Table des matières

Chapitre 1

CONCEPTS DE BASE

On fait appel aux probabilités pour décrire une expérience dont le résultat est impossible à prévoir avec certitude, mais dont on connait quand-même l'ensemble des résultats possibles (ce qui n'est pas le cas, par exemple, pour deviner le rêve d'un inconnu).

Expérience 1 : On jette un dé et on lit le numéro apparu sur la face supérieure.
Expérience 2 : On jette deux fois un dé et on note les numéros obtenus.

La notion de résultat d'une expérience n'est pas claire : c'est l'expérimentateur qui décide de ce qui mérite le nom de résultat en fonction de ses propres motivations.

Exemple 1 : Lors de l'expérience 1, on peut s'intéresser au numéro de la face supérieure –il y a alors 6 résultats possibles–, ou seulement à la parité de cette face –il y a alors 2 résultats possibles– (si, par exemple, pour débuter un match de foot, l'arbitre n'a pas de pièce mais un dé...)
Exemple 2 : Lors de la naissance d'un bébé, on peut s'intéresser au sexe de l'enfant (garçon ou fille !), ou bien à la couleur de ses yeux, ou encore à son poids, ou à sa taille.

Il est donc très important de définir avec précision les motivations de l'expérience et, par suite, ce que l'on entend par résultat.

Ce chapitre a pour but d'introduire les premières notions de probabilités et de se familiariser avec les outils qui seront utilisés tout au long du cours.

1.1 Introduction de la notion d'événement

Si, lorsqu'on répète l'expérience dans des conditions identiques, le résultat observé est susceptible de changer, l'expérience est dite *aléatoire*.

L'ensemble de tous les résultats possibles ou *états* est appelé *univers* de l'expérience : on le notera Ω.

Expérience 1 : $\Omega = \{1, 2, 3, 4, 5, 6\}$; $\mathrm{card}\Omega = 6$ (où card désigne le nombre d'éléments).
Expérience 2 : $\Omega = \{(1,1), \cdots, (1,6), (2,1), \cdots, (2,6), \cdots (6,1), \cdots, (6,6)\}$; $\mathrm{card}\Omega = 36$.

L'ensemble Ω peut être :
• fini (ensemble des 6 faces d'un dé, des 32 cartes d'un jeu,...) ;
• infini dénombrable (ensemble des entiers naturels, ou d'états que l'on peut numéroter) ;
• infini non dénombrable (position d'une particule dans un liquide, poids, taille,...).

Exemple : On lance une fléchette en direction d'une cible. On peut convenir d'appeler résultat :
→ le gain correspondant à la zone du point d'impact : $\Omega = \{0, 100, 200, 500, 1000\}$;
→ la distance du point d'impact au centre de la cible, mesurée à 1cm près par défaut : $\Omega = \mathbb{N}$;
→ le point d'impact : Ω partie de $\mathbb{R}^2$ correspondant au mur.

Lorsqu'on effectue une expérience aléatoire, certains faits liés à cette expérience peuvent se produire ou non : on les appelle *événements*.

Exemples d'événements :
Pour l'expérience 1, A_1 "le numéro obtenu est pair" ; A_2 "le numéro obtenu est $\geqslant 4$"
Pour l'expérience 2, B "la somme des numéros obtenus est 6".

Chaque résultat possible est appelé *événement simple*.
Les événements susceptibles d'intéresser ne sont pas seulement les événements simples.
Exemple : L'événement "le poids du bébé est compris entre 3 kg et 3,2 kg" est plus intéressant que l'événement simple "le poids du bébé est 3,124 kg". De même, pour les boxeurs, la catégorie est plus importante que le poids précis.

Un événement est lié à une expérience associée à Ω si, pour tout résultat $\omega \in \Omega$, on sait dire si cet événement a lieu ou non. On convient d'identifier un tel événement à l'ensemble des $\omega \in \Omega$ pour lequel il a lieu. Un événement sera donc identifié à une partie de Ω.

Exemple :
Pour l'expérience 1, A_1 est réalisé si et seulement si $\omega \in \{2, 4, 6\}$. On notera $A_1 = \{2, 4, 6\}$.
Pour l'expérience 2, B est réalisé si et seulement si (ω_1, ω_2) vérifie $\omega_1 + \omega_2 = 6$.
On notera $B = \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$.

Plus généralement, à chaque expérience, on peut associer un ensemble Ω tel que chaque événement puisse être représenté par une partie de Ω.

Rappel sur le vocabulaire ensembliste :

1- Soit A une partie de Ω.
On note $\overline{A}$ le *complémentaire* de A : c'est l'ensemble de tous les états qui ne sont pas dans A.

Propriété 1.1 : $\overline{\overline{A}} = A$.

2- Soient A et B deux parties de Ω.
On note $A \cap B$ *l'intersection* de A et de B : c'est l'ensemble des états qui sont à la fois dans A et dans B.
On note $A \cup B$ *la réunion* de A et de B : c'est l'ensemble des états qui sont dans A ou dans B (ils peuvent être dans les deux).
On note $A \subset B$ et on dit que A est *inclus* dans B si tous les états de A sont dans B.

Propriétés 1.2 :
1) $\cap$ et $\cup$ sont commutatives et associatives.
2) Si $A \subset B$, alors $A \cap B = A$, $A \cup B = B$ et $\overline{B} \subset \overline{A}$.

3- Si $A_1, \cdots, A_n, \cdots$ sont une infinité dénombrable de parties de Ω, on note $\bigcup_n A_n$ (ou, de façon plus précise, $\bigcup_{n=1}^{+\infty} A_n$ la *réunion dénombrable* des A_n : c'est l'ensemble des états qui sont au moins dans l'un des A_n et on note $\bigcap_n A_n$ l'*intersection dénombrable* des A_n : c'est l'ensemble des états qui sont dans tous les A_n à la fois.

Propriétés 1.3 :
1) $\overline{\bigcup_n A_n} = \bigcap_n \overline{A_n}$; $\overline{\bigcap_n A_n} = \bigcup_n \overline{A_n}$.
2) $A \cap \left(\bigcup_n A_n \right) = \bigcup_n (A \cap A_n)$: distributivité de l'intersection par rapport à la réunion ;
$A \cup \left(\bigcap_n A_n \right) = \bigcap_n (A \cup A_n)$: distributivité de la réunion par rapport à l'intersection.

On appelle $\emptyset$ l'*événement impossible* car il n'est jamais réalisé et Ω l'*événement certain* car il est toujours réalisé.
Si $A \cap B = \emptyset$, on dit que A et B sont *incompatibles* car ils ne peuvent avoir lieu en même temps.

1.2 Tribus

1.2.1 Définition, premières propriétés

Définition 1.1 : On appelle *tribu* $\mathcal{A}$ sur Ω, tout sous-ensemble de parties de Ω tel que :
i) $\Omega \in \mathcal{A}$;
ii) si $A \in \mathcal{A}$, alors $\overline{A} \in \mathcal{A}$;
iii) si pour tout $n \in \mathbb{N}$, $A_n \in \mathcal{A}$, alors $\bigcup_{n=0}^{+\infty} A_n \in \mathcal{A}$.

Le couple $(\Omega, \mathcal{A})$ est appelé *espace probabilisable*.

Propriétés 1.4 :
1) $\emptyset \in \mathcal{A}$;
2) si, pour tout $n \in \mathbb{N}$, $A_n \in \mathcal{A}$, alors $\bigcap_{n=0}^{+\infty} A_n \in \mathcal{A}$.

Démonstration. 1) Par i) $\Omega \in \mathcal{A}$ et par ii), $\overline{\Omega} \in \mathcal{A}$; or $\overline{\Omega} = \emptyset$.

2) Par ii) $\overline{A_i} = B_i \in \mathcal{A}$ et par iii) $\bigcup_{n=0}^{+\infty} B_n \in \mathcal{A}$. Enfin, de nouveau par ii), $\overline{\bigcup_{n=0}^{+\infty} B_n} \in \mathcal{A}$.

Or $\overline{\bigcup_{n=0}^{+\infty} B_n} = \bigcap_{n=0}^{+\infty} \overline{B_n} = \bigcap_{n=0}^{+\infty} A_n$. $\square$

Propriété 1.5 : L'intersection d'une famille quelconque (dénombrable ou non) de tribus de Ω est une tribu de Ω.

Exemples de tribus :
- $\{\emptyset, \Omega\}$: la plus petite ;
- $\mathcal{P}(\Omega)$: la plus grande ;
- $\{\emptyset, \Omega, A, \overline{A}\}$.

Cas particulier très important : lorsque Ω est fini ou infini dénombrable, on prend toujours $\mathcal{A} = \mathcal{P}(\Omega)$. (Ce ne sera pas le cas lorsque $\Omega = \mathbb{R}$: la tribu considérée dans ce cas sera en général la tribu des boréliens $\mathcal{B}_{\mathbb{R}}$ qui est la plus petite tribu contenant les intervalles réels).

