

PROBABILITÉS ET STATISTIQUES

Peter Haïssinsky

Introduction

Le calcul des probabilités a eu historiquement pour premier objectif de modéliser des jeux de hasard, pour ensuite étudier des expériences aléatoires variées et plus complexes. Les progrès technologiques actuels requièrent beaucoup de modélisations probabilistes, c'est le cas notamment en finance et en biologie. En même temps, de plus en plus de données complexes et variées sont stockées sur les ordinateurs. On voudrait traiter ces grosses quantités de données, pour en extraire une information utile, ce qui requiert des méthodes statistiques complexes. En outre, beaucoup d'algorithmes utilisent des tirages aléatoires.

La théorie des probabilités modélise des expériences où plusieurs issues sont possibles, la réalisation n'étant pas déterminée à l'avance, par exemple un lancer de dés, la roulette au casino, l'évolution du cours d'une action à la bourse, le chromosome transmis par un parent à son enfant... Cela en vue notamment d'évaluer les risques de paris et de mettre sur pied des stratégies pour faire face aux aléas. Le calcul des probabilités ne va pas permettre de prédire quelle issue va se réaliser, mais quelle chance a chaque issue de se réaliser. Le but des mathématiques financières n'est pas de gagner plus d'argent en spéculant mieux, mais de se prémunir contre le risque inhérent aux investissements en bourse. On va attribuer à chaque issue possible un pourcentage de réalisation, ou plutôt, un nombre entre 0 et 1 qui indique la chance qu'a cette issue de se réaliser : plus le nombre est petit, moins l'issue aura de chances de se réaliser. On appelle ce nombre la *probabilité* de cette issue. On appelle « événement » un ensemble d'issues. La probabilité qu'on associe à un événement est la somme des probabilités de chaque issue de cet ensemble ; un événement de probabilité nulle sera improbable alors qu'un événement de probabilité 1 sera certain.

Le but de la statistique est de traiter des données (en général en grand nombre) en vue d'en tirer une information utile. Ces données proviennent d'un sondage, de mesures, etc., et sont donc des réalisations d'un phénomène aléatoire. Elle se décline en deux champs : la statistique descriptive (ou analyse de données) et la statistique inférentielle.

- La statistique descriptive a pour objet d'extraire des informations pertinentes à partir de listes complexes de nombres. La moyenne des notes est un tel indicateur.
- La statistique inférentielle part de données d'un échantillon et cherche à généraliser ces résultats à la population entière, comme le fait les sondages. On peut aussi s'interroger sur la qualité d'un produit fabriqué à la chaîne.

L'objet du cours est de présenter une introduction rigoureuse au calcul des probabilités et aux statistiques. Dans le cadre des probabilités, nous partirons de situations simples qui nous permettront de développer les notions centrales de la théorie afin d'aborder ensuite des situations plus complexes. En ce qui concerne les statistiques, nous nous intéresserons à la statistique inférentielle et notamment aux tests.

Chapitre I. Objets centraux¹

1 Espace de probabilité

Nous commençons par présenter des exemples qui nous serviront à illustrer le cours.

Exemple 1.1 *On considère un dé à 6 faces, numérotés de 1 à 6. Lorsque l'on lance ce dé, les six numéros peuvent apparaître ; chacun d'eux représente une « issue » possible.*

On suppose que le dé est équilibré, ce qui veut dire que les 6 faces ont la même chance de sortir. Donc chaque numéro a une chance sur six d'être tiré. Cela veut dire (en un sens qui sera précisé plus loin) que si l'on jette n fois de suite le dé (n « grand »), on obtient en gros $n/6$ fois 1.

On peut considérer par exemple l'événement « le numéro est pair », ce qui correspond à ce que le numéro soit 2, 4 ou 6. Comme la moitié des numéros est paire, on aura une chance sur deux d'obtenir un tel numéro.

Exemple 1.2 *On considère le même dé, sauf que sur la sixième face, le nombre 6 a été remplacé par 5. Il y a donc deux faces où 5 est inscrit. L'ensemble des issues est différent : $\{1; 2; 3; 4; 5\}$. Comme le dé est équilibré, les faces 1,2,3 et 4 ont chacune la probabilité $1/6$ de sortir, alors que pour 5, on aura deux fois plus de chance de l'obtenir, à savoir $2/6 = 1/3$.*

Exemple 1.3 *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 résultats des deux dés : ainsi la personne II n'aura accès qu'à un nombre entre 2 et 12.*

Quelles sont les issues ? Il y a une certaine ambiguïté puisque I et II ne voient pas la même chose. La personne I voit les résultats des lancers 1 et 2, d'où

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

Dans cet espace, chaque couple a la même chance de sortir : pour tous les i, j dans $\{1, \dots, 6\}$, i a une chance sur six de sortir, et j aussi, donc la probabilité que i et j soient tirés dans cet ordre sera $(1/6) \cdot (1/6) = 1/36$.

Pour la personne II, les issues possibles sont

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

Par exemple si la somme des deux dés est 4, cela englobe les couples $\{(1, 3), (2, 2), (3, 1)\}$ de I. Ainsi, la probabilité que II voit un quatre sera

$$p_4^{II} = p_{(1,3)}^I + p_{(2,2)}^I + p_{(3,1)}^I = 3/36 = 1/12.$$

¹16 avril 2012

Exemple 1.4 On joue à pile ou face avec une pièce telle que pile apparaît avec probabilité p , donc face avec probabilité $1 - p$. Si on joue n fois, alors chaque issue sera une liste de n issues pile ou face, c'est-à-dire un n -uplet $\omega = (\omega_1, \dots, \omega_n)$ où chaque ω_j désignera le résultat du lancer numéro j , soit pile ou face. Notons s le nombre de piles sur n lancers consécutifs. Ce tirage avait une probabilité de $p^s(1 - p)^{n-s}$ d'arriver.

Exemple 1.5 Nous restons dans le jeu de n lancers d'une pièce, mais on ne s'intéresse qu'au nombre de piles. Ce nombre se trouvera dans l'ensemble $\{0, 1, \dots, n\}$. On remarque que chaque n -uplet qui contient exactement k piles se produit avec probabilité $p^k(1 - p)^{n-k}$. Par ailleurs, le nombre de telles configurations sera le nombre de n -uplets qui contiennent exactement k piles : autrement dit, il correspondra au nombre de manières de choisir k éléments dans un ensemble de n : soit C_n^k . Au final, la probabilité d'avoir k piles sera $C_n^k p^k (1 - p)^{n-k}$.

Exemple 1.6 Maintenant, on joue à pile ou face jusqu'à ce que l'on tire un pile. On s'intéresse alors au nombre nécessaire de lancers et à sa probabilité. Ce nombre peut être a priori n'importe quel entier naturel non nul. Si on doit tirer k fois, cela signifie que l'on a eu $(k - 1)$ faces, ce qui arrive avec probabilité $(1 - p)^{k-1}$, puis un pile, avec probabilité p . Au total, cela nous fait $(1 - p)^{k-1} p$.

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 qu'une issue arbitraire sera notée ω (petit omega).

Définition 1.1 (Distribution de probabilité) Soit Ω un ensemble au plus dénombrable. Une distribution de probabilité sur Ω est la donnée d'un nombre $p_\omega \geq 0$ pour chaque $\omega \in \Omega$ tel que

$$\sum_{\omega \in \Omega} p_\omega = 1.$$

Définition 1.2 (Espace de probabilité (au plus dénombrable)) Un espace de probabilité est un triplet $(\Omega, \mathcal{A}, \mathbb{P})$ où

- Ω , qui désigne l'ensemble des issues, est un ensemble au plus dénombrable muni d'une distribution de probabilité $(p_\omega)_{\omega \in \Omega}$,
- $\mathcal{A} = \mathcal{P}(\Omega)$ est l'ensemble des événements et
- $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ est la mesure de probabilité sur (Ω, \mathcal{A}) , c'est-à-dire l'application définie ainsi : pour tout événement $E \subset \Omega$, on pose

$$\mathbb{P}[E] = \sum_{\omega \in E} p_\omega.$$

Illustrons ces définitions par nos exemples.

Exemple 1.1.— L'ensemble des issues possibles d'un lancer est $\Omega = \{1; 2; 3; 4; 5; 6\}$ et on associe à chaque issue la probabilité $1/6$. Donc on pose, pour chaque $j \in \Omega$, $p_j = 1/6$. On vérifie

facilement que $p_1 \geq 0, \dots, p_6 \geq 0$ et $p_1 + \dots + p_6 = 1$. L'événement « le tirage est pair » est représenté par l'ensemble $E = \{2; 4; 6\}$ et on trouve $\mathbb{P}(E) = p_2 + p_4 + p_6 = 1/2$.

Dans cet exemple, on dit que la probabilité est *uniforme* ou que les tirages sont *équiprobables*.

Définition 1.3 (Loi équiprobable/uniforme sur un ensemble fini) *Sur un ensemble fini Ω , la loi uniforme, notée $\mathcal{U}(\Omega)$, est la loi qui donne le même poids à chaque issue : pour tout $\omega \in \Omega$, on a $p_\omega = 1/|\Omega|$.*

Exemple 1.2.— L'ensemble des issues est différent du précédent : $\Omega = \{1; 2; 3; 4; 5\}$. Comme le dé est équilibré, les faces 1, 2, 3 et 4 ont chacune la probabilité $1/6$ de sortir donc $p_1 = p_2 = p_3 = p_4 = 1/6$, mais 5 a deux chances de sortir donc $p_5 = 2/6 = 1/3$. On vérifie encore que $p_1 \geq 0, \dots, p_5 \geq 0$ et $p_1 + \dots + p_5 = 1$.

Exemple 1.4.— Dans un premier temps, en posant P =pile et F =face, on aura $\Omega = \{P, F\}$, $p_P = p$ et $p_F = 1 - p$.

Définition 1.4 (Loi de Bernoulli) *Si $p \in [0, 1]$, la loi de Bernoulli, notée $\mathcal{B}(p)$ est la probabilité sur $\Omega = \{0, 1\}$ où $p_1 = p$ et $p_0 = 1 - p$.*

Si on joue n fois, alors on considère

$$\Omega_n = \Omega^n = \{\omega = (\omega_1, \dots, \omega_n), \omega_j \in \{P, F\}\}.$$

En notant $s_n(\omega)$ le nombre d'indices $j \in \{1, \dots, n\}$ pour lesquels $\omega_j = P$, on aura $p_\omega = p^{s_n(\omega)}(1 - p)^{n - s_n(\omega)}$. S'il est facile de vérifier que $p_\omega \geq 0$ pour tout ω , il est en revanche plus difficile de vérifier que la somme vaut 1 :

$$\begin{aligned} \sum_{\omega \in \Omega} p_\omega &= \sum_{\omega \in \Omega} p^{s_n(\omega)}(1 - p)^{n - s_n(\omega)} \\ &= \sum_{k=0}^n \sum_{\omega, s_n(\omega)=k} p^{s_n(\omega)}(1 - p)^{n - s_n(\omega)} \\ &= \sum_{k=0}^n p^k(1 - p)^{n - k} |\{\omega, s_n(\omega) = k\}| \\ &= \sum_{k=0}^n C_n^k p^k(1 - p)^{n - k} \\ &= [p + (1 - p)]^n = 1. \end{aligned}$$

Pour ce calcul, nous avons utilisé que le nombre de piles possible variait entre 0 et n et qu'on avait C_n^k issues différentes qui produisaient exactement k piles. Ensuite, on a utilisé la formule du binôme de Newton.

Exemple 1.5.— Maintenant, nous nous intéressons à la valeur de s_n . Donc $\Omega = \{0, \dots, n\}$. Nous avons vu que pour chaque k , nous avons C_n^k listes de tirages possibles qui réalisaient ce nombre de piles, donc $p_k = C_n^k p^k(1 - p)^{n - k}$. Le calcul précédent nous montre que c'est bien une distribution de probabilité.

Définition 1.5 (Loi binomiale) *Etant donnés $p \in [0, 1]$ et $n \geq 1$, la loi binomiale $\mathcal{B}(n, p)$ est la loi sur $[0, \dots, n]$ telle que $p_k = C_n^k p^k (1-p)^{n-k}$.*

Exemple 1.6.— Nous avons $\Omega = \mathbb{N} \setminus \{0\}$, et nous avons trouvé $p_k = p(1-p)^{k-1}$ pour $p > 0$. Nous obtenons ainsi

$$\sum_{k \geq 1} p_k = p \sum_{k \geq 1} (1-p)^{k-1} = p \cdot \frac{1}{1-(1-p)} = 1.$$

Ici, nous avons reconnu la somme d'une suite géométrique de raison plus petite que 1.

Définition 1.6 (Loi géométrique) *Etant donné $q \in]0, 1]$, la loi géométrique $\mathcal{G}(q)$ est la loi sur \mathbb{N} telle que $p_0 = 0$ et si $k \geq 1$, $p_k = q(1-q)^{k-1}$.*

Rappels sur les ensembles.— Si A et B sont deux sous-ensembles de Ω , on note

$$A \cup B = \{\omega \in \Omega; \omega \in A \text{ ou } \omega \in B\},$$

$$A \cap B = \{\omega \in \Omega; \omega \in A \text{ et } \omega \in B\},$$

$$A \setminus B = \{\omega \in \Omega; \omega \in A \text{ et } \omega \notin B\},$$

$${}^c A = \Omega \setminus A.$$

Si $(A_i)_{i \in I}$ et $(B_j)_{j \in J}$ sont deux familles de sous-ensembles de Ω alors

$$\left(\bigcup_{i \in I} A_i \right) \cap \left(\bigcup_{j \in J} B_j \right) = \bigcup_{(i,j) \in I \times J} A_i \cap B_j$$

et

$$\left(\bigcap_{i \in I} A_i \right) \cup \left(\bigcap_{j \in J} B_j \right) \subset \bigcap_{(i,j) \in I \times J} A_i \cup B_j.$$

Cependant, si on fixe B , alors

$$\left(\bigcap_{i \in I} A_i \right) \cup B = \bigcap_{i \in I} (A_i \cup B).$$

On interprète $A \cap B$ comme « A **et** B se réalisent » ; $A \cup B$ comme « A **ou** B se réalisent ». On appelle ${}^c A$ le complémentaire de A . L'évènement ${}^c A$ s'interprète comme « l'évènement A ne se réalise pas ».

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

Une famille d'ensembles $(A_i)_{i \in I}$ est formée d'*ensembles (deux à deux) disjoints* si pour deux indices différents $i \neq j$, A_i et A_j sont disjoints.

Une *partition de Ω* est une collection finie d'ensembles A_1, \dots, A_n deux à deux disjoints telle que $\cup_{i \in I} A_i = \Omega$. On parle de *partition dénombrable* si $(A_i)_{i \in I}$ est une collection dénombrable d'ensembles deux à deux disjoints telle que $\Omega = \cup_{i \in I} A_i$.

Rappels sur les séries.— Si $(u_n)_{n \geq 0}$ est une suite numérique (à valeurs complexes), la série de terme générale (u_n) est la suite de ses sommes partielles $(S_n)_{n \geq 0}$ où

$$S_n = u_0 + \dots + u_n = \sum_{k=0}^n u_k.$$

Lorsque cette suite est convergente, on dit que la série est convergente et on note sa limite

$$\sum_{n \geq 0} u_n.$$

Lorsque (u_n) est à valeurs positives, alors (S_n) est une suite croissante. Donc, ou bien elle est majorée et convergente, ou bien elle n'est pas majorée et alors la suite tend vers l'infini. En tout état de cause, on peut toujours donner un sens à la série

$$\sum_{n \geq 0} u_n \in [0, +\infty].$$

On a les propriétés suivantes :

- **Commutativité.** Si $\sigma : \mathbb{N} \rightarrow \mathbb{N}$ est une bijection, alors

$$\sum_{n \geq 0} u_n = \sum_{n \geq 0} u_{\sigma(n)}.$$

- **Sommation par paquets.** Si (n_k) est strictement croissante, alors

$$\sum_{n \geq 0} u_n = \sum_k \left(\sum_{n=n_k}^{n_{k+1}-1} u_n \right).$$

Plus généralement, si (A_k) est une partition dénombrable de \mathbb{N} , alors

$$\sum u_n = \sum_k \left(\sum_{n \in A_k} u_n \right).$$

Si (u_n) n'est pas de signe constant, alors ces propriétés restent vraies **uniquement** lorsque la série est absolument convergente, c'est-à-dire si la série de terme général $(|u_n|)_n$ est finie.

Du coup, si I est un ensemble au plus dénombrable, on peut indexer une famille $(u_i)_{i \in I}$ par des entiers $(u_n)_n$. Si cette série est absolument convergente ou de signe constant, alors la somme ne dépend pas de l'ordre de sommation, et on peut l'écrire $\sum_{i \in I} u_i$; si la série n'est pas absolument convergente alors la notation « $\sum_{i \in I} u_i$ » n'a pas de sens.

Si $(u_n)_{n \geq 0}$ et $(v_n)_{n \geq 0}$ sont deux suites positives, alors

$$\left(\sum_{n \geq 0} u_n \right) \cdot \left(\sum_{k \geq 0} v_k \right) = \sum_{\ell \geq 0} \left(\sum_{n+k=\ell} u_n v_k \right) = \sum_{n \geq 0} u_n \cdot \left(\sum_{k \geq 0} v_k \right) = \sum_{k \geq 0} v_k \cdot \left(\sum_{n \geq 0} u_n \right).$$

Du coup, on peut écrire plus simplement

$$\left(\sum_{n \geq 0} u_n \right) \cdot \left(\sum_{k \geq 0} v_k \right) = \sum_{n, k \geq 0} u_n v_k.$$

Comme ci-dessus, ces identités restent vraies pour des suites de signe quelconque **uniquement** si les deux séries sont absolument convergentes. Si on a deux familles au plus dénombrables $(u_i)_{i \in I}$ et $(v_j)_{j \in J}$, si elles sont toutes deux de signe constant ou absolument intégrables alors on peut écrire

$$\left(\sum_{i \in I} u_i \right) \cdot \left(\sum_{j \in J} v_j \right) = \sum_{(i, j) \in I \times J} u_i v_j.$$

On peut aussi utiliser les formules précédentes. Si aucune des conditions ci-dessus n'est satisfaite, alors la notation $\ll \sum_{(i, j) \in I \times J} u_i v_j \gg$ n'a pas de sens.

On conclut cette parenthèse en rappelant quelques résultats bien pratiques :

– Si $z \in \mathbb{C}$, alors

$$\sum_{n \geq 0} \frac{z^n}{n!} = \exp(z).$$

– Si $z \in \mathbb{C}$ et $|z| < 1$ alors

$$\sum_{n \geq 0} z^n = \frac{1}{1 - z}.$$

Retournons maintenant aux probabilités et établissons quelques propriétés générales de \mathbb{P} .

Proposition 1.7 *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité associé à une distribution $(p_\omega)_{\omega \in \Omega}$. Alors les propriétés suivantes sont vérifiées.*

1. $\mathbb{P}(\Omega) = 1$; $\mathbb{P}(\emptyset) = 0$.
2. Si A et B sont disjoints alors $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B]$. Plus généralement, si $(A_i)_{i \in I}$ est une famille au plus dénombrable d'évènements disjoints, alors

$$\mathbb{P}[\cup_{i \in I} A_i] = \sum_{i \in I} \mathbb{P}[A_i]. \quad (1)$$

3. Pour tout évènement A , $\mathbb{P}(A) + \mathbb{P}(^c A) = 1$. Plus généralement, si $(A_i)_{i \in I}$ est une partition au plus dénombrable d'évènements alors

$$\sum_{i \in I} \mathbb{P}[A_i] = 1.$$

4. Si A et B sont deux évènements tels que $A \subset B$, alors $\mathbb{P}[A] \leq \mathbb{P}[B]$.

5. Si $(A_n)_{n \geq 1}$ est une suite croissante d'évènements (i.e $A_n \subset A_{n+1}$), alors

$$\mathbb{P}[\cup_{n \geq 1} A_n] = \lim_{n \rightarrow \infty} \mathbb{P}[A_n].$$

Par ailleurs, si (B_n) est une suite décroissante d'évènements ($B_{n+1} \subset B_n$ pour tout n) alors

$$\mathbb{P}[\cap_{n \geq 1} B_n] = \lim_{n \rightarrow \infty} \mathbb{P}[B_n].$$

DÉMONSTRATION. On utilise les propriétés des séries de suites positives.

1. Par définition de la distribution de probabilité et de \mathbb{P} , on a

$$\mathbb{P}[\Omega] = \sum_{\omega \in \Omega} p_\omega = 1$$

et $\mathbb{P}[\emptyset] = 0$.

2. Supposons A et B disjoints, alors

$$\mathbb{P}[A \cup B] = \sum_{\omega \in A \cup B} p_\omega = \left(\sum_{\omega \in A} p_\omega \right) + \left(\sum_{\omega \in B} p_\omega \right) = \mathbb{P}[A] + \mathbb{P}[B].$$

Plus généralement, si $(A_i)_{i \in I}$ est une famille dénombrable d'évènements disjoints, alors

$$\begin{aligned} \mathbb{P}[\cup_{i \in I} A_i] &= \sum_{\omega \in \cup A_i} p_\omega \\ &= \sum_{i \in I} \left(\sum_{\omega \in A_i} p_\omega \right) \\ &= \sum_{i \in I} \mathbb{P}[A_i]. \end{aligned}$$

3. On remarque que A et ${}^c A$ sont deux évènements disjoints tels que $A \cup {}^c A = \Omega$. On a donc d'après ce qui précède

$$\mathbb{P}[A] + \mathbb{P}[{}^c A] = \mathbb{P}[\Omega] = 1.$$

De même, pour une partition (A_i) au plus dénombrable de Ω ,

$$\sum_{i \in I} \mathbb{P}[A_i] = \mathbb{P}[\Omega] = 1.$$

4. Notons $B \setminus A = C$. Alors $\{A, C\}$ est une partition de B et $\mathbb{P}[C]$ étant positif, on obtient

$$\mathbb{P}[B] = \mathbb{P}[A] + \mathbb{P}[C] \geq \mathbb{P}[A].$$

5. Si $(A_n)_{n \geq 1}$ est une suite croissante d'évènements, alors on remarque que, pour tout n , $A_n \subset A_{n+1} \subset A$, où on a posé $A = \cup A_i$ donc $(\mathbb{P}[A_n])_n$ est une suite croissante majorée par $\mathbb{P}[A]$. Notons $C_0 = A_1$ et $C_n = A_{n+1} \setminus A_n$ pour $n \geq 1$. Par construction, les (C_k) sont

disjoints. On montre par récurrence que $A_{n+1} = \cup_{0 \leq k \leq n} C_k$ et donc $A = \cup C_n$. Alors (C_n) est une partition dénombrable de A donc

$$\begin{aligned} \mathbb{P}[\cup A_n] &= \sum_n \mathbb{P}[C_n] \\ &= \lim_{n \rightarrow \infty} \sum_{0 \leq k \leq n} \mathbb{P}[C_k] \\ &= \lim_{n \rightarrow \infty} \mathbb{P}[\cup_{0 \leq k \leq n} C_k] \\ &= \lim_{n \rightarrow \infty} \mathbb{P}[A_n]. \end{aligned}$$

On remarque si (B_n) est décroissante, alors $({}^c B_n)$ est croissante, donc, en remarquant que ${}^c(\cap B_n) = \cup {}^c B_n$, on obtient

$$\begin{aligned} \lim \mathbb{P}[B_n] &= 1 - \lim \mathbb{P}[{}^c B_n] \\ &= 1 - \mathbb{P}[\cup {}^c B_n] \\ &= 1 - \mathbb{P}[{}^c(\cap B_n)] \\ &= \mathbb{P}[\cap B_n]. \end{aligned}$$

■

A la place de se donner une distribution de probabilité, on aurait pu se donner une mesure de probabilité, c'est-à-dire une application $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ qui vérifie les trois propriétés suivantes :

- (i) $\mathbb{P}[\Omega] = 1$.
- (ii) Si A et B sont disjoints alors $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B]$.
- (iii) Si (A_n) est une suite d'événements disjoints, alors

$$\mathbb{P}[\cup A_n] = \sum \mathbb{P}[A_n].$$

Comme Ω est au plus dénombrable, on récupère une distribution de probabilité en posant $p_\omega = \mathbb{P}[\{\omega\}]$. On a bien sûr $p_\omega \geq 0$ pour tout $\omega \in \Omega$, et comme $\{\omega\}_{\omega \in \Omega}$ est une partition au plus dénombrable de Ω , on a aussi

$$\sum p_\omega = \mathbb{P}[\Omega] = 1.$$

Remarque 1.8 *La condition (ii) découle de (iii). Pour cela, il suffit de poser $A_1 = A$, $A_2 = B$ et $A_n = \emptyset$ pour tout $n \geq 3$.*

2 Variables aléatoires

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité défini sur un ensemble Ω au plus dénombrable.

Définition 2.1 (variable aléatoire) *Une variable aléatoire (v.a.) est une application $X : \Omega \rightarrow \mathbb{R}$. On utilise en général des lettres majuscules de la fin de l'alphabet pour les désigner (X, Y, Z).*

Comme Ω est au plus dénombrable c'est aussi le cas de $X(\Omega)$ — les valeurs prises par X . On pourra les noter $\{x_n, n \in \mathbb{N}\}$ le cas échéant (mais pas forcément dans l'ordre croissant).

Exemple 1.3.— On définit

$$\Omega = \{1, 2, 3, 4, 5, 6\}^2 = \{(i, j) \in \{1, \dots, 6\}^2\}$$

muni de la loi uniforme. On pose alors $X : \Omega \rightarrow \mathbb{R}$ défini par $X(i, j) = i + j$.

Fonctions indicatrices.— Les fonctions indicatrices fournissent une très large classe de v.a. qui jouent un rôle très important. Etant donné un événement A , on définit sa fonction indicatrice ainsi :

$$\begin{aligned} \mathbb{I}_A : \quad \Omega &\rightarrow \{0, 1\} \\ \omega \in A &\mapsto 1 \\ \omega \notin A &\mapsto 0 \end{aligned}$$

Proposition 2.2 Soient A, B des événements.

1. On a $\mathbb{I}_{A \cap B} = \mathbb{I}_A \times \mathbb{I}_B$.
2. Si $A \cap B = \emptyset$, alors $\mathbb{I}_{A \cup B} = \mathbb{I}_A + \mathbb{I}_B$.
3. $\mathbb{I}_{c_A} = 1 - \mathbb{I}_A$.
4. Si $(A_i)_{i \in I}$ est une famille d'événements deux à deux disjoints, alors $\mathbb{I}_{\cup A_i} = \sum \mathbb{I}_{A_i}$. En particulier, si (A_i) est une partition de Ω , alors $1 = \sum_{i \in I} \mathbb{I}_{A_i}$.

Notations.— Si X, Y sont des v.a., $x \in \mathbb{R}$ et $A, B \in \mathcal{A}$. On note

$$\begin{aligned} (X \in A) &= X^{-1}(A) = \{\omega \in \Omega, X(\omega) \in A\}, \\ (X \in A; Y \in B) &= X^{-1}(A) \cap Y^{-1}(B), \end{aligned}$$

ainsi que

$$\begin{aligned} (X = x) &= X^{-1}(\{x\}) = \{\omega \in \Omega, X(\omega) = x\}, \\ (X \leq x) &= X^{-1}(]-\infty, x]) = \{\omega \in \Omega, X(\omega) \leq x\}, \\ (X < x) &= X^{-1}(]-\infty, x[) = \{\omega \in \Omega, X(\omega) < x\}, \\ (X \geq x) &= X^{-1}([x, \infty[) = \{\omega \in \Omega, X(\omega) \geq x\}, \\ (X > x) &= X^{-1}(]x, \infty]) = \{\omega \in \Omega, X(\omega) > x\}, \end{aligned}$$

etc.

