramètres n et α. On note $S_n \sim B(n, \alpha)$. On se souviendra qu'une variable Binomiale $S_n \sim B(n,p)$ est la somme de n variables indépendantes de Bernoulli de paramètres p.

2.3. Probabilités Conditionnelles

2.3.1. Motivation. — Nous allons introduire la notion de probabilité conditionnelle. On imagine un nombre infini de lancers indépendants de deux pièces. On modélise cette situation en prenant

$$\Omega = \{\omega = \{(\omega_i, \tilde{\omega}_i), \ i \in \mathbb{N}\}, \ \omega_i, \tilde{\omega}_i \in \{p, f\}\},$$

et pour tout entier n et tous les choix a_i, b_i dans $\{p, f\}$, pour $i = 1, \dots, n$

$$P(\{\omega : (\omega_i, \tilde{\omega}_i) = (a_i, b_i), \ i = 1, \dots, n\}) = P_1(a_1, b_1) \dots P_1(a_n, b_n).$$

où P_1 est une probabilité sur $\Omega_1 = \{(p,p), (p,f), (f,p), (f,f)\}$. Nous verrons au cours X que lorsque ces lancers sont indépendants, alors

$$\frac{|\{i \le n : \ \omega_i = p, \tilde{\omega}_i = p\}|}{n} \longrightarrow P_1(p,p).$$

On peut donc penser à $P_1(p,p)$ comme à la proportion asymptotique de lancers où $\{\omega_i = p, \tilde{\omega}_i = p\}$. On se pose la question : quelle est la probabilité que la 2ème pièce donne pile sachant que la 1ère a donné pile ? Intuitivement cela va correspondre à la proportion asymptotique de lancers où $\{\omega_i = p$ et $\tilde{\omega}_i = p\}$ sur ceux où $\{\omega_i = p\}$. En d'autres termes,

$$P(\tilde{\omega}_1 = p | \omega_1 = p) = \lim_{n\to\infty} \frac{|\{i \leq n : \ \omega_i = p, \tilde{\omega}_i = p\}|}{|\{i \leq n : \ \omega_i = p\}|}$$

On réécrit alors

$$\frac{|\{i \leq n : \ \omega_i = p, \tilde{\omega}_i = p\}|}{|\{i \leq n : \ \omega_i = p\}|} = \frac{|\{i \leq n : \ \omega_i = p, \tilde{\omega}_i = p\}|}{n} . \frac{n}{|\{i \leq n : \ \omega_i = p\}|}.$$

Cette dernière quantité converge vers

$$\frac{P_1(\omega_1 = p, \tilde{\omega}_1 = p)}{P_1(\omega_1 = p)}.$$

Cette discussion intuitive motive la définition suivante.

2.3.2. Définitions. —

Définition 13. — . Soit A et B deux évènements sur $(\Omega, \mathcal{F}, P)$ et $P(B) > 0$. La probabilité conditionnelle de A sachant B est

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

On dit aussi que $P(A|B)$ est la probabilité de A sachant B.

2.3.3. Exemples. —

Exemple 10. — . On modélise le lancer de deux pièces semblables indépendantes par

$$\Omega = \{\omega = (\omega_1, \omega_2) \ \omega_i \in \{p, f\}\} \quad \text{et} \quad P(\omega_1, \omega_2) = P_1(\omega_1)P_1(\omega_2),$$

avec $P_1(p) = \alpha$ et $P_1(f) = 1 - \alpha$. On s'imagine qu'on a lancé la 1ère pièce et que $\omega_1 = p$. On associe à chaque issue du lancer de la 2ème pièce une probabilité (sachant que $\omega_1 = p$).

$$P(\omega_2 = p | \omega_1 = p) = \frac{P(\{p,p\})}{P(\{\omega : \omega_1 = p\})} = \frac{\alpha\alpha}{\alpha(1-\alpha) + \alpha\alpha} = \alpha.$$

Intuitivement, l'indépendance des deux pièces signifie que la connaissance du résultat du lancer de la 1ère ne nous apprend rien sur la seconde.

Exemple 11. — . On suppose qu'on a deux paires de pièces différentes : la première paire consiste en deux pièces indépendantes et pile sort 99 fois sur 100 en moyenne ; la seconde paire consiste en deux pièces indépendantes et pile sort 10 fois sur 100 en moyenne. A chaque coup, on tire au hasard une des paires, disons avec probabilité 1/2, et on lance les deux pièces de la paire choisie. On modélise l'espace des issue d'un tel jeu par

$$\Omega = \{\omega = (\omega_1, \omega_2, n) : \ \omega_i \in \{p, f\}, \ n = 1, 2\},$$

et

$$P(\omega_1, \omega_2, 1) = \frac{1}{2} P_1(\omega_1) P_1(\omega_2) \quad \text{et} \quad P_1(p) = \alpha = 0,99$$

et

$$P(\omega_1, \omega_2, 2) = \frac{1}{2} P_2(\omega_1) P_2(\omega_2) \quad \text{et} \quad P_2(p) = \beta = 0,1$$

On calcule ici

$$\begin{aligned} P(\omega_2 = p | \omega_1 = p) &= \frac{P(\{\omega : \omega_1 = p, \omega_2 = p\})}{P(\{\omega : \omega_1 = p\})} \\ &= \frac{P(p,p,1) + P(p,p,2)}{P(p,p,1) + P(p,f,1) + P(p,p,2) + P(p,f,2)} \\ &= \frac{\alpha^2/2 + \beta^2/2}{\alpha/2 + \beta/2} = \frac{\alpha^2 + \beta^2}{\alpha + \beta}. \end{aligned}$$

Par ailleurs, si on ne savait rien sur la première pièce

$$P(\omega_2 = p) = P\left(\{p,p,1\}, \{f,p,1\}, \{p,p,2\}, \{f,p,2\}\right) = \frac{1}{2}(\alpha + \beta).$$

On note que $P(\omega_2 = p | \omega_1 = p) > P(\omega_2 = p)$. Intuitivement, comme pile est sortie la 1ère fois, on a plus de chance que la paire choisie soit 1.

2.3.4. Formule de Bayes. —

Proposition 6. — *. Soit $\{A_i, i = 1, \ldots, n\}$ une partition, telle que pour tout i $P(A_i) > 0$. Pour tout $B \in \mathcal{F}$*

$$P(B) = \sum_{i=1}^{n} P(B|A_i) P(A_i).$$

Démonstration. — Notons que $B = B \cap (\cup_{i=1}^n A_i) = \cup_{i=1}^n B \cap A_i$ et que comme les A_i sont disjoints, les $A_i \cap B$ le sont aussi. Donc,

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B|A_i) P(A_i).$$

∎

Lemme 1. — *. Soient $A_1, \ldots, A_n \in \mathcal{F}$, tels que $P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$ alors,*

$$P(A_1 \cap \cdots \cap A_n) = P(A_n | A_1 \cap \cdots \cap A_{n-1}) P(A_{n-1} | A_1 \cap \cdots \cap A_{n-2}) \ldots P(A_2|A_1) P(A_1).$$

Pour démontrer ce Lemme, il suffit d'écrire la définition des termes de gauche ($P(A_n | A_1 \cap \cdots \cap A_{n-1})$,...) et de voir que tous les dénominateurs se simplifient.

CHAPITRE 3

ESPÉRANCE, VARIANCE, ET EXEMPLES DE LOIS

3.1. Espérance

Soient Ω l'espace des issues d'un jeu de dé (donc $\Omega = \{1, \ldots, 6\}$), et P la probabilité uniforme sur Ω qui traduit que le dé est bien équilibré. Soit $A \subset \Omega$, (par exemple les issues paires $\{2, 4, 6\}$) et imaginons que pour chaque issue ω, on gagne 1 euro chaque fois que $\omega \in A$, et qu'on ne gagne rien sinon. Si l'issue ω se réalise, alors le gain est $G(\omega) = 1_A(\omega)$. En probabilité, ce qui nous intéresse vraiement est combien on gagne en moyenne ? Supposons qu'on joue 100 fois, et que l'on note les issues $\omega_1, \ldots, \omega_{100}$. Ainsi, le gain total est

$$G(\omega_1) + G(\omega_2) + \cdots + G(\omega_{100}) = 1_A(\omega_1) + \cdots + 1_A(\omega_{100}),$$

alors que le gain moyen par jeu est pour $n = 100$

$$\frac{1_A(\omega_1) + \cdots + 1_A(\omega_n)}{n}.$$

Nous verons que plus n est grand, moins cette moyenne fluctue : on veut dire par-là que deux jeux de 1000 lancers donneront à peu près la même moyenne. C'est la limite de ces moyennes que nous appellerons la probabilité qu'a A de se réaliser. Ainsi le gain moyen pour un jeu infini est la probabilité de A : on note cette quantité $E[1_A]$, que l'on appelle aussi l'espérance mathématique de 1_A, on pose donc par définition $E[1_A] = P(A)$.

Plus généralement, supposons que $\{A_i, i = 1, \ldots, 7\}$ soit une partition de Ω en 7 parts, et que l'on gagne x_i euros si $\omega \in A_i$. Le gain est donc

$$X(\omega) = x_1 1_{A_1}(\omega) + x_2 1_{A_2}(\omega) + \cdots + x_7 1_{A_7}(\omega).$$

Combien gagne-t-on en moyenne ? Sur 1000 lancers, on gagne en moyenne

$$\begin{aligned} \frac{X(\omega_1) + \cdots + X(\omega_{1000})}{1000} &= \big(x_1 \frac{1_{A_1}(\omega_1) + \cdots + 1_{A_1}(\omega_{1000})}{1000} \\ &\quad + \cdots + x_7 \frac{1_{A_7}(\omega_1) + \cdots + 1_{A_7}(\omega_{1000})}{1000}\big). \end{aligned}$$

On voit donc que lorsque le nombre de lancers est très grand, on gagne x_i euros avec probabilités $P(A_i)$ pour chaque i de 1 à 7. Ainsi, le gain moyen, noté $E[X]$, est

$$E[X] = x_1 P(A_1) + x_2 P(A_2) + \cdots + x_7 P(A_7). \tag{3.1.1}$$

Nous allons définir le gain moyen autrement, puis vérifier que cela correspond à la même formule que (3.1.1).

Définition 14. — . Soit Ω une espace fini, soient P une probabilité sur Ω et X une variable aléatoire, alors

$$E[X] = \sum_{\omega\in\Omega} X(\omega)P(\omega).$$

Il est facile de voir que cette définition donne $E[1_A] = P(A)$. En effet

$$E[1_A] = \sum_{\omega\in\Omega} 1_A(\omega)P(\omega) = \sum_{\omega\in A} P(\omega) = P(A).$$

Reprenons l'exemple précédent, et imaginons que l'on gagne x_i euros si $\omega \in A_i$ pour $\{A_i,\ i = 1, \ldots, n\}$ une partition de Ω. Combien gagne-t-on en moyenne ? Le gain est

$$X(\omega) = x_1 1_{A_1}(\omega) + \cdots + x_n 1_{A_n}(\omega).$$

Par définition le gain moyen $E[X]$ est

$$\begin{aligned} E[X] &= \sum_{\omega\in\Omega} X(\omega)P(\omega) = \sum_{i=1}^{n} \sum_{\omega\in A_i} X(\omega)P(\omega) \\ &= \sum_{i=1}^{n} x_i \left(\sum_{\omega\in A_i} P(\omega) \right) = x_1 P(A_1) + \cdots + x_n P(A_n). \end{aligned} \tag{3.1.2}$$

Remarque 3. — . On retiendra aussi que si $\{A_i,\ i = 1, \ldots, n\}$ est une partition de Ω, et si $\{x_i,\ i = 1, \ldots, n\}$ sont n réels, alors l'espérance mathématique a la propriété suivante

$$E[x_1 1_{A_1} + \cdots + x_n 1_{A_n}] = x_1 P(A_1) + \cdots + x_n P(A_n).$$

Aussi, si $X = x_1 1_{A_1} + \cdots + x_n 1_{A_n}$ et f une fonction réelle, alors en utilisant (1.2.1) on a

$$E[f(X)] = f(x_1)P(A_1) + \cdots + f(x_n)P(A_n).$$

Remarque 4. — . Pour calculer l'espérance de $f(X)$, il suffit de connaitre la loi de X.

3.1.1. Exemples. —

Exemple 12. — . On continue notre exemple 1, d'un dé équilibré. Si un nombre pair sort, on gagne 2 euros, alors que si un nombre impair sort, on perd 3 euros. Combien gagne-t-on en moyenne ? On modélise notre jeu par

$$\Omega = \{1, 2, 3, 4, 5, 6\} \quad \text{et} \quad P(i) = 1/6,\ i = 1, 2, 3, 4, 5, 6.$$

Le gain est une variable aléatoire X, avec si $\omega \in \{1,3,5\}$, alors $X(\omega) = -3$, alors que si $\omega \in \{2,4,6\}$, alors $X(\omega) = 2$. Ainsi donc

$$X = 2.1_{\{\omega \in \{2,4,6\}\}} - 3.1_{\{\omega \in \{1,3,5\}\}}.$$

Ainsi $E[X] = 2.1/2 - 3.1/2 = -1/2$.

Exemple 13. — . On continue l'exemple 4 des deux dés, et de la personne II qui n'en voit que la somme. On décide que si la somme est un nombre premier (ce sont les nombres qui n'ont pas de diviseurs autres que 1 et eux-mêmes) alors on gagne 2 euros, si la somme est un multiple de 2 (autre que 2, qui lui est un nombre premier) on perd 3 euros, et que si la somme est un autre nombre, on gagne 5 euros. Quel est le gain moyen ?

On va travailler avec Ω_{II}. Si $\omega \in \{2,3,5,7,11\}$, $X(\omega) = 2$, si $\omega = 9$, $X(\omega) = 5$, et dans le reste des cas $X(\omega) = -3$. Finalement,

$$\begin{aligned} E[X] &= 2P_{II}(\omega : \ X(\omega) = 2) - 3P_{II}(\omega : \ X(\omega) = -3) + 5P_{II}(9) \\ &= 2\frac{(1+2+4+6+2)}{36} - 3\frac{(3+5+5+3+1)}{36} + 5\frac{4}{36} = \frac{30-51+20}{36} = -1/36. \end{aligned}$$

3.1.2. Propriétés de l'Espérance. —

Proposition 7. — . *Soit X et Y deux variables aléatoires et $\lambda \in \mathbb{R}$. Alors,*

1. $E[\lambda X] = \lambda E[X]$.
2. $E[X+Y] = E[X] + E[Y]$.

Démonstration. — 1) Par définition

$$E[\lambda X] = \sum_{\omega \in \Omega} \lambda X(\omega) P(\omega) = \lambda \sum_{\omega \in \Omega} X(\omega) P(\omega) = \lambda E[X].$$

2) Nous allons présenter deux preuves de ce résultat. D'abord, en utilisant la définition

$$E[X+Y] = \sum_{\omega \in \Omega} (X(\omega) + Y(\omega)) P(\omega) = \sum_{\omega \in \Omega} X(\omega) P(\omega) + \sum_{\omega \in \Omega} Y(\omega) P(\omega) = E[X] + E[Y].$$

L'autre preuve a l'air plus compliquée, mais d'une part, elle se généralise au cas où l'espace des issues n'est pas fini, et d'autre part, elle reflète la représentation qu'ont les probabilistes d'une variable aléatoire : comme somme de fonctions indicatrices. Soit $\{A_i, \ i = 1, \ldots, n\}$ la partition associée à X, et soit $\{B_i, \ i = 1, \ldots, m\}$ celle associée à Y. La première observation est que $C_{i,j} = A_i \cap B_j$ pour $i = 1, \ldots, n$ et $j = 1, \ldots, m$ est une partition (à vérifier !). On peut réécrire

$$X = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i 1_{A_i \cap B_j}, \quad \text{et} \quad Y = \sum_{i=1}^{n} \sum_{j=1}^{m} y_j 1_{A_i \cap B_j}.$$

Donc, en appliquant la définition

$$X + Y = \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i + y_j) 1_{A_i \cap B_j} \Longrightarrow E[X+Y] = \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i + y_j) P(A_i \cap B_j).$$

On utilise que pour tout i, $P(A_i \cap B_1) + \cdots + P(A_i \cap B_m) = P(A_i)$, et de même pour tout j, $P(A_1 \cap B_j) + \cdots + P(A_n \cap B_j) = P(B_j)$ pour conclure

$$E[X+Y] = \sum_i x_i P(A_i) + \sum_j y_j P(B_j) = E[X] + E[Y].$$

∎

Nous énonçons sans démonstration certaines propriétés simples de l'espérance. Il est important de les vérifier seul !

Proposition 8. — . *Soient X, Y deux variables aléatoires sur (Ω, P), et $\lambda \in \mathbb{R}$.*

1. $E[\lambda] = \lambda$.
2. $X \geq 0 \Longrightarrow E[X] \geq 0$.
3. $X \geq Y \Longrightarrow E[X] \geq E[Y]$.
4. $|E[X]| \leq E[|X|]$.

Exemple 14. — . Nous continuons l'exemple 5 (et 8). Quel est le nombre moyen de lettres utilisées ?
Solution : On a vu que le nombre de lettres utilisées dans le mot ω est

$$N = \sum_{x \in \mathcal{A}} 1_{A_x}(\omega).$$

On veut donc calculer la moyenne de N, ou encore son espérance. Ainsi,

$$E[N] = E[\sum_{i=1}^{26} 1_{A_i}] = \sum_{i=1}^{26} E[1_{A_i}] = \sum_{i=1}^{26} P(A_i) = 26(1 - (\frac{25}{26})^{15}).$$

Remarque 5. — . Un résultat utile concerne le développement d'un produit de deux sommes : soient m et n deux entiers et $a_1, \ldots, a_n$ et $b_1, \ldots, b_m$ deux suites de nombres réels. Alors

$$(\sum_{i=1}^{n} a_i)(\sum_{j=1}^{m} b_j) = \sum_{i,j} a_i b_j.$$

3.2. Variance

Définition 15. — . La variance d'une variable aléatoire X est une nombre positif donné par

$$\mathrm{var}(X) = E[(X - E[X])^2].$$

En développant le carré et en utilisant les propositions 7 et 8, on obtient

$$0 \leq (X - E[X])^2 = X^2 - 2X.E[X] + (E[X])^2 \Longrightarrow \mathrm{var}(X) = E[X^2] - (E[X])^2 \geq 0.$$

On note donc que pour toute variable aléatoire X, $E[X^2] \geq (E[X])^2$.

Proposition 9. — . *Propriétés de la variance.*

Soit X une variable aléatoire et $\lambda \in \mathbb{R}$.

1. $var(\lambda X) = \lambda^2 var(X)$.
2. $var(X + \lambda) = var(X)$.
3. $var(\lambda) = 0$.
4. $var(X) = 0$ *ssi* $\forall \omega \in \Omega$ *tel que* $P(\omega) > 0$, $X(\omega) = E[X]$.

Démonstration. — 1.
$$\begin{aligned} \mathrm{var}(\lambda X) &= E\left[(\lambda X - E[\lambda X])^2\right] \\ &= E\left[\lambda^2 (X - E[X])^2\right] \\ &= \lambda^2 \mathrm{var}(X) \text{ par linéarité de l'espérance.} \end{aligned}$$

2. $\mathrm{var}(X + \lambda) = E\left[(X + \lambda - E[X + \lambda])^2\right] = E\left[(X + \lambda - E[X] - \lambda)^2\right] = \mathrm{var}(X)$.

3. $\mathrm{var}(\lambda) = E\left[(\lambda - E[\lambda])^2\right] = E\left[(\lambda - \lambda)^2\right] = E[0] = 0$.

4. Comme $\mathrm{var}(X) = \sum_{\omega \in \Omega} (X(\omega) - E[X])^2 P(\omega)$ est une somme de termes positifs, pour que $\mathrm{var}(X) = 0$, il faut et il suffit que tous les termes de la somme soient nuls.