Système complet d'événements :

On appelle *système complet d'événements* de Ω, toute famille finie ou dénombrable $(A_i)_{i \in I}$ telle que :
i) $A_i \in \mathcal{A}$ pour tout $i \in I$;
ii) $\bigcup_{i \in I} A_i = \Omega$;
iii) si $i \neq j$, alors $A_i \cap A_j = \emptyset$.
(On dit aussi que les A_i forment une partition dénombrable de Ω).

1.2.2 Tribu engendrée par une famille de parties

Soit $\mathcal{C}$ une famille de parties de Ω. On souhaite définir $\sigma(\mathcal{C})$ la “plus petite” tribu de parties de Ω contenant $\mathcal{C}$, c'est-à-dire que $\sigma(\mathcal{C})$ doit être une tribu et que, si $\mathcal{D}$ est une tribu contenant $\mathcal{C}$, alors $\sigma(\mathcal{C}) \subset \mathcal{D}$.

On considère l'ensemble des tribus contenant $\mathcal{C}$. Cet ensemble n'est pas vide, puisqu'il contient $\mathcal{P}(\Omega)$, ensemble des parties de Ω. D'après la propriété 1.5, l'intersection de toutes les tribus contenant $\mathcal{C}$ est une tribu. Cette intersection contient $\mathcal{C}$ et, si $\mathcal{D}$ est une tribu contenant $\mathcal{C}$, alors $\mathcal{D}$ contient l'intersection des tribus contenant $\mathcal{C}$. Ceci prouve que $\sigma(\mathcal{C})$ existe, et est unique : c'est l'intersection de toutes les tribus contenant $\mathcal{C}$.

Définition 1.2 : On appelle *tribu engendrée* par une famille non vide $\mathcal{C}$ de parties de Ω, l'intersection de toutes les tribus contenant $\mathcal{C}$: on la note $\sigma(\mathcal{C})$.

Exemples :

1. Si $\mathcal{C} = \{A\}$ où $A \subset \Omega$, $A \neq \emptyset$ et $A \neq \Omega$, $\sigma(\mathcal{C}) = \{\emptyset, A, \overline{A}, \Omega\}$.

2. Si $\mathcal{C} = \{A, B, C\}$ est une partition de Ω, $\sigma(\mathcal{C}) = \{\emptyset, A, B, C, A \cup B, A \cup C, B \cup C, \Omega\}$ et $\text{card}\sigma(\mathcal{C}) = 2^3 = 8$. (Plus généralement, si $\mathcal{C}$ est une partition telle que $\text{card}\mathcal{C} = n$, alors $\text{card}\sigma(\mathcal{C}) = 2^n$).

3. Plus généralement, si $\mathcal{C} = \{A_i \; ; \; i \in I\}$ est une partition finie ou dénombrable de Ω ($I \subset \mathbb{N}$), alors $\sigma(\mathcal{C}) = \left\{ \bigcup_{i \in J} A_i \; ; \; J \subset I \right\}$. En effet, par définition même d'une tribu d'une part, et de $\sigma(\mathcal{C})$

d'autre part, on a $\mathcal{C} \subset \left\{ \bigcup_{i \in J} A_i \ ; \ J \subset I \right\} \subset \sigma(\mathcal{C})$. On montre ensuite que $\left\{ \bigcup_{i \in J} A_i \ ; \ J \subset I \right\}$ est une tribu en vérifiant les 3 axiomes de la définition.

4. Si $\Omega = \mathbb{R}^d$, on utilise le plus souvent la *tribu des boréliens* de $\mathbb{R}^d$. Par définition, il s'agit de la tribu engendrée par les pavés de $\mathbb{R}^d$, c'est-à-dire :

$$\mathcal{B}_{\mathbb{R}^d} = \sigma \left(\left\{ \prod_{i=1}^{d}]a_i, b_i[\ ; \ a_i, b_i \in \mathbb{R} \ ; \ 1 \leqslant i \leqslant d \right\} \right).$$

Dans le cas très important où $d = 1$, $\mathcal{B}_{\mathbb{R}}$ contient en particulier $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, tous les intervalles de $\mathbb{R}$ (ouverts, fermés, semi-ouvers, bornés ou non),... Cependant $\mathcal{B}_{\mathbb{R}} \neq \mathcal{P}(\mathbb{R})$ (la démonstration de ce dernier point sort du du cadre de ce cours).

1.3 Introduction de la notion de probabilité

Une probabilité P est une mesure qui permet d'évaluer les chances de réalisation des événements. Si $\mathcal{A}$ est une tribu sur Ω, à chaque événement A de $\mathcal{A}$, on associe un réel $P(A)$ compris entre 0 et 1, appelé probabilité de l'événement A.

Approche fréquentielle de la notion de probabilité :

Tous les événements liés à une même expérience n'ont pas la même "chance" d'être réalisés. Soit $(\Omega, \mathcal{A})$ l'espace probabilisable associé à l'expérience et soit $A \in \mathcal{A}$. Si on répète N fois l'expérience dans des conditions identiques et si A est réalisé N_A fois, le nombre $\dfrac{N_A}{N}$ est appelé fréquence de réalisation de A sur ces N coups.

En général, la fréquence de réalisation de A tend à se stabiliser lorsque $N \to +\infty$ et lorsque N est grand, $\dfrac{N_A}{N}$ est la valeur approchée de la mesure d'une grandeur associée à A : sa probabilité.

Remarques :
- $\dfrac{N_A}{N} \in [0, 1]$;
- Si A et B sont incompatibles, $\dfrac{N_{A \cup B}}{N} = \dfrac{N_A}{N} + \dfrac{N_B}{N}$.

Définition 1.3 : Soit $(\Omega, \mathcal{A})$ un espace probabilisable. On appelle **probabilité** sur $(\Omega, \mathcal{A})$ toute application P de $\mathcal{A}$ vers $[0, 1]$ telle que :
i) $P(\Omega) = 1$;
ii) pour toute suite d'événements $A_n \in \mathcal{A}$, incompatibles deux à deux, on a :

$$P\left(\bigcup_{n=0}^{+\infty} A_n \right) = \sum_{n=0}^{+\infty} P(A_n) \left(= \lim_{N \to +\infty} \sum_{n=0}^{N} P(A_n) \right).$$

Le triplet $(\Omega, \mathcal{A}, P)$ est appelé *espace probabilisé*.

Propriétés 1.6 :
1) Si $A \subset B$, alors $P(A) \leqslant P(B)$;
2) $P(\overline{A}) = 1 - P(A)$; $P(\emptyset) = 0$;
3) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;
4) $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \left(\sum_{1 \leqslant i_1 \cdots < i_k \leqslant n} P(A_{i_1} \cap \cdots \cap A_{i_k}) \right)$ (identité de Poincaré) ;
5) si $(A_n)_n$ est une suite croissante de $\mathcal{A}$ ($A_n \subset A_{n+1}$), alors $P\left(\bigcup_n A_n\right) = \lim_{n \to +\infty} P(A_n)$;
6) si $(B_n)_n$ est une suite décroissante de $\mathcal{A}$ ($B_{n+1} \subset B_n$), alors $P\left(\bigcap_n B_n\right) = \lim_{n \to +\infty} P(B_n)$.

Démonstration. **1)** $B = (B \cap A) \cup (B \cap \overline{A}) = A \cup (B \cap \overline{A})$ et $A \cap (B \cap \overline{A}) = \emptyset$ donc

$$P(B) = P(A) + P(B \cap \overline{A}) \geqslant P(A).$$

2) $P(\Omega) = P(A) + P(\overline{A})$ car $A \cup \overline{A} = \Omega$ et $A \cap \overline{A} = \emptyset$. Or $P(\Omega) = 1$ donc $P(\overline{A}) = 1 - P(A)$ et, en particulier pour $A = \Omega$, $P(\emptyset) = 0$.

3) On décompose $A \cup B$ en 3 événements incompatibles 2 à 2 :

$$A \cup B = (A \cap \overline{B}) \cup (A \cap B) \cup (\overline{A} \cap B).$$

De même, $A = (A \cap B) \cup (A \cap \overline{B})$ et $B = (A \cap B) \cup (\overline{A} \cap B)$.

