Probabilités de base (PRB-PSI1)

Bachir Bekka, Cours L3 CAPES et ING, Rennes 2021/2022

21 septembre 2021

Table des matières

Chapitre 1

Introduction

La théorie des probabilités est l'analyse du hasard, plus précisément, l'analyse mathématique de phénomènes où le hasard intervient. Il s'agit de l'étude d'évènements dépendant d'un grand nombre de facteurs qui échappe à notre contrôle et dont nous ne savons pas s'ils se réaliseront ou non. Voici quelques exemples :

- le résultat d'un jeu de hasard : pile ou face dans le lancer d'une pièce de monnaie, nombre obtenu dans le lancer d'un dé, résultat d'une loterie, ...
- la durée de vie d'un atome radioactif, d'un individu ou d'une ampoule électrique, ...
- le lancer d'une flèche sur un cible,
- la quantité de précipitations annuelles dans une région donnée,
- la position d'un grain de pollen en suspension dans l'eau,
- l'évolution du cours de la bourse,
- le résultat d'un match de football ; etc

La théorie des probabilités, issue à l'origine de la modélisation des jeux de hasard, occupe aujourd'hui une place centrale dans la plupart des sciences : analyse des données par la statistique (avec d'innombrables applications à pratiquement à toutes les disciplines scientifiques, de la physique aux sciences humaines, ainsi qu'à de nombreux domaines du monde économique, comme les assurances ou la finance), modélisation de nombreux phénomènes en sciences exactes (physique, chimie, biologie, informatique, ...), humaines (économie, sociologie, par exemple) ou médicales. Aujourd'hui, ses méthodes sont utilisées également dans de nombreux domaines des mathématiques, comme les EDP, l'analyse fonctionnelle ou même la géométrie et la théorie des nombres.

Chapitre 2

Espaces probabilisés

La première notion de la théorie des probabilités est celle d'*expérience aléatoire*. Il s'agit d'une épreuve avec un résultat incertain ou inconnu de l'expérimentateur avant que l'expérience n'ait lieu.

2.1 Modélisation d'une expérience aléatoire

La théorie des probabilités fournit une modélisation de l'expérience aléatoire.

2.1.1 Univers d'une expérience aléatoire

Une expérience aléatoire nécessite la donnée d'un ensemble Ω appelé *univers* de l'epreuve, décrivant l'ensemble des résultats possibles de l'expérience, chacun des éléments $\omega \in \Omega$ étant une *issue* ou eventualité de l'épreuve.

L'ensemble Ω peut être fini, dénombrable infini ou même infini non dénombrable. Dans les exemples plus haut, on peut choisir

— $\Omega = \{P, F\}$ dans un tirage à pile ou face ; $\Omega = \{PP, PF, FF, FP\}$ dans deux tirages à pile ou face ; $\Omega = \{P, F\}^{\mathbf{N}}$ pour une infinité de irages à pile ou face ;

— $\Omega = \{1, \cdots, 6\}$ dans le lancer d'un dé ;

— $\Omega = [0, +\infty[$ pour la durée de vie d'un atome radioactif, d'un individu ou d'une ampoule électrique ;

— $\Omega = \mathbf{R}^2$ pour le lancer d'une flèche sur une cible ;

— $\Omega = C([0, +\infty[, \mathbf{R}^3)$ pour le mouvement d'un grain de pollen dans l'eau.

2.1.2 Évènements, tribus

Une fois l'univers Ω choisi, on s'intéresse à la réalisation ou non d'évènements particuliers. Un *événement* est une propriété dont on peut dire si elle est vérifiée ou non une fois le résultat de l'expérience connu. Un événement est décrit par un sous-ensemble de Ω, qui représente les réalisations de l'événement. Par exemple, le lancer de deux dés est modélisé par

$$\Omega = \{1, \cdots, 6\} \times \{1, \cdots, 6\},$$

et l'évènement "la somme des deux lancers est égale à 4" correspond au sous-ensemble $\{(1,3),(3,1),(2,2)\}$.

L'ensemble des événements associés à une expérience aléatoire est donc décrit par un sous-ensemble $\mathscr{F}$ des parties de Ω. Les éléments $\omega \in \Omega$ sont les "évènements élémentaires". L' *évènement impossible* s'identifie à l'ensemble vide $\emptyset$ et l' *évènement certain* à Ω. Les opérations logiques telles que "ou" et "et" se traduisent par les opérations ensemblistes "réunion" et "intersection" Voici un tableau des correspondances entre ces deux langages

Symbole	Language ensembliste	Language probabiliste
Ω	ensemble plein	évènement certain
$\emptyset$	ensemble vide	évènement impossible
A	sous-ensemble de Ω	évènement
ω	élément de Ω	évènement élémentaire
$A \cup B$	réunion de A et B	A ou B
$A \cap B$	intersection de A et B	A et B
A^c	complémentaire de A	évènement contraire de A
$A \subset B$	A inclus dans B	A implique B
$A \cap B = \emptyset$	A et B sont disjoints	A et B sont incompatibles

Soit Ω l'univers d'une expérience aléatoire et soit $\mathscr{F} \subset \mathscr{P}(\Omega)$ l'ensemble des évènements. Il pourrait semblait raisonnable de prendre $\mathscr{F} = \mathscr{P}(\Omega)$. C'est ce choix qui sera fait en général quand Ω est dénombrable. Cependant, dans le cas où Ω n'est pas dénombrable, il sera en général impossible d'associer à chaque partie de Ω une probabilité de façon cohérente. Pour cette raison, on choisira pour $\mathscr{F}$ un sous-ensemble strict de $\mathscr{P}(\Omega)$, pour lequel on imposera des conditions naturelles de stabilité.

Définition 2.1.1 (Tribu) Un sous-ensemble $\mathscr{F}$ de $\mathscr{P}(\Omega)$ est appellé *tribu* (ou σ-algèbre) si les propriétés suivantes sont satisfaites :

(i) $\Omega \in \mathscr{F}$;
(ii) pour tout $A \in \mathscr{F}$, on a $A^c \in \mathscr{F}$ (stabilité par complémentaire) ;
(iii) pour toute *suite* $(A_n)_n$ d'éléments de $\mathscr{F}$, on a $\bigcup_n A_n \in \mathscr{F}$ (stabilité par réunion dénombrable).

Remarquons tout de suite qu'une tribu est stable par intersection dénombrable. Tout d'abord, rappelons les formules de Morgan et les formules de distributivité.

Remarque 2.1.2 (Rappel) Pour $(A_i)_{i\in I}$ une famille quelconque de parties de l'ensemble Ω et $B \subset \Omega$, on a

$$\left(\bigcup_{i\in I} A_i\right)^c = \bigcap_{i\in I} A_i^c \qquad \text{et} \qquad \left(\bigcap_{i\in I} A_i\right)^c = \bigcup_{i\in I} A_i^c$$

$$B \cap \left(\bigcup_{i\in I} A_i\right) = \bigcup_{i\in I} B \cap A_i \qquad \text{et} \qquad B \cup \left(\bigcap_{i\in I} A_i\right) = \bigcap_{i\in I} B \cup A_i.$$

Proposition 2.1.3 *Soit $\mathscr{F}$ une tribu de parties de Ω. Alors les propriétés suivantes sont satisfaites :*
(i) $\emptyset \in \mathscr{F}$;
(iii) pour toute suite $(A_n)_n$ *d'éléments de $\mathscr{F}$, on a $\bigcap_n A_n \in \mathscr{F}$ (stabilité par intersection dénombrable).*

Démonstration (i) Par les propriétés de tribu, on a $\Omega \in F$ et donc $\emptyset = \Omega^c \in \mathscr{F}$.
(ii) Par les propriétés de tribu, on a $A_n^c \in \mathscr{F}$ et donc $\bigcup_n A_n^c \in \mathscr{F}$. Il s'ensuit, par les formules de Morgan, que

$$\bigcap_n A_n = (\bigcup_n A_n^c)^c \in \mathscr{F}. \blacksquare$$

La proposition suivante se vérifie immédiatement à partir de la définition d'une tribu.

Proposition 2.1.4 *Soit $(\mathscr{F}_i)_{i\in I}$ une famille de tribus de Ω. Alors $\mathscr{F} := \cap_{i\in I}\mathscr{F}_i$ est une tribu de Ω.* $\blacksquare$

Corollaire 2.1.5 *Soit $\mathscr{C} \subset \mathscr{P}(\Omega)$. L'intersection des tribus de $\mathscr{P}(\Omega)$ contenant $\mathscr{C}$ est une tribu s'appelle la* tribu engendrée par $\mathscr{C}$ *et est notée $\sigma(\mathscr{C})$. On a donc $\sigma(\mathscr{C}) = \bigcap_{\mathscr{F}} \mathscr{F}$, où $\mathscr{F}$ parcourt l'ensemble de tribus contenant $\mathscr{C}$. C'est la plus petite tribu contenant $\mathscr{C}$.*

Exemple 2.1.6 (i) $\mathscr{P}(\Omega)$ est une tribu, appelée *tribu triviale*;
(ii) $\{\Omega, \emptyset\}$ est une tribu, appelée *tribu grossière*;
(iii) Soit $A \subset \Omega$; alors $\{\Omega, A, A^c, \emptyset\}$ est une tribu; il est clair que c'est la tribu $\sigma(A)$ engendrée par A;
(iv) Soit $\Omega = \mathbf{R}$ et $\mathscr{C}$ l'ensemble des intervalles $]-\infty, a]$ pour $a \in \mathbf{R}$. La tribu $\sigma(\mathscr{C})$ s'appelle la tribu des *boréliens* de $\mathbf{R}$ et contient tous les intervalles de $\mathbf{R}$.

2.1.3 Mesures de probabilité

Soit Ω un univers muni d'une tribu d'évènements $\mathscr{F}$. Un mesure de probabilité sur $\mathscr{F}$ permet d'attribuer à chaque évènement un nombre réel dans $[0,1]$ qui mesure le degré de confiance que l'on a quant à la réalisation de cet évènement.

Définition 2.1.7 **(Mesure de probabilité)** Une *mesure de probabilité* sur $(\Omega, \mathscr{F})$ est une application $\mathbf{P} : \mathscr{F} \to [0,1]$ possèdant les deux propriétés suivantes :

(i) $\boxed{\mathbf{P}(\Omega) = 1;}$

(ii) **(σ-additivité)** pour toute suite (finie ou dénombrable infinie) d'éléments A_n d'éléments deux-à-deux dispoints de $\mathscr{F}$, deux-à-deux dispoints de $\mathscr{F}$, on a $\boxed{\mathbf{P}(\bigcup_n A_n) = \sum_n \mathbf{P}(A_n).}$

On dit que le triplet $(\Omega, \mathscr{F}, \mathbf{P})$ est un *espace probabilisé* ou un *espace de probabilité.*

Voici quelques propriétés immédiates d'une mesure de probabilité.

Proposition 2.1.8 *Soit* $(\Omega, \mathscr{F}, \mathbf{P})$ *une espace de probabilité. Alors les propriétés suivantes sont satisfaites :*

(i) $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$ *pour tout* $A \in F$;

(ii) $\mathbf{P}(\emptyset) = 0$;

(iii) ***(monotonie)*** $\mathbf{P}(A) \leq \mathbf{P}(B)$ *pour tous* $A, B \in \mathscr{F}$ *avec* $A \subset B$;

(iv) $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$ *pour tous* $A, B \in \mathscr{F}$ *avec* $A \subset B$;

(v) ***(sous-σ-additivité)*** $\mathbf{P}(\cup_n A_n) \leq \sum_n \mathbf{P}(A_n)$ *pour toute suite (éventuellement finie)* $(A_n)_n$ *dans* $\mathscr{F}$;

(vi) $\mathbf{P}(A) = \sum_n \mathbf{P}(A \cap A_n)$ *pour tout* $A \in F$ *et toute partition* $(A_n)_n$ *de* Ω *par des évènements* $A_n \in \mathscr{F}$.

Démonstration (i) Les évènements A et A^c sont disjoints et donc, par les propriétés de $\mathbf{P}$,

$$1 = \mathbf{P}(\Omega) = \mathbf{P}(A \cup A^c) = \mathbf{P}(A) + \mathbf{P}(A^c).$$

(ii) est le cas particulier $A = \Omega$ de (i).
(iii) On a

$$\mathbf{P}(B) = \mathbf{P}(A \cup (B \setminus A)) = \mathbf{P}(A) + \mathbf{P}(B \setminus A) \geq \mathbf{P}(A).$$

(iv) On a des partitions $A \cup B = A \cup (B \setminus (A \cap B))$ et $B = (A \cap B) \cup (B \setminus (A \cap B))$ et donc

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B \setminus (A \cap B)) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B).$$

(v) On définit $B_0 = A_0$ et , de manière recursive, $B_{n+1} = A_n \setminus \bigcup_{k=0}^{n} A_k$ Alors $\cup_n A_n = \cup_n B_n$. Comme les B_n sont deux-à-deux disjoints et comme $B_n \subset A_n$, on a

$$\mathbf{P}(\cup_n A_n) = \mathbf{P}(\cup_n B_n) = \sum_n \mathbf{P}(B_n) \leq \sum_n \mathbf{P}(A_n).$$

(vi) découle du fait que $A = \cup_n (A \cap A_n)$ et de la σ-additivité de **P**.■

Souvent, nous aurons à calculer la probabilité d'une réunion croissante ou d'une intersection croissante d'évènements; le lemme suivant nous sera utile

Lemme 2.1.9 ***(Limites monotones)*** *(i) Soit* (A_n) *une suite croissante d'évènements, c-à-d* $A_n \subset A_{n+1}$ *pour tout n. Alors, pour* $A := \cup_n A_n$, *on a*

$$\boxed{\mathbf{P}(A) = \lim_n \mathbf{P}(A_n).}$$

(ii) Soit (B_n) *une suite décroissante d'évènements, c-à-d* $B_{n+1} \subset B_n$ *pour tout n. Alors, pour* $B := \cap_n B_n$, *on a*

$$\boxed{\mathbf{P}(B) = \lim_n \mathbf{P}(B_n).}$$

Démonstration (i) On définit $B_0 = A_0$ et , de manière recursive,

$$B_n = A_n \setminus \bigcup_{k=0}^{n-1} A_k \qquad \text{pour tout} \quad n \geq 1$$

Alors, pour tout $n \in \mathbf{N}$, on a $\cup_{i=0}^{n} A_i = \cup_{i=0}^{n} B_i$ et donc $A = \cup_{i \in \mathbf{N}} A_i$. Or $\cup_{i=0}^{n} A_i = A_n$, car la suite $(A_n)_n$ est croissante. Comme les B_n sont deux-à-deux disjoints on on a donc

$$\begin{aligned}
\mathbf{P}(A) = \mathbf{P}(\cup_n B_n) &= \sum_n \mathbf{P}(B_n) \\
&= \lim_n \sum_{i=0}^{n} \mathbf{P}(B_n) = \lim_n \mathbf{P}(\cup_{i=0}^{n} B_i) \\
&= \lim_n \mathbf{P}(A_n).
\end{aligned}$$

(ii) On pose $A_n := B_n^c$. Comme $(B_n)_n$ est décroissante, $(A_n)_n$ est croissante et on a donc, par (i), $\mathbf{P}(A) = \lim_n \mathbf{P}(A_n)$ pour $A = \cup_n A_n$. Or, par la formule de Morgan,

$$\cup_n A_n = \cup_n B_n^c = (\cap_n B_n)^c = B^c.$$

D'où

$$\begin{aligned}\mathbf{P}(B) &= 1 - \mathbf{P}(B^c) = 1 - \lim_n \mathbf{P}(A_n)\\ &= 1 - \lim_n \mathbf{P}(B_n^c) = 1 - (1 - \lim_n \mathbf{P}(B_n))\\ &= \lim_n \mathbf{P}(B_n). \blacksquare\end{aligned}$$

2.2 Espaces de probabilité finis ou dénombrables

Soit $\Omega = \{\omega_i \ : \ i \in I\}$ pour un ensemble I fini ou dénombrable infini, muni de la tribu $\mathscr{P}(\Omega)$. Une mesure de probabilité sur $\mathscr{P}(\Omega)$ est entièrement déterminée par sa valeur sur les évènements élémentaires ω_i. Commençons par une rappel sur les séries convergentes à termes positifs ; concernant sa preuve, voir le cours sur "Suites et séries numériques".

Lemme 2.2.1 *Soit $(a_n)_{n\in\mathbf{N}}$ une suite de nombres positifs tels que la série $\sum_{n\in\mathbf{N}} a_n$ est convergente et soit s sa somme.*
*(i)**(Commutativité)**Soit $\sigma : \mathbf{N} \to \mathbf{N}$ une bijection. Alors la série $\sum_{n\in\mathbf{N}} a_{\sigma(n)}$ est convergente de même somme s.*
*(ii)**(Convergence "en vrac")** Soit $\mathbf{N} = \cup_{i\in\mathbf{N}} N_i$ une partition de $\mathbf{N}$ en sous-ensembles $N_i = \{b_0^{(i)}, b_1^{(i)}, \ldots,\}$. Alors, pour chaque $i \in I$, la série $\sum_{n\geq 0} b_n^{(i)}$ est convergente et, on notant $s^{(i)}$ sa somme, la série $\sum_{n\in\mathbf{N}} s^{(i)}$ est convergente de même somme s. On peut donc écrire*

$$\sum_{n\in\mathbf{N}} a_n = \sum_{i\in\mathbf{N}} (\sum_{n\in N_i} a_n). \blacksquare$$

Remarque 2.2.2 Soit Ω un ensemble dénombrable et $f : \Omega \to [0, +\infty[$ une fonction à valeurs positives telles que la série $\sum_{n\in\mathbf{N}} f(\varphi(n))$ est convergente pour une bijection $\varphi : \mathbf{N} \to \Omega$ ("énumération" de Ω). Alors $\sum_{n\in\mathbf{N}} f(\psi(n))$ est convergente et de même somme, pour tout autre bijection $\psi : \mathbf{N} \to \Omega$. On peut donc écrire sans ambiguïté $\sum_{\omega\in\Omega} f(\omega)$ et on a

$$\sum_{\omega\in\Omega} f(\omega) = \sum_{i\in\mathbf{N}} (\sum_{\omega\in\Omega_i} f(\omega),$$

pour toute partition $\Omega = \cup_{i\in\mathbf{N}} \Omega_i$.

Proposition 2.2.3 *(i) Soit* $\mathbf{P}$ *une mesure de probabilité sur* $\mathscr{P}(\Omega)$. *On définit* $p : \Omega \to [0,1], \omega \mapsto p_\omega$ *par* $p_\omega = \mathbf{P}(\{\omega\})$. *Alors on a* $\sum_{\omega\in\Omega} p_\omega = 1$ *et*

$$\mathbf{P}(A) = \sum_{\omega\in A} p_\omega \qquad \textit{pour tout} \quad A \in \mathscr{P}(\Omega).$$

(ii) Réciproquement, soit $p : \Omega \to [0,1], \omega \mapsto p_\omega$ *une fonction telle que* $\sum_{\omega\in\Omega} p_\omega = 1$. *Alors la fonction* $\mathbf{P} : \mathscr{P}(\Omega) \to [0,1]$ *définie par*

$$\mathbf{P}(A) = \sum_{\omega\in A} p_\omega \qquad \textit{pour tout} \quad A \in \mathscr{P}(\Omega),$$

est une mesure de probabilité sur $\mathscr{P}(\Omega)$.

Démonstration (i) Comme $A = \cup_{\omega\in A}\{\omega\}$ est une partition de A, pour toute partie A de Ω, on a par σ-additivité

$$\mathbf{P}(A) = \mathbf{P}(\cup_{\omega\in A}\{\omega\}) = \sum_{\omega\in A} \mathbf{P}(\{\omega\}) = \sum_{\omega\in A} p_\omega$$

et en particulier $1 = \mathbf{P}(\Omega) = \sum_{\omega\in\Omega} p_\omega$.
(ii) On a tout d'abord $\mathbf{P}(\Omega) = \sum_{\omega\in\Omega} p_\omega = 1$. Ensuite, soit $A = \cup_n A_n$ une partition de $A \subset \Omega$. Alors, en utilisant le rappel précédent (Lemme 2.2.1), on a

$$\mathbf{P}(A) = \sum_{\omega\in A} p_\omega = \sum_n \Big(\sum_{\omega\in A_n} p_\omega\Big) = \sum_n \mathbf{P}(A_n).$$

Ceci montre la σ-additivité de $\mathbf{P}$.$\blacksquare$

2.2.1 Espaces de probabilité finis

Soit Ω un ensemble *fini*, de cardinal $\mathrm{Card}(\Omega) = n$. Un cas particulièrement important est celui où la même probabilité est associée à chaque résultat possible de l'expérience. (Ceci n'est pas possible quand $\Omega = \{\omega_1, \omega_2, \ldots\}$ est infini : on aurait alors $1 = \mathbf{P}(\Omega) = \sum_{n=1}^{\infty} p = +\infty$, si $\mathbf{P}(\omega_i) = p > 0$.)

Définition 2.2.4 On appelle *mesure de probabilité uniforme* sur Ω la mesure de probabilité définie par la fonction $p : \Omega \to [0,1], \omega \mapsto 1/n$. On dit dans ce cas qu'il y a *équiprobabilité.*

La probabilité d'un évènement $A \subset \Omega$ est alors donnée par la formule

$$\boxed{\mathbf{P}(A) = \frac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)}}$$

c-à-d, la probabilité de A est

$$\boxed{\mathbf{P}(\mathrm{A}) = \frac{\text{nombre de cas favorables}}{\text{nombre total de cas}}}$$

Exemple 2.2.5 On lance deux dés et on s'intéresse à la probabilité de l'évènement

$$\mathrm{A}: \text{“la somme des points obtenus est } 4\text{"}.$$

L'expérience est modélisée par l'univers $\Omega = \{1,\dots,6\}^2$, muni de la mesure uniforme

$$\mathbf{P}(\{\omega\} = \frac{1}{6} \qquad \text{pour tout} \quad \omega \in \Omega.$$

Alors A correspond à la partie $\{(1,3),(2,2),(3,1)\}$. On a donc

$$\mathbf{P}(\mathrm{A}) = \frac{\mathrm{Card}(\mathrm{A})}{\mathrm{Card}(\Omega)} = \frac{3}{36} = \frac{1}{12}.$$

2.2.2 Dénombrements

Les calculs de probabilités dans le cas d'équiprobabilité se ramènent souvent à des calcul de cardinaux d'ensembles, c-à-d à des problèmes de dénombrement.

Voici quelques formules simples concernant les cardinaux d'ensembles.

Proposition 2.2.6 *(i) Soit* E *un ensemble fini et* A *et* B *deux parties de* E.

1. *si* A *et* B *sont disjoints, alors* $\mathrm{Card}(\mathrm{A}\cup\mathrm{B}) = \mathrm{Card}(\mathrm{A}) + \mathrm{Card}(\mathrm{B})$;
2. $\mathrm{Card}(\mathrm{A}^c) = \mathrm{Card}\,\mathrm{E} - \mathrm{Card}\,\mathrm{A}$;
3. $\mathrm{Card}(\mathrm{A}\cup\mathrm{B}) = \mathrm{Card}\,\mathrm{A} + \mathrm{Card}\,\mathrm{B} - \mathrm{Card}(\mathrm{A}\cap\mathrm{B})$.

(ii) Soient $\mathrm{E}_1, \mathrm{E}_2, \dots, \mathrm{E}_n$ *des ensembles finis. Alors*

$$\mathrm{Card}(\mathrm{E}_1 \times \cdots \times \mathrm{E}_n) = \mathrm{Card}(\mathrm{E}_1) \cdots \times \mathrm{Card}(\mathrm{E}_n).$$

Exemple 2.2.7 On tire au hasard un nombre entre 1 et 1000. Quelle est la probabilité pour que ce nombre soit divisible par 2 ou 5 ?

On modélise l'expérience par l'univers $\Omega = \{1,2,\dots,1000\}$, muni de la probabilité uniforme. Soient A l'évènement “le nombre est divisible par 2" et et B l'évenement “le nombre est divisible par 5". Alors

$$\mathrm{Card}(\mathrm{A}) = 1000/2 = 500,\ \mathrm{Card}\,\mathrm{B} = 1000/5 = 200,\ \ \mathrm{Card}(\mathrm{A}\cap\mathrm{B}) = 1000/10 = 100$$

et donc

$$\mathrm{Card}(A \cup B) = \mathrm{Card}\, A + \mathrm{Card}\, B - \mathrm{Card}(A \cap B) = 600.$$

La probabilité cherchée est donc $p = 600/1000 = 0.6$.

Mentionnons quelques formules concernant les cardinaux d'ensembles d'applications.

Définition 2.2.8 (Listes et arrangements) Soient E un ensemble fini. Soit $p \in \mathbf{N}^*$.
(i) On appelle *p-liste* d'éléments de E tout p-tuplet $(x_1, \dots, x_p)$ dans E^p.
(ii) On appelle *p-arrangement* d'éléments de E tout p-tuplet $(x_1, \dots, x_p)$ dans E^p, avec $x_i \neq x_j$ pour tout $i \neq j$.

Proposition 2.2.9 *Soient* E *et* F *deux ensembles finis de cardinaux respectifs p et n.*
(i) Le nombre d'applications $f : E \to F$ est n^p et est égal au nombre de p-liste d'éléments de F.
(ii) Soit $p \leq n$. Le nombre d'applications injectives $f : E \to F$ *est*

$$\boxed{A_n^p := n \times (n-1) \times \cdots \times (n-p+1) = \frac{n!}{(n-p)!};}$$

ce nombre est égal au nombre de p-arrangements d'éléments de F.
(iii) Le nombre d'applications bijectives $f : E \to E$ *est*

$$n! := n \times (n-1) \times \cdots \times 1;$$

ce nombre est égal au nombre de permutations d'éléments de E.

Démonstration Soit $E = \{x_1, \dots, x_p\}$.
(i) Se donner une application $f : E \to F$ revient à choisir une p-liste $(f(x_1), \dots, f(x_p))$ d'éléments de F. Pour chaque $x_i \in E$, il y a n choix possibles pour $f(x_i)$ et donc n^p choix possibles pour f.
(ii) Se donner une application injective $f : E \to F$ revient à choisir un p-arrangement $(f(x_1), \dots, f(x_p))$ d'éléments de F. Pour x_1, il y a n choix possibles ; quand $f(x_1)$ est fixé, il y a $n-1$ choix possibles pour $f(x_2)$, car $f(x_2) \neq f(x_1)$ par injectivité de f ; de proche en proche, on voit qu'il y a

$$n \times (n-1) \times \cdots \times (n-p+1)$$

choix possibles pour f.
(iii) On applique (ii) à E = F et donc $p = n$.■

Exemple 2.2.10 On considère une urne contenant n boules numérotées de 1 à n (et donc discernables).
(i) **(Tirage avec remise et ordonné)** On effectue p-tirages successifs en remettant à chaque fois la boule tirée. Le nombre de tirages possibles est le nombre de p-listes et est donc n^p. Par exemple, le nombre de mots de 5 lettres formés avec l'alphabet latin est 26^5 (ici $n = 26, p = 5$).
(i) **(Tirage sans remise et ordonné)** On effectue p-tirages successifs en ne remettant pas la boule tirée. Le nombre de tirages possibles est le nombre de p-arrangements et est donc $\mathrm{A}_n^p = n \times (n-1) \times \cdots \times (n-p+1)$. Par exemple, la probabilité de gagner au tiercé avec 20 chevaux en course est

$$\frac{1}{\mathrm{A}_{20}^3} = \frac{1}{20.19.18} \sim 0.0002.$$

Passons maintenant au dénombrement de parties d'ensembles.
On rappelle que pour $0 \leq p \leq n$, le *coefficient binomial* $\binom{n}{p}$ est défini par

$$\boxed{\binom{n}{p} = \frac{n!}{p!(n-p)!} = \frac{n(n-1)\cdots(n-r+1)}{r!},}$$

avec la convention $0! = 1$.

Définition 2.2.11 **(Combinaisons)** Soit E un ensemble fini à n éléments et $0 \leq p \leq n$ un entier.
(i) On appelle *combinaison* de p-éléments de E toute partie de E à p-éléments.
(ii) On appelle *combinaison avec répétition* de p-éléments de E tout n-tuplet $(k_1, \ldots, k_n)$ dans $\mathbf{N}^n$ avec $k_1 + \cdots + k_n = p$.

Proposition 2.2.12 *Soit* E *un ensemble fini à n éléments et* $0 \leq p \leq n$ *un entier.*
(i) Le nombre de combinaisons de p-éléments de E *est* $\binom{n}{p} = \dfrac{n!}{p!(n-p)!}$.
(ii) Le nombre de combinaisons avec répétitions de p-éléments de E *est* $\binom{n+p-1}{n-1} = \dfrac{(n+p-1)!}{p!(n-1)!}$

Démonstration (i) Soit c_n^p le nombre de combinaisons de p-éléments. Le nombre de p-arrangements est $\mathrm{A}_n^p = \dfrac{n!}{(n-p)!}$. Pour choisir un p-arrangement, on choisit une combinaisons de p-éléments (pour laquelle il y a c_n^p choix possibles) et on ordonne ces éléments (il y a $p!$ choix possibles). D'où $\mathrm{A}_n^p = c_n^p \times p!$ et donc

$$c_n^p = \frac{n!}{p!(n-p)!} = \binom{n}{p}.$$

(ii) Soit Γ_n^p le nombre de combinaisons avec répétitions de p-éléments. Alors Γ_n^p est le nombre d'applications $f : \{1, \dots, n\} \to \{0, \dots, p\}$ vérifiant

$$f(1) + \cdots + f(n) = p.$$

On aligne $n + p - 1$ objets, disons des boules blanches ; on en colorie $n - 1$ en noir. À chaque coloriage correspond une et une seule application f vérifiant

$$f(1) + \cdots + f(n) = p.$$

En effet, on définit $f(1)$ comme le nombre de boules blanches à gauche de la 1ère boule noire ; $f(2)$ comme le nombre de boules blanches entre les 2 premières boules noires ; $f(3)$ comme le nombre de boules blanches entre les boules noires suivantes, et ainsi de suite.

Réciproquement, si $f : \{1, \dots, n\} \to \{0, \dots, p\}$ vérifiant $f(1) + \cdots + f(n) = p$ est donnée, on colorie la boule numéro $f(1) + 1$, puis la boule $f(1) + f(2) + 2$, etc, jusqu'à la la boule $f(1)+\cdots+f(n-1)+n-1$. Comme il y a $\binom{n+p-1}{n-1}$ choix possibles pour le coloriage, on a bien $\Gamma_n^p = \binom{n+p-1}{n-1}$.

Par exemple, si $n = 7$ et $p = 5$, on dessine 11 boules (carrées !)

□□□□□□□□□□□

On en colorie 6 en noir

■□□■□■■□■■□

L'application $f : \{1, \dots, n\} \to \{0, \dots, p\}$ correspondante est donnée par $f(1) = 0, f(2) = 2, f(3) = 1, f(4) = 0, f(5) = 1, f(6) = 0, f(7) = 1$■

Exemple 2.2.13 On considère une urne contenant n boules numérotées de 1 à n (et donc discernables)

(i) **(Tirage sans remise et non ordonné)** On effectue p tirages successifs en ne remettant pas la boule tirée et sans se soucier de l'ordre d'apparition des boules. Le nombre de tirages possibles est le nombre de p-combinaisons et est donc $\binom{n}{p} = \dfrac{n!}{p!(n-p)!}$.

Par exemple, le nombre de binômes dans une classe de 30 étudiants est $\binom{30}{2} = \dfrac{30!}{2!(28)!} = 30.29/2 = 435$.

(i) **(Tirage avec remise et non ordonné)** On effectue p-tirages successifs remettant à chaque fois la boule tirée. On ne s'intéresse pas à l'ordre d'apparition des boules, mais seulement au nombre $k_1, \dots, k_n$, où k_i désigne le nombre d'apparitions de la boule i. Il s'agit du nombre de p-combinaisons avec répétition de $\mathrm{E} = \{1, \dots, n\}$.

Considérons l'exemple suivant : 4 listes se présentent à des élections, où 9 sièges sont à pourvoir. Le nombre de répartitions de sièges possibles entre les listes est le nombre de 4-uplet $(k_1, \dots, k_4)$ dans $\mathbf{N}^4$ avec $k_1 + k_2 + k_3 + k_4 = 9$, où k_i représente le nombre de sièges de la liste $i \in \{1,2,3,4\}$, c-à-d du nombre de combinaisons avec répétition de 9-éléments de $\{1,2,3,4\}$ (ici, $n = 4$ et $p = 9$) qui est donc égale à

$$\binom{9+4-1}{4-1} = \binom{12}{3} = \frac{9 \times 8 \times 7}{1 \times 2 \times 3} = 84.$$

Exemple 2.2.14 (Nombre de piles dans un jeu de pile ou face) Pour $0 \le k \le n$, on cherche la probabilité de l'évènement

$$\mathrm{A}_k : \text{ “obtenir exactement } k \text{ piles"}$$

en n parties de pile ou face avec une pièce équilibrée. Cette expérience aléatoire est modélisée par $\Omega = \{0,1\}^n$, l'ensemble des suites $(a_1, a_2, \dots, a_n)$ avec $a_i = 1$ si on obtient pile au i-ième lancer et $a_i = 0$ sinon. L'évènement en question correspond à la partie

$$\mathrm{A}_k = \{(a_1, \dots, a_n) \in \Omega \, : \, a_1 + \dots + a_n = k\}.$$

Alors on a $\mathrm{Card}(\mathrm{A}_k) = \binom{n}{k}$, car chaque $(a_1, \dots, a_n) \in \mathrm{A}_k$ correspond à une partie unique de $\{1, \dots, n\}$ avec k éléments : l'ensemble des $i \in \{1, \dots, n\}$ tels que $a_i = 1$. Réciproquement, à une partie X de $\{1, \dots, n\}$ avec k éléments, on peut associer un unique élément de A_k : l'élément $(a_1, \dots, a_n) \in \Omega$ tel que $a_i = 1$ si et seulement si $i \in \mathrm{X}$.

Il s'ensuit que

$$\mathbf{P}(\mathrm{A}_k) = \frac{\mathrm{Card}(\mathrm{A}_k)}{\mathrm{Card}(\Omega)} = \frac{\binom{n}{k}}{2^n}.$$

Nous verrons plus tard que ceci est un cas particulier de la loi binomiale. Obervons que la formule plus haut permet de retrouver le fait que

$$\mathrm{Card}(\mathscr{P}(\mathrm{E})) = 2^n$$

pour tout ensemble fini E avec $\mathrm{Card}(E) = n$. En effet,

$$\mathrm{Card}(\mathscr{P}(E)) = \sum_{k=0}^{n} \mathrm{Card}(A_k) = \sum_{k=0}^{n} 2^n \mathbf{P}(A_k) = 2^n \sum_{k=0}^{n} \mathbf{P}(A_k) = 2^n.$$

Exemple 2.2.15 (Sondage simple) On considère une urne contenant N boules dont N_1 sont blanches et N_2 noires, avec $N = N_1 + N_2$. On effectue un tirage de $n \leq N$ boules sans remise. Pour $0 \leq k \leq n$, on cherche la probabilité de l'évènement A_k :"obtenir exactement k boules blanches." Il y a $\binom{N}{n}$ tirages possibles. Si on considère que ces tirages sont équiprobables, on peut modéliser l'expérience par l'univers Ω, l'ensemble des sous-ensembles à n ensembles de $\{1,\ldots,N\}$. Le nombres de tirages avec exactement k boules blanches est $\binom{N_1}{k}\binom{N_2}{n-k}$. D'où

$$\mathbf{P}(A_k) = \frac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N}{n}}.$$

C'est un cas particulier de la **loi hypergéométrique**.

Par exemple, dans une main de Poker, on tire 5 cartes parmi 52. La probabilité d'obtenir un paire d'As est

$$\mathbf{P}(\text{“paire d'As"}) = \frac{\binom{4}{2}\binom{48}{3}}{\binom{52}{5}} \sim 0.144.$$

Exemple 2.2.16 (Problème des anniversaires) On cherche la probabilité p_n que, parmi n personnes, au moins deux d'entre elles soient nées le même jour. On suppose que les naissances sont équidistribuées sur l'année et on ne tient pas compte des années bissextiles.

On modélise cette expérience par $\Omega = \{1,\ldots,d\}^n$ avec $d = 365$, qu'on munit de la mesure uniforme. On considère l'évènement

A : "au moins deux personnes sont nées le même jour".

Alors A^c est l'évènement “les n personnes sont nées en des jours différents"; $\mathrm{Card}(A^c)$ est donc le nombre de n-arrangements dans un ensemble à d éléments. D'où $\mathbf{P}(A^c) = \dfrac{d(d-1)\ldots(d-n+1)}{d^n}$ et donc

$$p_n = \mathbf{P}(A) = 1 - \mathbf{P}(A^c) = 1 - \frac{d(d-1)\ldots(d-n+1)}{d^n}.$$

On trouve

$$p_{20} \sim 0.41, \qquad p_{30} \sim 0.71, \qquad p_{40} \sim 0.89, \qquad p_{50} \sim 0.97.$$

2.2.3 Probabilités conditionnelles

Considérons le lancer d'un dé équilibré, modélisé par $\Omega = \{1, \dots, 6\}$ muni de la probabilité uniforme. La probabilité d'un évènement A (comme, par exemple, “obtenir 4 ou plus") est donnée par

$$\mathbf{P}(A) = \frac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)}$$

(donc 1/2 dans l'exemple). Supposons qu'on dispose de l'information qu'un évènement B s'est produit (comme, par exemple, “le résultat est pair" et donc $B = \{2, 4, 6\}$). Alors, l'univers des cas possibles ainsi que l'ensemble des cas favorables ont changé : Ω est remplacé par B et A par $A \cap B$. La nouvelle probabilité est

$$\frac{\mathrm{Card}(A \cap B)}{\mathrm{Card}(B)}$$

(donc 2/3 dans l'exemple).

Définition 2.2.17 (Probabilité conditionnelle) Soient $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilité et $B \in \mathscr{F}$ avec $\mathbf{P}(B) > 0$. Pour tout $A \in \mathscr{F}$, la *probabilité conditionnelle de* A *sachant* B, notée $\mathbf{P}(A|B)$ est le nombre

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

Voici quelques propriétés importantes des probabilités conditionnelles.

Proposition 2.2.18 *Soient* $(\Omega, \mathscr{F}, \mathbf{P})$ *un espace de probabilité,* $A, B \in \mathscr{F}$ *avec* $\mathbf{P}(A) > 0$ *et* $\mathbf{P}(B) > 0$. *Alors*

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

Démonstration On a, par définition, $\mathbf{P}(A|B) = \dfrac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}$ et $\mathbf{P}(B|A) = \dfrac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)}$. L'assertion s'ensuit. ■

Proposition 2.2.19 *Soient* $(\Omega, \mathscr{F}, \mathbf{P})$ *un espace de probabilité et* $B \in \mathscr{F}$ *avec* $\mathbf{P}(B) > 0$. *Alors l'application*

$$\mathbf{P}(\cdot|B) : \mathscr{F} \to [0, 1], \qquad A \mapsto \mathbf{P}(A|B)$$

est une mesure de probabilité, c-à-d
(i) $\mathbf{P}(\Omega|B) = 1$;
(ii) $\mathbf{P}(A^c|B) = 1 - \mathbf{P}(A|B)$ *pour tout* $A \in \mathscr{F}$;
(ii) $\mathbf{P}(\cup_n A_n|B) = \sum_n \mathbf{P}(A_n|B)$ *pour toute suite* $A_1, A_2, \dots$ *d'évènements disjoints de* $\mathscr{F}$.