∎

Proposition 10. — *. Soient deux variables aléatoires X et Y indépendantes. Alors $E[X \times Y] = E[X] \times E[Y]$.*

Démonstration. — Commençons par comprendre le cas le plus simple. Supposons que $X = 1_A$ et $Y = 1_B$, pour $A, B \subset \Omega$. Que signifie que X et Y sont indépendantes ? Cela veut dire que $\mathcal{A}(X)$ et $\mathcal{A}(Y)$ sont indépendantes. Or $\mathcal{A}(X) = \{A, A^c\}$ et $\mathcal{A}(Y) = \{B, B^c\}$. On doit donc avoir $P(A \cap B) = P(A) \times P(B)$. Maintenant $1_A \times 1_B = 1_{A \cap B}$ donc
$$E[XY] = E[1_{A \cap B}] = P(A \cap B) = P(A) \times P(B) = E[X] \times E[Y].$$
On va généraliser ceci à des variables plus compliquées. Supposons que $X : \Omega \to \{x_1, \ldots, x_M\}$, et $Y : \Omega \to \{y_1, \ldots, y_N\}$. Nous construisons les partitions $\mathcal{A}(X), \mathcal{A}(Y)$: pour $k = 1, \ldots, M$ et $l = 1, \ldots, N$, on pose $A_k = \{\omega : X(\omega) = x_k\}$ et $B_l = \{\omega : Y(\omega) = y_l\}$. Ainsi,
$$X = \sum_{i=1}^{M} x_i 1_{A_i}, \quad \text{et} \quad Y = \sum_{j=1}^{N} y_j 1_{B_j}.$$
Donc
$$E[XY] = \sum_{i,j} x_i y_j P(A_i \cap B_j).$$
On utilise maintenant que X et Y sont indépendantes : pour tous i, j on a $P(A_i \cap B_j) = P(A_i) P(B_j)$ et donc
$$\begin{aligned} E[XY] &= \sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j P(A_i) \times P(B_j) \\ &= \left(\sum_{i=1}^{n} x_i P(A_i)\right) \left(\sum_{j=1}^{m} y_j P(B_j)\right) = E[X] E[Y]. \end{aligned}$$

∎

Corollaire 1. — . *Soient X et Y des variables indépendantes, alors $var(X+Y) = var(X) + var(Y)$.*

Démonstration. — Il suffit d'écrire la définition de la variance et de développer.

$$\begin{aligned} var(X+Y) &= E[(X+Y-E[X+Y])^2] \\ &= E\left[(X-E[X])^2 + (Y-E[Y])^2 + 2(X-E[X])(Y-E[Y])\right], \end{aligned}$$

donc, par la Proposition 1,2) du cours I,

$$var(X+Y) = var(X) + var(Y) + 2E[(X-E[X])(Y-E[Y])].$$

Soit $f(x) = x - E[X]$ et $g(x) = x - E[Y]$ deux fonctions réelles. Notons d'abord que $E[f(X)] = 0 = E[g(Y)]$, puis que $f(X)$ et $g(Y)$ sont indépendantes. On applique la proposition précédente à ces fonctions :

$$E[(X-E[X])(Y-E[Y])] = E[f(X)g(Y)] = E[f(X)]E[g(Y)] = 0.$$

∎

3.2.1. Les inégalités de Markov.— Intuitivement, si l'on pense à une variable aléatoire positive comme au gain que l'on ferait à une jeu de chance, alors l'inégalité de Markov dit que si le gain moyen est petit, alors la probabilité de gagner beaucoup doit être petite.

Proposition 11. — .

Soit X une v.a. **positive**. *Alors, pour tout $c > 0$,*

$$P(X \geq c) \leq \frac{E[X]}{c}.$$

Démonstration. — On pose $A = \{X \geq c\}$ et on note que

$$c\,1_A(\omega) \leq X(\omega).$$

On prend l'espérance de chaque terme

$$E[c1_A] = cP(A) \leq E[X],$$

et le résultat en découle. ∎

Corollaire 2. — . *Inégalité de Markov. Soit X une v.a. de moyenne μ et de variance σ^2. Pour tout $\epsilon > 0$,*

$$P\left[|X-\mu| \geq \epsilon\right] \leq \frac{\sigma^2}{\epsilon^2}.$$

Démonstration. —

$$\begin{aligned} P\left[|X-\mu| \geq \epsilon\right] &= P\left[(X-\mu)^2 \geq \epsilon^2\right] \\ &\leq \frac{E\left[(X-\mu)^2\right]}{\epsilon^2} \quad \text{par la Proposition 11 appliquée à } Y = (X-\mu)^2, \\ &= \frac{\text{var}(X)}{\epsilon^2} \quad \text{car } E[X] = \mu. \end{aligned}$$

■

3.3. Exemples de lois

3.3.1. Loi de Bernoulli. — L'espace est celui d'une suite de n lancers d'une pièce. Ici n est un entier fixé.

$$\Omega_n = \{(\omega_1, \ldots, \omega_n) \text{ où pour tout } i\ \omega_i \in \{\text{pile,face}\}\}. \tag{3.3.1}$$

On fixe maintenant α dans $[0,1]$ qui va correspondre à la probabilité de voir pile : $P_1($ pile $) = \alpha$, et sur Ω_n on définit

$$\forall \omega = (\omega_1, \ldots, \omega_n) \in \Omega_n, \quad P\left(\omega_1, \ldots, \omega_n\right) = P_1(\omega_1) \times P_1(\omega_2) \times \cdots \times P_1(\omega_n).$$

On définit, sur Ω_n, n variables de Bernoulli comme suit

$$\forall i = 1, \ldots, n \qquad \forall \omega \in \Omega_n \quad X_i(\omega) = 1_{\{\omega_i = \text{ pile }\}}.$$

On peut aisément vérifier que $X_1, \ldots, X_n$ sont indépendantes.

3.3.2. Loi Binomiale. — Sur le même espace Ω_n, on définit

$$S_n = X_1 + \cdots + X_n.$$

$S_n : \Omega_n \to \{0, \ldots, n\}$ est une variable binomiale, et nous avons vu que pour tout $k \in [0,n]$

$$P(\omega : \ S_n(\omega) = k) = C_n^k \alpha^k (1-\alpha)^{n-k}.$$

On peut alors vérifier que

$$E[S_n] = n\alpha, \quad \text{et} \quad var(S_n) = n\alpha(1-\alpha).$$

3.3.3. Loi de Poisson. — On introduit une variable de Poisson comme limite de somme de variables de Bernoulli indépendantes. On fixe un entier N très grand, et on considère $X_1^{(N)}, X_2^{(N)}, \ldots, X_N^{(N)}$ N variables de Bernoulli indépendantes et de même loi telle que pour tout $i = 1, \ldots, N$

$$P(X_i^{(N)} = 1) = \frac{x}{N}, \quad \text{et} \quad P(X_i^{(N)} = 0) = 1 - \frac{x}{N},$$

où x est un nombre positif fixé. Notez qu'au rang $N+1$, on considèrera

$$X_1^{(N+1)}, X_2^{(N+1)}, \ldots, X_{N+1}^{(N+1)},$$

$N+1$ variables de Bernoulli indépendantes et de même loi différente de celle des $X_i^{(N)}$. En effet,

$$P(X_i^{(N+1)} = 1) = \frac{x}{N+1}, \quad \text{et} \quad P(X_i^{(N+1)} = 0) = 1 - \frac{x}{N+1}.$$

On pose

$$S_N = X_1^{(N)} + X_2^{(N)} + \cdots + X_N^{(N)},$$

et,

$$S_{N+1} = X_1^{(N+1)} + X_2^{(N+1)} + \cdots + X_{N+1}^{(N+1)},$$

et ainsi de suite...On sait que pour tout N, S_N est une variable binomiale et pour un entier k fixé (avec $k \leq N$ pour N assez grand)

$$\begin{aligned} P(S_N = k) &= C_N^k (\frac{x}{N})^k (1 - \frac{x}{N})^{N-k} \\ &= \frac{N(N-1)(N-2)\dots(N-k+1)}{k!N^k} x^k (1 - \frac{x}{N})^{-k} (1 - \frac{x}{N})^N \\ &= \frac{x^k}{k!}(1 - \frac{x}{N})^{-k} \frac{N(N-1)(N-2)\dots(N-k+1)}{N.N.N\dots N}(1 - \frac{x}{N})^N. \end{aligned} \tag{3.3.1}$$

On fixe k et on prend N très grand. Noter que

$$\lim_{N\to\infty} \frac{N(N-1)(N-2)\dots(N-k+1)}{N.N\dots N} = 1, \quad \text{et} \quad \lim_{N\to\infty} (1 - \frac{x}{N})^{-k} = 1.$$

Notons $f(x) = \lim_{N\to\infty}(1 - \frac{x}{N})^N$. Ainsi, en regroupant toutes les limites

$$\lim_{N\to\infty} P(S_N = k) = \frac{x^k}{k!} f(x).$$

Nous allons maintenant montrer que $f(x) = \exp(-x)$. Noter que

$$P(S_N = 0) + P(S_N = 1) + P(S_N = 2) + P(S_N = 3) + \dots = 1,$$

et quand on prend la limite

$$\lim_{N\to\infty} (P(S_N = 0) + P(S_N = 1) + P(S_N = 2) + P(S_N = 3) + \dots) = 1,$$

et donc

$$\frac{x^0}{0!} f(x) + \frac{x^1}{1!} f(x) + \frac{x^2}{2!} f(x) + \frac{x^3}{3!} f(x) + \dots = 1.$$

On factorise $f(x)$ et on se rappelle que $\exp(x)$ correspond à la série suivante

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots,$$

et donc $f(x)\exp(x) = 1$, ce qui donne le résultat. On appelle variable de Poisson, notée S ici, la limite des variables binomiale S_N, lorsque N tend vers l'infini. Ainsi une variable de Poisson est à valeurs dans $I\!N$ et sa loi est définie par la suite des probabilités : pour tout entier $k = 0, 1, 2, \dots$

$$P(S = k) = e^{-x} \frac{x^k}{k!}.$$

3.3.4. Loi Géométrique. — Soit $X_1, X_2, X_3, \dots$ une suite de variables de Bernoulli indépendantes et de même loi avec $P(X_1 = 1) = p = 1 - P(X_1 = 0)$. Soit T la première fois où apparait pile. Ainsi si la suite des lancers donne $ffppppfp\dots$ on aura que $T = 3$. On appelle T une variable géométrique. T peut être n'importe quelle valeur $1, 2, 3, \dots$ et pour tout $k = 1, 2, 3, \dots$

$$P(T = k) = p(1 - p)^{k-1}$$

En utilisant que pour $0 \leq x < 1$, on a

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \dots$$

on vérifie que $P(T=1)+P(T=2)+P(T=3)+\cdots=1$

3.3.5. Loi Géométrique négative. — On fixe un entier $k \geq 1$, et dans le même contexte que celui de la loi géometrique, on définit T_k comme étant le premier instant où pile apparait pour la k-ème fois. Notez que $\{T_k = n\}$ signifie qu'au n-ème coup un pile sort, et que parmi les $n-1$ coups précédents 'pile' est sortie $k-1$ fois. On a donc pour $k \geq n$

$$P(T_k = n) = P(X_n = 1, \text{ et } X_1 + \cdots + X_{n-1} = k-1) = pC_{n-1}^{k-1} p^{k-1}(1-p)^{n-k}.$$

T_k est appelée variable géométrique négative.

Considérons $T^{(1)}, T^{(2)}, \ldots, T^{(k)}$ k variables géometriques indépendantes, et montrez que la loi de T_k est la même que celle de $T^{(1)} + \cdots + T^{(k)}$. Ceci devrait vous permettre de calculer sans difficulté l'espérance et la variance de T_k.

3.3.6. Loi Multinomiale. — Nous allons généraliser ici la loi Binomiale. Nous allons d'abord proposer une autre modélisation de l'expérience du lancer de n pièces. Il suffit de donner la position des piles parmi les n lancers pour décrire complètement les n lancers. Ainsi, on peut penser aux issues comme aux sous-ensembles de $\{1, \ldots, n\}$; un tel sous-ensemble représente la position des piles.

$$\Omega_n = \{\omega = A \subset \{1, \ldots, n\}\}.$$

et si $|A|$ représente le nombre d'éléments dans A, alors

$$P(\omega = A) = p^{|A|}(1-p)^{n-|A|}, \quad \text{et} \quad P(|\omega| = k) = C_n^k p^k (1-p)^{n-k}.$$

Considérons n tirages avec remise dans une urne qui contient R_1 boules numérotées 1, R_2 boules numérotées 2, ..., jusqu'à R_d boules numérotées d. On pensera aux numéros comme à des couleurs. Une issue va spécifier les sous-ensembles occupés par chaque couleur. Ainsi,

$$\Omega_n = \{\omega = (A_1, \ldots, A_d) : \quad \text{où } \{A_1, \ldots, A_d\} \text{ forment une partition de } \{1, \ldots, n\}\}.$$

Pour $\omega = (A_1, \ldots, A_d)$, la donnée d'une partition de $\{1, \ldots, n\}$, on note que $|A_x|$ est le nombre de boules colorées x. On appellera $p(x)$ la probabilité de tirer la couleur x à un tirage :

$$p(x) = \frac{R_x}{R_1 + \cdots + R_d}.$$

Maintenant la probabilité de voir $\omega = (A_1, \ldots, A_d)$ quelconque est

$$P(\omega) = p(1)^{|A_1|} \times \cdots \times p(d)^{|A_d|}.$$

Noter que pour deux configurations ω et ω' telles que $|A_x| = |A'_x|$ pour tous les $x = 1, \ldots, d$, on a $P(\omega) = P(\omega')$. Ainsi si on fixe d entiers n_1,...,n_d tels que $n_1 + \cdots + n_d = n$, alors

$$P(\omega : \ |A_x| = n_x, \ \forall x = 1, \ldots, d) = \prod_{x=1}^{d} p(x)^{n_x} \times |\{\omega : \ |A_x| = n_x, \ \forall x = 1, \ldots, d\}|.$$

Noter que si $d = 2$, alors d'abord $p(1) + p(2) = 1$ et ensuite lorsque $n_1 + n_2 = n$ $\{\omega : |A_1| = n_1, |A_2| = n_2\}$ s'ecrit aussi $\{\omega : |A_1| = n_1\}$, et nous avons une loi binomiale :

$$P(\{\omega : |A_1| = n_1, |A_2| = n_2\}) = \frac{n!}{k!(n-k)!} p(1)^{n_1}(1-p(1))^{n-n_1}.$$

Nous laissons au lecteur le soin de vérifier par récurrence, sur $d \geq 2$, que

$$|\{\omega : \ |A_x| = n_x, \ \forall x = 1, \dots, d\}| = \frac{n!}{\prod_{x=1}^{d}(n_x)!}.$$

On dit que les variables $(|A_1|, \dots, |A_d|)$ suivent la loi multinomiale. On pourra montrer que pour $x = 1, \dots, d$, on a que $|A_x|$ suit une loi binomiale de paramètre $p(x)$.

Question : $(|A_1|, \dots, |A_d|)$ sont-elles indépendantes ?

Résponse : Non. Un suggestion pour le démontrer : $|A_1| + \cdots + |A_d| = n$.

3.3.7. Loi Hypergéométrique. — Dans la population totale N, on suppose que N_F personnes ont la propriété F, (être fumeur par exemple). Dans un échantillon de taille n, soit S_n le nombre de personnes qui ont la propriété F. Soit un entier $k \in [0, n]$. Alors, $P(S_n = k)$ est égale au nombre de tirages où k parmi les n sont F, divisé par le nombre totale de tirages de n éléments parmi N. Ainsi,

$$P(S_n = k) = \frac{C_{N_F}^{k} C_{N-N_F}^{n-k}}{C_N^n}. \tag{3.3.2}$$

On appelle S_n une variable hypergéométrique, et on note sa loi $\mathcal{H}(N, p, n)$: elle dépend de trois paramètres.

CHAPITRE 4

ESTIMATEURS

4.1. Motivation

Une usine a produit un million de postes de radio. On veut estimer la proportion p de postes défectueux sans les vérifier tous. On va donc choisir un échantillon de taille n très petite par rapport à 10^6, ce que l'on note $n \ll 10^6$, et on note $x_k = 1$ si le k^{eme} poste de l'échantillon est défectueux, et $x_k = 0$ sinon. La proportion des postes défectueux sur l'échantillon est

$$\bar{x}_n = \frac{x_1 + \cdots + x_n}{n}.$$

Il est naturel d'estimer l'inconnu p à l'aide de $\bar{x}_n$. On va étudier de façon quantitative si c'est une bonne façon d'estimer p, et quelle taille doit avoir l'échantillon pour que "$\bar{x}_n$ soit proche de p" ?

Commençons par préciser cette dernière question. Imaginez par exemple, qu'il y ait cent mille postes défectueux sur un million. Imaginez aussi qu'on mesure $\bar{x}_n$ sur plusieurs échantillons $I_1, \ldots, I_N$ tous de taille n. Chaque échantillon produira une valeur différente de $\bar{x}_n$. Lorsqu'on demandera que "$\bar{x}_n$ soit proche de p", ce sera donc en moyenne. Ainsi, on pensera à une des valeurs de $\bar{x}_n$ comme à la réalisation particulière d'une issue, disons ω,

$$\bar{x}_n = \frac{X_1(\omega) + \cdots + X_n(\omega)}{n} = \frac{S_n(\omega)}{n} \equiv \bar{X}_n(\omega)$$

où ω appartient à l'espace de tous les états possibles de n postes

$$\Omega_n = \{(\omega_1, \ldots, \omega_n),\ \omega_i \in \{D, M\}\},$$

où D signifie défectueux et M signifie qui marche, et les X_i sont des variables de Bernoulli indépendantes de paramètre p que l'on cherche à évaluer.

L'estimateur sera donc une **variable aléatoire**, $\bar{X}_n$ dans l'exemple ci-dessus, et la valeur d'un échantillon sera interprétée comme $\bar{X}_n(\omega)$.

Pour que l'on ait de bonnes chances d'estimer p à l'aide de $\bar{X}_n$ il faudrait qu'il y ait peu d'issues ω telle que $\bar{X}_n(\omega)$ soit loin de p. En d'autres termes, il faudrait qu'il y ai peu d'echantillons où $\bar{x}_n$ soit loin de p. Par exemple, si $p = 0,1$ et $n = 20$, quelle est la probabilité de $\{\omega : |\bar{X}_n(\omega) - p| \leq 0,05\}$?

$$|\bar{X}_n(\omega) - p| \leq 0,05 \Longleftrightarrow |S_n(\omega) - 2| \leq 1 \Longleftrightarrow S_n(\omega) \in \{1,2,3\}.$$

En faisant le calcul, on trouve $P(\{\omega : S_n(\omega) \in \{1,2,3\}\}) \simeq 0,75$. On dira donc que dans 75% des cas, ou pour 75 issues ω sur 100, $|\bar{X}_n(\omega) - p| \leq 0,05$. Il y a un gros problème : nous avons supposé connu p, ce qui n'est pas le cas en pratique.

4.2. Erreur

Définition 16. — . L'erreur carrée moyenne de l'estimateur $\bar{X}_n$ lorsque p est le vrai paramètre est

$$R_2(n,p) = E_p[(\bar{X}_n - p)^2]. \tag{4.2.1}$$

Si $\{X_1, \ldots, X_n\}$ sont des variables indépendantes de Bernoulli de paramètre p, alors

$$R_2(n,p) = E_p[\left(\frac{X_1 + \cdots + X_n}{n} - p\right)^2] = \frac{1}{n}E_p[(X_1 - p)^2] = \frac{p(1-p)}{n}. \tag{4.2.2}$$

L'indice p dans la probabilité est là pour préciser que les $\{X_n\}$ sont des variables des Bernoulli de paramètre p. Notez que pour $0 \leq p \leq 1$, on a $p(1-p) \leq 1/4$ et donc que $R_2(n,p) \leq 1/4n$. En général, lorsque les $\{X_i\}$ ont la même loi et sont indépendantes, et qu'on prend $\bar{X}_n$ comme estimateur de p, alors $R_2(n,p)$ décroit avec n : le calcul (4.2.2) donne que $R_2(n,p) = var(X_1)/n$.