En passant aux probabilités, on obtient :

$$P(A \cup B) = P(A \cap \overline{B}) + P(A \cap B) + P(\overline{A} \cap B),$$

avec $P(A) = P(A \cap B) + P(A \cap \overline{B})$ et $P(B) = P(A \cap B) + P(\overline{A} \cap B)$. On a donc bien :

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

4) Pour $n = 3$, on a $P(A_1 \cup A_2 \cup A_3) = P((A_1 \cup A_2) \cup A_3) = P(A_1 \cup A_2) + P(A_3) - P((A_1 \cup A_2) \cap A_3)$ d'après 3). Mais, toujours d'après 3),

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

et $(A_1 \cup A_2) \cap A_3 = (A_1 \cap A_3) \cup (A_2 \cap A_3)$ union à laquelle on applique encore 3) :

$$P((A_1 \cup A_2) \cap A_3) = P(A_1 \cap A_3) + P(A_2 \cap A_3) - P(A_1 \cap A_2 \cap A_3)$$

et en regroupant tout, on a bien le résultat.

On procède ensuite par récurrence sur n (on suppose la propriété vraie au rang n et on applique 3) à $A = \bigcup_{i=1}^{n} A_i$ et à $B = A_{n+1}$, puis l'hypothèse de récurrence pour A, la distributivité, et de nouveau l'hypothèse de récurrence aux $A_i \cap A_{n+1}$...)

5) Posons $C_0 = A_0$ et pour $n \geqslant 1$, $C_n = A_n \cap \overline{A}_{n-1}$ (on a alors $A_n = A_{n-1} \cup C_n$). On décompose la preuve en plusieurs étapes :
→ <u>Les C_i sont 2 à 2 disjoints</u>

En effet, si, par exemple, $i > j$, supposons qu'il existe $\omega \in C_i \cap C_j$. On a alors $\omega \in C_i = A_i \cap \overline{A}_{i-1} \subset \overline{A}_{i-1}$ et $\omega \in C_j \subset A_j \subset A_{i-1}$ car $j \leqslant i - 1$, ce qui est contradictoire.

→ <u>$A_N = \bigcup_{n=0}^{N} C_n$</u> : démonstration par récurrence sur N ;

C'est vrai pour $N=0$ et si $A_N = \bigcup_{n=0}^{N} C_n$, alors

$$\bigcup_{n=0}^{N+1} C_n = A_N \cup C_{N+1} = A_N \cup (A_{N+1} \cap \overline{A}_N) = A_{N+1}.$$

$\rightarrow$ $\underline{\bigcup_n A_n = \bigcup_n C_n}$: en effet, d'abord $C_n \subset A_n$ donc $\bigcup_n C_n \subset \bigcup_n A_n$; d'autre part $A_N = \bigcup_{n=0}^{N} C_n \subset \bigcup_{n\geqslant 0} C_n$ pour tout $N \geqslant 0$ donc $\bigcup_{N\geqslant 0} A_N \subset \bigcup_{n\geqslant 0} C_n$.

$\rightarrow$ <u>On en déduit</u> $P\left(\bigcup_n A_n\right) = P\left(\bigcup_n C_n\right) = \sum_{n=0}^{+\infty} P(C_n)$ car les C_n sont 2 à 2 disjoints. Or, par définition d'une série, $\sum_{n=0}^{+\infty} P(C_n) = \lim_{N\to+\infty} \sum_{n=0}^{N} P(C_n)$ et comme

$$P(A_N) = P\left(\bigcup_{n=0}^{N} C_n\right) = \sum_{n=0}^{N} P(C_n),$$

on a bien $P\left(\bigcup_{n\geqslant 0} A_n\right) = \lim_{n\to+\infty} P(A_n)$.

6) Si (B_n) est décroissante $(B_{n+1} \subset B_n)$ et si $A_n = \overline{B}_n$, alors (A_n) est croissante et, d'après 5), on a $P\left(\bigcup_{n\geqslant 0} \overline{B}_n\right) = \lim_{n\to+\infty} P(\overline{B}_n) = 1 - \lim_{n\to+\infty} P(B_n)$. Or

$$P\left(\bigcup_{n\geqslant 0} \overline{B}_n\right) = 1 - P\left(\overline{\bigcup_{n\geqslant 0} \overline{B}_n}\right) = 1 - P\left(\bigcap_{n\geqslant 0} B_n\right).$$

Donc $P\left(\bigcap_{n\geqslant 0} B_n\right) = \lim_{n\to+\infty} P(B_n)$. □

<u>**Cas particuliers très importants :**</u>

1- Ω fini : $\Omega = \{\omega_1, \cdots, \omega_n\}$ et $\mathcal{A} = \mathcal{P}(\Omega)$.
Soient $p_1, \cdots, p_n$, n nombres réels. Il existe une probabilité P sur $(\Omega, \mathcal{A})$ telle que, pour tout $i \in \{1, \cdots, n\}$, $P(\{\omega_i\}) = p_i$ si et seulement si, pour tout $i \in \{1, \cdots, n\}$, $p_i \geqslant 0$ et $\sum_{i=1}^{n} p_i = 1$.
La probabilité P est alors unique et, pour tout $A \in \mathcal{A}$, $P(A) = \sum_{i;\omega_i\in A} p_i$.

Définition 1.4 : On dit qu'il y a <u>*équiprobabilité*</u> lorsque les probabilités de tous les événements simples sont égales.

Théorème 1.1 : S'il y a équiprobabilité, alors, pour tout événement A, on a $P(A) = \dfrac{\mathrm{card}A}{\mathrm{card}\Omega}$.

Démonstration. Supposons que $\mathrm{card}\Omega = n$. Il existe λ tel que $p_i = \lambda$ pour tout i. On a alors $\sum_{i=1}^{n} p_i = n\lambda = 1$, d'où $\lambda = \dfrac{1}{n}$ et, pour tout $A \in \mathcal{A}$, $P(A) = \sum_{i;\omega_i\in A} p_i = (\mathrm{card}A) \times \lambda = \dfrac{\mathrm{card}A}{\mathrm{card}\Omega}$. □

Exemples de calculs de probabilités dans le cas où Ω est fini.
Exemple 1 : On lance un dé non truqué et on considère les événements A “le résultat est pair” et B “le résultat est multiple de 3”.

$\Omega = \{1,2,3,4,5,6\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ et, pour $1 \leqslant i \leqslant 6$, $P(\{i\}) = \dfrac{1}{6}$ (le dé étant non truqué, il y a équiprobabilité).

$$A = \{2,4,6\}\ ;\ B = \{3,6\}\ ;\ A \cap B = \{6\}\ ;\ A \cup B = \{2,3,4,6\}.$$

$P(A) = \dfrac{\mathrm{card}A}{\mathrm{card}\Omega} = \dfrac{3}{6} = \dfrac{1}{2}$ et de même $P(B) = \dfrac{2}{6} = \dfrac{1}{3}$; $P(A \cap B) = \dfrac{1}{6}$; $P(A \cup B) = \dfrac{4}{6} = \dfrac{2}{3}$.

On peut vérifier que l'on a bien $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Exemple 2 : Dans une salle qui contient 4 rangs de 10 personnes, on place une personne au hasard. Quelle chance a-t-elle d'être au premier rang ? (1/4). Au premier rang, à la première place ? (1/40).

Exemple 3 : Quelle est la probabilité, en tapant successivement 3 lettres de l'alphabet au hasard, d'écrire le mot “TFC” ? ($1/26^3$).

2- Ω infini dénombrable : $\Omega = \{\omega_i\ ;\ i \in \mathbb{N}\}$ et $\mathcal{A} = \mathcal{P}(\Omega)$.

Il existe une probabilité P sur $(\Omega, \mathcal{A})$ telle que, pour tout $i \in \mathbb{N}$, $P(\{\omega_i\}) = p_i$ si et seulement si, pour tout $i \in \mathbb{N}$, $p_i \geqslant 0$ et $\displaystyle\sum_{i=0}^{+\infty} p_i = 1$.
La probabilité P est alors unique et, pour tout $A \in \mathcal{P}$, $P(A) = \displaystyle\sum_{i;\omega_i \in A} p_i$.

Remarque : On ne peut pas avoir équiprobabilité dans le cas infini dénombrable car, pour qu'une série converge, il est nécessaire que son terme général tende vers 0 et si $p_i = \lambda$ pour tout i, on doit avoir $\lambda = 0$; mais alors $\displaystyle\sum_i p_i = 0 \neq 1$.

1.4 Événements indépendants

Approche intuitive : 2 événements sont indépendants si la réalisation de l'un est sans effet sur la réalisation de l'autre. En termes de probabilités, on a la définition suivante :

Définition 1.5 : Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé.
1) Deux événements A et B de $\mathcal{A}$ sont *indépendants* si $P(A \cap B) = P(A) \times P(B)$.
2) Une famille d'événements $(A_n)_n$ est dite *famille d'événements indépendants* si, pour tout $p \in \mathbb{N}^*$ et pour tout $\{i_1, \cdots, i_p\} \subset \mathbb{N}$,

$$P(A_{i_1} \cap \cdots \cap A_{i_p}) = P(A_{i_1}) \times \cdots \times P(A_{i_p}). \qquad (*)$$

Mises en garde :
1- Ne pas confondre événements incompatibles ($A \cap B = \emptyset$) et événements indépendants ($P(A \cap B) = P(A)P(B)$).