Les fonctions indicatrices permettent de décrire toutes les v.a. qui prennent au plus une quantité dénombrable de valeurs :

Lemme 2.3 Soit X une v.a. définie sur un espace au plus dénombrable. Soit $\{x_i\}_{i \in I}$ les valeurs prises par X . Alors $(\{X = x_i\})_{i \in I}$ est une partition au plus dénombrable de Ω et

$$X = \sum_{i \in I} x_i \mathbb{I}_{(X=x_i)}.$$

DÉMONSTRATION. Si $x_i \neq x_j$, alors $X^{-1}(x_i) \cap X^{-1}(x_j) = \emptyset$ par définition d'une fonction, donc $\{(X = x_i), i \in I\}$ est une famille d'ensembles disjoints. De plus, pour tout $\omega \in \Omega$, $X(\omega) \in X(\Omega)$ donc il existe un (unique) $i \in I$ tel que $X(\omega) = x_i$. Il s'agit bien d'une partition de Ω .

Soit $\omega \in \Omega$. Il existe un unique $i_0 \in I$ tel que $X(\omega) = x_{i_0}$. Par conséquent, on a $\mathbb{1}_{(X=x_{i_0})}(\omega) = 1$ et $\mathbb{1}_{(X=x_i)}(\omega) = 0$ si $i \neq i_0$. Du coup,

$$\sum_{i \in I} x_i \mathbb{1}_{(X=x_i)}(\omega) = x_{i_0} = X(\omega).$$

■

Définition et théorème 2.4 Soit X une v.a. définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ et notons $\{x_i, i \in I\}$ les valeurs prises par X . Pour $i \in I$, les quantités $p_i = \mathbb{P}[X = x_i]$ définissent une distribution de probabilité sur $X(\Omega) \subset \mathbb{R}$ dont la loi associée s'appelle la loi de X que l'on note \mathbb{P}_X .

Si X est définie sur un espace qui peut être éventuellement abstrait, sa loi, quant à elle, est définie sur \mathbb{R} . Cela permet en général de faire des calculs une fois que la loi de X est connue.

DÉMONSTRATION. Par construction, $p_i \geq 0$ pour tout $i \in X(\Omega)$. Par ailleurs, comme les $\{X = x_i\}, i \in I$ forment une partition au plus dénombrable de Ω (voir le lemme 2.3), on a

$$\sum_{i \in I} p_i = \sum \mathbb{P}[X = x_i] = 1.$$

■

Déterminer la loi d'une v.a. revient donc à calculer toutes les probabilités $\mathbb{P}[X = x_i]$ et à vérifier que la somme vaut bien 1, afin de s'assurer que l'on n'a pas omis de valeurs ni fait d'erreur(s) de calculs!

Exemple 1.3.— Notre v.a. prend ses valeurs dans $V = \{2, \dots, 12\}$. Pour calculer sa loi, nous déterminons d'abord les événements $(X = v)$ pour $v \in V$: On a

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Nous pouvons en déduire la loi de X . En effet, comme la loi sur Ω est uniforme, il suffit de compter le nombre de jets de dé qui donnent la valeur donnée. Ainsi, on trouve $\mathbb{P}[X = 2] = \mathbb{P}[X =$

$12] = 1/36$, $\mathbb{P}[X = 3] = \mathbb{P}[X = 11] = 2/36 = 1/18$, $\mathbb{P}[X = 4] = \mathbb{P}[X = 10] = 3/36 = 1/12$, $\mathbb{P}[X = 5] = \mathbb{P}[X = 9] = 4/36 = 1/9$, $\mathbb{P}[X = 6] = \mathbb{P}[X = 8] = 5/36$ et $\mathbb{P}[X = 7] = 6/36 = 1/6$.

Exemple 1.5.— On considère $\Omega = \{P, F\}^n$ muni de la distribution (p_ω) telle que $p_\omega = p^k(1-p)^{n-k}$ si ω contient exactement k piles, où $p \in [0, 1]$. Posons, pour $j = 1, \dots, n$, $X_j(\omega) = 1$ si $\omega_j = P$ et $X_j(\omega) = 0$ sinon. Autrement dit, $X_j = \mathbb{1}_{(\omega_j=P)}$. On note enfin $s_n(\omega) = \sum X_j$ qui compte le nombre de piles dans le tirage.

La v.a. s_n prend ses valeurs dans $V = \{0, \dots, n\}$. La loi de s_n est la loi binomiale $\mathcal{B}(n, p)$. En effet, l'événement $(s_n = k)$ a exactement C_n^k éléments (on choisit k piles parmi n possibilités) et, pour tout $\omega \in (s_n = k)$, on a $p_\omega = p^k(1-p)^{n-k}$. Par conséquent

$$\mathbb{P}[s_n = k] = C_n^k p^k (1-p)^{n-k}.$$

3 Espérance, variance et écart-type

3.1 Espérance

L'espérance d'une v.a. généralise la notion de moyenne. Si on joue à pile ou face, avec une pièce qui fait pile avec probabilité $p \in [0, 1]$ et que l'on gagne s euros sur piles et que l'on perd t euros sinon, la première question est de savoir si on a des « chances » de gagner ou non : autrement dit, on a une probabilité p de gagner s et $(1-p)$ de perdre t , soit au total, on s'attend à un gain moyen de $ps - (1-p)t$. Pour accepter de jouer, il faut que ce gain soit positif ou nul : s'il est strictement positif, alors nous sommes avantagés. et si le résultat est nul, alors le jeu est équilibré.

Définition 3.1 (Espérance) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité, où Ω est au plus dénombrable et soit $X : \Omega \rightarrow \mathbb{R}$ une v.a.

– (cas positif) Si X ne prend que des valeurs positives, on pose

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega)p_\omega$$

qui est une série à termes positifs. Donc $\mathbb{E}[X] \in [0, \infty]$.

– (cas général) Si $\mathbb{E}[|X|]$ est finie alors on pose

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

Si $\mathbb{E}[|X|]$ n'est pas finie, l'espérance de X n'est pas définie.

Testons cette définition sur quelques exemples.

Avec l'exemple précédent, on pose $G : \{P, F\} \rightarrow \mathbb{R}$ la fonction gain, donc $G(P) = s$ et $G(F) = -t$. On obtient $\mathbb{E}[G] = sp - (1-p)t$.

Exemple 1.3.— Dans ce cas, X prend un nombre fini de valeurs positives. On a

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=2}^{12} k\mathbb{P}[X = k] \\ &= 2(1/36) + 3(2/36) + 4(3/36) + 5(4/36) + 6(5/36) + 7(6/36) + 8(5/36) \\ &\quad + 9(4/36) + 10(3/36) + 11(2/36) + 12(1/36) \\ &= (2 + 6 + 12 + 20 + 30 + 42 + 40 + 36 + 30 + 22 + 12)/36 \\ &= 242/36 = 121/18.\end{aligned}$$

Exemple 1.4.— Lorsque l'on joue à pile ou face et que l'on note $X : \{P, F\} \rightarrow \{0, 1\}$ la v.a. définie par $X(P) = 1$ et $X(F) = 0$, on obtient $\mathbb{E}[X] = 1 \cdot \mathbb{P}[P] + 0 \cdot \mathbb{P}[F] = p$.

Exemple 1.5.— Si on s'intéresse au nombre de piles après n lancers, on obtient

$$\mathbb{E}[X] = \sum_{k=0}^n k\mathbb{P}[s_n = k] = \sum_{k=1}^n C_n^k k p^k (1-p)^{n-k}.$$

Nous ne sommes pas très avancés !

Voici deux propriétés pratiques.

Lemme 3.2 1. Soient X, Y deux v.a. positives. Si $X \leq Y$ alors $\mathbb{E}[X] \leq \mathbb{E}[Y]$. En particulier, si $\mathbb{E}[Y]$ est finie, il en est de même de $\mathbb{E}[X]$.

2. Si $X = \sum_{j \geq 0} X_j$, avec $(X_j)_{j \geq 0}$ une suite de v.a. positives alors

$$\mathbb{E}[X] = \sum \mathbb{E}[X_j].$$

DÉMONSTRATION. On applique les propriétés des séries positives.

1. On a, pour tout $\omega \in \Omega$, $0 \leq X(\omega) \leq Y(\omega)$ donc $0 \leq X(\omega)p_\omega \leq Y(\omega)p_\omega$ et en sommant $0 \leq \mathbb{E}[X] \leq \mathbb{E}[Y]$.

2. Par définition, on a

$$\begin{aligned}\mathbb{E}[X] &= \sum_{\omega \in \Omega} \left(\sum_{j \geq 0} X_j(\omega) \right) p_\omega \\ &= \sum_{\omega \in \Omega} \left(\sum_{j \geq 0} X_j(\omega) p_\omega \right) \\ &= \sum_{j \geq 0} \left(\sum_{\omega \in \Omega} X_j(\omega) p_\omega \right) \\ &= \sum_{j \geq 0} \mathbb{E}[X_j].\end{aligned}$$

■

Lemme 3.3 *Si l'espérance de X est bien définie et si on note $X(\Omega) = \{x_i\}_{i \in I}$ où I est au plus dénombrable, alors*

$$\mathbb{E}[X] = \sum_{i \in I} x_i \mathbb{P}[X = x_i].$$

En particulier, pour un événement $A \in \mathcal{A}$, on a $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}[A]$ et si $X = \lambda$ est constante, alors $\mathbb{E}[X] = \lambda$.

DÉMONSTRATION. D'après le lemme 2.3, on a

$$X = \sum_{i \in I} x_i \mathbb{1}_{(X=x_i)}$$

donc le lemme 3.2 implique

$$\mathbb{E}[X] = \sum_{i \in I} \mathbb{E}[x_i \mathbb{1}_{(X=x_i)}].$$

Or, pour tout $i \in I$, on a

$$\mathbb{E}[x_i \mathbb{1}_{(X=x_i)}] = \sum_{\omega \in \Omega} x_i \mathbb{1}_{(X=x_i)}(\omega) p_\omega = x_i \sum_{\omega \in (X=x_i)} p_\omega = x_i \mathbb{P}[X = x_i]$$

donc

$$\mathbb{E}[X] = \sum_{i \in I} x_i \mathbb{P}[X = x_i].$$

■

Remarque 3.4 *D'après le lemme 3.3, l'espérance d'une v.a., lorsqu'elle est définie, ne dépend que de la loi de X et non de X elle-même. Autrement dit, si X et Y sont des v.a. de même loi, alors $\mathbb{E}[|X|] = \mathbb{E}[|Y|]$ et, lorsque cette espérance est finie, on a aussi $\mathbb{E}[X] = \mathbb{E}[Y]$.*

L'espérance d'une v.a. est donc « bien définie » si X ne prend que des valeurs positives ou si $\mathbb{E}[|X|]$ est finie. Cela nous conduit à la définition suivante :

Définition 3.5 (Espace L^1) *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. On note L^1 l'ensemble des variables aléatoires X telles que $\mathbb{E}[|X|]$ est finie. Si $X \in L^1$, on dit que X est intégrable.*

Exemple 3.1 *Si X est une v.a. bornée, alors $X \in L^1$. En effet, si M est une constante telle que $|X(\omega)| \leq M$ pour tout $\omega \in \Omega$, alors le lemme 3.2 implique $\mathbb{E}[|X|] \leq \mathbb{E}[M] = M$. Plus précisément, on a*

$$\mathbb{E}[|X|] = \sum_{\omega \in \Omega} |X(\omega)| p_\omega \leq \sum_{\omega \in \Omega} M p_\omega \leq M \sum_{\omega \in \Omega} p_\omega = M.$$

Lemme 3.6 Si $f : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction bornée, alors $f(X) \in L^1$ et

$$\mathbb{E}[f(X)] = \sum_{i \in I} f(x_i) \mathbb{P}[X = x_i].$$

Ceci implique aussi que $\mathbb{E}[f(X)]$ ne dépend que de la loi de X .

DÉMONSTRATION. Comme f est bornée, $f(X)$ aussi, donc $f(X) \in L^1$. On remarque que

$$f(X) = \sum_{i \in I} f(x_i) \mathbb{I}_{(X=x_i)}$$

donc

$$\mathbb{E}[f(X)] = \sum_{i \in I} \mathbb{E}[f(x_i) \mathbb{I}_{(X=x_i)}] = \sum_{i \in I} f(x_i) \mathbb{P}[X = x_i].$$

■

Théorème 3.7 L'espace L^1 est un espace vectoriel et $\mathbb{E} : L^1 \rightarrow \mathbb{R}$ est une forme linéaire positive, c'est-à-dire

- si $X, Y \in L^1$ et $\lambda \in \mathbb{R}$, alors $\mathbb{E}[X + \lambda Y] = \mathbb{E}[X] + \lambda \mathbb{E}[Y]$;
- si $X \geq 0$ alors $\mathbb{E}[X] \geq 0$.

DÉMONSTRATION. Soient X, Y positives et $\lambda \geq 0$.

$$\begin{aligned} \mathbb{E}[X + \lambda Y] &= \sum_{\omega \in \Omega} (X(\omega) + \lambda Y(\omega)) p_\omega \\ &= \sum_{\omega \in \Omega} (X(\omega) p_\omega + \lambda Y(\omega) p_\omega) \\ &= \left(\sum_{\omega \in \Omega} X(\omega) p_\omega \right) + \lambda \left(\sum_{\omega \in \Omega} Y(\omega) p_\omega \right) \\ &= \mathbb{E}[X] + \lambda \mathbb{E}[Y]. \end{aligned}$$

Par conséquent, si $X, Y \in L^1$ et $\lambda \in \mathbb{R}$, en utilisant le fait que $|X(\omega) + \lambda Y(\omega)| \leq |X(\omega)| + |\lambda| \cdot |Y(\omega)|$ pour tout $\omega \in \Omega$, on aura

$$\begin{aligned} \mathbb{E}[|X + \lambda Y|] &= \sum_{\omega \in \Omega} |X(\omega) + \lambda Y(\omega)| p_\omega \\ &\leq \sum_{\omega \in \Omega} (|X(\omega)| + |\lambda| \cdot |Y(\omega)|) p_\omega \\ &\leq \mathbb{E}[|X| + |\lambda| |Y|] = \mathbb{E}[|X|] + |\lambda| \mathbb{E}[|Y|] \end{aligned}$$

ce qui implique que $(X + \lambda Y) \in L^1$. Donc L^1 est un sous-espace vectoriel de l'espace des applications $f : \Omega \rightarrow \mathbb{R}$.

Par ailleurs, le même calcul que dans le cas positif montre que \mathbb{E} est linéaire. Par définition de \mathbb{E} , si $X \geq 0$ alors $\mathbb{E}[X] \geq 0$. ■

Exemple 1.5.— On rappelle que nous avons écrit $s_n = \sum_{j=1}^n X_j$, où $X_j = 1$ si $\omega_j = P$ et $X_j = 0$ sinon. Dans ce cas, en utilisant ce qu'on a montré pour l'Exemple 1.4, on obtient par linéarité

$$\mathbb{E}[s_n] = \sum_{j=1}^n \mathbb{E}[X_j] = \sum_{j=1}^n p = np.$$

Proposition 3.8 *On a les propriétés suivantes.*

1. Si $X, Y \in L^1$ et si $X \leq Y$ alors $\mathbb{E}[X] \leq \mathbb{E}[Y]$.
2. Si $X \in L^1$, alors $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.
3. Si $X \geq 0$ et $\mathbb{E}[X] = 0$ alors $X(\omega) = 0$ sur $\{\omega \in \Omega, p_\omega > 0\}$.

DÉMONSTRATION.

1. Si $X \leq Y$, on aura $(Y - X) \in L^1$, $(Y - X) \geq 0$ donc

$$\mathbb{E}[Y] - \mathbb{E}[X] = \mathbb{E}[Y - X] \geq 0.$$

2. On remarque que, pour toute série numérique $(a_n)_{n \geq 0}$ absolument convergente, on a

$$\left| \sum_{n \geq 0} a_n \right| \leq \sum_{n \geq 0} |a_n|$$

car, pour tout $k \geq 1$, on a clairement

$$\left| \sum_{n=0}^k a_n \right| \leq \sum_{n=0}^k |a_n| \leq \sum_{n \geq 0} |a_n|.$$

Comme cette relation est vraie pour tout k , on peut passer à la limite. Du coup,

$$\begin{aligned} |\mathbb{E}[X]| &= \left| \sum_{\omega \in \Omega} X(\omega) p_\omega \right| \\ &\leq \sum_{\omega \in \Omega} |X(\omega)| p_\omega \\ &\leq \mathbb{E}[|X|]. \end{aligned}$$

3. Par définition, on a $\mathbb{E}[X] = \sum X(\omega) p_\omega$ qui est une série à termes positifs. Donc, si la somme est nulle, chaque terme est nul. Ceci conduit à $X(\omega) p_\omega = 0$ pour tout $\omega \in \Omega$, donc $X(\omega) = 0$ si $p_\omega > 0$. ■

3.2 Variance

Définition 3.9 (Espace L^2) *On note L^2 l'espace des v.a X telles que $\mathbb{E}[X^2]$ est finie. Si $X \in L^2$, on dit que X est de carré intégrable.*

Proposition 3.10 *L'espace L^2 est un sous-espace vectoriel de L^1 et*

$$(X, Y) \in L^2 \times L^2 \mapsto \mathbb{E}[XY]$$

est bien définie et une forme bilinéaire symétrique positive.

De plus, si on suppose que $p_\omega > 0$ pour tout $\omega \in \Omega$, alors c'est une forme définie positive.

Rappels sur l'algèbre bilinéaire.— Soit E un espace vectoriel réel. Une *forme bilinéaire symétrique* est une application $b : E \times E \rightarrow \mathbb{R}$ qui vérifie les propriétés suivantes.

B1 pour tous $x, y, z \in E$ et $\lambda \in \mathbb{R}$, $b(x + \lambda y, z) = b(x, z) + \lambda b(y, z)$.

B2 pour tous $x, y \in E$, $b(x, y) = b(y, x)$.

On associe à b la *forme quadratique* $q : E \rightarrow \mathbb{R}$ définie par $q(x) = b(x, x)$. On dit qu'une forme bilinéaire symétrique b est *positive* si $b(x, x) \geq 0$ pour tout $x \in E$ (ou $q(x) \geq 0$) et on dit que b est *définie positive* si $q(x) = 0$ seulement pour $x = 0$.

Inégalité de Cauchy-Schwarz.— *Si b est une forme bilinéaire symétrique et positive sur un espace vectoriel réel, alors, pour tous $x, y \in E$,*

$$|b(x, y)| \leq \sqrt{q(x)} \times \sqrt{q(y)}.$$

DÉMONSTRATION DE LA PROPOSITION 3.10. Soient X, Y deux v.a. On a, pour tout $\omega \in \Omega$,

$$0 \leq (|X(\omega)| - |Y(\omega)|)^2 = X^2(\omega) - 2|X(\omega)Y(\omega)| + Y^2(\omega)$$

donc

$$|X(\omega)Y(\omega)| \leq \frac{1}{2}(X^2(\omega) + Y^2(\omega)).$$

Par conséquent, si $X, Y \in L^2$ alors $\mathbb{E}[|XY|] \leq (1/2)(\mathbb{E}[X^2] + \mathbb{E}[Y^2])$ donc $(XY) \in L^1$ d'après le théorème 3.7 et le lemme 3.2.

Supposons maintenant $X, Y \in L^2$ et $\lambda \in \mathbb{R}$. On a $(X + \lambda Y)^2 = X^2 + 2\lambda XY + \lambda^2 Y^2$. Comme chaque terme de droite est dans L^1 , le théorème 3.7 nous dit $(X + \lambda Y) \in L^2$, donc L^2 est un espace vectoriel.

Montrons que $(X, Y) \in L^2 \times L^2 \mapsto \mathbb{E}[XY]$ est bilinéaire symétrique. D'après ce qui précède, le résultat est bien défini. Par commutativité de la multiplication sur \mathbb{R} , on a bien $\mathbb{E}[XY] = \mathbb{E}[YX]$. De plus, Si $X, Y, Z \in L^2$ et $\lambda \in \mathbb{R}$, alors $(X + \lambda Y) \in L^2$ donc $XZ, \lambda YZ, (X + \lambda Y)Z \in L^1$ et

$$\begin{aligned} \mathbb{E}[(X + \lambda Y)Z] &= \sum_{\omega \in \Omega} [(X(\omega) + \lambda Y(\omega))Z(\omega)]p_\omega \\ &= \sum_{\omega \in \Omega} [X(\omega)Z(\omega)p_\omega + \lambda Y(\omega)Z(\omega)p_\omega] \\ &= \left(\sum_{\omega \in \Omega} X(\omega)Z(\omega)p_\omega \right) + \lambda \left(\sum_{\omega \in \Omega} Y(\omega)Z(\omega)p_\omega \right) \\ &= \mathbb{E}[XZ] + \lambda \mathbb{E}[YZ] \end{aligned}$$

ce qui montre que c'est bien une forme bilinéaire symétrique. De plus $X^2 \geq 0$ donc $\mathbb{E}[X^2] \geq 0$ et cette forme est positive. Par l'inégalité de Cauchy-Schwarz appliquée à $|X|$ et $\mathbb{1}_\Omega$, on trouve que $\mathbb{E}[|X|] \leq \sqrt{\mathbb{E}[X^2]}$ donc $L^2 \subset L^1$.

Si $\mathbb{E}[X^2] = 0$ alors, pour tout $\omega \in \Omega$, on a $X(\omega)p_\omega = 0$. En supposant $p_\omega > 0$ pour tout ω , il vient $X = 0$. ■

Définition 3.11 (Variance, covariance et écart-type) On définit sur L^2 les fonctions suivantes.

– La covariance :

$$\begin{aligned} cov : L^2 \times L^2 &\rightarrow \mathbb{R} \\ (X, Y) &\mapsto \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \end{aligned}$$

– La variance : pour $X \in L^2$, on pose $var(X) = cov(X, X)$.

– L'écart-type : pour $X \in L^2$, on pose $\sigma(X) = \sqrt{var(X)}$.

Ces notions sont bien définies par la proposition 3.10.

Proposition 3.12 La covariance définit une forme bilinéaire symétrique et positive sur L^2 de forme quadratique la variance. Par suite, on a, pour $X, Y \in L^2$,

$$|cov(X, Y)| \leq \sqrt{var(X)}\sqrt{var(Y)}.$$

De plus, si on suppose que $p_\omega > 0$ pour tout $\omega \in \Omega$, alors $\sigma(X) = 0$ si et seulement si X est constante (pour toute constante λ , $var(\lambda) = 0$ sans hypothèses supplémentaires).

DÉMONSTRATION. Montrons que si $X, Y, Z \in L^2$ et $\lambda \in \mathbb{R}$ alors $cov(X + \lambda Y, Z) = cov(X, Z) + \lambda cov(Y, Z)$. Tout d'abord, par linéarité de \mathbb{E} , on a

$$\mathbb{E}[X + \lambda Y] = \mathbb{E}[X] + \lambda \mathbb{E}[Y]$$

donc

$$\begin{aligned} (X + \lambda Y - \mathbb{E}[X + \lambda Y])(Z - \mathbb{E}[Z]) &= [(X + \lambda Y) - (\mathbb{E}[X] + \lambda \mathbb{E}[Y])](Z - \mathbb{E}[Z]) \\ &= [(X - \mathbb{E}[X]) + \lambda(Y - \mathbb{E}[Y])](Z - \mathbb{E}[Z]) \\ &= (X - \mathbb{E}[X])(Z - \mathbb{E}[Z]) + \lambda(Y - \mathbb{E}[Y])(Z - \mathbb{E}[Z]). \end{aligned}$$

Donc, en utilisant encore la linéarité de \mathbb{E} :

$$\begin{aligned} cov(X + \lambda Y, Z) &= \mathbb{E}[(X - \mathbb{E}[X])(Z - \mathbb{E}[Z]) + \lambda(Y - \mathbb{E}[Y])(Z - \mathbb{E}[Z])] \\ &= \mathbb{E}[(X - \mathbb{E}[X])(Z - \mathbb{E}[Z])] + \lambda \mathbb{E}[(Y - \mathbb{E}[Y])(Z - \mathbb{E}[Z])] \\ &= cov(X, Z) + \lambda cov(Y, Z). \end{aligned}$$

On a clairement $cov(X, Y) = cov(Y, X)$ donc cov est une forme bilinéaire symétrique. En prenant $Y = X$, on trouve

$$cov(X, X) = \mathbb{E}[(X - \mathbb{E}[X])^2] \geq 0$$

car on calcule l'espérance d'une v.a. positive.

Si $X = \lambda$, on a $\mathbb{E}[\lambda] = \lambda$ donc $\text{var}(\lambda) = 0$. Réciproquement, si $\text{var}(X) = 0$ alors on a

$$\sum_{\omega \in \Omega} (X(\omega) - \mathbb{E}[X])^2 p_\omega = 0.$$

Comme chaque terme est positif, cela ne peut arriver que s'ils sont tous nuls, c'est-à-dire si, pour tout $\omega \in \Omega$, on a

$$(X(\omega) - \mathbb{E}[X])^2 p_\omega = 0.$$

Si on suppose $p_\omega > 0$ alors $X(\omega) - \mathbb{E}[X] = 0$ donc X est constante (égale à son espérance!). ■

Proposition 3.13 (Propriétés de la variance) Soient X, Y deux variables aléatoires de carré intégrable et $\lambda, \mu \in \mathbb{R}$.

1. (formule de Huygens) $\text{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ et $\text{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.
2. $\text{var}(\lambda) = 0$.
3. $\text{var}(X + Y) = \text{var}(X) + 2\text{cov}(X, Y) + \text{var}(Y)$.
4. $\text{var}(\lambda X + \mu) = \lambda^2 \text{var}(X)$ et $\sigma(\lambda X + \mu) = |\lambda| \sigma(X)$.

DÉMONSTRATION.

1. On développe la formule définissant la covariance en utilisant la linéarité de l'espérance. Tout d'abord,

$$(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) = XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y].$$

Ensuite, on utilise que l'espérance est un scalaire et que $\mathbb{E}[\lambda] = \lambda$ pour les constantes :

$$\begin{aligned} \text{cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[X\mathbb{E}[Y]] - \mathbb{E}[Y\mathbb{E}[X]] + \mathbb{E}[\mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[Y]\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \end{aligned}$$

En prenant $Y = X$, il vient $\text{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

2. Ce point est établi dans la proposition précédente.
3. Cela découle de la bilinéarité de cov :

$$\begin{aligned} \text{var}(X + Y) = \text{cov}(X + Y, X + Y) &= \text{cov}(X, X + Y) + \text{cov}(Y, X + Y) \\ &= \text{cov}(X, X) + \text{cov}(X, Y) + \text{cov}(Y, X) + \text{cov}(Y, Y) \\ &= \text{var}(X) + 2\text{cov}(X, Y) + \text{var}(Y). \end{aligned}$$

4. Avec les formules précédentes, on a $\text{var}(\mu) = 0$ et

$$\text{cov}(\lambda X, \mu) = \mathbb{E}[\mu \lambda X] - \mathbb{E}[\mu]\mathbb{E}[\lambda X] = \mu \mathbb{E}[\lambda X] - \mu \mathbb{E}[\lambda X] = 0$$

donc $\text{var}(\lambda X + \mu) = \lambda^2 \text{var}(X)$. En prenant la racine carrée, on obtient $\sigma(\lambda X + \mu) = |\lambda| \sigma(X)$. ■

On remarque que la variance ne dépend que de la loi de la variable aléatoire : si X et Y sont de même loi et $X \in L^2$, alors $Y \in L^2$ et $\text{var}(X) = \text{var}(Y)$.

Chapitre II. Conditionnement et indépendance²

Dans ce chapitre, on suppose donné un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$, où Ω est un ensemble fini ou dénombrable d'issues, \mathcal{A} est l'ensemble de toutes les parties de Ω et \mathbb{P} est une probabilité sur \mathcal{A} .