Supposez que l'usine affirme que $p = p_0$, et s'engage à payer une pénalité si jamais l'évaluation de $\bar{X}_n$ sur une échantillon de grande taille donne une valeur qui diffère de p_0 par plus de 0,05. Dans ce cas l'usine a pris un risque, car il se peut que p_0 soit la vraie proportion de postes déféctueux mais que par malchance on soit tombé sur un échantillon où l'on ait une plus grande proportion de postes déféctueux. Le risque est

$$\alpha = P_{p_0}(\omega : |\bar{X}_n(\omega) - p_0| > 0,05). \tag{4.2.3}$$

4.3. Estimateur du maximum de vraisemblace (e.m.v)

4.3.1. Lancers de pièces. — Dans un jeu de pile ou face, on nous donne un pièce de monnaie sans nous dire si elle est bien équilibrée ou pas. On va donc d'abord estimer $p_0 = P($ la pièce tombe sur pile $)$? On lance la pièce dix fois, et l'issue observée est notée $\omega_0 = (p,f,f,p,f,f,p,p,f,f)$. Il est assez naturel de supposer que $p_0 \approx 4/10$. Une base mathématique pour ce choix vient du modèle suivant : on suppose que les lancers sont indépendants de paramètre p pour l'instant inconnu. La probabilité que notre issue se réalise est

$$P_p(\omega_0) = p^4(1-p)^6. \tag{4.3.1}$$

On choisit le paramètre p (la seule inconnue du problème) qui donne à cette issue la plus grande probabilité. On pense donc à $P_p(\omega_0)$ comme à une fonction de $p \mapsto P_p(\omega_0)$, et souvenez-vous qu'une fonction $x \mapsto f(x)$ est maximale en un point x_0, si sa dérivée $f'(x_0) = 0$ et si $f'(x) \geq 0$ dans un voisinnage à gauche de x_0, alors que $f'(x) \leq 0$ dans un voisinnage à droite de x_0 : c'est-à-dire que la fonction croit avant x_0 et decroit après x_0 : il suffit de faire un dessin pour le voir. Ainsi, on calcule la dérivée de $p \mapsto P_p(\omega_0)$, et recherche les points p qui annulent cette dérivée et correspondent à un maximum

$$P_p(\omega_0)' = p^3(1-p)^5\,(4(1-p) - 6p) = 0 \Longrightarrow p^* = \frac{4}{10}. \tag{4.3.2}$$

Notez que les points $p = 0$ et $p = 1$ annulent la dérivée, mais correspondent à $P_p(\omega_0) = 0$ donc à des minima de la fonction. Le paramètre trouvé p^* correspond bien à notre intuition.

4.3.2. Principe de l'e.m.v. — On fait un modèle d'une expérience aléatoire dont la loi dépend de certains paramètres : disons $(p_1, p_2, ..., p_n)$. Le but est d'estimer ces paramètres à l'aide des résultats de l'expérience. On observe un certain évènement A, et on suppose que pour chaque choix de paramètres $(p_1, p_2, ..., p_n)$, on sache calculer $P_{(p_1,p_2,...,p_n)}(A)$. Les estimateurs du maximum de vraisemblace sont des nombres $(p_1^*, p_2^*, ..., p_n^*)$ qui maximise $P_{(p_1,p_2,...,p_n)}(A)$.

$$P_{(p_1^*,p_2^*,...)}(A) = \max P_{(p_1,p_2,...)}(A) \tag{4.3.3}$$

4.3.3. Exemples. — On suit l'évolution d'une équipe de football. On modélise le nombre de buts marqués pendant l'année i par une variable de Poisson X_i de paramètre inconnu λ. On suppose que les $\{X_i\}$ sont indépendants. On observe sur N années les scores suivant $(n_1, n_2, \ldots, n_N)$. Quel est l'e.m.v. de λ ?

Intuitivement, un estimateur “naturel” de la moyenne du nombre de buts marqués $(E[X] = \lambda)$ sera

$$\bar{n}_N = \frac{n_1 + \cdots + n_N}{N}.$$

En suivant le principe de la section précédente, on doit maximiser pour $\lambda > 0$,

$$P_\lambda(X_1 = n_1, X_2 = n_2, \ldots, X_N = n_N) = (e^{-\lambda}\frac{\lambda^{n_1}}{n_1!}) \ldots (e^{-\lambda}\frac{\lambda^{n_N}}{n_N!}). \tag{4.3.4}$$

Il revient au même de maximiser $f(\lambda) = \log P_\lambda(X_1 = n_1, X_2 = n_2, \ldots, X_N = n_N)$. Donc

$$f'(\lambda^*) = -N + \frac{1}{\lambda^*}\sum_{i=1}^N n_i = 0 \Longrightarrow \lambda^* = \frac{\sum_{i=1}^N n_i}{N} = \bar{n}_N. \tag{4.3.5}$$

4.4. Test d'adéquation du χ^2.

4.4.1. Erreur carrée. — Imaginons que l'on veuille tester si un dé est vraiment équilibré, ou s'il est truqué, auquel cas la probabilité de sortie des faces est $q := (q(1), \ldots, q(6))$, avec q distinct de la loi uniforme, que nous appelons $p := (\frac{1}{6}, \ldots, \frac{1}{6})$, sur $\Omega = \{1, \ldots, 6\}$. Nous faisons l'hypothèse que le dé *est équilibré*, et nous allons tenter de decider si cette hypothèse est vraie ou fausse au vu d'un grand échantillon.

Plus généralement, supposons que notre espace d'état $\Omega_1 = \{1, \ldots, d\}$ contient d couleurs, et soient p et q deux lois de probabilité sur Ω_1. Nous voulons tester laquelle est la bonne loi, en répétant le jeu Ω_1 un grand nombre de fois, et en mesurant la fréquence de sortie des différentes couleurs. Ainsi, on pourra choisir comme espace des n lancers

$$\Omega_n = \{\omega = (\omega_1, \ldots, \omega_n) : \ \forall i = 1, \ldots, n \ \omega_i \in \Omega_1\}.$$

Ainsi, une issue nous renseigne sur le résultat de chaque lancer. Lorsque nous effectuons un grand nombre de lancers, nous considérons les résultats comme des réalisations particulières de n variables indépendantes de même loi inconnue (soit p soit q), disons $X_1(\omega), \ldots, X_n(\omega)$. L'idée naturelle est de mesurer la fréquence de sortie de chaque couleur. On s'intéresse donc à

$$(\frac{N_n^{(1)}(\omega)}{n}, \ldots, \frac{N_n^{(d)}(\omega)}{n}) \quad \text{où} \quad \forall x \in \{1, \ldots, d\}, \quad N_n^{(x)} = 1\{\omega_1 = x\} + \cdots + 1\{\omega_n = x\}.$$

Noter que si la loi des X_i est $(p(1), \ldots, p(d))$, alors $(N_n^{(1)}, \ldots, N_n^{(d)})$ est multinomiale, avec

$$P_H\left(N_n^{(1)} = n_1, \ldots, N_n^{(d)} = n_d\right) = \frac{n!}{(n_1!) \ldots (n_d!)} \prod_{i=1}^{d} p(i)^{n_i}.$$

L'indice H dans P_H signifie que nous nous plaçons dans le cas où p est la loi des X_i. Nous notons l'espérance avec un indice A (A pour *alternative*) dans le cas où q est la loi des X_i. Ainsi,

$$P_A\left(N_n^{(1)} = n_1, \ldots, N_n^{(d)} = n_d\right) = \frac{n!}{(n_1!) \ldots (n_d!)} \prod_{i=1}^{d} q(i)^{n_i}.$$

Pour mesurer l'écart entre les fréquences et p, on introduit

$$Z_n = \sum_{x=1}^{d} \frac{(N_n^{(x)} - np(x))^2}{np(x)}.$$

Calculons la moyenne de Z_n. Notez que pour tout x fixé, $N_n^{(x)}$ suit une loi binomiale de paramètre n et $p(x)$ (si l'hypothèse est vraie). Ainsi

$$E_H[Z_n] = \sum_{x=1}^{d} \frac{1}{np(x)} E[(N_n^{(x)} - np(x))^2] = \sum_{x=1}^{d} \frac{var(N_n^{(x)})}{np(x)} = \sum_{x=1}^{d} (1 - p(x)) = d - 1.$$

Nous avons utilisé que $N_n^{(x)}$ est une loi binomiale de variance $np(x)(1 - p(x))$.

Faisons le calcul de $E_A[Z_n]$ dans le cas l'hypothèse est fausse, c'est-à-dire que les X_i suivent plutôt la loi $(q(1), \ldots, q(d))$.

$$N_n^{(x)} - np(x) = N_n^{(x)} - nq(x) + n(q(x) - p(x)).$$

En élevant au carré, et en prenant l'espérance

$$\begin{aligned} E_A[(N_n^{(x)} - np(x))^2] =& E_A[(N_n^{(x)} - nq(x))^2] \\ &+ 2n(q(x) - p(x))E_A[(N_n^{(x)} - nq(x))] + n^2(q(x) - p(x))^2. \end{aligned}$$

Et donc, en utilisant que $E_A[(N_n^{(x)} - nq(x))] = 0$, on obtient

$$E_A[(N_n^{(x)} - np(x))^2] = E_A[(N_n^{(x)} - nq(x))^2] + n^2(q(x) - p(x))^2.$$

Ainsi

$$E_A[(N_n^{(x)} - np(x))^2] \geq n^2[q(x) - p(x)]^2.$$

et donc

$$E_A[Z_n] \geq n \sum_{x \in \Omega} \frac{[q(x) - p(x)]^2}{p(x)}.$$

Ainsi, si l'alternative est vraie, et n assez grand, on devrait avoir $E_A[Z_n] \gg d - 1$.

4.4.2. Test du χ^2 (ki-deux).— Le test d'adéquation de la loi $p(.)$, appelé test du χ^2 de Pearson, consiste au vu de l'echantillon $X_1, \ldots, X_n$, à rejeter ou à accepter l'hypothèse que p est la loi commune des X_n, selon que $Z_n > a$ ou $Z_n \leq a$, où a est un seuil qu'il faudra determiner à l'aide d'une table : la table du χ^2. Le seuil a est calculé typiquement pour que $P_H(Z_n > a) = 0,05$.

En fait, la probabilité que $P_H(Z_n > a)$ converge vers une loi qui ne depend pas de $p(.)$. On appelle cette loi, la loi du χ^2 (ki-2) de degré $d - 1$.

4.4.3. Lecture de la table du χ^2. — Si Z est une variable qui suit une loi du χ^2 de degré n, et si l'on choisit un nombre α parmi ceux qui figurent sur la première ligne, alors cette table donne le z tel que $\alpha = P(\omega : Z(\omega) > z)$. Par exemple, si $n = 18$ et $\alpha = 0,80$, alors, on lit sur la table $z = 12,9$.

Noter que $P(\omega : Z(\omega) \leq 0) = 0$ et $P(\omega : Z(\omega) \geq 0) = 1$. Ainsi, pour $z > 0$

$$P(\omega : Z(\omega) > z) + P(\omega : 0 \leq Z(\omega) \leq z) = 1.$$

Notez que si l'on vous demandait de trouver t_1 et t_2 tels que $P(\omega : t_1 \leq Z(\omega) \leq t_2) = 0,7$ par exemple, alors on aurait plusieurs choix possibles. Par exemple, on peut prendre t_1 et t_2 tels que

$$P(\omega : Z(\omega) \leq t_1) = 0,1 \quad \text{et} \quad P(\omega : Z(\omega) \geq t_2) = 0,2,$$

ou encore

$$P(\omega : Z(\omega) \leq t_1) = 0,2 \quad \text{et} \quad P(\omega : Z(\omega) \geq t_2) = 0,1,$$

ainsi qu'une infinité d'autres choix possibles. (verifiez que ces deux choix donnent bien $P(\omega : t_1 \leq Z(\omega) \leq t_2) = 0,7$ quand-même). C'est cette ambiguité qui fait que nous avons décidé plus haut que lorsqu'on cherche t_1 et t_2 tels que $P(\omega : t_1 \leq Z(\omega) \leq t_2) = 0,7$, on choisit

$$P(\omega : Z(\omega) \leq t_1) = 0,3/2 \quad \text{et} \quad P(\omega : Z(\omega) \geq t_2) = 0,3/2.$$

CHAPITRE 5

FONCTIONS GÉNÉRATRICES

5.1. Définitions et Propriétés

Nous introduisons dans ce cours un outil très utile : la fonction génératrice d'une variable aléatoire. Cette fonction contiendra tout ce qui est intéressant de savoir sur une variable aléatoire. L'inconvénient est que cette notion requière quelque base d'analyse.

Ainsi, on aura besoin tout au long de ce cours de la fonction puissance. On fixe un réel t avec $1 \geq t \geq 0$, et on se rappellera quelques propriétés de la fonction $n \mapsto t^n$ définit pour $n \in \mathbb{N}$: c'est une fonction décroissante (pour $t \geq 0$), et si a, b sont des entiers

$$t^0 = 1, \qquad t^{a+b} = t^a t^b, \qquad \text{et pour } |t| < 1, \quad \lim_{n\to\infty} t^n = 0.$$

Définition 17. — . Soit X une variable aléatoire prenant un nombre fini $\{0, 1, 2 \cdots, N\}$ de valeurs dans $\mathbb{N}$. La **fonction génératrice** de X est la fonction $G_X : [0,1] \to \mathbb{R}$ définie par :

$$G_X(t) = E[t^X] = \sum_{k=0}^{N} t^k P(X = k)\,.$$

G_X est donc un polynôme. On notera que

$$G_X(1) = 1\,, \text{ et } G_X(0) = P(X = 0)\,.$$

L'importance de la fonction génératrice vient de la proposition suivante :

Proposition 12. — *. La fonction génératrice d'une variable aléatoire X à valeurs dans un sous-ensemble fini, E inclus dans $\mathbb{N}$, caractérise sa loi. Autrement dit, si X et Y sont deux variables $X, Y : \Omega \to E$, qui vérifient*

$$G_X = G_Y \Leftrightarrow \forall k \in E,\ P(X = k) = P(Y = k)\,.$$

Démonstration. — Si X et Y ont même loi, il est clair sur la définition de la fonction génératrice que X et Y ont même fonction génératrice.

Réciproquement, G_X et G_Y sont deux polynômes dont les coefficients d'ordre k sont respectivement $P(X = k)$ et $P(Y = k)$. Ainsi, si $G_X = G_Y$, en identifiant les coefficients des polynômes, on obtient que pour tout k, $P(X = k) = P(Y = k)$. ∎

Remarque 6. — . Si $X : \Omega \to \{1, \ldots, N\}$, on obtient, en dérivant G_X,

$$G'_X(t) = \sum_{k=1}^{N} k t^{k-1} P(X = k)\,,$$

et

$$G'_X(0) = P(X = 1)\,, \quad G'_X(1) = E[X]\,.$$

De la même façon, en dérivant k fois, on obtient

$$\forall k \in \mathbb{N} \quad P(X = k) = \frac{G_X^{(k)}(0)}{k!}\,.$$

Cette expression montre qu'il suffit de connaitre G_X au voisinage de 0 pour caractériser la loi de X.

Remarque 7. — . On a défini la fonction génératrice pour des variables prenant un nombre fini de valeurs entières. On peut étendre cette définition à des variables prenant un nombre infini (dénombrable) de valeurs entières (voir plus loin l'exemple de la loi de Poisson). On pose alors

$$G_X(t) = \sum_{k=0}^{+\infty} t^k P(X = k)\,.$$

Cette expression est bien définie, au moins pour $t \in [0; 1]$, puisque $\sum_{k=0}^{\infty} P(X = k) = 1$. La proposition 12 reste valable dans ce cas.

La raison fondamentale pour laquelle les fonctions génératrices sont tellement utiles est la suivante.

Lemme 2. — . *Si X et Y sont indépendantes, alors $G_{X+Y}(t) = G_X(t)G_Y(t)$.*

Démonstration. —

$$G_{X+Y}(t) = E[t^X t^Y] = E[t^X]E[t^Y] = G_X(t)G_Y(t).$$

∎

Exemple 15. — . Si X est une variable de Bernoulli de paramètre α

$$P(X = 1) = \alpha \quad \text{et} \quad P(X = 0) = 1 - \alpha,$$

alors,

$$G_X(t) = t^0 P(\{\omega : \ X(\omega) = 0\}) + t^1 P(\{\omega : \ X(\omega) = 1\}) = (1 - \alpha) + \alpha t.$$

Exemple 16. — . Si S_n est une variable binomiale ($S_n \sim B(n, \alpha)$), (pensez-y comme à la somme de n variables indépendantes de Bernoulli $X_1, \ldots, X_n$), alors l'indépendance des $\{X_i, i = 1, \ldots, n\}$, donne l'indépendance de $\left\{t_i^X, i = 1, \ldots, n\right\}$ et

$$G_{S_n}(t) = E[t^{X_1+\cdots+X_n}] = E[t^{X_1} t^{X_2} \ldots t^{X_n}] = E[t^{X_1}] \ldots E[t^{X_n}] = ((1-\alpha) + \alpha t)^n.$$

Exemple 17. — . Si $X \sim \mathcal{P}(\lambda)$ est une variable de Poisson ($P(X = k) = \exp(-\lambda)\lambda^k/k!$), alors

$$G_X(t) = \sum_{k=0}^{\infty} t^k P(\{\omega : \ X(\omega) = k\}) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} t^k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(t\lambda)^k}{k!} = e^{\lambda(t-1)}.$$

5.1.1. Exemple. — Une poule pond N oeufs. On suppose que N est une variable de Poisson $\mathcal{P}(\lambda)$. Chaque oeuf, indépendamment de tout le reste, éclot avec probabilité p. Soit K le nombre de poussins. Quelle est la loi de K ?

Pour répondre de façon rigoureuse à cette question, il faut d'abord définir Ω. Une issue nous informera sur le nombre d'oeufs et l'état de chaque oeuf. Intuitivement, on voudrait écrire $\omega = (n, o_1, \ldots, o_n)$ où n est un entier correspondant au nombre d'oeufs pondus et $o_i \in \{poussin,\ oeuf\}$ pour $i = 1, \ldots, n$. En pratique, il est malaisé de traiter avec des issues dont la taille est variable : $(1, poussin)$ et $(3, poussin, oeuf, poussin)$. En première lecture, vous pouvez directement aller à la définition de la probabilité de pondre n_0 oeufs de type $(o_1, o_2, \ldots, o_{n_0})$ qui est donné dans l'equation (5.1.1). Une façon (artificielle) très commode est d'introduire l'espace des états d'une suite infinie d'oeufs pondus ayant chacun une valeur dans $E = \{\text{poussin,oeuf}\}$

$$\tilde{\Omega} = \{\tilde{\omega} = \{\tilde{\omega}_i,\ i \in \mathbb{N}\},\ \tilde{\omega}_i \in E\}.$$

C'est un peu comme si on se plaçait dans la situation où notre poule pouvait à priori pondre un nombre infini d'oeufs ! Puis, on introduit une probabilité $\tilde{P}$ sur $\tilde{\Omega}$, telle que pour chaque entier k et chaque choix $o_1, \ldots, o_k \in E$

$$\tilde{P}(\{\tilde{\omega} :\ \tilde{\omega}_1 = o_1, \ldots, \tilde{\omega}_k = o_k\}) = p^{S_k}(1-p)^{k-S_k},$$

où S_k est le nombre de poussins parmi $o_1, o_2, \ldots, o_k$. Ceci correspondrait à la probabilité d'avoir $(o_1, o_2, \ldots, o_k)$ sachant que notre poule a pondu k oeufs. Finalement, on écrit notre vrai espace en tenant compte du nombre d'oeufs pondus

$$\Omega = \{\omega = (n, \tilde{\omega}),\ n \in \mathbb{N}, \quad \tilde{\omega} \in \tilde{\Omega}\}.$$

La probabilité associée à l'issue (n_0, ω^o) est

$$P((n_0, \omega^o)) = e^{-\lambda}(\frac{(\lambda)^{n_0}}{n_0!})\tilde{P}(\{\tilde{\omega} :\ \tilde{\omega}_1 = \omega_1^o, \ldots, \tilde{\omega}_k = \omega_k^o\}). \tag{5.1.1}$$

On définit sur l'espace Ω, la variable de Poisson $N(n, \tilde{\omega}) = n$ et les variables de Bernoulli

$$\forall i \in \mathbb{N}, \quad X_i(n, \tilde{\omega}) = 1 \text{ si } \tilde{\omega}_i = poussin, \quad X_i(n, \tilde{\omega}) = 0 \text{ sinon.}$$

(Vérifier que N et les $\{X_n,\ n \in \mathbb{N}\}$ sont indépendants.) Ainsi, si $N(\omega) = 0$ alors $K(\omega) = 0$, par contre, si $N(\omega) = n$ alors $K(\omega) = X_1(\omega) + \cdots + X_n(\omega)$. On écrira donc

$$K(\omega) = X_1(\omega) + \cdots + X_{N(\omega)}(\omega).$$

La difficulté de cette expression est que l'on somme un nombre aléatoire de X_i. Il est plus simple de l'exprimer comme

$$K(\omega) = 0.1_{\{N=0\}}(\omega) + \sum_{k=1}^{\infty} 1_{\{N=k\}}(\omega)\left(X_1(\omega) + \cdots + X_k(\omega)\right).$$

Calculons $G_K(t)$. Notons d'abord que

$$t^K = 1_{\{N=0\}} t^0 + \sum_{k=1}^{\infty} 1_{\{N=k\}} t^{X_1 + \cdots + X_k}.$$

Ainsi, en utilisant que l'espérance d'une somme de termes est la somme des espérances de ces termes (ce qui marche même pour une somme infinie de termes positifs)

$$G_K(t) = E[t^K] = E[1_{\{N=0\}}] + E\left[\sum_{k=1}^{\infty} 1_{\{N=k\}} t^{X_1+\cdots+X_k}\right]$$

$$= P(N=0) + \sum_{k=1}^{\infty} E\left[1_{\{N=k\}} t^{X_1+\cdots+X_k}\right] = P(N=0) + \sum_{k=1}^{\infty} P(N=k)\left(E[t^{X_1}]\right)^k$$

$$= \sum_{k=0}^{\infty} P(N=k)(G_X(t))^k.$$

On a utilisé la convention que pour tout nombre positif x, on a $x^0 = 1$. Finalement,

$$G_K(t) = G_N(G_X(t)).$$

Après ce long calcul, utilisons l'expression de G_X et G_N que nous avions obtenu respectivement dans l'exemple 1 et 3.

$$G_X(t) = 1 - p + pt, \quad \text{et} \quad G_N(y) = \exp(\lambda(y-1)).$$

Finalement,

$$G_K(t) = G_N(G_X(t)) = \exp(\lambda(1 - p + pt - 1)) = \exp(p\lambda(t-1)).$$

Maintenant, le nombre moyen d'oeufs s'obtient en dérivant une fois $G_K(t)$ par rapport à t

$$E[K] = G'_K(0) = G'_N(G_X(0).G'_X(0) = \lambda p,$$

d'après les exemples 15 et 17. De la même façon, on montre le théorème suivant.