2- L'indépendance dépend de la probabilité considérée.
Exemple : Soit P_1 la probabilité définie sur $\Omega = \{1, 2, 3, 4, 5, 6\}$ par :

$$P_1(\{1\}) = P_1(\{2\}) = \frac{1}{6} \ ; \ P_1(\{3\}) = \frac{1}{3} \ ; \ P_1(\{4\}) = P_1(\{5\}) = P_1(\{6\}) = \frac{1}{9}$$

et soit P_2 l'équiprobabilité sur Ω. Soient $A = \{1, 2\}$ et $B = \{2, 3\}$.

On a $P_1(A) = \frac{2}{6}$; $P_1(B) = \frac{1}{2}$; $P_1(A \cap B) = \frac{1}{6}$ donc $P_1(A)P_1(B) = P_1(A \cap B)$.

D'autre part, $P_2(A) = P_2(B) = \frac{1}{3}$; $P_2(A \cap B) = \frac{1}{6}$; $P_2(A \cap B) = \frac{1}{6}$ donc $P_2(A)P_2(B) = \frac{1}{9} \neq P_2(A \cap B)$.

Les événements A et B sont indépendants pour P_1 mais pas pour P_2.

3- Il faut vérifier $(*)$ pour toutes les sous-familles : en particulier, l'indépendance 2 à 2 d'événements n'implique pas leur indépendance.

Exemple 1 : Soit $\Omega = \{1, 2, 3, 4, 5, 6\}$ et P l'équiprobabilité sur Ω.

Soient $A = \{1, 2, 3\}$; $B = \{2, 4, 6\}$ et $C = \{1, 2, 4, 5\}$.

On a $P(A) = P(B) = \frac{1}{2}$; $P(C) = \frac{2}{3}$; $P(A \cap B) = P(A \cap B \cap C) = P(\{2\}) = \frac{1}{6}$.

On a bien $P(A)P(B)P(C) = \frac{1}{6} = P(A \cap B \cap C)$ mais $P(A)P(B) = \frac{1}{4} \neq P(A \cap B)$ et les événements A, B et C ne sont donc pas indépendants.

Exemple 2 : On jette deux fois un dé et on considère les événements A"le premier lancer est pair", B"le deuxième lancer est pair" et C"la somme des 2 lancers est paire".

On a $P(A) = P(B) = \frac{3 \times 6}{36} = \frac{1}{2}$ et comme $A \cap B = A \cap C = B \cap C = A \cap B \cap C$, on a de plus $P(A \cap B) = P(A \cap C) = P(B \cap C) = P(A \cap B \cap C) = \frac{3 \times 3}{36} = \frac{1}{4}$. De plus, $P(C) = \frac{9 + 9}{36} = \frac{1}{2}$.

On a $P(A \cap B) = P(A)P(B)$; $P(A \cap C) = P(A)P(C)$; $P(B \cap C) = P(B)P(C)$ mais A, B et C ne sont pas indépendants car $P(A \cap B \cap C) = \frac{1}{4} \neq P(A)P(B)P(C)$.

Le concept d'indépendance est donc très délicat : ainsi, il sera parfois difficile de deviner ou de pressentir l'indépendance, qui devra donc être vérifiée par le calcul.

Théorème 1.2 : Si (A, B) est un couple d'événements indépendants, il en est de même des couples $(\overline{A}, B)$, $(A, \overline{B})$ et $(\overline{A}, \overline{B})$.

Démonstration.

$$P(\overline{A})P(B) - P(\overline{A} \cap B) = (1 - P(A))P(B) - (P(B) - P(A \cap B)) = P(A \cap B) - P(A)P(B)$$

donc, si $P(A \cap B) = P(A)P(B)$, alors $P(\overline{A} \cap B) = P(\overline{A})P(B)$.

Comme A et B jouent des rôles symétriques, on a le même résultat pour $(A, \overline{B})$, puis, en remplaçant B par $\overline{B}$, pour $(\overline{A}, \overline{B})$. □

On peut généraliser ce résultat :

Théorème 1.3 : Si $(A_n)_n$ est une famille d'événements indépendants, il en est de même de $(A'_n)_n$, avec $A'_n = A_n$ ou bien $A'_n = \overline{A_n}$.

Démonstration. Elle se fait par récurrence sur le nombre de complémentaires. □

La notion d'indépendance sera reprise au Chapitre 4, à propos des probabilités conditionnelles.

Chapitre 2

RAPPELS D'ANALYSE COMBINATOIRE

Lorsque l'univers Ω d'une expérience est fini, on utilise l'équiprobabilité sur $(\Omega, \mathcal{P}(\Omega))$ chaque fois qu'aucun événement simple n'a de privilège sur les autres.

Dans ce cas, le calcul des probabilités se ramène donc au calcul du nombre d'éléments de Ω et de ses sous-ensembles.

L'analyse combinatoire est précisément l'ensemble des méthodes permettant de compter les éléments d'un ensemble.

Définition 2.1 : L'ensemble des p-uplets $(y_1, \cdots, y_p)$ où $y_i \in N_i$ pour tout $i \in \{1, \cdots, p\}$ est appelé *produit cartésien* des N_i. Il est noté $N_1 \times \cdots \times N_p$, ou bien $\prod_{i=1}^{p} N_i$.

Si $N_1 = \cdots = N_p = N$, alors $\prod_{i=1}^{p} N_i$ est noté N^p.

Propriété 2.1 : $\text{card}(N_1 \times \cdots \times N_p) = \text{card}N_1 \times \cdots \times \text{card}N_p$ et $\text{card}(N^p) = (\text{card}N)^p$.

On notera $A_p, B_p, \cdots$ des ensembles de cardinal p.

2.1 Tirages ordonnés avec remise (ou applications)

On note $\mathcal{F}(E_p, F_n)$ l'ensemble des applications de E_p vers F_n.

Théorème 2.1 : $\text{card}\mathcal{F}(E_p, F_n) = n^p$.

Démonstration. Soit $E_p = \{e_1, \cdots, e_p\}$. À tout e_i de E_p, on associe, de façon unique $f(e_i)$ dans F_n. La donnée de f équivaut à celle du p-uplet $(f(e_1), \cdots, f(e_p))$ de $(F_n)^p$ et, d'après la propriété 2.1, $\text{card}\mathcal{F}(E_p, F_n) = (\text{card}F_n)^p = n^p$. $\square$

Exemple 2.1 : Combien de "mots" (ayant un sens ou pas) de 5 lettres peut-on former ?

On prend 26 boules gravées de A à Z que l'on place dans une urne. On fait 5 tirages **successifs** (on veut un mot donc l'ordre des lettres a de l'importance), **avec remise** de la boule après chaque tirage (une lettre peut figurer plusieurs fois dans un mot).

L'ensemble des résultats possibles est l'ensemble des 5-uplets $(u_1, \cdots, u_5)$ avec $u_i \in \{A, \cdots, Z\}$. On a alors $\text{card}\Omega = 26^5$.

2.2 Tirages ordonnés sans remise (ou injections)

On note $\mathcal{I}(E_p, F_n)$ l'ensemble des injections de E_p vers F_n, lorsque $n \geqslant p$.

Théorème 2.2 : $\text{card}\mathcal{I}(E_p, F_n) = n(n-1)\cdots(n-(p-1))$.

Démonstration. Il est nécessaire que $n \geqslant p$ car chaque élément de E_p doit avoir une image distincte.

La donnée de j dans $\mathcal{I}(E_p, F_n)$ équivaut à celle du p-uplet $(j(e_1), \cdots, j(e_p))$ formé de p éléments **distincts** de F_n : il y a n choix possibles pour $j(e_1)$; le choix de $j(e_1)$ étant fait, il ne reste plus que $n-1$ choix possibles pour $j(e_2), \cdots$, et, pour $j(e_p)$, il ne reste plus que $n-(p-1)$ choix ($p-1$ éléments de F_n ayant déjà été utilisés).

Ainsi, on a bien $\text{card}\mathcal{I}(E_p, F_n) = n \times (n-1) \times \cdots \times (n-(p-1))$. □

Exemple 2.2 : Combien de mots de 5 lettres peut-on former sans utiliser 2 fois la même lettre ?

On reprend les boules de l'exemple 2.1 : on fait 5 tirages **successifs** (il s'agit toujours d'un mot, donc l'ordre a toujours de l'importance) mais **on ne remet pas** les boules déjà tirées dans l'urne (afin de ne pas retomber sur la même lettre).