Démonstration Les propriétés (i), (ii) et (iii) sont des conséquences immédiates du fait que $\mathbf{P}$ est une mesure de probabilité sur $\mathscr{F}$.■

Théorème 2.2.20 ***(Formule des probabilités totales)*** *Soit* $(\Omega, \mathscr{F}, \mathbf{P})$ *un espace de probabilité et soit* $\Omega = \cup_n B_n$ *une partition de* Ω *pour une suite (finie ou infinie) d'évènements* $B_n \in \mathscr{F}$.
(i) ***Formule des probabilités totales- 1ère version)*** *Pour tout* $A \in \mathscr{F}$, *on a*

$$\boxed{\mathbf{P}(A) = \sum_n \mathbf{P}(A \cap B_n).}$$

(ii) ***Formule des probabilités totales- 2ème version)*** *On suppose que* $\mathbf{P}(B_n) > 0$ *pour tout n. Pour tout* $A \in \mathscr{F}$, *on a*

$$\boxed{\mathbf{P}(A) = \sum_n \mathbf{P}(A|B_n)\mathbf{P}(B_n).}$$

En particulier, pour tout $B \in \mathscr{F}$ *avec* $0 < \mathbf{P}(B) < 1$, *on a*

$$\boxed{\mathbf{P}(A) = \mathbf{P}(A|B)\mathbf{P}(B) + \mathbf{P}(A|B^c)\mathbf{P}(B^c)}, \qquad \textit{pour tout} \quad A \in \mathscr{F}$$

Démonstration (i) On a $A = \cup_n A \cap B_n$, car $\Omega = \cup_n B_n$ est une partition. Les évènements $A \cap B_n$ étant disjoints, on a donc

$$\mathbf{P}(A) = \mathbf{P}(\cup_n A \cap B_n) = \sum_n \mathbf{P}(A \cap B_n).$$

(ii) La formule découle de (i) et du fait que $\mathbf{P}(A \cap B_n) = \mathbf{P}(A|B_n)\mathbf{P}(B_n)$, par définition de $\mathbf{P}(A|B_n)$.■

Le conditionnement est souvent utilisé dans les expériences aléatoires qui comportent plusieurs étapes.

Exemple 2.2.21 Dans une population donnée, on pense qu'une personne sur 1000 est atteinte d'une certaine maladie. On met au point un test médical pour la dépister. Il s'avère que le test déclare positifs 99% des malades testés ainsi que 2% des personnes saines. On choisit une personne au hasard (1ère étape : elle peut être malade ou non) et on la soumet au test (2ème étape : le test peut être positif ou non).

Soient A l'évènement "le test est positif" et B l'évènement "la personne est malade". Alors

$$\mathbf{P}(B) = \frac{1}{1000}, \ \mathbf{P}(A|B) = \frac{99}{100}, \ \mathbf{P}(A|B^c) = \frac{2}{100}.$$

La probabilité ("totale") de l'évènement A est

$$\begin{aligned}\mathbf{P}(A) &= \mathbf{P}(A|B)\mathbf{P}(B) + \mathbf{P}(A|B^c)\mathbf{P}(B^c)\\ &= \frac{99}{100}\frac{1}{1000} = \frac{2}{100}\frac{999}{1000} \sim 0.02.\end{aligned}$$

La 1ère étape dans l'exemple plus haut peut être vue comme une "cause" (maladie ou non) et la 2ème comme une "conséquence" (test positif ou non). On peut essayer de retrouver la cause en partant de la conséquence. C'est la formule de Bayes (ou "formule de la probabilité des causes").

Théorème 2.2.22 ***(Formule de Bayes)*** *Soit* $A, B \in \mathscr{F}$ *avec* $\mathbf{P}(A) > 0$ *et* $0 < \mathbf{P}(B) < 1$. *Alors*

$$\boxed{\mathbf{P}(B|A) = \frac{\mathbf{P}(A|B)\mathbf{P}(B)}{\mathbf{P}(A\cap B) + \mathbf{P}(A\cap B^c)} = \frac{\mathbf{P}(A|B)\mathbf{P}(B)}{\mathbf{P}(A|B)\mathbf{P}(B) + \mathbf{P}(A|B^c)\mathbf{P}(B^c)}}$$

Démonstration On a, d'après la formule des probabilités totales,

$$\begin{aligned}\mathbf{P}(A|B)\mathbf{P}(B) &= \mathbf{P}(B|A)\mathbf{P}(A)\\ &= \mathbf{P}(B|A)(\mathbf{P}(A\cap B) + \mathbf{P}(A\cap B^c)\\ &= \mathbf{P}(B|A)(\mathbf{P}(A|B)\mathbf{P}(B) + \mathbf{P}(A|B^c)\mathbf{P}(B^c)).\end{aligned}$$

Exemple 2.2.23 Reprenons l'exemple plus haut. Calculons la probabilité $\mathbf{P}(B|A)$ qu'une personne testée positive (évènement B) est effectivement malade (évènement A). On a

$$\begin{aligned}\mathbf{P}(B|A) &= \frac{\mathbf{P}(A|B)\mathbf{P}(B)}{\mathbf{P}(A\cap B) + \mathbf{P}(A\cap B^c)}\\ &= \frac{0.99 \times 0.001}{0.99 \times 0.001 + 0.02 \times 0.999}\\ &\sim 0.05,\end{aligned}$$

ce qui est une probabilité très faible.

Théorème 2.2.24 ***(Formule des probabilités composées)*** *Soit* $(\Omega, \mathscr{F}, \mathbf{P})$ *un espace de probabilité et soient* $A_1, \ldots, A_n \in \mathscr{F}$ *tels que* $\mathbf{P}(A_1 \cap \cdots \cap A_n) > 0$. *Alors, on a*

$$\boxed{\mathbf{P}(A_1 \cap \cdots \cap A_n) = \mathbf{P}(A_1)\mathbf{P}(A_2|A_1)\mathbf{P}(A_3|A_1 \cap A_2)\cdots\mathbf{P}(A_n|A_1 \cap \cdots \cap A_{n-1}).}$$

Démonstration On a

$$\begin{aligned}
\mathbf{P}(A_1)\mathbf{P}(A_2|A_1)\mathbf{P}(A_3|A_1 \cap A_2)\cdots\mathbf{P}(A_n|A_1 \cap \cdots \cap A_n) &= \\
&= \mathbf{P}(A_1)\frac{\mathbf{P}(A_2 \cap A_1)}{\mathbf{P}(A_1)}\frac{\mathbf{P}(A_3 \cap A_2 \cap A_1)}{\mathbf{P}(A_2 \cap A_1)}\cdots\frac{\mathbf{P}(A_1 \cap \cdots \cap A_n)}{\mathbf{P}(A_1 \cap \cdots \cap A_{n-1})} \\
&= \mathbf{P}(A_1 \cap \cdots \cap A_n).
\end{aligned}$$

On observera que toutes les probabilités conditionnelles sont bien définies : on a $\mathbf{P}(A_1 \cap \cdots \cap A_i) > 0$ pour tout i car $\mathbf{P}(A_1 \cap \cdots \cap A_n) > 0$ par hypothèse.■

2.2.4 Indépendance

Etant donnés deux évènements A et B dans un espace de probabilité $(\Omega, \mathscr{F}, \mathbf{P})$, supposons que la réalisation de A ne dépende pas de celle de B; cela veut dire que la probabilité de A sachant B est simplement celle de A : $\mathbf{P}(A|B) = \mathbf{P}(A)$, c-à-d $\dfrac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \mathbf{P}(A)$. Nous dirons alors que A est B sont indépendants.

Définition 2.2.25 **(Indépendance de deux évènements)** Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilité. Deux évènements A et B dans $\mathscr{F}$ sont dits *indépendants* si

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

Remarque 2.2.26 Deux évènements A et B avec $\mathbf{P}(A) > 0$ et $\mathbf{P}(B) > 0$ sont indépendants si et seulement $\mathbf{P}(A|B) = \mathbf{P}(A)$ ou bien si et seulement $\mathbf{P}(B|A) = \mathbf{P}(B)$.

Exemple 2.2.27 Considérons le lancer d'un dé équilibré, modélisé par l'univers $\Omega = \{1, 2, 3, 4, 5, 6\}$, la tribu $\mathscr{P}(\Omega)$ et la mesure de probabilité uniforme. Les évènements A : « le résultat est ≤ 4 » et B : « le résultat est pair » sont indépendants : en effet, on a

$$\mathbf{P}(A) = \frac{4}{6} = \frac{2}{3}, \ \mathbf{P}(B) = \frac{3}{6} = \frac{1}{2}, \ \mathbf{P}(A \cap B) = \frac{2}{6} = \frac{1}{3}$$

et donc $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$.

Proposition 2.2.28 *Soient* A *et* B *deux évènements indépendants. Alors* A *et* B^c *sont indépendants.*

Démonstration On a

$$A = (A \cap B) \cup (A \cap B^c)$$

et donc, la réunion étant disjointe et les évènements A et B indépendants,

$$\begin{aligned}\mathbf{P}(A \cap B^c) &= \mathbf{P}(A) - \mathbf{P}(A \cap B)\\ &= \mathbf{P}(A) - \mathbf{P}(A)\mathbf{P}(B) = \mathbf{P}(A)(1 - \mathbf{P}(B))\\ &= \mathbf{P}(A)\mathbf{P}(B^c),\end{aligned}$$

ce qui prouve l'indépendance de A et B^c.■

Définition 2.2.29 (Indépendance d'une famille d'évènements) Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilité et soit $(A_i)_{i\in I}$ une famille (finie ou infinie) d'évènements dans $\mathscr{F}$.

— On dit que les évènements A_i sont *mutuellement indépendants* (ou bien indépendants) si, pour toute partie finie $J \subset I$, on a

$$\mathbf{P}(\cap_{j\in J} A_j) = \prod_{j\in J} \mathbf{P}(A_j).$$

En particulier, des évènements $A_1, \ldots, A_n$ sont dits mutuellement indépendants, si pour toute partie finie $J \subset \{1, \ldots, n\}$, on a $\mathbf{P}(\cap_{j\in J} A_j) = \prod_{j\in J} \mathbf{P}(A_j)$.

— On dit que les évènements A_i sont *deux à deux indépendants* (ou bien indépendants) si, A_i et A_j sont indépendants pour tout $i \neq j$ (c-à-d $\mathbf{P}(A_i \cap A_j) = \mathbf{P}(A_i)\mathbf{P}(A_j)$).

Remarque 2.2.30 Il est clair que, si des évènements A_i sont mutuellement indépendants, alors ils sont deux à deux indépendants. La réciproque est fausse, comme le montre l'exemple suivant.

Exemple 2.2.31 Considérons le lancer d'un dé équilibré deux fois, modélisé par l'univers $\Omega = \{1, 2, 3, 4, 5, 6\}^2$, la tribu $\mathscr{P}(\Omega)$ et la mesure de probabilité uniforme. Considérons les évènements A : « le premier résultat est pair », B : « le deuxième résultat est impair» et C : « la somme des deux résultats est paire». Alors, en comptant les cas possibles, on trouve $\mathbf{P}(A) = \mathbf{P}(B) = \mathbf{P}(C) = \dfrac{1}{2} = \dfrac{2}{3}$,

$\mathbf{P}(A \cap B) = \mathbf{P}(A \cap C) = \mathbf{P}(B \cap C) = \frac{9}{36} = \frac{1}{4}$. Par conséquent, A, B, C sont deux à deux indépendants. Or, comme $A \cap B \cap C = \emptyset$, on a

$$\mathbf{P}(A \cap B \cap C) = 0 \neq \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C).$$

Ceci implique que A, B, C ne sont *pas* mutuellement indépendants.

Remarque 2.2.32 Soient A et B deux évènements ; il importe de pas confondre la propriété « A et B sont incompatibles » avec la propriété « A et B sont indépendants ». La première notion est purement ensembliste et signifie que $A \cap B = \emptyset$; elle s'interprète en disant que A et B ne peuvent pas arriver simultanément. La seconde nécessite une mesure de probabilité : intuitivement, elle signifie que savoir que A a eu lieu ne donne aucune information sur B.

Il faut noter que, si A et B sont à la fois indépendants et incompatibles alors $\mathbf{P}(A)\mathbf{P}(B) = \mathbf{P}(A \cap B) = \mathbf{P}(\emptyset) = 0$ et donc $\mathbf{P}(A) = 0$ ou $\mathbf{P}(B) = 0$

2.2.5 Produits d'espaces probabilisés-Epreuves de Bernoulli

On désire modéliser l'expérience aléatoire qui consiste à répéter n fois une épreuve, dans les mêmes conditions, le résultat de chacune des épreuves n'influançant pas celui des autres épreuves. Des exemples de telles expériences aléatoires sont :

— des tirages successifs avec remise dans une urne ;
— des lancers successifs d'un dé ;
— le contrôle de n pièces issues d'une usine ...

A chaque épreuve, on associe le même espace de probabilité $(\Omega, \mathscr{F}, \mathbf{P})$. L'univers attaché à l'épreuve global est le produit cartésien

$$\Omega^n = \{(\omega_1, \ldots, \omega_n) : \omega_i \in \Omega \qquad \text{pour tout} \quad i = 1, \ldots, n\},$$

muni de la tribu $\mathscr{F}^n$ engendrée par les évènements du type $A_1 \times \cdots \times A_n$ avec $A_i \in \mathscr{F}$ et de la mesure de probabilité $\mathbf{P} : \mathscr{F}^n \to [0, 1]$, définie par

$$\mathbf{P}(A_1 \times \cdots \times A_n) = \mathbf{P}(A_1) \cdots \mathbf{P}(A_n).$$

Un cas particulier important est celui où il n'y a que deux issues possibles (0 ou 1 ; “oui" ou “non" ; “succès" ou “échec",...).

Définition 2.2.33 (Schéma de Bernoulli)
(i) On appelle *épreuve de Bernoulli* une expérience aléatoire ne comportant que deux issues possibles et donc modelisée par $(\Omega, \mathscr{F}, \mathbf{P})$, où $\Omega = \{0,1\}$, $\mathscr{F} = \mathscr{P}(\Omega)$ et $\mathbf{P} : \mathscr{P}(\Omega) \to [0,1]$ est donnée par $\mathbf{P}(\{1\}) = p$ et $\mathbf{P}(\{0\}) = 1-p$ pour un nombre $p \in [0,1]$.
(ii) Soit $n \geq 1$. On appelle *schéma de Bernoulli* la répétition de n épreuves de Bernoulli indépendantes; un tel schéma est modelisé par $(\Omega, \mathscr{F}, \mathbf{P})$, où $\Omega = \{0,1\}^n$, $\mathscr{F} = \mathscr{P}(\Omega)$ et $\mathbf{P} : \mathscr{P}(\Omega) \to [0,1]$ donnée par

$$\mathbf{P}(\{(\omega_1, \omega_n)\}) = \mathbf{P}(\{\omega_1\}) \cdots \mathbf{P}(\{\omega_n\}) = p^k (1-p)^{n-k},$$

ou k est le nombre de 1 dans la suite $(\omega_1, \ldots, \omega_n)$ et $n-k$ le nombre de 0.

2.3 Exercices

Exercice 2.3.1 Soient Ω un ensemble et A et B des parties de Ω.
Simplifier les expressions

$$\mathrm{E} = (\mathrm{A} \cup \mathrm{B}) \cap (\mathrm{A}^c \cup \mathrm{B}) \cap (\mathrm{A} \cap \mathrm{B}^c) \qquad \text{et} \qquad \mathrm{F} = (\mathrm{A} \cup \mathrm{B}) \cap (\mathrm{A}^c \cup \mathrm{B}) \cap (\mathrm{A} \cup \mathrm{B}^c).$$

Exercice 2.3.2 Soient Ω un ensemble et A, B, C des parties de Ω.
Parmi les propositions suivantes, prouver celles qui sont vraies et donner des contre-exemples pour les autres :

1. $(\mathrm{A} \cup \mathrm{B} = \Omega) \Longrightarrow (\mathrm{A} \subset \mathrm{B}^c)$
2. $(\mathrm{A} \cup \mathrm{B} = \Omega) \Longrightarrow (\mathrm{A}^c \subset \mathrm{B})$
3. $(\mathrm{A} \cup \mathrm{B} = \Omega \text{ et } \mathrm{A} \cap \mathrm{B} = \emptyset) \Longrightarrow (\mathrm{A} = \mathrm{B}^c)$
4. $(\mathrm{A} \subset \mathrm{B}^c) \Longrightarrow (\mathrm{A} \cup \mathrm{B} = \Omega)$
5. $(\mathrm{A} \cap \mathrm{B}) \cup \mathrm{C} = \mathrm{A} \cap (\mathrm{B} \cup \mathrm{C})$.

Exercice 2.3.3 Soit Ω un ensemble; pour toute partie A de Ω, on rappelle que $\mathbf{1}_\mathrm{A}$ est la fonction indicatrice de A.
(i) Montrer que, pour toutes parties A, B de Ω, on a

$$\mathbf{1}_{\mathrm{A}\cap\mathrm{B}} = \mathbf{1}_\mathrm{A}\mathbf{1}_\mathrm{B},\ \mathbf{1}_{\mathrm{A}\cup\mathrm{B}} = \mathbf{1}_\mathrm{A} + \mathbf{1}_\mathrm{B} - \mathbf{1}_{\mathrm{A}\cap\mathrm{B}},$$
$$\mathbf{1}_{\mathrm{A}^c} = \mathbf{1}_\Omega - \mathbf{1}_\mathrm{A},\ \mathbf{1}_\mathrm{A}\mathbf{1}_{\mathrm{A}^c} = 0,\ \mathbf{1}_\mathrm{A}\mathbf{1}_\mathrm{A} = \mathbf{1}_\mathrm{A}.$$

(ii) Utiliser (i) pour simplifier les expressions

$$(\mathrm{A}^c \cup \mathrm{B}) \cap (\mathrm{A} \cap \mathrm{B}^c) \qquad \text{et} \qquad (\mathrm{A}^c \cup \mathrm{B}) \cap (\mathrm{A} \cup \mathrm{B}^c).$$

Exercice 2.3.4 Soit $(\mathrm{A}_n)_{n\in\mathbf{N}}$ une suite infinie d'évènements d'un univers Ω.

(i) Montrer que

$$(\liminf \mathrm{A}_n)^c = \limsup \mathrm{A}_n^c \qquad \text{et} \qquad (\limsup \mathrm{A}_n)^c = \liminf \mathrm{A}_n^c.$$

(ii) Montrer que

$$\liminf \mathrm{A}_n = \{\sum_n \mathbf{1}_{\mathrm{A}_n^c} < \infty\} \qquad \text{et} \qquad \limsup \mathrm{A}_n = \{\sum_n \mathbf{1}_{\mathrm{A}_n} = \infty\}.$$

(iii) On pose $\mathrm{A}_\infty = \liminf \mathrm{A}_n$ et $\mathrm{A}^\infty = \limsup \mathrm{A}_n$. Montrer que

$$\mathbf{1}_{\mathrm{A}_\infty} = \liminf \mathbf{1}_{\mathrm{A}_n} \qquad \text{et} \qquad \mathbf{1}_{\mathrm{A}^\infty} = \limsup \mathbf{1}_{\mathrm{A}_n}.$$

Exercice 2.3.5 Soient E et F deux ensembles de cardinaux respectifs n et k. On note F^E l'ensemble des applications de E dans F : $\mathrm{F}^\mathrm{E} = \{f \mid f : \mathrm{E} \to \mathrm{F}\}$. Quel est le cardinal $\mathrm{Card}(\mathrm{F}^\mathrm{E})$ de F^E ?

Exercice 2.3.6 Soit Ω un ensemble. Decrire une bijection entre $\mathscr{P}(\Omega)$ et $\{0,1\}^\Omega$.

Exercice 2.3.7 On jette 10 fois successivement une pièce de monnaie. Comment peut-on représenter le résultat d'une telle épreuve ? Combien a t-on de résultats différents ? Dans combien de résultats obtient-on exactement trois piles ? Dans combien de résultats obtient-on au moins un pile ?

Exercice 2.3.8 Une urne contient 5 boules blanches numérotées de 1 à 5 et 8 boules noires numérotées de 6 à 13. On tire successivement 6 boules de l'urne ; à chaque fois, on note le numéro de la boule tirée et on la remet ensuite dans l'urne.
Comment peut représenter le résultat d'une telle épreuve ?
Combien y a-t-il de tirages possibles ?
Dans combien de tirages obtient-on :

1. 1 boule noire au plus ?
2. 3 boules blanches exactement ?
3. 1 boule blanche au moins ?
4. 5 boules noires et une blanche, dans cet ordre ?
5. 5 boules noires et une blanche, dans n'importe quel ordre ?

Exercice 2.3.9 Reprendre l'exercise précédent en supposant cette fois que les boules ne sont pas remises dans l'urne après tirage.

Exercice 2.3.10 On dispose de 13 livres différents : 4 de mathématique, 6 de physique et 3 de chimie.
On les range au hasard. Combien de façons y a-t-il de les ranger sur une même étagère ?
Combien de façons y a-t-il de les ranger en regroupant les livres de mathématique ?
Combien de façons y a-t-il de les ranger en les regroupant par matière ?

Exercice 2.3.11 On considère un groupe de 2N personnes qui veulent prendre place autour d'une table ronde.

1. Quel est le nombre de manières différentes de répartir ces personnes ?
2. On suppose qu'il y a N femmes et N hommes et on souhaite avoir une alternance (femme, homme, femme...). Quel est le nombre de manières différentes de répartir ces personnes ?

Exercice 2.3.12 Après une marée noire en Bretagne, l'organisme de protection des oiseaux de mer a évalué à 20 000 la population de sternes au large du Finistère. 500 d'entre eux ont été bagués. Un an après, on capture 100 sternes dans cette zone. Calculer la probabilité :

1. de ne pas avoir d'oiseau bagué ?
2. d'avoir au moins deux sternes baguées ?

Exercice 2.3.13 On lance trois fois une pièce de monnaie équilibrée. Expliciter l'espace de probabilité $(\Omega, \mathscr{F}, \mathbf{P})$ qui modélise cette expérience aléatoire.

— Quelle est la probabilité d'obtenir exactement une fois face ?
— Quelle est la probabilité d'obtenir au moins une fois face ?
— Quelle est la probabilité d'obtenir pile au 1er lancer et au moins une fois face lors des deux suivants ?

Exercice 2.3.14 Soient$(\Omega, \mathscr{F}, \mathbf{P})$ un espace de probabilité et $(A_1, A_2, A_3) \in \mathscr{F}^3$. On pose, pour $i, j \in \{1,2,3\}$, $i \neq j$,

$$p_i = \mathbf{P}(A_i),\ p_{ij} = \mathbf{P}(A_i \cap A_j) \text{ et } p_{123} = \mathbf{P}(A_1 \cap A_2 \cap A_3).$$

Exprimez en fonction de ces probabilités les probabilités suivantes :
1) les trois évènements se réalisent, 2) au moins l'un des évènements se réalise, 3) au moins deux des évènements se réalisent, 4) A_1 seul se réalise,

5) A_1 et A_2 se réalisent mais pas A_3, 6) deux évènements au plus se réalisent, 7) un seul évènement se réalise, 8) deux évènements seulement se réalisent, 9) deux évènements ou plus se réalisent, 10) aucun des trois évènements ne se réalise.

Exercice 2.3.15 (i) Combien de fois faut-il lancer un dé équilibré pour avoir au moins une chance sur deux d'obtenir un « six »?
(ii) Même question avec deux dés pour obtenir un « double-six »
(iii) Quel est l'évènement le plus probable : obtenir au moins un « six » en lançant 4 fois un dé ou bien obtenir au moins un « double-six » en lançant 24 fois une paire de dés?

Exercice 2.3.16 Au loto, on tire six numéros entre 1 et 49, deux-à-deux distincts et sans tenir compte de leur ordre. Calculez les probabilités des évènements suivants, pour $0 \le k \le 6$:

— "Avoir exactement k bons numéros"
— "Avoir zéro, un ou deux bons numéros"
— "Avoir au moins trois bons numéros"

Exercice 2.3.17 Pour $r \le n$, on répartit aléatoirement r boules à l'intérieur de n urnes, chaque urne pouvant contenir plusieurs boules. Expliciter l'espace de probabilité $(\Omega, \mathscr{F}, \mathbf{P})$ qui modélise cette expérience aléatoire.
(i) Déterminer la probabilité de l'évènement « chaque urne contient au plus une boule ».
(ii) Déterminer la probabilité de l'évènement « il existe une urne contenant au moins deux boules ».

Exercice 2.3.18 Une urne contient N boules, dont N_1 sont blanches et N_2 noires. On opère à des tirages successifs avec remise. Soient $n \ge 1$ et $k \le n$.
(i) Quelle est la probabilité que la première boule blanche tirée apparaisse lors du n-ième tirage?
(ii) Quelle est la probabilité que la k-ième boule blanche tirée apparaisse lors du n-ième tirage?

Exercice 2.3.19 On considère un jeu de 32 cartes. On en distribue 5 au hasard à chacun des deux joueurs A et B.

1. Calculer la probabilité pour que A ait au moins un as.
2. Sachant que B a exactement un as, calculer la probabilité pour que A ait au moins un as.

Exercice 2.3.20 Deux joueurs A et B s'exercent au tir à l'arc. Le joueur A ne tire qu'une fois sur 3 et atteint sa cible 9 fois sur 10 quand il tire. Le joueur B, moins adroit, n'atteint sa cible que 6 fois sur 10. Un des joueurs tire.

1. Quelle est la probabilité pour que la cible soit atteinte ?
2. Sachant que la cible est atteinte, quelle est la probabilité pour que ce soit par B?

Exercice 2.3.21 Une urne contient 4 boules rouges et 3 vertes.

1. On tire successivement sans remise deux boules dans l'urne. Sachant qu'au premier tirage, on a obtenu une boule rouge, quelle est la probabilité qu'on obtienne une boule verte au deuxième tirage ?
2. Répondre à la même question quand on suppose que le tirage se fait avec remise.

Exercice 2.3.22 Dans une usine, 1% des articles produits sont défectueux. Un contrôle de qualité permet de rejeter 95% des articles lorsqu'ils sont défectueux mais aussi de rejeter 2% des articles qui ne le sont pas.

1. Quelle est la probabilité qu'il y ait un erreur dans le contrôle de qualité ?
2. Quelle est la probabilité pour qu'un article accepté soit en réalité défectueux ?

Exercice 2.3.23 Dans un étang, il y a des poissons rouges et des poissons verts. Les poissons trop petits sont remis à l'eau par les pêcheurs.
On estime qu'il y a 60 pour cent de poissons rouges dans l'étang, que la moitié des poissons rouges et le tiers des poissons verts sont trop petits.
Quelle est la probabilité de pêcher un poisson trop petit ? Sachant qu'on a pêché un poisson trop petit, quelle est la probabilité que ce soit un poisson rouge ?

Exercice 2.3.24 On considère $n \geq 1$ individus $A_1, \ldots, A_n$. On lance une pièce de monnaie et on transmet le résultat ("Pile" ou "Face") à A_1. Le résultat est transmis par A_1 à A_2, ensuite par A_2 à A_3, etc. On suppose que tous ces individus mentent avec la probabilité p et qu'ils le font de manière mutuellement indépendante. Quelle est la probabilité p_n que le résultat reçu par A_n soit le bon ? Quelle est la limite $\lim_{n\to\infty} p_n$?

Exercice 2.3.25 Pour $N \geq 1$, on considère $N+1$ urnes numérotées $0, 1, \ldots, N$. L'urne numéro k contient k boules blanches et $N - k$ boules noires. On choisit une urne au hasard, tous les choix étant mutuellement indépendants.

1. Quelle est la probabilité p_n que la $(n+1)$-ième boule tirée soit blanche, sachant que les n précédentes le sont toutes.
2. Donner un équivalent de p_n pour $N \to \infty$.

Exercice 2.3.26 Soient $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé et A et B des évènements dans $\mathscr{F}$ tels que $\mathbf{P}(A) > 0, \mathbf{P}(B) > 0$ et $A \cap B = \emptyset$.
(i) Montrer que A et B ne sont pas indépendants.
(ii) Montrer que A n'est pas indépendant avec lui-même.

Exercice 2.3.27 On considère une urne contenant initialement une boule blanche et une boule noire. On effectue des tirages successifs de la façon suivante : on tire une boule de l'urne, on la remet en y ajoutant une boule de la même couleur que la boule qui a été tirée.

Pour $k \geq 1$, on considère les évènements
N_k : "une boule noire apparaît au k-ième tirage"
A_k : "une boule blanche apparaît pour la 1ère fois au k-ième tirage"
(i) Calculer les probabilités de A_1 et N_1.
(ii) Soit $k \geq 2$. Calculer les probabilités conditionnelles $\mathbf{P}(N_k | N_1 \cap \cdots \cap N_{k-1})$ et $\mathbf{P}(N_k^c | N_1 \cap \cdots \cap N_{k-1})$
(iii) Soit $n \geq 1$. Calculer la probabilité $\mathbf{P}(A_n)$.

Exercice 2.3.28 Au poker, une **main** désigne un ensemble de 5 cartes piochées dans un paquet de 52 cartes.

Il y a 4 **couleurs** (pique-coeur-carreau-trèfle) et 13 **hauteurs** consécutives (as-roi-dame-valet-dix-neuf-huit-sept-six-cinq-quatre-trois-deux).

On se propose de calculer les probabilités de piocher certaines combinaisons classiques au poker.
(i) Modéliser l'expérience aléatoire par un espace probabilisé adéquat. Quelle est le nombre total de mains ?
(ii) Une **quinteflush** est une main constituée de 5 cartes de même couleur et de hauteurs consécutives (i.e. as-roi-dame-valet-dix de pique ou as-roi-dame-valet-dix de coeur etc...). Quelle est la probabilité d'obtenir une quinteflush ?
(iii) Un **carré** est une main constituée de 4 cartes de même hauteur (i.e. 4 as, 4 rois etc...) et d'une cinquième carte. Quelle est la probabilité de piocher un carré ?
(iv) Un **full** est une main constituée de 3 cartes de même hauteur et de deux cartes de même hauteur. Quelle est la probabilité de piocher un full ?

(v) Une **couleur** est constituée de 5 cartes de même couleur et qui n'est pas une quinteflush. Quelle est la probabilité de piocher une couleur ?
(vi) Une **suite** est une main constituée de 5 cartes de hauteurs consécutives (i.e. as, roi, dame, valet, dix etc..) et qui n'est pas une quinteflush. Quelle est la probabilité de piocher une suite ?
(vii) Un **brelan** est une main constituée de 3 cartes de même hauteur et de deux autres cartes de hauteur différentes et qui n'est pas un carré. Quelle est la probabilité d'obtenir un brelan ?
(viii) Classer les combinaisons ci-dessus par ordre de rareté.

Exercice 2.3.29 (Problème de rencontres)
On suppose que n personnes déposent leurs chapeaux dans un vestiaire ; à la sortie, chacun prend un chapeau au hasard. Expliciter l'espace de probabilité $(\Omega, \mathscr{F}, \mathbf{P})$ qui modélise cette expérience aléatoire. Déterminer la probabilité p_n de l'évènement « au moins une personne récupère son propre chapeau ». Quelle est la limite $\lim_{n\to+\infty} p_n$?

Exercice 2.3.30 Soient $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé et $A, B, C \in \mathscr{F}$ trois évènements.
(i) On suppose que A et B sont indépendants ainsi que A et C. Les évènements A et $B \cup C$ sont-ils indépendants ? Même question pour A et $B \cap C$.
(ii) On suppose que A et $B \cup C$ sont indépendants ainsi que A et $B \cap C$. Les évènements A et B sont-ils indépendants ?

Exercice 2.3.31 Soient $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé et $A_1, \dots, A_n \in \mathscr{F}$ des évènements mutuellement indépendants. Pour chaque $i = 1, \dots, n$, soit $B_i = A_i$ ou $B_i = A_i^c$. Montrer que $B_1, \dots, B_n \in \mathscr{F}$ sont des évènements mutuellement indépendants.

Exercice 2.3.32 Soient $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé et $A_1, \dots, A_n \in \mathscr{F}$ des évènements mutuellement indépendants. Montrer que la probabilité qu'aucun des A_i ne soit réalisé est inférieure à $\exp(-\sum_{i=1}^n \mathbf{P}(A_i))$.
Indication : Utiliser l'Exercice 2.3.31.

Chapitre 3

Variables aléatoires

Supposons qu'on lance n fois une pièce équilibrée et qu'on s'intéresse au nombre de fois où on a obtenu “pile"; ce nombre X est compris entre 0 et n et varie selon le résultat de l'expérience. Ceci correspond à un schéma de Bernoulli (voir Definition 3.2.2). On trouve que alors, pour $0 \leq k \leq n$:

$$\mathbf{P}(X = k) = \frac{\binom{n}{k}}{2^n}.$$

De manière similaire, dans de nombreux problèmes, on s'intéresse aux aspects quantitatifs d'une expérience aléatoire (gain dans un jeu de cartes ou à la roulette, temps d'attente à un guichet, durée de vie d'une pièce,...). Ceci nous amène à la notion de variable aléatoire.

3.1 Variables aléatoires : généralités

Remarque 3.1.1 (Rappel : Propriétés de l'image réciproque par une application) Soient E et F deux ensembles et $f : E \to F$ une application et $B \subset F$ une partie de F. L'image réciproque de B par f est la partie de E définie par

$$f^{-1}(B) = \{x \in E \ : \ f(x) \in B\}.$$

On remarquera que f n'est pas nécessairement une bijection et que donc $f^{-1}(B)$ est un symbole qui ne signifie pas l'image de B par la fonction réciproque de f (qui n'existe pas nécessairement;. bien sûr, si f est une bijection, $f^{-1}(B)$ est effectivement l'image de B par f^{-1}.

Pour toute partie B de F et toute famille $(B_i)_{i\in I}$ de parties de F, on a les formules suivantes

$$f^{-1}(B^c) = f^{-1}(B)^c, \qquad f^{-1}(\cup_{i\in I} B_i) = \cup_{i\in I} f^{-1}(B_i)\,, \qquad f^{-1}(\cap_{i\in I} B_i) = \cap_{i\in I} f^{-1}(B_i).$$

Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé.

Définition 3.1.2 (Variable aléatoire) Une application $X : \Omega \to \mathbf{R}$ est appelée *variable aléatoire réelle* (en abrégé v.a.r) si, pour tout intervalle $I \subset \mathbf{R}$, l'ensemble

$$\{X \in I\} := X^{-1}(I) = \{\omega \in \Omega \,:\, X(\omega) \in I\}$$

appartient à la tribu $\mathscr{F}$.

Exemple 3.1.3 (i) L'expérience consistant en n lancers indépendants d'une pièce de pièce monnaie équilibrée est modélisée par $\Omega = \{0,1\}^n$ avec 1 représentant “pile", $\mathscr{F} = \mathscr{P}(\Omega)$ et $\mathbf{P}$ la probabilité uniforme. L'appliciation $X : \Omega \to \mathbf{R}$ définie par

$$X(\omega) = \sum_{i=1}^{n} \omega_i \qquad \text{pour tout} \quad \omega_i = (\omega_1, \dots, \omega_n) \in \Omega$$

prend comme valeurs le nombre nombre de fois où on a obtenu “pile" à l'épreuve ω. Cette fonction est bien une variable aléatoire, puisque $\{X \in I\} \in \mathscr{P}(\Omega)$ pour toute partie I de $\mathbf{R}$.
(ii) Dans l'exemple précédent, la v.a.r. prenait ses valeurs dans un ensemble fini. Ceci n'est pas toujours le cas : l'ensemble des valeurs d'une variable aléatoire peut être infini; par exemple, si on joue à pile ou face une infinité de fois, le numéro (ou l'instant) n où on obtient la première fois un “pile" est une v.a.r X qui peut prendre toute les valeurs dans $\mathbf{N}^* \cup \{\infty\}$.

Comme autre exemple, considérons un roue de loterie qu'on fait tourner en la lançant. L'angle X dont tourne la roue est une v.a.r qui peut prendre toutes les valeurs dans $[0, 2\pi[$.

Définition 3.1.4 (Variable aléatoire discrète) Etant donné $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé, nous dirons qu'une v.a.r. $X : \Omega \to \mathbf{R}$ est *discrète* si $X(\Omega)$, l'ensemble des valeurs de X, est contenu dans une partie finie ou dénombrable infinie de $\mathbf{R}$.

Pour vérifier qu'une fonction est un v.a.r, il n'est pas nécessaire de vérifier que $\{X \in I\} \in \mathscr{F}$ pour *tous* les intervalles $I \subset \mathbf{R}$.

Proposition 3.1.5 *(i) Une application* $X : \Omega \to \mathbf{R}$ *est une v.a.r. si et seulement si* $\{X \leq x\} = X^{-1}(]-\infty, x]) \in \mathscr{F}$ *pour tout* $x \in \mathbf{R}$.
(ii) Soit $X : \Omega \to \mathbf{R}$ *une application à valeurs dans une partie dénombrable (ou finie)* D *de* $\mathbf{R}$. *Alors* X *est une v.a.r. si et seulement si* $\{X = x\} \in \mathscr{F}$ *pour tout* $x \in D$.

Démonstration (i) Soit I un intervalle du type $I =]a, b]$. Alors $I =]-\infty, b] \cap]-\infty, a]^c$ et donc

$$X^{-1}(I) = X^{-1}(]-\infty, b]) \cap X^{-1}(]-\infty, a])^c \in \mathscr{F}.$$

Si $I = [a, b]$, alors

$$I =]-\infty, b] \cap \cap_{n\geq 1}]-\infty, a - \frac{1}{n}]^c$$

et donc

$$X^{-1}(I) = X^{-1}(]-\infty, b]) \cap \cap_{n\geq 1} X^{-1}(]-\infty, a - \frac{1}{n}])^c \in \mathscr{F}.$$

On traite les cas $I = [a, b[,]a, b[, [a, +\infty[, \ldots$ de manière similaire.
(ii) Supposons que $\{X = x\} \in \mathscr{F}$ pour tout $x \in D$ et soit $I \subset \mathbf{R}$ un intervalle. Alors, commeD est dénombrable et comme $\mathscr{F}$ est une tribu, on a

$$\{X \in I\} = \bigcap_{x \in D \cap I} \{X = x\} \in \mathscr{F}$$

et X est bien une v.a.r.

Réciproquement, supposons que X est une v.a.r. Soit $x \in D$ fixé. Alors $\{x\} = \cap_{k\in\mathbf{N}^*} I_k$ avec $I_k = [x, x + \frac{1}{k}]$. Comme $\{X \in I_k\} \in \mathscr{F}$ pour tout k et comme $\mathscr{F}$ est une tribu, on a

$$\{X = x\} = \cap_{k\in\mathbf{N}^*} \{X \in I_k\} \in \mathscr{F}. \blacksquare$$

En partant de certaines v.a.r, on peut obtenir de nouvelles en utilisant certaines opérations.