Théorème 1. — *. Soient $\{X_n,\ n \geq 1\}$ une suite de variables entières indépendantes de même loi, et N une variable à valeurs entières indépendante des $\{X_n,\ n \geq 1\}$, alors*

$$S(\omega) = \begin{cases} 0, \text{ si } N(\omega) = 0 \\ X_1(\omega) + \cdots + X_{N(\omega)}(\omega), \text{ si } N(\omega) > 0 \end{cases}$$

a pour fonction génératrice $G_S(t) = G_N(G_X(t))$.

Voici un excercice que je conseille vivement de chercher au moins 30 mn, avant de lire la solution.

5.1.1.0.1. Exercice Corrigé :— Une pièce de monnaie est telle que *pile* sorte avec probabilité α. Cette pièce est lancée un nombre aléatoire de fois : N fois où N est une variable de Poisson de paramètre λ. Ceci veut dire que la probabilité qu'elle soit lancée 10 fois par exemple est

$$P(\omega : N(\omega) = 10) = e^{-\lambda}\frac{\lambda^{10}}{10!}.$$

Appelons S le nombre de *pile* obtenu, et T le nombre de *face* obtenu.

1. Quelle est la loi de S et T ?
2. Quelle est la fonction génératrice de S ?

Solution : Le gros problème vient du fait que le nombre de lancers est aléatoire. En effet, supposons qu'on sache qu'on ai lancé la pièce 10 fois. Quelle serait la loi de S ? Pour cela, on va associer au i-ème lancer un variable de Bernoulli X_i qui vaudra 1 si c'est *pile* qui sort (au i-ème lancer), et $X_i = 0$ si c'est *face* qui sort. Sachant que $N = 10$ on ecrirait

$$S = X_1 + X_2 + \cdots + X_{10}.$$

D'après le cours II, on aurait que S est une variable binomiale $b(10; \alpha)$: Ainsi pour $k = 0, \ldots, 10$

$$P(S = k \text{ sachant que } N = 10) = C_{10}^k \alpha^k (1-\alpha)^{10-k}.$$

Comment faire si N est variable ? On peut dire la chose toute simple suivante : si $N = 0$ alors $S = 0$, si $N = 1$ alors $S = X_1$, si $N = 2$ alors $S = X_1 + X_2$, si $N = 3$ alors $S = X_1 + X_2 + X_3$, et ainsi de suite... Un façon concise de dire cela est d'utiliser les fonctions indicatrices :

$$S = 1_{N=0}.0 + 1_{N=1}X_1 + 1_{N=2}(X_1 + X_2) + 1_{N=3}(X_1 + X_2 + X_3) + \ldots$$

$$\cdots + 1_{N=1000}(X_1 + X_2 + \cdots + X_{999} + X_{1000}) + \ldots \tag{5.1.2}$$

ou encore si l'on veut décrire l'évènement $\{S = k\}$, on dira qu'il est égal à $\{0 = k\}$ si $N = 0$, à $\{X_1 = k\}$ si $N = 1$, à $\{X_1 + X_2 = k\}$ si $N = 2$, etc... ou encore de façon plus concise

(5.1.3)

$$\{S = k\} = \{0 = k \text{ et } N = 0\} \cup \{X_1 = k \text{ et } N = 1\} \cup \{X_1 + X_2 = k \text{ et } N = 2\} \cup \ldots$$

Nous avons donc décomposé $\{S = k\}$ comme une union d'ensembles disjoints. Notez que si l'on veut voir k *piles*, il faut au moins k lancers, donc on pourrait faire commencer l'union dans (5.1.3) à $N = k$. Ainsi, losqu'on passe au probabilité, et qu'on utilise la formule de définition des probabilités conditionnelles $P(A \cap B) = P(A|B)P(B)$, on obtient

$$\begin{aligned} P(\{S = k\}) &= P(\{0 = k \text{ et } N = 0\}) + P(\{X_1 = k \text{ et } N = 1\}) \\ &\quad + P(\{X_1 + X_2 = k \text{ et } N = 2\}) + \ldots \\ &= \sum_{i=k}^{\infty} P(\{X_1 + X_2 + \cdots + X_i = k \text{ et } N = i\}) \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=k}^{\infty} P(\{X_1 + X_2 + \cdots + X_i = k | \{N = i\})P(\{N = i\}) \\
&= \sum_{i=k}^{\infty} \left(C_i^k \alpha^k (1-\alpha)^{i-k}\right) e^{-\lambda} \frac{\lambda^i}{i!}.
\end{aligned}$$

Il faut maintenant essayer de simplifier cette horrible formule :

$$C_i^k \alpha^k (1-\alpha)^{i-k} e^{-\lambda} \frac{(\lambda)^i}{i!} = e^{-\lambda} \frac{(\lambda\alpha)^k}{k!} \frac{(\lambda(1-\alpha))^{i-k}}{(i-k)!}.$$

Ainsi, en séparant ce qui dépend de k seulement et ce qui dépend de i

$$\begin{aligned}
P(\{S = k\}) &= e^{-\lambda} \frac{(\lambda\alpha)^k}{k!} \sum_{i=k}^{\infty} \frac{(\lambda(1-\alpha))^{i-k}}{(i-k)!} = e^{-\lambda} \frac{(\lambda\alpha)^k}{k!} \sum_{i=k}^{\infty} \frac{(\lambda(1-\alpha))^{i-k}}{(i-k)!} \\
&= e^{-\lambda} \frac{(\lambda\alpha)^k}{k!} \sum_{i=0}^{\infty} \frac{(\lambda(1-\alpha))^{i}}{i!} = e^{-\lambda} \frac{(\lambda\alpha)^k}{k!} e^{\lambda(1-\alpha)} = e^{-\lambda\alpha} \frac{(\lambda\alpha)^k}{k!}.
\end{aligned}$$

On reconnait donc une loi exponentielle de paramètre $\lambda\alpha$. Cela à l'air très compliqué. Essayons de voir si le calcul de la fonction génératrice ne nous aurait pas simplifié la vie. A partir de (5.1.2), on a

$$t^S = 1_{N=0} t^0 + 1_{N=1} t^{X_1} + 1_{N=2} t^{(X_1+X_2)} + 1_{N=3} t^{(X_1+X_2+X_3)} + \cdots + \tag{5.1.4}$$

Pour écrire (5.1.4), il suffit de vous demander : que vaut t^S lorsque $N = 0$, que vaut t^S lorsque $N = 1$, et ainsi de suite... C'est en faisant cela que l'on devrait commencer à comprendre la commodité des fonctions indicatrices. Prenons l'espérance de t^S.

$$E[t^S] = \sum_{k=0}^{\infty} E[1_{N=k} t^{(X_1+\cdots+X_k)}].$$

On utilise maintenant que N et les X_i sont indépendants, donc

$$\begin{aligned}
E[1_{N=k} t^{(X_1+\cdots+X_k)}] &= E[1_{N=k}] E[t^{(X_1+\cdots+X_k)}] \\
&= P(N = k) E[t^{X_1} \dots t^{X_k}] = P(N = k) E[t^{X_1}] E[t^{X_2}] \dots E[t^{X_k}] \\
&= P(N = k) \left(E[t^{X_1}]\right)^k.
\end{aligned}$$

Posons $\zeta = E[t^{X_1}]$. On a donc

$$E[t^S] = \sum_{k=0}^{\infty} P(N = k) \zeta^k = E[\zeta^N].$$

Maintenant, on a vu dans l'exemple (15) que $\zeta = E[t^{X_1}] = 1 - \alpha + t\alpha$, et dans l'exemple (17) que $E[\zeta^N] = \exp(\lambda(\zeta - 1))$. Ce qui donne au total

$$G_S(t) = E[t^S] = e^{\lambda(1-\alpha+t\alpha-1)} = e^{\lambda\alpha(t-1)}.$$

Il est facile de prendre la dérivée n-ème de $G_S(t)$ et d'obtenir

$$P(\omega : \ S(\omega) = n) = \frac{1}{n!} \frac{d^n G_S(t)}{dt^n}|_{t=0} = e^{-\lambda\alpha} \frac{(\lambda\alpha)^n}{n!}.$$

Ceci permet de conclure que S est une variable de Poisson de paramètre $\lambda\alpha$.

Exemple 18. — . Soient X et Y deux variables indépendantes de Poisson de paramètres λ et μ, respectivement. Quelle est la distribution de $X+Y$? Calculons la fonction génératrice de $X+Y$. Par le lemme 2,

$$G_{X+Y}(t) = G_X(t)G_Y(t) = \exp(\lambda(t-1))\exp(\mu(t-1)) = \exp((\lambda+\mu)(t-1)).$$

On reconnait donc la fonction génératrice d'une variable de Poisson de paramètre $\lambda+\mu$ et l'argument précédent nous permet de conclure que $X+Y$ est bien une variable de Poisson de paramètre $\lambda+\mu$.

Exemple 19. — . Dans l'exemple de la poule et des poussins éclos K, on a vu que

$$G_K(t) = \exp(p\lambda(t-1)).$$

Ce qui veut dire, par la proposition 12, que K est une variable de Poisson de paramètre $p\lambda$.

5.2. Évolution des Populations

5.2.1. Motivation. — Nous allons décrire un modèle d'évolution d'une population, et montrer en quoi les fonctions génératrices sont utiles. La génération 0 est composée d'un individu. Cet individu fait $\eta_1^{(1)}$ enfants ; $\eta_1^{(1)}$ est une variable aléatoire, dont la loi est appelée *loi de la progéniture*

$$P(\eta_1^{(1)} = k) = p_k, \quad \forall k \in \mathbb{N}.$$

Ces enfants forment la génération 1. Chacun d'eux, indépendamment du reste, fait un nombre aléatoire d'enfants : $\eta_1^{(2)}$ est le nombre d'enfants du 1er, $\eta_2^{(2)}$ est le nombre d'enfants du 2ème, etc...L'indice supérieur 2 rappelle que ces enfants forment la génération 2. Et ainsi de suite...
Soit Z_n le nombre d'individus à la génération n. Pour obtenir Z_{n+1}, il faut additionner le nombre d'enfants de chaque individu de la génération n : si $Z_n = 0$ alors $Z_{n+1} = 0$, sinon

$$Z_{n+1} = \eta_1^{(n+1)} + \eta_2^{(n+1)} + \cdots + \eta_{Z_n}^{(n+1)}.$$

La façon la plus simple de comprendre cette somme avec un indice aléatoire est de la *décomposer* sur toutes les valeurs possibles de Z_n

$$Z_{n+1} = 1_{\{Z_n=1\}}(\eta_1^{(n+1)}) + \cdots + 1_{\{Z_n=k\}}(\eta_1^{(n+1)} + \cdots + \eta_k^{(n+1)}) + \ldots$$

Quelle est la probabilité que la population finisse par s'éteindre ? Notez que cet évènement ne dépend pas de l'état d'un nombre fini de générations. L'espèce s'éteind signifie qu'il existe une génération n où $Z_n = 0$. Ainsi "extinction"$= \{\omega : \exists n \in \mathbb{N}$ tel que $Z_n(\omega) = 0\}$. En d'autres termes

$$\text{“extinction”} \quad = \bigcup_{n=0}^{\infty} \{\omega : \ Z_n(\omega) = 0\}.$$

Le symbole ∞ qui apparait au-dessus de $\cup$ est une source de confusion : il ne signifie rien d'autre que

$$\bigcup_{n=0}^{\infty} A_n = \{x : \ \exists n \in I\!N \text{ tel que } x \in A_n\}.$$

Par contre nous n'avons pas vu comment traiter $P(\bigcup_{n=0}^{\infty} A_n)$. Nous commençons donc par cela.

5.2.2. Résultat Préliminaire. — Jusqu'à présent, on n'a a priori défini que des probabilités sur des ensembles Ω de cardinal fini, même si dans les exemples on a utilisé des probabilités sur $I\!N$. En fait, la définition d'une probabilité s'étend facilement au cas d'un ensemble dénombrable.

Définition 18. — . Soit Ω un ensemble dénombrable, et $\mathcal{F} = \mathcal{P}(\Omega)$ l'ensemble des parties de Ω. Une probabilité P sur Ω est une fonction de $\mathcal{F}$ dans $[0,1]$ telle que
- $P(\Omega) = 1$;
- $P(A^c) = 1 - P(A)$.
- Si $(A_i)_{i\in I}$ est une famille dénombrable d'évènements disjoints, (i.e. $A_i \cap A_j = \emptyset$ si $i \neq j$), alors
$$P(\cup_{i\in I} A_i) = \sum_{i\in I} P(A_i)\,.$$

De cette définition, découle la propriété suivante :

Proposition 13. — . *Si $(A_n)_{n\geq 1}$ est une suite croissante d'évènements (i.e $A_n \subset A_{n+1}$), alors $P(\bigcup_{n\geq 1} A_n) = \lim_{n\to\infty} P(A_n)$.*

Démonstration. — On pose $B_1 := A_1$ et pour $n \geq 2$, $B_n := A_n \backslash A_{n-1}$. Les évènements B_n sont deux à deux disjoints. De plus, pour tout $n \geq 1$, $\bigcup_{k=1}^n B_k = \bigcup_{k=1}^n A_k = A_n$, et $\bigcup_{n\geq 1} B_n = \bigcup_{n\geq 1} A_n$. Par conséquent,

$$\begin{aligned} P(\bigcup_{n\geq 1} A_n) &= P(\bigcup_{n\geq 1} B_n) = \sum_{n\geq 1} P(B_n) = \lim_{n\to\infty} \sum_{k=1}^{n} P(B_k) = \lim_{n\to\infty} P(\bigcup_{k=1}^{n} B_k) \\ &= \lim_{n\to\infty} P(A_n)\,. \end{aligned}$$

∎

Nous aurons besoin de quelques propriétés générales de $G_{Z_1}(t)$, que nous notons $g_1(t)$ par simplicité.

$$g_1(t) = P(\eta_1^{(1)} = 0) + \sum_{n=1}^{\infty} t^n P(\eta_1^{(1)} = n), \quad g_1(0) = P(\eta = 0), \quad \text{et} \quad g_1(1) = 1.$$

On admettra que G_{Z_1} est continue et deux fois dérivable sur $[0,1]$.

$$g_1'(t) = \sum_{n=1}^{\infty} n t^{n-1} P(\eta_1^{(1)} = n) \Longrightarrow E[\eta_1^{(1)}] = g_1'(1).$$

Aussi, g_1' est croissante car sa dérivée est positive

$$g_1''(t) = \sum_{n=2}^{\infty} n(n-1) t^{n-2} P(\eta_1^{(1)} = n) \geq 0.$$

Donc, en résumé, on retiendra que G_{Z_1} est convexe (c'est-à-dire que sa dérivée seconde est positive) et croissante avec

$$G_{Z_1}(0) = P(\eta_1^{(1)} = 0), \quad G_{Z_1}(1) = 1, \quad G'_{Z_1}(1) = E[\eta_1^{(1)}].$$

5.2.3. Probabilité d'extinction. —

Proposition 14. — *. Soit la suite de variables entières $\{Z_n,\ n \geq 0\}$ définie plus haut. Alors la probabilité que la population finisse par s'éteindre est le plus petit réel positif, disons $\chi > 0$, tel que $G_{Z_1}(\chi) = \chi$.*

Démonstration. — Nous exprimons l'évènement A :="la population finit par s'éteindre" en terme des $\{Z_n\}$.

$$A = \bigcup_{n=1}^{\infty} \{Z_n = 0\}.$$

Comme $\{Z_n = 0\} \subset \{Z_{n+1} = 0\}$, on peut appliquer la proposition 13

$$P(A) = \lim_{n\to\infty} P(\{Z_n = 0\}) = \lim_{n\to\infty} G_{Z_n}(0) \leq 1.$$

Cette limite existe bien et nous l'appelons χ. Nous avons donc besoin de calculer $G_{Z_n}(0)$. Pour $n > 1$, on utilise le fait que Z_n et les $\{\eta_k^{(n+1)}, k = 1, \ldots\}$ sont indépendants.

$$G_{Z_{n+1}}(z) = \sum_{k=1}^{\infty} E[1_{\{Z_n=k\}} z^{\eta_1^{(n+1)} + \cdots + \eta_k^{(n+1)}}] = \sum_{k=1}^{\infty} P(Z_n = k)(E[z^{\eta_1^{(1)}}])^k = G_{Z_n}(G_{Z_1}(z)).$$

Donc, si on pose $g_1(z) := G_{Z_1}(z)$, et $g_n(z) := G_{Z_n}(z)$, alors

$$g_{n+1}(z) = g_n(g_1(z)) \Longrightarrow g_n(z) = g_1^{(n)}(z) := g_1(g_1(g_1(\ldots \text{n fois } \ldots g_1(z)\ldots)))).$$

Notons que g_1 est continue sur $[0,1]$, et donc

$$\lim_{n\to\infty} g_n(0) = \chi \Longrightarrow \lim_{n\to\infty} g_1(g_n(0)) = g_1(\chi).$$

Or, comme $g_1(g_n) = g_{n+1}$, on a que $g_1(\chi) = \chi$. Il nous reste à voir que χ est le plus petit réel positif vérifiant $g_1(\chi) = \chi$. Nous allons raisonner par l'absurde : Soit $\zeta \in [0,1]$ tel que $g_1(\zeta) = \zeta$. Montrons que $\chi \leq \zeta$. Comme $\zeta \geq 0$, et g_1 est croissante, $\zeta = g_1(\zeta) \geq g_1(0)$, et en prenant g_1 de chaque coté $\zeta \geq g_1^{(2)}(0)$. En répétant le procédé, on a que $\zeta \geq g_1^{(n)}(0)$ et donc en passant à la limite $\zeta \geq \chi$. ∎

5.2.4. Exemple. — Supposons que la loi de la progéniture soit une loi géométrique de paramètre p

$$P(\eta_1^{(1)} = k) = p(1-p)^k, \quad \text{pour} \quad k = 0, 1, \ldots,$$

Donc

$$g_1(z) = \sum_{n=0}^{\infty} z^k P(\eta_1^{(1)} = k) = \frac{p}{1 - z(1-p)}.$$

Donc $g_1(0) = p$ et $g_1'(1) = (1-p)/p$. On cherche la plus petite solution de $g_1(\chi) = \chi$. On résoud

$$\frac{p}{1 - z(1-p)} = z \Longrightarrow 0 = z^2(1-p) - z + p = (1-p)(z-1)(z - \frac{p}{1-p}).$$

Donc si $p/(1-p) < 1$ on a $\chi = p/(1-p)$, sinon $\chi = 1$. Ainsi, la population finit par s'éteindre avec probabilité 1 lorsque $p \leq 1/2$, c'est à dire lorsque le nombre moyen d'enfants est plus petit ou égal à 1.

Exercice : Démontrer que si $E[\eta] > 1$, alors $\chi < 1$. Puis, que si $E[\eta] < 1$, alors $\chi = 1$.