On a alors $\text{card}\Omega = 26 \times 25 \times 24 \times 23 \times 22$.

Remarque : Si $n = p$, les injections sont en fait des bijections et on a alors :

$$\text{card}\mathcal{I}(E_n, F_n) = n(n-1) \times \cdots \times 2 \times 1 = n!$$

Dans le cas général, $(n \geqslant p)$, $\text{card}\mathcal{I}(E_p, F_n) = \dfrac{n!}{(n-p)!}$ que l'on note A_n^p.

2.3 Tirages non ordonnés sans remise (ou combinaisons)

Définition 2.2 : Si $n \geqslant p$, on appelle *coefficient binomial* C_n^p (ou parfois $\binom{n}{p}$), le nombre

$$\frac{n!}{p!(n-p)!}.$$

Propriétés 2.2 :
1) $C_n^p = C_n^{n-p}$;
2) $C_n^p = C_{n-1}^p + C_{n-1}^{p-1}$ (formule de Pascal)

Démonstration. **1)** $C_n^{n-p} = \dfrac{n!}{(n-p)!(n-(n-p))!} = \dfrac{n!}{(n-p)!p!} = C_n^p$.

2) $C_{n-1}^p + C_{n-1}^{p-1} = \dfrac{(n-1)!}{p!(n-1-p)!} + \dfrac{(n-1)!}{(p-1)!(n-p)!} = \dfrac{(n-1)!}{p!(n-p)!}(n-p+p) = \dfrac{n!}{p!(n-p)!} = C_n^p$.

On peut aussi faire une justification directe des propriétés 2.2 :

1) Évident : il revient au même de choisir p éléments parmi n que d'en éliminer $n-p$.

2) Parmi les n éléments, on en considère un en particulier, que l'on note a. Si Ω désigne l'ensemble des choix de p objets, $\Omega = \Omega_1 \cup \Omega_2$, où Ω_1 désigne l'ensemble des choix de p objets dont a fait partie, et Ω_2 désigne l'ensemble des choix de p objets sans a.

$\mathrm{card}\Omega_1 = C_{n-1}^{p-1}$ (il ne faut choisir que $p-1$ éléments parmi $n-1$, puisqu'on a déjà a).

$\mathrm{card}\Omega_2 = C_{n-1}^{p}$ (il faut choisir p objets parmi les $n-1$ autres que a).

Comme $\mathrm{card}\Omega = C_n^p$ et comme $\mathrm{card}\Omega = \mathrm{card}\Omega_1 + \mathrm{card}\Omega_2$ (car Ω_1 et Ω_2 sont disjoints), on a bien le résultat. □

Définition 2.3 : Une <u>*p-combinaison*</u> de F_n est une partie de F_n à p éléments.

Théorème 2.3 : Le nombre de p-combinaisons de F_n est C_n^p.

Démonstration. À toute injection j de $\mathcal{I}(E_p, F_n)$, on associe la partie $\{j(e_1), \cdots, j(e_p)\}$.

Il y a $p!$ injections qui donnent la même partie $\{j(e_1), \cdots, j(e_p)\}$ (autant de façons que de classer les $j(e_i)$ pour $1 \leqslant i \leqslant p$).

On utilise alors le "**principe des bergers**" : pour compter les moutons de son troupeau, le berger compte les pattes, et, chaque mouton ayant 4 pattes, il divise par 4. Ici, pour compter les parties, on compte les injections (rôle des pattes) et, chaque partie (rôle du mouton) donnant lieu à $p!$ injections, on divise par $p!$.

Pour $n \geqslant p$, le nombre de parties est donc $\dfrac{A_n^p}{p!} = \dfrac{n!}{p!(n-p)!} = C_n^p$. □

<u>*Exemple 2.3*</u> : Combien de "mains" de 5 cartes différentes existe-t-il dans un jeu de 32 ?

Il y en a C_{32}^5, qui correspond au nombre de façons de choisir 5 cartes parmi 32 (l'ordre des cartes dans la main n'a pas d'importance !)

2.4 Tirages non ordonnés avec remise (ou combinaisons avec répétitions)

Définition 2.4 : Une <u>*p-combinaison avec répétition*</u> de F_n est une liste de p éléments de F_n, les répétitions étant autorisées et l'ordre dans la liste n'intervenant pas.

Par exemple $[a, a, b, c, c, f]$ est une 6-c.a.r.

Théorème 2.4 : Le nombre de p-combinaisons avec répétitions de F_n est C_{p+n-1}^{n-1}.

Démonstration. Il revient au même de considérer les applications f de $F_n = \{x_1, \cdots, x_n\}$ dans $\mathbb{N}$ vérifiant $f(x_1) + \cdots + f(x_n) = p$. En effet, à chaque élément de F_n, on associe son nombre (éventuellement nul) d'apparitions dans la p-c.a.r..

Il revient aussi au même de considérer toutes les répartitions possibles de p boules identiques dans n tiroirs $x_1, \cdots, x_n$. En effet, à chaque tiroir, on associe le nombre de boules qu'il contient. On va utiliser cette dernière modélisation pour démontrer le théorème, mais on va procéder à l'envers, c'est-à-dire aligner d'abord les p boules et placer ensuite les $n-1$ cloisons des tiroirs.

Pour placer la première cloison, on a $p+1$ choix possibles car les p boules délimitent $p+1$ espaces libres. Une fois placée la première cloison, on a $p+2$ choix pour la deuxième car la première cloison a partagé un espace libre en $2, \cdots$ Ainsi, pour placer la $(n-1)$-ième et dernière cloison, on a $p+n-1$ choix.

On a donc $(p+1) \times \cdots \times (p+n-1)$ façons de placer les cloisons. Mais les $n-1$ cloisons étant identiques, l'ordre de placement n'intervient pas et une même configuration peut être obtenue de $(n-1)!$ façons différentes.

D'après le principe des bergers, le nombre de configurations est donc :

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

Une autre façon plus rapide de procéder est de considérer que l'on a au total $n-1+p$ objets à placer dont $n-1$ cloisons identiques et p boules identiques. Les places occupées par les $n-1$ cloisons dans un tel $n-1+p$-uplet d'objets déterminent une unique configuration et il y en a en tout $C_{n-1+p}^{n-1} = C_{n-1+p}^{p}$. □

Exemple 2.4 : Au jeu des "chiffres et des lettres", combien y-a-t-il de tirages possibles pour les 9 lettres ?

Il s'agit de 9-c.a.r. de 26 éléments donc ce nombre de choix est $C_{26+9-1}^{9} = C_{34}^{9}$.

2.5 Partages d'ensembles

Problème : Etant donné p entiers positifs ou nuls, $n_1, \cdots, n_p$, vérifiant $n_1 + \cdots + n_p = n$, on cherche le nombre de partages de F_n en p parties $A_1, \cdots, A_p$ telles que $\mathrm{card} A_i = n_i$.

Théorème 2.5 : Le nombre de partages de F_n en p parties $A_1, \cdots, A_p$ telles que $\mathrm{card} A_i = n_i$ est $\dfrac{n!}{n_1! \cdots n_p!}$.

Démonstration. On commence par choisir les éléments de A_1 : il faut en choisir n_1 parmi n, soit $C_n^{n_1}$ choix. Puis, on choisit, parmi les $n - n_1$ qui restent, les n_2 éléments de A_2, soit $C_{n-n_1}^{n_2}$ choix...

Une fois choisis les éléments de $A_1, \cdots, A_{p-1}$, il ne reste plus que $C_{n-n_1-\cdots-n_{p-1}}^{n_p} = C_{n_p}^{n_p} = 1$ choix pour ceux de A_p (tous ceux qui restent).

Le nombre de partages est donc :

$$\begin{aligned} C_n^{n_1} C_{n-n_1}^{n_2} \cdots C_{n-n_1-\cdots-n_{p-1}}^{n_p} &= \frac{n!}{n_1!(n-n_1)!} \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \cdots \frac{(n-n_1-\cdots-n_{p-1})!}{n_p!0!} \\ &= \frac{n!}{n_1! \cdots n_p!} \end{aligned}$$

□

Exemple 2.5 : Au jeu "des chiffres et des lettres", combien y-a-t-il de tirages de 9 lettres possibles contenant n_A fois la lettre $A, \cdots,$ n_Z fois la lettre Z, avec $n_A + \cdots + n_Z = 9$?

Le nombre de choix est donc $\dfrac{9!}{n_A! \cdots n_Z!}$.

Remarques :

1) Un certain nombre de "n_i" sont nuls (ici au moins 17) et on a alors $n_i! = 1$.

2) Il s'agit d'un tirage **avec remise** (une lettre peut apparaître plusieurs fois) et **ordonné** (on fait des mots donc l'ordre des lettres est imposé). La différence avec le 2.1 est qu'ici, la fréquence d'apparition des lettres est imposée dès le départ, alors que dans 2.1., tout est permis.