Proposition 3.1.6 ***(Propriétés de stabilité des v.a.r)*** *Soit* $(\Omega, \mathscr{F}, \mathbf{P})$ *un espace probabilisé.*
(i) Soient $X : \Omega \to \mathbf{R}$ *une v.a.r sur* Ω *et* $f : \mathbf{R} \to \mathbf{R}$ *une fonction continue; alors* $f(X) = f \circ X : \Omega \to \mathbf{R}$ *est une v.a.r sur* Ω.
(ii) Soient X *et* Y *deux v.a.r sur* Ω *et* $a \in \mathbf{R}$. *Alors* $X + Y, XY$ *et* aX *sont v.a.r sur* Ω

Démonstration (i) Nous supposons que X est une v.a.r. discrète prenant ses valeurs dans un ensemble dénombrable ou fini D et $f : \mathbf{R} \to \mathbf{R}$ quelconque et admettrons le résultat général. Alors $f(\mathrm{X})$ prend ses valeurs dans l'ensemble dénombrable (ou fini)

$$f(\mathrm{D}) = \{f(x) \ : \ x \in \mathrm{D}\}.$$

On peut donc utiliser la proposition précédente pour montrer que $f(\mathrm{X})$ est une v.a.r. Pour tout $y \in f(\mathrm{D})$, on a

$$\{f(\mathrm{X}) = y\} = \{\mathrm{X} \in f^{-1}(y)\} = \bigcup_{x \in \mathrm{D}: f(x) = y} \{\mathrm{X} = x\} \in \mathscr{F},$$

car $\{x \in \mathrm{D} \ : \ f(x) = y\}$ est au plus dénombrable (étant une partie de D) et car $\{\mathrm{X} = x\} \in \mathscr{F}$ pour tout $x \in \mathrm{D}$. .
(ii) Nous supposons que X et Y sont des v.a.r. discrètes avec ensemble de valeurs dénombrables (ou finis) A et B et admettrons le résultat général. Alors $\mathrm{X} + \mathrm{Y}$ prend ses valeurs dans l'ensemble dénombrable (ou fini)

$$\mathrm{C} = \{x + y \ : \ x \in \mathrm{A}, y \in \mathrm{B}\}.$$

Pour tout $z \in \mathrm{C}$, on a

$$\{\mathrm{X} + \mathrm{Y} = z\} = \bigcup_{x \in \mathrm{A}} \bigcup_{y \in \mathrm{B}: x + y = z} \{\mathrm{X} = x\} \cap \{\mathrm{Y} = y\} \in \mathscr{F}.$$

On montre de manière similaire que XY et aX sont des v.a.r.■

3.1.1 Loi et fonction de répartition d'une variable aléatoire

Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé et $\mathrm{X} : \Omega \to \mathbf{R}$ une v.a.r.

On rappelle que ceci signifie que $\mathrm{X}^{-1}(\mathrm{I}) \in \mathscr{F}$ pour tout intervalle $\mathrm{I} \subset \mathbf{R}$. La proposition suivante montre que ceci reste vrai si on prend pour I n'importe quelle partie dans la tribu $\mathscr{B}(\mathbf{R})$ des boréliens de $\mathbf{R}$; rappellons que $\mathscr{B}(\mathbf{R})$ est la tribu engendrée par les intervalles $\mathrm{I} \subset \mathbf{R}$ (voir Exemple 2.1.6).

Proposition 3.1.7 *Soit* X *une v.a.r. sur l'espace probabilisé* $(\Omega, \mathscr{F}, \mathbf{P})$. *Alors, pour tout* $\mathrm{A} \in \mathscr{B}(\mathbf{R})$, *on a*

$$\mathrm{X}^{-1}(\mathrm{A}) = \{\mathrm{X} \in \mathrm{A}\} \in \mathscr{F}.$$

Démonstration Soit $\mathscr{B}$ l'ensemble des parties $A \subset \mathbf{R}$ telles que

$$X^{-1}(A) = \{X \in A\} \in \mathscr{F}.$$

Montrons que $\mathscr{B}$ est une tribu. En effet, on a $\mathbf{R} \in \mathscr{B}$ car $X^{-1}(\mathbf{R}) = \Omega \in \mathscr{F}$. De plus, supposons que $A \in \mathscr{B}$; alors $X^{-1}(A) \in \mathscr{F}$ et donc $X^{-1}(A^c) = (X^{-1}(A))^c \in \mathscr{F}$. Ainsi $A^c \in B$. Finalement, soit $(A_n)_n$ une suite d'éléments de $\mathscr{B}$. Alors, pour tout n, on a $X^{-1}(A_n) \in \mathscr{F}$ et donc

$$X^{-1}(\cup_n A_n) = \cup_n X^{-1}(A_n) \in \mathscr{F}.$$

Ainsi, $\cup_n A_n \in \mathscr{B}$. Donc $\mathscr{B}$ est bien une tribu.

Par hypothèse, $\mathscr{B}$ contient tous les intervalles $I \subset \mathbf{R}$. Comme $\mathscr{B}$ est une tribu et comme $\mathscr{B}(\mathbf{R})$ est la plus petite tribu contenant les intervalles, on a $\mathscr{B}(\mathbf{R}) \subset \mathscr{B}$. Ceci signifie que $X^{-1}(A) = \{X \in A\} \in \mathscr{F}$, pour tout $A \in \mathscr{B}(\mathbf{R})$.■

Définition 3.1.8 (Loi d'une variable aléatoire) L'application $P_X : \mathscr{B}(\mathbf{R}) \to [0,1]$, définie par

$$\boxed{P_X(A) = \mathbf{P}(X \in A) = \mathbf{P}(X^{-1}(A)) \qquad \text{pour tout} \quad A \in \mathscr{B}(\mathbf{R})},$$

s'appelle la *loi de probabilité* de la v.a.r X.

Proposition 3.1.9 *Soient* $(\Omega, \mathscr{F}, \mathbf{P})$ *un espace probabilisé et* $X : \Omega \to \mathbf{R}$ *une v.a.r. Alors* P_X *est une mesure de probabilité sur* $(\mathbf{R}, \mathscr{B}(\mathbf{R}))$.

Démonstration On a

$$P_X(\mathbf{R}) = \mathbf{P}(X \in \mathbf{R}) = \mathbf{P}(\Omega) = 1.$$

Soit $A_1, A_2, \ldots$ une suite de parties disjointes dans $\mathscr{B}(\mathbf{R})$. Alors $\{X \in A_1\}, \{X \in A_2\}, \ldots$ sont des évènements disjoints dans $\mathscr{F}$ et donc

$$P_X(\cup_n A_n) = \mathbf{P}(\cup_n \{X \in A_n\}) = \sum_n \mathbf{P}(X \in A_n) = \sum_n P_X(A_n).$$

ceci montre que P_X est une mesure de probabilité sur $(\mathbf{R}, \mathscr{B}(\mathbf{R}))$.■

Remarque 3.1.10 (Loi d'une v.a.r. discrète) Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé et $X : \Omega \to \mathbf{R}$ une v.a.r. discrète, prenant ses valeurs dans un espace fini ou dénombrable $E = \{x_1, x_2, \ldots\} \subset \mathbf{R}$. Alors P_X est déterminée par la donnée des nombres

$$\boxed{p_n := \mathbf{P}(X = x_n) \qquad \text{pour tout} \quad n = 1, 2, \ldots.}$$

En effet, on a, pour tout $A \in \mathscr{B}(\mathbf{R})$,

$$P_X(A) = \mathbf{P}(X \in A) = \sum_{n : x_n \in A} p_n.$$

Caractériser la loi d'une v.a.r. discrète, c'est donc se donner les valeurs x_n de X et les coefficients $p_n = \mathbf{P}(X = x_n)$ pour tout $n = 1, 2, \ldots$. On dira, par abus de langage, que la suite des couples $(x_n, p_n)_n$ est la loi de X.

Exemple 3.1.11 (i) Soit $a \in \mathbf{R}$ fixé et $X : \Omega \to \mathbf{R}$ la v.a.r constante $X = a$. Alors P_X est la mesure de probabilité sur $\mathbf{R}$ définie par $P_X(A) = 1$ si $a \in A$ et $P_X(A) = 0$ sinon. Elle est notée δ_a et est appelé **mesure de Dirac** en a.
(ii) On dit qu'une v.a.r. discrète X suit la **loi uniforme** sur l'ensemble fini $\{x_1, \ldots, x_n\}$ de $\mathbf{R}$ lorsque $P(X = x_k) = 1/n$ pour tout $1 \leq k \leq n$.
(iii) Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé et soit $A \in \mathscr{F}$ avec $\mathbf{P}(A) = p \in]0, 1[$. Alors $X = \mathbf{1}_A$ est une v.a.r sur Ω, avec $X(\Omega) = \{0, 1\}$. Alors P_X est la mesure de probabilité sur $\mathbf{R}$ définie, pour tout iintervalle $I \subset \mathbf{R}$, par

$$P_X(I) = \begin{cases} p & \text{si } 1 \in I \text{ et } 0 \notin I \\ 1 - p & \text{si } 0 \in I \text{ et } 1 \notin I \\ 1 & \text{si } 0 \in I \text{ et } 1 \in I \\ 0 & \text{si } 0 \notin I \text{ et } 1 \notin I. \end{cases}$$

Définition 3.1.12 (Fonction de répartition) Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé et $X : \Omega \to \mathbf{R}$ une v.a.r. On appelle *fonction de répartition* de X la fonction $F_X : \mathbf{R} \to \mathbf{R}$ définie par

$$\boxed{F_X(x) = \mathbf{P}(X \leq x) = \mathbf{P}(X^{-1}(]-\infty, x])) \qquad \text{pour tout} \quad x \in \mathbf{R}.}$$

Exemple 3.1.13 (i) La fonction de répartition d'une v.a.r. constante $X = a$ est donnée par $F_X(x) = 0$ pour $x < a$ et $F_X(x) = 1$ pour $x \geq a$. Donc $F_X = \mathbf{1}_{[a, +\infty[}$.
(ii) Soit X un v.a.r. discrète suivant la loi uniforme sur l'ensemble fini $\{x_1, \ldots, x_n\} \subset \mathbf{R}$ avec $x_1 < x_2 < \cdots < x_n$. Alors F_X est constante sur les intervalles

$$]-\infty, x_1[\,,\ [x_1, x_2[\,, \ldots,\ [x_{n-1}, x_n[\,,\ [x_n, +\infty[,$$

avec les valeurs successives $0, 1/n, 2/n, \dots, (n-1)/n, 1$.
(iii) Soientt $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé, $A \in \mathscr{F}$ avec $\mathbf{P}(A) = p \in]0,1[$. et $X = \mathbf{1}_A$ comme dans l'Exemple 3.1.11.iii. Alors $F_X : \mathbf{R} \to \mathbf{R}$ est donnée par

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1-p & \text{si } 0 \le x < 1 \\ 1 & \text{si } x \ge 1. \end{cases}$$

Proposition 3.1.14 ***(Propriétés d'une fonction de répartition)*** *Soit* $(\Omega, \mathscr{F}, \mathbf{P})$ *un espace probabilisé et* $X : \Omega \to \mathbf{R}$ *une v.a.r. La fonction de répartition* F_X *de* X *possède les propriétés suivantes :*

1. F_X *est croissante sur* $\mathbf{R}$;
2. $\lim_{x\to-\infty} F_X(x) = 0$ *et* $\lim_{x\to+\infty} F_X(x) = 1$;
3. F_X *est continue à droite en tout point* $x \in \mathbf{R}$: $\lim_{t\to x, t>x} F_X(t) = F(x)$;
4. *pour tout* $x \in \mathbf{R}$, *on a* $F_X(x) - \lim_{t\to x, t<x} F_X(t) = \mathbf{P}(X = x)$.

Démonstration 1. Pour $x \le y$, on a $]-\infty, x] \subset]-\infty, y]$ et donc

$$F_X(x) = P_X(]-\infty, x]) \le P_X(]-\infty, y]) = F_X(y).$$

2. Soit $(x_n)_n$ une suite décroissante avec $\lim_n x_n = -\infty$. Alors $(]-\infty, x_n])_n$ est une suite décroissante d'intervalles et $\cap_n]-\infty, x_n] = \emptyset$. On a donc, par continuité monotone des probabilités (voir Lemme 2.1.9),

$$\lim_n F_x(x_n) = \lim_n P_X(]-\infty, x_n]) = P_X(\emptyset) = 0.$$

De même ,si $(x_n)_n$ est une suite croissante avec $\lim_n x_n = +\infty$. Alors $(]-\infty, x_n])_n$ est une suite croissante d'intervalles et $\cup_n(]-\infty, x_n])_n = \mathbf{R}$; par continuité monotone des probabilités, on a donc

$$\lim_n F_x(x_n) = \lim_n P_X((]-\infty, x_n]) = P_X(\mathbf{R}) = 1.$$

3. Soit $x \in \mathbf{R}$. Soit $(x_n)_n$ une suite décroissante avec $\lim_n x_n = x$. Alors $(]-\infty, x_n])_n$ est une suite décroissante avec $\cap_n]-\infty, x_n] =]-\infty, x]$ et donc, par continuité monotone,

$$\lim_n F_x(x_n) = \lim_n P_X(]-\infty, x_n]) = P_X(]-\infty, x]) = F_X(x).$$

4. Soit $x \in \mathbf{R}$. Soit $(x_n)_n$ une suite croissante avec $\lim_n x_n = x$. Alors $(]-\infty, x_n])_n$ est une suite croissante avec $\cap_n]-\infty, x_n] =]-\infty, x[$ et donc, par continuité monotone,

$$\lim_n \mathrm{F}_x(x_n) = \lim_n \mathrm{P_X}(]-\infty, x_n]) = \mathrm{P_X}(]-\infty, x[);$$

Comme

$$\mathrm{P_X}(]-\infty, x[= \mathrm{P_X}(]-\infty, x]) - \mathrm{P_X}(\{x\}) = \mathrm{F_X}(x) - \mathbf{P}(\mathrm{X} = x),$$

on donc bien

$$\mathrm{F_X}(x) - \lim_n \mathrm{F}_x(x_n) = \mathbf{P}(\mathrm{X} = x). \blacksquare$$

Remarque 3.1.15 (Lien entre loi et fonction de répartition) (i) Soit $\mathrm{X} : \Omega \to \mathbf{R}$ une v.a.r. sur un espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$. Il est clair que, par définition, la loi $\mathrm{P_X}$ détermine $\mathrm{F_X}$. En fait, la réciproque est également vraie : la fonction de répartition détermine $\mathrm{P_X}$. Montrons-le dans le cas d'une variable aléatoire discrète X avec un ensemble de valeurs D. En effet, d'après le 4. de la proposition précédente, on a pour tout $x \in \mathrm{D}$,

$$\boxed{\mathbf{P}(\mathrm{X} = x) = \mathrm{F_X}(x) - \lim_{t \to x, t < x} \mathrm{F_X}(t)}$$

et ceci détermine $\mathrm{P_X}$.

(ii) Soit $\mathrm{X} : \Omega \to \mathbf{R}$ une v.a.r. discrète, prenant ses valeurs dans un espace fini ou dénombrable $\{x_1, x_2, \ldots\} \subset \mathbf{R}$, avec les probabilités $p_n := \mathbf{P}(\mathrm{X} = x_n)$. Alors $\mathrm{F_X}$ est donnée par

$$\boxed{\mathrm{F_X}(x) = \sum_{n \,:\, x_n \leq x} p_n \qquad \text{pour tout} \quad x \in \mathbf{R}.}$$

3.1.2 Espérance d'une variable aléatoire discrète

Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé et $\mathrm{X} : \Omega \to \mathbf{R}$ une v.a.r. que nous supposerons **discrète** dans toute cette section. Soit D l'ensemble (fini ou dénombrable) des valeurs de X. Pour se faire une idée du comportement d'une v.a.r. X, dont on observe n valeurs successives $x_1, \ldots, x_n \in \mathrm{D}$, il est naturel de considérer la moyenne arithmétique

$$\mathrm{M}_n = \frac{1}{n}(x_1 + \cdots + x_n).$$

En regroupant selon les différents résultats possibles de l'expérience, on a

$$\mathrm{M}_n = \sum_{i \geq 1} f_n(x_i) x_i,$$

où $f_n(x_i)$ est la fréquence d'apparition de l'évènement $\{X = x_i\}$. Avec l'idée (qui reste encore à justifier) que ces fréquences $f_n(x_i)$ sont des approximations de la probabilité $p_i = P(X = x_i)$, on est amené à définir la notion d'espérance de X.

Définition 3.1.16 (Espérance d'une v.a.r) Soit $D = \{x_1, x_2, \ldots\}$ l'ensemble des valeurs de la v.a.r. discrète X. On dit que X possède une *espérance* si la série $\sum_n x_n P(X = x_n)$ est absolument convergente (c-à-d si la série $\sum_n |x_n| P(X = x_n)$ est convergente). Dans ce cas, on appelle *espérance* de X le nombre réel $\mathbb{E}(X)$ qui est somme de cette série :

$$\boxed{\mathbb{E}(X) := \sum_n x_n \mathbf{P}(X = x_n) = \sum_n x_n p_n,}$$

avec $p_n = \mathbf{P}(X = x_n)$.

Remarque 3.1.17 (i) Si l'ensemble des valeurs de X est un ensemble fini $D = \{x_1, \ldots, x_n\}$, alors $\mathbb{E}(X)$ existe et vaut

$$\mathbb{E}(X) := \sum_{k=1}^{n} x_k \mathbf{P}(X = x_k).$$

(ii) Si $X \geq 0$ (c-à-d si $x_n \geq 0$ pour toute valeur $x_n \in X(\Omega)$), alors X possède une espérance si et seulement si la série $\sum_n x_n P(X = x_n)$ est convergente et, dans ce cas, $\mathbb{E}(X)$ est la somme de cette série.

Exemple 3.1.18 (i) Soit $X = a$ une v.a.r. constante ; alors $\mathbb{E}(X) = a$.
(ii) Soit X une v.a.r de Bernoulli, c-à-d X prend les valeurs $\{0, 1\}$ avec les probabilités $\mathbf{P}(X = 1) = p$ et $\mathbf{P}(X = 0) = 1 - p$. Alors

$$\mathbb{E}(X) = 1 \times p + 0 \times (1 - p) = p.$$

(iii) Soit X une v.ar. avec $\mathbf{N}^*$ comme ensemble de valeurs et les probabilités

$$p_n = \mathbf{P}(X = n) = \frac{6}{\pi^2 n^2}.$$

(On a bien $\sum_{n\geq 1} p_n = 1$, car $\sum_{n\geq 1} \dfrac{1}{n^2} = \pi^2/6$.) La v.a.r. X n'a pas d'espérance car la série (harmonique) $\sum_{n\geq 1} n p_n = \sum_{n\geq 1} \dfrac{1}{n}$ n'est pas convergente

Voici quelques propriétés couramment utilisées de l'espérance.

Proposition 3.1.19 *Soit* $(\Omega, \mathscr{F}, \mathbf{P})$ *un espace probabilisé.*
(i) ***(Linéarité)*** *Soient* $X, Y : \Omega \to \mathbf{R}$ *deux v.a.r. discrètes et* $a \in \mathbf{R}$; *alors* $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ *et* $\mathbb{E}(aX) = a\mathbb{E}(X)$.
(ii) ***(Formule de transfert)*** *Soient* $X : \Omega \to \mathbf{R}$ *une v.a. r. discrète, avec ensemble (fini ou dénombrable) de valeurs* E *et* $f : \mathbf{R} \to \mathbf{R}$; *alors*

$$\mathbb{E}(f(X)) = \sum_{x \in E} f(x)\mathbf{P}(X = x),$$

à condition que la série $\sum_{x\in E} f(x)\mathbf{P}(X = x)$ *converge absolument.*

Démonstration (i) Soient $E = X(\Omega)$ et $F = Y(\Omega)$. On pose $Z = X + Y$. L'ensemble des valeurs de Z est

$$G = \{x + y \ : (x, y) \in E \times F.$$

Pour $z \in G$, posons

$$A_z = \{(x, y) \in E \times F \ : \ x + y = z\}.$$

Alors

$$\{Z = z\} \ = \bigcup_{(x,y)\in A_z} \{X = x, Y = y\}$$

est une partition en la famille (au plus dénombrable) des parties $\{X = x, Y = y\} \in \mathscr{F}$. On a donc

$$\boxed{\mathbf{P}(Z = z) \ = \sum_{(x,y)\in A_z} \mathbf{P}(X = x, Y = y).}$$

Montrons que la série $\sum_{z\in G} |z|\mathbf{P}(Z = z)$ converge. Notons qu'on a également la partition

$$E \times F = \cup_{z\in G} A_z$$

de $E \times F$ en une famille (au plus dénombrable) de parties. On a alors

$$\begin{aligned}
\sum_{z\in Z} |z|\mathbf{P}(Z = z) &= \sum_{z\in G} |z| \sum_{(x,y)\in A_z} \mathbf{P}(X = x, Y = y) = \sum_{z\in G} \sum_{(x,y)\in A_z} |x + y| \, \mathbf{P}(X = x, Y = y) \\
&\le \sum_{z\in G} \sum_{(x,y)\in A_z} (|x| + |y|) \, \mathbf{P}(X = x, Y = y) \\
&= \sum_{z\in G} \sum_{(x,y)\in A_z} |x| \, \mathbf{P}(X = x, Y = y) + \sum_{z\in G} \sum_{(x,y)\in A_z} |y| \, \mathbf{P}(X = x, Y = y) \\
&= \sum_{(x,y)\in E\times F} |x| \, \mathbf{P}(X = x, Y = y) + \sum_{(x,y)\in E\times F} |x| \, \mathbf{P}(X = x, Y = y) \\
&= \sum_{x\in E} |x| \, \mathbf{P}(X = x) + \sum_{y\in F} |y| \, \mathbf{P}(Y = y).
\end{aligned}$$

(Dans la dernière égalité, on a utilisé le fait que $\sum_{(x,y)\in E\times F}|x|\,\mathbf{P}(X=x,Y=y)=\sum_{x\in E}|x|\sum_{y\in F}\mathbf{P}(X=x,Y=y)=\sum_{x\in E}|x|\,\mathbf{P}(X=x)$ et de même $\sum_{(x,y)\in E\times F}|y|\,\mathbf{P}(X=x,Y=y)=\sum_{y\in F}|y|\,\mathbf{P}(Y=y)$.)

Comme ces deux dernières séries sont convergentes, il s'ensuit que $\mathbb{E}(X+Y)$ existe. Le même raisonnement montre que

$$\begin{aligned}
\mathbb{E}(X+Y)=\sum_{z\in Z} z\mathbf{P}(Z=z) &= \sum_{z\in G} z\sum_{(x,y)\in A_z}\mathbf{P}(X=x,Y=y)=\sum_{z\in G}\sum_{(x,y)\in A_z}(x+y)\mathbf{P}(X=x,Y=y)\\
&=\sum_{z\in G}\sum_{(x,y)\in A_z} x\,\mathbf{P}(X=x,Y=y)+\sum_{z\in G}\sum_{(x,y)\in A_z} y\mathbf{P}(X=x,Y=y)\\
&=\sum_{(x,y)\in E\times F} y\,\mathbf{P}(X=x,Y=y)+\sum_{(x,y)\in E\times F} y\,\mathbf{P}(X=x,Y=y)\\
&=\sum_{x\in E} x\,\mathbf{P}(X=x)+\sum_{y\in F} y\mathbf{P}(Y=y)=\mathbb{E}(X)+\mathbb{E}(Y).
\end{aligned}$$

Soit $a\in\mathbf{R}$. Alors l'ensemble des valeurs de aX est $\{ax\ :\ x\in E\}$. Si $a=0$, alors $aX=0$ et donc $\mathbb{E}(aX)=0=a\mathbb{E}(X)$. On peut donc supposer que $a\neq 0$; il est alors clair que $\{aX=ax\}=\{X=x\}$ et donc

$$\mathbb{E}(aX)=\sum_{x\in E} ax\mathbf{P}(aX=ax)=a\sum_{x\in E} x\mathbf{P}(X=x)=a\mathbb{E}(X).$$

(ii) L'ensemble des valeurs de $f(X)$ est $f(E)$. Pour chaque $y\in f(E)$, soit $E_y=f^{-1}(\{y\})=\{x\in E\ :\ f(x)=y\}$. On a $\{f(X)=y\}=\cup_{x\in E_y}\{X=x\}$, une partition en une famille au plus dénombrable de parties. D'où

$$\mathbf{P}(f(X)=y)=\sum_{x\in E_y}\mathbf{P}(X=x).$$

Alors $E=\cup_{y\in f(E)} f^{-1}(\{y\})$ est une partition et en supposant que la série

$$\sum_{x\in E} f(x)\mathbf{P}(X=x)$$

converge absolument, on a alors

$$\begin{aligned}
\mathbb{E}(f(X))=\sum_{y\in f(E)} y\mathbf{P}(f(X)=y) &= \sum_{y\in f(E)} y\sum_{x\in E_y}\mathbf{P}(X=x)\\
&\sum_{y\in f(E)}\sum_{x\in E_y} y\mathbf{P}(X=x)\\
&\sum_{y\in f(E)}\sum_{x\in E_y} f(x)\mathbf{P}(X=x)\\
&\sum_{x\in E} f(x)\mathbf{P}(X=x).\blacksquare
\end{aligned}$$

Remarque 3.1.20 Si X admet un espérance, alors la v.a.r $Y = X - \mathbb{E}(X)$ admet un espérance et $\mathbb{E}(Y) = \mathbb{E}(X) - \mathbb{E}(\mathbb{E}(X)) = 0$. Une v.a.r Y avec $\mathbb{E}(Y) = 0$ est dite *centrée.*

3.1.3 Moments d'ordre quelconque et variance

Á côté de l'espérance, qui correspond à la valeur moyenne de la v.a.r. en question, on introduit d'autres indicateurs qui tiennent compte de la dispersion de cette v.a.r. autour de la moyenne.

Définition 3.1.21 (Moments) Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé et X une v.a.r. discrère sur Ω, avec loi donnée par $(x_n, p_n)_{n\geq 1}$. Soit $k \in \mathbf{N}$. On dit que X admet un moment d'ordre k, si la v.a.r X^k possède une espérance (c-à-d si $\sum_{n\geq 1} x_n^k p_n$ est absolument convergente). Le *moment d'ordre k* de X est alors le réel $\mathbb{E}(X^k)$.

Remarque 3.1.22 Si X admet un moment d'ordre k, alors X admet un moment d'ordre k' pour tout $1 \leq k' \leq k$. En effet, comme $x^{k'} \leq x^k + 1$ pour tout $x \geq 0$ (car $x^{k'}(x^{k-k'} - 1) \geq -1$ pour $x \geq 0$, on a $|x_n|^{k'} p_n \leq |x_n|^k p_n + p_n$. La série $\sum_{n\geq 1} |x_n|^k p_n + p_n$ étant convergente, l'assertion s'ensuit.

Définition 3.1.23 (Variance) Soit X une v.a.r. discrète sur l'espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$ admettant un moment d'ordre 2. La *variance* Var(X) de X est le moment d'ordre 2 de la v.a.r. centrée $X - \mathbb{E}(X)$, c-à-d

$$\boxed{\mathrm{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2).}$$

La racine carrée $\sqrt{\mathrm{Var}(X)}$ s'appelle *écart type* de X et est noté $\sigma(X)$. (On observera que, par la remarque plus haut, si X admet un moment d'ordre 2, alors $\mathbb{E}(X)$ existe automatiquement.)

Remarque 3.1.24 Soit $X(\Omega) = \{x_1, x_2, \ldots,\}$ l'ensemble des valeurs de X. Alors Var(X) existe si et seulement si la série $\sum_n x_n^2 p_n$ converge, où $p_n = \mathbf{P}(X = x_n)$. Dans ce cas, on a, en posant $m = \mathbb{E}(X)$:

$$\mathrm{Var}(X) = \sum_n (x_n - m)^2 p_n.$$

Voici quelques propriétés de la variance.

Proposition 3.1.25 *Soit* X *une v.a.r. discrète sur l'espace probabilisé* $(\Omega, \mathscr{F}, \mathbf{P})$ *admettant un moment d'ordre* 2.
(i) Pour tout $a \in \mathbf{R}$, *on a*

$$\mathrm{Var}(a\mathrm{X}) = a^2\,\mathrm{Var}(\mathrm{X}) \quad \textit{et} \quad \mathrm{Var}(\mathrm{X}+a) = \mathrm{Var}(\mathrm{X}).$$

(ii) ***(Formule de König-Huygens)*** *On a*

$$\boxed{\mathrm{Var}(\mathrm{X}) = \mathbb{E}(\mathrm{X}^2) - \mathbb{E}(\mathrm{X})^2.}$$

Démonstration (i) Ces formules sont immédiates :

$$\mathrm{Var}(a\mathrm{X}) = \mathbb{E}((a\mathrm{X} - \mathbb{E}(a\mathrm{X}))^2) = \mathbb{E}(a^2(\mathrm{X} - \mathbb{E}(\mathrm{X}))^2) = a^2\mathbb{E}((\mathrm{X} - \mathbb{E}(\mathrm{X}))^2) = a^2\,\mathrm{Var}(\mathrm{X})$$

et

$$\mathrm{Var}(\mathrm{X}+a) = \mathbb{E}((\mathrm{X}+a-\mathbb{E}(\mathrm{X}+a))^2) = \mathbb{E}((\mathrm{X}-\mathbb{E}(\mathrm{X}))^2) = \mathrm{Var}(\mathrm{X}).$$

(ii) On a

$$\begin{aligned}\mathrm{Var}(\mathrm{X}) &= \mathbb{E}((\mathrm{X}-\mathbb{E}(\mathrm{X}))^2) = \mathbb{E}\left(\mathrm{X}^2 - 2\mathrm{X}\mathbb{E}(\mathrm{X}) + \mathbb{E}(\mathrm{X})^2\right)\\ &= \mathbb{E}(\mathrm{X}^2) - 2\mathbb{E}(\mathrm{X})\mathbb{E}(\mathrm{X}) + \mathbb{E}(\mathrm{X})^2\\ &= \mathbb{E}(\mathrm{X}^2) - \mathbb{E}(\mathrm{X})^2.\blacksquare\end{aligned}$$

Il est parfois utile de savoir que, si $\mathrm{Var}(\mathrm{X}) = 0$, alors X est essentiellement constante.

Proposition 3.1.26 *Soit* X *une v.a.r. discrète sur l'espace probabilisé* $(\Omega, \mathscr{F}, \mathbf{P})$, *admettant un moment d'ordre* 2. *Alors* $\mathrm{Var}(\mathrm{X}) = 0$ *si et seulement s'il existe* $a \in \mathbf{R}$ *tel que* $\mathbf{P}(\mathrm{X} = a) = 1$.

Démonstration il est clair que, si $\mathbf{P}(\mathrm{X} = a) = 1$, alors $\mathbb{E}(\mathrm{X}) = a$ et $\mathbf{P}((\mathrm{X}-\mathbb{E}(\mathrm{X}))^2 = 0) = 1$ et donc $\mathrm{Var}(\mathrm{X}) = 0$.

Réciproquement, si $\mathrm{Var}(\mathrm{X}) = 0$, alors, en notant $\mathrm{D} = \mathrm{X}(\Omega)$ et $a = \mathbb{E}(\mathrm{X})$, on a

$$0 = \mathrm{Var}(\mathrm{X}) = \sum_{x \in \mathrm{D}} (x-a)^2\mathbf{P}(\mathrm{X} = x)$$

et donc $(x-a)^2\mathbf{P}(\mathrm{X} = x) = 0$ pour tout $x \in \mathrm{D}$, car $(x-a)^2\mathbf{P}(\mathrm{X} = x) \geq 0$. Comme $\Omega = \cup_{x\in\mathrm{D}}\{\mathrm{X} = x\}$ est une partition, on a $a \in \mathrm{D}$ et $\mathbf{P}(\mathrm{X} = a) = 1.\blacksquare$

3.2 Variables aléatoires discrètes classiques

Nous donnons un “catalogue" des principales lois discrètes.

3.2.1 Loi uniforme

Nous l'avons déjà rencontrée plus haut. On tire une boule au hasard dans une urne contient n boules numérotées de 1 à n; le numéro tiré X est une v.a.r discrète avec les valeurs $1,\dots,n$, de probabilité uniforme $p_k = \mathbf{P}(X = k) = 1/n$.

Définition 3.2.1 (Loi uniforme). Soit $E = \{x_1, x_2, \dots, x_n\}$ une partie finie à n éléments de $\mathbf{R}$. On dit qu'une v.a.r. X suit la *loi uniforme* sur E si l'ensemble des valeurs de X est E et si $p_k = \mathbf{P}(X = x_k) = 1/n$ pour tout $k \in [\![1, n]\!]$. On écrit $X \sim \mathscr{U}(E)$.

Supposons que $E = [\![1, n]\!] = \{1,\dots,n\}$ et que $X \sim \mathscr{U}([\![1, n]\!])$. Alors

$$\mathbb{E}(X) = \sum_{k=1}^{n} k\frac{1}{n} = \frac{n+1}{2}, \qquad \mathbb{E}(X^2) = \sum_{k=1}^{n} k^2\frac{1}{n} = \frac{(n+1)(2n+1)}{6}$$

et, avec la formule de König-Huygens,

$$\operatorname{Var}(X) = \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{n^2-1}{12}.$$

3.2.2 Loi de Bernoulli

Nous l'avons déjà rencontrée plus haut; c'est la loi de réalisation ou non d'un évènement (comme “obtenir Six" dans le lancer d'un dé).

Définition 3.2.2 (Loi de Bernoulli). On dit qu'une v.a.r. X suit une *loi de Bernoulli* de paramètre $p \in]0,1[$ si l'ensemble des valeurs de X est $\{0,1\}$ et si

$$\mathbf{P}(X = 1) = p$$

et donc $\mathbf{P}(X = 0) = 1 - p$. On écrit $X \sim \mathscr{B}(p)$.

Alors, comme vu plus haut,

$$\boxed{\mathbb{E}(X) = p.}$$

On a

$$\mathbb{E}(X^2) = 1^2\mathbb{E}(X = 1) + 0^2\mathbb{E}(X = 0) = p,$$

et, avec la formule de König-Huygens,

$$\boxed{\operatorname{Var}(X) = pq, \text{ avec } q = 1 - p.}$$

3.2.3 Loi binomiale

Soient $p \in]0,1[$ et $n \geq 1$. On lance une pièce donnant pile avec probabilité p et face avec probabilité $q = 1-p$. On considère la v.a.r. X égale au nombre de piles après n lancers indépendants.

L'ensemble des valeurs possibles pour X est $[\![0,n]\!] = \{0,1,\dots,n\}$. Soit $k \in [\![0,n]\!]$ fixé. Alors $\{X = k\}$ est réalisé quand exactement k lancers donnent pile. Il y a $\binom{n}{k}$ façons de choisir les rangs de ces k lancers. Pour un choix fixé de ces rangs, la probabilité de cette séquence est $p^k q^{n-k}$, par indépendance des lancers. On a donc

$$p_k := \mathbf{P}(X = k) = \binom{n}{k} p^k q^{n-k}.$$

Définition 3.2.3 (Loi binomiale). On dit qu'une v.a.r. X suit une *loi binomiale* de paramètres (n,p) avec $n \in \mathbf{N}$ et $p \in]0,1[$, si l'ensemble des valeurs de X est $\{0,1,\dots,n\}$ et si, pour tout $k \in [\![0,n]\!]$, on a

$$\mathbf{P}(X = k) = \binom{n}{k} p^k q^{n-k} \quad \text{avec } q = 1-p.$$

On écrit $X \sim \mathscr{B}(n,p)$.

Calculons l'espérance et la variance d'une telle loi.

Proposition 3.2.4 ***(Espérance et variance d'une loi binomiale)*** *Soit* X *une v.a.r suivant une loi* $\mathscr{B}(n,p)$. *Alors on a*
(i) $\boxed{\mathbb{E}(X) = np;}$
(i) $\boxed{\mathrm{Var}(X) = npq, \text{ avec } q = 1-p.}$

Démonstration (i) On a

$$\mathbb{E}(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} = \sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k}.$$

Comme, pour $1 \leq k \leq n$, on a la relation

$$k\binom{n}{k} = k\frac{n(n-1)\dots n-k+1}{k!} = n\frac{(n-1)\dots(n-1)-(k-1)+1}{(k-1)!} = n\binom{n-1}{k-1},$$

il s'ensuit que

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{k=1}^{n} n\binom{n-1}{k-1} p^k q^{n-k} \\
&= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} q^{(n-1)-(k-1)} \\
&= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{(n-1)-k} \\
&= np(p+q)^{n-1} = np.
\end{aligned}
$$

(ii) On calcule d'abord $\mathbb{E}(X(X-1)) = \sum_{k=2}^{n} k(k-1)\binom{n}{k} p^k q^{n-k}$: en utilisant deux fois la relation $k\binom{n}{k} = n\binom{n-1}{k-1}$, on a

$$
\begin{aligned}
\mathbb{E}(X(X-1)) &= \sum_{k=2}^{n} k(k-1)\binom{n}{k} p^k q^{n-k} \\
&= \sum_{k=2}^{n} n(k-1)\binom{n-1}{k-1} p^k q^{n-k} \\
&= \sum_{k=2}^{n} n(n-1)\binom{n-2}{k-2} p^k q^{n-k} \\
&= n(n-1)p^2 \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2} q^{n-k} \\
&= n(n-1)p^2 \sum_{k=0}^{n-2} \binom{n-2}{k} p^k q^{(n-2)-k)} \\
&= n(n-1)p^2(p+q)^{n-2} = n(n-1)p^2.
\end{aligned}
$$

Comme

$$\mathrm{Var}(X) = \mathbb{E}(X(X-1)) + \mathbb{E}(X) - \mathbb{E}(X)^2,$$

il s'ensuit que

$$\mathrm{Var}(X) = n(n-1)p^2 + np - n^2p^2 = np(1-p) = npq. \blacksquare$$

3.2.4 Loi hypergéométrique

Pour $a, b \geq 1$, considérons une urne avec N boules, dont a sont blanches et $b = \mathrm{N} - a$ noires. On tire $n \leq \mathrm{N}$ boules simultanément et on considère la v.a.r. X donnant le nombre de boules blanches tirées. La valeur minimale de X est 0 si $n \leq b$ et $n - b$ si $n > b$. La valeur maximale de X est n si $n \leq a$ et a si $n > a$. L'ensemble des valeurs possibles pour X est donc $[\![\max\{0, n-b\}, \min\{n, a\}]\!]$. Pour $k \in [\![0, n[\![$ fixé, on a

$$p_k := \mathbf{P}(\mathrm{X} = k) = \frac{\binom{a}{k}\binom{b}{n-k}}{\binom{\mathrm{N}}{n}}.$$

Définition 3.2.5 (Loi hypergéométrique). On dit qu'une v.a.r. X suit une *loi hypergéométrique* de paramètres (N, n, p) avec $1 \leq n \leq \mathrm{N}$ et $p \in]0, 1[$, si l'ensemble de ses valeurs est $\mathrm{E} = [\![\max\{0, n - \mathrm{N}q\}, \min\{n, \mathrm{N}p\}]\!]$ avec $q = 1 - p$ et et si, pour tout $k \in \mathrm{E}$, on a

$$\mathbf{P}(\mathrm{X} = k) = \frac{\binom{\mathrm{N}p}{k}\binom{\mathrm{N}q}{n-k}}{\binom{\mathrm{N}}{n}}.$$

On écrit $\mathrm{X} \sim \mathscr{H}(\mathrm{N}, n, p)$.