4 Conditionnement

Reprenons notre exemple 1.3.

Exemple 1.3.— On rappelle que $\Omega = \{1, \dots, 6\}^2$ muni de la loi uniforme et $X(i, j) = i + j$. La première personne communique à la seconde $X = 4$. La seconde personne se demande alors quelle est la probabilité que le tirage était $(1, 3)$? L'événement $(X = 4) = \{(1, 3), (2, 2), (3, 1)\}$ contient trois issues. Chacune est équiprobable, donc la probabilité que le tirage était $(1, 3)$ est $1/3$. Il s'agit de la probabilité conditionnelle que le tirage soit $(1, 3)$, sachant que la somme vaut 4.

On peut aussi s'interroger sur la probabilité d'avoir un double. Si on sait que X est impaire, alors la probabilité est nulle puisque, pour un double, X est toujours paire. En revanche si on sait que X est paire, on augmentera les chances d'avoir un double.

On pose plus généralement la

Définition 4.1 (Probabilité conditionnelle) *Etant donnés deux événements A et B , tels que $\mathbb{P}(B) > 0$, la probabilité conditionnelle de A sachant B est*

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

On utilisera souvent la formule qui définit $\mathbb{P}[A|B]$ sous la « forme plate » $\mathbb{P}[A \cap B] = \mathbb{P}[A|B]\mathbb{P}[B]$. On vérifie aisément la

Proposition 4.2 *Si $\mathbb{P}[B] > 0$, alors $\mathbb{P}[\cdot|B]$ est une probabilité qui vérifie $\mathbb{P}[B|B] = 1$.*

DÉMONSTRATION. On vérifie que

$$\mathbb{P}[\Omega|B] = \frac{\mathbb{P}[\Omega \cap B]}{\mathbb{P}[B]} = \frac{\mathbb{P}[B]}{\mathbb{P}[B]} = 1.$$

De même, on a aussi $\mathbb{P}[B|B] = 1$. Si $(A_n)_n$ est une famille d'ensemble disjoints, alors $(B \cap A_n)_n$ aussi donc, comme $(\cup_n A_n) \cap B = \cup_n (A_n \cap B)$, on obtient

$$\mathbb{P}[\cup_n A_n|B] = \frac{\mathbb{P}[\cup_n (A_n \cap B)]}{\mathbb{P}[B]} = \sum_n \frac{\mathbb{P}[A_n \cap B]}{\mathbb{P}[B]} = \sum_n \mathbb{P}[A_n|B].$$

Ceci montre que la probabilité conditionnelle est bien une probabilité, cf. la remarque suivant la proposition 1.7. ■

On a les propriétés suivantes.

Proposition 4.3 1. (première formule de Bayes) Si $\mathbb{P}[A] > 0$ et $\mathbb{P}[B] > 0$, alors

$$\mathbb{P}[B|A] = \mathbb{P}[A|B] \frac{\mathbb{P}[B]}{\mathbb{P}[A]}.$$

2. (formule des probabilités totales) Si $\{B_1, \dots, B_n\}$ forme une partition de telle que $\mathbb{P}[B_i] > 0$, pour $1 \leq i \leq n$, alors

$$\mathbb{P}[A] = \sum_{i=1}^n \mathbb{P}[A|B_i] \mathbb{P}[B_i].$$

3. (seconde formule de Bayes) Sous les mêmes conditions, pour tout $1 \leq i \leq n$, on a

$$\mathbb{P}[B_i|A] = \frac{\mathbb{P}[A|B_i] \mathbb{P}[B_i]}{\sum_{k=1}^n \mathbb{P}[A|B_k] \mathbb{P}[B_k]}.$$

En particulier, si $0 < \mathbb{P}[B] < 1$, alors

$$\mathbb{P}[B|A] = \frac{\mathbb{P}[A|B] \mathbb{P}[B]}{\mathbb{P}[A|B] \mathbb{P}[B] + \mathbb{P}[A|cB] \mathbb{P}[cB]}.$$

4. (formule des probabilités composées) Si $(A_j)_{1 \leq j \leq n}$ sont des événements tels que $\mathbb{P}[A_1 \cap \dots \cap A_n] > 0$, alors

$$\mathbb{P}[A_1 \cap \dots \cap A_n] = \mathbb{P}[A_n | A_1 \cap \dots \cap A_{n-1}] \mathbb{P}[A_{n-1} | A_1 \cap \dots \cap A_{n-2}] \times \dots \times \mathbb{P}[A_2 | A_1] \mathbb{P}[A_1].$$

DÉMONSTRATION.

1. Il suffit de remarquer que $\mathbb{P}[A \cap B] = \mathbb{P}[A|B] \mathbb{P}[B]$ puis de diviser par $\mathbb{P}[A]$.

2. Notons que $A \cap (\cup_i B_i) = \cup_i (A \cap B_i)$. Par l'additivité de \mathbb{P} , on a

$$\mathbb{P}[A] = \mathbb{P}[\cup_{i=1}^n A \cap B_i] = \sum_{i=1}^n \mathbb{P}[A \cap B_i]$$

et d'utiliser n fois la formule ci-dessus.

3. En combinant la formule des probabilités totales avec la première formule de Bayes, on obtient la seconde formule de Bayes :

$$\mathbb{P}[B_i|A] = \mathbb{P}[A|B_i] \frac{\mathbb{P}[B_i]}{\mathbb{P}[A]} = \frac{\mathbb{P}[A|B_i] \mathbb{P}[B_i]}{\sum_{k=1}^n \mathbb{P}[A|B_k] \mathbb{P}[B_k]}.$$

4. On part du membre de droite, et on écrit chacune des probabilités conditionnelles comme une fraction. On s'aperçoit alors que le dénominateur du premier facteur se simplifie avec le numérateur du second facteur, etc., le dernier dénominateur se simplifie avec le dernier terme. Il reste le premier numérateur, qui est bien le membre de gauche. Une démonstration par récurrence fait aussi bien l'affaire. ■

5 Indépendance

Deux événements A et B sont indépendants si la probabilité conditionnelle de A sachant B est égale à la probabilité de A , c'est-à-dire le fait de savoir si B est ou non réalisé ne modifie pas la probabilité que A le soit ; ceci s'écrit $\mathbb{P}[A|B] = \mathbb{P}[A]$, ou encore $\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$.

Définition 5.1 *Deux événements A et B sont dits indépendants si*

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B].$$

En fait cette égalité est vraie dès que $\mathbb{P}[A] \in \{0, 1\}$ ou $\mathbb{P}[B] \in \{0, 1\}$, et si les deux probabilités sont non nulles, cette égalité est bien équivalente à $\mathbb{P}[A|B] = \mathbb{P}[A]$ et à $\mathbb{P}[B|A] = \mathbb{P}[B]$.

Remarque 5.2 Attention! *A et B disjoints entraîne une formule simple pour la probabilité de $A \cup B$, A et B indépendants entraîne une formule simple pour la probabilité de $A \cap B$. Ces deux propriétés n'ont rien à voir l'une avec l'autre. Ne pas confondre!*

Définition 5.3 *Deux v.a. X et Y sont dites indépendantes si pour tous $x, y \in \mathbb{R}$, les événements $\{X = x\}$ et $\{Y = y\}$ sont indépendants.*

Exemple 1.4.— On a $\Omega = \{P, F\}^n$ muni de la loi $p_\omega = p^{s_n(\omega)}(1-p)^{n-s_n(\omega)}$, $p \in [0, 1]$. Ici, nous pouvons considérer que chaque tirage est indépendant des autres : pour $x, y \in \{P, F\}$, on a $\mathbb{P}[\omega_1 = x; \omega_2 = y] = \mathbb{P}[\omega_1 = x] \cdot \mathbb{P}[\omega_2 = y]$.

Exemple 5.1 (Espace produit) *On considère deux espaces de probabilité $(\Omega_j, \mathcal{A}_j, \mathbb{P}_j)$, $j = 1, 2$, et on construit un nouvel espace $(\Omega, \mathcal{A}, \mathbb{P})$ ainsi. On note $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{A} = \mathcal{P}(\Omega)$ et si $\omega = (\omega_1, \omega_2) \in \Omega$, on pose $p_\omega = p_{\omega_1} \times p_{\omega_2}$. On vérifie*

$$\begin{aligned} \sum_{\omega \in \Omega} p_\omega &= \sum_{(\omega_1, \omega_2) \in \Omega} p_{\omega_1} \times p_{\omega_2} \\ &= \sum_{\omega_1 \in \Omega_1} \sum_{\omega_2 \in \Omega_2} p_{\omega_1} \times p_{\omega_2} \\ &= \left(\sum_{\omega_1 \in \Omega_1} p_{\omega_1} \right) \times \left(\sum_{\omega_2 \in \Omega_2} p_{\omega_2} \right) \\ &= 1. \end{aligned}$$

On remarque que si $A \in \mathcal{A}_1$ et $B \in \mathcal{A}_2$, alors $(A \times \Omega_2)$ et $(\Omega_1 \times B)$ sont indépendants. En effet, on a

$$\begin{aligned} (A \times \Omega_2) \cap (\Omega_1 \times B) &= A \times B, \\ \mathbb{P}[A \times \Omega_2] &= \sum_{\omega_1 \in A} p_{\omega_1} \sum_{\omega_2 \in \Omega_2} p_{\omega_2} = \mathbb{P}_1[A], \\ \mathbb{P}[\Omega_1 \times B] &= \sum_{\omega_1 \in \Omega_1} p_{\omega_1} \sum_{\omega_2 \in B} p_{\omega_2} = \mathbb{P}_2[B], \\ \mathbb{P}[A \times B] &= \sum_{\omega_1 \in A} p_{\omega_1} \sum_{\omega_2 \in B} p_{\omega_2} = \mathbb{P}_1[A]\mathbb{P}_2[B], \end{aligned}$$

donc $\mathbb{P}[(A \times \Omega_2) \cap (\Omega_1 \times B)] = \mathbb{P}[A \times \Omega_2]\mathbb{P}[\Omega_1 \times B]$.

Plus généralement, si $Y_j : \Omega_j \rightarrow \mathbb{R}$, $j = 1, 2$ sont deux variables aléatoires, et que l'on définit $X_j : \Omega \rightarrow \mathbb{R}$, $j = 1, 2$ en posant $X_j(\omega) = Y_j(\omega_j)$, alors X_1 et X_2 sont indépendantes. En effet, si $x, y \in \mathbb{R}$, alors $(X_1 = x) = \{Y_1 = x\} \times \Omega_2$, $(X_2 = y) = \Omega_1 \times \{Y_2 = y\}$, $(X_1 = x; X_2 = y) = (Y_1 = x) \times (Y_2 = y)$. Du coup, le calcul précédent permet de conclure.

Lemme 5.4 Si X et Y sont indépendantes et si A et B sont deux événements, alors $\mathbb{P}[X \in A; Y \in B] = \mathbb{P}[X \in A]\mathbb{P}[Y \in B]$.

DÉMONSTRATION. On décompose $(X \in A)$ et $(Y \in B)$ selon les valeurs de X et Y :

$$(X \in A) = \bigcup_{a \in A} (X = a) \quad \text{et} \quad (Y \in B) = \bigcup_{b \in B} (Y = b).$$

Comme Ω est au plus dénombrable, on peut supposer que A et B aussi. Donc on obtient des partitions au plus dénombrable de Ω . On a aussi

$$(X \in A; Y \in B) = \bigcup_{(a,b) \in A \times B} (X = a; Y = b).$$

Du coup

$$\begin{aligned} \mathbb{P}[X \in A; Y \in B] &= \mathbb{P} \left[\bigcup_{(a,b) \in A \times B} (X = a; Y = b) \right] \\ &= \sum_{(a,b) \in A \times B} \mathbb{P}[X = a; Y = b] \\ &= \sum_{(a,b) \in A \times B} \mathbb{P}[X = a] \mathbb{P}[Y = b] \\ &= \left(\sum_{a \in A} \mathbb{P}[X = a] \right) \left(\sum_{b \in B} \mathbb{P}[Y = b] \right) \\ &= \mathbb{P}[X \in A] \mathbb{P}[Y \in B] \end{aligned}$$

■

Proposition 5.5 Soient X, Y des v.a. indépendantes, et f, g des applications de \mathbb{R} dans \mathbb{R} . Alors $f(X)$ et $g(Y)$ sont indépendantes.

DÉMONSTRATION. Soient X et Y les deux v.a. indépendantes, f et g les deux fonctions de \mathbb{R} dans \mathbb{R} . Si $\alpha, \beta \in \mathbb{R}$, alors

$$\begin{aligned} (f \circ X = \alpha) &= (X \in f^{-1}(\alpha)) \\ (g \circ Y = \beta) &= (Y \in g^{-1}(\beta)) \\ (f \circ X = \alpha; g \circ Y = \beta) &= (X \in f^{-1}(\alpha); Y \in g^{-1}(\beta)) \end{aligned}$$

donc le lemme 5.4 implique

$$\begin{aligned}\mathbb{P}[f(X) = \alpha; g(Y) = \beta] &= \mathbb{P}[X \in f^{-1}(\alpha); Y \in g^{-1}(\beta)] \\ &= \mathbb{P}[X \in f^{-1}(\alpha)]\mathbb{P}[Y \in g^{-1}(\beta)] \\ &= \mathbb{P}[f(X) = \alpha]\mathbb{P}[g(Y) = \beta].\end{aligned}$$

■

On a le résultat essentiel suivant.

Théorème 5.6 *Soient X, Y deux v.a. indépendantes et intégrables. Alors le produit XY est intégrable et $\mathbb{E}[XY] = \mathbb{E}[X] \times \mathbb{E}[Y]$.*

De plus, si $f, g : \mathbb{R} \rightarrow \mathbb{R}$ sont bornées alors $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)] \times \mathbb{E}[g(Y)]$.

DÉMONSTRATION. On suppose d'abord X, Y positives et on note $\{x_i\}_{i \in I}$ (resp. $\{y_j\}_{j \in J}$) les valeurs prises par X (resp. par Y), de sorte que $X = \sum_{i \in I} x_i \mathbb{1}_{(X=x_i)}$ (resp. $Y = \sum_{j \in J} y_j \mathbb{1}_{(Y=y_j)}$) et

$$\begin{aligned}XY &= \left(\sum_{i \in I} x_i \mathbb{1}_{(X=x_i)} \right) \times \left(\sum_{j \in J} y_j \mathbb{1}_{(Y=y_j)} \right) \\ &= \sum_{(i,j) \in I \times J} x_i y_j \mathbb{1}_{(X=x_i)} \mathbb{1}_{(Y=y_j)} \\ &= \sum_{(i,j) \in I \times J} x_i y_j \mathbb{1}_{(X=x_i; Y=y_j)}.\end{aligned}$$

Par conséquent, d'après les lemmes 3.2 et 3.3,

$$\begin{aligned}\mathbb{E}[XY] &= \sum_{(i,j) \in I \times J} x_i y_j \mathbb{E}[\mathbb{1}_{(X=x_i; Y=y_j)}] \\ &= \sum_{(i,j) \in I \times J} x_i y_j \mathbb{P}[X = x_i; Y = y_j] \\ &= \sum_{(i,j) \in I \times J} x_i y_j \mathbb{P}[X = x_i] \times \mathbb{P}[Y = y_j] \\ &= \left(\sum_{i \in I} x_i \mathbb{P}[X = x_i] \right) \times \left(\sum_{j \in J} y_j \mathbb{P}[Y = y_j] \right) \\ &= \mathbb{E}[X] \times \mathbb{E}[Y].\end{aligned}$$

Ceci montre que si X et Y sont indépendantes et de signe quelconque, alors, d'après la proposition précédente, $|X|$ et $|Y|$ sont aussi indépendantes et XY est intégrable. En refaisant le même calcul que ci-dessus, on obtient aussi $\mathbb{E}[XY] = \mathbb{E}[X] \times \mathbb{E}[Y]$.

Si on prend f et g bornées, alors $f(X)$ et $g(Y)$ sont indépendantes et intégrables donc le résultat suit. ■

Remarque 5.7 Si on prend $f = \mathbb{1}_{x_i}$ et $g = \mathbb{1}_{y_j}$ ci-dessus, on obtient

$$\mathbb{P}[X = x_i; Y = y_j] = \mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)] \times \mathbb{E}[g(Y)] = \mathbb{P}[X = x_i] \times \mathbb{P}[Y = y_j].$$

Pour montrer que deux v.a. X et Y sont indépendantes, il faut et il suffit de montrer que pour toutes fonctions bornées f et g , on a $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)] \times \mathbb{E}[g(Y)]$.

Corollaire 5.8 Si $X, Y \in L^2$ sont indépendantes alors

1. on a $\text{cov}(X, Y) = 0$ (mais la réciproque n'est pas vraie) et
2. on a $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$.

DÉMONSTRATION. Comme $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, la formule de Huygens implique $\text{cov}(X, Y) = 0$. De plus, on a

$$\text{var}(X + Y) = \text{var}(X) + 2\text{cov}(X, Y) + \text{var}(Y) = \text{var}(X) + \text{var}(Y).$$

■

On est amené en général à manipuler plus de deux événements ou v.a. à la fois. On pose les

Définition 5.9 Les événements A_1, \dots, A_n sont dits indépendants si pour tout $2 \leq k \leq n$ et toutes les sous-suites $1 \leq j_1 < j_2 < \dots < j_k \leq n$,

$$\mathbb{P}[\cap_{i=1}^k A_{j_i}] = \prod_{i=1}^k \mathbb{P}[A_{j_i}].$$

Définition 5.10 Les v.a. X_1, \dots, X_n sont dites indépendantes si pour tous les $x_1, \dots, x_n \in \mathbb{R}$, on a

$$\mathbb{P}[\cap_{k=1}^n (X_k = x_k)] = \prod_{i=1}^n \mathbb{P}[X_k = x_k].$$

Une suite infinie de v.a. $(X_n)_{n \geq 1}$ est dite indépendante si toutes ses sous-suites finies le sont.

Remarquons que dans chacune de ces deux définitions, il y a un « pour tout », mais il ne porte pas sur le même objet ! Dans le cas des événements, on est obligé de vérifier une identité pour chaque sous-ensemble de $\{1, 2, \dots, n\}$. Dans le second cas, le « pour tout » sur les x_i nous dispense de faire varier les sous-ensembles de $\{1, 2, \dots, n\}$. En effet, en sommant sur toutes les valeurs possibles de X_i , on obtient une identité qui porte sur les indices autres que i , etc. On peut en particulier noter que l'indépendance d'une suite de v.a. indicatrices est équivalente à l'indépendance des événements correspondants (bien que les définitions ne s'écrivent pas de la même façon !).

Remarque 5.11 *Il est important de noter que l'indépendance d'une suite d'événements est une propriété plus forte que l'indépendance deux à deux des mêmes événements (idem pour des v.a.). Ceci se voit sur l'exemple suivant. Soit X_1, X_2 et X_3 trois v.a. indépendantes, telles que pour $i = 1, 2, 3$, on ait $\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = -1] = 1/2$. Posons $Y_1 = X_1X_2$, $Y_2 = X_2X_3$ et $Y_3 = X_3X_1$. Les trois v.a. Y_1, Y_2, Y_3 ne sont pas indépendantes, puisque $Y_3 = Y_1Y_2$ donc connaissant Y_1 et Y_2 , on connaît Y_3 et l'indépendance globale des trois v.a. est mise en défaut (exercice). Par contre, ces trois v.a. sont deux à deux indépendantes. Il suffit de vérifier que Y_1 et Y_2 sont indépendantes (les autres vérifications se font de manière identique). On remarque d'abord que les Y_i ont chacune la même loi que les X_i , puis que pour $a, b = +1$ ou -1 , $\mathbb{P}[Y_1 = a; Y_2 = b] = 1/4$ (exercice).*

Les résultats précédents se déclinent comme suit.

Théorème 5.12 *Soient X_1, \dots, X_n des v.a. indépendantes, A_1, \dots, A_n des événements et f_1, \dots, f_n des applications de \mathbb{R} dans \mathbb{R} .*

1. *On a $\mathbb{P}[\cap(X_i \in A_i)] = \prod \mathbb{P}[X_i \in A_i]$.*
2. *La suite $f_1(X_1), \dots, f_n(X_n)$ est une suite de v.a. indépendantes.*
3. *Si les f_j sont bornées, alors*

$$\mathbb{E}[f_1(X_1) \times \dots \times f_n(X_n)] = \mathbb{E}[f_1(X_1)] \times \dots \times \mathbb{E}[f_n(X_n)].$$

4. *Si X_1, \dots, X_n sont de carré intégrables, alors*

$$\text{var}(X_1 + \dots + X_n) = \text{var}(X_1) + \dots + \text{var}(X_n).$$

Exemple 1.5.— Si X suit la loi binomiale $\mathcal{B}(n, p)$, alors $\text{var}(X) = np(1-p)$. En effet, X suit la même loi que la v.a. s_n définie dans l'exemple 1.5. Du coup, $\text{var}(X) = \text{var}(s_n)$. Or $s_n = \sum_{j=1}^n X_j$ où $(X_j)_{1 \leq j \leq n}$ est une famille de v.a. indépendantes de loi $\mathcal{B}(p)$, donc de variance $p(1-p)$. Par conséquent

$$\text{var}(X) = \text{var}(s_n) = \sum_{j=1}^n \text{var}(X_j) = np(1-p).$$

Chapitre III. Fonctions génératrices³

On se fixe un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ où l'espace des issues Ω est au plus dénombrable. Dans tout ce chapitre, nous nous intéresserons aux v.a. qui prennent leurs valeurs dans \mathbb{N} . Pour une telle v.a. X , on aura

$$X = \sum_{k \geq 0} k \mathbb{1}_{(X=k)}.$$

6 Généralités

Définition 6.1 Soit X une variable aléatoire prenant ses valeurs dans \mathbb{N} . La fonction génératrice de X est la fonction G_X définie par :

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

On notera que

$$G_X(1) = 1 \text{ et } G_X(0) = \mathbb{P}[X = 0].$$

Exemple 1.4.— Si X est une variable de Bernoulli de paramètre p ,

$$G_X(t) = t^0 \mathbb{P}[X = 0] + t^1 \mathbb{P}[X = 1] = (1 - p) + pt.$$

Remarque 6.2 Si la variable X ne prend qu'un nombre fini de valeurs, alors G_X est définie pour tout $t \in \mathbb{R}$, et est un polynôme.

Proposition 6.3 La série génératrice d'une v.a. à valeurs dans \mathbb{N} est une série entière de rayon de convergence au moins 1.

DÉMONSTRATION. Si $z \in \mathbb{C}$ est de module au plus 1, alors $|z^k \mathbb{P}[X = k]| \leq \mathbb{P}[X = k]$, et $\sum_{k=0}^{\infty} \mathbb{P}[X = k] = 1$. Donc le rayon de convergence vaut au moins 1. ■

Rappels sur les séries entières.— Soit $(a_n)_{n \geq 0}$ une suite de complexes. On s'intéresse à la série dite entière

$$f(z) = \sum_{n \geq 0} a_n z^n.$$

La théorie nous dit qu'il existe un réel maximal $R_f \geq 0$ qui peut être infini, appelé le *rayon de convergence de f* , tel que, pour tout complexe z de module $|z| < R_f$, $f(z)$ est absolument convergente. Sur $\{|z| < R_f\}$, la fonction f est de classe C^∞ et, pour tout $t \in]-R_f, R_f[$, on a

$$f^{(k)}(t) = \sum_{n \geq k} \frac{n!}{(n-k)!} a_n t^{n-k}.$$

³16 avril 2012

Cette série a même rayon de convergence que f . En particulier, on a, pour tout $k \geq 0$,

$$a_k = \frac{f^{(k)}(0)}{k!}.$$

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

Proposition 6.4 *La fonction génératrice d'une variable aléatoire X à valeurs dans \mathbb{N} ne dépend que de sa loi et la caractérise complètement. Plus précisément, si X et Y sont deux variables à valeurs dans \mathbb{N} ,*

$$G_X = G_Y \Leftrightarrow \forall k \in \mathbb{N}, \mathbb{P}[X = k] = \mathbb{P}[Y = k]$$

et

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

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.

Comme G_X est une série entière de rayon de convergence strictement positif, on peut identifier les coefficients de la série avec les dérivées successives en zéro de la série : on a, pour tout $k \in \mathbb{N}$,

$$\mathbb{P}[X = k] = \frac{G_X^{(k)}(0)}{k!}.$$

Supposons maintenant que $G_X = G_Y$. On aura, pour tout $k \geq 0$,

$$\mathbb{P}[X = k] = \frac{G_X^{(k)}(0)}{k!} = \frac{G_Y^{(k)}(0)}{k!} = \mathbb{P}[Y = k].$$

■

Remarque 6.5 *Cette expression montre qu'il suffit de connaître G_X au voisinage de 0 pour caractériser la loi de X .*

Corollaire 6.6 *Soit X une v.a. intégrable prenant ses valeurs dans \mathbb{N} .*

1. On a

$$\mathbb{E}[X] = G'_X(1).$$

2. Si X est de carré intégrable, alors

$$\text{var}(X) = G''_X(1) + G'_X(1)(1 - G'_X(1)).$$

DÉMONSTRATION. On utilise le fait que G_X est de classe C^2 . On a donc

$$G'_X(t) = \sum_{k \geq 1} k \mathbb{P}[X = k] t^{k-1}.$$

Quand $t = 1$, on trouve

$$G_X(1) = \sum_{k \geq 1} k \mathbb{P}[X = k] = \mathbb{E}[X].$$

En redérivant une seconde fois, on obtient

$$G_X''(t) = \sum_{k \geq 1} k(k-1) \mathbb{P}[X = k] t^{k-2}.$$

Si X est de carré intégrable, alors le terme de droite est bien définie pour $t = 1$:

$$\sum_{k \geq 1} k(k-1) \mathbb{P}[X = k] = \sum_{k \geq 1} k^2 \mathbb{P}[X = k] - \sum_{k \geq 1} k \mathbb{P}[X = k] = \mathbb{E}[X^2] - \mathbb{E}[X].$$

Par suite, en utilisant la formule de Huygens, on obtient

$$\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = G_X''(1) + G_X'(1)(1 - G_X'(1)).$$

■

7 Fonctions génératrices et indépendance

Si X, Y sont deux v.a. indépendantes, alors t^X et t^Y sont indépendantes pour tout $t > 0$ d'après la proposition 5.5. Donc le théorème 5.6 affirme $\mathbb{E}[t^{X+Y}] = \mathbb{E}[t^X] \mathbb{E}[t^Y]$. On en déduit plus généralement

Proposition 7.1 *Soient X_1, \dots, X_n des v.a. indépendantes prenant des valeurs dans \mathbb{N} . Alors*

$$G_{\sum X_j} = G_{X_1} \times \dots \times G_{X_n}.$$

En particulier, si elles ont toutes même loi, alors

$$G_{\sum X_j}(t) = G_{X_1}^n.$$

DÉMONSTRATION. Grâce au théorème 5.12. points 2. et 3., on écrit, pour $t > 0$,

$$\begin{aligned} G_{\sum X_j}(t) &= \mathbb{E}[t^{X_1} \dots t^{X_n}] \\ &= \mathbb{E}[t^{X_1}] \dots \mathbb{E}[t^{X_n}] \\ &= G_{X_1}(t) \times \dots \times G_{X_n}(t). \end{aligned}$$

Si les v.a. suivent toutes la même loi, alors elles ont même fonction génératrice. ■

Exemple 1.5.— Montrons comment déterminer la loi de s_n en calculant sa fonction génératrice. On rappelle que s_n est la somme de n v.a. indépendantes de loi de Bernoulli $\mathcal{B}(p)$. Si on désigne par Y une v.a. de loi de Bernoulli $\mathcal{B}(p)$, alors la proposition 7.1 implique

$$G_{s_n}(t) = G_Y^n(t) = (1 - p + tp)^n = \sum_{k=0}^n C_n^k p^k (1 - p)^{n-k} t^k.$$

Par conséquent, on trouve $\mathbb{P}[s_n = k] = C_n^k p^k (1-p)^{n-k}$ si $k \in \{0, \dots, n\}$, et 0 sinon.