Chapitre 3

VARIABLES ALÉATOIRES, LOIS CLASSIQUES SUR $\mathbb{R}$

Soit $\mathcal{E}$ une expérience aléatoire et $(\Omega, \mathcal{A}, P)$ l'espace probabilisé qui en rend compte.
Il arrive très souvent qu'à chaque résultat de $\mathcal{E}$, on associe une valeur numérique (par exemple, le poids d'un bébé lors d'une naissance ou bien la somme des chiffres obtenus après le lancer de 2 dés) ; on définit ainsi une application X de Ω vers $\mathbb{R}$.
On sera amené à considérer l'ensemble des résultats $\omega \in \Omega$ tel que $X(\omega) = a$, ou encore $X(\omega) < a$, ou encore $X(\omega) \in [a, b[$... On voudrait parler de la probabilité de telle ou telle situation. Or, dans un espace probabilisé, seuls les événements ont une probabilité (et la tribu n'est pas toujours formée de toutes les parties de Ω). Donc, pour pouvoir considérer la probabilité de ces ensembles, il faut que ceux-ci soient dans la tribu $\mathcal{A}$. Dans ce chapitre, après avoir présenté quelques généralités sur les variables aléatoires et leurs lois (partie un peu théorique, que l'on peut sauter en première lecture), on s'intéressera plus particulièrement aux variables aléatoires réelles.

3.1 Généralités

3.1.1 Tribu engendrée par une application

Rappel :
Si X est une application d'un espace Ω dans un espace Ω' et si $A' \subset \Omega'$, l'ensemble $X^{-1}(A')$ est le sous-ensemble de Ω, appelé image réciproque de A' par X, défini par :

$$X^{-1}(A') = \{\omega \in \Omega \ ; \ X(\omega) \in A'\} \text{ noté aussi } [X \in A']$$

(en d'autres termes, il s'agit de l'ensemble des éléments de Ω qui ont leur image par X dans A').

Propriété 3.1 : Si X est une application de Ω vers Ω' et si $\mathcal{A}'$ est une tribu de parties de Ω', alors

$$\mathcal{B}_X = X^{-1}(\mathcal{A}') = \{X^{-1}(A') \ ; \ A' \in \mathcal{A}'\}$$

est une tribu de parties de Ω.

Démonstration. • $X^{-1}(\Omega') = \{\omega \in \Omega \ ; \ X(\omega) \in \Omega'\} = \Omega \in \mathcal{B}_X$.
• Si $A' \in \mathcal{A}'$, alors

$$\overline{X^{-1}(A')} = \{\omega \in \Omega \ ; \ X(\omega) \notin A'\} = \{\omega \in \Omega \ ; \ X(\omega) \in \overline{A'}\} = X^{-1}(\overline{A'}) \in \mathcal{B}_X$$

car $\overline{A'} \in \mathcal{A}'$.

• Si $(A'_i)_{i\in I}$ est une famille dénombrable d'événements de $\mathcal{A}'$, alors

$$\bigcup_{i\in I} X^{-1}(A'_i) = X^{-1}\left(\bigcup_{i\in I} A'_i\right) \in \mathcal{B}_X$$

car $\bigcup_{i\in I} A'_i \in \mathcal{A}'$. □

Définition 3.1 : La tribu $\mathcal{B}_X$ est appelée *tribu engendrée par X*.

Propriété 3.2 : Si X est une application de Ω vers Ω', si $\mathcal{C}'$ est une famille de parties de Ω' et si $\mathcal{A}' = \sigma(\mathcal{C}')$, alors

$$\mathcal{B}_X = \sigma\left(X^{-1}(\mathcal{C}')\right)$$

c'est-à-dire $X^{-1}\left(\sigma(\mathcal{C}')\right) = \sigma\left(X^{-1}(\mathcal{C}')\right)$.

Démonstration. Admise (l'égalité s'obtient en fait par double inclusion). □

3.1.2 Variables aléatoires

Définition 3.2 : Soit $X : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$. On dit que X est une *variable aléatoire* (en abrégé v.a.), ou encore *application mesurable* sur $(\Omega, \mathcal{A})$ et à valeurs dans $(\Omega', \mathcal{A}')$, si

$$\mathcal{B}_X \subset \mathcal{A}$$

c'est-à-dire si, pour tout $A' \in \mathcal{A}'$, $X^{-1}(A') \in \mathcal{A}$.

Propriété 3.3 : Si X est une application de Ω vers Ω', et si $\mathcal{A}' = \sigma(\mathcal{C}')$ où $\mathcal{C}'$ est une famille de parties de Ω', alors

$$X \text{ est une variable aléatoire} \iff \text{pour tout } A' \in \mathcal{C}',\ X^{-1}(A') \in \mathcal{A}$$

c'est-à-dire qu'il suffit de considérer les images réciproques des éléments d'une partie engendrant la tribu $\mathcal{A}'$.

Démonstration. $\Rightarrow$ Évident, puisque $\mathcal{C}' \subset \mathcal{A}'$

$\Leftarrow$ D'après la propriété 3.2, on a $\mathcal{B}_X = \sigma\left(X^{-1}(\mathcal{C}')\right.$. Or $\sigma\left(X^{-1}(\mathcal{C}')\right) \subset \mathcal{A}$ puisque la famille de parties $X^{-1}(\mathcal{C}')$ et contenue dans la tribu $\mathcal{A}$ par hypothèse, et que la tribu engendrée par cette famille (plus petite tribu la contenant) est ainsi contenue dans $\mathcal{A}$. On a donc bien $\mathcal{B}_X \subset \mathcal{A}$ donc X est bien une variable aléatoire. □

Propriété 3.4 : Si X est une v.a. de $(\Omega, \mathcal{A})$ dans $(\Omega', \mathcal{A}')$ et si Y est une v.a. de $(\Omega', \mathcal{A}')$ dans $(\Omega'', \mathcal{A}'')$, alors $Y \circ X$ est une v.a. de $(\Omega, \mathcal{A})$ dans $(\Omega'', \mathcal{A}'')$.

Démonstration. Soit $A'' \in \mathcal{A}''$. On a :

$$\begin{aligned}(Y\circ X)^{-1}(A'') &= \{\omega \in \Omega\ ;\ Y\circ X(\omega) = Y(X(\omega)) \in A''\} \\ &= \{\omega \in \Omega\ ;\ X(\omega) \in A' = Y^{-1}(A'')\} = X^{-1}(A') = A \in \mathcal{A}\end{aligned}$$

car $A' \in \mathcal{A}'$ (Y v.a.) et $A = X^{-1}(A') \in \mathcal{A}$ (X v.a.), ce qui prouve bien que $Y \circ X$ est une v.a. □

3.1.3 Loi de probabilité d'une variable aléatoire

Propriété 3.5 : Si X est une v.a. définie sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, à valeurs dans $(\Omega', \mathcal{A}')$, alors l'application $P_X : A' \in \mathcal{A}' \mapsto P_X(A') = P([X \in A'])$ est une probabilité sur $(\Omega', \mathcal{A}')$.

Démonstration. • On a $[X \in \Omega'] = \{\omega \in \Omega \ ; \ X(\omega) \in \Omega'\} = \Omega$ donc

$$P_X(\Omega') = P([X \in \Omega']) = P(\Omega) = 1.$$

• Soit $(A'_i)_{i\in I}$ une famille dénombrable d'événements deux à deux disjoints de $\mathcal{A}'$. Alors $([X \in A'_i])_{i\in I}$ est une famille dénombrable d'événements deux à deux disjoints de $\mathcal{A}$ car $[X \in A'_i] \cap [X \in A'_j] = [X \in A'_i \cap A'_j] = [X \in \emptyset] = \emptyset$. On a donc

$$\begin{aligned} P_X\left(\bigcup_{i\in I} A'_i\right) &= P\left(X \in \bigcup_{i\in I} A'_i\right) = P\left(\bigcup_{i\in I} [X \in A'_i]\right) \\ &= \sum_{i\in I} P([X \in A'_i]) = \sum_{i\in I} P_X(A'_i) \end{aligned}$$

ce qui prouve bien que P_X est une probabilité. □

Définition 3.3 : La probabilité $P_X : A' \in \mathcal{A}' \mapsto P([X \in A'])$ est appelée <u>*loi de probabilité*</u> de X (on dit aussi <u>*image*</u> de P par X).

Remarque : Si $\mathcal{A}' = \sigma(\mathcal{C}')$, il suffit de connaître $P_X(A')$ pour tout $A' \in \mathcal{C}'$.