Proposition 3.2.6 ***(Espérance et variance d'une loi hypergéométrique)*** *Soit* X *une v.a.r suivant une loi* $\mathscr{H}(\mathrm{N}, n, p)$. *Alors on a*

(i) $\boxed{\mathbb{E}(\mathrm{X}) = np;}$

(i) $\boxed{\mathrm{Var}(\mathrm{X}) = npq\frac{\mathrm{N}-n}{\mathrm{N}-1}, \text{ avec } q = 1 - p.}$

Démonstration

Remarquons d'abord la formule (dite *formule de Vandermonde*)

$$\sum_{k=0}^{n} \binom{a}{k}\binom{b}{n-k} = \binom{a+b}{n},$$

valable pour tous $a, b, n \in \mathbf{N}$, avec $n \leq a + b = \mathrm{N}$. Cette formule est claire : le membre de droite est le nombre d'échantillons de n boules choisies dans un ensemble de N boules. On suppose que cet ensemble de N boules est composé de a boules blanches et de b boules noires. Alors un échantillon de n boules choisies parmi ces N boules est déterminé de manière unique par un échantillon de $k \leq n$ boules blanches choisies parmi les a boules blanches et par un échantillon de $n-k$ boules noires choisies parmi les b boules noires. Le membre de gauche est égal au nombre de choix de ces deux échantillons.

Posons $a = \mathrm{N}p$ et $b = \mathrm{N}q$, de sorte que $a+b=\mathrm{N}$.

(i) On a

$$\mathbb{E}(\mathrm{X}) = \sum_{k=1}^{n} k \frac{\binom{a}{k}\binom{b}{n-k}}{\binom{\mathrm{N}}{n}}.$$

En utilisant la relation

$$k\binom{n}{k} = n\binom{n-1}{k-1},$$

il s'ensuit que

$$\begin{aligned}\mathbb{E}(\mathrm{X}) &= a\sum_{k=1}^{n} \frac{\binom{a-1}{k-1}\binom{b}{n-k}}{\binom{\mathrm{N}}{n}}\\ &= a\sum_{k=0}^{n-1} \frac{\binom{a-1}{k}\binom{b}{(n-1)-k}}{\binom{\mathrm{N}}{n}}\end{aligned}$$

En utilisant la formule de Vandermonde, on obtient

$$\begin{aligned}\mathbb{E}(\mathrm{X}) &= a\frac{\binom{a+b-1}{n-1}}{\binom{\mathrm{N}}{n}}\\ &= a\frac{(\mathrm{N}-1)!}{(n-1)!(\mathrm{N}-n)!}\frac{n!(\mathrm{N}-n)!}{\mathrm{N}!} = \mathrm{N}p\frac{n}{\mathrm{N}}\\ &= np.\end{aligned}$$

(ii) On calcule d'abord

$$\mathbb{E}(\mathrm{X}(\mathrm{X}-1)) = \sum_{k=2}^{n} k(k-1)\frac{\binom{a}{k}\binom{b}{n-k}}{\binom{\mathrm{N}}{n}}.$$

En utilisant deux fois la relation $k\binom{n}{k} = n\binom{n-1}{k-1}$, ainsi que la formule de Vandermonde, on a

$$\begin{aligned}
\mathbb{E}(X(X-1)) &= a(a-1)\sum_{k=2}^{n}\frac{\binom{a-2}{k-2}\binom{b}{n-k}}{\binom{N}{n}}\\
&= a(a-1)\sum_{k=0}^{n-2}\frac{\binom{a-2}{k}\binom{b}{(n-2)-k}}{\binom{N}{n}}\\
&= a(a-1)\frac{\binom{a+b-2}{n-2}}{\binom{N}{n}}\\
&= a(a-1)\frac{(N-2)!}{(n-2)!(N-n)!}\frac{n!(N-n)!}{N!} = a(a-1)\frac{n(n-1)}{N(N-1)}.
\end{aligned}$$

Il s'ensuit que

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}(X(X-1)) + \mathbb{E}(X) - \mathbb{E}(X)^2\\
&= Np(Np-1)\frac{n(n-1)}{N(N-1)} + np - (np)^2\\
&= np\left(\frac{(Np-1)(n-1)}{N-1} + 1 - np\right)\\
&= np\left(\frac{N+np-nNp}{N-1}\right) = np\left(\frac{(1-p)(N-n)}{N-1}\right)\\
&= npq\frac{N-n}{N-1}.\blacksquare
\end{aligned}$$

3.2.5 Loi géométrique

Une pièce de monnaie amène “Pile" avec probabilité $p \in]0,1[$ et “Face" avec probabilité $q = 1-p$. On effectue des lancers successifs jusqu'à l'apparition de “Pile". Soit x le nombre de lancers effectués. L'ensemble des valeurs possibles pour X est donc $[\![1,+\infty[\![$ et, pour $k \in [\![1,+\infty[\![$, on a

$$p_k := \mathbf{P}(X=k) = q^{k-1}p.$$

Définition 3.2.7 (Loi géométrique). On dit qu'une v.a.r. X suit une *loi géométrique* de paramètre $p \in]0,1[$, si l'ensemble de ses valeurs est $[\![1,+\infty[\![$ et si pour tout $k \in [\![1,+\infty[\![$, on a

$$p_k := \mathbf{P}(X=k) = q^{k-1}p.$$

On écrit $X \sim \mathscr{G}(p)$.

Proposition 3.2.8 ***(Espérance et variance d'une loi géométrique)*** *Soit* X *une v.a.r suivant une loi* $\mathcal{G}(p)$. *Alors on a*

(i) $\boxed{\mathbb{E}(X) = \frac{1}{p};}$

(i) $\boxed{\mathrm{Var}(X) = \frac{q}{p^2},\ \text{avec } q = 1-p.}$

Démonstration Le calcul repose sur la considération de la fonction

$$f : x \mapsto \frac{1}{1-x} = \sum_{k\geq 0} x^k$$

qui est définie sur $]-1,1[$; cette fonction est indéfiniment dérivable sur $]-1,1[$. Sa dérivée est donnée par

$$f'(x) = \frac{1}{(1-x)^2} = \sum_{k\geq 1} kx^{k-1}$$

et sa dérivée seconde par

$$f''(x) = \frac{2}{(1-x)^3} = \sum_{k\geq 2} k(k-1)x^{k-2}$$

(les séries étant convergentes pour $|x| < 1$). On en déduit que, pour $|x| < 1$, on a

$$\sum_{k\geq 1} k^2 x^{k-1} = \sum_{k\geq 1} (k(k-1)+k)x^{k-1} = \frac{1+x}{(1-x)^3}.$$

En appliquant ce qui précède à $x = q$ on voit que $\mathbb{E}(X)$ existe (car $|q| < 1$) et que

$$\mathbb{E}(X) = p\sum_{k\geq 1} kq^{k-1} = p\frac{1}{(1-q)^2} = p\frac{1}{p^2} = \frac{1}{p}.$$

ce qui précède montre également que $\mathbb{E}(X^2)$ existe et que

$$\mathbb{E}(X^2) = p\sum_{k\geq 1} k^2 q^{k-1} = p\frac{1+q}{(1-q)^3} = \frac{2-p}{p^2}.$$

Avec la formule de König-Huygens, on conclut que

$$\mathrm{Var}(X) = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}.\blacksquare$$

3.2.6 Loi de Poisson

La loi de Poisson ne correspond pas à une situation-type comme les lois précédentes ; elle modélise souvent des situations où on s'intéresse au nombre d'apparitions d'un évènement rare (c-à-d de petite probabilité). Elle sert aussi comme loi-limite (comme approximation de la loi binomiale, par exemple : voir Proposition 4.5.1 plus loin).

Définition 3.2.9 (Loi de Poisson). On dit qu'une v.a.r. X suit une *loi Poisson* de paramètre $\lambda \in \mathbf{R}, \lambda > 0$ si l'ensemble de ses valeurs est $\mathbf{N}$ et si pour tout $k \in \mathbf{N}$, on a

$$p_k := \mathbf{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

On écrit $X \sim \mathscr{P}(\lambda)$.

On observera que, comme $e^{\lambda} = \sum_{k=0}^{\infty} \dfrac{\lambda^k}{k!}$, on a

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

de sorte que X est bien une v.a.r.

Proposition 3.2.10 ***(Espérance et variance d'une loi géométrique)*** *Soit* X *une v.a.r suivant une loi* $\mathscr{P}(\lambda)$. *Alors on a*
(i) $\boxed{\mathbb{E}(X) = \lambda;}$
(i) $\boxed{\mathrm{Var}(X) = \lambda.}$ ■

Démonstration (i) On a

$$\begin{aligned}\mathbb{E}(X) &= e^{-\lambda}\sum_{k=0}^{\infty} k\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty} k\frac{\lambda^k}{k!}\\ &= e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}\\ &= \lambda e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{(k)!} = \lambda e^{-\lambda}e^{\lambda} = \lambda.\end{aligned}$$

(ii) On calcule d'abord $\mathbb{E}(X(X-1))$:

$$\begin{aligned}\mathbb{E}(X(X-1)) &= e^{-\lambda}\sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=2}^{\infty} k(k-1)\frac{\lambda^k}{k!}\\ &= e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^k}{(k-2)!} = \lambda^2 e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!}\\ &= \lambda^2 e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{(k)!} = \lambda^2 e^{-\lambda}e^{\lambda} = \lambda^2.\end{aligned}$$

D'où

$$\mathbb{E}(X^2) = \mathbb{E}(X(X-1)) + \mathbb{E}(X) = \lambda^2 + \lambda$$

et donc

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda. \blacksquare$$

3.3 Vecteurs aléatoires discrets

Une expérience aléatoire fait souvent intervenir non pas une mais plusieurs v.a.r. Par exemple, on considère une urne contenant n boule numérotées de 1 à n. On en tire deux successivement avec remise et on s'interesse au plus petit numéro obtenu X ainsi qu'au plus grand Y. ces deux variables prennent leurs valeurs dans $\{1,\ldots,n\}$ et on s'intéresse aux évènements du type

$$\{X = i, Y_j\} := \{X = i\} \cap \{Y = j\}.$$

Définition 3.3.1 (Vecteur aléatoire) Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé. Soit $n \geq 1$. On appelle *vecteur aléatoire* de dimension n toute application

$$V : \Omega \to \mathbf{R}^n, \omega \mapsto (X_1(\omega), \ldots, X_n(\omega)),$$

où $X_1, \ldots, X_n$ sont des v.a.r. On note en général V par $V = (X_1, \ldots, X_n)$.

Quand $X_1, \ldots, X_n$ sont des v.a.r discrètes, on dit que V est *vecteur aléatoire discret*. Dans le cas $n = 2$, on parlera de *couple aléatoire discret*, et on le notera souvent $V = (X, Y)$.

Définition 3.3.2 (Loi d'un couple aléatoire) Soit $V = (X, Y)$ un couple aléatoire discret; notons E, F les ensembles des valeurs de X et de Y. Les *lois marginales* du couple V sont les lois P_X et P_Y de X et Y.

La *loi conjointe* du couple V est la famille de nombres réels $(p_{x,y})_{(x,y)\in E\times F}$ définie par

$$\boxed{p_{x,y} = \mathbf{P}(X = x, Y = y) := \mathbf{P}(\{X = y\} \cap \{Y = y\}).}$$

Exemple 3.3.3 Dans l'exemple cité en introduction, l'ensemble des valeurs de X et Y est $[\![1, n]\!]$ Soit $(i, j) \in [\![1, n]\!]^2$ et $p_{ij} = \mathbf{P}(X = i, Y = j)$. Envisageons plusieurs cas.

1e cas $i > j$: il est clair que $p_{ij} = 0$ car $X \leq Y$.

2e cas $i = j$: on a $p_{ii} = \dfrac{1}{n^2}$, par indépendance des deux tirages.

3e cas $i < j$: on a $p_{ij} = 2\dfrac{1}{n^2}$, car on doit obtenir (i, j) ou (j, i).

Remarque 3.3.4 Supposons que les v.a.r X et Y prennent leurs valeurs dans des ensembles finis $\{x_1, \ldots, x_n\}$ et $\{y_1, \ldots, y_m\}$. Il est alors pratique de présenter la loi du couple (X, Y) sous forme d'un tableau à double entrée; à l'intersection de la ligne i (correspondant à x_i) et de la colonne j (correspondant à y_j), on fait figurer la probabilité $p_{ij} = \mathbf{P}(X = x_i, Y = y_j)$:

X/Y	y_1	y_2	$\ldots$	y_m	P_X
x_1	p_{11}	p_{12}	$\ldots$	p_{1m}	$p_{1,.}$
x_2	p_{21}	p_{22}	$\ldots$	p_{2m}	$p_{2,.}$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_n	p_{n1}	p_{n2}	$\ldots$	p_{nm}	$p_{n,.}$
P_Y	$p_{.,1}$	$p_{.,2}$	$\ldots$	$p_{.m}$	1

On observera que la somme de probabilités des lignes (colonnes) permet de trouver la loi de X (respectivement de Y) :

$$p_{i,.} := \mathbf{P}(X = x_i) = \sum_j \mathbf{P}(X = i, Y = j) = \sum_j p_{ij}$$

et

$$p_{.,j} := \mathbf{P}(Y = y_j) = \sum_i \mathbf{P}(X = i, Y = j) = \sum_i p_{ij}.$$

On introduit la notion de v.a.r. indépendantes.

Définition 3.3.5 (Variables aléatoires indépendantes) Soient $X_1,\dots,X_n$ des v.a.r discrètes sur un espace probabilisé $(\Omega,\mathscr{F},\mathbf{P})$, avec ensemble de valeurs $D_1,\dots,D_n$. On dit que $X_1,\dots,X_n$ sont *(mutuellement) indépendantes* si, pour tout $(x_1,\dots,x_n)\in D_1\times\cdots\times D_n$, on a

$$\boxed{\mathbf{P}(X_1=x_1,X_2=x_2,\dots,X_n=x_n)=\mathbf{P}(X_1=x_1)\mathbf{P}(X_2=x_2)\dots\mathbf{P}(X_n=x_n).}$$

Remarque 3.3.6 Soit (X,Y) un couple de v.a.r prenant leurs valeurs dans des ensembles finis $\{x_1,\dots,x_n\}$ et $\{y_1,\dots,y_m\}$. Soient

$$p_{ij}=\mathbf{P}(X=x_i,Y=y_j),\qquad ,p_{i,.}=\mathbf{P}(X=x_i),\qquad p_{.,j}=\mathbf{P}(Y=y_j).$$

Alors, X et Y sont indépendants si et seulement si

$$\boxed{p_{ij}=p_{i,.}p_{.,j}\qquad \text{pour tout}\quad 1\le i,j\le n.}$$

Exemple 3.3.7 (i) On considère un tirage successif de deux boules, avec remise, dans une urne contenant des boules numérotées de 1 à n. Soient X_1 et X_2 les v.a.r. égales aux numéros de la 1ère et de la 2ème boules tirées. On a, pour tous $i,j\in\{1,\dots,n\}$,

$$\mathbf{P}(X_1=i,X_2=j)=\frac{1}{n}\times\frac{1}{n}=\mathbf{P}(X_1=i)\mathbf{P}(X_2=j).$$

Les v.a.r. X_1 et X_2 sont donc indépendantes.
(ii) Dans l'exemple précédent, considérons maintenant deux tirages successifs sans remise. Alors, pour les mêmes v.a.r. X_1 et X_2, on a $\mathbf{P}(X_1=i,X_2=i)=0$, pour tout $i\in\{1,\dots,n\}$. D'autre part, on a $\mathbf{P}(X_1=i)=\dfrac{1}{n}$ ainsi que

$$\begin{aligned}\mathbf{P}(X_2=i)&=\sum_{k=1}^{n}\mathbf{P}(X_2=i|X_1=k)\mathbf{P}(X_1=k)=\sum_{k\neq i}\mathbf{P}(X_2=i|X_1=k)\mathbf{P}(X_1=k)\\&=\sum_{k\neq i}\frac{1}{n-1}\times\frac{1}{n}=\frac{1}{n}.\end{aligned}$$

D'où

$$\mathbf{P}(X_1=i)\mathbf{P}(X_2=i)=\frac{1}{n^2}\neq\mathbf{P}(X_1=i,X_2=i)$$

et les v.a.r. X_1 et X_2 ne sont pas indépendantes.

La formule transfert suivante généralise la formule de transfert de la Proposition 3.4.6 et se démontre de manière similaire.

Proposition 3.3.8 *Soient* X *et* Y *deux v.a.r. discrètes, avec ensembles de valeurs* $E = X(\Omega)$ *et* $F = Y(\Omega)$. *Soit* $f : E \times F \to \mathbf{R}$ *une fonction.*
(i) La fonction

$$f(X,Y) : \Omega \to \mathbf{R}, \quad \omega \mapsto f(X(\omega), Y(\omega))$$

est une v.a.r.
(ii) ***(Formule de transfert généralisée)*** *Supposons que la série*

$$\sum_{(x,y)\in E\times F} |f(x,y)|\mathbf{P}(X = x, Y = y)$$

est convergente. Alors $f(X,Y)$ *possède une espérance et*

$$\boxed{\mathbb{E}(f(X,Y)) = \sum_{(x,y)\in E\times F} f(x,y)\mathbf{P}(X = x, Y = y).}$$

Démonstration On pose $Z = f(X,Y)$. L'ensemble des valeurs de Z est

$$G = \{f(x,y) \ : (x,y) \in E \times F\}.$$

Pour $z \in G$, posons

$$A_z = \{(x,y) \in E \times F \ : \ z = f(x,y)\} = f^{-1}(\{z\}).$$

(i) On a une partition

$$\{Z = z\} = \bigcup_{(x,y)\in A_z} \{X = x, Y = y\}$$

et donc $\{Z = z\} \in \mathscr{F}$ car $\{X = x, Y = y\} \in \mathscr{F}$ pour tout (x,y) et car A_z est dénombrable.

Notons qu'on a également la partition

$$E \times F = \cup_{z\in G} A_z$$

(ii) Pour $z \in G$, on a, avec la relation plus haut,

$$z\mathbf{P}(Z = z) = z \sum_{(x,y)\in A_z} \mathbf{P}(X = x, Y = y)$$

et donc

$$\begin{aligned}
\sum_{z\in G} z\mathbf{P}(Z = z) &= \sum_{z\in G} z \sum_{(x,y)\in A_z} \mathbf{P}(X = x, Y = y) \\
&= \sum_{z\in G} \sum_{(x,y)\in A_z} f(x,y)\mathbf{P}(X = x, Y = y) \\
&= \sum_{(x,y)\in E\times F} f(x,y)\mathbf{P}(X = x, Y = y). \blacksquare
\end{aligned}$$

Corollaire 3.3.9 *Soient* X *et* Y *deux v.a.r. , possédant chacune une variance. Alors*
(i) la v.a.r. XY *possède une espérance;*
(ii) si X *et* Y *sont* indépendantes, *on a*

$$\boxed{\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).}$$

Démonstration Soient E et F les ensembles de valeurs de X et Y. Soit $f : \mathbf{R} \times \mathbf{R} \to \mathbf{R}, (x, y) \mapsto xy$, de sorte que $f(X, Y) = XY$.
(ii) Montrons que la série $\sum_{(x,y)\in E\times F} |xy|\mathbf{P}(X = x, Y = y)$ est convergente, Comme $ab \leq \frac{1}{2}(a^2 + b^2)$ pour tous $a, b \in \mathbf{R}^*$, on a

$$|xy|\mathbf{P}(X = x, Y = y) = |x||y|\mathbf{P}(X = x, Y = y) \leq \frac{1}{2}(x^2 + y^2)\mathbf{P}(X = x, Y = y).$$

et donc

$$\begin{aligned}
\sum_{(x,y)\in E\times F} |xy|\mathbf{P}(X = x, Y = y) &\leq \sum_{(x,y)\in E\times F} \frac{1}{2}(x^2 + y^2)\mathbf{P}(X = x, Y = y)\\
&= \frac{1}{2}\left(\sum_{(x,y)\in E\times F} x^2\mathbf{P}(X = x, Y = y) + \sum_{(x,y)\in E\times F} y^2\mathbf{P}(X = x, Y = y)\right)\\
&= \frac{1}{2}\left(\sum_{x\in E} x^2\mathbf{P}(X = x) + \sum_{y\in F} y^2\mathbf{P}(Y = y)\right)\\
&= \frac{1}{2}(\mathrm{Var}(X) + \mathrm{Var}(Y)) < \infty.
\end{aligned}$$

Donc XY possède une espérance.
(ii) Supposons que X et Y sont indépendantes. Avec la proposition précédente, on a alors

$$\begin{aligned}
\mathbb{E}(XY) &= \sum_{(x,y)\in E\times F} xy\mathbf{P}(X = x, Y = y)\\
&= \sum_{(x,y)\in E\times F} xy\mathbf{P}(X = x)\mathbf{P}(Y = y)\\
&\sum_{x\in E} x\mathbf{P}(X = x) \sum_{y\in F} y\mathbf{P}(Y = y)\\
&= \mathbb{E}(X)\mathbb{E}(Y). \blacksquare
\end{aligned}$$

3.3.1 Lois conditionnelles

Soient X et Y deux v.a.r; sur l'espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$, avec $X(Omega) = \{x_1, x_2, \ldots\}$ et $Y(\Omega) = \{y_1, y_2, \ldots\}$.

Définition 3.3.10 (Lois conditionnelles) Soit $x_i \in X(\Omega)$ fixé avec $\mathbf{P}(X = x_i) > 0$. La *loi conditionnelle de* Y *sachant l'évènement* $\{X = x_i\}$ est la mesure probabilité $\mathbf{P}^{x_i}$ sur $Y(\Omega)$ définie par

$$\mathbf{P}(\{y_j\}) = \mathbf{P}(Y = y_j | X = x_i) = \frac{\mathbf{P}(Y = y_j, X = x_i)}{\mathbf{P}(X = x_i)}.$$

On définit de même les lois conditionnelles de X sachant $\{Y = y_j\}$.

On note $X|_{(Y=y_j)}$ la v.a.r. obtenue à partir de X en conditionnant par l'évènement $\{Y = y_j\}$.

Exemple 3.3.11 (i) Considérons, comme dans l'exemple 3.3.7.ii, un tirage successif de deux boules dans une urne contenant des boules numérotées de 1 à n. Soient X et Y les v.a.r. égales aux numéros de la 1ère et de la 2ème boules tirées. Supposons que le tirage soit sans remise. Soient $i, j \in \{1, \ldots, n\}$. On a $\mathbf{P}(X = i | Y = j) = 0$ si $i = j$ et $\mathbf{P}(X = i | Y = j) = \dfrac{1}{n-1}$ sinon. La loi de $X|_{(Y=j)}$ est donc la loi uniforme sur $\{1, \ldots, n\} \setminus \{j\}$.

Supposons que le tirage soit avec remise. Pour tous $i, j \in \{1, \ldots, n\}$, on a $\mathbf{P}(X = i | Y = j) = \dfrac{1}{n}$. La loi de $X|_{(Y=j)}$ est donc la loi uniforme sur $\{1, \ldots, n\}$, qui est également la loi de X.

(ii) On considère une urne contenant a boules blanches et b boules noires et on procède à un tirage successif de deux boules, sans remise. Soient X_i, pour $i = 1, 2$, la v.a.r. égale à 1 si la i-ème boule est blanche et 0 sinon.

Cherchons la loi de $X_2|_{(X_1=1)}$: on a, en posant $N = a + b$,

$$\mathbf{P}(X_2 = 1 | X_1 = 1) = \frac{a-1}{N-1}, \qquad \mathbf{P}(X_2 = 0 | X_1 = 1) = \frac{b}{N-1},$$

de sorte que la loi $X_2|_{(X_1=1)}$ est une loi de Bernoulli $\mathscr{B}(\dfrac{a-1}{N-1})$.

Cherchons la loi de $X_1|_{(X_2=0)}$: avec la formule de Bayes, on a

$$\mathbf{P}(X_1 = 1 | X_2 = 0) = \frac{\mathbf{P}(X_2 = 0 | X_1 = 1)\mathbf{P}(X_1 = 1)}{\mathbf{P}(X_2 = 0)} = \frac{b}{N-1} \times \frac{a}{N} \times \frac{N}{b} = \frac{a}{N-1}$$

de sorte que la loi $X_1|_{(X_2=0)}$ est une loi de Bernoulli $\mathscr{B}(\dfrac{a}{N-1})$.

Exprimons la notion d'indépendance au moyen des lois conditionnelles.

Proposition 3.3.12 *Soient* X *et* Y *deux v.a.r. discrètes sur l'espace de probabilité* $(\Omega, \mathscr{F}, \mathbf{P})$ *. Le propriétés suivantes sont équivalentes :*

(i) X *et* Y *sont indépendantes;*

(ii) Pour tout $y \in Y(\Omega)$ *avec* $\mathbf{P}(Y = y) > 0$*, la loi conditionnelle de* X *sachant* $\{Y = y\}$ *est la même que la loi de* X.

(iii) Pour tout $x \in X(\Omega)$ *avec* $\mathbf{P}(X = x) > 0$*, la loi conditionnelle de* Y *sachant* $\{X = x\}$ *est la même que la loi de* Y.

Démonstration Il suffit de montrer que (i) et (ii) sont équivalents (l'équivalence de (i) et (iii) s'en déduit, les rôles de X et Y par rapport à l'indépendance étant symétriques).

Supposons que X et Y sont indépendantes et soit $y \in Y(\Omega)$ avec $\mathbf{P}(Y = y) > 0$. Alors, pour tout $x \in X(\Omega)$, on a

$$\mathbf{P}(X = x|Y = y) = \frac{\mathbf{P}(X = x, Y = y)}{\mathbf{P}(Y = y)} = \frac{\mathbf{P}(X = x)\mathbf{P}(Y = y)}{\mathbf{P}(Y = y)} = \mathbf{P}(X = x).$$

Ceci montre que la loi de $X|_{(Y=y)}$ est la même que la loi de X.

Réciproquement, supposons que (ii) est vraie. Soient $x \in X(\Omega)$ et $y \in Y(\Omega)$. Si $\mathbf{P}(Y = y) = 0$, alors $\mathbf{P}(X = x, Y = y) = 0$ (car $\{X = x, Y = y\} \subset \{Y = y\}$) et donc $\mathbf{P}(X = x, Y = y) = \mathbf{P}(X = x)\mathbf{P}(Y = y)$. Si $\mathbf{P}(Y = y) > 0$, alors $\mathbf{P}(X = x|Y = y) = \mathbf{P}(X = x)$ et donc

$$\mathbf{P}(X = x, Y = y) = \mathbf{P}(X = x|Y = y)\mathbf{P}(Y = y) = \mathbf{P}(X = x)\mathbf{P}(Y = y).$$

Ceci montre que X et Y sont indépendantes ■

3.3.2 Covariance de deux variables aléatoires

Soit (X, Y) un couple de v.a.r. discrètes sur un espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$. Jusqu'à présent, les indicateurs que avons introduits (espérance, variance, ...) concernaient chaque v.a.r. séparément. Nous alors définir maintenant un indicateur faisant intervenir la loi conjointe du couple (X, Y) et qui mesurera le degré de dépendance de X et Y.

Si X et Y possédent chacune un moment d'ordre 2, alors elles possèdent chacunes un espérances et les v.a. centrées $X^* = X - \mathbb{E}(X)$ et $Y^* = Y - \mathbb{E}(Y)$ associées possèdent des moments d'ordre 2 également. Rappelons que ceci implique que la v.a. r. produit X^*Y^* possède une espérance (voir Corolaire 3.3.9).

Définition 3.3.13 Soient X et Y deux v.a.r. discrètes, possédant chacune un moment d'ordre 2. On appelle *covariance* de X et Y,, notée Cov(X, Y), le nombre

$$\boxed{\mathrm{Cov}(X,Y) := \mathbb{E}(X^*Y^*) = \mathbb{E}\left((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\right).}$$

Avec $X(\Omega) = \{x_1, x_2, \ldots\}$, $Y(\Omega) = \{y_1, y_2, \ldots\}$, $p_{ij} = \mathbf{P}(X = x_i, Y = y_j)$ et $m_X = \mathbb{E}(X)$, $m_Y = \mathbb{E}(Y)$, on a (en utilisant la formule du transfert)

$$\mathrm{Cov}(X,Y) = \sum_{i,j}(x_i - m_X)(y_j - m_Y)p_{ij}.$$

Les v.a.r. sont dites *non corrélées* si $\mathrm{Cov}(X, Y) = 0$.

Voici quelques propriétés de la covariance.

Proposition 3.3.14 *Soient* X, Y, Z *des v.a.r sur* $(\Omega, \mathscr{F}, \mathbf{P})$, *possédant chacune un moment d'ordre* 2. *Alors*

(i) $\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$;

(ii) ***(Bilinéarité de la covariance)*** $\mathrm{Cov}(sX + tY, Z) = s\,\mathrm{Cov}(X,Z) + t\,\mathrm{Cov}(Y,Z)$ *pour tous* $s, t \in \mathbf{R}$;

(iii) ***(Symétrie de la covariance)*** $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$;

(iv) $\mathrm{Cov}(X,X) = \mathrm{Var}(X) \geq 0$;

(v) ***(Inégalité de Cauchy-Schwarz)*** $|\mathrm{Cov}(X,Y)| \leq \sigma(X)\sigma(Y)$, *où* $\sigma(X), \sigma(Y)$ *sont les écarts-types de* X *et* Y;

(vi) *lorsque* $\sigma(X) \neq 0$, *on a* $|\mathrm{Cov}(X,Y)| = \sigma(X)\sigma(Y)$ *si et seulement* Y *est une fonction affine de* X : *il existe* $\alpha, \beta \in \mathbf{R}$ *tels que* $\mathbf{P}(Y = \alpha X + \beta) = 1$.

Démonstration (i) On a, en développant et en utilisant la linéarité de l'espérance

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathbb{E}\left((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\right) = \mathbb{E}\left(XY - X\mathbb{E}(Y) - \mathbb{E}(X)Y + \mathbb{E}(X)\mathbb{E}(Y)\right)\\ &= \mathbb{E}(XY) - \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).\end{aligned}$$

Les propriétés (ii), (iii), (iv) sont aisément vérifiées.
(v) et (vi) Posons $a := \mathrm{Var}(X)$, $b := \mathrm{Cov}(X,Y)$ et $c := \mathrm{Var}(Y)$; on veut montrer que $b^2 \leq ac$. Pour tout $t \in \mathbf{R}$, on a

$$\mathrm{Cov}(Y + tX, Y + tX) = \mathrm{Var}(Y + tX) \geq 0.$$

D'autre part, on a

$$\mathrm{Cov}(Y + tX, Y + tX) = \mathrm{Var}(Y) + 2t\,\mathrm{Cov}(X,Y) + t^2\,\mathrm{Var}(X) = c + 2bt + at^2.$$

Le polynôme $P : t \mapsto at^2 + 2bt + c$ de degré ≤ 2 est à coefficients réels et prend ses valeurs dans $\mathbf{R}_+$; son discriminant $b^2 - ac$ est donc ≤ 0.

Supposons que $\sigma(X) \neq 0$, c-à-d $a \neq 0$. Le cas d'égalité $b^2 = ac$ a lieu si et seulement si P possède $t_0 = -b/a$ comme une unique racine. Donc $P(t) = \mathrm{Var}(Y + t_0 X) = 0$. Ceci implique (voir Proposition 3.1.26) qu'il existe $\beta \in \mathbf{R}$ telque $\mathbf{P}(Y + t_0 X = \beta) = 1$, c-à-d $\mathbf{P}(Y = -t_0 X + \beta) = 1$.■

Corollaire 3.3.15 *Soient* $X_1, \dots, X_n$ *des v.a.r sur* $(\Omega, \mathscr{F}, \mathbf{P})$, *possédant chacune un moment d'ordre* 2.
(i) Alors $\mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2\sum_{1\leq i<j\leq n} \mathrm{Cov}(X_i, X_j)$;
(ii) Si les v.a.r. $X_1, \dots, X_n$ *sont non corrélées deux-à-deux, alors*

$$\mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^n \mathrm{Var}(X_i).$$

(iii) Si $X_1, \dots, X_n$ *sont indépendantes deux-à-deux, alors elles sont non corrélées deux-à-deux.*

Démonstration

(i) En utilisant la bilinéarité et la symétrie de la covariance (Proposition 3.3.14.ii et iii), on a

$$\begin{aligned}
\mathrm{Var}(X_1 + \cdots + X_n) &= \mathrm{Cov}(X_1 + \cdots + X_n, X_1 + \cdots + X_n) \\
&= \sum_{1\leq i,j\leq n} \mathrm{Cov}(X_i, X_j) = \sum_{i=1}^n \mathrm{Cov}(X_i, X_i) + \sum_{1\leq i\neq j\leq n} \mathrm{Cov}(X_i, X_j) \\
&= \sum_{i=1}^n \mathrm{Cov}(X_i, X_i) + 2 \sum_{1\leq i<j\leq n} \mathrm{Cov}(X_i, X_j)
\end{aligned}$$

(ii) découle de (i).
(iii) On a $\mathbb{E}(X_i X_j) = \mathbb{E}(X_i)\mathbb{E}(X_j)$ quand X_i et X_j sont indépendantes (voir Corollaire 3.3.9). L'assertion découle donc de (i) de la Proposition 3.3.14.■

Remarque 3.3.16 Il n'est pas vrai que des v.a.r. non corrélées sont nécessairement indépendantes. Soit par exemple $\Omega = \{-1, 0, 1\}$, muni de la probabilité uniforme; soit X la v.a.r. sur Ω définie par $X(\omega) = \omega$ pour tout ω. Soit $Y = |X|$. Alors $\mathbb{E}(X) = 0, \mathbb{E}(Y) = 2/3$ et $\mathbb{E}(XY) = 0$. Par conséquent X et Y sont non corrélées. Cependant, elles ne sont manifestement pas indépendantes (car, par exemple $\mathbf{P}(X = 0, Y = 0) = \mathbf{P}(X = 0) = 1/3$, mais $\mathbf{P}(X = 0)\mathbf{P}(Y = 0) = (1/3)^2 = 1/9$).

Exemple 3.3.17 On considère une urne contenant a boules blanches et b boules noires et on procède à un tirage successif de deux boules, sans remise. Soient X_i, pour $i = 1, 2$, la v.a.r. égale à 1 si la i-ème boule est blanche et 0 sinon (voir Exemple 3.3.11). Calculons

$$\mathrm{Cov}(X_1, X_2) = \mathbb{E}(X_1X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2).$$

On a, par la formule de transfert (Proposition 3.3.8),

$$\mathbb{E}(X_1X_2) = \sum_{(\varepsilon_1,\varepsilon_2)\in\{0,1\}^2} \varepsilon_1\varepsilon_2 \mathbf{P}(X_1 = \varepsilon_1, X_2 = \varepsilon_2).$$

En posant $N = a + b$, on calcule que

$$\begin{aligned}\mathbf{P}(X_1 = 1, X_2 = 1) &= \frac{a(a-1)}{N(N-1)}, \mathbf{P}(X_1 = 1, X_2 = 0) = \frac{ab}{N(N-1)}\\ \mathbf{P}(X_1 = 0, X_2 = 1) &= \frac{ab}{N(N-1)}, \mathbf{P}(X_1 = 0, X_2 = 0) = \frac{b(b-1)}{N(N-1)}.\end{aligned}$$

et donc $\mathbb{E}(X_1X_2) = \dfrac{a(a-1)}{N(N-1)}$. D'autre part, on a

$$\mathbb{E}(X_1) = \mathbb{E}(X_2) = \frac{a}{N}.$$

D'où

$$\mathrm{Cov}(X_1, X_2) = \frac{a(a-1)}{N(N-1)} - \frac{a^2}{N^2} = -\frac{a(N-a)}{N^2(N-1)}.$$

3.4 Variables aléatoires à densité

Jusqu'à présent, nous avons étudié des v.a.r. dont l'ensemble des valeurs était fini ou dénombrable infini. Nous allons maintenant nous intéresser à des v.a.r. dont l'ensemble des valeurs est non dénombrable, par exemple un intervalle borné ou non borné de $\mathbf{R}$, et dont les fonctions de répartition sont continues. Une telle v.a.r. décrit, par exemple, la durée de vie d'un atome radioactif, d'un individu ou d'une ampoule électrique, ...

3.4.1 Variables aléatoires à densité : définitions

Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé et $X : \Omega \to \mathbf{R}$ une v.a.r. Rappelons (voir Definition 3.1.12) que la fonction de répartition $F_X : \mathbf{R} \to \mathbf{R}$ de X est définie par $F_X(x) = \mathbf{P}(X \leq x)$.

Définition 3.4.1 (Variable aléatoire à densité) On dit que X est une *variable aléatoire à densité* (v.a.r à densité) ou *variable aléatoire continue* (v.a.r continue) si sa fonction de répartition F_X s'écrit

$$F_X(x) = \int_{-\infty}^{x} f(t)\,dt \qquad \text{pour tout} \quad x \in \mathbf{R}$$

pour une fonction $f : \mathbf{R} \to \mathbf{R}$, appelée *densité de* X, avec les propriétés suivantes :

1. $f \geq 0$;
2. f est continue sur $\mathbf{R}$, sauf éventuellement en un nombre fini de points ;
3. $\int_{-\infty}^{+\infty} f(x)\,dx = 1$.

Remarque 3.4.2 Soit X une v.a.r. continue, avec densité f.

(i) La densité f d'une v.a.r. continue n'est pas unique : par exemple, si $g = f$ sur $\mathbf{R}$, à l'exception d'un nombre fini de points, alors

$$\int_{-\infty}^{x} g(t)\,dt = \int_{-\infty}^{x} f(t)\,dt = F_X(x) \qquad \text{pour tout} \quad x \in \mathbf{R}.$$

(ii) Pour tous $a \leq b$, on a

$$\mathbf{P}(a < X \leq b) = \int_{a}^{b} f(t)\,dt.$$

En effet, on a

$$\mathbf{P}(a < \mathrm{X} \leq b) = \mathrm{F_X}(b) - \mathrm{F}_x(a) = \int_{-\infty}^{b} f(t)dt - \int_{-\infty}^{a} f(t)dt = \int_{a}^{b} f(t)dt.$$

(ii) La fonction de répartition $\mathrm{F_X}$ de X est continue sur $\mathbf{R}$; en effet, nous savons déjà que, pour X une v.a.r quelconque, $\mathrm{F_X}$ est continue à droite, et possède une limite à gauche en tout point (voir Proposition 3.1.14). Quand X possède une densité f, alors, pour tout $x \in \mathbf{R}$ et toute suite croissante $(x_n)_n$ avec $\lim_n x_n = x$, on a

$$\lim_n \mathrm{F_X}(x) - \mathrm{F_X}(x_n) = \lim_n \int_{x_n}^{x} f(t)dt = 0.$$

(iii) Pour tout $x \in \mathbf{R}$, on a $\mathbf{P}(\mathrm{X} = x) = 0$. en effet, avec (x_n) comme en (ii), on a $\{x\} = \bigcap_n]x_n, x]$ et donc, par monotonie décroissante (voir Lemme 2.1.9)

$$\mathbf{P}(\mathrm{X} = x) = \lim_n \mathbf{P}(]x_n, x]) = \lim_n \mathrm{F_X}(x) - \mathrm{F_X}(x_n) = 0.$$

(iv) On déduit de (i) et (iii) que, pour tous $a \leq b$, on a

$$\mathbf{P}(a < \mathrm{X} < b) = \mathbf{P}(a \leq \mathrm{X} < b) = \mathbf{P}(a \leq \mathrm{X} \leq b) = \mathbf{P}(a < \mathrm{X} \leq b) = \int_{a}^{b} f(t)dt.$$

(v) Pour tout $a, b \in \mathbf{R}$ avec $a > 0$, la v.a.r $a\mathrm{X} + b$ possède une densité donnée par

$$x \mapsto \frac{1}{a} f\left(\frac{x-b}{a}\right).$$

En effet, on a, pour tout $x \in \mathbf{R}$, par le changement de variable $s = at + b$:

$$\begin{aligned}\mathbf{P}(a\mathrm{X} + b \leq x) &= \mathbf{P}(\mathrm{X} \leq \frac{x-b}{a}) = \int_{-\infty}^{\frac{x-b}{a}} f(t)dt \\ &= \frac{1}{a}\int_{-\infty}^{x} f\left(\frac{s-b}{a}\right) ds.\end{aligned}$$

3.4.2 Moments d'une variable aléatoire à densité

Soit X une v.a.r. continue sur l'espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$, avec densité f.