Plus généralement, si X suit la loi binomiale $\mathcal{B}(n, p)$, alors X est de même loi que s_n . Donc la proposition 6.4 affirme

$$G_X(t) = G_{s_n}(t) = (1 - p + tp)^n.$$

Supposons maintenant que X' soit une v.a. de loi $\mathcal{B}(m, p)$ indépendante de X . Alors

$$G_{X+X'}(t) = G_X(t)G_{X'}(t) = (1 - p + tp)^{m+n}.$$

On reconnaît la fonction génératrice d'une v.a. de loi $\mathcal{B}(m+n, p)$, donc, puisque la fonction génératrice détermine la loi, on en déduit que $X + X'$ suit la loi $\mathcal{B}(m+n, p)$.

Exemple 1.3.— Notons X_1 et X_2 les projections de $\Omega = \{1, \dots, 6\}^2$ sur la première et la seconde coordonnée de sorte que $X = X_1 + X_2$. Par définition, X_1 et X_2 sont indépendantes et de même loi (uniforme). Donc

$$G_{X_1}(t) = G_{X_2}(t) = \frac{1}{6} \sum_{j=1}^6 t^j.$$

D'après la proposition 7.1, on trouve

$$\begin{aligned} G_X(t) &= \frac{1}{36} \left(\sum_{i=1}^6 t^i \right) \cdot \left(\sum_{j=1}^6 t^j \right) \\ &= \frac{1}{36} \sum_{k=2}^{12} \left(\sum_{\substack{i+j=k \\ 1 \leq i, j \leq 6}} t^k \right) \\ &= \frac{1}{36} \left(\sum_{k=2}^7 (k-1)t^k + \sum_{k=8}^{12} (13-k)t^k \right). \end{aligned}$$

On retrouve ainsi la loi de X .

On se pose le problème suivant : peut-on piper les deux dés de sorte que la loi de X soit maintenant la loi uniforme sur $\{2, \dots, 12\}$? Plus précisément, peut-on trouver une distribution de probabilité $(p_j)_{1 \leq j \leq 6}$ pour le premier dé et une distribution de probabilité $(q_j)_{1 \leq j \leq 6}$ pour le second dé de sorte que la loi de $X = X_1 + X_2$ soit uniforme? Dans ce cas, la proposition 7.1 nous dirait

$$\left(\sum_{i=1}^6 p_i t^i \right) \cdot \left(\sum_{j=1}^6 q_j t^j \right) = \frac{1}{11} \sum_{k=2}^{12} t^k.$$

On peut simplifier par t^2 cette équation afin d'obtenir

$$\left(\sum_{i=0}^5 p_{i+1} t^i \right) \cdot \left(\sum_{j=0}^5 q_{j+1} t^j \right) = \frac{1}{11} \sum_{k=0}^{10} t^k.$$

Le terme de gauche est un produit de deux polynômes de degré impair. En effet, l'événement $(X = 12)$ se réalise avec probabilité $1/11$, ce qui force p_6 et q_6 à être non nulles. Ces polynômes s'annulent donc tous les deux. En revanche, le polynôme de droite ne s'annule pas. En effet, en 1, sa valeur vaut 1 ; par ailleurs, on reconnaît la somme partielle d'une suite géométrique de raison t donc

$$\sum_{k=0}^{10} t^k = \frac{1 - t^{11}}{1 - t}$$

ce qui implique que ce terme de l'équation ne s'annule pas. Ceci montre que cette équation n'admet pas de solution et donc qu'il est impossible de piper les dés pour que X suive une loi uniforme.

8 Somme d'un nombre aléatoire de variables aléatoires

Exemple 8.1 *Une poule pond N œufs. On suppose que N est une variable aléatoire (à valeurs dans \mathbb{N}). Chaque œuf, indépendamment de tout le reste, éclôt 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'œufs pondus, et l'état de chaque œuf. Intuitivement, on voudrait écrire $\omega = (n, o_1, \dots, o_n)$ où n est un entier correspondant au nombre d'œufs pondus et

$$o_i = \begin{cases} 1 & \text{si le } i\text{-ème œuf donne un poussin;} \\ 0 & \text{sinon.} \end{cases}$$

En pratique, il est malaisé de traiter avec des issues dont la taille est variable : $(1, 0)$ et $(3, 1, 0, 1)$... Une façon (artificielle) très commode est de se donner une première v.a. aléatoire N à valeurs dans \mathbb{N} qui donnera le nombre d'œufs pondus et de traiter les œufs pondus comme une suite de variable aléatoires $(o_n)_{n \geq 1}$ indépendantes et de même loi (de Bernoulli $\mathcal{B}(p)$). C'est un peu comme si on se plaçait dans la situation où notre poule pouvait a priori pondre un nombre infini d'œufs !

Cependant, on s'intéresse à la loi de $K(\omega) = o_1(\omega) + \dots + o_{N(\omega)}(\omega)$.

Le contexte de ce paragraphe est donc le suivant : on se donne $N, X_1, \dots, X_n, \dots$ une suite de variables aléatoires indépendantes à valeurs dans \mathbb{N} . En général, on supposera la suite (X_n) de v.a. « indépendantes et identiquement distribuées » (i.i.d.), c'est-à-dire indépendantes et de même loi.

On pose ensuite $S_0 = 0$ et $S_n = X_1 + \dots + X_n$, $n \geq 1$, et on considère la v.a.

$$S_N(\omega) = \sum_{j=1}^{N(\omega)} X_j(\omega) = \sum_{n \geq 0} S_n(\omega) \mathbb{1}_{(N=n)}(\omega),$$

où $\sum_{j=1}^0 = 0$ par convention.

On commence par établir un résultat préliminaire.

Lemme 8.1 *Pour tout n , S_n et N sont indépendantes.*

DÉMONSTRATION. Soient $k, \ell \in \mathbb{N}$. Tout d'abord, on peut supposer $k, \ell \geq 1$. On décompose $(S_n = k)$ selon les valeurs possibles des X_j , $1 \leq j \leq n$:

$$(S_n = k) = \bigcup_{\substack{x_1 + \dots + x_n = k \\ x_j \in \mathbb{N}, 1 \leq j \leq n}} (X_j = x_j).$$

On a ainsi décomposé cet événement en événements disjoints pour lesquels nous pourrions utiliser notre hypothèse d'indépendance pour obtenir

$$\mathbb{P}[X_j = x_j, j = 1, \dots, n; N = \ell] = \left(\prod_{j=1}^n \mathbb{P}[X_j = x_j] \right) \mathbb{P}[N = \ell] = \mathbb{P}[X_j = x_j, j = 1, \dots, n] \mathbb{P}[N = \ell].$$

Du coup

$$\begin{aligned} \mathbb{P}[S_n = k; N = \ell] &= \mathbb{P} \left[\bigcup_{\substack{x_1 + \dots + x_n = k \\ x_j \in \mathbb{N}, 1 \leq j \leq n}} (X_j = x_j; N = \ell) \right] \\ &= \sum_{\substack{x_1 + \dots + x_n = k \\ x_j \in \mathbb{N}, 1 \leq j \leq n}} \mathbb{P}[X_j = x_j, 1 \leq j \leq n; N = \ell] \\ &= \sum_{\substack{x_1 + \dots + x_n = k \\ x_j \in \mathbb{N}, 1 \leq j \leq n}} \mathbb{P}[X_j = x_j, 1 \leq j \leq n] \mathbb{P}[N = \ell] \\ &= \mathbb{P}[N = \ell] \sum_{\substack{x_1 + \dots + x_n = k \\ x_j \in \mathbb{N}, 1 \leq j \leq n}} \mathbb{P}[X_j = x_j, 1 \leq j \leq n] \\ &= \mathbb{P}[N = \ell] \mathbb{P}[S_n = k] \end{aligned}$$

■

Nous pouvons maintenant établir le résultat clef de ce contexte.

Proposition 8.2 *Soit $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$. Alors*

$$\mathbb{E}[f(S_N)] = \sum_{n \geq 0} \mathbb{E}[f(S_n)] \mathbb{P}[N = n].$$

DÉMONSTRATION. Par indépendance de S_n et N , on a $\mathbb{E}[f(S_n)\mathbb{I}_{(N=n)}] = \mathbb{E}[f(S_n)]\mathbb{P}[N = n]$ donc

$$\begin{aligned}\mathbb{E}[f(S_N)] &= \mathbb{E}\left[\sum_{n \geq 0} f(S_n)\mathbb{I}_{(N=n)}\right] \\ &= \sum_{n \geq 0} \mathbb{E}[f(S_n)\mathbb{I}_{(N=n)}] \\ &= \sum_{n \geq 0} \mathbb{E}[f(S_n)]\mathbb{P}[N = n].\end{aligned}$$

■

Nous allons maintenant montré comment calculer la fonction génératrice de S_N .

Proposition 8.3 *Si la suite (X_n) est i.i.d. alors $G_{S_N} = G_N \circ G_{X_1}$.*

DÉMONSTRATION. Comme la suite est i.i.d., on a $G_{S_n} = G_{X_1}^n$ pour chaque $n \geq 0$ d'après la proposition 7.1. Par ailleurs, si $t \in [0, 1]$, alors la proposition 8.2 affirme

$$\begin{aligned}G_{S_N}(t) &= \mathbb{E}[t^{S_N}] = \sum_{n \geq 0} \mathbb{E}[t^{S_n}] \mathbb{P}[N = n] \\ &= \sum_{n \geq 0} G_{S_n}(t) \mathbb{P}[N = n] = \sum_{n \geq 0} \mathbb{P}[N = n] G_{X_1}(t)^n \\ &= G_N(G_{X_1})(t).\end{aligned}$$

■

Corollaire 8.4 *Sous les hypothèses de la proposition 8.3, si toutes les v.a. sont de plus de carré intégrable, alors*

$$\mathbb{E}[S_N] = \mathbb{E}[X]\mathbb{E}[N]$$

et

$$\text{var}(S_N) = \text{var}(X_1)\mathbb{E}[N] + \mathbb{E}[X_1]^2 \text{var}(N).$$

DÉMONSTRATION. Prenons X de même loi que X_1 . On dérive G_{S_N} : on trouve d'abord $G'_{S_N} = (G'_N \circ G_X) \cdot G'_X$ puis

$$G''_{S_N} = (G''_N \circ G_X) \cdot (G'_X)^2 + (G'_N \circ G_X) \cdot G''_X$$

donc la proposition 6.6 nous dit

$$\mathbb{E}[S_N] = G'_{S_N}(1) = (G'_N \circ G_X)(1) \cdot G'_X(1) = \mathbb{E}[X]\mathbb{E}[N]$$

en utilisant $G_X(1) = 1$, et

$$\begin{aligned}\text{var}(S_N) &= G''_{S_N}(1) + G'_{S_N}(1)(1 - G'_{S_N}(1)) \\ &= (G''_N \circ G_X(1)) \cdot (G'_X)^2(1) + (G'_N \circ G_X(1)) \cdot G''_X(1) \\ &\quad + (G'_N \circ G_X(1)) \cdot G'_X(1)\{1 - (G'_N \circ G_X(1)) \cdot G'_X(1)\} \\ &= G''_N(1)\mathbb{E}[X]^2 + \mathbb{E}[N]G''_X(1) + G'_N(1)\mathbb{E}[X]\{1 - G'_N(1) \cdot \mathbb{E}[X]\} \\ &= \{G''_N(1) + G'_N(1)(1 - G'_N(1))\}\mathbb{E}[X]^2 + \mathbb{E}[N]\{G''_X(1) + G'_X(1)(1 - G'_X(1))\} \\ &= \text{var}(X_1)\mathbb{E}[N] + \mathbb{E}[X_1]^2 \text{var}(N).\end{aligned}$$



Remarque 8.5 *On aurait pu faire ces calculs sans passer par les fonctions génératrices en utilisant directement la proposition 8.2. Faisons-le pour l'espérance.*

Comme la suite est de même loi, on a $\mathbb{E}[X_j] = \mathbb{E}[X_1]$ pour tout $j \geq 1$ et donc $\mathbb{E}[S_n] = n\mathbb{E}[X_1]$. D'après la proposition 8.2, on a

$$\mathbb{E}[S_N] = \sum_{n \geq 0} \mathbb{E}[S_n] \mathbb{P}[N = n] = \mathbb{E}[X_1] \sum_{n \geq 0} n \mathbb{P}[N = n] = \mathbb{E}[X] \mathbb{E}[N].$$

9 Evolution des Populations

9.1 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*

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

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

$$Z_{n+1} = \eta_1^{(n+1)} + \eta_2^{(n+1)} + \dots + \eta_{Z_n}^{(n+1)} = \sum_{j=1}^{Z_n} \eta_j^{(n+1)}.$$

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

$$Z_{n+1} = \sum_{k \geq 0} \mathbb{1}_{\{Z_n=k\}} \left(\sum_{j=1}^k \eta_j^{(n+1)} \right)$$

Le problème est de déterminer si la population va survivre ou non. Autrement dit, on veut déterminer la probabilité que la population finisse par s'éteindre.

9.2 Probabilité d'extinction.

Proposition 9.1 *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'événement $A := \ll \text{la population finit par s'éteindre} \gg$ en termes des $\{Z_n\}$.

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

Comme $\{Z_n = 0\} \subset \{Z_{n+1} = 0\}$, on peut appliquer la propriété 5. de la proposition 1.7

$$\mathbb{P}[A] = \lim_{n \rightarrow \infty} \mathbb{P}[Z_n = 0] = \lim_{n \rightarrow \infty} G_{Z_n}(0) \leq 1.$$

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 $\{\eta_k^{(n+1)}\}_{k \geq 1}$ sont indépendants et les $\{\eta_k^{(n+1)}\}_{k \geq 1}$ de même loi. D'après la proposition 8.3, on a

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

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

$$g_{n+1}(z) = g_n(f(z)) \implies g_n(z) = \underbrace{f \circ \dots \circ f}_{n \text{ fois}}(z) := f^{\circ n}(z).$$

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

$$\lim_{n \rightarrow \infty} g_n(0) = \chi \implies \lim_{n \rightarrow \infty} f(g_n(0)) = f(\chi).$$

Or, comme $f(g_n) = g_{n+1}$, on a que $f(\chi) = \chi$. Il nous reste à voir que χ est le plus petit réel positif vérifiant $f(\chi) = \chi$. Soit $\zeta \in [0, 1]$ tel que $f(\zeta) = \zeta$. Montrons que $\chi \leq \zeta$. Pour cela, il suffit de remarquer que f est croissante sur $[0, 1]$. En effet, si $0 \leq z_1 \leq z_2 \leq 1$, alors $0 \leq z_1^{Z_1} \leq z_2^{Z_1} \leq 1$, et en passant aux espérances, $f(z_1) \leq f(z_2)$. Par conséquent, si $\zeta \in [0, 1]$ est tel que $f(\zeta) = \zeta$, $\zeta = f(\zeta) \geq f(0)$, et en prenant f de chaque côté $\zeta \geq f^{(2)}(0)$. En répétant le procédé, on a que $\zeta \geq f^{(n)}(0)$ et donc en passant à la limite $\zeta \geq \chi$. ■

Complément. —

1. Si $\mathbb{E}[\eta] > 1$, alors $\chi < 1$.
2. Si $\mathbb{E}[\eta] < 1$, alors $\chi = 1$.

DÉMONSTRATION. On s'intéresse à la fonction $h : x \mapsto f(x) - x$ sur $[0, 1]$. Par définition de f , on a $h(1) = 0$, et on veut déterminer si elle peut s'annuler sur $[0, 1[$. Pour cela, on remarque que $\mathbb{E}[\eta] = f'(1)$.

Si $\mathbb{E}[\eta] > 1$, alors $f'(1) > 1$. Donc il existe $\delta > 0$ tel que si $x \in [1 - \delta, 1]$, alors $f'(x) > 1$, donc en intégrant entre $1 - \delta$ et 1 , on trouve $f(1) - f(1 - \delta) > 1 - (1 - \delta)$. Du coup, on a $h(1 - \delta) < 0$. Par ailleurs, on a $h(0) \geq 0$. Donc, ou bien $h(0) = 0$ auquel cas $\chi = 0$, ou bien le théorème des valeurs intermédiaires nous fournit l'existence de $t \in]0, 1 - \delta[$ tel que $h(t) = 0$. Par conséquent, $\chi \leq t < 1$.

En revanche, si $\mathbb{E}[\eta] < 1$, alors $f'(t) < 1$ pour tout $t \in [0, 1[$ car f' est croissante. Par conséquent, $f(1) - f(t) < 1 - t$ sur $[0, 1[$ donc $t < f(t)$; ceci implique que $h(t) > 0$ pour $t < 1$. Du coup, $h(t) = 0$ uniquement pour $t = 1$: on obtient ainsi $\chi = 1$. ■

9.3 Exemple

Supposons que la loi de la progéniture soit une loi géométrique de paramètre p

$$\mathbb{P}[\eta_1^{(1)} = k] = p(1-p)^k, \quad \text{pour } k = 0, 1, \dots,$$

Donc

$$G_{Z_1}(z) = \sum_{n=0}^{\infty} z^n \mathbb{P}[\eta_1^{(1)} = n] = \frac{p}{1 - z(1-p)}.$$

Donc $G_{Z_1}(0) = p$ et $G'_{Z_1}(1) = (1-p)/p$. On cherche la plus petite solution de $G_{Z_1}(\chi) = \chi$. On résoud

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

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.

Chapitre IV. Loïs (absolument) continues⁴

10 Espace de probabilité revisité

Jusqu'à présent, nous n'avons considéré que des phénomènes aléatoires prenant un nombre fini ou dénombrable de valeurs. Dans bien des situations, ce n'est pas le cas. Par exemple, considérons une aiguille que l'on jette sur un parquet dont les lames sont de largeur constante d . On se demande quelle est la probabilité que l'aiguille touche deux lames du parquet (problème de l'aiguille de Buffon). L'espace des issues est ici $\Omega = [0, d] \times [0, 2\pi]$ qui décrit la position de la tête de l'aiguille relative aux lames du parquet et l'angle qu'elle fait avec l'horizontale. Dans ce cas, il est impossible de construire une probabilité \mathbb{P} sur Ω , comme nous l'avons fait jusqu'à présent en partant d'une distribution de probabilité qui associe une probabilité à chaque issue possible. Ainsi, dans l'exemple précédent, et si la probabilité de tomber dans un sous-ensemble A de Ω est proportionnelle à l'aire de A , alors $\mathbb{P}[\{\omega\}] = 0$ pour tout $\omega \in \Omega$. Du coup, nous ne pourrons plus utiliser la théorie des séries.

A partir de la proposition 1.7, on peut oublier la notion de distribution de probabilité et n'utiliser que la mesure de probabilité \mathbb{P} ; celle-ci doit vérifier les propriétés suivantes pour que l'on puisse généraliser ce que l'on a fait jusqu'ici :

- (i) $\mathbb{P}[\Omega] = 1$.
- (ii) Si (A_n) est une suite d'événements disjoints, alors

$$\mathbb{P}[\cup A_n] = \sum \mathbb{P}[A_n].$$

Il se trouve, et c'est là un point délicat de la théorie, qu'une fonction qui vérifie (ii) pour **toute partie** de Ω n'existe pas en général. Par exemple, il n'existe pas d'application $\mathbb{P} : \mathcal{P}([0, 1]) \rightarrow [0, 1]$ telle que $\mathbb{P}[a, b] = |b - a|$ pour tout intervalle $[a, b] \subset [0, 1]$. Du coup, on est obligé de réduire l'ensemble des événements pour lesquels nous pourrons calculer leur probabilité de se réaliser.

A cette fin, on introduit la notion suivante.

Définition 10.1 (tribu) Soit Ω un espace d'issues. On appelle **tribu** d'événements toute famille \mathcal{A} de sous-ensembles de Ω vérifiant

1. $\Omega \in \mathcal{A}$.
2. Si $A \in \mathcal{A}$, ${}^cA \in \mathcal{A}$.
3. Pour tout famille finie ou dénombrable $(A_i)_{i \in I}$ d'éléments de \mathcal{A} , $\cup_{i \in I} A_i \in \mathcal{A}$.

Les propriétés 2 et 3 entraînent que toute intersection dénombrable d'éléments de \mathcal{A} est encore dans \mathcal{A} .

Exemple 10.1 1. L'ensemble des parties de Ω est une tribu. C'est la plus grande. C'est celle que nous avons de façon implicite toujours considérée dans le cas où Ω est fini ou dénombrable.

⁴16 avril 2012

2. $\{\emptyset, \Omega\}$ est une tribu. C'est la plus petite.
3. Si $A \subset \Omega$, $\{\emptyset, A, A^c, \Omega\}$ est une tribu. C'est la plus petite tribu contenant A . On dit que c'est la tribu engendrée par A .

Nous pouvons maintenant définir un espace de probabilité en toute généralité.

Définition 10.2 (Espace de probabilité (cas général)) *Un espace de probabilité est un triplet $(\Omega, \mathcal{A}, \mathbb{P})$ où*

- Ω , qui désigne l'ensemble des issues, est un ensemble quelconque,
- \mathcal{A} est une tribu sur Ω qui désigne l'ensemble des événements et
- $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ est une mesure de probabilité sur (Ω, \mathcal{A}) , c'est-à-dire une application qui vérifie
 - (i) $\mathbb{P}[\Omega] = 1$.
 - (ii) Si (A_n) est une suite d'événements disjoints, alors

$$\mathbb{P}[\cup A_n] = \sum \mathbb{P}[A_n].$$

Pour travailler sur des espaces d'issues définis sur \mathbb{R} , on montre qu'il existe une plus petite tribu qui contient les intervalles : c'est la *tribu des boréliens* notée $\mathcal{B}(\mathbb{R})$. En pratique, les boréliens que nous considérerons seront tous des réunions et intersections au plus dénombrables (et la plupart du temps finies) d'intervalles.

Définition 10.3 (Loi absolument continue et densité)

- Une fonction continue par morceaux (c'est-à-dire ayant un ensemble de discontinuité au plus dénombrable) $f : \mathbb{R} \rightarrow \mathbb{R}$ définit une densité de probabilité si $f(x) \geq 0$ pour tout $x \in \mathbb{R}$ et si

$$\int_{\mathbb{R}} f(x) dx = 1.$$

Il existe alors une unique probabilité $\mathbb{P} : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ telle que, pour tous $a \leq b$, on ait

$$\mathbb{P}[a, b] = \int_a^b f(x) dx.$$

- Une probabilité \mathbb{P} sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est dite absolument continue s'il existe une densité f telle que, pour tous $a \leq b$, on ait

$$\mathbb{P}[a, b] = \int_a^b f(x) dx.$$

Remarque 10.4 *La densité d'une probabilité absolument continue n'est pas définie de manière unique : on peut changer ses valeurs sur n'importe quel ensemble dénombrable. Plus précisément, deux densités définissent la même probabilité si elles sont égales en dehors d'un ensemble dénombrable.*

Exemple 10.2 (loi uniforme) Soient $a < b$ deux réels et $f = (1/(b - a))\mathbb{I}_{[a,b]}$ la fonction valant $1/(b - a)$ sur $[a, b]$, et 0 partout ailleurs. Cette fonction f est positive, continue (sauf en a et b) et $\int f(t)dt = 1$ donc f est bien une densité de probabilité. On note la loi $\mathcal{U}([a, b])$.

Exemple 10.3 (loi exponentielle) Soit λ un réel strictement positif et posons $f(x) = \lambda \exp(-\lambda x) \mathbb{I}_{\mathbb{R}_+}(x)$. On vérifie que f est une densité de probabilité. En effet, f est positive, et

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^{\infty} \lambda \exp(-\lambda x)dx = [-\exp(-\lambda x)]_0^{\infty} = 1.$$

On note la loi $\mathcal{E}(\lambda)$.

Exemple 10.4 (loi normale) La loi normale ou gaussienne est une loi qui apparaît de manière naturelle comme on le verra dans le chapitre suivant. Soient $m \in \mathbb{R}$ et σ un réel strictement positif et soit

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp -\frac{1}{2} \left(\frac{x - m}{\sigma} \right)^2.$$

On vérifie que f est une densité de probabilité dont on note la loi $\mathcal{N}(m, \sigma^2)$. La loi $\mathcal{N}(0, 1)$ s'appelle la loi normale réduite centrée.

On a les mêmes propriétés que dans celles de la proposition 1.7 :

Proposition 10.5 Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Alors les propriétés suivantes sont vérifiées.

1. Si A et B sont disjoints alors $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B]$.
2. Pour tout évènement A , $\mathbb{P}(A) + \mathbb{P}(^cA) = 1$. Plus généralement, si $(A_i)_{i \in I}$ est une partition au plus dénombrable d'évènements alors

$$\sum_{i \in I} \mathbb{P}[A_i] = 1.$$

3. Si A et B sont deux évènements tels que $A \subset B$, alors $\mathbb{P}[A] \leq \mathbb{P}[B]$.
4. Si $(A_n)_{n \geq 1}$ est une suite croissante d'évènements (i.e $A_n \subset A_{n+1}$), alors

$$\mathbb{P}[\cup_{n \geq 1} A_n] = \lim_{n \rightarrow \infty} \mathbb{P}[A_n].$$

Par ailleurs, si (B_n) est une suite décroissante d'évènements ($B_{n+1} \subset B_n$ pour tout n) alors

$$\mathbb{P}[\cap_{n \geq 1} B_n] = \lim_{n \rightarrow \infty} \mathbb{P}[B_n].$$

Définition 10.6 (Fonction de répartition) Soit \mathbb{P} une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On définit la fonction de répartition $F_{\mathbb{P}} : \mathbb{R} \rightarrow \mathbb{R}$ par

$$F_{\mathbb{P}}(x) = \mathbb{P}[\] - \infty, x].$$

Théorème 10.7 Si \mathbb{P} est une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, alors

1. la fonction $F_{\mathbb{P}}$ est croissante ;
2. on a $\lim_{-\infty} F_{\mathbb{P}} = 0$ et $\lim_{+\infty} F_{\mathbb{P}} = 1$ donc $F_{\mathbb{P}}(\mathbb{R}) \subset [0, 1]$;
3. si \mathbb{P} est absolument continue alors $F_{\mathbb{P}}$ est continue sur \mathbb{R} et C^1 par morceaux.

Réciproquement, si $F : \mathbb{R} \rightarrow \mathbb{R}$ vérifie les propriétés ci-dessus, alors F est la fonction de répartition d'une probabilité absolument continue de densité F' là où la fonction est dérivable.

DÉMONSTRATION. Si $x \leq x'$ alors $] - \infty, x] \subset] - \infty, x']$ donc $\mathbb{P}[] - \infty, x]] \leq \mathbb{P}[] - \infty, x']]$ par la proposition 10.5 et $F_{\mathbb{P}}$ est croissante. De même, cette proposition implique aussi le point 2.

Si \mathbb{P} est à densité, alors $F_{\mathbb{P}}$ est une primitive de la densité f , donc $F_{\mathbb{P}}$ est dérivable partout où f est continue, et sa dérivée vaut f en ces points. Ceci montre bien que $F_{\mathbb{P}}$ est continue et C^1 par morceaux.