3.1.4 Variables aléatoires réelles

On rappelle que, si X est une application de Ω vers Ω', on désigne par $X(\Omega)$ l'ensemble $\{X(\omega) \ ; \ \omega \in \Omega\}$; c'est l'ensemble des valeurs susceptibles d'être prises par X.

Une variable aléatoire de $(\Omega, \mathcal{A})$ dans $(\Omega', \mathcal{A}')$ où $(\Omega', \mathcal{A}') = (\mathbb{R}, \mathcal{B}_{\mathbb{R}}$ (avec $\mathcal{B}_{\mathbb{R}}$ tribu des boréliens de $\mathbb{R}$) est dite réelle. On rappelle que $\mathcal{B}_{\mathbb{R}} = \sigma\left(\{[a,b[\ ; \ a, b \in \mathbb{R}\}\right)$.
En fait, on a aussi $\mathcal{B}_{\mathbb{R}} = \sigma\left(\{I \ ; \ I \text{ intervalle de } \mathbb{R}\}\right)$ puisque

$$\{[a,b[\ ; \ a, b \in \mathbb{R}\} \subset \{I \ ; \ I \text{ intervalle de } \mathbb{R}\} \subset \mathcal{B}_{\mathbb{R}}.$$

On peut ainsi écrire de manière équivalente :

Définition 3.4 : On appelle <u>*variable aléatoire réelle*</u> (en abrégé v.a.r.) toute application X de $(\Omega, \mathcal{A})$ dans $\mathbb{R}$ telle que, pour tout intervalle I de $\mathbb{R}$, on ait $X^{-1}(I) \in \mathcal{A}$.

On rappelle que $X^{-1}(I) = \{\omega \in \Omega \ ; \ X(\omega) \in I\}$ est noté par commodité $[X \in I]$.
On écrira ainsi $[a < X \leqslant b]$ pour $X^{-1}(]a,b])$; $[X \leqslant x]$ pour $X^{-1}(]-\infty, x])$ et $[X = x]$ pour $X^{-1}(\{x\})$.

Cas particulier important : Si $\mathcal{A} = \mathcal{P}(\Omega)$, toute application de Ω vers $\mathbb{R}$ est une v.a.r.

Théorème 3.1 : Une application X de $(\Omega, \mathcal{A})$ dans $\mathbb{R}$ est une variable aléatoire réelle si et seulement si, pour tout $x \in \mathbb{R}$, on a $[X \leqslant x] \in \mathcal{A}$.

Démonstration. On sait, par définition que X est une v.a.r. si $[X \in I] \in \mathcal{A}$ pour tout I intervalle de $\mathbb{R}$ donc, si X est une v.a.r., en appliquant la définition pour $I_x =]-\infty, x]$, on a bien $[X \in I_x] = [X \leqslant x] \in \mathcal{A}$.

Réciproquement, tout intervalle I de $\mathbb{R}$ s'écrit comme réunion, intersection, complémentaire, d'une suite d'intervalles de la forme $]-\infty, x]$.

Par exemple, si $a < b$, $]a, b[= \left(\bigcup_{n \geqslant 1} \left]-\infty, b - \frac{1}{n}\right] \right) \cap \overline{]-\infty, a]}$. On a donc aussi :

$$\begin{aligned} X^{-1}(]a,b[) &= X^{-1}\left(\left(\bigcup_{n \geqslant 1} \left]-\infty, b - \frac{1}{n}\right] \right) \cap \overline{]-\infty, a]} \right) \\ &= \left(\bigcup_{n \geqslant 1} X^{-1}\left(\left]-\infty, b - \frac{1}{n}\right] \right) \right) \cap \overline{X^{-1}(]-\infty, a])}. \end{aligned}$$

On a alors $\bigcup_{n \geqslant 1} X^{-1}\left(\left]-\infty, b - \frac{1}{n}\right] \right) \in \mathcal{A}$ car $\mathcal{A}$ est stable par union dénombrable ($\mathcal{A}$ tribu) ; $\overline{X^{-1}(]-\infty, a])} \in \mathcal{A}$ car $\mathcal{A}$ est stable par complémentation ; et finalement $X^{-1}(]a,b[) \in \mathcal{A}$ car $\mathcal{A}$ est stable par intersection. Les autres cas se traiteraient de manière analogue. □

3.2 Fonctions de répartition

Définition 3.5 : Soit X une v.a.r. définie sur un espace probabilisé $(\Omega, \mathcal{A}, P)$. On appelle *fonction de répartition* de X l'application F_X de $\mathbb{R}$ sur $\mathbb{R}$ définie par

$$F_X(x) = P([X \leqslant x]).$$

Remarque : F_X est bien définie sur $\mathbb{R}$ car $[X \leqslant x] \in \mathcal{A}$ pour tout $x \in \mathbb{R}$.

Propriétés 3.6 :
1) Pour tout $x \in \mathbb{R}$, $F_X(x) \in [0, 1]$;
2) F_X est croissante ;
3) F_X est continue à droite ;
4) $\lim_{x \to -\infty} F_X(x) = 0$; $\lim_{x \to +\infty} F_X(x) = 1$;
5) pour tout $(a, b) \in \mathbb{R}^2$, si $a < b$, alors $P([a < X \leqslant b]) = F_X(b) - F_X(a)$;
6) $P([X = x]) = F_X(x) - F_X(x_-)$ ($= 0$ si F_X est continue en x).

Démonstration. **1)** Évident : $F_X(x) = P([X \leqslant x]) \in [0, 1]$ car c'est la probabilité d'un événement.

2) et **5)** Si $x' \leqslant x$, alors $[X \leqslant x'] = [X \leqslant x] \cup [x < X \leqslant x']$ (réunion de 2 ensembles disjoints) ; donc $P([X \leqslant x']) = P([X \leqslant x]) + P([x < X \leqslant x'])$, c'est-à-dire

$$F_X(x') = F_X(x) + P([x < X \leqslant x']).$$

Comme $P([x < X \leqslant x']) \geqslant 0$, on a bien $F_X(x) \leqslant F_X(x')$, c'est-à-dire F_X croissante, et on obtient 5) en prenant $x = a$ et $x' = b$.

3) On utilise le résultat d'analyse suivant :

Toute fonction monotone F admet en tout point x des limites à gauche (notée $F(x_-)$) et à droite (notée $F(x_+)$) ;

si $\lim_{n, x_n < x} x_n = x$, alors $\lim_{n \to +\infty} F(x_n) = F(x_-)$

si $\lim\limits_{n, x_n > x} x_n = x$, alors $\lim\limits_{n\to+\infty} F(x_n) = F(x_+)$.

Ici, on écrit $]-\infty, x] = \bigcap\limits_n \left]-\infty, x + \frac{1}{n}\right]$. On a alors $[X \leqslant x] = \bigcap\limits_n D_n$ avec $D_n = \left[X \leqslant x + \frac{1}{n}\right]$.

(D_n) est une suite décroissante d'événements, donc, d'après la propriété 1.6 6), on a alors $P\left[\bigcap\limits_n D_n\right] = \lim\limits_n P(D_n)$, c'est-à-dire $P([X \leqslant x]) = \lim\limits_n P\left(\left[X \leqslant x + \frac{1}{n}\right]\right)$, soit $F_X(x) = \lim\limits_n F_X\left(x + \frac{1}{n}\right) = F_X(x_+)$, d'après le résultat d'analyse que l'on vient d'évoquer.

On a donc $F_X(x) = F_X(x_+)$ et F_X est bien continue à droite.

4) $\Omega = \bigcup\limits_n [X \leqslant n]$ donc $P(\Omega) = 1 = \lim\limits_n P([X \leqslant n]) = \lim\limits_n F_X(n) = \lim\limits_{x\to+\infty} F_X(x)$ par limite croissante.
De même, $\emptyset = \bigcap\limits_n [X \leqslant -n]$ donc $P(\emptyset) = 0 = \lim\limits_{n\to+\infty} F_X(-n) = \lim\limits_{x\to-\infty} F_X(x)$ par limite décroissante.

6) $[X \leqslant x] = [X < x] \cup [X = x]$. Or $[X < x] = \bigcup\limits_n \left[X \leqslant x - \frac{1}{n}\right]$ et, par propriété de limite croissante,

$$P([X < x]) = \lim_n P\left(\left[X \leqslant x - \frac{1}{n}\right]\right) = \lim_n F_X\left(x - \frac{1}{n}\right) = F_X(x_-).$$

Donc, $F_X(x) = F_X(x_-) + P([X = x])$. □

Réciproque très importante : Toute fonction F vérifiant 2), 3) et 4) peut être considérée comme la fonction de répartition d'une variable réelle.

Recommandation : L'étude d'une v.a.r. commence toujours par la détermination de l'ensemble $X(\Omega)$ des valeurs prises par X.