Définition 3.4.3 (Espérance et moments d'une v.a.r. à densité) On dit que X admet une espérance si l'intégrale $\int_{-\infty}^{+\infty} |x| f(x)dx$ existe ; dans ce cas, on définit son *espérance*, notée $\mathbb{E}(\mathrm{X})$, par

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x)\,dx.$$

On dit que X admet un moment d'ordre $k \geq 1$ si l'intégrale $\int_{-\infty}^{+\infty} |x|^k f(x)\,dx$ existe ; dans ce cas, on définit son *moment d'ordre k*, noté m_k, par

$$m_k = \mathbb{E}(X^k) = \int_{-\infty}^{+\infty} x^k f(x)\,dx.$$

Remarque 3.4.4 Si X admet un moment d'ordre $k \geq 1$, alors X admet tout moment d'ordre $k' \leq k$. En effet, ceci découle du fait (voir Remarque 3.1.22) que $|x|^{k'} \leq |x|^k + 1$ pour tout $x \in \mathbf{R}$.

Définition 3.4.5 (Variance d'une v.a.r à densité) Supposons que X admet un espérance. On appelle *moment centré d'ordre* $k \geq 1$ de X, noté μ_k, le moment d'ordre k de la v.a.r. centrée $X - \mathbb{E}(X)$:

$$\mu_k = \mathbb{E}\Big((X - \mathbb{E}(X))^k\Big) = \int_{-\infty}^{+\infty} (x - \mathbb{E}(X))^k f(x)\,dx,$$

sous réserve de l'existence de l'intégrale.

La *variance* de X, notée Var(X) est μ_2, le moment centrée d'ordre 2 :

$$\mathrm{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \int_{-\infty}^{+\infty} (x - \mathbb{E}(X))^2 f(x)\,dx,$$

sous réserve de l'existence de l'intégrale. On appelle $\sigma(X) = \sqrt{\mathrm{Var}(X)}$ *écart-type* de X.

Voici quelques propriétés et formules couramment utilisées concernant l'espérance et la variance.

Proposition 3.4.6 *Soit* $(\Omega, \mathscr{F}, \mathbf{P})$ *un espace probabilisé.*
(i) ***(Linéarité)*** *Soient* $X, Y : \Omega \to \mathbf{R}$ *deux v.a.r. à densité et* $a \in \mathbf{R}$; *alors* $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ *et* $\mathbb{E}(aX) = a\mathbb{E}(X)$.
(ii) Soit $X : \Omega \to \mathbf{R}$ *une v.a.r. à densite admettant un moment d'ordre 2. Pour tout* $a, b \in \mathbf{R}$, *on a* $\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$.
(iii) ***(Formule de König-Huyghens)*** *Soit* $X : \Omega \to \mathbf{R}$ *une v.a.r. à densite admettant un moment d'ordre 2. Alors*

$$\boxed{\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.}$$

Démonstration (i) Nous admettrons la preuve de la linéarite de l'espérance qui dépasse le cadre du cours.
(ii) Cette formule découle directement de la définition de la variance et des propriétés de linéarité de l'intégrale :

$$\begin{aligned}\mathrm{Var}(a\mathrm{X}+b) &= \int_{-\infty}^{+\infty}(ax+b-\mathbb{E}(a\mathrm{X}+b))^2 f(x)\,dx = \int_{-\infty}^{+\infty}(a(x-\mathbb{E}(\mathrm{X})))^2 f(x)\,dx \\ &= a^2\int_{-\infty}^{+\infty}(x-\mathbb{E}(\mathrm{X}))^2 f(x)\,dx = a^2\,\mathrm{Var}(\mathrm{X}).\end{aligned}$$

(ii) La formule de de König-Huygens se démontre comme dans le cas de v.a.r. discrètes (voir Proposition 3.1.25) :

$$\begin{aligned}\mathrm{Var}(\mathrm{X}) &= \mathbb{E}((\mathrm{X}-\mathbb{E}(\mathrm{X}))^2) = \mathbb{E}\left(\mathrm{X}^2-2\mathrm{X}\mathbb{E}(\mathrm{X})+\mathbb{E}(\mathrm{X})^2\right) \\ &= \mathbb{E}(\mathrm{X}^2)-2\mathbb{E}(\mathrm{X})\mathbb{E}(\mathrm{X})+\mathbb{E}(\mathrm{X})^2 \\ &= \mathbb{E}(\mathrm{X}^2)-\mathbb{E}(\mathrm{X})^2. \blacksquare\end{aligned}$$

3.5 Variables aléatoires à densité classiques

Voici une liste des principales lois de probabilité à densité.

3.5.1 Loi uniforme

Soit X une v.a.r sur l'espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$.

Définition 3.5.1 (Loi uniforme sur un intervalle) On dit que X suit une *loi uniforme* sur l'intervalle $[a,b]$ de $\mathbf{R}$, si X admet comme densité la fonction f définie par

$$\boxed{f(x) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(x) \qquad \text{pour tout} \quad x \in \mathbf{R}.}$$

On dit alors que X est de loi $\mathscr{U}([a,b])$ et on écrit souvent $\mathrm{X} \sim \mathscr{U}([a,b])$.

Remarque 3.5.2 On remarquera qu'on a bien $f \geq 0$ et $\int_{-\infty}^{+\infty} f(x)\,dx = 1$.

Proposition 3.5.3 *Soit* X *de loi* $\mathscr{U}([a,b])$.

(i) La fonction de répartition de X *est*

$$\boxed{F_X(x) = \begin{cases} 0 & si\ x < a \\ \dfrac{x-a}{b-a} & si\ a \le x \le b \\ 1 & si\ x > b. \end{cases}}$$

(ii) $\boxed{\mathbb{E}(X) = \dfrac{a+b}{2}.}$

(ii) $\boxed{\mathrm{Var}(X) = \dfrac{(b-a)^2}{12}.}$

Démonstration (i) On a

$$F_X(x) = \int_{-\infty}^{x} f(t)dt = \frac{1}{b-a}\int_{-\infty}^{x} \mathbf{1}_{[a,b]}(t)dt$$

et donc $F_X(x) = 0$ si $x \le a$, $F_X(x) = \dfrac{1}{b-a}\int_a^x dt = \dfrac{x-a}{b-a}$ si $x \in [a,b]$ et $F_X(x) = \dfrac{1}{b-a}\int_a^b dt = 1$ si $x \ge b$.

(ii) On a

$$\begin{aligned}\mathbb{E}(X) &= \int_{-\infty}^{\infty} xf(x)dx = \frac{1}{b-a}\int_a^b xdx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b \\ &= \frac{a+b}{2}.\end{aligned}$$

et

(iii) On a

$$\begin{aligned}\mathbb{E}(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)dx = \frac{1}{b-a}\int_a^b x^2dx = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b \\ &= \frac{a^2+ab+b^2}{3},\end{aligned}$$

et donc

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}.\blacksquare$$

Exemple 3.5.4 Les bus passent à un arrêt donné à 7h, 7h15 et 7h30. Un usager se présente entre 7h et 7h30 à cet arrêt, l'heure exacte de son arrivée étant une variable uniforme sur cette période. Cherchons la probabilité qu'il doive attendre moins de 5 minutes, puis plus de 10 minutes.

Soit X le nombre de minutes s'écoulant entre 7h et l'arrivée de l'usager. Alors $X \sim \mathscr{U}([0,30])$. L'attente est inférieure à 5 mns si l'usager arrive entre 7h10 et 7h15 ou entre 7h25 et 7h30. La probabilité d'attendre moins de 5 minutes est donc

$$\mathbf{P}(10 < X < 15) + \mathbf{P}(25 < X < 30) = 2\frac{5}{30} = \frac{1}{3} = 0,33...$$

De même, la probabilité d'attendre plus de 10 minutes vaut

$$\mathbf{P}(0 < X < 5) + \mathbf{P}(15 < X < 20) = \frac{1}{3} = 0,33...$$

3.5.2 Loi exponentielle

Soit X une v.a.r sur l'espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$.

Définition 3.5.5 (Loi exponentielle) On dit que X suit une *loi exponentielle* de paramètre $\lambda > 0$, si X admet comme densité la fonction f définie par

$$\boxed{f(x) = \lambda e^{-\lambda x}\mathbf{1}_{[0,+\infty[}(x) \qquad \text{pour tout} \quad x \in \mathbf{R}.}$$

On dit alors que X est de loi $\mathscr{E}(\lambda)$ et on écrit souvent $X \sim \mathscr{E}(\lambda)$.

Remarque 3.5.6 On remarquera qu'on a bien $f \geq 0$ et

$$\int_{-\infty}^{+\infty} f(x)dx = \lambda \int_0^{+\infty} e^{-\lambda t}dt = -\left[e^{-\lambda t}\right]_0^{+\infty} = 1.$$

Proposition 3.5.7 *Soit* X *de loi* $\mathscr{E}(\lambda)$. *Alors :*
(i) la fonction de répartition de X *est*

$$\boxed{F_X(x) = (1 - e^{-\lambda x})\mathbf{1}_{[0,+\infty[}(x) \qquad \textit{pour tout} \quad x \in \mathbf{R}.}$$

(ii) $\boxed{\mathbb{E}(X) = \frac{1}{\lambda}.}$

(ii) $\boxed{\mathrm{Var}(X) = \frac{1}{\lambda^2}.}$

Démonstration (i) On a

$$F_X(x) = \int_{-\infty}^{x} f(t)dt = \lambda \int_{-\infty}^{x} e^{-\lambda t}\mathbf{1}_{[0,+\infty[}(t)dt$$

et donc $F_X(x) = 0$ si $x \leq 0$, $F_X(x) = \lambda \int_0^x e^{-\lambda t} dt = \left[-e^{-\lambda t}\right]_0^x = 1 - e^{-\lambda x}$ si $x \geq 0$.
(ii) On a, en intégrant par parties,

$$\begin{aligned}\mathbb{E}(X) &= \lambda \int_0^{\infty} xe^{-\lambda x}dx = -\left[-xe^{-\lambda x}\right]_0^{\infty} + \int_0^{\infty} e^{-\lambda x}dx \\ &= -\frac{1}{\lambda}\left[e^{-\lambda x}\right]_0^{\infty} = \frac{1}{\lambda}.\end{aligned}$$

(iii) On a, en intégrant par parties,

$$\begin{aligned}\mathbb{E}(X^2) &= \lambda^2 \int_0^{\infty} x^2 e^{-\lambda x}dx = -\lambda\left[-x^2 e^{-\lambda x}\right]_0^{\infty} + 2\lambda\int_0^{\infty} xe^{-\lambda x}dx \\ &= \frac{2}{\lambda^2}\end{aligned}$$

et donc

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{1}{\lambda^2}.\blacksquare$$

Remarque 3.5.8 Dans la pratique, la loi exponentielle modélise souvent une durée de vie ou le temps d'attente avant l'arrivée d'un événement spécifique. Par exemple la durée de vie d'une bactérie, la durée d'une conversation téléphonique ou l'instant de désintégration d'un noyau radioactif peuvent être considérées comme des variables aléatoires de loi exponentielle.

Exemple 3.5.9 En moyenne, un atome d'uranium 235 se désintègre au bout d'un milliard d'années. Supposons que sa durée de vie X, mesurée en années, soit une variable aléatoire exponentielle de paramètre λ; on est donc amené à prendre $\lambda = \frac{1}{10^9}$. La période T de l'uranium 235 est le laps de temps nécessaire à la désintégration de la moitié d'un échantillon donné de cette substance; T est donc solution de l'équation $\mathbf{P}(X > T) = 1/2$, c-à-d

$$\int_T^{+\infty} \lambda e^{-\lambda x}dx = e^{-\lambda T} = \frac{1}{2}$$

et donc

$$T = \frac{\log 2}{\lambda} \sim 10^9 \times 0,7 = 700 \text{ millions d'années.}$$

3.5.3 Loi normale

Soit X une v.a.r sur l'espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$. Nous allons introduire la loi la plus importante de la théorie des probabilités, la loi normale de Gauss ou loi de Laplace-Gauss.

Définition 3.5.10 (Loi normale ou loi de Gauss) On dit que X suit une *loi normale* ou *loi de Gauss-Laplace* de paramètres $m \in \mathbf{R}, \sigma > 0$, si X admet comme densité la fonction $f_{m,\sigma}$, définie par

$$\boxed{f_{m,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \qquad \text{pour tout} \quad x \in \mathbf{R}.}$$

On dit alors que X est de loi normale $\mathcal{N}(m, \sigma)$ et on écrit souvent $X \sim \mathcal{N}(m, \sigma)$.

La loi $\mathcal{N}(0,1)$ est appelée *loi normale centrée-réduite*; sa densité est

$$\boxed{f_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad \text{pour tout} \quad x \in \mathbf{R}.}$$

Remarque 3.5.11 On remarquera qu'on a bien $f_{m,\sigma} \geq 0$; montrons d'abord que $\int_{-\infty}^{+\infty} f_{0,1}(x)dx = 1$. Il suffit de montrer que $\int_{-\infty}^{+\infty} e^{-x^2}\, dx = \sqrt{\pi}$.

Pour cela, on considère la fonction $\varphi : (x, y) \mapsto e^{-(x^2+y^2)}$ définie sur $\mathbf{R}^2$. On a, d'une part,

$$\int_{\mathbf{R}^2} \varphi(x, y)\, dx\, dy = \int_{\mathbf{R}} e^{-x^2} dx \int_{\mathbf{R}} e^{-y^2} dy = (\int_{-\infty}^{+\infty} e^{-x^2}\, dx)^2.$$

D'autre part, en passant aux coordonnées polaires et avec le changement de variable $\rho = r^2$, on a

$$\begin{aligned}\int_{\mathbf{R}^2} \varphi(x, y)\, dx\, dy &= \int_{r=0}^{+\infty} \int_{\theta=0}^{2\pi} r e^{-r^2}\, dr\, d\theta \\ &= \frac{1}{2} \int_{\rho=0}^{+\infty} \int_{\theta=0}^{2\pi} e^{-\rho}\, d\rho\, d\theta = \pi \int_{\rho=0}^{+\infty} e^{-\rho}\, d\rho \\ &= \pi.\end{aligned}$$

L'assertion s'ensuit.

Soient maintenant m et $\sigma > 0$ quelconques. Alors, avec le changement de variables $y = \dfrac{x-m}{\sigma}$, on a

$$\int_{-\infty}^{+\infty} f_{m,\sigma}(x)\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}}\, dy = 1.$$

Proposition 3.5.12 *Soit* X *de loi* $\mathcal{N}(m,\sigma)$. *Alors :*
(i) la v.a.r. $Y = \dfrac{X-m}{\sigma}$ *est de loi* $\mathcal{N}(0,1)$.
(ii) $\boxed{\mathbb{E}(X) = m.}$
(ii) $\boxed{\mathrm{Var}(X) = \sigma^2.}$

En particulier, si X *suit la loi normale* $\mathcal{N}(0,1)$, X *est centrée et réduite :* $\mathbb{E}(X) = 0$ *et* $\mathrm{Var}(X) = 1$.

Démonstration Pour $x \in \mathbf{R}$, on a, avec le changement de variable $s = \dfrac{t-m}{\sigma}$,

$$\begin{aligned} F_Y(x) &= \mathbf{P}(Y \le x) = \mathbf{P}(X \le \sigma x + m) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\sigma x + m} e^{-\frac{(t-m)^2}{2\sigma^2}} dt \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{s^2}{2}} ds \\ &= \int_{-\infty}^{x} f_{0,1}(s) ds. \end{aligned}$$

Donc $Y \sim \mathcal{N}(0,1)$.

Par les formules de Proposition 3.4.6, il suffit donc de calculer $\mathbb{E}(X)$ et $\mathrm{Var}(X)$ dans le cas $m = 0, \sigma = 1$.

Tout d'abord, $\mathbb{E}(X)$ existe, car l'intégrale $\int_{-\infty}^{+\infty} |x| e^{-\frac{x^2}{2}} dx$ existe, car

$$|x| e^{-x^2/2} = O(1/x^2) \qquad \text{pour } x \to \pm\infty.$$

D'autre part, $x \mapsto x e^{-x^2/2} dx$ étant impaire, on a alors

$$\int_{-\infty}^{+\infty} x e^{-\frac{x^2}{2}} dx = 0.$$

et donc $\mathbb{E}(X) = 0$.

Comme $x^2 e^{-\dfrac{x^2}{2}} = O(1/x^2)$ pour $x \to \pm\infty$, l'intégrale $\int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} dx$ existe. On a alors, par parité et en intégrant par parties (en remarquant que $\dfrac{d}{dx}(e^{-x^2/2}) =$

$-xe^{-x^2/2}$) :

$$\begin{aligned}\int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} dx &= 2\int_0^{+\infty} x^2 e^{-\frac{x^2}{2}} dx \\ &= 2\left[-xe^{-\dfrac{x^2}{2}}\right]_0^{\infty} + 2\int_0^{\infty} e^{-\frac{x^2}{2}} dx \\ &= \sqrt{2\pi}.\end{aligned}$$

D'où

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}(X^2) = 1.$$

Remarque 3.5.13 (i) La fonction de répartition $\Phi_{0,1}$ de la loi normale centrée-réduite des donnée sous forme intégrale :

$$\Pi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$$

Cette intégrale n'est pas exprimable en termes de fonctions usuelles; ses valeurs sont tabulées dans une table, appelée *table de la loi normale* (voir la table en Annexe, Chapitre 8; ces valeurs sont également disponibles dans certaines calculatrices). Ces tables permettent de déterminer des seuils à des probabilités fixées. On a, par exemple,

$$\begin{aligned}\Pi(1.24) &= \mathbf{P}(X \le 1.24) = 0.8925 \\ 1 - \Pi(1.24) &= \mathbf{P}(X > 1.24) = 0.1185 \\ \Pi(1.96) &= \mathbf{P}(X \le 1.96) = 0.975 \\ 1 - \Pi(1.96) &= \mathbf{P}(X > 1.96) = 0.025 \\ \Pi(2.58) &= \mathbf{P}(X < 2.58) = 0.995\end{aligned}$$

(ii) Par parité, on a, pour tout $x \in \mathbf{R}$, $\Pi(-x) = 1 - \Pi(x)$ et donc, si $X \sim \mathcal{N}(0,1)$ et $x \ge 0$,

$$\boxed{\mathbf{P}(|X| \le x) = 2\Pi(x) - 1},$$

car $\mathbf{P}(|X| \le x) = \mathbf{P}(X \le x) - (\mathbf{P}(X \le -x)) = \Pi(x) - \Pi(-x) = 2\Pi(x) - 1$. Ainsi,

$$\mathbf{P}(|X| \le 1.96) = 2\Pi(1.96) - 1 = 1.950 - 1 = 0.95,$$

c-à-d avec une probabilité de 95%, X prend ses valeurs dans l'intervalle $[-1.96, 1.96]$.

De même,

$$\mathbf{P}(|X| \le 2.58) = 2\Pi(2.58) - 1 = 1.99 - 1 = 0.99,$$

c-à-d avec une probabilité de 99%, X prend ses valeurs dans l'intervalle $[-2.58, 2.58]$.

3.6 Exercices

Exercice 3.6.1 Un joueur jette simultanément deux dés. A l'issue du jeu, il gagne une somme X égale à la différence entre le plus grand et le plus petit des points marqués.
(i) Déterminer la loi de la v.a.r X ainsi que sa fonction de répartition F_X. Tracer le graphe de F_X.
(ii) Calculer l?espérance et la variance de X.

Exercice 3.6.2 On lance un dé 20 fois. Soit V la v.a.r égale au nombre de 5 obtenus. Déterminer la loi de V . Quelle est la probabilité d'obtenir un 5 trois fois ?

Exercice 3.6.3 On dispose de trois urnes et de trois boules. On place chacune des boules au hasard dans l'une des urnes. Soit X la v.a.r égale au nombre d'urnes qui ne sont pas vides. Déterminer la loi de X et sa fonction de répartition.

Exercice 3.6.4 Soit $n \geq 1$ un entier. Une urne contient n boules blanches et n boules rouges. On tire simultanément n boules. Soit X la v.a.r égale au nombre de boules rouges obtenues. Déterminer la loi de X, son espérance et sa variance.

Exercice 3.6.5 Soit $N \geq 1$ un entier. Une urne contient N boules numérotées de 1 à N. On effectue $n \geq 1$ tirages successifs avec remise. Soit X la v.a.r égale au plus grand des numéros obtenus.

Déterminer la fonction de répartition de X. En déduire la loi de X.

Exercice 3.6.6 Soient $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé et $A, B \subset \mathscr{F}$ deux évènements. On considère la v.a.r. X sur Ω définie par $X(\omega) = 1$ si ω réalise un et un seul des évènements A ou B et $X(\omega) = 0$ sinon. On pose $p_1 = \mathbf{P}(A)$, $p_2 = \mathbf{P}(B)$, $p_3 = \mathbf{P}(A \cap B)$.
(i) Déterminer la loi de X ainsi que son espérance et sa variance, en fonction de p_1, p_2, p_3.
(ii) En déduire les inégalités

$$\frac{p_1 + p_2 - 1}{2} \leq p_3 \leq \frac{p_1 + p_2}{2}.$$

Etudier les cas d'égalité.

Exercice 3.6.7 Soient X une variable aléatoire de Bernoulli de paramètre p et $P \in \mathbf{R}[x]$ un polynôme. Montrer que

$$P(X) = (P(1) - P(0))X + P(0)$$

et en déduire l'espérance et la variance de X.

Exercice 3.6.8 Soient $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé et X une v.a.r. sur Ω à valeurs dans $\mathbf{N}$. On suppose qu'il existe $q \in]0,1[$ tel que

$$\mathbf{P}(X = n) = q\mathbf{P}(X \geq n) \qquad \text{pour tout} \quad n \in \mathbf{N}.$$

Déterminer la loi de X.

Exercice 3.6.9 Soient $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé et $(X_n)_{n\in\mathbf{N}}$ une suite de v.a.r. sur Ω. Soit N une v.a.r sur Ω à valeurs dans $\mathbf{N}$. On définit $Y : \Omega \to \mathbf{R}$ par $Y(\omega) = X_{N(\omega)}(\omega)$ pour tout $\omega \in \Omega$. Montrer que Y est une v.a.r. sur Ω.

Exercice 3.6.10 Soient $a \neq b$ deux nombres réels non nuls, X une v. a. r. de loi uniforme sur $\{-a, a, b, -b\}$ et $Y = X^2$. Calculer les variances de X, Y et $X + Y$ respectivement. Les v.a.r. X et Y sont-elles indépendantes ?

Exercice 3.6.11 Soient X et Y deux v. a. r indépendantes suivant toutes les deux une loi de Bernoulli de paramètre p. On note $q = 1 - p$, $S = X + Y$ et $D = X - Y$. Déterminer les loi de S et D ainsi que leurs espérances et variances. Calculer $\mathbf{E}[SD]$. Les variables S et D sont-elles indépendantes ?

Exercice 3.6.12 Soient $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé et $X_1, \ldots, X_k$ des v.a.r. indépendantes prenant leurs valeurs dans $\{1, \ldots, N\}$ suivant la loi uniforme. Montrer que $Y = \max_{1\leq i\leq k} X_i$ et $Z = \min_{1\leq i\leq k} X_i$ sont des v.a.r. et donner leurs lois.

Exercice 3.6.13 Soit X une v.a.r sur l'espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$ à valeurs dans $\mathbf{N}$. Montrer que X possède une espérance si et seulement si la série $\sum_{n=0}^{+\infty} \mathbf{P}(X > n)$ converge et qu'alors

$$\mathbb{E}(X) = \sum_{n=0}^{+\infty} \mathbf{P}(X > n).$$

Exercice 3.6.14 Soient X et Y deux v.a.r indépendantes sur l'espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$ et suivant toutes deux une même loi géométrique $\mathscr{G}(p)$. Déterminer la loi de la v.a.r $Z = X + Y$.

Exercice 3.6.15 Soient X et Y deux v.a.r indépendantes sur l'espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$ et suivant des lois de Poisson de paramètres λ et μ. Déterminer la loi de la v.a.r $Z = X + Y$.

Exercice 3.6.16 Soit X une v.a.r sur l'espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$ suivant une loi géométrique $\mathscr{G}(p)$. Calculer $\mathbb{E}\left(\frac{1}{X}\right)$.

Exercice 3.6.17 Soit X une v.a.r sur l'espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$ suivant une loi de Poisson $\mathscr{P}(\lambda)$. Calculer $\mathbb{E}\left(\frac{1}{X+1}\right)$.

Exercice 3.6.18 Soient X et Y deux v.a.r à valeurs dans $\mathbf{N}$. On suppose qu'il existe $a \in \mathbf{R}$ tel que

$$\mathbf{P}(X = i, Y = j) = \frac{a}{i!j!} \qquad \text{pour tout} \quad (i, j) \in \mathbf{N}^2.$$

(i) Déterminer la valeur de a.
(ii) Quelles sont les lois marginales de X et de Y?
(iii) X et Y sont elles indépendantes ?

Exercice 3.6.19 Soit X une v.a.r. sur l'espace de probabilité $(\Omega, \mathscr{F}, \mathbf{P})$ suivant la loi géométrique $\mathscr{G}(p)$ de paramètre $p \in]0,1[$, c-à-d X prend ses valeurs dans $\mathbf{N}^*$ et pour tout $k \in \mathbf{N}^*$:

$$\mathbf{P}(X = k) = q^{k-1} p, \qquad \text{avec} \quad q = 1 - p.$$

(i) Déterminer la fonction de répartition F_X de X.
(ii) Pour $x \in \mathbf{R}$, calculer $\mathbf{P}(X > x)$.

Soient $X_1, \dots X_n$ des v.a.r. sur $(\Omega, \mathscr{F}, \mathbf{P})$ suivant toutes la loi $\mathscr{G}(p)$. On suppose que $X_1, \dots, X_n$ sont indépendantes, c-à-d

$$\mathbf{P}(X_1 = k_1, \dots, X_n = k_n) = \prod_{i=1}^{n} \mathbf{P}(X_i = k_i) \qquad \text{pour tout} \quad (k_1, \dots, k_n) \in \mathbf{N}.$$

On considère les v.a.r

$$U = \min(X_1, \dots, X_n) \quad \text{et} \quad V = \max(X_1, \dots, X_n).$$

(iii) Déterminer la fonction de répartition F_V de V et en déduire la loi de V.
(iv) Déterminer la loi de U.
(v) *Application :* On jette 5 dés. On met de côté les dés qui affichent un “6" et on recommence. De nouveau, on met de côté les dés qui affichent un “6" et ainsi de suite. Le jeu s'arrête quand il n'y a plus de dé. En associant au dé de numéro $i \in \{1, \dots, 5\}$ la v.a.r X_i qui est égale au moment d'apparition du “6' sur ce dé, déterminer la v.a.r N égale à la durée du jeu ainsi que son espérance $\mathbb{E}(N)$.

Exercice 3.6.20 À un péage autoroutier n voitures franchissent, au hasard et indépendamment les unes des autres, l'une des trois barrières numérotées $1, 2, 3$ mises à leur disposition. Soient X_1, X_2 et X_3 les v.a.r dénombrant les voitures ayant franchi ces barrières.
(i) Déterminer la loi de X_1.
(ii) Calculer les variances de X_1, X_2 et de $X_1 + X_2$.
(iii) En déduire la covariance de X_1 et X_2.

Exercice 3.6.21 Soient X et Y deux v.a.r sur l'espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$ possédant toutes deux un moment d'ordre 2. On suppose que $\mathrm{Var}(X) > 0$. Déterminer $(a, b) \in \mathbf{R}^2$ tels que la quantité $\mathbb{E}\left((Y - (aX + b))^2\right)$ soit minimale.

Exercice 3.6.22 Soit X une v.a.r suivant la loi uniforme sur $[a, b]$ et soit $(\alpha, \beta) \in \mathbf{R}^* \times \mathbf{R}$. Déterminer la loi de la v.a.r $Y = \alpha X + \beta$.

Exercice 3.6.23 (i) Soit X une v.a.r suivant une loi exponentielle de paramère λ. Déterminer la loi de la partie entière $[X]$ de X.

On dit qu'une v.a.r X sur l'espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$ est sans mémoire si $\mathbf{P}(X > t + s | X > s) = \mathbf{P}(X > t)$ pour tous t, s.
(ii) Montrez que si une v.a.r X suit une loi exponentielle, alors X est sans mémoire.
(iii) Montrez que si une v.a.r X est positive, sans mémoire et à densité, alors X suit une loi exponentielle.

Exercice 3.6.24 Soit X une v.a.r suivant la loi normale $\mathscr{N}(0, 1)$. On pose $Y = X^2$.
(i) Déterminer une densité g de Y. On dit que Y suit une *loi du* χ^2 *à un degré de liberté*
(ii) Calculer, en justifiant leur existence, $\mathbb{E}(Y)$ et $\mathrm{Var}(Y)$.

Exercice 3.6.25 Soient X et Y deux v.a.r indépendantes sur l'espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$ et suivant toutes deux une même loi géométrique $\mathscr{G}(p)$. Déterminer la loi de la v.a.r $Z = X + Y$.

Exercice 3.6.26 Soient X et Y deux v.a.r indépendantes sur l'espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$ et suivant des lois de Poisson de paramètres λ et μ. Déterminer la loi de la v.a.r $Z = X + Y$.

Exercice 3.6.27 Soit X une v.a.r sur l'espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$ suivant une loi géométrique $\mathscr{G}(p)$. Calculer $\mathbb{E}\left(\frac{1}{X}\right)$.

Exercice 3.6.28 Soit X une v.a.r sur l'espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$ suivant une loi de Poisson $\mathscr{P}(\lambda)$. Calculer $\mathbb{E}\left(\frac{1}{X+1}\right)$.

Exercice 3.6.29 Soient X et Y deux v.a.r à valeurs dans $\mathbf{N}$. On suppose qu'il existe $a \in \mathbf{R}$ tel que

$$\mathbf{P}(X = i, Y = j) = \frac{a}{i!j!} \qquad \text{pour tout} \quad (i, j) \in \mathbf{N}^2.$$

(i) Déterminer la valeur de a.
(ii) Quelles sont les lois marginales de X et de Y?
(iii) X et Y sont elles indépendantes ?

Exercice 3.6.30 On considère une urne contenant N boules numérotées de 1 à N. On effectue des tirages successifs, avec remise. Soit $k \geq 2$ fixé. On note T la v.a.r. égale au nombre de tirages effectués jusqu'à l'obtention de k boules identiques. L'ensemble des valeurs possibles de T est $\{l \in \mathbf{N} : l \geq k\} \cup \{+\infty\}$. On admettra que $\mathbf{P}(T = +\infty) = 0$.
(i) Déterminer $\mathbf{P}(T = k)$ et $\mathbf{P}(T = k+1)$.
(ii) Soit $n \geq 1$. Montrer que

$$\mathbf{P}(T = n + k) = \frac{N-1}{N^k}\mathbf{P}(T > n).$$

(iii) Montrer que T possède une espérance et la déterminer.

Exercice 3.6.31 Soit $f : \mathbf{R} \to \mathbf{R}$, $t \mapsto \frac{1}{2}e^{-|t|}$.
(i) Montrer que f définit une densité de probabilité.
Soit X une v.a.r de densité f.
(ii) Calculer la fonction de répartition F_X de X.
(iii) Montrer que $\mathbb{E}(X)$ et $\mathrm{Var}(X)$ existent et les calculer.

Exercice 3.6.32 Soit $f : \mathbf{R} \to \mathbf{R}$ la fonction définie par $f(t) = \frac{1}{\pi}\frac{1}{\sqrt{t(1-t)}}$ pour tout $t \in]0,1[$ et $f(t) = 0$ si $t \notin]0,1[$.
(i) Montrer que f définit une densité de probabilité.
Soit X une v.a.r de densité f.
(ii) Calculer la fonction de répartition F_X de X.
(iii) Montrer que $\mathbb{E}(X)$ et $\mathrm{Var}(X)$ existent.

Exercice 3.6.33 Soit $f : \mathbf{R} \to \mathbf{R}$, $t \mapsto \frac{1}{\pi(1+x^2)}$.
(i) Montrer que f définit une densité de probabilité.
Soit X une v.a.r de densité f.
(ii) Calculer la fonction de répartition F_X de X.
(iii) Montrer que X ne possède pas d'espérance.

Exercice 3.6.34 Soit X une v.a.r suivant la loi uniforme sur $[0,\pi]$. Montrer que $Y = \cos(X)$ suit une loi continue dont on déterminera la densité.

Exercice 3.6.35 **(Loi log-normale)** Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé et soit Y une v.a.r. sur Ω suivant une loi normale $\mathscr{N}(m,\sigma)$ avec $m \in \mathbf{R}$ et $\sigma > 0$. Soit

$$X := e^{Y}.$$

(i) Montrer que X suit une loi continue dont on déterminera la densité $g_{m,\sigma}$; la loi de X est appelé *loi log-normale* $\mathscr{L}\mathrm{N}(m,\sigma)$ de paramètres m, σ.
(ii) Montrer que X possède une espérance et calculer $\mathbb{E}(X)$.
(iii) Montrer que X possède une variance et calculer $\mathrm{Var}(X)$.
(iv) Soit $\alpha > 0, \beta \neq 0$. Déterminer la loi de la v.a.r. αX^{β}.

On dispose d'un tas de sable, composé de n grains homogènes sphériques. On admet que que le diamètre X (mesuré en mm) d'un grain de sable pris au hasard suit la loi $\mathscr{L}\mathrm{N}(-0.5; 0; 3)$. On passe la sable au crible d'un tamis, dont les trous circulaires sont de diamètre $0.5mm$.

(v) Calculer, au moyen d'une table de la loi normale, la probabilité p qu'un grain de sable passe à travers le tamis.
(vi) Calculer la proportion moyenne p_n des grains de sable passant à travers le tamis.
(vii) Quelle est la loi de la v.a.r. Y donnant le volume des grains de sable ?

Chapitre 4

Théorèmes limites

Dans certains cas, il arrive qu'une suite X_n de v.a.r. converge vers une v.a.r. X, dans un sens à préciser. Un tel résultat est appelé "Théorème limite" ; il permet de remplacer X_n par son approximation X, dont la loi est souvent plus facile à étudier.

4.1 Inégalité de Bienaymé-Tchebychev

Nous allons d'abord établir une inégalité classique très utile. Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé et $X : \Omega \to \mathbf{R}$ une v.a.r.

Proposition 4.1.1 ***(Inégalité de de Bienaymé-Tchebychev)*** *On suppose que* X *possède un moment d'ordre* 2 *Alors, pour tout* $a > 0$, *on a*

$$\boxed{\mathbf{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}.}$$

Démonstration Posons $m := \mathbb{E}(X)$.
1er cas : On suppose que X est une v.a.r. discrète, de sorte que $E = X(\Omega)$ est fini ou dénombrable infini. Posons $A = \{x \in E \ : \ |x - m| \geq a\}$. Alors

$$\mathbf{P}(|X - \mathbb{E}(X)| \geq a) = \mathbf{P}(|X - m| \geq a) = \sum_{x \in A} \mathbf{P}(X = x).$$

D'autre part, on a

$$\begin{aligned}\mathrm{Var}(X) &= \sum_{x\in E}(x-m)^2\mathbf{P}(X=x)\\ &\geq \sum_{x\in A}(x-m)^2\mathbf{P}(X=x)\\ &\geq a^2\sum_{x\in A}\mathbf{P}(X=x)\end{aligned}$$

D'où l'assertion dans ce cas.

2ème cas : On suppose que X est une v.a.r. continue, de densité f. Comme $\{|X-m|\geq a\}=\{X\leq m-a\}\cup\{X\geq m+a\}$, on a

$$\mathbf{P}(|X-m|\geq a)=\mathbf{P}(\{X\leq m-a\}\cup\{X\geq m+a\})=\int_{-\infty}^{m-a}f(x)dx+\int_{m+a}^{+\infty}f(x)dx.$$

D'autre part, on a

$$\begin{aligned}\mathrm{Var}(X) &= \int_{-\infty}^{+\infty}(x-m)^2f(x)dx\\ &\geq \int_{-\infty}^{m-a}(x-m)^2f(x)dx+\int_{m+a}^{+\infty}(x-m)^2f(x)dx\\ &\geq a^2\left(\int_{-\infty}^{m-a}f(x)dx+\int_{m+a}^{+\infty}f(x)dx\right).\end{aligned}$$

D'où l'assertion dans ce cas également.■

4.2 Convergence en probabilité de suites de v.a.r.

Soit $X_n:\Omega\to\mathbf{R}$ une suite de v.a.r. ainsi qu'une autre v.a.r. $X:\Omega\to\mathbf{R}$ sur l'espace probabilisé $(\Omega,\mathscr{F},\mathbf{P})$. Nous allons introduire une première notion de convergence vers X pour une telle suite $(X_n)_n$.

Définition 4.2.1 (Convergence en probabilité) On dit que $(X_n)_n$ *converge en probabilité* vers X si :

$$\boxed{\text{pour tout } \varepsilon>0, \text{on a } \lim_{n\to+\infty}\mathbf{P}(|X_n-X|\geq\varepsilon)=0.}$$

On écrit alors, de façon abrégée,

$$X_n\xrightarrow[n\to+\infty]{\mathbf{P}}X.$$

Exemple 4.2.2 Supposons que X_n suive une loi exponentielle de paramètres $n\lambda$ pour un $\lambda > 0$ fixé. Montrons que $X_n \xrightarrow[n\to+\infty]{\mathbf{P}} 0$.

En effet, soit $\varepsilon > 0$. Alors

$$\begin{aligned}\mathbf{P}(|X_n| \geq \varepsilon) &= \int_{\varepsilon}^{+\infty} n\lambda e^{-n\lambda x} dx \\ &= \left[-e^{-n\lambda x}\right]_{\varepsilon}^{+\infty} = e^{-n\lambda\varepsilon}.\end{aligned}$$

Comme $\lim_{n\to+\infty} e^{-n\lambda\varepsilon} = 0$, on a bien $X_n \xrightarrow[n\to+\infty]{\mathbf{P}} 0$.

4.3 Loi faible des grands nombres

Nous allons montrer que, quand on répète une expérience aléatoire un grand nombre de fois, la fréquence de réalisations d'un événement converge vers la probabilité de réalisation de cet événement. Ce résultat, appelé "Loi des grands nombres", est un outil d'une grande importance.