Réciproquement, si F est continue et C^1 par morceaux, notons f la dérivée de F là où elle est définie et 0 sinon. Tout d'abord, f est continue par morceaux car F est C^1 par morceaux. Comme F est croissante, on a bien $f \geq 0$. De plus,

$$\lim_{x \rightarrow +\infty, y \rightarrow -\infty} \int_y^x f(t) dt = \lim_{x \rightarrow +\infty, y \rightarrow -\infty} (F(x) - F(y)) = 1$$

donc f est bien une densité de probabilité. Notons-la \mathbb{P} . Par le point 4. de la proposition 10.5, on a

$$\mathbb{P}[] - \infty, x] = \lim_{y \rightarrow -\infty} \int_y^x f(t) dt = F(x) - \lim_{y \rightarrow -\infty} F(y) = F(x).$$

Donc F est bien la fonction de répartition de \mathbb{P} . ■

11 Variables aléatoires

Définition 11.1 (Variable aléatoire) Une v.a. sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ est une application $X : \Omega \rightarrow \mathbb{R}$ telle que, pour tout $x \in \mathbb{R}$, $\{X \leq x\} \in \mathcal{A}$.

Exemple 11.1 Lorsque Ω est au plus dénombrable et que l'on choisit $\mathcal{A} = \mathcal{P}(\Omega)$ alors la définition correspond bien avec celle avec laquelle nous avons travaillé jusqu'à présent.

Remarque 11.2 La définition implique que $\{X \in B\} \in \mathcal{A}$ pour tout borélien $B \in \mathcal{B}(\mathbb{R})$.

La loi d'une v.a. X est une probabilité \mathbb{P}_X définie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ par

$$\mathbb{P}_X[B] = \mathbb{P}[X \in B]$$

pour tout borélien B . Pour déterminer la loi d'une v.a., le théorème 10.7 nous dit qu'il suffit de déterminer sa fonction de répartition. Du coup, on définit la *fonction de répartition d'une v.a.* X comme la fonction de répartition de sa loi, à savoir

$$F_X(x) = \mathbb{P}[X \leq x].$$

On dira que la v.a. X est à densité (f) si la loi de \mathbb{P}_X est à densité (f), c'est-à-dire si

$$F_X(x) = \int_{-\infty}^x f(t)dt.$$

Définition 11.3 (Espérance) Soit X une v.a. de densité f .

– Si $X \geq 0$, on pose

$$\mathbb{E}[X] = \int_{\mathbb{R}} xf(x)dx.$$

Cette valeur peut être finie ou infinie.

– Si X est de signe quelconque et si $\mathbb{E}[|X|]$ est finie, on pose

$$\mathbb{E}[X] = \int_{\mathbb{R}} xf(x)dx.$$

Dans ce cas, l'espérance est toujours finie.

On peut définir les espaces L^1 et L^2 comme dans le chapitre I. Les propriétés restent inchangées dans ce contexte :

Définition 11.4 (Espaces L^1 et L^2) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. On note L^1 l'ensemble des variables aléatoires X telles que $\mathbb{E}[|X|]$ est finie et L^2 celles telles que $\mathbb{E}[X^2]$ est finie. Si $X \in L^1$, on dit que X est intégrable et si $X \in L^2$, on dit que X est de carré intégrable.

Si X est une v.a. bornée, alors $X \in L^1$. On a les résultats suivants.

Théorème 11.5 L'espace L^1 est un espace vectoriel et $\mathbb{E} : L^1 \rightarrow \mathbb{R}$ est une forme linéaire positive, c'est-à-dire

- si $X, Y \in L^1$ et $\lambda \in \mathbb{R}$, alors $\mathbb{E}[X + \lambda Y] = \mathbb{E}[X] + \lambda \mathbb{E}[Y]$;
- si $X \geq 0$ alors $\mathbb{E}[X] \geq 0$.

Proposition 11.6 On a les propriétés suivantes.

1. Si $X, Y \in L^1$ et si $X \leq Y$ alors $\mathbb{E}[X] \leq \mathbb{E}[Y]$.
2. Si $X \in L^1$, alors $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.
3. Si $A \in \mathcal{A}$, alors $\mathbb{E}[1_A] = \mathbb{P}[A]$ et si $\lambda \in \mathbb{R}$, $\mathbb{E}[\lambda] = \lambda$.
4. Si $X \geq 0$ et $\mathbb{E}[X] = 0$ alors $\mathbb{P}[X = 0] = 1$.

Si $h : \mathbb{R} \rightarrow \mathbb{R}$ est continue par morceaux et bornée, alors

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x)f(x)dx.$$

Par exemple, si $X \in L^2$, on peut définir sa variance :

$$\text{var}(X) = \int_{\mathbb{R}} (x - \mathbb{E}[X])^2 f(x)dx = \int_{\mathbb{R}} x^2 f(x)dx - \left(\int_{\mathbb{R}} xf(x)dx \right)^2.$$

Proposition 11.7 *L'espace L^2 est un sous-espace vectoriel de L^1 et*

$$(X, Y) \in L^2 \times L^2 \mapsto \mathbb{E}[XY]$$

est bien définie et une forme bilinéaire symétrique positive.

Tous les résultats concernant variance, écart-type et covariance restent vraies.

12 Indépendance

La notion de conditionnement est beaucoup plus délicate dans le cadre général, donc nous ne l'évoquerons pas ici. En revanche, l'indépendance s'adapte très bien.

Définition 12.1 *Les v.a. X_1, \dots, X_n sont dites indépendantes si pour tous les $x_1, \dots, x_n \in \mathbb{R}$, on a*

$$\mathbb{P}[\cap_{k=1}^n (X_k \leq x_k)] = \prod_{i=1}^n \mathbb{P}[X_k \leq x_k].$$

Une suite infinie de v.a. $(X_n)_{n \geq 1}$ est dite indépendante si toutes ses sous-suites finies le sont.

Exemple 12.1 *Tout réel $x \in [0, 1]$ admet une écriture en base 10, ce qui signifie qu'il existe une suite (x_n) à valeurs dans $\{0, 1, \dots, 9\}$ telle que*

$$x = \sum_{n \geq 0} x_n 10^{-n}.$$

Lorsque x n'est pas décimal, cette suite est unique. Lorsque x est décimal, on a deux écritures du fait que

$$\sum_{n \geq 0} 9 \times 10^{-n} = \frac{9 \times 10^{-1}}{1 - 10^{-1}} = 1.$$

Donc, pour un décimal, on a une écriture telle que $x_n = 0$ à partir d'un certain rang et une écriture telle que $x_n = 9$ à partir d'un certain rang. Choisissons la première par commodité. On peut alors définir $X_n(x) = x_n$. J'affirme que $(X_n)_n$ est une suite de v.a. indépendantes.

Les résultats précédents se déclinent comme suit.

Théorème 12.2 *Soient X_1, \dots, X_n des v.a. indépendantes, A_1, \dots, A_n des boréliens et f_1, \dots, f_n des applications de \mathbb{R} dans \mathbb{R} continues par morceaux.*

1. *On a $\mathbb{P}[\cap (X_i \in A_i)] = \prod \mathbb{P}[X_i \in A_i]$.*
2. *La suite $f_1(X_1), \dots, f_n(X_n)$ est une suite de v.a. indépendantes.*
3. *Si les f_j sont bornées, alors*

$$\mathbb{E}[f_1(X_1) \times \dots \times f_n(X_n)] = \mathbb{E}[f_1(X_1)] \times \dots \times \mathbb{E}[f_n(X_n)].$$

4. *Si X_1, \dots, X_n sont de carré intégrable, alors*

$$\text{var}(X_1 + \dots + X_n) = \text{var}(X_1) + \dots + \text{var}(X_n).$$

13 L'aiguille de Buffon

L'aiguille de Buffon est une expérience de probabilité proposée en 1733 par Georges-Louis Leclerc de Buffon, un scientifique français du XVIII^e siècle. Cette expérience fournit une approximation du nombre π . Son analyse met en œuvre un cas simple d'espace de probabilité bidimensionnel et continu.

On s'intéresse au problème suivant : on se donne une aiguille de longueur $\ell > 0$ que l'on jette sur un parquet dont les lames font une largeur $d > \ell$. On cherche à déterminer la probabilité que l'aiguille soit à cheval sur deux lames.

Modélisons ce problème : le parquet peut être modélisé par des droites horizontales sur le plan euclidien \mathbb{R}^2 espacées de la distance d . La position de l'aiguille sera donnée par les coordonnées (x, y) de sa tête et l'angle $\theta \in [0, 2\pi[$ que fait l'aiguille avec l'horizontale.

L'abscisse x n'a aucun intérêt dans notre problème (le parquet est régulier) et ce qui nous importe dans l'ordonnée est uniquement sa position par rapport aux lames : on peut donc considérer que y prend ses valeurs dans l'intervalle $[0, d[$. Notre espace des issues sera

$$\Omega = [0, d[\times [0, 2\pi[.$$

Comme c'est un espace continu, on le munit de la tribu des boréliens $\mathcal{B}(\Omega)$: la plus petite tribu qui contient les pavés $[a, b] \times [s, t]$, où $0 \leq a < b < d$ et $0 \leq s < t < 2\pi$.

On introduit les deux v.a. de positions : Y et Θ sur Ω . Nous n'avons pas de raison de privilégier une position plutôt qu'une autre, donc nous supposons que Y et Θ sont deux v.a. indépendantes et que chacune suivent une loi uniforme. Par conséquent, si $0 \leq a < b < d$ et $0 \leq s < t < 2\pi$, alors

$$\mathbb{P}[(Y, \Theta) \in [a, b] \times [s, t]] = \mathbb{P}[Y \in [a, b]] \times \mathbb{P}[\Theta \in [s, t]] = \frac{|b - a|}{d} \times \frac{|t - s|}{2\pi}.$$

Sous ces conditions, on a

Théorème 13.1 *Soit E l'événement « l'aiguille est à cheval sur deux lames ». On a*

$$\mathbb{P}[E] = \frac{2\ell}{\pi d}.$$

Nous proposons une démonstration due à Erdős qui repose sur très peu de calculs.

Notations.— De fait, la position de l'aiguille est donnée par trois coordonnées (x, y, θ) dans \mathbb{R}^3 . Mais toutes nos considérations ne dépendront ni de x , ni de la position relative de y par rapport à d ni de θ par rapport à 2π . Il sera cependant plus pratique parfois de considérer que nos v.a. sont définies sur \mathbb{R}^3 . Si $X : \Omega \rightarrow \mathbb{R}$, on peut définir $\tilde{X} : \mathbb{R}^3 \rightarrow \mathbb{R}$ en posant $\tilde{X}(x, y + kd, \theta + 2\pi n) = X(y, \theta)$, où $(y, \theta) \in \Omega$, $x \in \mathbb{R}$ et $k, n \in \mathbb{Z}$. Réciproquement, si $\tilde{X} : \mathbb{R}^3 \rightarrow \mathbb{R}$ est une fonction borélienne telle que $\tilde{X}(x, y + d, \theta + 2\pi) = \tilde{X}(x', y, \theta)$ pour tous $x, x', y, \theta \in \mathbb{R}$, on peut définir $X : \Omega \rightarrow \mathbb{R}$ en posant $X(y, \theta) = \tilde{X}(0, y, \theta)$. Nous écrirons $\tilde{X} = X$ pour ne pas alourdir les notations.

Réduction à un calcul d'espérance. — La première chose à remarquer est que si $N_\ell : \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{N}$ désigne le nombre d'intersection d'une aiguille de longueur ℓ avec les rainures du parquet, alors, comme $\ell < d$, on a $\mathbb{P}[E] = \mathbb{E}[N_\ell]$. En effet, ou bien l'aiguille ne touche qu'une seule lame, alors $\mathbb{1}_E = 0 = N_\ell$ ou bien elle est à cheval sur deux lames, donc une seule rainure donc $N_\ell = 1 = \mathbb{1}_E$.

Nous allons donc calculer $\mathbb{E}[N_\ell]$. Pour ce calcul, nous ne nous restreindrons pas aux petites valeurs de ℓ .

La démonstration s'appuie sur les deux propositions suivantes.

Proposition 13.2 *On a $\mathbb{E}[N_\ell] = \ell \mathbb{E}[N_1]$.*

Proposition 13.3 *Si A est une aiguille « tordue » (une ligne brisée) de longueur $\ell > 0$ que l'on lance sous les mêmes conditions que l'aiguille droite, et si on note N_A le nombre d'intersection que fait A avec les rainures, alors $\mathbb{E}[N_A] = \mathbb{E}[N_\ell]$.*

Montrons maintenant comment cela implique le théorème : considérons un cercle C de diamètre d . Pour chaque $n \geq 3$, on considère les polygones réguliers à n côtés P_n et Q_n qui sont inscrit et exinscrit à C .

On voit P_n et Q_n comme des aiguilles tordues que l'on lance aléatoirement sur notre parquet. A chaque position de P_n , le cercle C qu'il inscrit intersecte toujours les rainures exactement deux fois. Par conséquent, on a toujours $N_{P_n} \leq 2$. De même, quelle que soit la position de Q_n , le cercle inscrit intersecte aussi les rainures exactement deux fois, donc on a toujours $2 \leq N_{Q_n}$.

Donc $N_{P_n} \leq 2 \leq N_{Q_n}$. En prenant les espérances on trouve $\mathbb{E}[N_{P_n}] \leq 2 \leq \mathbb{E}[N_{Q_n}]$, c'est-à-dire, par les propositions 13.2 et 13.3,

$$\ell(P_n) \times \mathbb{E}[N_1] \leq 2 \leq \ell(Q_n) \times \mathbb{E}[N_1].$$

Or, quand n tend vers l'infini, on a

$$\lim_{n \rightarrow \infty} \ell(P_n) = \lim_{n \rightarrow \infty} \ell(Q_n) = \ell(C) = \pi d$$

donc

$$\mathbb{E}[N_1] = \frac{2}{\pi d}$$

et pour $\ell < d$, on trouve

$$\mathbb{P}[E] = \mathbb{E}[N_\ell] = \frac{2\ell}{\pi d}.$$

Nous allons maintenant démontrer les propositions.

Il est pratique d'introduire la notation suivante : si $\theta \in [0, 2\pi[$, notons u_θ le vecteur unitaire faisant un angle θ avec l'axe $0x$. De plus, puisque N ne dépend pas de x , nous oublierons de le mentionner dans la suite.

Lemme 13.4 *Soient $\alpha, t \in \mathbb{R}$ et $\tau_{(t,\alpha)} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ définie par $\tau_{(t,\alpha)}(x, y, \theta) = (y + t, \theta + \alpha)$. On considère une fonction $s : \mathbb{R} \rightarrow \mathbb{R}$. On a*

$$\mathbb{E}[N_\ell \circ \tau_{(s(\Theta), \alpha)}] = \mathbb{E}[N_\ell].$$

Pour montrer ce lemme, il nous faut énoncer un résultat général que nous n'avons pas encore vu : Si X, Y sont deux v.a. indépendantes de densité f_X et f_Y respectivement, et si $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ est borélienne bornée, alors on a

$$\mathbb{E}[h(X, Y)] = \int_{\mathbb{R}^2} h(x, y) f_X(x) f_Y(y) dx dy.$$

DÉMONSTRATION. (lemme 13.4) Cela découle de la formule de changement de variables : on a

$$\mathbb{E}[N_\ell \circ \tau_{(s(\Theta), \alpha)}] = \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{d} \left(\int_0^d N_\ell(y + s(\theta), \theta + \alpha) dy \right) d\theta;$$

Or, à θ fixé,

$$\begin{aligned} \int_0^d N_\ell(y + s(\theta), \theta + \alpha) dy &= \int_0^{d-s(\theta)} N_\ell(y + s(\theta), \theta + \alpha) dy + \int_{d-s(\theta)}^d N_\ell(y + s(\theta), \theta + \alpha) dy \\ &= \int_{s(\theta)}^d N_\ell(z, \theta + \alpha) dz + \int_0^{s(\theta)} N_\ell(w, \theta + \alpha) dw \\ &= \int_0^d N_\ell(y, \theta + \alpha) dy \end{aligned}$$

Dans la première intégrale, on a posé $z = y + s(\theta)$ et dans la seconde, on a posé $w = y - (d - s(\theta))$ et on a utilisé que $N_\ell(w + d, \theta) = N_\ell(w, \theta)$ pour tous $w, \theta \in \mathbb{R}$ (le nombre d'intersection de l'aiguille ne change pas si on la translate de la largeur des lames du parquet). Donc on a, en procédant de la même manière avec θ :

$$\begin{aligned} \mathbb{E}[N_\ell \circ \tau_{(s(\Theta), \alpha)}] &= \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{d} \left(\int_0^d N_\ell(y, \theta + \alpha) dy \right) d\theta \\ &= \frac{1}{d} \int_0^d \left(\frac{1}{2\pi} \int_0^{2\pi} N_\ell(y, \theta + \alpha) d\theta \right) dy \\ &= \frac{1}{d} \int_0^d \left(\frac{1}{2\pi} \int_0^{2\pi} N_\ell(y, \theta) d\theta \right) dy \\ &= \mathbb{E}[N_\ell]. \end{aligned}$$

■

Remarque.— On a montré implicitement dans le lemme 13.4 que $Z = Y + s(\Theta)$ suit la même loi que Y et qu'elle est indépendante de Θ .

Lemme 13.5 Si $\ell = \ell_1 + \ell_2$, alors $\mathbb{E}[N_\ell] = \mathbb{E}[N_{\ell_1}] + \mathbb{E}[N_{\ell_2}]$.

DÉMONSTRATION. On considère une aiguille de longueur ℓ que l'on voit comme deux aiguilles : une de longueur ℓ_1 suivie d'une autre de longueur ℓ_2 . Du coup, si l'aiguille initiale a pour coordonnées (x, y, θ) alors la première aiguille a pour coordonnées (x, y, θ) et la seconde $((x, y) +$

$\ell_1 u_\theta$), θ). Si on note $\ell_1 u_\theta = (a(\theta), b(\theta))$ alors, pour une issue $\omega = (y, \theta)$, on aura $N_\ell(\omega) = N_{\ell_1}(\omega) + N_{\ell_2}(\tau_{(b(\theta), 0)}(\omega))$, donc

$$\mathbb{E}[N_\ell] = \mathbb{E}[N_{\ell_1}] + \mathbb{E}[N_{\ell_2}(\tau_{(b(\theta), 0)})].$$

D'après le lemme 13.4, on a $\mathbb{E}[N_{\ell_2}(\tau_{(b(\theta), 0)})] = \mathbb{E}[N_{\ell_2}]$ donc

$$\mathbb{E}[N_\ell] = \mathbb{E}[N_{\ell_1}] + \mathbb{E}[N_{\ell_2}].$$

■

DÉMONSTRATION. (proposition 13.2) Soit $\ell > 0$. Une récurrence à partir du lemme 13.5 montre que si $n \geq 1$ alors $\mathbb{E}[N_{n\ell}] = n\mathbb{E}[N_\ell]$ et donc, en prenant $\ell = 1/n$, on obtient $\mathbb{E}[N_{1/n}] = \mathbb{E}[N_1]/n$ et, pour $p, q \geq 1$, $\mathbb{E}[N_{(p/q)}] = p\mathbb{E}[N_{(1/q)}] = (p/q)\mathbb{E}[N_1]$. Si $\ell > 0$ est réel, on trouve deux suites de rationnels (p_n/q_n) et (p'_n/q'_n) qui tendent vers ℓ de sorte que $(p_n/q_n) \leq \ell \leq (p'_n/q'_n)$. Donc on trouve $(p_n/q_n)\mathbb{E}[N_1] \leq \mathbb{E}[N_\ell] \leq (p'_n/q'_n)\mathbb{E}[N_1]$. Le résultat s'obtient en passant à la limite. ■

DÉMONSTRATION. (proposition 13.3) La démonstration est similaire à celle du lemme 13.5. Une aiguille « tordue » peut être décrite à partir de ses points de brisure (z_1, \dots, z_{n-1}) : on se donne chaque fois leur position ainsi que les angles que font les aiguilles avec l'axe Ox . L'aiguille est donc décrite par sa tête z_0 , et une suite finie (ℓ_j, α_j) telle que la j ème partie de l'aiguille a pour longueur ℓ_j et fait un angle α_j avec Ox . Si on connaît la position de la tête z_0 , on trouve la position des points de brisure $(z_j)_{1 \leq j \leq n-1}$ par récurrence : on aura $z_j = z_{j-1} + \ell_j u_{\alpha_j}$. Si on tourne d'un angle θ l'aiguille, les nouvelles positions seront $(\ell_j, \alpha_j + \theta)$ et les points de brisures seront $z_0(\theta) = z_0$ et $z_j(\theta) = z_{j-1}(\theta) + \ell_j u_{\alpha_j + \theta}$. Les données qui nous intéressent sont, outre les angles, les ordonnées $y_j(\theta)$ des points $z_j(\theta)$. Pour chaque θ et chaque point de brisure $j \in \{1, \dots, n-1\}$, nous avons un nombre $s_j(\theta)$ tel que $y_j(\theta) = y_j(0) + s_j(\theta)$.

Ceci nous permet d'écrire pour une issue $\omega = (y_0, \theta)$,

$$N_A(\omega) = \sum_{j=1}^{n-1} N_{\ell_j}(\tau_{(s_j(\theta), \alpha_j)}(\omega))$$

et donc

$$\mathbb{E}[N_A] = \sum_{j=1}^{n-1} \mathbb{E}[N_{\ell_j}(\tau_{(s_j(\theta), \alpha_j)})].$$

D'après le lemme 13.4 et la proposition 13.2, on a

$$\mathbb{E}[N_A] = \sum_{j=1}^{n-1} \mathbb{E}[N_{\ell_j}] = \mathbb{E}[N_\ell].$$

■

Chapitre V. Théorèmes limites

On se fixe un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. On s'intéresse dans ce chapitre au comportement asymptotique de suites de v.a. (X_n) .

14 Loi des grands nombres

La loi des grands nombres traduit l'idée que si on lance un grand nombre de fois un dé non pipé, la proportion de 2 qui apparaît est de l'ordre de $1/6$. Elle affirme la chose suivante :

Théorème 14.1 *Soit (X_n) une suite de v.a.i.i.d. intégrables d'espérance m . Alors, pour tout $\varepsilon > 0$, on a*

$$\lim_{n \rightarrow \infty} \mathbb{P}[|\overline{X}_n - m| \geq \varepsilon] = 0$$

où

$$\overline{X}_n = \frac{1}{n} \sum_{j=0}^{n-1} X_j.$$

Nous allons montrer ce résultat lorsque la suite est de carré intégrable. Cela nous permettra d'établir des résultats intermédiaires utiles.

Proposition 14.2 *Soit X une v.a. et soit $\lambda > 0$.*

– **Inégalité de Markoff.** *Si X est intégrable, alors on a*

$$\mathbb{P}[|X| \geq \lambda] \leq \frac{1}{\lambda} \mathbb{E}[|X|].$$

– **Inégalité de Bienaymé-Tchebycheff.** *Si X est de carré intégrable, alors on a*

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq \lambda] \leq \frac{1}{\lambda^2} \text{var}(X).$$

DÉMONSTRATION. Supposons d'abord X intégrable. On remarque

$$\begin{aligned} |X| &= |X| \mathbb{1}_{|X| \geq \lambda} + |X| \mathbb{1}_{|X| < \lambda} \\ &\geq \lambda \mathbb{1}_{|X| \geq \lambda}. \end{aligned}$$

En utilisant la positivité de l'espérance, il vient

$$\mathbb{E}[|X|] \geq \mathbb{E}[\lambda \mathbb{1}_{|X| \geq \lambda}];$$

or

$$\mathbb{E}[\lambda \mathbb{1}_{|X| \geq \lambda}] = \lambda \mathbb{P}[|X| \geq \lambda]$$

d'où l'inégalité de Markoff.

Si X est de carré intégrable, alors nous allons appliquer l'inégalité de Markoff à $(X - \mathbb{E}[X])^2$ en remarquant que les événements $\{|X - \mathbb{E}[X]| \geq \lambda\}$ et $\{|X - \mathbb{E}[X]|^2 \geq \lambda^2\}$ coïncident :

$$\begin{aligned} \mathbb{P}[|X - \mathbb{E}[X]| \geq \lambda] &= \mathbb{P}[|X - \mathbb{E}[X]|^2 \geq \lambda^2] \\ &\leq \frac{1}{\lambda^2} \text{var}(X). \end{aligned}$$

■

DÉMONSTRATION. (loi des grands nombres) Nous supposons donc la suite de carré intégrable. Notons m leur espérance et σ^2 leur variance (communes).

On remarque que, pour tout $n \geq 1$,

$$\mathbb{E}[\overline{X}_n] = \frac{1}{n} \sum_{j=0}^{n-1} \mathbb{E}[X_j] = m$$

et, puisque les (X_j) sont indépendants,

$$\text{var}(\overline{X}_n) = \frac{1}{n^2} \text{var} \left(\sum_{j=0}^{n-1} X_j \right) = \frac{1}{n^2} \sum_{j=0}^{n-1} \text{var}(X_j) = \frac{1}{n} \sigma^2.$$

Par conséquent, si $\varepsilon > 0$, alors l'inégalité de Bienaymé-Tchebycheff nous dit

$$\mathbb{P}[|\overline{X}_n - m| \geq \varepsilon] \leq \frac{\sigma^2}{n\varepsilon^2}$$

donc la limite est nulle.

■

Il existe une version forte de la loi des grands nombres que nous énonçons maintenant :

Théorème 14.3 Soit (X_n) une suite de v.a.i.i.d. intégrables d'espérance m . Alors on a

$$\mathbb{P}[\lim_{n \rightarrow \infty} \overline{X}_n = m] = 1$$

où

$$\overline{X}_n = \frac{1}{n} \sum_{j=0}^{n-1} X_j.$$

Aiguille de Buffon. — La loi des grands nombres affirme donc que l'on peut approcher la valeur de π en lançant un grand nombre de fois une aiguille sur un parquet, cf. le théorème 13.1. En fait, la convergence est très lente, comme le montre les résultats suivants (avec $\ell = d$) :

Nombre de lancers	1000	1000	1000	10000	10000	10000
approximation de π	3.17	2.96	3.09	3.136	3.099	3.168

La page <http://www-sop.inria.fr/mefisto/java/tutorial1/node14.html> contient un petit programme de simulation.

15 Convergence en loi

Nous allons illustrer la notion de convergence en loi, qui traduit l'évolution de la loi de X_n quand n tend vers l'infini.

Définition 15.1 Soient $(X_n)_n$ une suite de v.a. et X une autre v.a. On dira que (X_n) tend en loi vers X si

1. pour tout $x \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \mathbb{P}[X_n \leq x] = \mathbb{P}[X \leq x]$$

si X est à densité;

2. pour tout $k \in \mathbb{N}$,

$$\lim_{n \rightarrow \infty} \mathbb{P}[X_n = k] = \mathbb{P}[X = k]$$

si (X_n) et X sont à valeurs dans \mathbb{N} .

Dans cet énoncé, seule la loi de X importe : le rôle de X est mineur ; on interprète la loi de X comme le comportement asymptotique de la loi de X_n pour n grand. Cette convergence ne donne aucun contrôle sur la convergence de $X_n(\omega)$ vers $X(\omega)$. On dira parfois que la loi de X_n tend vers « telle loi » pour dire qu'il existe une v.a. X de cette loi-là telle que (X_n) tend en loi vers X .

Nous allons maintenant illustrer cette notion sur quelques exemples.

Proposition 15.2 Si X_n suit la loi uniforme sur $\{0, \dots, n\}$, alors la loi de (X_n/n) tend vers la loi uniforme sur $[0, 1]$.