CHAPITRE 6

VARIABLES CONTINUES

6.1. Motivation

Les variables aléatoires que nous avons considérées jusqu'à présent, appelées *variables discrètes*, sont insuffisantes pour décrire un jeu de fléchettes. Ce jeu est le suivant : on jette *uniformement au hasard* une fléchette sur une cible, et on s'intéresse à la position de cette fléchette. Pour plus de simplicité, imaginons que la cible soit une carré de côté 1, et $\Omega = \{\omega = (x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$ représente les positions possibles de cette fléchette. Notez que, dans ce jeu, la fléchette finit toujours dans la cible. La question est d'associer une probabilité à chaque issue de Ω. Jetter *uniformement au hasard* signifie que chaque point à la même chance d'être touché. Imaginons un quadrillage de ce carré en n^2 petits carrés disjoints de côté $1/n$, et appelons ces petits carrés $Q_1^{(n)}, Q_2^{(n)}, \ldots$. L'évènement $\left\{\dagger \in Q_1^{(n)}\right\}$ signifie que la fléchette tombe dans le carré $Q_1^{(n)}$, et jetter *uniformement au hasard* signifie que pour tout $i = 1, \ldots, n^2$

$$P(\left\{\dagger \in Q_i^{(n)}\right\}) = P(\left\{\dagger \in Q_1^{(n)}\right\}).$$

Ainsi, comme les petits carrés forment une partition de Ω

$$1 = P(\Omega) = P(\bigcup_{i=1}^{n^2} \left\{\dagger \in Q_i^{(n)}\right\}) = \sum_{i=1}^{n^2} P(\left\{\dagger \in Q_i^{(n)}\right\}) = n^2 P(\left\{\dagger \in Q_1^{(n)}\right\}).$$

Ainsi, la probabilité que la fléchette tombe dans un carré de superficie $1/n^2$ est exactement $1/n^2$.

Soit A une région faite d'une union de petits carrés $\{Q_i^{(n)}, i \in I\}$, où I est un sous-ensemble de $\{1, \ldots, n^2\}$, et si $|A|$ désigne la superficie de A, alors

$$|A| = \sum_{i \in I} |Q_i^{(n)}|, \quad \text{et} \quad P\left(\{\dagger \in A\}\right) = P\left(\bigcup_{i \in I} \left\{\dagger \in Q_i^{(n)}\right\}\right) = \sum_{i \in I} P\left(\left\{\dagger \in Q_i^{(n)}\right\}\right) = |A|.$$

On obtient par un tel procédé que pour une région A assez générale, $P(\{\dagger \in A\}) = |A|$, ce que nous admettrons. En général, pour une fléchette lancée dans une cible Ω de surface totale $|\Omega|$, nous dirons que la fléchette est lancée uniformément au hasard si pour toute région $A \subset \Omega$ (pour laquelle on peut calculer la surface $|A|$) on a

$$P(\{\dagger \in A\}) = \frac{|A|}{|\Omega|}.$$

Notez que $(0,0) \in Q_1^{(n)} = [0, \frac{1}{n}] \times [0, \frac{1}{n}]$, et donc pour tout n

$$P(\{\dagger = (0,0)\}) \leq P\left(\left\{\dagger \in Q_1^{(n)}\right\}\right) = \frac{1}{n^2}.$$

Ceci entraine que $P(\dagger = (0,0)) = 0$. On interprète cela en disant que la probabilité que la fléchette tombe exactement sur $(0,0)$ est nulle, et ce sera vrai pour tous les points de Ω. On associera une probabilité positive aux régions dont la surface est non-nulle.

On retient donc que la grande nouveauté liée au jeu de fléchette est que l'on ne peut caractériser un probabilité par ses seules valeurs sur les issues de Ω (puisqu'ici $P(\omega) = 0$ pour tout $\omega \in \Omega$), mais que l'on doit considérer une probabilité comme une fonction sur une famille de sous-ensembles de Ω : dans cet exemple, nous pourrions choisir pour n entier

$$\mathcal{F}_n := \{\bigcup_{i \in I} Q_i^{(n)}, \quad \text{pour tout } I \subset \{1, \ldots, n^2\}\}.$$

$\mathcal{F}_n$ est la collection des évènements concernant la position de la fléchette *à une resolution de* $1/n$ *près*. Noter que $\mathcal{F}_n$ n'est pas la collection de toutes les parties de Ω comme c'etait le cas jusqu'à présent.

Imaginons qu'on s'intéresse à l'absisse de la fléchette : si $\omega = (x, y)$, on définit la variable aléatoire $X(\omega) = x$. Les valeurs possibles de X sont les réels de $[0, 1]$. La loi de X pose problème : pour tout $a \in [0, 1]$

$$P(\omega : X(\omega) = a) = |\{(x, y) : x = a\}| = 0,$$

car la surface de la section $\{(x, y) : x = a\}$ est nulle. Par contre, pour $0 \leq a < b \leq 1$

$$P(\omega : X(\omega) \in [a, b[) = |\{(x, y) : a \leq x < b\}| = b - a.$$

Nous avons donc besoin de concepts nouveaux car la loi de la variable n'est pas adéquate.

6.2. Définitions

Définition 19. — . On définit une probabilité sur une collection d'évènements $\mathcal{F}$, nommée *tribu*, qui vérifie les propriétés suivantes :

1. $\mathcal{F} \subset \mathcal{P}(\Omega)$.
2. $\Omega \in \mathcal{F}$.
3. Si $A \in \mathcal{F}$, alors le complementaire de A, noté $A^c \in \mathcal{F}$.
4. Si $A_1, A_2, \ldots$ est une suite infinie d'éléments de $\mathcal{F}$, alors $A_1 \cup A_2 \cup \ldots := \{\omega : \exists n \in \mathbb{N},\ \omega \in A_n\}$ est aussi dans $\mathcal{F}$.

Une probabilité est alors une fonction de $\mathcal{F}$ dans $[0,1]$ qui vérifie

1. $P(\emptyset) = 0$ et $P(\Omega) = 1$.
2. $P(A^c) = 1 - P(A)$.
3. Pour une suite infinie $A_1, A_2, \ldots$ dans $\mathcal{F}$ et disjoints,

$$P(\bigcup_{n\geq 1} A_n) = \lim_{n\to\infty} (P(A_1) + \cdots + P(A_n)) = \sum_{n\geq 1} P(A_n).$$

Remarque 8. — . Notez que dans l'exemple des fléchettes, nous n'aurions pas pu choisir comme *tribu*

$$\mathcal{F} = \bigcup_{n\geq 1} \mathcal{F}_n := \{A : \ \exists n \in I\!N \text{ tel que } \ A \in \mathcal{F}_n\}.$$

En effet, nous aurions eu alors que $\cap Q_1^{(n)} = (0,0)$ appartient à $\mathcal{F}$ (expliquez pourquoi à l'aide des points 2,et 3 de la définition d'une tribu), mais ce point n'appartenant à aucun des $\mathcal{F}_n$, il n'est pas dans leur union.

Définition 20. — . Une variable aléatoire continue est une fonction $X : \Omega \to \mathbb{R}$ qui a la propriété suivante : pour tout $x \in \mathbb{R}$

$$P(\omega : X(\omega) = x) = 0.$$

Elle est caractérisée par une fonction $f_X : \mathbb{R} \to \mathbb{R}$, sa *densité*, de la façon suivante. Pour deux réels $a < b$

$$P(\omega : a < X(\omega) < b) = \int_a^b f_X(x)dx. \qquad (6.2.1)$$

Cette densité a deux propriétés importantes que nous démontrons plus bas :
- $f_X(x) \geq 0$.
- $\int_{-\infty}^{\infty} f_X(x)dx = 1$.

On fera en outre l'hypothèse que f_X est continue par morceaux et qu'il y a un nombre fini de points de discontinuité. Ce qui veut dire qu'il y a un certain nombre (fini) de points $x_1, \ldots, x_n$ en dehors desquels f_X est continue et en chaque point x_i on a que f est continue à droite de x_i ou à gauche de x_i.

Dans l'exemple où X est l'absisse de la fléchette, notez que

$$\int_a^b f_X(u)du = b - a, \quad \text{et donc} \quad \int_0^x f_X(u)du = x,$$

pour tout $x \in [0,1]$, et donc $f_X(u) = 1$ si $u \in [0,1]$ et f_X est nulle en dehors de $[0,1]$.

6.2.1. Propriétés de la densité. — Notez d'abord que $\{\omega : a < X(\omega) < b\} \cup \{\omega : X(\omega) = a\} = \{\omega : a \leq X(\omega) < b\}$ et donc

$$\begin{aligned} P(\{\omega : a \leq X(\omega) < b\}) &= P(\{\omega : a < X(\omega) < b\}) + P(\{\omega : X(\omega) = a\}) \\ &= P(\{\omega : a < X(\omega) < b\}). \end{aligned}$$

Ainsi les évènements suivants ont tous la même probabilité $\int_a^b f_X(u)du$.

$$\{\omega : a < X(\omega) < b\}, \ \{\omega : a \leq X(\omega) < b\}, \ \{\omega : a < X(\omega) \leq b\}, \ \{\omega : a \leq X(\omega) \leq b\}.$$

Aussi, on peut faire tendre b vers l'infini dans (6.2.1) pour obtenir

$$P(\omega : a < X(\omega)) = P(\omega : a < X(\omega) < \infty) = \int_a^\infty f_X(x)dx. \tag{6.2.2}$$

Il faut comprendre que comme $X(\omega)$ n'a que des valeurs réelles, vérifier $a < X(\omega)$ est équivalent à vérifier $a < X(\omega) < \infty$, et donc les probabilités de ces deux évènements sont égales. De même, en faisant tendre a vers $-\infty$ on obtient

$$P(\omega : X(\omega) < b) = P(\omega : -\infty < X(\omega) < b) = \int_{-\infty}^b f_X(x)dx.$$

Aussi, si $a_1 < b_1 < \cdots < a_n < b_n$ sont des nombres réels. On peut calculer la probabilité de $\{\omega : X(\omega) \in]a_i, b_i[, i = 1, 2, \ldots, n\}$ par la formule

$$P(\{\omega : X(\omega) \in \bigcup_{i=1}^n]a_i, b_i[\}) = \sum_{i=1}^n P(\{\omega : X(\omega) \in]a_i, b_i[\}) = \sum_{i=1}^n \int_{a_i}^{b_i} f_X(x)dx.$$

Cette formule est justifiée car les n évènements $\{\{\omega : X(\omega) \in]a_i, b_i[\},\ i = 1, \ldots, n\}$ sont disjoints. Il suffit donc de savoir intégrer f. Faisons l'hypothèse que f est continue à droite : s'il y avait un point x_0 tel que $f_X(x_0) < 0$, on aurait par continuité que f est négative sur un petit intervalle, disons $[x_0, x_0 + \epsilon[$ pour ϵ petit, et la relation (6.2.1) entrainerait une contradiction :

$$0 \leq P(\omega : X(\omega) \in [x_0, x_0 + \epsilon[) = \int_{x_0}^{x_0+\epsilon} f_X(x)dx < 0.$$

Ainsi, une densité est positive. Une autre relation très importante est

$$\int_{-\infty}^{\infty} f_X(x)dx = 1. \tag{6.2.3}$$

Cela vient du fait que pour tout ω, $-\infty < X(\omega) < \infty$, et donc $\Omega = \{\omega : -\infty < X(\omega) < \infty\}$ et $P(\Omega) = 1$.

Il est très utile de visualiser l'intégrale d'une densité : cela vous aidera à lire les tables. Dessinez donc deux axes perpendiculaires : l'axe des x horizontal, et l'axe des y vertical, se recontrant au point $(x, y) = (0, 0)$. Dessinez une densité $x \mapsto y = f(x)$ positive, donc au-dessus de l'axe des x, et qui tend vers 0 lorsque $|x|$ tend vers l'infini. l'intégrale de f entre a et b représente la surface délimitée sur les cotés par les droites verticales $x = a$ et $x = b$, en haut par la courbe $y = f(x)$ et en bas par l'axe des x. Colorez cette région dont la surface est $\int_a^b f(x)dx$. Dessinez maintenant une densité f qui soit paire, c'est-à-dire que $f(x) = f(-x)$ ou encore qui soit symétrique par rapport à l'axe des y. Des exemples de fonctions paires sont $1/(1 + x^2)$, $\exp(-|x|)$ ou encore $\exp(-x^2)$. Sur le dessin, observez que $\int_{-b}^{-a} f(x)dx = \int_a^b f(x)dx$.

Exemple 20. — . Soit f la densité d'une variable X, avec

$$\forall x \in [-1, 1] \quad f(x) = cx^2, \qquad \text{et} \quad \forall x \notin [-1, 1], \quad f(x) = 0.$$

Trouver c ? On doit donc vérifier (6.2.3).

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-1}^{1} cx^2 dx = 2c/3 = 1$$

Ainsi $c = 3/2$. Maintenant que f est complètement déterminée, on peut calculer la probabilité de tout évènement du genre $\{\omega : X(\omega) \in]a, b[\}$, par exemple que $X > 1/2$

$$P(\omega : X(\omega) > 1/2) = \int_{1/2}^{\infty} f(x)dx = 3/2 \int_{1/2}^{1} x^2 dx = 7/16.$$

Exemple 21. — . Soit la densité $f(x) = 1$ pour $x \in [0, 1[$ et $f(x) = 0$ si $x \notin [0, 1[$. La variable associée à f est dite variable uniforme sur $[0, 1]$. On vérifira que (6.2.3) est vraie. Quelle est la probabilité que $X \in [3/4, 5]$?

$$P(\omega : 3/4 \leq X(\omega) \leq 5) = \int_{3/4}^{1} f(x)dx + \int_{1}^{5} 0dx = 1/4.$$

Exemple 22. — . Soit λ un réel positif et soit la densité $f(x) = \lambda \exp(-\lambda x)$ pour $x \geq 0$ et $f(x) = 0$ pour $x < 0$. Vérifions d'abord (6.2.3)

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{\infty} \lambda \exp(-\lambda x)dx = \big[-\exp(-\lambda x)\big]_0^{\infty} = 1.$$

On appelle la variable aléatoire associée à $f(x) = \lambda \exp(-\lambda x)$, une variable exponentielle de paramètre λ. On peut, par exemple, calculer la probabilité que $X > s$, pour $s > 0$

$$P(\omega : X(\omega) > s) = \int_{s}^{\infty} \lambda \exp(-\lambda x)dx = \exp(-\lambda s).$$

Exemple 23. — . Soit X une variable exponentielle de paramètre λ, sachant que $X > s$, quelle est la probabilité que $X > t + s$, pour t et s positifs ? Il s'agit de faire un calcul d'espérance conditionnelle : il suffit donc de connaitre $P(\{\omega : X(\omega) > s\})$ et $P(\{\omega : X(\omega) > s\} \cap \{\omega : X(\omega) > t + s\})$ et d'utiliser la définition

$$P(\{\omega : X(\omega) > t + s\}|\{\omega : X(\omega) > s\}) = \frac{P(\{\omega : X(\omega) > s\} \cap \{\omega : X(\omega) > t + s\})}{P(\{\omega : X(\omega) > s\})}.$$

Notons que si ω est telle que $X(\omega) > t + s$, alors $X(\omega) > s$, et donc $\{\omega : X(\omega) > t + s\} \subset \{\omega : X(\omega) > s\}$. On obtient donc

$$\begin{aligned} P(\{\omega : X(\omega) > t + s\}|\{\omega : X(\omega) > s\}) &= \frac{P(\{\omega : X(\omega) > t + s\})}{P(\{\omega : X(\omega) > s\})} \\ &= \frac{\exp(-\lambda(s + t))}{\exp(-\lambda s)} = e^{-\lambda t}. \end{aligned} \tag{6.2.4}$$

Les variables exponentielles jouent un rôle fondamentale en modélisation, en raison de la propriété (6.2.4). En effet, X modélise un temps d'apparition d'un phénomène imprévisible tel l'emission de particules radioactives. L'égalité (6.2.4) siginifie que l'observation du système pendant la période $[0, t]$ n'apporte aucune information sur ce que le sysème va faire pendant l'intervalle successif $[t, t + s]$.

Exemple 24. — . Si f est la densité de X, quelle est la densité de $2X+1$? Posons $Y = 2X+1$, et cherchons une fonction positive g telle que pour tous $a < b$

$$P(\{\omega : a < Y(\omega) < b\}) = \int_a^b g(x)dx.$$

Il suffit de réécrire Y en fonction de X : $\{\omega : a < Y(\omega) < b\} = \{\omega : (a-1)/2 < X(\omega) < (b-1)/2\}$, et ainsi

$$\int_a^b g(x)dx = P(\{\omega : (a-1)/2 < X(\omega) < (b-1)/2\}) = \int_{(a-1)/2}^{(b-1)/2} f(x)dx.$$

On veut retrouver dans cette dernière intégrale, les mêmes bornes d'intégration a et b que pour g. On fait donc un changement de variable $y = 2x+1$ (ou encore $x = (y-1)/2$) de façon à ce que quand $x = (a-1)/2$ alors $y = a$, et lorsque $x = (b-1)/2$ alors $y = b$. Ainsi

$$\int_{(a-1)/2}^{(b-1)/2} f(x)dx = \int_a^b f(\frac{y-1}{2})\frac{1}{2}dy.$$

Ainsi, $g(y) = f((y-1)/2)/2$.

Exemple 25. — . On s'intéresse maintenant à une fléchette lancée sur une cible triangulaire. Ainsi, $\Omega = \{(x,y) : -2 < x < 2, 0 < y < 2 - |x|\}$, $\mathcal{F}$ est la collection des région de Ω pour lesquels on peut calculer la surface, et nous posons que la probabilité d'une région $A \in \mathcal{F}$ est proportionnelle à sa surface de façon à ce que $P(\Omega) = 1$:

$$P(A) = \frac{|A \cap \Omega|}{|\Omega|}, \quad \text{notez qu'ici } |\Omega| = 4).$$

On considère comme variable aléatoire l'ordonnée de la fléchette : $X((x,y)) = y$ et la question est de trouver la loi de X ? D'abord les valeurs possibles de X sont les réels entre 0 et 2, et pour tout $b \in [0,2]$ et toute distance infinitésimale db

$$P(\{(x,y) : X((x,y)) \in [b, b+db[\}) \sim \frac{2(2-b)\ db}{4} = f_Y(b)db.$$

Ainsi la densité est $f_Y(b) = (1 - b/2)$ pour $b \in [0,2]$. On pourra vérifier que f_Y est bien une densité.

Exemple 26. — . On s'intéresse au même jeu que dans l'exemple 25, sauf que l'on va prendre en compte que la fléchette tombe une fois sur deux en dehors de la cible : on considère $\Omega' = \Omega \cup \{\delta\}$ où l'issue δ symbolise que l'on est en dehors de la cible. Les probabilités ne sont plus les mêmes : pour une région A de Ω on a

$$P(A) = \frac{1}{2}\frac{|A|}{|\Omega|}, \quad \text{alors que} \quad P(\{\delta\}) = \frac{1}{2}.$$

On s'intéresse à la variable $\tilde{X}$ qui vaut X sur Ω et $\tilde{X}(\delta) = 0$. Notez que $\tilde{X}$ est bien une variable aléatoire, mais n'est pas une variable continue car $P(\omega :\ X(\omega) = 0) = 1/2$.

6.2.2. Indépendance. —

Définition 21. — . Deux variables continues X et Y sont indépendantes si pour tous $a < b$ et $c < d$ réels, on a

$$\begin{aligned}[P(\{\omega : a < X(\omega) < b\} &\cap \{\omega : c < Y(\omega) < d\}) \\ &= P(\{\omega : a < X(\omega) < b\})P(\{\omega : c < Y(\omega) < d\}).\end{aligned} \tag{6.2.5}$$

Remarque 9. — . Il est facile de voir qu'il suffit en fait de vérifier que pour tout réels a et c

$$P(\{\omega : a < X(\omega)\} \cap \{\omega : c < Y(\omega)\}) = P(\{\omega : a < X(\omega)\})P(\{\omega : c < Y(\omega)\}). \tag{6.2.6}$$

En effet, notez que comme X est continue, on a

$$P(\{\omega : b < X(\omega)\}) = P(\{\omega : b \leq X(\omega)\}),$$

et que l'on a

$$\{a < X < b\} = \{a < X\} \cap \{X < b\} = \{a < X\} \backslash \{X \leq b\},$$

et donc

$$P(\{\omega : a < X(\omega) < b\}) = P(\{\omega : a < X(\omega)\}) - P(\{\omega : b \leq X(\omega)\}).$$

Puis il faut transformer $\{a < X < b\} \cap \{c < Y < d\}$ en deux étapes :

$$\begin{aligned}\{a < X < b\} \cap & \{c < Y < d\} \\ & = (\{a < X\} \cap \{c < Y < d\}) \setminus (\{X \leq b\} \cap \{c < Y < d\}),\end{aligned}$$

ce qui implique que

$$\begin{aligned}P\big(\{a < X < b\} &\cap \{c < Y < d\}\big) \\ &= P(\{a < X\} \cap \{c < Y < d\}) - P(\{X \leq b\} \cap \{c < Y < d\}),\end{aligned}$$

puis l'on fait les mêmes manipulations pour transformer $\{c < Y < d\}$ en deux évènements $\{c < Y\}$ et $\{d \leq Y\}$. Il est alors facile de voir que (6.2.6) implique (6.2.5).