Théorème 4.3.1 ***(Loi faible des grands nombres)*** *Soit* $(X_n)_{n\geq 1}$ *une suite de v.a.r. sur l'espace probabilisé* $(\Omega, \mathscr{F}, \mathbf{P})$ *. On suppose que ces v.a.r. admettent toutes une même espérance* m*, un même écart-type* $\sigma > 0$ *et qu'elles sont deux-à-deux non corrélées. Posons* $M_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$. *Alors* $M_n \xrightarrow[n\to+\infty]{\mathbf{P}} m$ *; plus précisément,*

$$\forall \varepsilon > 0 : \ \mathbf{P}(|M_n - m| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

Démonstration On a

$$\mathbb{E}(M_n) = \mathbb{E}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(X_k) = \frac{1}{n}\sum_{k=1}^{n} m = m,$$

ainsi que (voir Corollaire 3.3.15)

$$\mathrm{Var}(M_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n^2}\sum_{k=1}^{n}\mathrm{Var}(X_k) + 2\sum_{1\leq i<j\leq n}\mathrm{Cov}(X_i, X_j)$$

et donc, comme les X_k sont deux-à-deux non corrélées et $\mathrm{Var}(X_k) = \sigma^2$,

$$\mathrm{Var}(M_n) = \frac{1}{n^2}\sum_{k=1}^{n}\mathrm{Var}(X_k) = \frac{\sigma^2}{n}.$$

Soit $\varepsilon > 0$. Par l'inégalité de Bienaymé-Tchebychev (Proposition 4.1.1), on a

$$\mathbf{P}(|M_n - m| \geq \varepsilon) \leq \frac{\mathrm{Var}(M_n)}{\varepsilon^2}.$$

Il s'ensuit que

$$\mathbf{P}(|M_n - m| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

Comme $\lim_{n\to+\infty} \frac{\sigma^2}{n\varepsilon^2} = 0$, on a bien $M_n \xrightarrow[n\to+\infty]{\mathbf{P}} m$.■

Remarque 4.3.2 Les hypothèses du théorème précédent sont satisfaites si les X_n sont indépendantes, ont tous la même loi et admettent un moment d'ordre 2. C'est souvent pour une telle suite que le théorème est appliqué.

Exemple 4.3.3 (Cas de variables aléatoires de Bernoulli) Soit $(X_n)_n$ une suite de v.a.r. indépendantes et suivant une loi de Bernoulli $\mathscr{B}(p)$ pour $p \in]0,1[$; on dira que l'évènement $X_n = 1$, dont la probabilté est égale à p, est un *succès*. La suite $M_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ représente alors la proportion de succès parmi les n premières épreuves. Comme $\mathbb{E}(X_n) = p$ et $\sigma^2 = \mathrm{Var}(X_n) = p(1-p)$, on a, par la loi des grands nombres : pour tout $\varepsilon > 0$,

$$\mathbf{P}(|M_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2}.$$

Comme la fonction $[0,1] \to \mathbf{R}, x \mapsto x(x-1)$ prend son maximum en $x = 1/2$, où elle vaut $1/4$, on a donc

$$\mathbf{P}(|M_n - p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}.$$

Considérons, par exemple, le lancer d'une pièce de monnaie équilibrée, répété plusieurs fois. Soit X_n la v.a.r égale à 1 si on obtient "Pile" au n-ième lancer et 0 sinon. Alors $S_n = X_1 + \cdots + X_n$ est égale au nombre de "Piles" obtenus lors des n premiers lancers et $X_n \sim \mathscr{B}(1/2)$. Ce qui précède montre que, pour $n = 10^3 = 1000$ lancers et $\varepsilon = 0.05$, on a

$$\mathbf{P}(|S_{1000} - 500| \geq 50) = \mathbf{P}(|M_{1000} - \frac{1}{2}| \geq \varepsilon|) \leq \frac{1}{4 \times 10^3 \times 25 \times 10^{-4}} = 0.1$$

On s'attend donc à obtenir un nombre de "Piles" compris entre 450 et 550, avec une probabilité supérieure à 90%. En fait, cette borne est assez pessimiste : on peut montrer (voir plus loin Exemple 4.6.3) que ce nombre est compris entre 484 et 516, avec une probabilité de 95%.

4.4 Converge en loi d'une suite de v.a.r

Soient X_n et X des v.a.r. dont on ne suppose *pas* qu'elles sont définies sur un même espace probabilisé.

Définition 4.4.1 (Convergence en loi) On dit que $(X_n)_n$ *converge en loi* vers X si, pour tout $x \in \mathbf{R}$ tel que la fonction de répartition F_X de X est continue en x, on a $\lim_{n\to+\infty} F_{X_n}(x) = F_X(x)$. On écrit alors, de façon abrégée,

$$X_n \xrightarrow[n\to+\infty]{\mathscr{L}} X.$$

Remarque 4.4.2 (i) Supposons que $X_n \xrightarrow[n\to+\infty]{\mathscr{L}} X$. Alors, pour tous points de continuité a, b de F_X avec $a \leq b$, on a

$$\lim_n \mathbf{P}(a < X_n \leq b) = \lim_n F_{X_n}(b) - F_{X_n}(a) = F_X(b) - F_X(a) = \mathbf{P}(a < X \leq b).$$

(ii) Un cas important (voir plus loin le théorème central limite) sera la cas où X suit une loi normale $\mathscr{N}(0,1)$. La condition pour que $X_n \xrightarrow[n\to+\infty]{\mathscr{L}} X$ signifie alors que, pour tout $a \leq b$, on a

$$\lim_n \mathbf{P}(a < X_n \leq b) = \Pi(b) - \Pi(a),$$

où $\Pi(x) = \dfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$.

Exemple 4.4.3 Supposons que la v.a.r. X_n suive une loi normale $\mathscr{N}(0, \sigma_n)$ et que X suive une loi normale $\mathscr{N}(0, \sigma)$. Si $\lim_n \sigma_n = \sigma$, alors $X_n \xrightarrow[n\to+\infty]{\mathscr{L}} X$.

Dans le cas de v.a.r. discrètes, la proposition suivante donne une condition nécessaire et suffisante pour prouver la convergence en loi.

Proposition 4.4.4 *Supposons que les v.a.r.* X_n *et* X *prennent leurs valeurs un ensemble contenu dans* $\mathbf{Z}$. *Alors, la suite* $X_n \xrightarrow[n\to+\infty]{\mathscr{L}} X$ *si et seulement si*

$$\boxed{\lim_n \mathbf{P}(X_n = k) = \mathbf{P}(X = k) \qquad \textit{pour tout} \quad k \in \mathbf{Z}.}$$

Démonstration : voir Poly complet du cours.

4.5 Approximations de lois discrètes

4.5.1 Approximation de la loi binomiale par la loi de Poisson

Etant donnée une v.a.r qui suit une loi binomiale $\mathscr{B}(n,p)$, le calcul des quantités $\mathbf{P}(X=k)$ est souvent fastidieux quand n est “grand" , car il requiert celui des coefficients binomiaux $\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$. La proposition suivante montre que, sous certaines conditions, on peut remplacer la loi binomiale par une loi de Poisson.

Proposition 4.5.1 ***(Approximation de lois binomiales)*** *Soit* $(X_n)_n$ *une suite de v.a.r. de loi* $\mathscr{B}(n,p_n)$. *Supposons que* $\lim_n np_n = \lambda \in \mathbf{R}$. *Alors,* $X_n \xrightarrow[n\to+\infty]{\mathscr{L}} X$ *pour une v.a.r de loi de Poisson* $\mathscr{P}(\lambda)$, *c-à-d pour tout* $k \in \mathbf{N}$, *on a*

$$\lim_n \mathbf{P}(X_n = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Démonstration Soit k fixé. On a, pour $n \geq k$,

$$\begin{aligned}\mathbf{P}(X_n = k) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} \\ &= \frac{n(n-1)\dots(n-k+1)}{k!} p_n^k e^{(n-k)\log(q_n)}.\end{aligned}$$

Comme $p_n \sim \lambda/n$, on a $\lim_n p_n = 0$ et

$$(n-k)\log(q_n) = (n-k)\log(1-p_n) \sim -np_n \sim -\lambda.$$

D'ou $\lim_n (n-k)\log(1-p_n) = -\lambda$ et donc $\lim_n e^{(n-k)\log(q_n)} = e^{-\lambda}$. Par conséquent, on a

$$\mathbf{P}(X_n = k) \sim \frac{n^k}{k!} p_n^k e^{-\lambda} = \frac{(np_n)^k}{k!} e^{-\lambda} \to_{n\to\infty} e^{-\lambda}\frac{\lambda^k}{k!}. \blacksquare$$

Remarque 4.5.2 Dans la pratique, on considère que la loi binomiale $\mathscr{B}(n,p)$ peut être approchée de manière acceptable par une loi de Poisson avec $\lambda = np$, quand $n > 30$ et np de l'ordre de quelques unités (par exemple, $np \leq 10$).

Exemple 4.5.3 On considère l'épreuve consistant en un lancer de 5 dés plusieurs fois de suite, et Y le nombre de fois où on obtient 5 fois "6". Alors Y une v.a.r de loi $\mathscr{B}(n,p)$ avec $p = 1/6^5 = 1/7776$. Pour $n = 10000$, le calcul exact donne

$$\mathbf{P}(\mathrm{Y}=4) = \binom{10000}{4}\left(\frac{1}{7776}\right)^4\left(\frac{7775}{7776}\right)^9 996 \sim 0.03149.$$

En posant $\lambda = np = 10^5/7776 = 1,286$, on a la bonne approximation par une v.a.r $\mathrm{X} \sim \mathscr{P}(\lambda)$:

$$\mathbf{P}(\mathrm{X}=4) = e^{-\lambda}\frac{\lambda^k}{k!} \sim 0.0315.$$

4.6 Théorème central limite

Soit $(\mathrm{X}_n)_{n\geq 1}$ une suite de v.a.r. sur l'espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$, indépendantes, de même loi et admettant une espérance m et une variance σ^2. La loi des grands nombres montre que $\mathrm{M}_n = \dfrac{\mathrm{S}_n}{n}$ tend vers m en probabilté, où

$$\mathrm{S}_n = \mathrm{X}_1 + \cdots + \mathrm{X}_n.$$

L'erreur faite en estimant m par M_n est la v.a.r. centrée $\mathrm{E}_n = \mathrm{M}_n - m$; la variance de E_n est

$$\mathrm{Var}(\mathrm{E}_n) = \mathrm{Var}(\mathrm{M}_n) = \frac{\sigma^2}{n}.$$

La v.a.r. centrée-réduite correspondante est donc

$$\mathrm{Z}_n = \frac{\sqrt{n}}{\sigma}(\mathrm{M}_n - m) = \frac{\mathrm{S}_n - nm}{\sigma\sqrt{n}}.$$

Le résultat suivant, qui mérite son nom de Théorème Central Limite ("TCL"), dit que le comportement asymptotique de Z_n est décrit par la loi normale $\mathscr{N}(0,1)$.

Théorème 4.6.1 ***(Théorème Central Limite)*** *Soit* $(\mathrm{X}_n)_{n\geq 1}$ *une suite de v.a.r. sur l'espace probabilisé* $(\Omega, \mathscr{F}, \mathbf{P})$, *indépendantes, de même loi et admettant une espérance m et une variance* σ^2. *Soit* $\mathrm{Z}_n = \dfrac{\mathrm{S}_n - nm}{\sigma\sqrt{n}}$. *Alors* $\mathrm{Z}_n \xrightarrow[n\to+\infty]{\mathscr{L}} \mathrm{X}$ *pour une v.a.r.* X *de loi normale centrée-réduite* $\mathscr{N}(0,1)$, *c-à-d pour tout a < b, on a*

$$\mathbf{P}(a < \mathrm{Z}_n \leq b) \to_{n\to+\infty} \int_a^b \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt.$$

Remarque 4.6.2 La preuve du Théorème Central Limite dépasse le cadre de ce cours.

Exemple 4.6.3 (Approximation de la loi binomiale par la loi normale)

Soit $(X_n)_{n\geq 1}$ une suite de v.a.r indépendantes, de loi de Bernoulli $\mathscr{B}(p)$. Alors $S_n = X_1 + \cdots + X_n$ suit la loi binomiale $\mathscr{B}(n,p)$. Comme $\mathbb{E}(X_n) = p$ et $\mathrm{Var}(X_n) = p(1-p)$, on a, par le Théorème Central Limite,

$$\mathbf{P}\left(a < \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) \to_{n\to+\infty} \Pi(b) - \Pi(a)$$

En pratique, on admet l'approximation

$$\mathbf{P}\left(a < \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) = \Pi(b) - \Pi(a)$$

dès que $n \geq 30$ et $np(1-p) \geq 9$.

Pour $a = 1.96$, on a $\Pi(a) = 0.975$ et $\Pi(-a) = 1 - \Pi(-a) = 0.025$; si n est suffisamment grand, on a donc

$$\mathbf{P}\left(-\sqrt{np(1-p)}a < S_n - np \leq \sqrt{np(1-p)}a\right) = \Pi(a) - \Pi(-a) = 0.95.$$

Reprenons l'Exemple 4.3.3. Soit X_n la v.a.r égale à 1 si on obtient “Pile" au n-ième lancer d'une pièce équilibrée et 0 sinon. Alors $S_n = X_1 + \cdots + X_n$ est égal au nombre de “Piles" obtenus lors des n premiers lancers. On a $p = 1/2$ et $p(1-p) = 1/4$. et donc, pour n suffisamment grand,

$$\mathbf{P}\left(-\frac{\sqrt{n}a}{4} < S_n - \frac{n}{2} \leq \frac{\sqrt{n}a}{4}\right) = \Pi(a) - \Pi(-a) = 0.95$$

Pour $n = 1000$, on a ainsi

$$\mathbf{P}(-15.49 < S_{1000} - 500 \leq 15.49) = 0.95,$$

c-à-d $\mathbf{P}(S_{1000} \in [486, 516]) = 0.95$.

4.7 Application du TCL : intervalles de confiance

On considère une suite $X_1, X_2, \ldots, X_n$ de v.a.r idépendantes, suivant toutes une même loi de Bernoulli $\mathscr{B}(p)$ pour un paramètre inconnu $p \in]0,1[$. On cherche à **estimer** p **à travers un échantillon** $(x_1, \ldots, x_n)$ **de** $X_1, \ldots, X_n$; un tel échantillon est une réalisation de $X_1, \ldots, X_n$, c-à-d

$$(x_1, \ldots, x_n) = (X_1(\omega), \ldots, X_n(\omega))$$

pour un $\omega \in \Omega$.

Soit

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Pour n suffisamment grand (dans la pratique $n \geq 30$ suffit), on peut considérer, par le Théorème 6.2.3 que

$$Z := \frac{\overline{X} - p}{\sqrt{p(1-p)}/\sqrt{n}} = \sqrt{n}\frac{\overline{X} - p}{\sqrt{p(1-p)}} \sim \mathscr{N}(0,1).$$

Soit $\alpha \in]0,1[$ fixé (**"seuil de risque"**). On note z_α le nombre réel ≥ 0 tel que $\Pi(-z_\alpha) = \alpha$, c-à-d $\mathbf{P}(Z \geq z_\alpha) = \alpha$ ou bien $\mathbf{P}(Z \leq z_\alpha) = 1 - \alpha$. La valeur z_α est lue sur la table de la loi normale (voir table dans l'Annexe 8 plus loin). Alors

$$\mathbf{P}(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = \Pi(z_{\alpha/2}) - \Pi(-z_{\alpha/2}) = 2\Pi(z_{\alpha/2}) - 1 = 1 - \alpha.$$

et donc

$$\begin{aligned}
1 - \alpha &= \mathbf{P}(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) \\
&= \mathbf{P}\left(-z_{\alpha/2} \leq \sqrt{n}\frac{\overline{X} - p}{\sqrt{p(1-p)}} \leq z_{\alpha/2}\right) \\
&= \mathbf{P}\left(\overline{X} - z_{\alpha/2}\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq p \leq \overline{X} + z_{\alpha/2}\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right).
\end{aligned}$$

Comme $\sqrt{p(1-p)} \leq 1/2$, on obtient

$$\mathbf{P}\left(\overline{X} - z_{\alpha/2}\frac{1}{2\sqrt{n}} \leq p \leq \overline{X} + z_{\alpha/2}\frac{1}{2\sqrt{n}}\right) \geq 1 - \alpha.$$

Posons $\overline{x} = \dfrac{1}{n}\sum_{i=1}^n x_i$, qui est la **moyenne observée** au vu de l'échantillon $x_1, \ldots, x_n$. **L'intervalle de confiance**, au risque α, est alors donné par

$$\boxed{\mathrm{I}_\alpha = \left[\overline{x} - z_{\alpha/2}\frac{1}{2\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{1}{2\sqrt{n}}\right].}$$

Ceci signifie que $p \in \mathrm{I}_\alpha$ avec probabilité au moins $1-\alpha$. Pour $\alpha = 0.05$, on a $z_{\alpha/2} = 1.96$ et donc $p \in \left[\overline{x} - 1.96\dfrac{1}{2\sqrt{n}}, \overline{x} + 1.96\dfrac{1}{2\sqrt{n}}\right]$ avec probabilité de 95%

Exemple 4.7.1 ((**Sondage**)) Au soir d'une élection, on veut estimer la proportion (inconnue) p des électeurs du candidat A dans l'ensemble de la population, à travers un sondage portant sur 2500 électeurs. Ici, $n = 2500$ est le nombre d'électeurs sondés (taille de l'échantillon du sondage) et, pour chaque $i \in \{1, \ldots, 2500\}$, on a $\mathrm{X}_i = 1$ si l'électeur i dit avoir voté pour A et $\mathrm{X}_i = 0$ sinon.

Supposons que le sondage a révélé 1350 votes pour le candidat A. On a donc

$$\overline{x} = \frac{1300}{2500} = 0.54$$

et l'intervalle de confiance au risque $\alpha = 5\%$ est

$$[0.54 - 1.96 \times (1/2\sqrt{2500}), 0.54 + 1.96 \times (1/2\sqrt{2500})] = [0.52, 0.56];$$

on a donc une fourchette entre 52% et 56% dans la prédiction de votes pour le candidat A, avec probabilité de 95%.

Chapitre 5

Couples de variables aléatoires à densité

Nous avons étudié à la Section 3.3 des couples de variables aléatoires *discrets*. Dans ce chapitre, nous abordons les couples de variables aléatoires *continus*.

5.1 Vecteurs aléatoires à densité

Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé.

Soit $n \geq 1$. On rappelle qu'un *vecteur aléatoire* de dimension n est une application

$$Z : \Omega \to \mathbf{R}^n, \omega \mapsto (X_1(\omega), \ldots, X_n(\omega)),$$

où $X_1, \ldots, X_n$ sont des v.a.r. On note en général Z par $Z = (X_1, \ldots, X_n)$.

Dans le cas $n = 2$, on parlera de *couple aléatoire*, et on le notera souvent $Z = (X, Y)$.

On considère la tribu $\mathscr{B}(\mathbf{R}^n) \subset \mathscr{P}(\mathbf{R}^n)$ engendrée par les "rectangles" $I_1 \times I_2 \times \cdots \times I_n$, où les I_i sont des intervalles de $\mathbf{R}$. Chaque partie de $\mathscr{B}(\mathbf{R}^n)$ est dite partie borélienne ou (borélien) de $\mathbf{R}^n$. On dit que $\mathscr{B}(\mathbf{R}^n)$ est la tribu des boréliens de $\mathbf{R}^n$.

Définition 5.1.1 (Loi et fonction de répartition d'un vecteur aléatoire) Soit $Z = (X_1, \ldots, X_n)$ un vecteur aléatoire.
(i) La *loi* de Z (ou la *loi conjointe* de $X_1, \ldots, X_n$) est la mesure de probabilité $P_Z : \mathscr{B}(\mathbf{R}^n) \to [0, 1]$ définie par

$$P_Z(I_1 \times I_2 \times \cdots \times I_n) = \mathbf{P}(X_1 \in I_1, X_2 \in I_2, \ldots, X_n \in I_n).$$

pour tous intervalles $I_1, I_2, \dots, I_n$ de $\mathbf{R}$.
(ii) La *fonction de répartition* de Z est la fonction $F_Z : \mathbf{R}^n \to [0,1]$ définie par

$$F_Z(x_1, x_2, \dots, x_n) = \mathbf{P}(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n).$$

Nous allons introduire maintenant les vecteurs aléatoires continus.

Définition 5.1.2 (Vecteur aléatoire à densité) Soit $Z = (X_1, \dots, X_n)$ un vecteur aléatoire. On dit que Z est une *vecteur aléatoire à densité* ou *vecteur aléatoire continue* si sa fonction de répartition F_Z s'écrit sous la forme

$$F_Z(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(t_1, \dots, t_n) dt_1 \cdots dt_n \qquad \text{pour tout} \quad (x_1, \dots, x_n) \in \mathbf{R}^n$$

pour une fonction $f : \mathbf{R}^n \to \mathbf{R}$, appelée *densité de* Z, avec les propriétés suivantes :

1. $f(x_1, \dots, x_n) \geq 0$ pour tout $(x_1, \dots, x_n) \in \mathbf{R}^n$;
2. f est intégrable sur $\mathbf{R}^n$;
3. $\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(t_1, \dots, t_n) dt_1 \cdots dt_n = 1$.

Exemple 5.1.3 Soit $Z = (X, Y)$ de loi uniforme sur le triangle

$$T = \{(x, y) \in \mathbf{R}^2 \mid x \geq 0, y \geq 0, x + y \leq 1\};$$

comme l'aire de T est $1/2$, la densité de Z est $f = 2\mathbf{1}_T$, c-à-d

$$f(x, y) = \begin{cases} 2 & \text{si } (x, y) \in T \\ 0 & \text{si } (x, y) \notin T. \end{cases}$$

Remarque 5.1.4 Soit $Z = (X_1, \dots, X_n)$ un vecteur aléatoire à densité. On peut calculer ainsi la probabilité d'évènements associés à Z au moyen de la densité f de Z :

$$\forall B \in \mathscr{B}(\mathbf{R}^n) \qquad : \qquad \mathbf{P}(Z \in B) = P_Z(B) = \int_B f(t_1, \dots, t_n) dt_1 \cdots dt_n.$$

5.2 Lois marginales

Dans toute la suite, on se limitera au cas $\boxed{n=2}$, c-à-d à celui d'un couple aléatoire $Z = (X, Y) : \Omega \to \mathbf{R}^2$.

Les *lois marginales* du couple Z sont les lois P_X et P_Y de X et Y.

Proposition 5.2.1 *Soit* $Z = (X, Y)$ *un vecteur aléatoire à densité* f. *Alors les v.a.r.* X *et* Y *sont continues, de densités* f_X *et* f_Y *données par*

$$\boxed{f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy} \qquad \textit{et} \qquad \boxed{f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)dx}$$

pour $x, y \in \mathbf{R}$.

Démonstration Soit $x \in \mathbf{R}$. On a $\{X \le x\} = \{X \le x, Y \in \mathbf{R}\}$. Il s'ensuit que, pour la fonction de répartition F_X de X, on a

$$\begin{aligned}
F_X(x) &= \mathbf{P}(X \le x) = \mathbf{P}(X \le x, Y \in]-\infty, +\infty[) \\
&= P_Z(]-\infty, x] \times]-\infty, +\infty[) = \int_{-\infty}^{x}\int_{-\infty}^{+\infty} f(t,y)dtdy = \int_{-\infty}^{x}\int_{-\infty}^{+\infty} f(t,y)dydt \\
&= \int_{-\infty}^{x} f_X(t)dy,
\end{aligned}$$

où $f_X(t) = \int_{-\infty}^{+\infty} f(t,y)dy$. Ceci montre que X est continue à densité f_X.

De manière similaire, on montre que Y est continue à densité f_Y.■

Exemple 5.2.2 Reprenons l'Exemple 5.1.3. On rappelle que la densité de $Z = (X, Y)$ est $f = 2\mathbf{1}_T$, où T est le triangle $\{(x, y) \in \mathbf{R}^2 \mid x \ge 0, y \ge 0, x + y \le 1\}$.

La loi marginale f_X de X est donnée par $f_X(x) = 0$ pour $x \notin [0,1]$ et

$$f_X(x) = 2\int_0^{1-x} \mathbf{1}_T(x,y)dxdy = 2(1-x)$$

pour $x \in [0,1]$. De manière similaire, $f_Y(y) = 0$ pour $y \notin [0,1]$ et $f_Y(y) = 2(1-y)$ pour $y \in [0,1]$.

On rappelle que les v.a.r. X et Y sont indépendantes si et seulement si

$$\forall (x,y) \in \mathbf{R}^2 : \qquad \mathbf{P}(X \le x, Y \le y) = \mathbf{P}(X \le x)\mathbf{P}(Y \le y).$$

Proposition 5.2.3 *Soit* $Z = (X, Y)$ *un vecteur aléatoire à densité* f. *Les deux propriétés suivantes sont équivalentes :*
(i) X *et* Y *sont indépendantes ;*
(ii) $f(x, y) = f_X(x) f_Y(y)$ *pour tout* $(x, y) \in \mathbf{R}^2$.*Pro-IndCouple*

Démonstration Par définition, X et Y sont indépendantes si et seulement si $F_Z(x, y) = F_X(x)F_Y(y)$, c-à-d si et seulement si

$(*)\ \int_{-\infty}^{x}\int_{-\infty}^{y} f(t, s)\,dt\,ds = (\int_{-\infty}^{x} f_X(t)\,dt)(\int_{-\infty}^{y} f_X(s)\,ds)$

pour tout $(x, y) \in \mathbf{R}^2$.

Supposons que $f(x, y) = f_X(x) f_Y(y)$ pour tout $(x, y) \in \mathbf{R}^2$; alors la relation $(*)$ est satisfaite et X et Y sont donc indépendantes.

Réciproquement, supposons que X et Y sont indépendantes et que donc la relation $(*)$ est satisfaite. Alors, on a

$$\int_{-\infty}^{x}\int_{-\infty}^{y} f(t, s)\,dt\,ds = \int_{-\infty}^{x}\int_{-\infty}^{y} (f_X(t) f_X(s))\,dt\,ds$$

pour tout $(x, y) \in \mathbf{R}^2$. Par identification, on a $f(x, y) = f_X(x) f_Y(y)$ pour tout $(x, y) \in \mathbf{R}^2$.$\blacksquare$

On peut démontrer une formule de transfert analogue à celle pour un couple de v.a.r. discrètes.

Proposition 5.2.4 ***(Formule de transfert)*** *Soit* $Z = (X, Y)$ *un vecteur aléatoire à densité* f. *Soit* $g : \mathbf{R}^2 \to \mathbf{R}$ *une fonction continue (ou plus généralement, mesurable). On suppose que l'intégrale* $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |g(x, y)| f(x, y)\,dx\,dy$ *est finie. Alors la v.a.r.* $g(X, Y)$ *possède une espérance et on a*

$$\boxed{\mathbb{E}(g(X, Y)) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x, y) f(x, y)\,dx\,dy}$$

Démonstration Omise.

5.3 Covariance

Comme pour un couple de v.a.r. discrètes, on peut définir la covariance d'un couple de variables continues.

Définition 5.3.1 (Covariance d'un couple aléatoire à densité) Soit $Z = (X, Y)$ un vecteur aléatoire à densité. On suppose que X et Y possèdent des variances. La *covariance* de Z (ou de X et Y), notée $\text{Cov}(X, Y)$, est le nombre

$$\boxed{\text{Cov}(X, Y) := \mathbb{E}\left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right)}$$

Les v.a.r. X et Y sont dites *non corrélées* si $\text{Cov}(X, Y) = 0$.

La covariance a des propriétés similaires à celle pour un couple de v.a.r. discrètes.

Proposition 5.3.2 *Soit* $Z = (X, Y)$ *un vecteur aléatoire à densité* f. *On suppose que* X *et* Y *possèdent une variance.*
(i) ***(Formule de Koenig-Huygens)*** *On a*

$$\boxed{\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).}$$

(ii) On a

$$\boxed{\text{Cov}(X, Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x, y)\,dx\,dy - \left(\int_{-\infty}^{+\infty} xf_X(x)\,dx\right)\left(\int_{-\infty}^{+\infty} yf_X(y)\,dy\right).}$$

(iii) Si X *et* Y *sont indépendantes, alors* $\text{Cov}(X, Y) = 0$; *la réciproque est fausse en général.*

Démonstration
(i) On a, en développant et en utilisant la linéarité de l'espérance

$$\begin{aligned}\text{Cov}(X, Y) &= \mathbb{E}\left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right) = \mathbb{E}\left(XY - X\mathbb{E}(Y) - \mathbb{E}(X)Y + \mathbb{E}(X)\mathbb{E}(Y)\right)\\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).\end{aligned}$$

(ii) Par la formule de transfert (Proposition 5.2.4), on a

$$\mathbb{E}(XY) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x, y)\,dx\,dy.$$

ceci montre que (ii) est une conséquence de (i) .

(iii) Supposons que X et Y sont indépendantes. Alors, en utilisant la Proposition 5.2.3, on a

$$\begin{aligned}
\mathbb{E}(XY) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf_X(x)f_Y(y)dxdy \\
&= \left(\int_{-\infty}^{+\infty} xf_X(x)dx\right)\left(\int_{-\infty}^{+\infty} yf_X(y)dy\right) \\
&= \mathbb{E}(X)\mathbb{E}(Y),
\end{aligned}$$

c-à-d $\mathrm{Cov}(X,Y) = 0.$■

Exemple 5.3.3 Reprenons l'Exemple 5.1.3 et 5.2.2 On rappelle que la loi f de $Z = (X,Y)$ est la loi uniforme sur le triangle T ; les lois marginales f_X et f_Y sont données par $f_X(x) = f_Y(x) = 2\mathbf{1}_{[0,1]}(x)(1-x)$. On calcule que

$$\mathbb{E}(X) = \mathbb{E}(Y) = 2\int_0^1 x(1-x)dx = 1/3$$

et que

$$\mathbb{E}(XY) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy = 2\int_0^1\int_0^{1-x} xydxdy = \int_0^1 x(1-x)^2dx = 1/12$$

D'où $\mathrm{Cov}(X,Y) = (1/3)^2 - 1/12 = -1/36.$

5.4 Lois et espérance conditionnelles

Soit $Z = (X, Y)$ un couple aléatoire à densité f. Comme dans le cas de couples discrets (voir Definition 3.3.10), on peut définir la loi conditionnelle de Y sachant $X = x$. On rappelle que f_X et f_Y sont les densités de X et Y.

Définition 5.4.1 (Densité conditionnelle) Soit $x \in \mathbf{R}$ tel que $f_X(x) > 0$. La fonction $f_{Y|X=x} : \mathbf{R} \to \mathbf{R}$ définie par

$$\forall y \in \mathbf{R} : \qquad \boxed{f_{Y|X=x}(y) = \frac{f(x,y)}{f_X(x)}}$$

est une densité, appelée *densité conditionnelle* de Y sachant $X = x$. De manière similaire, si $y \in \mathbf{R}$ tel que $f_Y(y) > 0$, la densité conditionnelle $f_{X|Y=y}$ de X sachant $Y = y$ est définie par $f_{X|Y=y}(x) = \dfrac{f(x,y)}{f_Y(y)}$.

Remarque 5.4.2 On remarque que $f_{Y|X=x}$ est bien une densité pour tout $x \in \mathbf{R}$ tel que $f_X(x) > 0$; en effet, en utilisant la Proposition 5.2.1, on a

$$\int_{-\infty}^{+\infty} f_{Y|X=x}(y)\,dy = \frac{1}{f_X(x)} \int_{-\infty}^{+\infty} f(x,y)\,dy = \frac{1}{f_X(x)} f_X(x) = 1.$$

Définition 5.4.3 (Espérance conditionnelle) On suppose maintenant que Y possède une espérance.
(i) Pour tout $x \in \mathbf{R}$ tel que $f_X(x) > 0$, le nombre réel $\mathbb{E}(Y|X = x)$, défini par

$$\boxed{\mathbb{E}(Y|X = x) = \int_{-\infty}^{+\infty} y f_{Y|X=x}(y)\,dy},$$

est appelé *espérance conditionnelle* de Y sachant $X = x$.
(ii) Soit $N = \{x \in \mathbf{R} \mid f_X(x) = 0\}$ et définissons $\psi : \mathbf{R} \to \mathbf{R}$ par $\psi(x) = \mathbb{E}(Y|X = x)$ pour $x \in \mathbf{R} \setminus N$ et $\psi(x) = 0$ pour $x \in N$. La variable aléatoire $\mathbb{E}(Y|X)$, définie par

$$\boxed{\mathbb{E}(Y|X) = \psi(X)},$$

est appelée *espérance conditionnelle* de Y sachant X.

Remarque 5.4.4 (i) Soit $N = \{x \in \mathbf{R} \mid f_X(x) = 0\}$. Alors $\mathbf{P}(X \in N) = \int_N f_X(t)\,dt = 0$. En notant $A = \{X \notin N\}$, on a donc $\mathbf{P}(A) = 1$. Ainsi, on a

$$\mathbb{E}(Y|X)(\omega) = \psi(X(\omega)) = \mathbb{E}(Y|X = X(\omega))$$

pour $\omega \in A$ et $\mathbb{E}(Y|X)(\omega) = 0$ pour $\omega \notin A$.
(ii) En échangeant les rôles de X et Y, on définit de manière similaire $\mathbb{E}(X|Y)$, l'espérance conditionnelle de X sachant Y.

L'espérance conditionnelle possède la propriéte fondamentale suivante.

Proposition 5.4.5 *Soit* $Z = (X, Y)$ *un couple aléatoire à densité* f. *On suppose que* Y *possède une espérance. Alors, on a*

$$\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X)) = \int_{-\infty}^{+\infty} \mathbb{E}(Y|X = x) f_X(x) dx.$$

Démonstration On a, par la formule de transfert (Proposition 5.2.4)

$$\begin{aligned}
\mathbb{E}(\mathbb{E}(Y|X)) &= \int_{-\infty}^{+\infty} \psi(x) f_X(x) dx = \int_{-\infty}^{+\infty} \mathbb{E}(Y|X = x) f_X(x) dx \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y f_{Y|X=x}(y) f_X(x) dy dx = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y \frac{f(x,y)}{f_X(x)} f_X(x) dy dx \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y f(x,y) dy dx = \int_{-\infty}^{+\infty} y \left(\int_{-\infty}^{+\infty} f(x,y) dx\right) dy \\
&= \int_{-\infty}^{+\infty} y f_Y(y) dy = \mathbb{E}(Y).
\end{aligned}$$

Exemple 5.4.6 Reprenons l'Exemple 5.1.3 (voir aussi 5.2.2 et 5.4.6). On rappelle que la loi f de $Z = (X, Y)$ est la loi uniforme sur le triangle

$$T = \{(x, y) \in \mathbf{R}^2 \mid x \geq 0, y \geq 0, 0 \leq y \leq 1 - x\}.$$

Pour $x \in]0, 1[$, on a $f_X(x) = 2 - 2x > 0$ et

$$f_{Y|X=x}(y) = 2\frac{\mathbf{1}_T(x, y)}{2 - 2x} = \frac{1}{1 - x}\mathbf{1}_{[0,1-x]}(y)$$

La loi de Y sachant $X = x$ est donc la loi uniforme $\mathscr{U}([0, 1 - x])$. On a ainsi

$$\mathbb{E}(Y|X = x) = \frac{1 - x}{2}$$

et donc

$$\mathbb{E}(Y|X) = \frac{1 - X}{2}.$$

Rappelons (vor Exemple 5.3.3) que $\mathbb{E}(Y) = \mathbb{E}(X) = 1/3$. On a donc bien

$$\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}\left(\frac{1 - X}{2}\right) = \frac{1 - 1/3}{2} = 1/3 = \mathbb{E}(Y).$$

5.5 Sommes de variables aléatoires indépendantes

Soit $Z = (X, Y)$ un couple aléatoire à densité. Supposons que X et Y sont **indépendantes.**

Nous voulons déterminer la densité de la v.a.r. $X + Y$.

Proposition 5.5.1 *Soient* X *et* Y *des v.a.r. indépendantes, de densité f_X et f_Y respectivement. Alors* $X + Y$ *est une v.a.r. continue de densité f_{X+Y} donnée par*

$$\forall x \in \mathbf{R} : \qquad \boxed{f_{X+Y}(x) = \int_{-\infty}^{+\infty} f_X(x-y) f_Y(y) dy.}$$

Démonstration Soit f la densité de $Z = (X, Y)$. Rappelons (voir Proposition 5.2.3) que, comme X et Y sont indépendants, on a $f(x, y) = f_X(x) f_Y(y)$ pour tout $(x, y) \in \mathbf{R}^2$. Calculons la fonction de répartition F_{X+Y} de $X + Y$. Soit $t \in \mathbf{R}$. Considérons le demi-plan

$$D = \{(x, y) \in \mathbf{R}^2 \mid x + y \leq t\} = \{(x, y) \in \mathbf{R}^2 \mid x \leq t - y\}.$$

On a

$$\begin{aligned}
F_{X+Y}(t) &= \mathbf{P}(X + Y \leq t) \\
&= \int_D f(x, y) dx dy = \int_D f_X(x) f_Y(y) dx dy = \\
&= \int_{-\infty}^{+\infty} (\int_{-\infty}^{t-y} f_X(x) f_Y(y) dx) dy \\
&= \int_{-\infty}^{+\infty} f_Y(y) (\int_{-\infty}^{t-y} f_X(x) dx) dy \\
&\underset{(x=s-y)}{=} \int_{-\infty}^{+\infty} f_Y(y) (\int_{-\infty}^{t} f_X(s-y) ds) dy \\
&= \int_{-\infty}^{t} \left(\int_{-\infty}^{+\infty} f_X(s-y) f_Y(y) ds \right) dy
\end{aligned}$$

Il s'ensuit que la densité f_{X+Y} de $X + Y$ est $s \mapsto \int_{-\infty}^{+\infty} f_X(s-y) f_Y(y) ds$.■

Remarque 5.5.2 La fonction $x \mapsto \int_{-\infty}^{+\infty} f_X(x-y) f_Y(y) dy$ comme dans la Proposition 5.5.1 est appelé le **produit de convolution** des fonctions f_X et f_Y.

Exemple 5.5.3 Soient X et Y des v.a.r. indépendantes, suivant toutes deux une loi normale centrée-réduite $\mathcal{N}(0,1)$. On a donc

$$f_X(x) = f_Y(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

En se rappelant que $\int_{-\infty}^{+\infty} e^{-t^2} dt = \sqrt{\pi}$, on a, pour la densité f_{X+Y} de $X+Y$:

$$\begin{aligned}
f_{X+Y}(x) &= \int_{-\infty}^{+\infty} f_X(x-y) f_Y(y) dy \\
&= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-\frac{(x-y)^2}{2}} e^{-\frac{y^2}{2}} dy \\
&= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-\frac{(x-y)^2+y^2}{2}} dy \\
&= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-(\frac{x^2}{2}-xy+y^2)} dy \\
&= \frac{1}{2\pi} e^{-x^2/4} \int_{-\infty}^{+\infty} e^{-\left(y-\frac{x}{2}\right)^2} dy \\
&\underset{(t=y-x/2)}{=} \frac{1}{2\pi} e^{-x^2/4} \int_{-\infty}^{+\infty} e^{-t^2} dt \\
&= \frac{1}{2\pi} e^{-x^2/4} \sqrt{\pi} = \frac{1}{2\sqrt{\pi}} e^{-x^2/4}.
\end{aligned}$$

Ceci montre que $X+Y \sim \mathcal{N}(0,2)$. Plus généralement, avec un calcul similaire, on peut montrer que, si $X \sim \mathcal{N}(m_1, \sigma_1)$ et $Y \sim \mathcal{N}(m_2, \sigma_2)$, alors $X+Y \sim \mathcal{N}(m_1 + m_2, \sigma_1 + \sigma_2)$.