DÉMONSTRATION. On note $\lfloor x \rfloor$ la partie entière du réel x , c'est-à-dire l'entier relatif n tel que $n \leq x < n + 1$. Prenons $0 \leq a < b \leq 1$. On a

$$\left\{ a < \frac{X_n}{n} \leq b \right\} = \{na \leq X_n \leq nb\} = \{\lfloor na \rfloor + 1 \leq X_n \leq \lfloor nb \rfloor\}.$$

Donc

$$\mathbb{P} \left[a < \frac{X_n}{n} \leq b \right] = \mathbb{P}[\lfloor na \rfloor + 1 \leq X_n \leq \lfloor nb \rfloor] = \frac{\lfloor nb \rfloor - \lfloor na \rfloor - 1}{n + 1}.$$

Or

$$\frac{(nb - 1) - na - 1}{n} \leq \frac{\lfloor nb \rfloor - \lfloor na \rfloor - 1}{n} \leq \frac{nb - (na - 1) - 1}{n + 1}$$

donc

$$\lim \mathbb{P} \left[a < \frac{X_n}{n} \leq b \right] = (b - a).$$

■

Proposition 15.3 Si X_n suit la loi binomiale $\mathcal{B}(n, p_n)$ et si (np_n) tend vers $\lambda > 0$, alors la loi de (X_n) tend vers la loi de Poisson $\mathcal{P}(\lambda)$.

DÉMONSTRATION. Nous allons calculer un équivalent. On remarque que puisque (np_n) tend vers une valeur finie, $(p_n)_n$ tend forcément vers zéro. Prenons $k \in \mathbb{N}$ et $n \geq k$. On a

$$\begin{aligned} \mathbb{P}[X_n = k] &= \frac{n!}{k!(n-k)!} p_n^k (1-p_n)^{n-k} \\ &= \frac{\lambda^k}{k!} \left(\frac{np_n}{\lambda}\right)^k \frac{n(n-1)\dots(n-k+1)}{n^k} e^{(n-k)\ln(1-p_n)} \\ &\sim \frac{\lambda^k}{k!} e^{-\lambda} \end{aligned}$$

car

$$(n-k)\ln(1-p_n) \sim -np_n + kp_n \sim -\lambda.$$

■

Proposition 15.4 *On suppose que $(X_k^n)_k$ est une suite i.i.d. de loi $\mathcal{B}(p_n)$ et on suppose que (np_n) tend vers $\lambda > 0$. Notons pour chaque $n \geq 1$*

$$T_n = \inf\{k, X_k^n = 1\}.$$

La loi de (T_n/n) tend vers la loi exponentielle de paramètre λ .

DÉMONSTRATION. La loi de T_n est la loi géométrique de paramètre p_n . Du coup

$$\begin{aligned} \mathbb{P}[T_n \geq k] &= \sum_{j \geq k} p_n (1-p_n)^{j-1} \\ &= p_n \frac{(1-p_n)^{k-1}}{1-(1-p_n)} = (1-p_n)^{k-1}. \end{aligned}$$

Du coup, si $t > 0$,

$$\begin{aligned} \mathbb{P}\left[\frac{T_n}{n} > t\right] &= \mathbb{P}[T_n > nt] \\ &= (1-p_n)^{\lfloor nt \rfloor} = e^{\lfloor nt \rfloor \ln(1-p_n)} \\ &\sim e^{-p_n \lfloor nt \rfloor}. \end{aligned}$$

Maintenant,

$$np_n t \leq p_n \lfloor nt \rfloor \leq (nt+1)p_n$$

donc

$$\lim \mathbb{P}\left[\frac{T_n}{n} > t\right] = e^{-\lambda t} = \int_t^\infty \lambda e^{-\lambda s} ds,$$

ce qui suffit pour conclure.

■

16 Théorème central limite

Nous nous intéressons à la limite en loi d'une somme de v.a. (X_n) i.i.d. de carré intégrable. Notons $S_n = \sum_{j=0}^{n-1} X_j$, $m = \mathbb{E}[X_1]$ et $\sigma^2 = \text{var}(X_1)$. On remarque tout d'abord que

$$\mathbb{E}[S_n] = \sum_{j=0}^{n-1} \mathbb{E}[X_j] = nm$$

et

$$\text{var}(S_n) = \sum_{j=0}^{n-1} \text{var}(X_j) = n\sigma^2.$$

Si $m, \sigma \neq 0$, alors la suite (S_n) ne pourra pas avoir de limite en loi de carré intégrable ni même intégrable. En revanche, la suite

$$\left(\frac{S_n - nm}{\sigma\sqrt{n}} \right)_n$$

est centrée (espérance nulle et variance normalisée à 1). La question de la limite en loi de cette suite devient maintenant raisonnable.

16.1 Loi de Bernoulli équilibrée

On commence par le théorème de de Moivre (1733) :

Théorème 16.1 *Si (X_n) est une suite de v.a.i.i.d. de loi $\mathcal{B}(1/2)$ alors $\{(S_n - n/2)/\sqrt{n/4}\}_n$ tend en loi vers une v.a. normale réduite centrée.*

DÉMONSTRATION. La démonstration s'appuiera sur la formule de Stirling fournissant un équivalent de $n!$:

$$n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n}.$$

Notons que de Moivre avait établi cet équivalent où la constante $\sqrt{2\pi}$ lui était restée inconnue. Stirling, un contemporain de de Moivre, détermina la constante inconnue.

On sait que S_n suit la loi $\mathcal{B}(n, 1/2)$. Par conséquent, on a, pour $0 \leq k \leq n$,

$$\mathbb{P}[S_n = k] = C_n^k \frac{1}{2^n}.$$

Du coup, si $a < b$, alors

$$\begin{aligned} \mathbb{P}\left[a \leq \frac{S_n - n/2}{\sqrt{n/4}} \leq b\right] &= \mathbb{P}\left[\frac{a\sqrt{n} + n}{2} \leq S_n \leq \frac{b\sqrt{n} + n}{2}\right] \\ &= \frac{1}{2^n} \sum_{\frac{a\sqrt{n} + n}{2} \leq k \leq \frac{b\sqrt{n} + n}{2}} C_n^k. \end{aligned}$$

Pour appliquer la formule de Stirling aux C_n^k , on remarque que k et $(n - k)$ sont équivalents à $n/2$ indépendamment de k . Cela nous permet d'écrire

$$\begin{aligned} C_n^k &= \frac{n!}{k!(n-k)!} \\ &\sim \frac{\left(\frac{n}{e}\right)^n \sqrt{2\pi n}}{\left(\frac{k}{e}\right)^k \sqrt{2\pi k} \left(\frac{n-k}{e}\right)^{n-k} \sqrt{2\pi(n-k)}} \\ &\sim \frac{1}{\sqrt{2\pi n}} \frac{1}{\sqrt{(k/n)(1-k/n)[(k/n)^{k/n}(1-k/n)^{1-k/n}]^n}}. \end{aligned}$$

Si

$$\frac{a\sqrt{n} + n}{2} \leq k \leq \frac{b\sqrt{n} + n}{2}$$

alors

$$\frac{a}{2\sqrt{n}} + \frac{1}{2} \leq \frac{k}{n} \leq \frac{b}{2\sqrt{n}} + \frac{1}{2}.$$

Pour chaque $0 \leq k \leq n$ dans cet intervalle, on trouve donc $x_k \in [a, b]$ tel que

$$\frac{k}{n} = \frac{1}{2} \left(1 + \frac{x_k}{\sqrt{n}}\right).$$

On a aussi

$$1 - \frac{k}{n} = \frac{1}{2} \left(1 - \frac{x_k}{\sqrt{n}}\right).$$

On remarque que les x_k sont bornés indépendamment de k et n . Du coup, en écrivant temporairement $x = x_k$.

$$\sqrt{\frac{k}{n} \left(1 - \frac{k}{n}\right)} = \frac{1}{2} \sqrt{1 - \frac{x^2}{n}} \sim \frac{1}{2} \left(1 - \frac{x^2}{2n}\right) \sim \frac{1}{2}$$

et

$$\left[\left(\frac{k}{n}\right)^{k/n} \left(1 - \frac{k}{n}\right)^{1-k/n} \right]^n = \frac{1}{2^n} \left[\left(1 + \frac{x}{\sqrt{n}}\right)^{1+x/\sqrt{n}} \left(1 - \frac{x}{\sqrt{n}}\right)^{1-x/\sqrt{n}} \right]^{n/2}.$$

Or

$$\left(1 + \frac{x}{\sqrt{n}}\right) \ln \left(1 + \frac{x}{\sqrt{n}}\right) = \left(1 + \frac{x}{\sqrt{n}}\right) \left(\frac{x}{\sqrt{n}} - \frac{x^2}{2n} + O\left(\frac{1}{n\sqrt{n}}\right)\right)$$

et

$$\left(1 - \frac{x}{\sqrt{n}}\right) \ln \left(1 - \frac{x}{\sqrt{n}}\right) = \left(1 - \frac{x}{\sqrt{n}}\right) \left(-\frac{x}{\sqrt{n}} - \frac{x^2}{2n} + O\left(\frac{1}{n\sqrt{n}}\right)\right)$$

donc

$$\begin{aligned} \frac{n}{2} \left[\left(1 + \frac{x}{\sqrt{n}}\right) \ln \left(1 + \frac{x}{\sqrt{n}}\right) + \left(1 - \frac{x}{\sqrt{n}}\right) \ln \left(1 - \frac{x}{\sqrt{n}}\right) \right] \\ = \frac{n}{2} \left[\frac{x^2}{n} + O\left(\frac{1}{n\sqrt{n}}\right) \right] = \frac{x^2}{2} + O\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

On en déduit ainsi

$$\mathbb{P}[S_n = k] \sim \frac{2}{\sqrt{2\pi n}} e^{-x_k^2/2}$$

et

$$\mathbb{P}\left[a \leq \frac{S_n - n/2}{\sqrt{n/4}} \leq b\right] \sim \frac{1}{\sqrt{2\pi}} \frac{2}{\sqrt{n}} \sum_{\frac{a\sqrt{n}+n}{2} \leq k \leq \frac{b\sqrt{n}+n}{2}} e^{-x_k^2/2}.$$

Rappelons que $x_k = 2\frac{k}{\sqrt{n}} - \sqrt{n}$, donc cette suite est espacée de $2/\sqrt{n}$ avec pour premier terme $a + O(1/\sqrt{n})$ et dernier terme $b + O(1/\sqrt{n})$. Ecrivons $f(x) = e^{-x^2/2}$ et posons $h = 2/\sqrt{n}$ alors on obtient

$$\mathbb{P}\left[a \leq \frac{S_n - n/2}{\sqrt{n/4}} \leq b\right] \sim \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\lfloor (b-a)/h \rfloor} f(a + kh).$$

On reconnaît ainsi les sommes de Riemann de la fonction f prises entre a et b . Par conséquent on obtient

$$\mathbb{P}\left[a \leq \frac{S_n - n/2}{\sqrt{n/4}} \leq b\right] \longrightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$$

■

16.2 La version générale

En fait, le théorème de de Moivre est un cas particulier d'un théorème beaucoup plus général —le théorème central limite :

Théorème 16.2 *Soit $(X_n)_n$ une suite de v.a.i.i.d. de carré intégrable, de moyenne m et variance σ^2 . Notons*

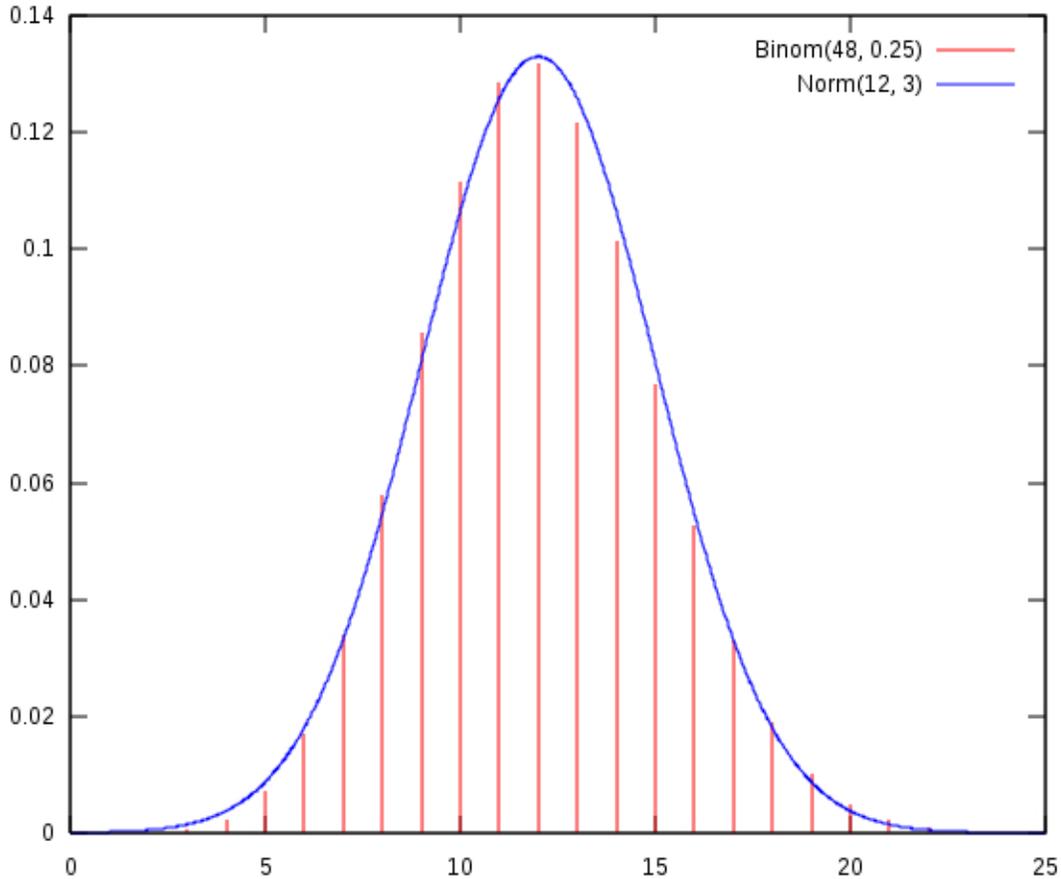
$$S_n = \sum_{j=0}^{n-1} X_j \quad \text{et} \quad Z_n = \frac{S_n - nm}{\sqrt{n}\sigma}.$$

Alors (Z_n) tend en loi vers une v.a. normale réduite centrée.

La démonstration est très différente, mais reste assez technique si on n'utilise pas de théories plus élaborées (la formule de Stirling n'est plus utile).

Un point important est d'estimer la vitesse de convergence dans le théorème central limite (voir plus bas).

Voici une illustration du théorème : on répète un nombre important de fois l'expérience suivante : on tire à pile ou face 48 fois une pièce qui a une chance sur quatre de faire pile et on compte le nombre de piles. Chaque bâton correspond à la proportion du nombre d'expériences ayant comptabilisé le même nombre de piles. On trace sur le même graphique la densité de la gaussienne de moyenne $48/4 = 12$ et de variance $(1/4)(3/4)48 = 9$ (la figure indique l'écart-type de la gaussienne).



Le théorème de Berry-Esséen permet d'estimer la vitesse de convergence dans le théorème central limite, mais il requiert un moment d'ordre 3 :

Théorème 16.3 Soit $(X_n)_n$ une suite de variables indépendantes identiquement distribuées, admettant des moments jusqu'à l'ordre 3, c'est-à-dire $\mathbb{E}[|X|^3] < \infty$. Notons $m = \mathbb{E}[X_1]$, $\sigma^2 = \text{var}(X_1)$,

$$S_n = \sum_{j=0}^{n-1} X_j \quad \text{et} \quad Z_n = \frac{S_n - nm}{\sqrt{n}\sigma}.$$

Alors, pour tout $x \in \mathbb{R}$, on a

$$\left| \mathbb{P}[Z_n \leq x] - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du \right| \leq \frac{1}{2} \frac{\mathbb{E}[|X_1 - m|^3]}{\sigma^3 \sqrt{n}}.$$

Une autre manière de formuler cette condition est la suivante :

$$\left| \mathbb{P}[S_n \leq x] - \frac{1}{\sqrt{2\pi n\sigma}} \int_{-\infty}^x e^{-\frac{(u-m)^2}{2\sigma^2}} du \right| \leq \frac{1}{2} \frac{\mathbb{E}[|X_1 - m|^3]}{\sigma^3 \sqrt{n}}.$$

La table de la loi normale nous indique que si X est normale réduite centrée alors

$$\mathbb{P}[|X| \geq 2] \leq 5\%.$$

Par conséquent, si (X_n) est une suite de v.a.i.i.d. de carré intégrable, alors, en reprenant les notations du théorème central limite, pour n assez grand, Z_n suit approximativement la loi $\mathcal{N}(0, 1)$, donc

$$\mathbb{P} \left[\left| \frac{S_n}{n} - m \right| \leq \frac{2\sigma}{\sqrt{n}} \right] \geq 95\% .$$

Cela nous informe sur la convergence de S_n/n vers la moyenne m dans la loi des grands nombres.

16.3 Somme de gaussiennes indépendantes

Nous allons conclure ce paragraphe par quelques résultats sur les v.a. gaussiennes :

Proposition 16.4 *La somme (finie) de v.a. gaussiennes indépendantes est aussi gaussienne.*

Si X_j suit la loi $\mathcal{N}(m_j, \sigma_j^2)$ alors $\sum_{j=1}^n X_j$ suivra la loi $\mathcal{N}(\sum m_j, \sum \sigma_j^2)$.

17 Approximation de la loi binomiale

La théorie central limite et la proposition 15.3 nous donnent deux comportements possibles d'une v.a. de loi binomiale $\mathcal{B}(n, p)$ avec n grand.

En pratique, si $n \geq 30$ et $n \min\{p, 1-p\} \geq 20$, alors on remplacera la loi binomiale par la loi $\mathcal{N}(np, np(1-p))$ et si $pn < 15$, $p \leq 0.1$ et $n \geq 30$, alors on considèrera plutôt la loi de Poisson $\mathcal{P}(np)$. La loi de Poisson est utilisée pour décrire des phénomènes rares (p petit).

Le théorème 16.3 et l'inégalité de Le Cam (proposition 17.1 ci-dessous) ne permettent pas de vérifier que ces approximations sont effectivement raisonnables.

17.1 Paradigme de Poisson

Le paradigme de Poisson affirme que la somme S_n d'un grand nombre de variables de Bernoulli indépendantes de petit paramètre suit approximativement la loi de Poisson de paramètre $\mathbb{E}[S_n]$, voir la proposition 15.3. Ce résultat est précisé par l'inégalité suivante de Le Cam (on énonce un cas particulier) :

Proposition 17.1 *Soit $p \in]0, 1[$ et soit $(X_n)_n$ une suite de v.a.i.i.d. de loi $\mathcal{B}(p)$; notons $S_n = \sum_{j=1}^n X_j$ On a*

$$\sum_{k \geq 0} \left| \mathbb{P}[S_n = k] - \frac{(np)^\ell e^{-np}}{\ell!} \right| \leq 2np^2$$

et pour tout ensemble $A \subset \mathbb{N}$,

$$\left| \mathbb{P}[S_n \in A] - \sum_{\ell \in A} \frac{(np)^\ell e^{-np}}{\ell!} \right| \leq np^2 .$$

En particulier, si $(np_n) \rightarrow \lambda > 0$ comme dans la proposition 15.3, alors on contrôle la convergence vers la loi de Poisson de paramètre λ (puisque alors np_n^2 tend vers 0).

DÉMONSTRATION. Nous allons mettre en œuvre la méthode dite de « couplage » très employée en probabilités qui consiste ici à coupler des v.a. de Bernoulli et de Poisson de façon à ce qu'elles soient égales en grande probabilité (au moins $1 - p^2$).

Posons sur $([0, 1], \mathcal{B}([0, 1]), \mathcal{U}([0, 1]))$,

$$X = \mathbb{1}_{[1-p, 1]}$$

et

$$Y = \sum_{k \geq 0} k \mathbb{1}_{A_k}$$

où

$$A_k = \left[e^{-p} \sum_{j=0}^{k-1} \frac{p^j}{j!}, e^{-p} \sum_{j=0}^k \frac{p^j}{j!} \right].$$

On vérifie facilement que X suit la loi $\mathcal{B}(p)$ et Y la loi $\mathcal{P}(p)$.

On constate que l'événement $(X = Y)$ correspond à $[0, 1-p]$ (car $1-p \leq e^{-p}$) où $X = Y = 0$ et à $[e^{-p}, (1+p)e^{-p}[$ où $X = Y = 1$. Du coup $\mathbb{P}[X = Y] = (1-p) + pe^{-p}$ et

$$\mathbb{P}[X \neq Y] = p - pe^{-p} \leq p^2.$$

Le couplage consiste à considérer la loi μ du couple $Z = (X, Y)$, c'est-à-dire la loi sur $\{0, 1\} \times \mathbb{N}$ telle que $\mu(k, \ell) = \mathbb{P}[X = k, Y = \ell]$ et de considérer une suite i.i.d. $(Z_n) = (X_n, Y_n)$ de loi μ . D'après ci-dessus, pour tout k , on a $\mathbb{P}[X_k \neq Y_k] = \mathbb{P}[X \neq Y] \leq p^2$.

On note $S_n = \sum_{j=1}^n X_j$ et $W_n = \sum_{j=1}^n Y_j$. La v.a. S_n suit la loi $\mathcal{B}(n, p)$ et W_n la loi $\mathcal{P}(np)$.

Si $k \geq 0$, alors $\{S_n = k\} = \{S_n = k; S_n = W_n\} \cup \{S_n = k; S_n \neq W_n\}$. Donc

$$\mathbb{P}[S_n = k] = \mathbb{P}[S_n = k; W_n = S_n] + \mathbb{P}[S_n = k; W_n \neq S_n]$$

et par symétrie

$$\mathbb{P}[W_n = k] = \mathbb{P}[W_n = k; W_n = S_n] + \mathbb{P}[W_n = k; W_n \neq S_n].$$

Par conséquent

$$\begin{aligned} |\mathbb{P}[S_n = k] - \mathbb{P}[W_n = k]| &= |\mathbb{P}[S_n = k; W_n \neq S_n] - \mathbb{P}[W_n = k; W_n \neq S_n]| \\ &\leq \mathbb{P}[S_n = k; W_n \neq S_n] + \mathbb{P}[W_n = k; W_n \neq S_n] \end{aligned}$$

Or $\cup_{k \geq 0} \{S_n = k; W_n \neq S_n\} = \{S_n \neq W_n\}$, donc

$$\sum_{k \geq 0} \mathbb{P}[S_n = k; W_n \neq S_n] = \mathbb{P}[W_n \neq S_n]$$

et de même, on a

$$\sum_{k \geq 0} \mathbb{P}[W_n = k; W_n \neq S_n] = \mathbb{P}[W_n \neq S_n].$$

Par conséquent,

$$\sum_{k \geq 0} |\mathbb{P}[S_n = k] - \mathbb{P}[W_n = k]| \leq 2\mathbb{P}[W_n \neq S_n].$$

Par ailleurs, si $S_n \neq W_n$ alors il existe au moins un $k \leq n$ tel que $X_k \neq Y_k$. On en déduit que

$$\{S_n \neq W_n\} \subset \bigcup_{1 \leq k \leq n} \{X_k \neq Y_k\}.$$

Finalement

$$\mathbb{P}[S_n \neq W_n] \leq \sum_{1 \leq k \leq n} \mathbb{P}[X_k \neq Y_k] \leq np^2.$$

De là, on obtient

$$\sum_{k \geq 0} |\mathbb{P}[S_n = k] - \mathbb{P}[W_n = k]| \leq 2np^2.$$

Si $A \subset \{0, 1, \dots, n\}$, en remarquant d'une part que $\{W_n \in A\} \supset \{W_n \in A \text{ et } S_n \in A\}$ et d'autre part que $\{S_n \in A\} \setminus \{W_n \in A \text{ et } S_n \in A\} = \{S_n \in A; W_n \notin A\} \subset \{S_n \neq W_n\}$, on a

$$\begin{aligned} \mathbb{P}[S_n \in A] - \mathbb{P}[W_n \in A] &\leq \mathbb{P}[S_n \in A] - \mathbb{P}[W_n \in A \text{ et } S_n \in A] \\ &= \mathbb{P}[S_n \in A \text{ et } W_n \notin A] \\ &\leq \mathbb{P}[S_n \neq W_n] \end{aligned}$$

et, en échangeant les rôles de W_n et de S_n ,

$$|\mathbb{P}[S_n \in A] - \mathbb{P}[W_n \in A]| \leq \mathbb{P}[S_n \neq W_n].$$

Comme ci-dessus, on conclut

$$|\mathbb{P}[S_n \in A] - \mathbb{P}[W_n \in A]| \leq \sum_{1 \leq k \leq n} \mathbb{P}[X_k \neq Y_k] \leq np^2.$$

■

17.2 L'état de Californie contre Collins

C'est un procès célèbre qui a eu lieu en 1968 car le procureur s'est appuyé officiellement sur un raisonnement probabiliste pour faire condamner les accusés.

Dans la ville de Los Angeles, une femme se fait arracher son sac à main dans la rue. Des témoins de la scène décrivent les assaillants comme étant un couple, dont l'homme est noir, portant moustache et barbe, accompagné d'une femme blanche, cheveux blonds attachés en queue de cheval, qui se seraient enfuis au volant d'une voiture jaune. La police trouve un couple satisfaisant cette description et les voilà conduits au tribunal.

Après avoir appelé à la barre un « instructeur » en mathématiques qui a expliqué la règle de « multiplication des probabilités » (la probabilité de l'intersection d'événements indépendants est le produit de leurs probabilités), le procureur s'est livré au raisonnement suivant.

Parmi la population de L.A., le procureur a estimé les proportions suivantes :

- Homme noir avec barbe : 1/10.
- Homme moustachu : 1/4.
- Femme blanche avec queue de cheval : 1/10.
- Femme blanche avec cheveux blonds : 1/3.
- Voiture jaune : 1/10.
- Couple interracial dans une même voiture : 1/1000.

Au total (en multipliant toutes ces probabilités), une chance sur 12.000.000 d'avoir un tel couple!

Sur ce, le couple a été condamné.

Finalement, la Cour Suprême de Californie a annulé ce jugement, estimant que les événements ci-dessus n'étaient pas indépendants, et qu'un raisonnement mathématique ne pouvait que servir à étayer d'autres preuves, mais ne pouvait être la cause centrale établissant la culpabilité.

En fait, la dépendance des événements est loin d'être le seul problème de ce raisonnement. Au mieux, ce que le procureur a calculé est la probabilité qu'un couple pris au hasard dans la population vérifie tous ces critères simultanément! La réponse est donc : c'est très rare. Autrement dit, la police a eu de la chance de trouver un tel couple. Mais, ce n'est pas parce que ces couples sont rares qu'ils sont forcément coupables!

La question qui aurait dû être posée est : sachant qu'il en existe un, quelle est la probabilité qu'il y en ait un autre? Désignons par X le nombre de couples à L.A. ayant ces caractéristiques. Si on considère que 5 millions de couples vivent dans le « grand Los Angeles » et que la probabilité qu'un couple ait ces caractéristiques vaut bien $1/12.000.000$, alors X devrait suivre une loi binomiale $\mathcal{B}(5.000.000, 1/12.000.000)$. Comme la probabilité est très faible, nous approximations cette loi par la loi de Poisson $\mathcal{P}(5/12)$. On s'intéresse alors à

$$\mathbb{P}[X \geq 2 | X \geq 1] = \frac{\mathbb{P}[X \geq 2]}{\mathbb{P}[X \geq 1]} \simeq 0.19.$$

Ce petit calcul montre qu'il y a un véritable doute sur la culpabilité de ce couple, si elle est uniquement basée sur ces critères.

Chapitre VI. Statistique

L'objet de la statistique est d'obtenir des informations sur une population entière à partir de données partielles (prises sur un échantillon). Ce type de problème apparaît dans de nombreuses situations.

18 La démarche statistique

Prenons quelques exemples.

Exemple 18.1 *On sonde 1000 personnes en leur demandant leur intention de vote au second tour des élections présidentielles afin de déterminer les intentions de vote de toute la population.*

Une autre situation similaire : une municipalité veut organiser le feu d'artifice du 14 juillet et doit acheter des pièces pyrotechniques. Elle a le choix entre plusieurs fabricants. Son souci est de s'assurer que les pièces fonctionneront bien. Elle achète donc des échantillons (réduits) à plusieurs producteurs et teste leur bon fonctionnement.