6.2.3. Espérance et Variance. —

6.2.3.1. Définition. — Soit X une v.a. et f_X sa densité. Soit φ une somme de fonctions indicatrices :

$$\varphi = \sum_{i=1}^{n} \alpha_i 1\{[a_i, b_i[\}, \quad \text{avec } n \text{ réels non-nuls et distincts} \quad \alpha_1, \ldots, \alpha_n.$$

Notez que $\varphi(X)$ est une variable discrète. En effet $\varphi(X)$ a comme valeurs possibles $\{0, \alpha_1, \ldots, \alpha_n\}$. On a donc

$$\begin{aligned}E[\varphi(X)] &= \sum_{i=1}^{n} \alpha_i P(\omega : \ \varphi(X(\omega)) = \alpha_i) = \sum_{i=1}^{n} \alpha_i P(\omega : \ X(\omega) \in [a_i, b_i[) \\ &= \sum_{i=1}^{n} \alpha_i \int_{a_i}^{b_i} f_X(x)dx = \int_{-\infty}^{\infty} \varphi(x) f_X(x)dx.\end{aligned}$$

Ce cas particulier motive la définition suivante.

Définition 22. — . Pour toute fonction φ pour laquelle on peut calculer $\int \varphi(x) f_X(x) dx$, on pose

$$E[\varphi(X)] = \int_{-\infty}^{\infty} \varphi(x) f_X(x) dx. \tag{6.2.7}$$

Ainsi, on a

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx, \quad \text{et} \quad E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx.$$

6.2.3.2. Propriétés. — Les propriétés de l'espérance que nous avions montrées dans le cas de variables à valeurs finies, sont vraies pour des variables continues. Nous en énonçons quelques unes

Proposition 15. — *. Soit X et Y deux variables aléatoires et $\lambda \in \mathbb{R}$. Alors,*

1. $E[\lambda X] = \lambda E[X]$.
2. $E[X+Y] = E[X] + E[Y]$.
3. *Si X et Y sont indépendantes, alors $E[XY] = E[X]E[Y]$.*
4. *Si X et Y sont indépendantes, alors $var[X+Y] = var[X] + var[Y]$.*

Nous présentons une formule très utile pour calculer l'espérance d'une variable positive X. Notez que pour tout t, si $\varphi(x) = 1\{x > t\}$, alors

$$E[\varphi(X)] = P(\omega : \ X(\omega) > t) = \int_0^{\infty} f_X(x) 1\{x > t\} dx.$$

Maintenant, notez que pour tout x positif

$$x = \int_0^x dt = \int_0^{\infty} 1\{x > t\} dt.$$

Nous appliquons cette égalité aux x à l'intérieur de l'intégrale, et échangeons deux intégrales positives : vous n'avez pas encore vu ce Théorème, et nous l'admettrons. Nous donnons dans la section suivante une autre démonstration de cette propriété.

$$\begin{aligned} E[X] &= \int_0^{\infty} x f_X(x) dx = \int_0^{\infty} \left(\int_0^{\infty} 1\{x > t\} dt \right) f_X(x) dx \\ &= \int_0^{\infty} \left(\int_0^{\infty} f_X(x) 1\{x > t\} dx \right) dt = \int_0^{\infty} P(\omega : \ X(\omega) > t) dt. \end{aligned} \tag{6.2.8}$$

Nous généralisons cette représentation à des fonctions φ telle que

$$\varphi(x) = \int_0^x \psi(t) dt = \int_0^{\infty} 1\{x > t\} \psi(t) dt, \quad \text{avec} \quad \psi(t) \geq 0.$$

Par défintion, et en utilisant ici aussi un échange d'intégrales positives

$$\begin{aligned} E[\varphi(X)] &= \int_0^{\infty} \varphi(x) f_X(x) dx = \int_0^{\infty} \left(\int_0^{\infty} 1\{x > t\} \psi(t) dt \right) f_X(x) dx \\ &= \int_0^{\infty} \left(\int_0^{\infty} 1\{x > t\} f_X(x) dx \right) \psi(t) dt \end{aligned}$$

$$= \int_0^{\infty} P(\omega : \ X(\omega) > t)\psi(t) \ dt. \tag{6.2.9}$$

Une application possible de cette formule est pour $\varphi(x) = x^2$ dont la fonction $\psi(x)$ correspondante est $2x$. Ainsi, pour une Y une variable aléatoire quelconque :

$$E[Y^2] = \int_0^{\infty} 2tP(\omega : \ |Y|(\omega) > t) \ dt.$$

6.2.3.3. Preuve de (6.2.8) et (6.2.9). — Commençons par donner des conditions équivalentes pour que $E[X] < \infty$, pour une variable positive X. En effet, toutes les variables continues ne sont pas d'espérance finie : considérez $f_X(x) = \frac{1}{x^2}$ pour $x \geq 1$, et $f_X(x) = 0$ pour $x < 0$. Alors f_X est bien une densité, par contre

$$E[X] = \int_1^{\infty} x f_X(x) dx = \int_1^{\infty} \frac{1}{x} dx = \infty.$$

Lemme 3. — *. Considérons $X \geq 0$. $E[X] < \infty$ est équivalent à ce que la série de terme général $a_n = nP(\omega : n \leq X(\omega) < n+1)$ converge. De plus, si $E[X] < \infty$, alors*

$$\lim_{n \to \infty} nP(\omega : n \leq X(\omega)) = 0. \tag{6.2.10}$$

Démonstration. — Notez que

$$\sum_{n=0}^{\infty} n 1_{\{n \leq X(\omega) < n+1\}} \leq X(\omega) \leq \sum_{n=0}^{\infty} (n+1) 1_{\{n \leq X(\omega) < n+1\}}. \tag{6.2.11}$$

Posons, $A_n = \{\omega : \ n \leq X(\omega) < n+1\}$, et prenons l'espérance de chaque terme dans (6.2.11)

$$\sum_{n=0}^{\infty} nP(A_n) \leq E[X] \leq \sum_{n=0}^{\infty} (n+1)P(A_n) \tag{6.2.12}$$

Notez aussi que $P(A_0) + P(A_1) + \cdots = 1$, et donc (6.2.12) devient

$$\sum_{n=0}^{\infty} a_n \leq E[X] \leq 1 + \sum_{n=0}^{\infty} a_n, \quad \text{où} \quad a_n = nP(A_n). \tag{6.2.13}$$

Le comportement de $E[X]$ est donc le même que celui de la série $\sum a_n$. Maintenant, si $E[X] < \infty$, alors le reste de la série tend vers 0, or

$$\sum_{k \geq n} a_k \geq n \sum_{k \geq n} P(A_k) = nP(\omega : \ X(\omega) \geq n). \tag{6.2.14}$$

Ainsi (6.2.10) suit de (6.2.14). ∎

Notons maintenant que pour tout $x > 0$

$$\int_x^{\infty} f_X(u) du = P(\omega : \ X(\omega) > x), \tag{6.2.15}$$

ce qui signifie que $P(\omega:\ X(\omega)>x)$ a pour dérivée $-f_X(x)$. Faisons un intégration par parties

$$\begin{aligned}\int_0^n x f_X(x)dx &= [-xP(\omega:\ X(\omega)>x)]_0^n + \int_0^n P(\omega:\ X(\omega)>x)dx \\ &= -nP(\omega:\ X(\omega)>n) + \int_0^n P(\omega:\ X(\omega)>x)dx. \qquad (6.2.16)\end{aligned}$$

En utilisant le lemme 3, et n qui tend vers l'infini, on obtient (6.2.8). Pour montrer (6.2.9) on observe que φ est positive, et croissante, et donc

$$\sum_{n=0}^{\infty}\varphi(n)1\{n\le X(\omega)<n+1\}\le\varphi(X(\omega))\le\sum_{n=0}^{\infty}\varphi(n+1)1\{n\le X(\omega)<n+1\}.$$

En suivant la même démonstration que celle du lemma 3, on obtient que $E[\varphi(X)]<\infty$ entraine que $\varphi(n)P(\omega:\ X(\omega)>n)$ tend vers 0, et l'on peut faire une intégration par parties

$$\int_0^n\varphi(x)f_X(x)dx=-\varphi(n)P(\omega:\ X(\omega)>n)+\int_0^n\varphi'(x)P(\omega:\ X(\omega)>x)dx. \qquad (6.2.17)$$

6.2.4. Loi d'une somme de variables continues indépendantes. —

Lemme 4. — *. Soient X et Y des variables continues indépendantes.*

$$P(X+Y\le t)=\int_{-\infty}^{\infty}f_X(x)P(Y\le t-x)dx=\int_{-\infty}^{\infty}f_Y(x)P(X\le t-x)dx. \qquad (6.2.18)$$

Démonstration. — La preuve rigoureuse est hors programme. Voici donc une façon intuitive de retrouver ce résultat. Nous traitons la première égalité.

$$\begin{aligned}P(X+Y\le t) &= \int_{-\infty}^{\infty}P(X\in[x,x+dx[\cap X+Y\le t) \\ &= \int_{-\infty}^{\infty}P(X\in[x,x+dx[\cap Y\le t-x) \\ &= \int_{-\infty}^{\infty}P(X\in[x,x+dx[)\ P(Y\le t-x) \\ &= \int_{-\infty}^{\infty}f_X(x)P(Y\le t-x)dx. \qquad (6.2.19)\end{aligned}$$

∎

6.3. Variable Normale

Il faudra avoir compris, après la lecture de cette section, comment lire les tables des variables normales qui accompagnent ce cours. Une fois qu'on a compris comment ça marche, cette lecture de table est un jeu d'enfant, dans lequel on ne peut pas se tromper.

6.3.1. Définition. — La densité d'une variable normale standard (appellée aussi variable Gaussienne ou de Gauss) est

$$f_X(x) = \exp(-x^2/2)/\sqrt{2\pi}. \tag{6.3.1}$$

La constante $\sqrt{2\pi}$ est telle que $\int f_X(x)dx = 1$. On notera souvent une variable normale standard Z. Soit $a_1 < b_1 < \cdots < a_n < b_n$ des nombres réels. On calcule la probabilité que $\{\omega : Z(\omega) \in]a_i, b_i[, i = 1, 2, \ldots, n\}$ par la formule

$$P(\{\omega : Z(\omega) \in \bigcup_{i=1}^{n}]a_i, b_i[\}) = \sum_{i=1}^{n} \int_{a_i}^{b_i} f_X(x)dx \quad \text{avec} \quad f_X(x) = \exp(-x^2/2)/\sqrt{2\pi}.$$

On pourra trouver la valeur des probabilités de $]a, b[$ à l'aide de la table 1.

6.3.2. Lecture de la table 1. — Prenons la table directe (table de gauche) et supposons qu'on veuille $P(\omega : Z(\omega) < 0,95)$ pour $Z \sim N(0,1)$? On écrit $x = 0,95 = 0,9 + 0,05$, et la deuxième décimale se trouve sur la ligne 1, et l'unité ainsi que la première décimale se trouvent sur la colonne 1. On cherche donc sur la table l'intersection de la colonne en dessous de 0,05 et de la ligne à droite de 0,9 et on trouve $P(\omega : Z(\omega) < 0,95) = 0,82894$.

Noter que cette table ne donne que les valeurs de x positives. Si $x < 0$, ce qu'on peut toujours écrire comme $x = -|x|$, alors par parité de f

$$P(\{\omega : Z(\omega) < -|x|\}) = \int_{-\infty}^{-|x|} f(y)dy = \int_{|x|}^{\infty} f(y)dy$$
$$= \int_{-\infty}^{\infty} f(y)dy - \int_{-\infty}^{|x|} f(y)dy = 1 - P(\{\omega : Z(\omega) < |x|\}).$$

Puis, on lit x sur la table directe. Pour calculer $P(\{\omega : Z(\omega) < -2,47\})$, on note que $|x| = 2,47 = 2,4 + 0,07$, et avec la table $P(\{\omega : Z(\omega) < -2,47\}) = 1 - 0,99324 = 0,00676$.

Pour la table inverse, on choisit une probabilité p, et dans la table on lit z tel que $P(\{\omega : |Z(\omega)| < z\}) = p$. Par exemple pour $p = 0,83 = 0,8 + 0,03$, on cherche donc sur la table l'intersection de la colonne en dessous de 0,03 et de la ligne à droite de 0,8 et on trouve $P(\omega : Z(\omega) < 0,215) = 0,83$.

6.3.3. Propriétés des variables normales. — Les propriétés simples de f sont : f est continue, dérivable, strictement positive, paire et

$$\lim_{\pm\infty} f(x) = 0, \quad \int_{-\infty}^{\infty} f(x)dx = 1 \quad \forall x,\ f(x) \le f(0) \le 1. \tag{6.3.2}$$

On a aussi, si f' désigne la dérivée, et f'' désigne la dérivée de f'

$$f'(x) = -x.f(x) \quad \text{et} \quad f''(x) = (x^2 - 1)f(x). \tag{6.3.3}$$

Corollaire 3. — *. Soit Z une variable normale standard, alors*

1. $E[Z] = 0$.
2. $E[Z^2] = 1$.

3. $\forall x,\ P(\{\omega : Z(\omega) = x\}) = 0$.
4. $\forall x \geq 0,\ P(\{\omega : Z(\omega) < -x\}) = P(\{\omega : Z(\omega) > x\})$.

Démonstration. — 1.) Si f est une fonction paire ($f(x) = f(-x)$), alors $x \mapsto xf(x)$ est impaire et donc

$$E[Z] = \int_{-\infty}^{\infty} xf(x)dx = 0.$$

2.) Il faut faire une integration par parties. On pose $u' = -x\exp(-x^2/2)$ et $v = x/\sqrt{2\pi}$. Donc

$$E[Z^2] = \int_{-\infty}^{\infty} -u'(x)v(x)dx = -[uv]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} u(x)v'(x)dx = \int_{-\infty}^{\infty} f(x)dx = 1.$$

3.)

$$\{x\} \subset]x - \frac{1}{n}, x + \frac{1}{n}[\quad \text{donc} \quad P(\{\omega : Z(\omega) = x\}) \leq P(\{\omega : Z(\omega) \in]x - \frac{1}{n}, x + \frac{1}{n}[).$$

or, c'est vraie pour tout n et

$$P(\{\omega : Z(\omega) \in]x - \frac{1}{n}, x + \frac{1}{n}[) = \int_{x-1/n}^{x+1/n} f(u)du \leq \frac{2}{n} \longrightarrow 0 \Longrightarrow P(\{x\}) = 0.$$

4.) C'est une conséquence simple de la parité de f. ∎

Soient Z une variable normale standard, et $\mu \in \mathbb{R}$ et $\sigma > 0$, soit $X = \sigma Z + \mu$. On appelle X une variable normale de moyenne μ et variance σ^2, et on note $X \sim N(\mu, \sigma^2)$. En statistique, σ est appelé l'écart-type. On vérifie aisèment que

$$E[X] = \mu \qquad \text{et} \qquad var[X] = \sigma^2.$$

Noter que si $X \sim N(\mu, \sigma^2)$, alors pour λ réel, on a $\lambda X \sim N(\lambda\mu, \lambda^2\sigma^2)$. Nous aurons besoin d'un résultat que nous énonçons sans démonstration :

Théorème 2. — *. Si X et Y sont deux variables normales indépendantes, alors $X + Y$ est aussi une variable normale.*

Soit $X \sim N(\mu_1, \sigma_1^2)$ et $Y \sim N(\mu_2, \sigma_2^2)$, essayons de trouver μ et σ tels que $Z = X + Y \sim N(\mu, \sigma^2)$. Il suffit de calculer la moyenne de Z et sa variance.

$$E[Z] = E[X + Y] = E[X] + E[Y] = \mu_1 + \mu_2 = \mu,$$

et,

$$E[Z^2] - (E[Z])^2 = E[(X + Y)^2] - (\mu_1 + \mu_2)^2 = E[X^2 + 2XY + Y^2] - (\mu_1^2 + 2\mu_1\mu_2 + \mu_2^2).$$

On note que comme X et Y sont indépendants, alors $E[XY] = E[X]E[Y]$ et donc $\sigma^2 = \sigma_1^2 + \sigma_2^2$. On aurait pu utiliser que $var(X + Y) = var(X) + var(Y)$ directement.

Corollaire 4. — *. Soient $\{X_i, i = 1, \ldots, n\}$ des variables $N(\mu, \sigma^2)$ indépendantes, alors*

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n} \sim N(\mu, \frac{\sigma^2}{n}) \tag{6.3.4}$$

Démonstration. — Nous faisons la preuve par récurrence. Pour $n = 2$, $\bar{X}_2 = X_1/2 + X_2/2$, et par le théorème 2 on a que $\bar{X}_2 \sim N(\mu, \sigma^2/2)$. Supposons la propriété (6.3.4) est vraie à l'ordre $n-1$. On écrit

$$\bar{X}_n = \frac{X_1 + \cdots + X_{n-1}}{n} + \frac{X_n}{n} = \frac{n-1}{n}\bar{X}_{n-1} + \frac{X_n}{n},$$

or $(n-1)/n\bar{X}_{n-1}$ est une variable normale par notre hypothèse de recurrence, alors que X_n/n est normale par hypothèse. Donc, par le théorème 2, $\bar{X}_n$ est une variable normale. Il est alors facile de calculer son espérance et sa variance. ∎

On voit donc qu'on peut représenter $\bar{X}_n$ comme

$$\bar{X}_n = \frac{\sigma}{\sqrt{n}} Z + \mu, \quad \text{avec} \quad Z \sim N(0,1) \Longrightarrow Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}\sigma}.$$

CHAPITRE 7

MODES DE CONVERGENCE

7.1. Convergence en loi

Avant de donner la définition de la convergence en loi, nous donnons deux exemples.

7.1.1. Binomiale vers Poisson. — Nous avons déjà rencontré la situation où une suite de variables aléatoires *convergeait* vers une variable aléatoire. Il s'agissait de variables Binomiales $\mathcal{B}(n, \frac{x}{n})$ que nous avions appelées S_n, et qui convergeaient vers une variable de Poisson de paramètre x. Notez que S_n était vu comme une somme de n Bernoulli indépendantes $S_n = X_1^{(n)} + \cdots + X_n^{(n)}$ définies sur $\Omega_n := \{\text{pile, face}\}^n$ et $P_n(X_i^{(n)} = 1) = \frac{x}{n}$. Ainsi chaque S_n était définie sur un espace de probabilité différent (Ω_n, P_n). Par *converger* on signifiait que pour tout $k \in \mathbb{N}$

$$P(S = k) = \lim_{n\to\infty} P(S_n = k) = e^{-x}\frac{x^k}{k!}.$$

Ceci correspond à la convergence des lois des variables S_n vers la loi de Poisson.

7.1.2. Géométrique vers exponentielle. — On rappelle qu'une variable exponentielle est une variable continue et $P(\omega :\ T(\omega) = x) = 0$ pour tout $x \in \mathbb{R}$. On considère donc plutôt sa densité $f_T(x) = \lambda e^{-\lambda x}$ pour $x \geq 0$, et $f_T(x) = 0$ pour $x < 0$. Ainsi, pour $x \geq 0$

$$P(\omega :\ T(\omega) > x) = \int_x^\infty \lambda e^{-\lambda t} dt = e^{-\lambda x}.$$

Soit une collection $\{J_n, n \in \mathbb{N}\}$ de lancers infini de pile et face, tous définis sur le même espace $\Omega = \{\omega = \{\omega_n, n \in \mathbb{N}\} \text{ avec } \omega_n \in \{\text{ pile, face}\}\}$. Au jeu J_n, la probabilité de voir pile est très petite et vaut λ/n. Finalement, T_n correspond au premier instant où pile apparait dans le jeu J_n : T_n est donc une variable géométrique et

$$P_n(T_n > k) = P_n(\omega_1 = \text{face}, \ldots, \omega_k = \text{face}) = (1 - \frac{\lambda}{n})^k.$$

Notez aussi que $E[T_n] = \frac{n}{\lambda}$ (à faire !). Nous montrons que $\frac{T_n}{n}$ *converge* vers T dans le sens suivant : pour tout $x \in \mathbb{R}$

$$\lim_{n\to\infty} P(\omega:\ \frac{T_n(\omega)}{n} > x) = P(\omega:\ T(\omega) > x) = e^{-\lambda x}. \tag{7.1.1}$$

En effet, soit $[nx]$ la partie entière de nx avec

$$nx - 1 < [nx] \le nx, \quad \text{et donc} \quad \lim_{n\to\infty} \frac{[nx]}{n} = x.$$

Notons d'abord que

$$P(\omega:\ \frac{T_n(\omega)}{n} > x) = (1 - \frac{\lambda}{n})^{[nx]}.$$

Puis, nous utilisons un résultat que nous démontrons dans le cours suivant.