5.6 Changements de variables

Avant d'aborder le cas d'un couple aléatoire, traitons d'abord le cas d'une v.a.r. dans un cas particulier.

Proposition 5.6.1 *Soit* X *une v.a.r. admettant une densité* f_X. *Soit* $\varphi : \mathbf{R} \to \mathbf{R}$ *une bijection dérivable et strictement croissante (respectivement, décroissante). Alors la v.a.r.* $\varphi(X)$ *admet un densité* f_Y *donnée par*

$$\forall\, y \in \mathbf{R} : \qquad f_Y(y) = \frac{1}{\varphi'(\varphi^{-1}(y))} f_X(\varphi^{-1}(y))$$

(respectivement,

$$\forall\, y \in \mathbf{R} : \qquad f_Y(y) = -\frac{1}{\varphi'(\varphi^{-1}(y))} f_X(\varphi^{-1}(y)).)$$

Démonstration Soit $y \in \mathbf{R}$. Supposons que φ est croissante Alors, en notant F_X et F_Y les fonctions de répartition de X et Y, on a

$$F_Y(y) = \mathbf{P}(\varphi(X) \leq y) = \mathbf{P}(X \leq \varphi^{-1}(y)) = F_X(\varphi^{-1}(y)).$$

En dérivant F_Y, on obtient $f_Y(y) = \dfrac{1}{\varphi'(\varphi^{-1}(y))} f_X(\varphi^{-1}(y))$.

Supposons que φ est décroissante. Alors

$$F_Y(y) = \mathbf{P}(\varphi(X) \leq y) = \mathbf{P}(X \geq \varphi^{-1}(y)) = 1 - F_X(\varphi^{-1}(y)).$$

En dérivant F_Y, on obtient $f_Y(y) = -\dfrac{1}{\varphi'(\varphi^{-1}(y))} f_X(\varphi^{-1}(y))$.$\blacksquare$

Soit $Z = (X, Y)$ un couple aléatoire à densité f. Posons $D = \{(x, y \in \mathbf{R}^2 \mid f(x, y) > 0\}$ et soit

$$\varphi : D \to \mathbf{R}^2, \qquad (x, y) \mapsto (\varphi_1(x, y), \varphi_2(x, y)).$$

On cherche à determiner (sous certaines conditions) la densité du couple aléatoire $\varphi(X, Y)$.

On supposera que $\varphi : D \to \varphi(D)$ est bijective, que φ est continûment dérivable. Soit $J(\varphi)$ le déterminant de la matrice **matrice jacobienne** de φ, c-à-d

$$J(\varphi)(x, y) = \det \begin{pmatrix} \frac{\partial \varphi_1(x,y)}{\partial x} & \frac{\partial \varphi_2(x,y)}{\partial x} \\ \frac{\partial \varphi_1(x,y)}{\partial y} & \frac{\partial \varphi_2(x,y)}{\partial y} \end{pmatrix}$$

Proposition 5.6.2 *Soit $f_{(X,Y)}$ la densité du couple* (X, Y). *Posons* $T = \varphi_1(X, Y)$ *et* $U = \varphi_2(X, Y)$. *La densité $f_{(T,U)}$ du couple aléatoire* $(T, U) = \varphi(X, Y)$ *est donnée par la formule suivante :*

$$\boxed{f_{(T,U)}(t,u)) = \begin{cases} |J(\varphi^{-1})(t,u)|\, f_{(X,Y)}(\varphi^{-1}(t,u)) = \dfrac{1}{|J(\varphi)(\varphi^{-1}(t,u))|} f_{(X,Y)}(\varphi^{-1}(t,u)) & \textit{si } (t,u) \in \varphi(D) \\ 0 & \textit{sinon}\,. \end{cases}}$$

Démonstration Omise

Exemple 5.6.3 **(Coordonnées polaires)** Soit (X, Y) un couple aléatoire à densité $f_{(X,Y)}$. Soit (R, Θ) le couple aléatoire te que $X = R\cos\Theta$ et $Y = R\sin\Theta$. Il s'agit ici du changement de variables en coordonnées polaires : φ est une bijection dérivable entre $\mathbf{R}^2 \setminus \{0\}$ et $]0, +\infty[\times[0, 2\pi[$; l'expression de φ^{-1} est simple : $\varphi^{-1}(r, \theta) = (r\cos\theta, r\sin\theta)$. On a

$$J(\varphi^{-1})(r,\theta) = \det\begin{pmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{pmatrix} = r$$

La densité de (R, Θ) est donc

$$f_{(R,\Theta)}(r,\theta)) = \begin{cases} r\, f_{(X,Y)}(r\cos\theta, r\sin\theta) & \text{si } (r,\theta) \in \varphi(D) \\ 0 & \text{sinon,} \end{cases}$$

où $D = \{(x, y \in \mathbf{R}^2 \mid f_{(X,Y)}(x, y) > 0\}$.

Supposons, par exemple, que X et Y soient indépendantes, toutes deux de loi $\mathcal{N}(0, 1)$.La densité de (X, Y) est donnée par $f_{(X,Y)}(x, y) = \dfrac{1}{2\pi} e^{-(x^2+y^2)/2}$ Alors la densité de (R, Θ) est

$$f_{(R,\Theta)}(r,\theta)) = \frac{1}{2\pi} r e^{-r^2/2}$$

pour $(r, \theta) \in]0, +\infty[\times[0, 2\pi[$.

5.7 Exercices

Exercice 5.7.1 Pour $n \geq 1$, on considère une v.a.r X_n suivant la loi uniforme sur $\{k/n \ : \ k = 0, \ldots, n\}$. Montrez que $X_n \xrightarrow{\mathscr{L}} X$ où X suit une loi uniforme sur $[0,1]$.

Exercice 5.7.2 Supposons que vous ayez en moyenne deux accrochages par an avec votre voiture. Devez-vous prévoir plus de deux constats à l'amiable pour les prochains 18 mois ?

Exercice 5.7.3 Dans un bureau de vote, une urne contient 1000 bulletins, dont 5% sont nuls. On prend 100 bulletins au hasard. Soit X la v.a.r égale au nombre de bullletins nuls contenus dans cet échantillon.
(i) Décrire la loi de X.
(ii) Approcher la loi de X par une loi binomiale, puis par une loi de Poisson et en déduire une valeur approchée des probabilités $\mathbf{P}(X \leq 5)$ et $\mathbf{P}(X = 0)$.

Exercice 5.7.4 On lance un dé jusqu'à ce que la somme des points obtenus soit supérieure à 300. Utiliser le théorème central limite pour estimer la probabilité qu'au moins 80 lancers soient nécessaires.

Exercice 5.7.5 Pour $n \in \mathbf{N}$, on pose $I_n = \int_0^{+\infty} x^n e^{-x} dx$.
(i) Justifier l'existence de I_n et établir une relation de récurrence entre I_{n+1} et I_n pour $n \in \mathbf{N}$.
(ii) Calculer I_n pour tout $n \in \mathbf{N}$.
Soit X une v.a.r. qui suit une loi exponentielle de paramètre $\lambda > 0$.
(iii) Calculer, pour tout $n \geq 1$, le moment $\mathbb{E}(X^n)$ d'ordre n de X.

Exercice 5.7.6 Des clients arrivent à un guichet de manière aléatoire. On suppose qu'il existe $\alpha > 0$ tel que, pour tout $t > 0$, le nombre de clients arrivant entre les instants 0 et $t > 0$ est une v.a.r. N_t qui suit une loi de Poisson de paramètre αt.
Soit X_1 l'instant d'arrivée du premier client.
(i) Déterminer la probabilité $\mathbf{P}(X_1 > t)$; en déduire que X_1 est une v.a. continue qui suit une loi exponentielle.
(ii) Calculer $\mathbb{E}(X_1)$ et $\mathrm{Var}(X_1)$.
Pour $n \geq 1$, soit X_n l'instant d'arrivée du n-ième client.
(iii) Déterminer la probabilité $\mathbf{P}(X_n > t)$ et en déduire la fonction de répartition de X_n.
(iv) Montrer que X_n est une v.a. continue et en déterminer une densité.
(v) En utilisant l'Exercice 5.7.5, calculer $\mathbb{E}(X_n)$ et $\mathrm{Var}(X_n)$.

Exercice 5.7.7 Soit $(\Omega, \mathscr{F}, \mathbf{P})$ un espace probabilisé ; on suppose qu'il existe $\Omega_1, \Omega_2 \in \mathscr{F}$ tels $\Omega = \Omega_1 \cup \Omega_2$, $\Omega_1 \cap \Omega_2 = \emptyset$ et $\mathbf{P}(\Omega_1) = \mathbf{P}(\Omega_2)$.

Pour $n \geq 1$, on considère les v.a.r. $X_{2n} = \mathbf{1}_{\Omega_1}$ et $X_{2n+1} = \mathbf{1}_{\Omega_2}$.

(i) Déterminer la loi de X_n.
(ii) Déduire de (i) que $X_n \xrightarrow{\mathscr{L}} X_1$.
(iii) Calculer $\mathbf{P}(|X_{2n} - X_1| \geq 1)$.
(iv) Déduire de (iii) que $(X_n)_n$ ne converge pas en probabilité vers X_1.

Exercice 5.7.8 La durée de vie T d'une particule radioactive est une v.a.r qui suit une loi exponentielle de paramètre $\lambda > 0$. On considère n particules dont les durées de vies $T_1, T_2, \ldots, T_n$ sont indépendantes. On note S_n le temps de désintégration de la 1ère particule qui se désintègre et R_n celui de la dernière.
(i) Déterminer les fonctions de répartitions de S_n et R_n.
(ii) Etudier la convergence en loi des suites $(nS_n)_{n\geq 1}$ et $\left(\frac{R_n}{n}\right)_{n\geq 1}$.
(iii) Montrer que la suite $(\overline{T}_n)_{n\geq 1}$ définie par $\overline{T}_n = \frac{1}{n}\sum_{k=1}^{n} T_k$ converge en probabilité vers la v.a.r. constante égale à $\frac{1}{\lambda}$.

Exercice 5.7.9 Soit X une v.a.r. suivant une loi binomiale $\mathscr{B}(1000, 0.03)$.
(i) Donner une approximation de $\mathbf{P}(X \leq 20)$, en approchant la loi de X par une loi de Poisson.
(ii) Donner une approximation de $\mathbf{P}(X \leq 20)$, en approchant la loi de X par une loi normale.
(iii) Essayer de calculer $\mathbf{P}(X \leq 20)$ avec votre calculatrice.

Exercice 5.7.10 Soient (X_n) et $(Y_n)_n$ des suites de v.a.r. sur l'espace de probabilité $(\Omega, \mathscr{F}, \mathbf{P})$ convergeant en probabilité vers deux v.a.r X et Y respectivement.
(i) Soit $\varepsilon > 0$. Montrer que l'évènement $(|(X_n + Y_n) - (X + Y)| \geq \varepsilon)$ entraîne l'évènement $(|X_n - X| \geq \varepsilon/2) \cup (|Y_n - Y| \geq \varepsilon/2)$.
(ii) Soient $a, b \in \mathbf{R}$. Montrer que $(aX_n + bY_n)_n$ converge en probabilité vers $aX + bY$.
(iii) On suppose qu'il existe une autre v.a.r X' telle que $(X_n)_n$ converge en probabilité vers X'. Déduire de (ii) que $\mathbf{P}(X = X') = 1$.
(iv) On suppose qu'il existe une constante $C > 0$ tel que, pour tout n, on ait $|X_n| \leq C$ et $|Y_n| \leq C$. Montrer que $(X_nY_n)_n$ converge en probabilité vers XY.

Chapitre 6

Éléments de statistique

Il n'est plus nécessaire aujourd'hui de se convaincre de l'importance de la statistique, tant dans notre vie quotidienne que dans les activités économiques, politiques ou scientifiques. La statistique s'intéresse à des caractéristiques d'une population au sens large (ensemble de personnes ou objets équivalents) ; elle recueille organise et traite les données (statistiques) obtenus à l'aide d'un échantillon issu de cette population pour les rendre utilisables (statistique descriptive) ; elle fournit des outils pour cerner les caractéristiques de la population-mère, en confrontant les données à des modèles probabilistes (statistique inférentielle). On cherche à étendre (inférer) les propriétés constatées sur l'échantillon à la population, en validant ou en infirmant des hypothèses formulées sur la population.

6.1 Statistique descriptive

L'ensemble de personnes ou d'objets équivalents étudié s'appelle la *population*; c'est en général un ensemble fini Ω, dont le cardinal s'appelle la *taille* de la population.

La série d'observations recueillies s'appelle *série statistique série de données* ou bien *échantillon.* Elle est général présentée sous la forme d'un *tableau de données.*

Définition 6.1.1 (i) Chaque objet d'une population s'appelle un *individu statistique* ou unité statistique.
(ii) L'ensemble des individus sur lesquels porte l'observation est appelé *echantillon.*

(iii) La grandeur ou caractéristique observé ou recueillie pour chaque individu est appelée *variable statistique* ou *caractère statistique.*
(iv) Il existe différents types de variables statistiques

- *Variables quantitatives :* ce sont des caractéristiques numériques (taille, poids age,...). Ils s'expriment par des nombres réels sur lesquels les opérations arithmétiques de base (somme, moyenne,...) ont un sens.
  Ces variables peuvent être *discrètes* : leurs valeurs sont prises dans nombre fini ou dénombrable de (comme par exemple l'âge) ; elles peuvent être *continues :* toutes les valeurs réelles sont susceptibles d'être prises (comme par exemple la taille) et peuvent être regroupées en sous-intervalles, appelées *classes*;
- *Variables qualitatives :* ce sont des caractéristiques non numériques dans le sens où les opérations arithmétiques de base n'ont pas de sens pour elles. Elles peuvent être *nominales* (sexe, couleur,...) ou *ordinales* lorsque l'ensemble des catégories est muni d'un ordre total (défavorable, favorable, très favorable,..). Les différentes valeurs d'une variable qualitative s'appellent des *modalités* (ou catégories).

6.1.1 Description d'une série de données

On présente les données en général sous forme d'un tableau, où $i = 1\dots,n$ désigne l'individu et x_i le caractère observé sur l'individu i :

i	1	$\dots$	i	$\dots$	n
x	x_1	$\dots$	x_i	$\dots$	x_n

Exemple 6.1.2 Voici une liste de de 10 notes d'examen $x_1,\dots,x_{10}$, attribuées à un groupe de 10 étudiants avec les numéros $i = 1,\dots 10$.

No	1	2	3	4	5	6	7	8	9	10
x_i	5	10	12	10	20	14	15	10	5	19

Les données statistiques brutes ne permettent pas de déduire des informations pertinentes sans un traitement préalable. On regrouper les données par effectifs ou bien par fréquence.

Définition 6.1.3 (Effectifs) Soit $x_1,\dots,x_n$ une série statistique.
(i) L'*effectif* d'une valeur (ou d'une modalité) x_i est le nombre n_i d'individus possédant la valeur x_i ; la *fréquence* de x_i est n_i/n, le quotient de l'effectif correspondant par l'effectif total de l'échantillon.

Soient $\xi_1, \dots, \xi_k$ les différentes valeurs prises par la série.
(ii) Pour chaque $1 \le j \le k$, la somme $n_1 + \cdots + n_j$ est appelée *effectif cumulé* et $n_1/n + \cdots + n_j/n$ est appelée *fréquence cumulée*.

Exemple 6.1.4 Pour l'Exemple 6.1.2 plus haut, on obtient le tableau suivant :

x_i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Effectifs	0	0	0	0	0	2	0	0	0	0	3	0	1	0	1	1	0	0	0	1	1
Fréquences Cumulées.	0	0	0	0	0	2	0	0	0	0	5	0	6	0	7	8	0	0	0	9	10

On peut regrouper les effectifs par classes :

x_i	$[0,5[$	$[5,10[$	$[10,15[$	$[15,20[$
Effectifs	0	2	5	3
Fréquences Cumulées	0	2	7	10

On visualise le tableau de données par un graphique :

— diagramme en bâton : on fait figurer sur un axe horizontal les individus $i = 1, \dots, n$ et on trace une barre verticale proportionnelle à l'effectif n_i ou à la fréquence n_i/n des individus ayant le caractère x_i, où n est le cardinal de l'échantillon.

— histogramme : pour des caractères quantitatifs regroupés en classes $[a_1, a_2[, \dots, [a_{k-1}, a_k[$, on trace un au dessus de chaque intervalle $[a_i, a_{i+1}[$ un rectangle dont l'aire est proportionnelle à l'effectif.

— diagramme circulaire ("camembert"),

— ....

6.1.2 Indicateurs numériques

Soit $\{x_1, x_2, \dots, x_n\}$ une série de données.

Définition 6.1.5 Moyenne, médiane (i) L'*étendue* de la série est l'écart $x_{\max} - x_{\min}$ entre la valeur maximale et la valeur minimale de l'échantillon.
(ii) On appelle *mode* (ou *classe modale*) toute valeur x_i de fréquence maximale. Si cette classe est unique, on dit que la série est *unimodale*. Sinon, on parle de série *plurimodale*.
(iii) La *moyenne* de l'échantillon est

$$\boxed{\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,}$$

Si $\xi_1, \dots, \xi_k$ sont les valeurs prises par un échantillon de données, alors

$$\boxed{\bar{x} = \frac{1}{n}\sum_{j=1}^{k} n_j \xi_j,}$$

où n_j est l'effectif de la valeur ξ_j.
(iv) La *médiane* de l'échantillon $\{x_1, x_2, \dots, x_n\}$ est toute valeur x_m telle que

$$\mathrm{Card}\{1 \leq i \leq n \ : x_i \leq x_m\} = \mathrm{Card}\{1 \leq i \leq n \ : x_i \geq x_m\};$$

elle permet de diviser l'échantillon en deux : une moitié des individus a des valeurs $\leq x_m$ et l'autre moitié des valeurs $\geq x_m$. si les données sont ordonnées $x_1 \leq x_2 \leq \cdots \leq x_n$, alors $x_m = x_{q+1}$ si $n = 2q+1$ est impair et x_m est n'importe quelle valeur dans $]x_q, x_{q+1}[$ si $n = 2q$ est pair.

Exemple 6.1.6 Dans l'Exemple **??** plus haut, l'étendue de l'échantillon est 16, la moyenne est $\bar{x} = 11,5$ et la médiane $x_m = 9$.

La moyenne (ou médiane) est un premier indicateur numérique de la série statistique ; on introduit la variance comme paramètre de dispersion.

Définition 6.1.7 (Variance, écart-type) La *variance* de la série statistique $\{x_1, \dots, x_m\}$ est

$$\boxed{s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{j=1}^{k} n_j(\xi_j - \bar{x})^2,}$$

où $\xi_1, \dots, \xi_k$ sont les valeurs prises par un échantillon de données.
L'*écart-type* de l'échantillon est s, la racine carrée de la variance.

Remarque 6.1.8 On a ici aussi une formule de Koenig-Huygens, similaire à celle de la Proposition 3.1.25 :

$$s^2 = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \bar{x}^2.$$

Exemple 6.1.9 Dans l'Exemple 6.1.2 plus haut, la variance est $s^2 = 30.35$ et l'écart-type $s \simeq 5.5$

La médiane d'un échantillon correspond à un découpage de cet échantillon en deux effectifs égaux. On peut affiner ce découpage et considérer un découpage par quart.

Définition 6.1.10 (Quartiles) Le *1er quartile* de la série statistique $\{x_1, \dots, x_m\}$ est la plus petite valeur x_i, notée q_1,

$$\mathrm{Card}\{1 \leq i \leq n \, : x_i \leq q_1\} \geq \frac{n}{4}$$

le *3ème quartile* de la série statistique $\{x_1, \dots, x_m\}$ est la plus petite valeur x_i, notée q_3,

$$\mathrm{Card}\{1 \leq i \leq n \, : x_i \leq q_3\} \geq \frac{3n}{4}$$

L'intervalle $[q_1, q_3]$ est appelé *intervalle interquartile* et $q_3 - q_1$ *écart interquartile.*

Exemple 6.1.11 Dans l'Exemple 6.1.2 plus haut, le 1er quartile est $q_1 = 10$, 3ème quartile est $q_3 = 15$ et l'écart interquartile est 5.

6.2 Estimation

On modélise la population à étudier par un espace probabilisé $(\Omega, \mathscr{F}, \mathbf{P})$; un caractère quantitatif de cette population correspond à une v.a.r X sur Ω; par exemple, Ω pourrait être l'ensemble des particpants à une élection, muni de la mesure uniforme, et X le nombre de votes obtenus par un candidat particulier.

Un echantillon de taille n de X est la donnée $X_1, \dots, X_n$ de n v.a.r. indépendantes et de même loi que X; l'échantillon observé $x_1, \dots, x_n$ est interprété comme étant une réalisation de $X_1, \dots, X_n$, c-à-d

$$(X_1(\omega), \dots, X_n(\omega)) = (x_1, \dots, x_n)$$

pour un $\omega \in \Omega$; dans l'exemple de l'élection, $x_1, \dots, x_n$ correspond au résultat d'un sondage, obtenu en interrogeant n électeurs choisis au hasard.

On désire obtenir des renseignements sur les paramètres de la loi de X (espérance, variance, ...) à travers les observations $x_1, \dots, x_n$ de l'échantillon $X_1, \dots, X_n$ de X. On introduit des estimateurs pour ces paramètres, en l'occurrence $\mu = \mathbb{E}(X)$ l'espérance de X et $\sigma^2 = \mathrm{Var}(X)$ la variance de X.

On appelle *statistique* sur $X_1, \dots, X_n$ toute fonction $f(X_1, \dots, X_n)$ de $(X_1, \dots, X_n)$, qui est donc une v.a.r.

Définition 6.2.1 (Estimateurs de la moyenne et de la variance) On appelle *moyenne empirique* et *variance empirique* les statistiques $\overline{X}$ et S^2 définies par

$$\boxed{\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i} \qquad \text{et} \qquad \boxed{S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.}$$

On introduit aussi la *variance empirique modifiée* S'^2 par

$$\boxed{S'^2 = \frac{n}{n-1}S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.}$$

La proposition suivante montre que $\overline{X}$, S^2 et $S'2$ sont de bons estimateurs de la moyenne μ et de la variance σ^2 de la v.a.r X.

Proposition 6.2.2 *(i) On a*

$$\mathbb{E}(\overline{X}) = \mu \qquad et \qquad \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}.$$

(ii) On suppose que X *possède un moment* μ_4 *d'ordre* 4; *alors*

$$\mathbb{E}(S^2) = \frac{n-1}{n}\mu \qquad et \qquad \mathrm{Var}(S^2) \sim_{n\to+\infty} \frac{\mu^4 - \sigma^3}{n}.$$

et

$$\mathbb{E}(S'^2) = \mu \qquad et \qquad \mathrm{Var}(S'^2) = (n/n-1)^2 S^2 \sim_{n\to+\infty} \frac{\mu^4 - \sigma^3}{n}.$$

Démonstration (i) On a, par linéarité de l'espérance,

$$\mathbb{E}(\overline{X}) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu.$$

Comme $X_1, \ldots, X_n$ sont indépendants, on a

$$\mathrm{Var}(\overline{X}) = \sum_{i=1}^{n}\mathrm{Var}(X_i/n) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n}.$$

(ii) On a

$$\begin{aligned} n\mathrm{S}^2 &= \sum_{i=1}^{n} (\mathrm{X}_i - \overline{\mathrm{X}})^2 = \sum_{i=1}^{n} \Big((\mathrm{X}_i - \mu) - (\overline{\mathrm{X}} - \mu) \Big)^2 \\ &= \sum_{i=1}^{n} (\mathrm{X}_i - \mu)^2 + \sum_{i=1}^{n} (\overline{\mathrm{X}} - \mu)^2 - 2(\overline{\mathrm{X}} - \mu) \sum_{i=1}^{n} (\mathrm{X}_i - \mu) \\ &= \sum_{i=1}^{n} (\mathrm{X}_i - \mu)^2 + n(\overline{\mathrm{X}} - \mu)^2 - 2n(\overline{\mathrm{X}} - \mu)^2 \\ &= \sum_{i=1}^{n} (\mathrm{X}_i - \mu)^2 - n(\overline{\mathrm{X}} - \mu)^2 \end{aligned}$$

et donc, par linéarité de l'espérance,

$$n\mathbb{E}(\mathrm{S}^2) = \sum_{i=1}^{n} \mathbb{E}((\mathrm{X}_i - \mu)^2) - n\mathbb{E}((\overline{\mathrm{X}} - \mu)^2) = n\sigma^2 - n\frac{\sigma^2}{n} = (n-1)\sigma^2.$$

L'assertion concernant $\mathrm{Var}(\mathrm{S}^2)$ sera admise ■

Voici une conséquence éminemment importante du TCL (Théorème 4.6.1).

Théorème 6.2.3 *(i) Soit* Z *une v.a.r de loi normale centre réduite* $\mathcal{N}(0,1)$ *; alors*

$$\frac{\overline{\mathrm{X}} - \mu}{\sigma/\sqrt{n}} \xrightarrow[n\to+\infty]{\mathscr{L}} \mathrm{Z}.$$

(ii) On a $\mathrm{S}'^2 \xrightarrow[n\to+\infty]{\mathbf{P}} \sigma^2$.

Démonstration (i) L'assertion découle immédiatement du TCL (Théorème 4.6.1), combiné au fait que $\mathbb{E}(\overline{\mathrm{X}}) = \mu$ et au fait que $\mathrm{Var}(\overline{\mathrm{X}}) = \dfrac{\sigma^2}{n}$.

(ii) Par l'inégalité de Bienaymé-Tchebychev (Proposition 4.1.1), on a, pour tout $\varepsilon > 0$,

$$\mathbf{P}(|\mathrm{S}'^2) - \mathbb{E}(\mathrm{S}'^2)| \geq \varepsilon) \leq \frac{\mathrm{Var}(\mathrm{S}'^2)}{\varepsilon^2}.$$

Comme $\mathbb{E}(\mathrm{S}'^2) = \sigma^2$ et

$$\mathrm{Var}(\mathrm{S}'^2) \sim_{n\to+\infty} \frac{\mu^4 - \sigma^3}{n} \to_{n\to+\infty} = 0,$$

l'assertion en découle. ■

Soit $X_1, \dots, X_n$ un échantillon de taille n de la v.a.r X, de moyenne μ et de variance σ^2.

Rappelons que la moyenne empirique et la variance empirique (modifiée) sont

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{et} \qquad S'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

6.2.1 Estimation d'une moyenne : cas d'une variance connue

Soit Z une v.a.r de loi normale $\mathcal{N}(0,1)$ et soit $\alpha \in]0,1[$ fixé ("seuil de risque"). On note z_α le nombre réel ≥ 0 tel que $\Pi(-z_\alpha) = \alpha$, c-à-d $\mathbf{P}(Z \geq z_\alpha) = \alpha$ ou bien $\mathbf{P}(Z \leq z_\alpha) = 1-\alpha$. La valeur z_α est lue sur la table de la loi normale. Alors

$$\mathbf{P}(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = \Pi(z_{\alpha/2}) - \Pi(-z_{\alpha/2}) = 2\Pi(z_{\alpha/2}) - 1 = 1-\alpha.$$

Pour n suffisamment grand (dans la pratique $n \geq 30$ suffit), on peut considérer, par le Théorème 6.2.3 que

$$Z := \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1).$$

D'où

$$\begin{aligned} 1-\alpha &= \mathbf{P}(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) \\ &= \mathbf{P}(-z_{\alpha/2} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}) \\ &= \mathbf{P}(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}). \end{aligned}$$

On suppose que la variance σ est connue. L'intervalle de confiance, au risque α, est alors donné par

$$\boxed{\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}},}$$

c-à-d

$$\boxed{I_\alpha = [\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}].}$$

Ceci signifie que $\mu \in I_\alpha$ avec probabilité $1-\alpha$. Pour $\alpha = 0.05$, on a $z_{\alpha/2} = 1.96$ et donc $\mu \in [\overline{x} - 1.96\frac{\sigma}{\sqrt{n}}, \overline{x} + 1.96\frac{\sigma}{\sqrt{n}}]$ avec probabilité de 95%

6.2.2 Estimation d'une moyenne : cas d'une variance inconnue

On suppose que la variance σ n'est pas connue.

Définition 6.2.4 (i) **(Loi du** χ^2**)** Soit $Z_1, \dots, Z_\nu$ des v.a.r indépendantes de même loi normale $\mathcal{N}(0,1)$; la v.a.r

$$Z^2 = Z_1^2 + \cdots + Z_\nu^2$$

possède une loi, appelée *loi du Khi-deux* à ν degrés de liberté et notée $\chi^2(\nu)$. La v.a.r. Z^2 possède une densité f donnée par

$$f(x) = \mathbf{1}_{[0,+\infty[}(x) \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\frac{\nu}{2}-1}\, e^{-\frac{x}{2}},$$

où Γ est la fonction Γ d'Euler définie par

$$\Gamma(r) = \int_0^{+\infty} x^{r-1} e^{-x} dx \qquad \text{pour tout} \quad r > 0.$$

On a $\mathbb{E}(Z^2) = \nu$ et $\mathrm{Var}(Z^2) = 2\nu$.
(i) **(Loi de Student)** Soit Z^2 une v.a.r. de loi $\chi^2(\nu)$ et soit Y une v.a.r de loi $\mathcal{N}(0,1)$; alors la v.a.r

$$T = \frac{Y}{\sqrt{Z^2/\nu}},$$

suit une loi appelée *loi de Student* à ν degrés de liberté. La v.a.r. T possède une densité f donnée par

$$f(x) = \frac{1}{\sqrt{\pi\nu}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\nu/2)} \frac{1}{(1+x^2/\nu)^{\frac{\nu+1}{2}}}.$$

On a $\mathbb{E}(T) = 0$ et $\mathrm{Var}(T) = \frac{\nu}{\nu-2}$ pour $n > 2$.

Les lois du χ^2 et de Student sont tabulées dans la littérature (voir Annexe 8, à la fin de ces notes).

En revenant à notre échantillon $X_1, \dots, X_n$, on remplace la variance inconnue σ^2 par la variance empirique (modifiée) S'^2. On démontre que

$$Z^2 = (n-1)S'^2/\sigma^2$$

suit une loi $\chi^2(n-1)$ et que donc

$$\frac{\overline{X}-\mu}{S'/\sqrt{n}} = \left(\frac{\overline{X}-\mu)}{\sigma/\sqrt{n}}\right)/\sqrt{Z^2/(n-1)}$$

suit une loi de Student à $n-1$ degrés de liberté.

Pour $\alpha \in]0,1[$ fixé, soit $t_{\alpha/2} = t_{\alpha/2;n-1}$ le nombre réel ≥ 0 tel que

$$\mathbf{P}(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) = 1-\alpha,$$

pour une variable de Student T à $n-1$ degrés de liberté..

L'intervalle de confiance, au risque α, est donné par

$$\boxed{\overline{x} - t_{\alpha/2}\frac{s'}{\sqrt{n}} \leq \mu \leq \overline{x} + t_{\alpha/2}\frac{s'}{\sqrt{n}},}$$

c-à-d

$$\boxed{I_\alpha = [\overline{x} - t_{\alpha/2}\frac{s'}{\sqrt{n}}, \overline{x} + t_{\alpha/2}\frac{s'}{\sqrt{n}}],}$$

ou $s'^2 = \dfrac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2$ est la variance empirique modifiée de l'échantillon.

6.2.3 Estimation d'une fréquence

On suppose la v.a.r X modélise une proportion inconnue $p \in]0,1[$, c-à-d $X \sim \mathscr{B}(p)$. Alors $\overline{X} = \dfrac{1}{n}\sum_{i=1}^n X_i \sim \mathscr{B}(n,p)$; on a donc

$$\mathbb{E}(\overline{X}) = p \qquad \text{et} \qquad \mathrm{Var}(\overline{X}) = \frac{pq}{n}.$$

On remplace $\mathrm{Var}(\overline{X})$ par

$$s'^2 = \frac{f(f-1)}{n-1},$$

où $f = \bar{x}$ est la fréquence observée. Pour n suffisamment grand ($n \geq 30$), on peut considérer que la v.a.r

$$Z := \frac{\overline{X}-p}{s'}$$

suit approximativement une loi normale $\mathcal{N}(0,1)$

Soit $\alpha \in]0,1[$ fixé (“seuil de risque"). Soit $z_{\alpha/2}$ le nombre réel ≥ 0 tel que

$$\mathbf{P}(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha.$$

Alors

$$\begin{aligned}1 - \alpha &= \mathbf{P}(-z_{\alpha/2} \leq Z = \frac{\overline{X} - p}{s'} \leq z_{\alpha/2}) \\ &= \mathbf{P}(f - z_{\alpha/2}\sqrt{\frac{f(f-1)}{n-1}} \leq p \leq f + z_{\alpha/2}\sqrt{\frac{f(f-1)}{n-1}}).\end{aligned}$$

L'intervalle de confiance, au risque α, est donné par

$$\boxed{f - z_{\alpha/2}\sqrt{\frac{f(f-1)}{n-1}} \leq p \leq f + \sqrt{\frac{f(f-1)}{n-1}},}$$

c-à-d

$$\boxed{I_\alpha = [f - z_{\alpha/2}\sqrt{\frac{f(f-1)}{n-1}}, f + z_{\alpha/2}\sqrt{\frac{f(f-1)}{n-1}}].}$$

Lorsque n est grand, on peut négliger la différence entre n et $n-1$ et écrire plus simplement

$$\boxed{I_\alpha = [f - z_{\alpha/2}\sqrt{\frac{f(f-1)}{n}}, f + z_{\alpha/2}\sqrt{\frac{f(f-1)}{n}}].}$$

Ceci signifie que $p \in I_\alpha$ avec probabilité $1-\alpha$. Pour $\alpha = 0.05$ on a $z_{\alpha/2} = 1.96$ et $p \in [f - 1.96\sqrt{\frac{f(f-1)}{n}}, f + 1.96\sqrt{\frac{f(f-1)}{n}}]$ avec une probabilité de 95%.

Exemple 6.2.5 Un sondage portant sur 1000 électeurs a donné 530 votes pour le candidat A. On a donc

$$f = \frac{530}{1000} \simeq 0.53, \qquad \text{et} \qquad s = \frac{f(f-1)}{n} \simeq 0.016$$

et l'intervalle de confiance au risque $\alpha = 5\%$ est

$$[0.53 - 1.96 \times 0.016, 0.53 + 1.96 \times 0.016] = [0.50, 0.56];$$

on a donc une fourchette entre 50% et 56% dans la prédiction de votes pour le candidat A, avec probabilité de 95%.

6.3 Tests d'hypothèse

On souhaite, à partir d'un échantillon observé $x_1, \ldots, x_n$, contredire ou au contraire confirmer une certaine hypothèse formulée sur la loi X de la population (le nouveau traitement médical apporte-t-il une amélioration de l'état de santé des malades par rapport à l'ancien ?; le nouveau carburant améliore-t-il le rendement du moteur ?...). L'hypothèse est appelé *hypothèse nulle* et est notée H_0 (dans les deux exemples, H_0 pourrait être : il n' y a pas de changement par rapport à l'ancien traitement ou l'ancien carburant). Elle est confrontée à une hypothèse alternative H_1 (dans les deux exemples, H_1 pourrait être : il y a une amélioration de l'état de santé des malades ou du rendement). On souhaite avoir "peu de chance" de rejeter H_0 s'il elle est vraie ; cette erreur, appelée erreur de première espèce, est estimée à travers un seuil de risque $\alpha > 0$, avec souvent $\alpha = 0.05 = 5\%$.

6.3.1 Test d'une moyenne

Soit $\overline{x}$ la moyenne de l'échantillon observé $x_1, \ldots, x_n$. On désire savoir si cette moyenne est conforme à la moyenne théorique annoncée μ. L'hypothèse H_0 est

$$H_0 : \text{ la moyenne théorique est } \mu.$$

Supposons qu'on choisisse

$$H_1 : \text{ la moyenne théorique est } < \mu.$$

Sous l'hypothèse H_0, l'espérance $\mathbb{E}(X)$ de la population est μ. Pour n suffisamment grand, on a

$$Z = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim \mathcal{N}(0,1).$$

D'où

$$\mathbf{P}\left(\frac{\overline{X} - \mu}{S/\sqrt{n}} \leq -z_\alpha\right) = \mathbf{P}(\overline{X} \leq \mu - z_\alpha \sqrt{n}/S) = \alpha.$$

On rejette H_0 si $\overline{x} \leq \mu - z_\alpha \sqrt{n}/s$; la zone

$$\{\frac{\overline{X} - \mu}{S/\sqrt{n}} \leq -z_\alpha\} = \{\overline{X} \leq \mu - z_\alpha \sqrt{n}/S\}$$

est appelé *zone de rejet.*

Voici les différentes étapes dans la mise en oeuvre d'un test.

Mise en oeuvre d'un test

1. On choisit les hypothèses H_0 et H_1 ;
2. on choisit le seuil α et on détermine une zone de rejet de H_0;
3. on décide au vu de l'échantillon du rejet ou non de l'hypothèse H_0.

Exemple 6.3.1 Le temps de fabrication d'une certaine pièce dans un usine dure en moyenne 270 s avec un écart-type de 24 s. On introduit une modification technique du montage susceptible diminuer ce temps. Pour le tester, on effectue $n = 38$ mesures avec cette modification technique et on calcule la moyenne empirique $\overline{x}$ du temps de fabrication. On choisit comme test $H_0 : \mu = 270$ et comme hypothèse alternative $H_1 : \mu < 270$. Pour $\alpha = 0.05$, on a $-z_{0.05} = -1.645$ et

$$\mathbf{P}(Z \leq -1.645) = \mathbf{P}\left(\frac{\overline{X} - 270}{24/\sqrt{38}} \leq -1.645\right) = 0.05,$$

c-à-d

$$\mathbf{P}(\overline{X} \leq (24 \times (-1.645)/\sqrt{38}) + 270) = \mathbf{P}(\overline{X} \leq 263.6) = 0.05.$$

On choisit comme zone de rejet : $\{\overline{X} \leq 263.6\}$

Si, par exemple, on a observé une moyenne empirique $\overline{x} = 267$, alors on accepte donc H_0 et on considère qu'il n'a pas d'amélioration du temps de fabrication.

Supposons, par exemple, qu'on ait trouvé $\overline{x} = 260$. On rejette H_0 et on considère qu'il y a une amélioration du temps de fabrication.

On peut même renforcer le test : on a $\dfrac{\overline{x} - 270}{24/\sqrt{38}} = -2.57$ et la valeur (dite *p*-valeur) $\mathbf{P}(Z \leq -2.57)$ est 0.005. Ceci signifie que, si H_0 est vraie (c-à-d $\mu = 270$), il y a une probabilité extrêmement faible (égale à 0.005) d'observer la valeur $\overline{x} = 260$.