Exemple 18.2 *Un fabricant d'ampoules veut tester leur durée de vie. Il s'agit donc d'en mesurer un certain nombre et d'en déduire la durée de tout son lot.*

Exemple 18.3 *La ville de Marseille mesure régulièrement la qualité de l'eau de mer afin d'autoriser ou non la baignade. Elle extrait un certain nombre d'échantillons à des endroits bien choisis. Il s'agit ensuite de mesurer la teneur en différents produits toxiques, afin d'en déduire la teneur générale pour enfin prendre une décision.*

Exemple 18.4 *L'agence sanitaire teste l'efficacité de médicaments avant d'autoriser leur mise sur le marché. Un certain nombre de tests seront faits sur des cobayes visant à établir leur efficacité et mesurer les effets secondaires.*

Dans toutes ces situations, nous avons un certain nombre de mesures x_1, \dots, x_n à partir desquelles nous allons travailler. Nous les interprétons comme la réalisation d'une suite de v.a.i.i.d X_1, \dots, X_n , c'est-à-dire, on aura une issue ω d'un espace de probabilité telle que $X_j(\omega) = x_j$, pour $1 \leq j \leq n$.

Définition 18.1 (échantillon) *Un échantillon de taille $n \geq 1$ est la donnée de n v.a.i.i.d. X_j . Leur loi commune (et inconnue) est la loi mère de l'échantillon. Les valeurs x_j représentent une réalisation de l'échantillon.*

On considèrera aussi des fonctions de l'échantillon.

Définition 18.2 (Statistique) *Une statistique est une v.a. qui ne dépend que de l'échantillon, c'est-à-dire une fonction $f(X_1, \dots, X_n)$ dont les arguments sont les v.a. qui forment l'échantillon.*

Le point important d'une statistique est que l'on peut l'évaluer dès que nous avons une réalisation (afin d'obtenir un nombre).

Voici deux exemples importants et classiques de statistiques.

1. La moyenne empirique est définie par

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j.$$

2. La variance empirique est définie par

$$\tilde{S}_n = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X}_n)^2.$$

19 Estimation

La donnée est un échantillon de taille n et on suppose que la loi mère fait partie d'une famille $(\mathbb{P}_\theta)_\theta$ bien identifiée. On souhaite déterminer la loi mère, donc le paramètre θ qui convient.

Définition 19.1 (Modèle statistique) *Un modèle statistique est le choix d'une famille de probabilités (\mathbb{P}_θ) parmi lesquelles on veut modéliser les observations.*

Exemple 18.1.— Dans ce cas, on supposera que l'échantillon suit une loi de Bernoulli $\mathcal{B}(\theta)$. Dans le premier cas, on choisira un des candidats prenant la valeur 0 et l'autre la valeur 1. Dans le second cas, on associera la valeur 1 si la pièce a bien explosé, et 0 sinon.

Exemple 18.2.— Ici, il est plus naturel de considérer que la loi mère suit une loi exponentielle.

Exemple 18.3.— On veut par exemple connaître la concentration μ en pesticides dans l'eau. Pour cela, on prélève dix échantillons de ce produit, dont on détermine la concentration. On obtient ainsi dix mesures (x_1, \dots, x_{10}) probablement différentes de μ , chaque dosage étant entaché d'erreurs. On a donc observé le vecteur aléatoire $X = (X_1, \dots, X_{10})$ avec $X_i = \mu + \epsilon_i$, où ϵ_i représente les erreurs de dosage. On dispose des informations *a priori* suivantes :

- la loi de chaque ϵ_i est indépendante de μ (les erreurs de dosages ne dépendent pas de la concentration des pesticides) ;
- les variables aléatoires ϵ_i sont indépendantes (les erreurs de mesure faites lors d'un dosage ne se répercutent pas sur les dosages suivants) ;
- les ϵ_i ont même loi (la méthode de dosage utilisée est toujours la même) ;
- la moyenne des ϵ_i est nulle (la méthode de dosage n'introduit pas une erreur systématique) ;
- la variance σ^2 des ϵ_i est connue ou non, ce qui signifie que l'on connaît la précision de la méthode de dosage utilisée ou non.

On suppose de plus que la loi commune des ϵ_i est la loi normale (de moyenne 0, et de variance σ^2). Cette hypothèse peut se justifier par le TCL, si l'on considère que les erreurs de dosage proviennent de l'accumulation d'un grand nombre de petites erreurs indépendantes.

Définition 19.2 *Un estimateur $\hat{\theta}_n$ de θ est une statistique dont la réalisation après expérience est envisagée comme estimation de θ . On obtient ainsi l'estimation $\hat{\theta}(x_1, \dots, x_n)$.*

Insistons sur le fait qu'un estimateur est avant tout une variable aléatoire. Elle ne dépend que de l'échantillon, ce qui fait qu'on peut l'évaluer dès qu'on nous en soumet une réalisation. Le nombre ainsi calculé produit une estimation de θ .

Nous introduisons deux mesures de pertinence d'un estimateur.

Définition 19.3 (biais) *Soit $\hat{\theta}_n$ un estimateur de θ . Son biais est*

$$b_\theta = \mathbb{E}_\theta[\hat{\theta}_n] - \theta.$$

On dit que $\hat{\theta}_n$ est sans biais si $b_\theta = 0$ pour tout θ .

Définition 19.4 (consistance) *On dit que $(\hat{\theta}_n)_n$ est un estimateur consistant si $\hat{\theta}_n$ tend vers θ quand n tend vers l'infini.*

La limite ci-dessus peut être comprise comme dans la loi des grands nombres.

Nous allons présenter deux manières de trouver des estimateurs : par la méthode des moments et par le maximum de vraisemblance.

19.1 Méthode des moments

La méthode des moments consiste à exprimer θ comme l'espérance d'une statistique, et cette statistique sera l'estimateur recherché. Autrement dit, on cherche à exprimer θ sous la forme

$$\theta = \mathbb{E}_\theta[\hat{\theta}_n].$$

Dans ce cas, l'estimateur est automatiquement sans biais. Dans de nombreux cas, la loi des grands nombres montrera sa consistance.

Exemple 18.1.— Si la loi mère est une loi de Bernoulli $\mathcal{B}(p)$, alors on a $\mathbb{E}[X_1] = p$ et donc $\mathbb{E}[\bar{X}_n] = p$ aussi. Du coup \bar{X}_n est un estimateur de θ , sans biais et consistant (par la loi des grands nombres).

Exemple 18.2.— Si la loi mère est une loi exponentielle $\mathcal{E}(\lambda)$, alors on a $\mathbb{E}[X_1] = 1/\lambda$. Il paraît plus judicieux de paramétrer la loi exponentielle par l'inverse de λ , c'est-à-dire poser $\theta = 1/\lambda$. Dans ce cas, la moyenne empirique est aussi un estimateur sans biais et consistant.

Exemple 18.3.— Dans le cas d'une loi normale $\mathcal{N}(m, \sigma^2)$, les paramètres sont m et σ^2 . Pour le premier, la moyenne empirique convient. Regardons maintenant si la variance empirique est un estimateur de σ^2 par la méthode des moments (autrement dit s'il est sans biais).

On calcule d'abord \tilde{S}_n :

$$\begin{aligned}
 \tilde{S}_n &= \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X}_n)^2 \\
 &= \frac{1}{n} \sum_{j=1}^n (X_j^2 - 2X_j\bar{X}_n + \bar{X}_n^2) \\
 &= \frac{1}{n} \sum_{j=1}^n X_j^2 - \frac{2}{n} \sum_{j=1}^n X_j\bar{X}_n + \frac{1}{n} \sum_{j=1}^n \bar{X}_n^2 \\
 &= \frac{1}{n} \sum_{j=1}^n X_j^2 - 2\bar{X}_n^2 + \bar{X}_n^2 \\
 &= \frac{1}{n} \sum_{j=1}^n X_j^2 - \bar{X}_n^2.
 \end{aligned}$$

Or

$$\begin{aligned}
 \bar{X}_n^2 &= \frac{1}{n^2} \left(\sum_{j=1}^n X_j \right)^2 \\
 &= \frac{1}{n^2} \left(\sum_{j=1}^n X_j^2 \right) + 2 \left(\sum_{1 \leq i < j \leq n} X_i X_j \right).
 \end{aligned}$$

Donc

$$\tilde{S}_n = \frac{n-1}{n^2} \left(\sum_{j=1}^n X_j^2 \right) - \frac{2}{n} \left(\sum_{1 \leq i < j \leq n} X_i X_j \right).$$

Du coup

$$\begin{aligned}
 \mathbb{E}[\tilde{S}_n] &= \frac{1}{n} \sum_{j=1}^n \mathbb{E}[(X_j - \bar{X}_n)^2] \\
 &= \frac{n-1}{n^2} \left(\sum_{j=1}^n \mathbb{E}[X_j^2] \right) - \frac{2}{n} \left(\sum_{1 \leq i < j \leq n} \mathbb{E}[X_i X_j] \right) \\
 &= \frac{n-1}{n^2} \left(\sum_{j=1}^n (\sigma^2 + m^2) \right) - \frac{2}{n} \left(\sum_{1 \leq i < j \leq n} \mathbb{E}[X_i] \mathbb{E}[X_j] \right) \\
 &= \frac{n-1}{n} (\sigma^2 + m^2) - \frac{2}{n} \left(\sum_{1 \leq i < j \leq n} m^2 \right) \\
 &= \frac{n-1}{n} (\sigma^2 + m^2) - \frac{2}{n} \left(\frac{(n-1)n}{2} m^2 \right) \\
 &= \frac{n-1}{n} \sigma^2.
 \end{aligned}$$

Du coup, \tilde{S}_n a un biais.

On considère alors *la variance de l'échantillon*

$$S_n = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2.$$

On trouve maintenant

$$\mathbb{E}[S_n] = \frac{n}{n-1} \mathbb{E}[\tilde{S}_n] = \sigma^2$$

de sorte que S_n est sans biais.

Il reste à montrer la consistance de S_n ou de sa version \tilde{S}_n . On a vu ci-dessus que

$$\tilde{S}_n = \frac{1}{n} \sum_{j=1}^n X_j^2 - \bar{X}_n^2.$$

Or la loi des grands nombres s'applique à chaque terme pour montrer qu'ils tendent respectivement vers $\mathbb{E}[X_1^2]$ et m^2 respectivement, donc $(\tilde{S}_n)_n$ tend bien vers σ^2 .

19.2 Méthode du maximum de vraisemblance

L'idée de cette méthode est de choisir θ de sorte que la probabilité de la réalisation soit maximale. Nous allons distinguer les deux situations où X_1 est à valeurs dans \mathbb{N} ou est à densité.

19.2.1 Échantillon à valeurs entières

Nous avons X_1, \dots, X_n un échantillon de loi mère définie sur \mathbb{N} et on suppose donnée une réalisation x_1, \dots, x_n .

On introduit la *fonction de vraisemblance*

$$V : \theta \mapsto \mathbb{P}_\theta[X_1 = x_1; X_2 = x_2, \dots, X_n = x_n].$$

On appelle *estimateur* du maximum de vraisemblance l'estimateur $\hat{\theta}_n$ qui maximise V lorsqu'il existe, c'est-à-dire

$$V(\hat{\theta}) = \sup_{\theta} \mathbb{P}_\theta[X_1 = x_1; X_2 = x_2, \dots, X_n = x_n].$$

Remarquez qu'ici, le calcul conduit en fait à une estimation, en fonction des x_i , de θ . On obtient l'estimateur en remplaçant la réalisation par l'échantillon.

En général, il est plus commode de travailler avec la fonction *log-vraisemblance* $L = \ln V$ car on transforme ainsi un produit en somme. Pour trouver le maximum, si on montre que V ou L est de classe C^2 et concave (dérivée seconde strictement négative) alors on aura un unique minimum qui sera la valeur —lorsqu'elle existe— qui annule la dérivée.

Prenons un exemple.

Exemple 18.1.— Notons $s_n = \sum x_j$. On note

$$L(\theta) = \ln \mathbb{P}_\theta[X_1 = x_1, \dots, X_n = x_n] = \ln \theta^{s_n} (1 - \theta)^{n - s_n} = s_n \ln \theta + (n - s_n) \ln(1 - \theta).$$

On calcule les dérivées successives :

$$L'(\theta) = \frac{s_n}{\theta} - \frac{n - s_n}{1 - \theta}$$

et

$$L''(\theta) = -\frac{s_n}{\theta^2} - \frac{n - s_n}{(1 - \theta)^2} < 0.$$

Donc L est concave et l'estimation est donnée par l'annulation de L' : $L'(\theta) = 0$ si

$$\theta = \frac{s_n}{n}.$$

On vient de montrer que la moyenne empirique était aussi l'estimateur du maximum de vraisemblance.

19.2.2 Echantillon de loi à densité

On suppose maintenant que la loi mère fait partie d'une famille de loi de densité $(f_\theta)_\theta$. La méthode du maximum de vraisemblance consiste maintenant à choisir θ pour maximiser la fonction de vraisemblance suivante

$$V : \theta \mapsto \prod_{j=1}^n f_\theta(x_j).$$

Considérons le problème de trouver un estimateur pour m et σ dans une modélisation par une loi gaussienne. On écrit

$$V(m, \sigma) = \prod_{j=1}^n \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_j - m)^2}{2\sigma^2}} \right).$$

On considère alors la fonction log-vraisemblance

$$L(m, \sigma) = -n \log \sqrt{2\pi}\sigma - \sum_{j=1}^n \frac{(x_j - m)^2}{2\sigma^2}.$$

En dérivant selon m , on obtient

$$\frac{\partial L}{\partial m} = 2 \sum_{j=1}^n \frac{x_j - m}{2\sigma^2}$$

et

$$\frac{\partial^2 L}{\partial m^2} = -\sum_{j=1}^n \frac{1}{\sigma^2} < 0$$

donc on a un unique minimum qui correspond à

$$\sum_{j=1}^m \frac{x_j - m}{\sigma^2} = 0$$

donc on trouve

$$m = \frac{1}{n} \sum_{j=1}^n x_j$$

c'est-à-dire la moyenne empirique.

Cherchons maintenant la variance :

$$\frac{\partial L}{\partial \sigma} = -\frac{n}{\sigma} + 2 \sum_{j=1}^m \frac{(x_j - \bar{x}_n)^2}{2\sigma^3}.$$

On vérifie que cette dérivée est positive pour $\sigma > 0$ petit et négative ensuite. On a bien un maximum là où la dérivée s'annule et on trouve

$$\sigma^2 = \frac{1}{n} \sum_{j=1}^m (x_j - \bar{x}_n)^2$$

soit la variance empirique.

20 Intervalle de confiance

On suppose donner un modèle statistique et que l'on a calculé un estimateur $\hat{\theta}_n$. La question qui se pose est : quelle est la fiabilité de cet estimateur ? Autrement dit, si le paramètre est θ , quelle est la probabilité

$$\mathbb{P}_\theta[|\theta - \hat{\theta}_n| \leq \varepsilon]$$

où $\varepsilon > 0$ est un terme d'erreur que l'on veut contrôler ?

Définition 20.1 (intervalle de confiance) Soit X_1, \dots, X_n un échantillon. Un processus d'intervalle de confiance de niveau $(1 - \alpha) \in [0, 1]$ est la donnée de deux statistiques $T_- < T_+$ telles que, pour tout θ ,

$$\mathbb{P}_\theta[T_- \leq \theta \leq T_+] \geq 1 - \alpha.$$

On obtient un intervalle de confiance en évaluant T_- et T_+ sur une réalisation de l'échantillon.

En général, on choisit α proche de 0 : 0.05 ou 0.01. On interprète ceci en disant, qu'avec la probabilité $(1 - \alpha)$, θ est dans l'intervalle $[T_-, T_+]$.

Lorsque $\hat{\theta}_n$ est sans biais, on peut utiliser l'inégalité de Bienaymé-Tchebycheff pour obtenir

$$\mathbb{P}_\theta[|\theta - \hat{\theta}_n| \geq \varepsilon] \leq \frac{1}{\varepsilon^2} \text{var}(\hat{\theta}_n).$$

Par exemple, si $\hat{\theta}_n$ est la moyenne empirique, alors

$$\text{var}(\hat{\theta}_n) = \frac{1}{n} \text{var}(X_1).$$

Donc, pour avoir un intervalle de confiance de niveau $1 - \alpha$, il suffit d'avoir

$$\frac{1}{n\varepsilon^2} \text{var}(X_1) \leq \alpha.$$

Le processus d'intervalle de confiance sera de la forme, es supposant $\text{var}(X_1)$ connu,

$$[\bar{X}_n - \varepsilon, \bar{X}_n + \varepsilon].$$

Loi de Bernoulli.— Lorsque la loi mère est une loi de Bernoulli, alors on a $\text{var}(X_1) = \theta(1 - \theta) \leq (1/4)$, donc, pour obtenir un intervalle de confiance de niveau $1 - \alpha$, il suffit d'avoir

$$n\varepsilon^2 \geq 4\alpha.$$

Selon que n est imposé ou non, on choisit ε pour avoir l'estimation demandée.

Dans les sondages en politique, on interroge environ 1000 personnes. Donc, pour avoir un niveau de confiance de 95%, on obtient

$$\varepsilon \simeq \sqrt{\frac{4}{0.05 \times 1000}} \simeq 3\%.$$

Autrement dit, avec 95% de chances, on obtient les intentions de vote dans un intervalle de 6%! C'est énorme, quand on entend les journalistes discuter des variations au point près.

Loi normale.— Supposons que l'on connaisse la variance de la loi mère (qui est une gaussienne). Alors, $\sqrt{n}(\bar{X}_n - m)/\sigma$ suit la loi $\mathcal{N}(0, 1)$. On peut alors lire sur la table que pour avoir un intervalle de niveau 95%, on doit choisir $\varepsilon = 1.96$. Du coup, on aura

$$\mathbb{P}[-1.96\sigma/\sqrt{n} + \bar{X}_n \leq m \leq \bar{X}_n + 1.96\sigma/\sqrt{n}] \geq 95\%.$$

On obtient ainsi l'intervalle de confiance $[-1.96\sigma/\sqrt{n} + \bar{X}_n, \bar{X}_n + 1.96\sigma/\sqrt{n}]$.

On dit que $\sqrt{n}(\bar{X}_n - m)/\sigma$ est un *pivot* pour deux raisons :

1. sa loi est connue (dans notre cas $\mathcal{N}(0, 1)$);
2. on peut trouver un encadrement de la donnée recherchée à partir de $a \leq \sqrt{n}(\bar{X}_n - m)/\sigma \leq b$.

Intervalle de confiance asymptotique.— L'idée est d'appliquer le théorème central limite pour obtenir un intervalle de confiance *asymptotique* de la moyenne empirique. On rappelle que la loi de

$$Z_n = \frac{n\bar{X}_n - nm}{\sqrt{n}\sigma}$$

est approximativement $\mathcal{N}(0, 1)$. Donc, comme ci-dessus, on aura un intervalle asymptotique

$$[\bar{X}_n - 1.96\sigma/\sqrt{n}, \bar{X}_n + 1.96\sigma/\sqrt{n}].$$

Si on reprend le cas Bernoulli avec $n = 1000$ et en utilisant que $\sigma \leq 1/2$ pour tout θ , alors

$$\varepsilon \simeq \frac{1.96}{2\sqrt{1000}} \simeq 3\%.$$

On obtient donc la même estimation (sauf que l'on n'a pas estimé l'erreur dans le théorème central limite).

21 Tests d'hypothèse

Un test d'hypothèse est une procédure statistique qui vise à donner une réponse « oui » ou « non » à une question. Par exemple, aux questions suivantes :

1. Une pièce est-elle juste ?
2. La fabrication qui sort de l'usine contient-elle une proportion acceptable d'appareils en état de marche ?
3. Le nouveau médicament qui vient d'être conçu est-il efficace pour soigner la maladie visée ?

21.1 Le principe des tests : le cas de deux hypothèses simples

Il s'agit du problème d'école, qui ne correspond pas vraiment aux problèmes que l'on rencontre dans la réalité, mais qui permet d'exposer simplement les principes des tests statistiques. Supposons pour fixer les idées que l'on joue n fois à Pile ou Face, avec une pièce dont on ne sait pas si la probabilité d'obtenir Pile vaut 0 ou 1. On suppose par contre que le paramètre inconnu vaut l'une des deux valeurs connues θ_0 ou θ_1 , et ne peut pas valoir autre chose. Supposons par exemple que $\theta_0 < \theta_1$. Alors on a envie de prendre comme règle de décision d'accepter l'hypothèse $H_0 : \theta = \theta_0$ si

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n} < K;$$

et de rejeter H_0 (donc d'accepter l'hypothèse alternative $H_1 : \theta = \theta_1$) dans le cas contraire, *i. e.* si $\bar{X}_n \geq K$. Tout le problème maintenant est de choisir le seuil K . Le problème du test est que l'on voudrait minimiser la probabilité de prendre la mauvaise décision.

21.1.1 Deux types d'erreur

Dans un test, il y a deux façons de se tromper :

1. Soit H_0 est vraie, et on la rejette à tort.
2. Soit H_1 est vraie, et on accepte à tort H_0 .

La première erreur est appelée *erreur de 1ère espèce* et la seconde *de 2ième espèce*. La probabilité de commettre cette erreur (bien sûr lorsque H_0 est vraie) vaut

$$\mathbb{P}_{H_0}[\overline{X}_n \geq K]$$

et la probabilité de commettre une erreur de 2ème espèce vaut quant à elle

$$\mathbb{P}_{H_1}[\overline{X}_n < K].$$

Notons que plus K est grand, plus la probabilité de commettre une erreur de 1ère espèce est petite, et au contraire plus K est petit, plus la probabilité de commettre une erreur de seconde espèce est petite. On ne peut donc pas minimiser les deux types d'erreur à la fois ! Pour sortir de ce dilemme, Neymann a proposé de faire jouer un rôle dissymétrique aux deux types d'erreur. On va choisir K de sorte que

$$\mathbb{P}_{H_0}[\overline{X}_n \geq K] \leq \alpha$$

où le niveau $1 - \alpha$ est à choisir (typiquement, $\alpha = 0.05$ ou $\alpha = 0.01$). De ce fait, on contrôle la probabilité de commettre une erreur de première espèce, mais pas du tout celle de commettre une erreur de deuxième espèce ! Cette dissymétrie doit être prise en compte dans la modélisation. Si l'on décide qu'il y a plus de risque à livrer une fabrication douteuse, plutôt qu'à la mettre au rebut, et à mettre sur le marché un médicament inefficace, plutôt qu'à tirer un trait sur x années de recherches, on choisira

1. H_0 = la pièce n'est pas juste,
2. H_0 = la fabrication est défectueuse,
3. H_0 = le médicament n'est pas efficace,

avec l'espoir à chaque fois de rejeter cette hypothèse, parce que l'on contrôle ainsi la probabilité de mettre sur le marché la fabrication (resp. le médicament) alors qu'elle est défectueuse (resp. il est inefficace). De même, dans le problème de contrôle de la pollution, on doit choisir H_0 = la pollution a atteint un niveau dangereux pour la population. De même dans le cas du jeu de Pile ou Face, s'il s'agit de tester si oui ou non la pièce est juste, en suivant le principe que nous venons d'exposer, il faudrait choisir $H_0 : \theta = 0.5$, car en cas de doute, il vaut mieux la jeter et en tester une nouvelle, plutôt que de risquer d'utiliser une pièce qui n'est pas juste. Cependant, on verra que ce choix n'est pas possible, et comment traiter le problème (on étudiera la probabilité d'erreur de 2ème espèce).

Remarque 21.1 *Une fois que l'on a fixé la probabilité d'erreur de première espèce, que peut-on faire pour minimiser l'erreur de seconde espèce ? On peut faire deux remarques à ce sujet. D'une part, il faut baser le test sur une statistique (une fonction des données) qui discrimine le mieux possible les deux hypothèses, autrement dit qui a tendance à prendre des valeurs très différentes suivant que l'une ou l'autre des deux hypothèses est réalisée. D'autre part, on peut améliorer la situation en augmentant la taille de l'échantillon. On rediscutera ce point sur un exemple particulier à la section suivante.*

21.1.2 Le test entre deux hypothèses simples dans le jeu de Pile ou Face

Revenons au problème de tester l'hypothèse $H_0 = \{\theta = \theta_0\}$ contre l'alternative $H_1 = \{\theta = \theta_1\}$, dans le cas $\theta_0 < \theta_1$. D'après ce qui précède, on doit choisir K tel que

$$\mathbb{P}_{H_0}[\overline{X}_n \geq K] \leq \alpha.$$

Or

$$[\overline{X}_n \geq K] = \left[\frac{\overline{X}_n - \theta_0}{\sqrt{\theta_0(1-\theta_0)/n}} \geq \frac{K - \theta_0}{\sqrt{\theta_0(1-\theta_0)/n}} \right]$$

donc, dès que n atteint ou dépasse quelques dizaines, on peut utiliser l'approximation du théorème central limite, qui nous dit que

$$\begin{aligned} \mathbb{P}_{H_0}[\overline{X}_n \geq K] &= \mathbb{P}_{H_0} \left[\frac{\overline{X}_n - \theta_0}{\sqrt{\theta_0(1-\theta_0)/n}} \geq \frac{K - \theta_0}{\sqrt{\theta_0(1-\theta_0)/n}} \right] \\ &\simeq \int_{\frac{K - \theta_0}{\sqrt{\theta_0(1-\theta_0)/n}}}^{\infty} e^{-x^2/2} \frac{dx}{\sqrt{2\pi}} \leq \alpha \end{aligned}$$

et donc on peut choisir K en fonction de α , à partir de la table de la loi $\mathcal{N}(0, 1)$. C'est-à-dire que l'on détermine la région de rejet du test $[K; +\infty[$ en fonction de α . Si \overline{X}_n appartient à cette région de rejet, on rejette H_0 . Sinon, on accepte H_0 . Une autre approche, très prisée des biologistes, médecins et pharmaciens, est de s'appuyer pour prendre la décision sur ce que l'on appelle la p-valeur. Au lieu de comparer la valeur observée \overline{x}_n de la moyenne \overline{X}_n , à la valeur K , on peut, ce qui revient exactement au même, comparer la p-valeur

$$p = \mathbb{P}_{H_0}[\overline{X}_n > \overline{x}_n]$$

à α . Plus précisément, on rejette H_0 si et seulement si $p < \alpha$, autrement dit on accepte H_0 si et seulement si $p \geq \alpha$. En effet, si $p < \alpha$, alors cela signifie que l'on a l'inclusion

$$\{\overline{X}_n > K\} \subset \{\overline{X}_n > \overline{x}_n\}$$

et donc $\overline{x}_n > K$, et inversement si $p < \alpha$.

21.1.3 Test du rapport de vraisemblance

On a vu sur l'exemple ci-dessus qu'un test est basé sur la valeur prise par une statistique, par exemple un estimateur du paramètre inconnu. Le choix de cette statistique de test est important. Comme on l'a déjà dit, pour que la probabilité d'erreur de 2ème espèce soit la plus faible possible, il faut choisir une statistique qui ait toutes les chances de prendre des valeurs nettement différentes, suivant que $\theta = \theta_0$ ou $\theta = \theta_1$. Dans le cas du jeu de Pile ou Face, on ne voit guère quoi prendre d'autre que \overline{X}_n . Plus généralement, dans le cas du test entre deux hypothèses simples, Neymann et Pearson ont montré que le choix optimal consiste à prendre comme statistique de

test *le rapport de vraisemblance*, à savoir le quotient de la vraisemblance sous H_1 , divisée par la vraisemblance sous H_0 , soit

$$T = \frac{V(\theta_1)}{V(\theta_0)}.$$

De façon générale, on s'attend à ce que T prenne une valeur plus grande que 1 si les observations ont été recueillies sous la probabilité \mathbb{P}_{θ_1} , et au contraire plus petite que 1 si les observations ont été recueillies sous \mathbb{P}_{θ_0} .