Lemme 5. — . *Pour tout y réel, on a*

$$\lim_{n\to\infty} \left(1 + \frac{y}{n}\right)^n = e^y. \tag{7.1.2}$$

Ce résultat n'est pas trivial, car d'un côté $\frac{y}{n}$ tend vers 0, mais d'un autre, on prend une très grande puissance : si $y > 0$, on rajoute un tout petit peu à 1, mais on multiplie ce rajout un grand nombre de fois. Nous utilisons le lemme 5 de la façon suivante :

$$(1 - \frac{\lambda}{n})^{[nx]} = \left[\left(1 - \frac{\lambda}{n}\right)^n\right]^{[nx]/n} = b_n^{a_n}, \tag{7.1.3}$$

avec

$$b_n = \left(1 - \frac{\lambda}{n}\right)^n \longrightarrow b = e^{-\lambda} \quad \text{et} \quad a_n = \frac{[nx]}{n} \longrightarrow x.$$

Ainsi, $b_n^{a_n}$ converge vers $b^x = e^{-\lambda x}$. En effet,

$$b_n^{a_n} - b^x = b_n^{a_n - x} b_n^x - b_n^x + b_n^x - b^x = (b_n^{a_n - x} - 1) b_n^x + (b_n^x - b^x). \tag{7.1.4}$$

Il suffit de voir que chacun des termes converge vers 0 (à faire)

$$\lim_{n\to\infty} (b_n^{a_n - x} - 1) b_n^x = 0 \quad \text{et} \quad \lim_{n\to\infty} (b_n^x - b^x) = 0.$$

Ce dernier exemple motive la définition suivante.

Définition 23. — . On considère un espace de probabilité (Ω, P) et $X : \Omega \to \mathbb{R}$ une variable aléatoire, et pour chaque entier n, un espace (Ω_n, P_n) et $X_n : \Omega_n \to \mathbb{R}$ telles que pour tout $x \in \mathbb{R}$

$$\lim_{n\to\infty} P_n(\omega : X_n(\omega) > x) = P(\omega : X(\omega) > x). \tag{7.1.5}$$

On dit que X_n *converge en loi vers* X.

7.2. Convergence en Probabilité

Commençons par une exemple.

7.2.1. Discrétisation d'une variable continue. — Considérons une fléchette sur une cible carrée $\Omega = [0,1[\times[0,1[$, et P la probabilité qui à chaque région de Ω associe sa surface. Soit la variable $X(\omega) = x$ lorsque $\omega = (x,y)$. On a vu que X était une variable uniforme sur $[0,1[$, c'est-à-dire que pour $0 \leq a < b < 1$

$$P(\omega : \ a \leq X(\omega) < b) = b - a.$$

On découpe $[0,1[\times[0,1[$ en n sections $[\frac{i-1}{n}, \frac{i}{n}[\times[0,1[$, et on pose $X_n(\omega) = \frac{i-1}{n}$ si $\omega \in [\frac{i-1}{n}, \frac{i}{n}[\times[0,1[$ pour $i = 1, \ldots, n$. Notez alors que quelque soit ω, $|X(\omega) - X_n(\omega)| \leq \frac{1}{n}$, et donc pour tout ϵ

$$\lim_{n\to\infty} P(\omega : \ |X(\omega) - X_n(\omega)| > \epsilon) = 0. \tag{7.2.1}$$

Notez ici que **toutes les variables sont définies sur le même espace de probabilité**.

7.2.2. Définition. — L'exemple précédent motive la définition suivante.

Définition 24. — . Soit (Ω, P) un espace de probabilité sur lequel sont définies $X, X_1, \ldots, X_n, \ldots$ On dit que la suite $\{X_n\}$ *converge en probabilité* vers X si pour tout $\epsilon > 0$, on a (7.2.1).

Proposition 16. — . *Soient $X_1, X_2, \ldots$ des variables aléatoires définies sur un même espace de probabilité (Ω, P), et qui convergent en probabilité vers une fonction $X : \Omega \to \mathbb{R} \cup \{\infty\}$. On veut dire par cela que l'on n'exclut pas a priori que X puisse avoir la valeur $+\infty$. Alors,*

$$P(\omega : X(\omega) = \infty) = 0.$$

Démonstration. — Montrons que pour tout k fixé,

$$\lim_{n\to\infty} P(\omega : |X_k(\omega)| > n) = 0.$$

Posons $A_n = \{\omega : |X_k(\omega)| > n\}$. Notons que A_n sont des ensembles décroissants et que

$$\bigcap_{n\geq 1} A_n = \{\omega : |X_k(\omega)| = \infty\}.$$

Par une propriété du cours

$$P(\{\omega : |X_k(\omega)| = \infty\}) = P(\bigcap_{n\geq 1} A_n) = \lim_{n\to\infty} P(A_n).$$

X_k étant une variable aléatoire, on a que $P(\{\omega : |X_k(\omega)| = \infty\}) = 0$, et notre premier point est prouvé.

Montrons maintenant que pour tout n et k,

$$\{\omega : |X(\omega)| > n\} \subset \{\omega : |X(\omega) - X_k(\omega)| \geq 1\} \cup \{\omega : |X_k(\omega)| > n-1\}. \tag{7.2.2}$$

Si $|X(\omega)| > n$, alors soit $|X(\omega) - X_k(\omega)| \geq 1$, soit

$$|X(\omega) - X_k(\omega)| < 1, \quad \text{et} \quad |X(\omega)| > n.$$

Mais dans ce dernier cas, par l'inégalité triangulaire

$$|X_k(\omega)| \geq |X(\omega)| - |X(\omega) - X_k(\omega)| > n - 1,$$

et notre deuxième point est prouvé. En prenant les probabilités de chaque ensemble dans (7.2.2), on obtient

$$P(\{\omega : |X(\omega)| > n\}) \leq P(\{\omega : |X(\omega) - X_k(\omega)| \geq 1\}) + P(\{\omega : |X_k(\omega)| > n - 1\}).$$

On fait tendre n vers l'infini, et on obtient

$$\limsup_{n\to\infty} P(\{\omega : |X(\omega)| > n\}) \leq P(\{\omega : |X(\omega) - X_k(\omega)| \geq 1\}).$$

On fait alors tendre k vers l'infini, pour obtenir notre deuxième point.

On applique le raisonnement du premier point à $\tilde{A}_n = \{\omega : |X(\omega)| > n\}$ pour obtenir

$$P(\omega : X(\omega) = \infty) = P(\bigcap_{n\geq 1} \tilde{A}_n) = \lim_{n\to\infty} P(\tilde{A}_n) = 0.$$

∎

Proposition 17. — . *Soient $X_1, X_2, \ldots$ des variables aléatoires définies sur un même espace de probabilité (Ω, P), et qui convergent en probabilité vers X. Soit $g : \mathbb{R} \to \mathbb{R}$ une fonction continue. Alors, $g(X_n)$ converge vers $g(X)$ en probabilité.*

Démonstration. — Notons que sur un intervalle compact, g est uniformément continue, et donc pour $|X(\omega)| \leq n$, on a que pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que pour tout y vérifiant $|y - X(\omega)| \leq \delta$ et pour $|X(\omega)| \leq n$ on a

$$|g(y) - g(X(\omega))| \leq \epsilon.$$

Ainsi, la contraposée de cette propriété dit qu'il existe $\delta > 0$ tel que

$$|g(y) - g(X(\omega))| \geq \epsilon \text{ et } |X(\omega)| \leq n \Longrightarrow |y - X(\omega)| \geq \delta.$$

Ainsi,

$$\{\omega : |g(X_n(\omega)) - g(X(\omega))| \geq \epsilon \text{et } |X(\omega)| \leq n\} \subset \{\omega : |X_n(\omega) - X(\omega)| \geq \delta\}.$$

ce qui implique

$$P(\omega : |X(\omega)| \leq n \text{ et } |g(X_n(\omega)) - g(X(\omega))| \geq \epsilon) \leq P(\{\omega : |X_n(\omega) - X(\omega)| \geq \delta\}) \longrightarrow 0.$$

Finalement,

$$P(|g(X_n(\omega)) - g(X(\omega))| \geq \epsilon) \leq P(\omega : |X(\omega)| > n) + P(\omega : |X(\omega)| \leq n \text{ et } |g(X_n(\omega)) - g(X(\omega))| \geq \epsilon) \longrightarrow 0.$$

∎

Dans le cas où toutes les variables sont définies sur le même espace, nous aurions pû étudier d'autres modes de convergence :

1. *Convergence en moyenne* : $\lim_{n\to\infty} E[|X_n - X|] = 0$.
2. *Convergence quadratique* : $\lim_{n\to\infty} E[(X_n - X)^2] = 0$.
3. *Convergence ponctuelle* : Pour tout $\omega \in \Omega$, $\lim_{n\to\infty} |X_n(\omega) - X(\omega)| = 0$.

7.3. Liens entre la convergence en loi et en probabilté.

Proposition 18. — *. La convergence en probabilité vers une variable continue, entraine la convergence en loi. En d'autres termes, (7.2.1) est plus fort que (7.1.5).*

Démonstration. — Considérons une suite $\{X_n, n \in I\!N\}$ qui convergence en probabilité vers X, variable continue. Montrons que cela entraine la convergence en loi. Notez que si pour un ω et un x quelconques

$$X_n(\omega) > x, \quad \text{et} \quad |X_n(\omega) - X(\omega)| < \epsilon, \quad \text{alors} \quad X(\omega) > x - \epsilon. \tag{7.3.1}$$

Ceci entraine la seconde inégalité qui suit

$$\begin{aligned} P(\omega: X_n(\omega) > x) &\leq P(\omega: X_n(\omega) > x,\ |X_n(\omega) - X(\omega)| < \epsilon) \\ &\quad + P(\omega: |X_n(\omega) - X(\omega)| \geq \epsilon) \\ &\leq P(\omega: X(\omega) > x - \epsilon) + P(\omega: |X_n(\omega) - X(\omega)| \geq \epsilon). \end{aligned} \tag{7.3.2}$$

De même, en utilisant que

$$X(\omega) > x + \epsilon, \quad \text{et} \quad |X_n(\omega) - X(\omega)| < \epsilon, \quad \text{alors} \quad X_n(\omega) > x.$$

on obtient

$$P(\omega: X(\omega) > x + \epsilon) \leq P(\omega: X_n(\omega) > x) + P(\omega: |X_n(\omega) - X(\omega)| \geq \epsilon). \tag{7.3.3}$$

On combine (7.3.2) et (7.3.3) pour obtenir

$$\begin{aligned} P(\omega: X(\omega) > x + \epsilon) &- P(\omega: |X_n(\omega) - X(\omega)| \geq \epsilon) \leq P(\omega: X_n(\omega) > x) \\ &\leq P(\omega: X(\omega) > x - \epsilon) + P(\omega: |X_n(\omega) - X(\omega)| \geq \epsilon). \end{aligned} \tag{7.3.4}$$

On fait tendre n vers l'infini pour obtenir

$$P(\omega: X(\omega) > x + \epsilon) \leq \lim_{n \to \infty} P(\omega: X_n(\omega) > x) \leq P(\omega: X(\omega) > x - \epsilon). \tag{7.3.5}$$

Dans (7.3.5), on fait tendre ϵ vers 0, en utilisant que X est continue, c'est-à-dire que

$$\lim_{\epsilon \to 0} P(\omega: X(\omega) > x + \epsilon) = P(\omega: X(\omega) > x) = \lim_{\epsilon \to 0} P(\omega: X(\omega) > x - \epsilon).$$

∎

7.4. Liens entre convergence en Probabilité et *en moyenne*.

7.4.1. Convergence *en moyenne* **plus fort qu'***en probabilité***.** — D'abord, la convergence *en moyenne* entraine la convergence en probabilité. Cela découle de l'inégalité de Markov :

$$P(\omega: |X_n(\omega) - X(\omega)| \geq \epsilon) \leq \frac{E[|X_n(\omega) - X(\omega)|]}{\epsilon}. \tag{7.4.1}$$

Ainsi, si le terme de droite converge vers 0, le terme de gauche aussi.

Nous allons montrer que la convergence en probabilité n'est pas suffisante pour garantir la convergence *en moyenne*. Soit $\Omega = [0,1[$, et pour $0 \leq a < b < 1$, on pose $P([a,b[) = b - a$. On pose

$$X_n = n 1\{[0, \frac{1}{n}[\}, \qquad Y_n = n^2 1\{[0, \frac{1}{n}[\}, \quad \text{et} \quad Z_n = \sqrt{n} 1\{[0, \frac{1}{n}[\}. \tag{7.4.2}$$

Il est facile de vérifier que X_n, Y_n et Z_n convergent toutes trois vers 0 en probabilité. Calculons leur espérance.

(7.4.3)
$$E[X_n] = nP([0, \frac{1}{n}[) = 1, \qquad E[Y_n] = n^2 \frac{1}{n} = n, \quad \text{et} \quad E[Z_n] = \sqrt{n} \frac{1}{n} = \frac{1}{\sqrt{n}}.$$

Ainsi, l'espérance peut avoir n'importe quel type de comportement !

7.4.2. Une application aux populations. — Nous avions étudié dans le chapitre des fonctions génératrices, un exemple de population, avec un seul ancêtre, et où chaque individu avait un nombre aléatoire d'enfants η indépendant des autres individus. Nous avions étudié l'extinction de la population, et montré que si $E[\eta] \leq 1$, alors il y avait extinction avec probabilité 1. Si Z_n était la taille de la population à l'époque n, alors nous avions

$$P\left(\omega : \ \exists n, \ Z_n(\omega) = 0\right) = 1. \tag{7.4.4}$$

En des termes plus simple, *la taille de la population, Z_n, tend vers 0.* Mais de quel type de convergence s'agit-il ?

Notez que (7.4.4) signifie que

$$\lim_{n\to\infty} P\left(\omega : \ Z_n(\omega) = 0\right) = 1 \quad \Longrightarrow \quad \forall \epsilon > 0, \ \lim_{n\to\infty} P\left(\omega : \ |Z_n(\omega)| > \epsilon\right) = 0. \tag{7.4.5}$$

Ainsi, il s'agit de convergence en probabilité. Notez, qu'au vu de la section précédente, il aurait suffit de montrer que $E[Z_n]$ tendait vers 0. Essayons de calculer cette espérance. Rappelons, que si G_X désigne la fonction génératrice de X

$$Z_n = \eta_1^{(n)} + \cdots + \eta_{Z_{n-1}}^{(n)}, \quad \text{et} \quad G_{Z_n}(t) = G_{Z_{n-1}}(G_\eta(t)). \tag{7.4.6}$$

Pour ne pas alourdir les notations, posons $f(t) = G_\eta(t)$, et $G_n(t) = G_{Z_n}(t)$. Pour obtenir $E[Z_n]$ il suffit de prendre la dérivée de $G_n(t)$ pour $t = 1$. Ainsi, en utilisant la formule de dérrivation d'une composée de fonctions

$$E[Z_n] = \left(G_n(t)\right)'|_{t=1} = f'(1) G'_{n-1}(f(1)). \tag{7.4.7}$$

Or $f(1) = 1$, et $f'(1) = E[\eta]$, ce qui implique que

$$E[Z_n] = E[\eta] \times E[Z_{n-1}], \quad \text{et par récurrence} \quad E[Z_n] = \left(E[\eta]\right)^n. \tag{7.4.8}$$

Ainsi,

$$E[\eta] < 1 \Longrightarrow \lim_{n\to\infty} E[Z_n] = 0. \tag{7.4.9}$$

Dans le cas $E[\eta] < 1$, on a donc que Z_n converge *en moyenne* vers 0. Par contre si $E[\eta] = 1$, on a que $E[Z_n] = 1$, alors que nous savons que Z_n converge en probabilité vers 0, sauf dans le cas où $\eta(\omega) \equiv 1$.

7.4.3. Converger en probabilité en étant bornée. — Lorsqu'une suite $\{X_n\}$ converge en probabilité vers 0, tout en étant uniformément bornée, alors elle converge en *en moyenne* vers 0. D'abord être uniformément borné signifie qu'il existe $M > 0$ tel que pour tout entier n et tout $\omega \in \Omega$ on a

$$|X_n(\omega)| \leq M.$$

Fixons $\epsilon > 0$ arbitraire. Alors, en utilisant qu'une probabilité est toujours bornée par 1,

$$\begin{aligned} E[|X_n|] &= \int_0^\infty P(|X_n| > t)dt = \int_0^M P(|X_n| > t)dt \\ &= \int_0^{\epsilon/2} P(|X_n| > t)dt + \int_{\epsilon/2}^M P(|X_n| > t)dt \\ &= \frac{\epsilon}{2} + MP(|X_n| > \epsilon). \end{aligned} \tag{7.4.10}$$

Or, X_n converge en probabilité implique que $P(|X_n| > \epsilon)$ tend vers 0 lorsque n tend vers l'infini. Ainsi, on choisit n assez grand pour que le dernier terme dans (7.4.10) soit plus petit que $\epsilon/2$.

Une inspection de la preuve permet de relaxer l'hypothèse $|X_n(\omega)| \leq M$, tout en obtenant un résultat plus fort. Supposons que $|X_n(\omega)|$ est uniformément bornée par une variable aléatoire $Y(\omega)$, telle que $E[Y^2] < \infty$. Alors, $E[X_n^2]$ tend vers 0.

En effet, $|X_n(\omega)| \leq Y(\omega)$ implique $P(\omega : X(\omega) > t) \leq P(\omega : Y(\omega) > t)$, ce qui a son tour implique

$$\int_M^\infty 2tP(\omega : X(\omega) > t)dt \leq \int_M^\infty 2tP(\omega : Y(\omega) > t)dt \underset{M\to\infty}{\longrightarrow} 0. \tag{7.4.11}$$

On a utilisé que

$$E[Y^2] = \int_0^\infty 2tP(\omega : Y(\omega) > t)dt < \infty \Longrightarrow \int_M^\infty 2tP(\omega : Y(\omega) > t)dt \underset{M\to\infty}{\longrightarrow} 0.$$

Ainsi, on reprend (7.4.10) en prenant l'espérance du carré de X.

$$\begin{aligned} E[X_n^2] &= \int_0^\infty 2t\ P(|X_n| > t)dt = \int_0^M 2t\ P(|X_n| > t)dt + \int_M^\infty 2t\ P(|X_n| > t)dt \\ &= \int_0^{\epsilon/2} P(|X_n| > t)dt + \int_{\epsilon/2}^M P(|X_n| > t)dt + \int_M^\infty 2t\ P(Y > t)dt \\ &= \frac{\epsilon}{2} + MP(|X_n| > \epsilon) + \int_M^\infty 2t\ P(Y > t)dt. \end{aligned} \tag{7.4.12}$$

Les deux derniers termes de (7.4.12) tendent vers 0 lorsque M tend vers l'infini.

CHAPITRE 8

COMPORTEMENT D'UNE SOMME DE VARIABLES

8.1. Loi des grands nombres

8.1.1. Motivation. — Supposons que l'on veuille connaitre la proportion p de fumeurs dans une population. Une façon naturelle d'avoir une idée de p, est de compter le nombre de fumeurs dans un (grand) échantillon de taille n dans cette population. Si on note $\hat{p}_n$ la proportion de fumeurs dans l'échantillon, il est naturel de penser que $\hat{p}_n \approx p$ quand n est grand. Le théorème qui justifie cette intuition est la célèbre loi des grands nombres. Notons que p est un nombre, alors que $\hat{p}_n$ dépend de l'échantillon. Ainsi, on peut se demander quel sens donner à l'affirmation $\hat{p}_n \approx p$? Doit-on comprendre que

1. $\hat{p}_n \approx p$ pour n'importe quel grand échantillon choisi ?
2. si on refaisait l'expérience sur un grand nombre d'échantillons, il y aurait beaucoup d'échantillons pour lesquels $\hat{p}_n \approx p$?

Chacune de ces questions correspond à **une** notion de convergence de variables aléatoires, et à **une** version de la loi des grands nombres. Pour être un peu plus précis, modélisons l'expérience précédente. Comme on veut prendre n grand, on considère en théorie qu'on peut interroger un nombre infini de personnes. Ainsi, l'espace des issues est

$$\Omega = \{\omega = (\omega_i,\ i \in \mathbb{N}^*) :\ \omega_i \in \{1, 0\}\},$$

où $\omega_i = 1$ signifie que la $i^{\text{ème}}$ personne interrogée est un fumeur.

La probabilité P sur Ω est telle que pour tout n, et tout $(\epsilon_1, \cdots, \epsilon_n) \in \{0,1\}^n$,

$$P(\{\omega :\ \omega_1 = \epsilon_1, \cdots, \omega_n = \epsilon_n\}) = P_1(\epsilon_1) \cdots P_1(\epsilon_n),$$

où $P_1(1) = p$ et $P_1(0) = 1 - p$.