Remarque 6.3.2 Dans l'exemple plus haut, nous avons choisi un *test unilatéral* : la zone de rejet était du type $\{\overline{X} \leq c\}$. Ceci était raisonnable, car il s'agissait de tester

$$H_0 : \mathbb{E}(X) = \mu \quad \text{contre} \quad H_1 : \mathbb{E}(X) < \mu.$$

Dans d'autres problèmes, on peut être amené à tester une hypothèse du type

$$H_0 : \mathbb{E}(X) = \mu \quad \text{contre} \quad H_1 : \mathbb{E}(X) > \mu$$

ou bien

$$H_0 : \mathbb{E}(X) = \mu \quad \text{contre} \quad H_1 : \mathbb{E}(X) \neq \mu.$$

Pour un risque d'erreur α, on prendra dans le 1e cas une zone de rejet du type $\{\overline{X} \geq c\}$ correspondant à une zone $\{Z \geq z_\alpha\}$ et dans le 2e cas, appelé *test bilatéral*, une zone de rejet du type $\{|\overline{X}| \geq c\}$ correspondant à une zone $\{|Z| \geq z_{\alpha/2}\}$. Dans tous ces cas, on a donc

$$\mathbf{P}(H_0 \text{ est rejetée alors qu'elle est vraie }) = \alpha.$$

6.3.2 Test de comparaison à une fréquence théorique

Soit $\overline{x}$ la fréquence d'un certain caractère observée sur un échantillon de taille n. On désire tester si $\overline{x}$ est conforme à une fréquence théorique annoncée p. On adopte l'hypothèse

$$H_0 : \text{ la fréquence est } p.$$

Supposons qu'on choisisse (test bilatéral)

$$H_1 : \text{ la fréquence est } \neq p.$$

Sous l'hypothèse H_0, on a pour n suffisamment grand

$$Z = \frac{\overline{X} - p}{\sqrt{p(1-p/n}} \sim \mathcal{N}(0,1).$$

Pour un seuil de risque α, on a

$$\mathbf{P}\left(\frac{|\overline{X} - p|}{\sqrt{p(1-p)/n}} \geq z_{\alpha/2}\right) = \alpha.$$

La zone de risque est alors

$$\boxed{\left\{\frac{|\overline{X} - p|}{\sqrt{p(1-p)/n}} \geq z_{\alpha/2}\right\},}$$

c-à-d la zone

$$\left\{\overline{X} \geq p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right\} \bigcup \left\{\overline{X} \leq p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right\}$$

Si on choisit

$$\mathrm{H}_1: \text{ la fréquence est } > p,$$

on a un test unilatéral avec une zone de risque

$$\boxed{\left\{\frac{\overline{\mathrm{X}} - p}{\sqrt{p(1-p)/n}} \geq z_{\alpha/2}\right\},}$$

c-à-d la zone

$$\left\{\overline{\mathrm{X}} \geq p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right\}.$$

Exemple 6.3.3 A la veille d'une élection entre deux candidats A et B, un sondage portant sur $n = 1021$ électeurs a donné 520 électeurs en faveur de A. Peut on considérer que le résultat de l'élection est complètement indécis? Soit p la la proportion des électeurs de A dans la population. On choisit de tester l'hypothèse $\mathrm{H}_0 : p = 1/2$ contre $\mathrm{H}_1 : p \neq 0.5$, On a $\overline{x} = 520/1021 = 0.51$. Avec $\alpha = 0.05$, on a $z_{\alpha/2} = 1.96$; la zone de risque est

$$\{\overline{\mathrm{X}} \geq 0.5 + 1.96\frac{1}{2\sqrt{1021}}\} \bigcup \{\overline{\mathrm{X}} \leq 0.5 - 1.96\frac{1}{2\sqrt{1021}}\} = \{\overline{\mathrm{X}} \geq 0.53\} \bigcup \{\overline{\mathrm{X}} \leq 0.47\}.$$

Comme $\overline{x} = 0.51$ est à l'extérieur de cette zone, l'hypothèse H_0 n'est pas rejetée.

6.3.3 Tests d'homogénéité

On étudie un même caractère dans deux populations P_1 et et P_2 et on cherche à savoir si elles sont homogènes par rapport à ce caractère.

Test de comparaison de moyennes

Soient X_1 et X_2 des variables aléatoires représentant le caractère dans chaque population, de moyennes respectives μ_1 et μ_2. De P_1 et P_2, on extrait des échantillons de tailles n_1 et n_2, de moyennes $\overline{x_1}$ et $\overline{x_2}$ et d'écart-type s_1 et s_2. On veut tester l'hypothèse $\mathrm{H}_0 : \mu_1 = \mu_2$ contre $\mathrm{H}_1 : \mu_1 \neq \mu_2$. Alors, pour n_1 et n_2 suffisamment grands ($n_1 \geq 30, n_2 \geq 30$), sous l'hypothèse H_0,

$$\frac{\overline{\mathrm{X}_1} - \overline{\mathrm{X}_2}}{\sqrt{\frac{\mathrm{S}_1^2}{n_1} + \frac{\mathrm{S}_2^2}{n_2}}} \sim \mathcal{N}(0,1).$$

Pour un seuil de risque α, la zone de risque (rejet de H_0) est

$$\boxed{\left\{\frac{|\overline{X_1}-\overline{X_2}|}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \geq z_{\alpha/2}\right\}}.$$

Exemple 6.3.4 Un nouveau carburant moins polluant est testé sur des automobiles. On désire savoir s'il ne réduit pas leur puissance. Un échantillon de $n_1 = 1000$ voitures circulant avec ce nouveau carburant a donné une puissance moyenne $\overline{x_1} = 78$ CV avec un écart-type $s_1 = 20$ CV. Un deuxième échantillon avec $n_2 = 700$ voitures circulant avec l'ancien carburant a donné une puissance moyenne $\overline{x_2} = 79$ CV avec un écart-type $s_2 = 21$ CV. On désire tester l'hypothèse que le nouveau carburant n'affecte pas la puissance des voitures, c–à-d $\mu_1 = \mu_2$, pour les moyennes théoriques des puissances avec et sans le nouveau carburant. On a

$$\frac{\overline{x_1}-\overline{x_2}}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} = \frac{1}{\sqrt{\frac{400}{1000}+\frac{441}{700}}} = 0.985$$

Prenons $\alpha = 0.05$ et donc $z_{\alpha/2} = 1.96$; comme $0.985 < 1.96$, on ne rejette pas l'hypothèse H_0 (le nouveau carburant n'affecte pas la puissance).

Test de comparaison de fréquences

On étudie les fréquence p_1 et p_2 d'un même caractère dans deux populations P_1 et et P_2. De P_1 et P_2, on extrait deux échantillons de tailles n_1 et n_2, avec les fréquences f_1 et f_2 pour le caractère étudié. On veut tester l'hypothèse

$$H_0 : p_1 = p_2 \text{ contre } H_1 : p_1 \neq p_2.$$

Alors, pour n_1 et n_2 suffisamment grands ($n_1 \geq 30, n_2 \geq 30$), sous l'hypothèse H_0,

$$\frac{\overline{X_1}-\overline{X_2}}{\sqrt{\frac{f_1(1-f_1))}{n_1}+\frac{f_2(1-f_2)}{n_2}}} \sim \mathcal{N}(0,1).$$

Pour un seuil de risque α, la zone de risque (rejet de H_0) est

$$\boxed{\frac{|\overline{f_1}-\overline{f_2}|}{\sqrt{\frac{f_1(1-f_1))}{n_1}+\frac{f_2(1-f_2)}{n_2}}} \geq z_{\alpha/2}.}$$

Exemple 6.3.5 Lors du naufrage du Titanic, on a constaté les nombres suivants de survivants par classe de voyageurs :

Classe	1ère	2ème	3ème
Nombre	322	280	711
Survivants	193	119	138

Les proportions de survivants sont donc pour les 3 classes :

$$f_1 = 0.6, \qquad f_2 = 0.425, \qquad f_3 = 0.2.$$

On a

$$F_{12} := \frac{\overline{f_1} - \overline{f_2}}{\sqrt{\frac{f_1(1-f_1))}{n_1} + \frac{f_2(1-f_2)}{n_2}}} = 4.35$$

et de même $F_{13} = 12.8$ et $F_{23} = 6.8$. Ces valeurs sont largement supérieures à $z_{\alpha/2} = 2.58$ pour $\alpha = 0.01$. L'hypothèse que l'appartenance à la classe n'a pas eu d'influence sur les chances de survie est donc à rejeter.

6.4 Séries statistiques à deux variables

On veut étudier deux caractères numériques X et Y d'une population (comme par exemple, le poids et la taille). On observe un échantillon $x = (x_1, \dots, x_n)$ de X et un échantillon $y = (y_1, \dots, y_n)$ de Y.

Définition 6.4.1 (i) Le *nuage de points* correspondant à $\hat{x} = (x_1, \dots, x_n)$ et $\hat{y} = (y_1, \dots, y_n)$ est l'ensemble des points $M_i = (x_i, y_i)$ pour $i \in \{1, \dots, n\}$.
(ii) Le *point moyen* du nuage de points correspondant à $\hat{x} = (x_1, \dots, x_n)$ et $\hat{y} = (y_1, \dots, y_n)$ est le point $G = (\overline{x}, \overline{y})$, où

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \text{et} \qquad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

sont les moyennes des échantillons $\hat{x}$ et $\hat{y}$.
(ii) La *covariance* de $(\hat{x}, \hat{y})$ est le nombre

$$\mathrm{Cov}(\hat{x}, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}).$$

(iii) Le *coefficient de corrélation* de $(\hat{x}, \hat{y})$ est le nombre

$$\rho(\hat{x}, \hat{y}) = \frac{\mathrm{Cov}(\hat{x}, \hat{y})}{\sigma_{\hat{x}}\sigma_{\hat{y}}},$$

où $\sigma_{\hat{x}}$ et $\sigma_{\hat{y}}$ sont les écarts-type des des échantillons $\hat{x}$ et $\hat{y}$.

Remarque 6.4.2 Par l'inégalité de Cauchy-Schwarz (voir Proposition 3.3.14), on a $\mathrm{Cov}(\hat{x},\hat{y}) \leq \sigma_{\hat{x}}\sigma_{\hat{y}}$; il s'ensuit qu'on a toujours $|\rho(\hat{x},\hat{y})| \leq 1$, c-à-d $\rho(\hat{x},\hat{y}) \in [-1,1]$. De plus, on a $\rho(\hat{x},\hat{y}) = \pm 1$ si $\hat{y} = a\hat{x}+b$ pour un $(a,b) \in \mathbf{R}^2$, c-à-d si et seulement si les points $\mathrm{M}_1 = (x_1, y_1), \dots, \mathrm{M}_n = (x_n, y_n)$ sont situés sur une droite.

On cherche la droite qui, en un certain sens, approche le mieux le nuage de points de $(\hat{x},\hat{y})$.

Proposition 6.4.3 ***(Méthode des moindres carrés)*** *La droite d'équation*

$$\boxed{y - \overline{y} = \frac{\mathrm{Cov}(\hat{x},\hat{y})}{\sigma^2_{\hat{x}}}(x - \overline{x})}$$

passe par le point moyen $\mathrm{G} = (\overline{x}, \overline{y})$ *et est, parmi les droites d'équation* $y = ax+b$ *pour* $(a,b) \in \mathbf{R}^2$*, celle qui minimise la somme*

$$\sum_{i=1}^{n} (y_i - ax_i - b)^2.$$

Démonstration Le fait que la droite passe par G est clair. Montrons que cette droite minimise la somme

$$\mathrm{S}(a,b) := \sum_{i=1}^{n} (y_i - ax_i - b)^2$$

pour $(a,b) \in \mathbf{R}^2$.

Soit $(a,b) \in \mathbf{R}^2$. Posons $z = \hat{y} - a\hat{x} - b$. Alors $\mathrm{S}(a,b) = \sum_{i=1} z_i^2$. On a, par le formule de Koenig,

$$\mathrm{Var}(z) = \left(\frac{1}{n}\sum_{i=1} z_i^2\right) - \overline{z}^2,$$

avec

$$\overline{z} = \frac{1}{n}\sum_{i=1} z_i = \overline{y} - a\overline{x}.$$

Donc

$$\mathrm{S}(a,b) = n(\mathrm{Var}(z) + \overline{z}^2). \qquad (*).$$

D'autre part, on a

$$\mathrm{Var}(z) = \mathrm{Var}(\hat{y} - a\hat{x} - b) = \mathrm{Var}(\hat{y} - a\hat{x}) = \mathrm{Var}(\hat{y}) + a^2\,\mathrm{Var}(\hat{x}) - 2a\,\mathrm{Cov}(\hat{x},\hat{y}),$$

qui est donc un trinôme de degré 2 en a. On l'écrit sous forme canonique, en posant $\alpha = \sigma_x, \beta = \mathrm{Cov}(\hat{x}, \hat{y})$ et $\gamma = \sigma_y$:

$$\begin{aligned}\mathrm{Var}(z) &= \left(\alpha a - \frac{\beta}{\alpha}\right)^2 + \gamma^2 - \frac{\beta^2}{\alpha^2}^2 \\ &= \left(\alpha a - \frac{\beta}{\alpha}\right)^2 + \frac{\alpha^2\gamma^2 - \beta^2}{\alpha^2}.\end{aligned}$$

Comme

$$\alpha^2\gamma^2 = \mathrm{Var}(\hat{x})\,\mathrm{Var}(\hat{y}) \geq \mathrm{Cov}(\hat{x}, \hat{y})^2 = \beta^2,$$

les deux termes précédents dans $\mathrm{Var}(z)$ sont positifs ; il s'ensuit que $\mathrm{Var}(z)$ est minimal si $\left(\alpha a - \frac{\beta}{\alpha}\right)^2 = 0$, c-à-d si

$$a = \frac{\beta}{\alpha^2} = \frac{\mathrm{Cov}(\hat{x}, \hat{y})}{\mathrm{Var}(\hat{x})}.$$

Comme $\overline{z} = \overline{y} - a\overline{x} - b$, $\overline{z}^2$ est minimal (et égal à 0) pour

$$b = \overline{y} - a\overline{x}.$$

Sur l'expression $(*)$ de $S(a, b)$, on voit donc que $S(a, b)$ est minimal pour

$$a = \frac{\mathrm{Cov}(\hat{x}, \hat{y})}{\mathrm{Var}(\hat{x})} \qquad \text{et} \qquad b = \overline{y} - \frac{\mathrm{Cov}(\hat{x}, \hat{y})}{\mathrm{Var}(\hat{x})}\overline{x}.\blacksquare$$

Définition 6.4.4 Droite de régression La droite d'équation

$$y - \overline{y} = \frac{\mathrm{Cov}(\hat{x}, \hat{y})}{\sigma_{\hat{x}}^2}(x - \overline{x})$$

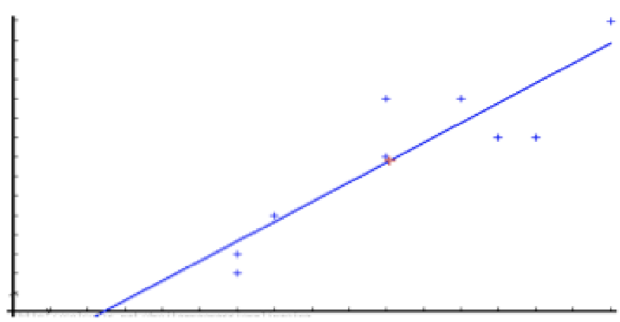
s'appelle *droite de régression* (ou droite d'ajustement) de $\hat{y}$ en $\hat{x}$.

Exemple 6.4.5 10 étudiants ont obtenus les notes suivantes au contrôle continu CC à l'examen final F :

CC : x_i	10	6	12	6	7	13	10	7	14	16
F : y_i	8	2	11	3	5	9	11	5	9	15

On calcule que
$$\overline{x} = 10.1 \qquad \text{et} \qquad \overline{y} = 7.8,$$
de sorte que le point moyen est $G = (10.1, 7.8)$. Les variances sont
$$\sigma^2_{\hat{x}} = 11.49 \qquad \text{et} \qquad \sigma^2_{\hat{y}} = 14.76$$
Le coefficient de corrélation est $\rho(\hat{x}, \hat{y}) = 0.8999$. C'est une valeur très proche de 1, qui montre une forte corrélation entre les notes des deux examens.

La droite de régression linéaire admet pour équation
$$y = bx + a, \qquad \text{avec} \qquad b = 1.0200, a = -2.5021.$$



6.5 Exercices

Exercice 6.5.1 Dans une population de $N = 30.000.000$ individus la proportion d'individus présentant le caractère c est $p = 0,4$. On interroge un échantillon de $n = 1.600$ personnes. Pour chaque échantillon possible ω, on note $S_n(\omega)$ le nombre de personnes présentant le caractère c.

(i) Décrire S_n comme une variable aléatoire.

(ii) Minorer les probabilités des événements suivants :
$$\mathbf{P}(0,3n \leq S_n \leq 0,5n), \quad \mathbf{P}(0,35n \leq S_n \leq 0,45n), \quad \mathbf{P}(0,38n \leq S_n \leq 0,42n)$$
(iii) Donner un intervalle de longueur minimale I tel que $\mathbf{P}(\frac{S_n}{n} \in I) \approx 0,95$.

Exercice 6.5.2 On a relevé le contenu d'essence d'un échantillon de 20 réservoirs d'essence. Les valeurs (en litres) sont les suivantes :

60	59	57	61	57
56	58	59	60	56
56	59	60	56	56
59	56	58	53	62

1. Déterminer les effectifs pour chaque valeur de la série.
2. Calculer la moyenne et l'écart-type de la série.
3. Donner la médiane de cette série.
4. Calculer les effectifs cumulés de la série.
5. Calculer le 1e et le 3e quartile de la série.
6. Tracer le diagramme en bâtons.

Exercice 6.5.3 Sur 429440 naissances, on a dénombré 221023 garçons. La proportion de garçons est-elle compatible avec l'hypothèse d'équiprobabilité des naissances de filles et de garçons au risque de 5%.

Exercice 6.5.4 20% des ampoules issues d'une certaine usine fonctionnent plus de 200 heures. Après un changement dans le processus de fabrication, on constate sur un échantillon de 100 ampoules que 30 fonctionnent plus de 200 heures. L'amélioration apparente est-elle significative au seuil de 5% ?

Exercice 6.5.5 Une machine produit des rondelles dont l'épaisseur est une v.a.r. X d'écart-type 0.3mm. La machine a été réglée pour obtenir des pièces d'épaisseur 5mm. Un contrôle portant sur un échantillon de 100 rondelles a donné 5.07 mm comme moyenne des épaisseurs de cet échantillon. Peut on affirmer que la machine est bien réglée au seuil de 5% ?

Exercice 6.5.6 On sait que la grippe touche 30% d'une population lors d'une épidémie. Pour tester l'efficacité d'un vaccin antigrippal, on vaccine préalablement 300 personnes. A la fin de la saison grippale, on dénombre 50 personnes vaccinées qui ont été atteintes par la grippe. Que peut-on dire de l'efficacité du vaccin, au seuil de 5% ?

Exercice 6.5.7 On prélève 226 pommes de la réserve d'un cidrier. Le poids moyen des pommes de l'échantillon est de 115g et l'écart-type de 15g. Trouver a tel que le poids moyen des pommes de la réserve soit compris entre $115 - a$ et $115 + a$ avec une probabilité de 50%.

Exercice 6.5.8 On veut tester l'autonomie de deux modèles de vélos électriques. On dispose de 101 exemplaires du premier et 121 exemplaires du deuxième modèle. Leurs autonomies moyennes sont respectivement de 85km avec un écart type de 5km et de 98km avec un écart type de 10km. Peut on considérer, avec un risque de 5%, que les deux modèles peuvent être distingués par leur autonomie ?

Exercice 6.5.9 On se propose d'étudier l'influence de la température T sur la durée d'incubation des œufs de grenouilles. On choisit six échantillons de deux cents œufs chacun. Le nombre X d'éclosions au vingt-deuxième jour est le suivant :

température t_i d'incubation	6	6,4	6,8	7,2	7,6	8
nombre d'éclosions x_i	131	144	157	170	190	189

(i) Dessinez le nuage de points des données. Tracer une droite semblant bien approcher ce nuage.
(ii) Calculer le coefficient de corrélation de l'observation et donner l'équation de la droite de régression. Comparer avec la droite précédente.
(iii) Donnez le nombre d'éclosions prédit par la droite de régression sur un échantillon de 200 œufs à 7,5 degrés Celsius.

Exercice 6.5.10 On a mesuré certaines variables x et y sur 10 individus et obtenu le tableau suivant :

individu i	1	2	3	4	5	6	7	8	9	10
x_i	18	20	19	16	19	16	19	21	15	17
y_i	43	110	70	17	91	29	80	134	15	34

On cherche une relation de la forme $y = a \ln x + b$. Pour cela on pose $z = \ln x$.
(i) Dessiner le nuage de points $(z_i, y_i)_{1\le i\le 10}$ et calculer la droite de régression.
(ii) Quelle valeur de y peut-on prédire pour un individu présentant pour x la valeur 22 ?

Chapitre 7

Annexe : Un tableau de lois classiques

Lois discrètes

Nom	Symbole	$X(\Omega)$	Paramètres	Loi	Espérance	Variance
Bernoulli	$\mathscr{B}(p)$	$\{0,1\}$	$p \in]0,1[$	$\mathbf{P}(X=1)=p$	p	$p(1-p)$
binomiale	$\mathscr{B}(n,p)$	$[\![0,n]\!]$	$n \in \mathbf{N}*, p \in]0,1[$	$\mathbf{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$
hypergéométrique	$\mathscr{H}(N,n,p)$	$[[0 \vee n-N(1-p), n \wedge Np]]$	$n \leq N \in \mathbf{N}*, p \in]0,1[$	$\mathbf{P}(X=k)=\frac{\binom{Np}{k}\binom{N(1-p)}{n-k}}{\binom{N}{n}}$	np	$np(1-p)\,\frac{N-n}{N-1}$
géométrique	$\mathscr{G}(p)$	$\mathbf{N}^*$	$p \in]0,1[$	$\mathbf{P}(X=k)=(1-p)^{k-1}p$	$1/p$	$(1-p)/p^2$
Poisson	$\mathscr{P}(\lambda)$	$\mathbf{N}$	$\lambda > 0$	$\mathbf{P}(X=k)=e^{-\lambda}\lambda^k/k!$	λ	λ

Lois continues

Nom	Symbole	Support	Paramètres	Densité	Espérance	Variance
uniforme	$\mathscr{U}([a,b])$	$[a,b]$	a<b	$\frac{1}{b-a}\mathbf{1}_{[a,b]}(x)$	$(a+b)/2$	$(b-a)^2/12$
exponentielle	$\mathscr{E}(\lambda)$	$[0,+\infty[$	$\lambda > 0$	$\lambda e^{-\lambda x}\mathbf{1}_{[0,+\infty[}(x)$	$1/\lambda$	$1/\lambda^2$
normale/Gauss	$\mathscr{N}(m,\sigma)$	$\mathbf{R}$	$m \in \mathbf{R}, \sigma > 0$	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}$	m	σ^2

Chapitre 8

Annexe : Tables statistiques

Tables statistiques

1 Fonction de répartition de la loi normale centrée réduite

Si U suit la loi normale centrée réduite, pour $x \geq 0$, la table donne la valeur $\phi(x) = P(U \leq x)$ avec $x = x_1 + x_2$ où x_1 et x_2 sont indiqués en marge. Pour $x < 0$, on utilise $\phi(x) = 1 - \phi(-x)$.

					x_1					
x_2	**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
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

2 Fractiles de la loi normale centrée réduite

u_p est le fractile d'ordre p de la loi normale centrée réduite. Donc $\phi(u_p) = p$.
La table donne la valeur u_p pour $p = p_1 + p_2$ avec p_1 et p_2 indiqués en marge.
Pour les valeurs $p < 0,5$, on utilise la relation $u_p = -u_{1-p}$.

p_1

p_2	**0.000**	**0.001**	**0.002**	**0.003**	**0.004**	**0.005**	**0.006**	**0.007**	**0.008**	**0.009**
0.50	0.0000	0.0025	0.0050	0.0075	0.0100	0.0125	0.0150	0.0175	0.0201	0.0226
0.51	0.0251	0.0276	0.0301	0.0326	0.0351	0.0376	0.0401	0.0426	0.0451	0.0476
0.52	0.0502	0.0527	0.0552	0.0577	0.0602	0.0627	0.0652	0.0677	0.0702	0.0728
0.53	0.0753	0.0778	0.0803	0.0828	0.0853	0.0878	0.0904	0.0929	0.0954	0.0979
0.54	0.1004	0.1030	0.1055	0.1080	0.1105	0.1130	0.1156	0.1181	0.1206	0.1231
0.55	0.1257	0.1282	0.1307	0.1332	0.1358	0.1383	0.1408	0.1434	0.1459	0.1484
0.56	0.1510	0.1535	0.1560	0.1586	0.1611	0.1637	0.1662	0.1687	0.1713	0.1738
0.57	0.1764	0.1789	0.1815	0.1840	0.1866	0.1891	0.1917	0.1942	0.1968	0.1993
0.58	0.2019	0.2045	0.2070	0.2096	0.2121	0.2147	0.2173	0.2198	0.2224	0.2250
0.59	0.2275	0.2301	0.2327	0.2353	0.2378	0.2404	0.2430	0.2456	0.2482	0.2508
0.60	0.2533	0.2559	0.2585	0.2611	0.2637	0.2663	0.2689	0.2715	0.2741	0.2767
0.61	0.2793	0.2819	0.2845	0.2871	0.2898	0.2924	0.2950	0.2976	0.3002	0.3029
0.62	0.3055	0.3081	0.3107	0.3134	0.3160	0.3186	0.3213	0.3239	0.3266	0.3292
0.63	0.3319	0.3345	0.3372	0.3398	0.3425	0.3451	0.3478	0.3505	0.3531	0.3558
0.64	0.3585	0.3611	0.3638	0.3665	0.3692	0.3719	0.3745	0.3772	0.3799	0.3826
0.65	0.3853	0.3880	0.3907	0.3934	0.3961	0.3989	0.4016	0.4043	0.4070	0.4097
0.66	0.4125	0.4152	0.4179	0.4207	0.4234	0.4261	0.4289	0.4316	0.4344	0.4372
0.67	0.4399	0.4427	0.4454	0.4482	0.4510	0.4538	0.4565	0.4593	0.4621	0.4649
0.68	0.4677	0.4705	0.4733	0.4761	0.4789	0.4817	0.4845	0.4874	0.4902	0.4930
0.69	0.4959	0.4987	0.5015	0.5044	0.5072	0.5101	0.5129	0.5158	0.5187	0.5215
0.70	0.5244	0.5273	0.5302	0.5330	0.5359	0.5388	0.5417	0.5446	0.5476	0.5505
0.71	0.5534	0.5563	0.5592	0.5622	0.5651	0.5681	0.5710	0.5740	0.5769	0.5799
0.72	0.5828	0.5858	0.5888	0.5918	0.5948	0.5978	0.6008	0.6038	0.6068	0.6098
0.73	0.6128	0.6158	0.6189	0.6219	0.6250	0.6280	0.6311	0.6341	0.6372	0.6403
0.74	0.6433	0.6464	0.6495	0.6526	0.6557	0.6588	0.6620	0.6651	0.6682	0.6713
0.75	0.6745	0.6776	0.6808	0.6840	0.6871	0.6903	0.6935	0.6967	0.6999	0.7031
0.76	0.7063	0.7095	0.7128	0.7160	0.7192	0.7225	0.7257	0.7290	0.7323	0.7356
0.77	0.7388	0.7421	0.7454	0.7488	0.7521	0.7554	0.7588	0.7621	0.7655	0.7688
0.78	0.7722	0.7756	0.7790	0.7824	0.7858	0.7892	0.7926	0.7961	0.7995	0.8030
0.79	0.8064	0.8099	0.8134	0.8169	0.8204	0.8239	0.8274	0.8310	0.8345	0.8381
0.80	0.8416	0.8452	0.8488	0.8524	0.8560	0.8596	0.8633	0.8669	0.8705	0.8742
0.81	0.8779	0.8816	0.8853	0.8890	0.8927	0.8965	0.9002	0.9040	0.9078	0.9116
0.82	0.9154	0.9192	0.9230	0.9269	0.9307	0.9346	0.9385	0.9424	0.9463	0.9502
0.83	0.9542	0.9581	0.9621	0.9661	0.9701	0.9741	0.9782	0.9822	0.9863	0.9904
0.84	0.9945	0.9986	1.0027	1.0069	1.0110	1.0152	1.0194	1.0237	1.0279	1.0322
0.85	1.0364	1.0407	1.0450	1.0494	1.0537	1.0581	1.0625	1.0669	1.0714	1.0758
0.86	1.0803	1.0848	1.0893	1.0939	1.0985	1.1031	1.1077	1.1123	1.1170	1.1217
0.87	1.1264	1.1311	1.1359	1.1407	1.1455	1.1503	1.1552	1.1601	1.1650	1.1700
0.88	1.1750	1.1800	1.1850	1.1901	1.1952	1.2004	1.2055	1.2107	1.2160	1.2212
0.89	1.2265	1.2319	1.2372	1.2426	1.2481	1.2536	1.2591	1.2646	1.2702	1.2759
0.90	1.2816	1.2873	1.2930	1.2988	1.3047	1.3106	1.3165	1.3225	1.3285	1.3346
0.91	1.3408	1.3469	1.3532	1.3595	1.3658	1.3722	1.3787	1.3852	1.3917	1.3984
0.92	1.4051	1.4118	1.4187	1.4255	1.4325	1.4395	1.4466	1.4538	1.4611	1.4684
0.93	1.4758	1.4833	1.4909	1.4985	1.5063	1.5141	1.5220	1.5301	1.5382	1.5464
0.94	1.5548	1.5632	1.5718	1.5805	1.5893	1.5982	1.6072	1.6164	1.6258	1.6352
0.95	1.6449	1.6546	1.6646	1.6747	1.6849	1.6954	1.7060	1.7169	1.7279	1.7392
0.96	1.7507	1.7624	1.7744	1.7866	1.7991	1.8119	1.8250	1.8384	1.8522	1.8663
0.97	1.8808	1.8957	1.9110	1.9268	1.9431	1.9600	1.9774	1.9954	2.0141	2.0335
0.98	2.0537	2.0749	2.0969	2.1201	2.1444	2.1701	2.1973	2.2262	2.2571	2.2904
0.99	2.3263	2.3656	2.4089	2.4573	2.5121	2.5758	2.6521	2.7478	2.8782	3.0902

3 Fractiles de la loi de Student

$t_{\nu,p}$ est le fractile d'ordre p de la loi de Student à ν degrés de liberté.
Pour les valeurs de $p \leq 0,5$, on utilise la relation $t_{\nu,p} = -t_{\nu,1-p}$.
Lorsque $\nu > 50$, on utilise l'approximation de la loi de Student par la loi normale $\mathcal{N}(0,1)$, ce qui revient à : $t_{\nu,p} \approx u_p$.

	p									
ν	**0.60**	**0.70**	**0.80**	**0.90**	**0.95**	**0.9750**	**0.9900**	**0.9950**	**0.9990**	**0.9995**
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657	318.309	636.619
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750	3.385	3.646
31	0.256	0.530	0.853	1.309	1.696	2.040	2.453	2.744	3.375	3.633
32	0.255	0.530	0.853	1.309	1.694	2.037	2.449	2.738	3.365	3.622
33	0.255	0.530	0.853	1.308	1.692	2.035	2.445	2.733	3.356	3.611
34	0.255	0.529	0.852	1.307	1.691	2.032	2.441	2.728	3.348	3.601
35	0.255	0.529	0.852	1.306	1.690	2.030	2.438	2.724	3.340	3.591
36	0.255	0.529	0.852	1.306	1.688	2.028	2.434	2.719	3.333	3.582
37	0.255	0.529	0.851	1.305	1.687	2.026	2.431	2.715	3.326	3.574
38	0.255	0.529	0.851	1.304	1.686	2.024	2.429	2.712	3.319	3.566
39	0.255	0.529	0.851	1.304	1.685	2.023	2.426	2.708	3.313	3.558
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704	3.307	3.551
41	0.255	0.529	0.850	1.303	1.683	2.020	2.421	2.701	3.301	3.544
42	0.255	0.528	0.850	1.302	1.682	2.018	2.418	2.698	3.296	3.538
43	0.255	0.528	0.850	1.302	1.681	2.017	2.416	2.695	3.291	3.532
44	0.255	0.528	0.850	1.301	1.680	2.015	2.414	2.692	3.286	3.526
45	0.255	0.528	0.850	1.301	1.679	2.014	2.412	2.690	3.281	3.520
46	0.255	0.528	0.850	1.300	1.679	2.013	2.410	2.687	3.277	3.515
47	0.255	0.528	0.849	1.300	1.678	2.012	2.408	2.685	3.273	3.510
48	0.255	0.528	0.849	1.299	1.677	2.011	2.407	2.682	3.269	3.505
49	0.255	0.528	0.849	1.299	1.677	2.010	2.405	2.680	3.265	3.500
50	0.255	0.528	0.849	1.299	1.676	2.009	2.403	2.678	3.261	3.496

4 Fractiles de la loi du χ^2

$\chi^2_{\nu,p}$ est le fractile d'ordre p de la loi du χ^2.

Pour les valeurs de $\nu > 50$, on utilise l'approximation $\chi^2_{\nu,p} \approx \dfrac{(u_p + \sqrt{2\nu - 1})^2}{2}$

p

ν	**0.001**	**0.005**	**0.010**	**0.025**	**0.05**	**0.1000**	**0.5000**	**0.9000**	**0.9500**	**0.9750**	**0.9900**	**0.9950**	**0.9990**
1	0.000	0.000	0.000	0.001	0.004	0.016	0.455	2.706	3.841	5.024	6.635	7.879	10.828
2	0.002	0.010	0.020	0.051	0.103	0.211	1.386	4.605	5.991	7.378	9.210	10.597	13.816
3	0.024	0.072	0.115	0.216	0.352	0.584	2.366	6.251	7.815	9.348	11.345	12.838	16.266
4	0.091	0.207	0.297	0.484	0.711	1.064	3.357	7.779	9.488	11.143	13.277	14.860	18.467
5	0.210	0.412	0.554	0.831	1.145	1.610	4.351	9.236	11.070	12.833	15.086	16.750	20.515
6	0.381	0.676	0.872	1.237	1.635	2.204	5.348	10.645	12.592	14.449	16.812	18.548	22.458
7	0.598	0.989	1.239	1.690	2.167	2.833	6.346	12.017	14.067	16.013	18.475	20.278	24.322
8	0.857	1.344	1.646	2.180	2.733	3.490	7.344	13.362	15.507	17.535	20.090	21.955	26.124
9	1.152	1.735	2.088	2.700	3.325	4.168	8.343	14.684	16.919	19.023	21.666	23.589	27.877
10	1.479	2.156	2.558	3.247	3.940	4.865	9.342	15.987	18.307	20.483	23.209	25.188	29.588
11	1.834	2.603	3.053	3.816	4.575	5.578	10.341	17.275	19.675	21.920	24.725	26.757	31.264
12	2.214	3.074	3.571	4.404	5.226	6.304	11.340	18.549	21.026	23.337	26.217	28.300	32.909
13	2.617	3.565	4.107	5.009	5.892	7.042	12.340	19.812	22.362	24.736	27.688	29.819	34.528
14	3.041	4.075	4.660	5.629	6.571	7.790	13.339	21.064	23.685	26.119	29.141	31.319	36.123
15	3.483	4.601	5.229	6.262	7.261	8.547	14.339	22.307	24.996	27.488	30.578	32.801	37.697
16	3.942	5.142	5.812	6.908	7.962	9.312	15.338	23.542	26.296	28.845	32.000	34.267	39.252
17	4.416	5.697	6.408	7.564	8.672	10.085	16.338	24.769	27.587	30.191	33.409	35.718	40.790
18	4.905	6.265	7.015	8.231	9.390	10.865	17.338	25.989	28.869	31.526	34.805	37.156	42.312
19	5.407	6.844	7.633	8.907	10.117	11.651	18.338	27.204	30.144	32.852	36.191	38.582	43.820
20	5.921	7.434	8.260	9.591	10.851	12.443	19.337	28.412	31.410	34.170	37.566	39.997	45.315
21	6.447	8.034	8.897	10.283	11.591	13.240	20.337	29.615	32.671	35.479	38.932	41.401	46.797
22	6.983	8.643	9.542	10.982	12.338	14.041	21.337	30.813	33.924	36.781	40.289	42.796	48.268
23	7.529	9.260	10.196	11.689	13.091	14.848	22.337	32.007	35.172	38.076	41.638	44.181	49.728
24	8.085	9.886	10.856	12.401	13.848	15.659	23.337	33.196	36.415	39.364	42.980	45.559	51.179
25	8.649	10.520	11.524	13.120	14.611	16.473	24.337	34.382	37.652	40.646	44.314	46.928	52.620
26	9.222	11.160	12.198	13.844	15.379	17.292	25.336	35.563	38.885	41.923	45.642	48.290	54.052
27	9.803	11.808	12.879	14.573	16.151	18.114	26.336	36.741	40.113	43.195	46.963	49.645	55.476
28	10.391	12.461	13.565	15.308	16.928	18.939	27.336	37.916	41.337	44.461	48.278	50.993	56.892
29	10.986	13.121	14.256	16.047	17.708	19.768	28.336	39.087	42.557	45.722	49.588	52.336	58.301
30	11.588	13.787	14.953	16.791	18.493	20.599	29.336	40.256	43.773	46.979	50.892	53.672	59.703
31	12.196	14.458	15.655	17.539	19.281	21.434	30.336	41.422	44.985	48.232	52.191	55.003	61.098
32	12.811	15.134	16.362	18.291	20.072	22.271	31.336	42.585	46.194	49.480	53.486	56.328	62.487
33	13.431	15.815	17.074	19.047	20.867	23.110	32.336	43.745	47.400	50.725	54.776	57.648	63.870
34	14.057	16.501	17.789	19.806	21.664	23.952	33.336	44.903	48.602	51.966	56.061	58.964	65.247
35	14.688	17.192	18.509	20.569	22.465	24.797	34.336	46.059	49.802	53.203	57.342	60.275	66.619
36	15.324	17.887	19.233	21.336	23.269	25.643	35.336	47.212	50.998	54.437	58.619	61.581	67.985
37	15.965	18.586	19.960	22.106	24.075	26.492	36.336	48.363	52.192	55.668	59.893	62.883	69.346
38	16.611	19.289	20.691	22.878	24.884	27.343	37.335	49.513	53.384	56.896	61.162	64.181	70.703
39	17.262	19.996	21.426	23.654	25.695	28.196	38.335	50.660	54.572	58.120	62.428	65.476	72.055
40	17.916	20.707	22.164	24.433	26.509	29.051	39.335	51.805	55.758	59.342	63.691	66.766	73.402
41	18.575	21.421	22.906	25.215	27.326	29.907	40.335	52.949	56.942	60.561	64.950	68.053	74.745
42	19.239	22.138	23.650	25.999	28.144	30.765	41.335	54.090	58.124	61.777	66.206	69.336	76.084
43	19.906	22.859	24.398	26.785	28.965	31.625	42.335	55.230	59.304	62.990	67.459	70.616	77.419
44	20.576	23.584	25.148	27.575	29.787	32.487	43.335	56.369	60.481	64.201	68.710	71.893	78.750
45	21.251	24.311	25.901	28.366	30.612	33.350	44.335	57.505	61.656	65.410	69.957	73.166	80.077
46	21.929	25.041	26.657	29.160	31.439	34.215	45.335	58.641	62.830	66.617	71.201	74.437	81.400
47	22.610	25.775	27.416	29.956	32.268	35.081	46.335	59.774	64.001	67.821	72.443	75.704	82.720
48	23.295	26.511	28.177	30.755	33.098	35.949	47.335	60.907	65.171	69.023	73.683	76.969	84.037
49	23.983	27.249	28.941	31.555	33.930	36.818	48.335	62.038	66.339	70.222	74.919	78.231	85.351
50	24.674	27.991	29.707	32.357	34.764	37.689	49.335	63.167	67.505	71.420	76.154	79.490	86.661