D'où la région de rejet de H_0 de la forme $\{T > t\}$. Le test que nous avons proposé pour le jeu de Pile ou Face est bien un test du rapport de vraisemblance. En effet, rappelons que

$$V(\theta) = \theta^{n\bar{X}_n} (1 - \theta)^{n - n\bar{X}_n} = \left(\frac{\theta}{1 - \theta} \right)^{n\bar{X}_n} (1 - \theta)^n.$$

La région de rejet du test du rapport de vraisemblance est de la forme

$$\left(\frac{\theta_1(1 - \theta_0)}{\theta_0(1 - \theta_1)} \right)^{n\bar{X}_n} > t \left(\frac{1 - \theta_0}{1 - \theta_1} \right)^n$$

soit

$$\bar{X}_n > \frac{\log t + n \log \frac{1 - \theta_0}{1 - \theta_1}}{n \log \frac{\theta_1(1 - \theta_0)}{\theta_0(1 - \theta_1)}}$$

qui est bien de la forme $\bar{X}_n > K$, puisque les deux logarithmes ci-dessus sont positifs dans le cas $\theta_0 < \theta_1$.

21.2 Test entre hypothèses multiples

Dans les situations importantes en pratique, les deux hypothèses H_0 et H_1 ne se réduisent pas à une seule valeur du paramètre. Des situations plus réalistes sont présentées dans les deux exemples suivants, dans lesquels on va généraliser le test du rapport de vraisemblance.

21.2.1 Test entre une hypothèse simple et une alternative multiple

Considérons le problème de tester si oui ou non une pièce est juste.

Suivant le principe énoncé au début, on devrait choisir $H_0 = \{\theta = 0.5\}$. Mais on aurait alors

$$\sup_{\theta \neq 0.5} \mathbb{P}_\theta[\text{rejeter } H_0] = \mathbb{P}_{0.5}[\text{rejeter } H_0] = 1$$

Donc on ne pourra pas contrôler l'erreur de première espèce! Du coup, nous allons tester l'hypothèse $H_0 = \{\theta = 0.5\}$ contre l'alternative $H_1 = \{\theta \neq 0.5\}$. Dans ce cas l'alternative est représentée par l'ensemble Θ_1 de toutes les valeurs de θ autres que 0.5, soit $\Theta_1 = [0, 1] \setminus \{0.5\}$. Nous allons d'abord montrer comment contrôler l'erreur de première espèce, et ensuite s'intéresser à l'erreur de seconde espèce.

On va alors remplacer le numérateur dans le rapport de vraisemblance par $\sup_{\theta \in \Theta_1} V(\theta)$ pour obtenir

$$T = \frac{\sup_{\theta \in \Theta_1} V(\theta)}{V(0.5)}.$$

La statistique du test (généralisé) du rapport de vraisemblance est donc

$$T = \frac{\overline{X}_n^{n\overline{X}_n} (1 - \overline{X}_n)^{n(1-\overline{X}_n)}}{0.5^n},$$

et la région de rejet $A_t := \{T > t\}$ est l'événement

$$\overline{X}_n \log \overline{X}_n + (1 - \overline{X}_n) \log(1 - \overline{X}_n) - \log 0.5 > \frac{\log t}{n}.$$

Notons que la fonction $\varphi(x) = x \log x + (1 - x) \log(1 - x) - \log 0.5$ vérifie $\varphi'(x) = \log\left(\frac{x}{1-x}\right)$, $\varphi''(x) = \frac{1}{x(1-x)}$, et, sur $[0, 1]$, la fonction φ atteint son minimum 0 au point $x = 0.5$, et φ est symétrique autour de $x = 0.5$ (c'est-à-dire $\varphi(x+1/2) = \varphi(1/2-x)$). Donc l'événement A_t est de la forme $\{\overline{X}_n \notin [0.5 - \delta; 0.5 + \delta]\}$, et le problème est de trouver $\delta > 0$ tel que $\mathbb{P}_{0.5}[|\overline{X}_n - 0.5| > \delta] \leq \alpha$.

Mais sous la probabilité $\mathbb{P}_{0.5}$, la v.a. $2\sqrt{n}(\overline{X}_n - 0.5)$ suit approximativement, dès que n est un peu grand, la loi $\mathcal{N}(0, 1)$.

On est plus concerné par l'erreur de seconde espèce que par celle de première espèce. Il nous faut donc étudier la fonction

$$\theta \mapsto \mathbb{P}_\theta[\text{accepter } H_0] = \mathbb{P}_\theta[\overline{X}_n \in [0.5 - \delta, 0.5 + \delta]],$$

que l'on voudrait aussi petite que possible pour $\theta \neq 0.5$, ou bien étudier la *puissance du test* qui est ici la fonction

$$\theta \mapsto \mathbb{P}_\theta[\text{rejeter } H_0] = \mathbb{P}_\theta[\overline{X}_n \notin [0.5 - \delta, 0.5 + \delta]],$$

que l'on voudrait aussi proche que possible de 1 pour $\theta \neq 0.5$.

Remarque 21.2 *La puissance d'un test vaut par définition : 1- la probabilité d'erreur de seconde espèce. Puisque l'alternative est une hypothèse multiple, la puissance (comme l'erreur de seconde espèce) est ici une fonction.*

Comme ces probabilités dépendent continûment de θ , si θ est proche de 1/2 alors on aura forcément une puissance de test proche de zéro. Il est donc raisonnable de voir si, en imposant $|\theta - 1/2| \geq \beta$, où $\beta > 0$, alors $\mathbb{P}_\theta[\overline{X}_n \in [0.5 - \delta, 0.5 + \delta]] \leq \alpha'$.

Ceci revient exactement à contrôler les quantités

$$\mathbb{P}_{0.5-\beta}[\overline{X}_n \in [0.5 - \delta, 0.5 + \delta]] \text{ et } \mathbb{P}_{0.5+\beta}[\overline{X}_n \in [0.5 - \delta, 0.5 + \delta]]. (*)$$

Or si la loi mère est $\mathcal{B}(0.5 - \beta)$, alors $(X_j + \beta)$ suit la loi $\mathcal{B}(0.5)$, et $\overline{X}_n \in [0.5 - \delta, 0.5 + \delta]$ implique $\overline{X}_n - \beta \leq -(\beta - \delta)$; et si la loi mère est $\mathcal{B}(0.5 + \beta)$, alors $(X_j - \beta)$ suit la loi $\mathcal{B}(0.5)$ et $\overline{X}_n \in [0.5 - \delta, 0.5 + \delta]$ implique $\overline{X}_n + \beta \geq (\beta - \delta)$.

Du coup, les événements $(\overline{X}_n \in [0.5 - \delta, 0.5 + \delta])$ sous $\mathcal{B}(n, 0.5 - \beta)$ et $(\overline{X}_n \in [0.5 - \delta, 0.5 + \delta])$ sous $\mathcal{B}(n, 0.5 + \beta)$ sont contrôlés par $\overline{X}_n \notin [0.5 - (\beta - \delta), 0.5 + (\beta - \delta)]$ avec $\delta < \beta$.

En conclusion, les deux quantités (*) sont plus petites ou égales à α' si et seulement si $\delta < \beta$ et

$$\mathbb{P}_{0.5}[|\overline{X}_n - 0.5| \geq \beta - \delta] \leq \alpha'.$$

Si l'on veut que lorsque la vraie valeur du paramètre θ ne soit pas dans l'intervalle $[0.5 - \beta, 0.5 + \beta]$, le test ait une probabilité au plus $\alpha' = 0.05$ d'accepter à tort l'hypothèse que la pièce est juste, il faut que $2\sqrt{n}(\beta - \delta) \geq 1.96$. Ceci impose que l'on accepte une probabilité d'erreur de première espèce $\alpha > \alpha'$, et que n soit suffisamment grand.

Par exemple, si l'on choisit $\alpha' = 0.05$, $\beta = 0.02$ et $n = 10^4$, alors la valeur α' impose de choisir $\beta - \delta = \frac{1.96}{2\sqrt{n}} = 0.0098$, soit $\delta = 0.0102$. Remarquons qu'alors $\alpha \simeq 0.05$. Ceci signifie qu'au vu d'un échantillon de taille $n = 10^4$, il existe un test capable de détecter une pièce fautive de plus de 0.02 *i.e.* avec une probabilité de Pile en dehors de l'intervalle $[0.48, 0.52]$, avec à la fois une probabilité d'accepter à tort l'hypothèse d'une pièce juste de 5%, et une probabilité de déclarer la pièce fautive alors qu'elle est juste également de 5%.

21.2.2 Test entre deux hypothèses multiples

Dans le problème de la qualité de fabrication, on veut tester si la proportion inconnue d'appareils défectueux est bien inférieure à 10%. Pour $1 \leq k \leq n$, on pose

$$X_k = \begin{cases} 1, & \text{si le } k\text{-ième appareil testé est défectueux,} \\ 0, & \text{si le } k\text{-ième appareil testé est en bon état.} \end{cases}$$

Si la taille n de l'échantillon testé est petite devant la taille de la livraison, on peut admettre que les v. a. X_1, \dots, X_n sont indépendantes, de loi de Bernoulli de paramètre θ . On pose $H_0 = \{\theta \geq 10\%\}$, $H_1 = \{\theta < 10\%\}$, et

$$\begin{aligned} \hat{\theta}_0 &= \arg \max_{\{\theta \geq 0.1\}} V(\theta) \\ \hat{\theta}_1 &= \arg \max_{\{\theta < 0.1\}} V(\theta). \end{aligned}$$

Le test du rapport de vraisemblance généralisé est à nouveau basé sur une région de rejet de la forme $\{T > t\}$, avec ici

$$T = \frac{V(\hat{\theta}_1)}{V(\hat{\theta}_0)}.$$

La log-vraisemblance vaut $n\overline{X}_n \log \theta + n(1 - \overline{X}_n) \log(1 - \theta)$. Cette fonction de θ est concave (sa dérivée seconde est négative), et atteint son unique maximum au point $\theta = \overline{X}_n$. Il est alors facile de vérifier que

$$\hat{\theta}_0 = \max\{\overline{X}_n, 0.1\}, \quad \hat{\theta}_1 = \min\{\overline{X}_n, 0.1\}.$$

Deux cas sont possibles. Si $\overline{X}_n > 0.1$, alors $\hat{\theta}_0 = \overline{X}_n$, $\hat{\theta}_1 = 0.1$,

$$T = \frac{0.1^{n\overline{X}_n} 0.9^{n(1-\overline{X}_n)}}{\overline{X}_n^{n\overline{X}_n} (1 - \overline{X}_n)^{n(1-\overline{X}_n)}}.$$

On a $\log(T) = n\psi(\overline{X}_n)$, avec $\psi(x) = x \log \frac{0.1}{x} + (1-x) \log \frac{0.9}{1-x}$; or, $\psi'(x) < 0$ pour $x > 0.1$, et $\psi(0.1) = 0$, d'où $n\psi(x) < 0$ pour $x > 0.1$. Donc $T < 1$ pour tout $\overline{X}_n > 0.1$, et on accepte l'hypothèse $H_0 = \{\theta > 0.1\}$, ce qui intuitivement est très naturel.

Si par contre $\overline{X}_n < 0.1$, alors $\hat{\theta}_0 = 0.1$, $\hat{\theta}_1 = \overline{X}_n$,

$$T = \frac{\overline{X}_n^{-n\overline{X}_n} (1 - \overline{X}_n)^{n(1-\overline{X}_n)}}{0.1^{n\overline{X}_n} 0.9^{n(1-\overline{X}_n)}}.$$

Donc $\log(T) = n\varphi(\overline{X}_n)$, avec $\varphi(x) = x \log \frac{x}{0.1} + (1-x) \log \frac{1-x}{0.9}$; or, $\varphi'(x) < 0$ pour $x < 0.1$, et $\varphi(0.1) = 0$, d'où $\{T > t\} = \{\overline{X}_n < 0.1 - \delta\}$, et il reste à calculer δ tel que $\mathbb{P}_{0.1}[\overline{X}_n - 0.1 < -\delta] = \alpha$.

Or, sous $\mathbb{P}_{0.1}$, puisque $\sqrt{0.1(1-0.1)} = 0.3$, $\frac{\sqrt{n}}{0.3}(\overline{X}_n - 0.1)$ suit approximativement la loi $\mathcal{N}(0, 1)$. Finalement on rejette H_0 au niveau 0.05 si $\overline{X}_n < 0.1 - \frac{0.495}{\sqrt{n}}$ et au niveau 0.99 si $\overline{X}_n < 0.1 - \frac{0.699}{\sqrt{n}}$ (si X suit la loi $\mathcal{N}(0, 1)$, $\mathbb{P}[X < -1.65] = \mathbb{P}[X > 1.65] = 0.05$ et $0.3 \times 1.65 = 0.495$, $\mathbb{P}[X < -2.33] = \mathbb{P}[X > 2.33] = 0.01$ et $0.3 \times 2.33 = 0.699$).

On voit que pour prendre la décision d'accepter l'hypothèse que la fabrication est conforme, après avoir déballé 100 appareils (notons que vu la valeur $\theta = 0.1$, $n = 100$ est une taille minimale d'échantillons pour utiliser l'approximation gaussienne), au niveau 0.95, il faut que pas plus de 5 appareils sur les 100 soient défectueux! Une autre façon de présenter la même règle de décision est de calculer la

$$\text{p-valeur} = \mathbb{P} \left[Z < \frac{\sqrt{n}}{0.3} (\overline{x}_n - 0.1) \right],$$

où Z suit la loi $\mathcal{N}(0, 1)$, et x_n est la proportion d'appareils défectueux parmi les appareils testés. On rejette H_0 au niveau α si et seulement si p-valeur $< \alpha$.

21.3 Etape de mise en place d'un test

En résumé, la mise en place d'un test se fait en trois étapes.

1. On fixe le niveau du test : $(1 - \alpha) \in \{95\%, 99\%\}$.
2. On choisit H_0 : il faut que ce soit l'hypothèse que l'on souhaite rejeter car on aura un bon contrôle de l'erreur de première espèce.
3. On détermine le seuil du test : pour cela, on calcule le rapport de vraisemblance et on cherche à estimer la zone de rejet $\{T > t\}$ en fonction d'un estimateur que l'on sait calculer. Dans nos exemples, il s'agissait de la moyenne empirique.

La méconnaissance de l'erreur de deuxième espèce a la conséquence suivante :

la conclusion d'un test n'a valeur de preuve que lorsque cette conclusion est le rejet de H_0 .

En effet, si à l'issue du test, on conclut au rejet de H_0 , on sait qu'on a une probabilité α de se tromper. Si au contraire, on conclut à l'acceptation de H_0 , on a peut-être grande chance de se tromper. Ainsi, il faut plutôt voir l'acceptation de H_0 comme un non-rejet de H_0 . La démarche est ici très empirique : sur l'expérience que j'ai faite, rien ne permet de dire que H_0 n'est pas vraie.

Chapitre VII. Quelques lois usuelles

On présente ici quelques lois usuelles.

22 Lois de v.a. entières

Nom et symbole	Support	Probabilités élémentaires $\mathbb{P}[X = k]$	Espérance	Variance	Fonction génératrice $G_X(z)$
Loi de Bernoulli $\mathcal{B}(p)$, $p \in [0, 1]$	$\{0, 1\}$	$\mathbb{P}[X = 0] = 1 - p$ $\mathbb{P}[X = 1] = p$	p	$p(1 - p)$	$1 - p + pz$
Loi binomiale $\mathcal{B}(n, p)$, $p \in [0, 1]$, $n \in \mathbb{N} \setminus \{0\}$	$\{0, 1, \dots, n\}$	$C_n^k p^k (1 - p)^{n-k}$	np	$np(1 - p)$	$(1 - p + pz)^n$
Loi binomiale négative $\mathcal{BN}(n, p)$, $p \in [0, 1]$, $n \in \mathbb{N} \setminus \{0\}$	$\{n, n + 1, \dots\}$	$C_{n-1}^{k-1} p^n (1 - p)^{k-n}$	$\frac{n}{p}$	$\frac{np(1-p)}{p^2}$	$\left(\frac{pz}{1-(1-p)z}\right)^n$
Loi de Poisson $\mathcal{P}(\lambda)$, $\lambda > 0$	\mathbb{N}	$\frac{\lambda^k}{k!} e^{-\lambda}$	λ	λ	$e^{\lambda(z-1)}$
Loi géométrique $\mathcal{G}(p)$, $p \in]0, 1]$	$\mathbb{N} \setminus \{0\}$	$p(1 - p)^{k-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pz}{1-(1-p)z}$
Loi hypergéométrique $\mathcal{H}(N, m, n)$, $N \in \mathbb{N} \setminus \{0\}$, $m, n \in \{1, \dots, N\}$	$\{0, \dots, \min\{m, n\}\}$	$\frac{C_m^k C_{N-m}^{n-k}}{C_N^n}$	$\frac{m}{n}$	$\frac{mn(N-n)(N-m)}{N^2(N-1)}$	

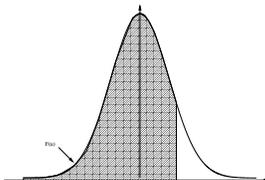
23 Lois de v.a. absolument continues

Nom et symbole	Support	Densité $f(x)$	Espérance	Variance
Loi uniforme $\mathcal{U}[a, b]$, $a < b$	$[a, b]$	$f(x) = \frac{1}{ b-a } \mathbb{I}_{[a,b]}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Loi normale $\mathcal{N}(m, \sigma^2)$, $m, \sigma \in \mathbb{R}$, $\sigma \neq 0$	\mathbb{R}	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$	m	σ^2
Loi exponentielle $\mathcal{E}(\lambda)$, $\lambda > 0$	$[0, \infty[$	$f(x) = \lambda e^{-\lambda x} \mathbb{I}_{\mathbb{R}_+}(x)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

24 Tables de la loi normale

1 Loi normale : fonction de répartition

Pour une valeur $u \geq 0$, la table ci-dessous renvoie la valeur $F(u)$ de la fonction de répartition F de la loi normale centrée réduite au point u .



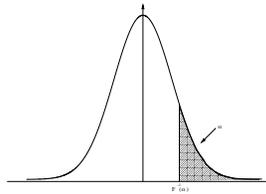
u	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Table pour les grandes valeurs de u :

u	3.0	3.1	3.2	3.3	3.4
$F(u)$	0.99865	0.999032	0.999313	0.999517	0.999663
u	3.5	3.6	3.7	3.8	3.9
$F(u)$	0.999767	0.999841	0.999892	0.999928	0.999952
u	4.0	4.1	4.2	4.3	4.4
$F(u)$	0.999968	0.999979	0.999987	0.999991	0.999995
u	4.5	4.6	4.7	4.8	4.9
$F(u)$	0.999997	0.999998	0.999999	0.999999	1

2 Loi normale : quantiles

Pour une valeur $\alpha \in]0; 0.5[$, la table ci-dessous renvoie la valeur $F^{-1}(\alpha)$ de la fonction quantile F^{-1} de la loi normale centrée réduite au point α .



α	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
0.00	∞	3.0902	2.8782	2.7478	2.6521	2.5758	2.5121	2.4573	2.4089	2.3656
0.01	2.3263	2.2904	2.2571	2.2262	2.1973	2.1701	2.1444	2.1201	2.0969	2.0749
0.02	2.0537	2.0335	2.0141	1.9954	1.9774	1.9600	1.9431	1.9268	1.9110	1.8957
0.03	1.8808	1.8663	1.8522	1.8384	1.8250	1.8119	1.7991	1.7866	1.7744	1.7624
0.04	1.7507	1.7392	1.7279	1.7169	1.7060	1.6954	1.6849	1.6747	1.6646	1.6546
0.05	1.6449	1.6352	1.6258	1.6164	1.6072	1.5982	1.5893	1.5805	1.5718	1.5632
0.06	1.5548	1.5464	1.5382	1.5301	1.5220	1.5141	1.5063	1.4985	1.4909	1.4833
0.07	1.4758	1.4684	1.4611	1.4538	1.4466	1.4395	1.4325	1.4255	1.4187	1.4118
0.08	1.4051	1.3984	1.3917	1.3852	1.3787	1.3722	1.3658	1.3595	1.3532	1.3469
0.09	1.3408	1.3346	1.3285	1.3225	1.3165	1.3106	1.3047	1.2988	1.2930	1.2873
0.10	1.2816	1.2759	1.2702	1.2646	1.2591	1.2536	1.2481	1.2426	1.2372	1.2319
0.11	1.2265	1.2212	1.2160	1.2107	1.2055	1.2004	1.1952	1.1901	1.1850	1.1800
0.12	1.1750	1.1700	1.1650	1.1601	1.1552	1.1503	1.1455	1.1407	1.1359	1.1311
0.13	1.1264	1.1217	1.1170	1.1123	1.1077	1.1031	1.0985	1.0939	1.0893	1.0848
0.14	1.0803	1.0758	1.0714	1.0669	1.0625	1.0581	1.0537	1.0494	1.0450	1.0407
0.15	1.0364	1.0322	1.0279	1.0237	1.0194	1.0152	1.0110	1.0069	1.0027	0.9986
0.16	0.9945	0.9904	0.9863	0.9822	0.9782	0.9741	0.9701	0.9661	0.9621	0.9581
0.17	0.9542	0.9502	0.9463	0.9424	0.9385	0.9346	0.9307	0.9269	0.9230	0.9192
0.18	0.9154	0.9116	0.9078	0.9040	0.9002	0.8965	0.8927	0.8890	0.8853	0.8816
0.19	0.8779	0.8742	0.8705	0.8669	0.8633	0.8596	0.8560	0.8524	0.8488	0.8452
0.20	0.8416	0.8381	0.8345	0.8310	0.8274	0.8239	0.8204	0.8169	0.8134	0.8099
0.21	0.8064	0.8030	0.7995	0.7961	0.7926	0.7892	0.7858	0.7824	0.7790	0.7756
0.22	0.7722	0.7688	0.7655	0.7621	0.7588	0.7554	0.7521	0.7488	0.7454	0.7421
0.23	0.7388	0.7356	0.7323	0.7290	0.7257	0.7225	0.7192	0.7160	0.7128	0.7095
0.24	0.7063	0.7031	0.6999	0.6967	0.6935	0.6903	0.6871	0.6840	0.6808	0.6776
0.25	0.6745	0.6713	0.6682	0.6651	0.6620	0.6588	0.6557	0.6526	0.6495	0.6464
0.26	0.6433	0.6403	0.6372	0.6341	0.6311	0.6280	0.6250	0.6219	0.6189	0.6158
0.27	0.6128	0.6098	0.6068	0.6038	0.6008	0.5978	0.5948	0.5918	0.5888	0.5858
0.28	0.5828	0.5799	0.5769	0.5740	0.5710	0.5681	0.5651	0.5622	0.5592	0.5563
0.29	0.5534	0.5505	0.5476	0.5446	0.5417	0.5388	0.5359	0.5330	0.5302	0.5273
0.30	0.5244	0.5215	0.5187	0.5158	0.5129	0.5101	0.5072	0.5044	0.5015	0.4987
0.31	0.4959	0.4930	0.4902	0.4874	0.4845	0.4817	0.4789	0.4761	0.4733	0.4705
0.32	0.4677	0.4649	0.4621	0.4593	0.4565	0.4538	0.4510	0.4482	0.4454	0.4427
0.33	0.4399	0.4372	0.4344	0.4316	0.4289	0.4261	0.4234	0.4207	0.4179	0.4152
0.34	0.4125	0.4097	0.4070	0.4043	0.4016	0.3989	0.3961	0.3934	0.3907	0.3880
0.35	0.3853	0.3826	0.3799	0.3772	0.3745	0.3719	0.3692	0.3665	0.3638	0.3611
0.36	0.3585	0.3558	0.3531	0.3505	0.3478	0.3451	0.3425	0.3398	0.3372	0.3345
0.37	0.3319	0.3292	0.3266	0.3239	0.3213	0.3186	0.3160	0.3134	0.3107	0.3081
0.38	0.3055	0.3029	0.3002	0.2976	0.2950	0.2924	0.2898	0.2871	0.2845	0.2819
0.39	0.2793	0.2767	0.2741	0.2715	0.2689	0.2663	0.2637	0.2611	0.2585	0.2559
0.40	0.2533	0.2508	0.2482	0.2456	0.2430	0.2404	0.2378	0.2353	0.2327	0.2301
0.41	0.2275	0.2250	0.2224	0.2198	0.2173	0.2147	0.2121	0.2096	0.2070	0.2045
0.42	0.2019	0.1993	0.1968	0.1942	0.1917	0.1891	0.1866	0.1840	0.1815	0.1789
0.43	0.1764	0.1738	0.1713	0.1687	0.1662	0.1637	0.1611	0.1586	0.1560	0.1535
0.44	0.1510	0.1484	0.1459	0.1434	0.1408	0.1383	0.1358	0.1332	0.1307	0.1282
0.45	0.1257	0.1231	0.1206	0.1181	0.1156	0.1130	0.1105	0.1080	0.1055	0.1030
0.46	0.1004	0.0979	0.0954	0.0929	0.0904	0.0878	0.0853	0.0828	0.0803	0.0778
0.47	0.0753	0.0728	0.0702	0.0677	0.0652	0.0627	0.0602	0.0577	0.0552	0.0527
0.48	0.0502	0.0476	0.0451	0.0426	0.0401	0.0376	0.0351	0.0326	0.0301	0.0276
0.49	0.0251	0.0226	0.0201	0.0175	0.0150	0.0125	0.0100	0.0075	0.0050	0.0025

Table des matières

I	Objets centraux	3
1	Espace de probabilité	3
2	Variables aléatoires	10
3	Espérance, variance et écart-type	13
3.1	Espérance	13
3.2	Variance	17
II	Conditionnement et indépendance	21
4	Conditionnement	21
5	Indépendance	23
III	Fonctions génératrices	29
6	Généralités	29
7	Fonctions génératrices et indépendance	31
8	Somme d'un nombre aléatoire de variables aléatoires	33
9	Evolution des Populations	36
9.1	Motivation	36
9.2	Probabilité d'extinction.	36
9.3	Exemple	38
IV	Lois (absolument) continues	39
10	Espace de probabilité revisité	39
11	Variables aléatoires	42
12	Indépendance	44

	82
13 L'aiguille de Buffon	45
V Théorèmes limites	49
14 Loi des grands nombres	49
15 Convergence en loi	51
16 Théorème central limite	53
16.1 Loi de Bernoulli équilibrée	53
16.2 La version générale	55
16.3 Somme de gaussiennes indépendantes	57
17 Approximation de la loi binomiale	57
17.1 Paradigme de Poisson	57
17.2 L'état de Californie contre Collins	59
VI Statistique	61
18 La démarche statistique	61
19 Estimation	62
19.1 Méthode des moments	63
19.2 Méthode du maximum de vraisemblance	65
19.2.1 Echantillon à valeurs entières	65
19.2.2 Echantillon de loi à densité	66
20 Intervalle de confiance	67
21 Tests d'hypothèse	69
21.1 Le principe des tests : le cas de deux hypothèses simples	69
21.1.1 Deux types d'erreur	69
21.1.2 Le test entre deux hypothèses simples dans le jeu de Pile ou Face	71
21.1.3 Test du rapport de vraisemblance	71
21.2 Test entre hypothèses multiples	72

	83
21.2.1 Test entre une hypothèse simple et une alternative multiple	72
21.2.2 Test entre deux hypothèses multiples	74
21.3 Etape de mise en place d'un test	75
VII Quelques lois usuelles	77
22 Lois de v.a. entières	77
23 Lois de v.a. absolument continues	77
24 Tables de la loi normale	78