Sur Ω, on définit les variables indépendantes de Bernoulli $\{X_i, i = 1, \ldots, n\}$, avec $X_i(\omega) = \omega_i$. Puis, la moyenne

$$\bar{X}_n = \frac{X_1 + \cdots X_n}{n}. \tag{8.1.1}$$

On peut alors reformuler les questions précédentes par :

1. est-ce-que pour tous les ω, $\lim_{n\to\infty} \bar{X}_n(\omega) = p$?
2. si $\epsilon > 0$, est-ce-que $\lim_{n\to\infty} P\left[|\bar{X}_n - p| \leq \epsilon\right] = 1$?

On pourrait aussi se demander si $E\left[|\bar{X}_n - p|\right]$ tend vers 0 quand $n \to \infty$, si $E\left[|\bar{X}_n - p|^2\right]$ tend vers 0 quand $n \to \infty$, ...

La réponse à toutes ces questions est affirmative. Nous allons étudier dans cette section la question 2.

Définition 25. — . Soient $\{X_i,\ i \geq 1\}$ des v.a. indépendantes et de même loi. La variable $\bar{X}_n$ donnée dans (8.1.1) s'appelle **la moyenne empirique** des X_i.

Il est facile de voir que

$$E[\bar{X}_n] = E[X_1], \quad \text{et} \quad \mathrm{var}(\bar{X}_n) = \frac{\mathrm{var}(X_1)}{n}. \tag{8.1.2}$$

8.1.2. Rappels : Les inégalités de Chebychev. —

Proposition 19. — *. Soit (Ω, P) et X une variable aléatoire. Pour tout $c > 0$, nous avons ce que l'on appelle aussi l'inégalité de Markov :*

$$P(\{\omega : |X(\omega)| \geq c\}) \leq E[|X|]/c. \tag{8.1.3}$$

Puis, une inégalité plus facile à utiliser :

$$P(\{\omega : |X(\omega)| \geq c\}) \leq E[X^2]/c^2. \tag{8.1.4}$$

Finalement, l'estimation de probabilité d'évènements rares utilise souvent l'inégalité suivante : pour tout $\lambda > 0$,

$$P(\{\omega : X(\omega) \geq c\}) \leq E[\exp(\lambda(X - c))]. \tag{8.1.5}$$

Démonstration. — Pour 1), on applique la proposition 19 à $|X|$. Pour 2), il suffit de voir que $\{\omega : |X(\omega)| \geq c\} = \{\omega : X^2(\omega) \geq c^2\}$, et d'appliquer la proposition 19 à X^2 et c^2. Pour 3), il suffit de voir que $\{\omega : X(\omega) \geq c\} = \{\omega : Y(\omega) \geq 1\}$, avec $Y = \exp(\lambda(X - c))$ et d'appliquer la proposition 19 à Y et 1. ∎

8.1.3. Loi faible des grands nombres. —

Théorème 3. — *. Soit $(X_i)_{i\geq 1}$ des v.a. indépendantes et de même loi, de moyenne μ et de variance σ^2. Pour tout $\epsilon > 0$,*

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 0\,.$$

Démonstration. — On applique l'inégalité (8.1.4) à la variable $Y = \bar{X}_n$. On a déjà vu dans (8.1.2) que $E(\bar{X}_n) = \mu$, et que $\mathrm{var}(\bar{X}_n) = \frac{\sigma^2}{n}$. Ainsi, par (8.1.4)

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\mathrm{var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \underset{n\to\infty}{\longrightarrow} 0\,.$$

∎

Nous pouvons reformuler différement la LGN. On centre d'abord les variables en définissant $Y_i = X_i - E[X_i]$, de façon à ce que $E[Y_i] = 0$. On a donc montré que

$$\frac{Y_1 + \cdots + Y_n}{n} \underset{\text{probabilité}}{\longrightarrow} 0.$$

Nous allons nous poser deux types de questions :

1. a-t-on $\frac{1}{n^\beta}(Y_1 + \cdots + Y_n)$ converge vers 0, pour $\beta < 1$?
2. Peut-on estimer la "petitesse" de $P(|Y_1 + \cdots + Y_n| > ny)$ pour y positif fixé ?

Pour répondre à ces questions nous avons besoin de quelques notions d'analyse.

8.2. Au-delà de la loi des grands nombres

On considère une suite de variables $\{Y_n,\ n \in \mathbb{N}\}$ indépendantes de même loi avec $E[Y_n] = 0$, $|Y_n(\omega)| \le 1$ pour tout ω, et on pose

$$Z_n = \frac{Y_1 + \cdots + Y_n}{\sqrt{n}} = \sqrt{n}\bar{Y}_n,$$

si $\bar{Y}_n$ est la moyenne empirique des $\{Y_n,\ n \in \mathbb{N}\}$. Le calcul de (8.1.2) donne que $E[Z_n^2] = var(Y_1)$. Cela implique que pour tout $\epsilon > 0$, Z_n/n^ϵ converge en probabilité vers 0. En effet, par l'inégalité de Chebychev (8.1.4), pour tout $\delta > 0$

$$P(\omega : |\frac{Z_n(\omega)}{n^\epsilon}| > \delta) \le \frac{E[Z_n^2]}{\delta^2 n^{2\epsilon}} = \frac{var(Y_1)}{\delta^2 n^{2\epsilon}} \underset{n\to\infty}{\longrightarrow} 0. \tag{8.2.1}$$

Ainsi, la réponse à la première question est affirmative pour $\beta > \frac{1}{2}$. Par contre, $\beta = \frac{1}{2}$ ne marche pas, c'est-à-dire que Z_n ne peut converger en probabilité vers 0. En effet, il est facile de calculer $E[Z_n^4]$ (à faire !) et d'obtenir pour un constante C_Y

$$E[Z_n^4] \le C_Y.$$

Puis de noter que

$$E[Z_n^2 1_{|Z_n|>x}] \le \big(E[Z_n^4]P(Z_n > x)\big)^{1/2} \le \big(C_Y P(Z_n > x)\big)^{1/2}.$$

Il reste a choisir un bon x : Notez que si $x^2 = E[Z_n^2]/2 = \mathrm{var}(Y_1)/2$, alors

$$E[Z_n^2 1_{|Z_n|>x}] = E[Z_n^2] - E[Z_n^2 1_{|Z_n|\le x}] \ge \frac{1}{2}E[Z_n^2] = \frac{1}{2}\mathrm{var}(Y_1).$$

D'ou

$$C_Y P\left(Z_n > \sqrt{\frac{\mathrm{var}(Y_1)}{2}}\right) \ge \frac{1}{4}\mathrm{var}(Y_1)^2. \tag{8.2.2}$$

Notez que si Z_n convergeait vers 0, alors $P(Z_n > x)$ tendrait vers 0 pour $x > 0$. Ce qui contredit l'inegalité précédente.

Nous allons aussi montrer une borne supérieure.

Proposition 20. — *. Pour tout $x > 0$, il existe une fonction $\epsilon(x,n)$ telle que $\lim_{n\to\infty} \epsilon(x,n) = 0$, et*

$$P(\omega : |Z_n| > x) \leq \exp(-\frac{x^2}{2E[Y_1^2]} + \epsilon(x,n)). \tag{8.2.3}$$

Concernant la deuxième question, nous allons montrer le résultat suivant.

Proposition 21. — *. Pour tout $y > 0$, il existe $c_1(y), c_2(y)$ positives, telles que*

$$\exp(-nc_1(y)) \leq P(|\bar{Y}_n| > yn) \leq \exp(-nc_2(y)). \tag{8.2.4}$$

8.2.1. Préliminaire d'analyse. — Nous avons besoin de faire une étude de quelques propriétés de la fonction exponentielle. Cette section suppose que vous ne savez pas ce qu'est un developpement limité.

Lemme 6. — *. Pour tout $x \in \mathbb{R}$, $e^x \geq 1 + x$.*

Démonstration. — Posons $f(x) = e^x - 1 - x$, et notons que $f(0) = 0$. Dérivons, $f'(x) = e^x - 1$. On a donc que $f'(x) \geq 0$ pour $x > 0$, et $f'(x) < 0$ pour $x < 0$. Cela entraine que f décroit avant 0, atteind la valeur 0 en 0, puis croit pour $x \geq 0$. Ainsi f reste toujours positive ou nulle. ∎

Posons $g_0(x) = e^x - 1 = \psi_0(x)$, et pour tout $n \geq 1$

$$g_n(x) = e^x - \left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}\right), \quad \text{et} \quad \psi_n(x) = \int_0^x \psi_{n-1}(u)du. \tag{8.2.5}$$

Notez que ψ_n est définie par récurrence. On veut montrer que pour x "petit", $g_n(x)$ est petit, et nous allons le faire à l'aide de l'identité suivante.

Lemme 7. — *. Pour tout entier n, $g_n(x) = \psi_n(x)$.*

Démonstration. — Nous utilisons un preuve par récurrence. La propriété est vraie pour $n = 0$. Supposons qu'elle est vraie à l'ordre $n - 1$. Notez que

$$g_n'(x) = g_{n-1}(x), \quad \text{et} \quad \psi_n'(x) = \psi_{n-1}(x). \tag{8.2.6}$$

Ainsi par hypothèse $g_n'(x) = \psi_n'(x)$. Aussi $g_n(0) = \psi_n(0) = 0$, et finalement

$$g_n(x) = g_n(0) + \int_0^x g_n'(u)du = \psi_n(0) + \int_0^x \psi_n'(u)du = \psi_n(x). \tag{8.2.7}$$

∎

Nous déduisons de lemma 7 la conséquence suivante

Corollaire 5. — *. Pour tout entier n, et réel x*

$$|e^x - \sum_{i=0}^{n} \frac{x^i}{i!}| \leq e^{|x|} \frac{|x|^{n+1}}{(n+1)!}. \tag{8.2.8}$$

Démonstration. — Considérons deux cas. Cas $x \geq 0$, alors montrons par récurrence que pour tout $u \in [0, x]$, $0 \leq \psi_n(u) \leq e^x u^n/n!$. Supposons que c'est vraie à l'ordre $n-1$. Alors

$$\psi_n(u) = \int_0^u \psi_{n-1}(y)dy \leq \int_0^u e^x \frac{y^{n-1}}{(n-1)!}dy \leq e^x \frac{u^n}{n!}. \tag{8.2.9}$$

De plus, on voit que $\psi_n(u) \geq 0$. Vérifions la propriété à l'ordre 0 : $\psi_0(u) = e^u - 1 \leq e^x$, et $\psi_0(u) \geq 0$ pour $0 \leq u \leq x$.

Cas $x \leq 0$. Montrons par récurrence que $|\psi_n(x)| \leq |x|^n/n!$ pour tout $x \leq 0$. Supposons que c'est vraie à l'ordre $n-1$. Notons

$$|\psi_n(x)| = |\int_x^0 \psi_{n-1}(y)dy| \leq \int_x^0 \frac{|y|^{n-1}}{(n-1)!}dy = \int_0^{|x|} \frac{y^{n-1}}{(n-1)!}dy = \frac{|x|^n}{n!}. \tag{8.2.10}$$

∎

Une applications. Fixons x et choisissons n très grand ($n \geq |x|$), alors

$$|e^{\frac{x}{n}} - 1 - \frac{x}{n}| \leq e^{\frac{|x|}{n}} \frac{|x|^2}{2!n^2} \leq C\frac{x^2}{n^2}, \quad \text{avec} \quad C = \frac{e}{2}. \tag{8.2.11}$$

Lemme 8. — . *Pour tout réel x*

$$\lim_{n\to\infty} (1 + \frac{x}{n})^n = e^x. \tag{8.2.12}$$

Démonstration. — Nous utilisons l'inégalité

$$a^n - b^n = (a-b)\left(a^{n-1} + a^{n-2}b^1 + \cdots + a^1 b^{n-2} + b^{n-1}\right), \tag{8.2.13}$$

appliquée à $a = e^{x/n}$, et $b = 1 + \frac{x}{n}$. Ainsi,

$$\begin{aligned} e^x - (1+\frac{x}{n})^n &= a^n - b^n = \left(e^{x/n} - (1+\frac{x}{n})\right) a^{n-1} \left(\sum_{k=0}^{n-1} \frac{b^k}{a^k}\right) \\ &= e^{\frac{(n-1)}{n}x} \left(e^{x/n} - (1+\frac{x}{n})\right) \left(\sum_{k=0}^{n-1} \left(\frac{1+\frac{x}{n}}{e^{x/n}}\right)^k\right). \end{aligned} \tag{8.2.14}$$

On utilise alors (8.2.11) et le lemme 6, pour obtenir que

$$0 \leq e^x - (1+\frac{x}{n})^n \leq Ce^{|x|/n} \frac{x^2}{n^2} n \leq \frac{Ce^{|x|}x^2}{n}. \tag{8.2.15}$$

∎

On va généraliser le lemme 8 de la façon suivante

Lemme 9. — . *Soit $\psi(x,n)$ une fonction telle que* $\lim_{n\to\infty} n\psi(x,n) = 0$. *Alors*

$$\lim_{n\to\infty} (1 + \frac{x}{n} + \psi(x,n))^n = e^x. \tag{8.2.16}$$

Démonstration. — En utilisant le lemme 8, on voit qu'on a besoin de montrer que

$$\lim_{n\to\infty} |(1+\frac{x}{n}+\psi(x,n))^n - (1+\frac{x}{n})^n| = 0. \tag{8.2.17}$$

On utilise a nouveau (8.2.13) avec $a = 1+\frac{x}{n}+\psi(x,n)$, et $b = 1+\frac{x}{n}$, et l'on obtient

$$\begin{aligned} |a^n - b^n| &\leq |\psi(x,n)| \left(a^{n-1}+\cdots+b^{n-1}\right) \\ &\leq |\psi(x,n)| \times n \times \left(1+\frac{|x|}{n}+|\psi(x,n)|\right)^n \quad \text{puis le lemme 6} \\ &\leq n|\psi(x,n)| \exp\left(n(\frac{|x|}{n}+|\psi(x,n)|)\right) = n|\psi(x,n)|e^{x+n|\psi(x,n)|}. \end{aligned} \tag{8.2.18}$$

Comme $\lim_{n\to\infty} n\psi(x,n) = 0$, on déduit (8.2.16) de (8.2.18). ∎

8.2.2. Démonstration de la proposition 20.

— On utilise l'inégalité (8.1.5) pour obtenir pour tout $\lambda > 0$

$$\begin{aligned} P(\frac{Y_1+\cdots+Y_n}{\sqrt{n}} > x) &\leq e^{-\lambda x} E[\exp(\lambda \frac{Y_1+\cdots+Y_n}{\sqrt{n}})] \\ &= e^{-\lambda x}\left(E[e^{\frac{\lambda}{\sqrt{n}}Y_1}]\right)^n. \end{aligned} \tag{8.2.19}$$

Utilisons les notations de la démonstration du lemme 6,

$$g_2(\frac{\lambda}{\sqrt{n}}Y_1(\omega)) = \exp(\frac{\lambda}{\sqrt{n}}Y_1(\omega)) - 1 - \frac{\lambda}{\sqrt{n}}Y_1(\omega) - \frac{1}{2}\left(\frac{\lambda}{\sqrt{n}}Y_1(\omega)\right)^2.$$

On applique le corollaire 5, et l'hypothèse $|Y_i| \leq 1$ pour obtenir

$$|g_2(\frac{\lambda}{\sqrt{n}}Y_1(\omega))| \leq \exp\left(\frac{\lambda}{\sqrt{n}}|Y_1(\omega)|\right)\frac{\lambda^3|Y_1^3(\omega)|}{3!n\sqrt{n}} \leq \exp\left(\frac{\lambda}{\sqrt{n}}\right)\frac{\lambda^3}{3!n\sqrt{n}}. \tag{8.2.20}$$

Posons $\psi(\lambda,n) = E[g_2(\frac{\lambda}{\sqrt{n}}Y_1(\omega))]$, et notons que

$$n \times |\psi(\lambda,n)| \leq e^{\frac{\lambda}{\sqrt{n}}}\frac{\lambda^3}{3!\sqrt{n}} \underset{n\to\infty}{\longrightarrow} 0. \tag{8.2.21}$$

Nous pouvons donc réecrire le terme de droite dans (8.2.19)

$$\begin{aligned} E[e^{\frac{\lambda}{\sqrt{n}}Y_1}] &= 1+\frac{\lambda}{\sqrt{n}}E[Y_1]+\frac{\lambda^2}{2n}E[Y_1^2]+\psi(\lambda,n) \\ &= 1+\frac{\lambda^2}{2n}var(Y_1)+\psi(\lambda,n). \end{aligned} \tag{8.2.22}$$

En utilisant lemme 6, on obtient

$$\left(E[e^{\frac{\lambda}{\sqrt{n}}Y_1}]\right)^n \leq \exp\left(\frac{\lambda^2}{2}var(Y_1)+n\times\psi(\lambda,n)\right). \tag{8.2.23}$$

Finalement,

$$P(\omega:\ Z_n(\omega) > x) \leq \exp\left(-\left(\lambda x - \frac{\lambda^2}{2}var(Y_1) - n\times\psi(\lambda,n)\right)\right). \tag{8.2.24}$$

Il suffit de choisir $\lambda var(Y_1) = x$ dans (8.2.24), ce qui donne

$$P(\omega:\ Z_n(\omega) > x) \leq \exp\left(-\frac{x^2}{2var(Y_1)} + n\psi(\frac{x}{var(Y_1)}, n)\right). \tag{8.2.25}$$

8.2.3. Estimées de Grandes Déviations (21). — Nous établissons la borne inférieure. Notez que si pour tous $i = 1, \ldots, n$, on a $Y_i > y$, alors $Y_1 + \cdots + Y_n > ny$. Ainsi, en utilisant l'indépendance des Y_i, on obtient pour $y > 0$

$$P(Y_1 + \cdots + Y_n > ny) \geq P(Y_1 > y)^n. \tag{8.2.26}$$

Notez que $E[Y_1] = 0$ implique que $P(Y_1 > y) < 1$. En effet, si nous avions $P(Y_1 > y) = 1$, alors $Y_1(\omega) > y$ pour tous les ω dans un ensemble de probabilité 1. Ce qui entrainerait que $E[Y_1] > y$. Il existe donc $c_1(y) > 0$ telle que $\exp(-c_1(y)) = P(Y_1 > y)$. L'inégalité (8.2.26) implique

$$P(Y_1 + \cdots + Y_n > ny) \geq \exp(-n\ c_1(y)). \tag{8.2.27}$$

Pour la borne supérieure de (21), nous allons suivre de très prés la preuve de la proposition (20). On utilise l'inégalité (8.1.5) pour obtenir pour tout $\lambda > 0$

$$\begin{aligned} P(Y_1 + \cdots + Y_n > ny) &\leq e^{-\lambda ny} E[\exp(\lambda(Y_1 + \cdots + Y_n))] \\ &= e^{-\lambda ny}\left(E[e^{\lambda Y_1}]\right)^n. \end{aligned} \tag{8.2.28}$$

Utilisons les notations de la démonstration du lemme 6,

$$g_1(\lambda Y_1(\omega)) = \exp(\lambda Y_1(\omega)) - 1 - \lambda Y_1(\omega)$$

On applique le corollaire 5, et l'hypothèse $|Y_i| \leq 1$ pour obtenir

$$|g_1(\lambda Y_1(\omega))| \leq \exp(\lambda |Y_1(\omega)|)\frac{\lambda^2 |Y_1^2(\omega)|}{2!} \leq \frac{e^\lambda \lambda^2}{2}. \tag{8.2.29}$$

Nous pouvons donc réecrire le terme de droite dans (8.2.28)

$$\begin{aligned} E[e^{\lambda Y_1}] &= 1 + \lambda E[Y_1] + E[g_1(\lambda Y_1)] \\ &\leq 1 + \frac{e^\lambda \lambda^2}{2}. \end{aligned} \tag{8.2.30}$$

En utilisant lemme 6, on obtient

$$\left(E[e^{\lambda Y_1}]\right)^n \leq \exp\left(n\frac{e^\lambda \lambda^2}{2}\right). \tag{8.2.31}$$

Finalement,

$$P(\omega:\ Z_n(\omega) > x) \leq \exp\left(-n\left(\lambda y - \frac{e^\lambda \lambda^2}{2}\right)\right). \tag{8.2.32}$$

Il suffit de choisir λ tel que $4\lambda y \geq \exp(\lambda)\lambda^2$. En choisissant $\lambda < 1$, on voir qu'il suffit que $4ye^{-1} \geq \lambda$. On choisit donc $\lambda = \min(1, 4ye^{-1})$.