3.3 Variables aléatoires réelles discrètes

3.3.1 Définition, fonction de répartition et loi d'une v.a.r. discrète

Définition 3.6 : Une v.a.r. X sur $(\Omega, \mathcal{A}, P)$ est dite *discrète* si $X(\Omega)$ est fini ou dénombrable.

On note $X(\Omega) = \{x_1, \cdots, x_n\}$ ou bien $X(\Omega) = \{x_i\ ;\ i \in \mathbb{N}\}$ et $p_i = P([X = x_i])$.

Remarque : On a, pour tout $x \in \mathbb{R}$, $[X \leqslant x] = \bigcup\limits_{i; x_i \leqslant x} [X = x_i]$ et une application X de Ω vers $\mathbb{R}$ est donc une v.a.r. discrète à valeurs dans $\{x_i\ ;\ i \in \mathbb{N}\}$ si, pour tout $i \in \mathbb{N}$, on a $[X = x_i] \in \mathcal{A}$. Sa loi P_X est alors entièrement déterminée par les $P([X = x_i]) = p_i$. Ainsi, on peut reformuler dans ce cas :

Définition 3.7 : On appelle *loi de probabilité d'une v.a.r. discrète* X, l'ensemble des couples (x_i, p_i).

Remarque : On a $p_i \geqslant 0$ et $\sum\limits_i p_i = 1$. On notera également $p_i = P_X(\{x_i\})$.

Fonction de répartition : $F_X(x) = \sum\limits_{i; x_i \leqslant x} p_i$.

Sur $[x_{i-1}, x_i[$, on a $F_X(x) = F_X(x_{i-1})$ et $F_X(x_i) - F_X(x_{i-1}) = P([X = x_i]) = p_i$:

La fonction de répartition F_X d'une v.a.r. discrète X est ainsi une fonction en escalier croissante, présentant des sauts de p_i en chaque x_i.

3.3.2 Fonction d'une v.a.r. discrète

Théorème 3.2 : Soit X une v.a.r. discrète et φ une fonction numérique définie sur $X(\Omega)$; alors $Y = \varphi(X)$ est une v.a.r. discrète vérifiant $Y(\Omega) = \varphi(X(\Omega))$ et, pour tout $y \in Y(\Omega)$, on a :

$$P([Y = y]) = \sum_{x_i; \varphi(x_i)=y} P([X = x_i]).$$

Démonstration. Soit, pour $y \in Y(\Omega)$, $I_y = \{x_i \in X(\Omega) \ ; \ \varphi(x_i) = y\}$. On a alors

$$[Y = y] = \{\omega \in \Omega \ ; \ \varphi(X(\omega)) = y\} = \bigcup_{x_i \in I_y} \{\omega \in \Omega \ ; \ X(\omega) = x_i\} = \bigcup_{x_i \in I_y} [X = x_i].$$

D'où $P([Y = y]) = \sum_{x_i \in I_y} P([X = x_i]) = \sum_{x_i; \varphi(x_i)=y} P([X = x_i])$. □

3.3.3 Lois discrètes classiques

1) Lois discrètes finies.

a) Loi de Bernoulli $\mathcal{B}(p)$ (ou $\mathcal{B}(1, p)$) :

$$X(\Omega) = \{0, 1\} \ ; \ P([X = 1]) = p \ ; \ P([X = 0]) = 1 - p.$$

Schéma théorique : On considère une expérience $\mathcal{E}$ et un événement A lié à $\mathcal{E}$ tel que $P(A) = p$. On effectue une fois $\mathcal{E}$ et on appelle X le nombre de réalisation de A. On a donc $X = 1$ si A est réalisé et $X = 0$ si A n'est pas réalisé. d'où $[X = 1] = A$ et $[X = 0] = \overline{A}$. X est bien une v.a.r. car $A \in \mathcal{A}$: on dit que X est la variable indicatrice de l'événement A et on note $X = \mathbb{1}_A$. On a bien $X(\Omega) = \{0, 1\}$; $P([X = 1]) = p$; $P([X = 0]) = 1 - p$.

b) Loi binomiale $\mathcal{B}(n, p)$:

$$X(\Omega) = \{0, \cdots, n\} \ ; \ P([X = k]) = C_n^k p^k (1 - p)^{n-k} \text{ pour tout } k \in X(\Omega).$$

Schéma théorique : On considère l'expérience $\mathcal{E}$ du a) et l'événement A lié à $\mathcal{E}$ tel que $P(A) = p$. On effectue n fois $\mathcal{E}$ et on appelle X le nombre de réalisations de A au cours des n épreuves. On a bien $X(\Omega) = \{0, 1, \cdots, n\}$.

Du point de vue de A, le résultat des n expériences peut être représenté par un n-uplet formé de 0 et de 1. On a alors $X = k$ si et seulement si le n-uplet est formé de k "1" et de $n - k$ "0". Il y a C_n^k tels n-uplets (autant que de façons de placer exactement k "1" dans un n-uplet), et ils ont tous la même probabilité $p^k(1 - p)^{n-k}$ (les expériences étant effectuées de façon indépendante, les probabilités se multiplient).

On a donc bien $X(\Omega) = \{0, 1, \cdots, n\}$ et, pour tout $k \in X(\Omega)$,

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

Remarque : On vérifie que $\sum_{k=0}^{n} P([X=k]) = 1$: c'est la formule du binôme.

De plus, pour $n = 1$, $\mathcal{B}(1,p) = \mathcal{B}(p)$.

c) Loi hypergéométrique $\mathcal{H}(n, M, N)$:

$$X(\Omega) = \{\max(0,\ n-N+M), \cdots, \min(n,\ M)\}\ ;\ P([X=k]) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \text{ pour } k \in X(\Omega).$$

Schéma théorique : Soit E un ensemble constitué de N objets dont M de type 1 et $N-M$ de type 2. On effectue n tirages sans remise dans E ($n \leqslant N$). Soit X le nombre d'objets de type 1 obtenus.

On ne peut avoir $X = k$ que si $0 \leqslant k \leqslant M$ et $0 \leqslant n-k \leqslant N-M$, c'est-à-dire $0 \leqslant k \leqslant M$ et $n-N+M \leqslant k \leqslant n$. Les valeurs prises par X sont donc les entiers k tels que $\max(0, n-N+M) \leqslant k \leqslant \min(n, M)$.

Un élément ω de $[X=k]$ est constitué de k objets de type 1 parmi M et de $n-k$ objets de type 2 parmi $N-M$. Il y a C_M^k façons de choisir les objets de type 1 et C_{N-M}^{n-k} objets de type 2. Or il y a en tout C_N^n façons de choisir n objets parmi N.

On a donc bien $P([X=k]) = \dfrac{C_M^k C_{N-M}^{n-k}}{C_N^n}$ pour tout $k \in \{\max(0, n-N+M), \cdots, \min(n, M)\}$.

Remarque : Si $M \geqslant n$ et $N-M \geqslant n$ (cas le plus fréquent), alors $X(\Omega) = \{0, \cdots, n\}$.

Attention : Selon les ouvrages, l'ordre des paramètres dans la dénomination de la loi peut changer.

d) Loi équiprobable $\mathcal{U}(\{x_1, \cdots, x_n\})$:

$$X(\Omega) = \{x_1, \cdots, x_n\}\ ;\ P([X = x_k]) = \frac{1}{n} \text{ pour tout } x_k \in X(\Omega).$$

2) <u>Lois discrètes infinies.</u>

a) Loi géométrique sur $\mathbb{N}^*$ $\mathcal{G}(p)$ (ou $\mathcal{P}(1,p)$) :

$$X(\Omega) = \mathbb{N}^*\ ;\ P([X=k]) = p(1-p)^{k-1} \text{ pour } k \in \mathbb{N}^*.$$

Schéma théorique : On considère toujours l'expérience $\mathcal{E}$ du 1)a) et l'événement A lié à $\mathcal{E}$ tel que $P(A) = p$. On répète maintenant $\mathcal{E}$ dans les mêmes conditions jusqu'à ce que A soit réalisé. On note X le nombre d'épreuves effectuées.

On a bien $X(\Omega) = \mathbb{N}^*$ et, si l'on note A_i l'événement "A est réalisé à la i-ème épreuve", on a alors $[X=1] = A_1$ d'où $P([X=1]) = p$ et, pour $k \geqslant 2$, $[X=k] = \overline{A_1} \cap \cdots \cap \overline{A_{k-1}} \cap A_k$ d'où

$$P([X=k]) = P\left(\overline{A_1} \cap \cdots \cap \overline{A_{k-1}} \cap A_k\right) = p(1-p)^{k-1}.$$

D'où, finalement, $X(\Omega) = \mathbb{N}^*$ et, pour tout $k \in \mathbb{N}^*$, $P([X=k]) = p(1-p)^{k-1}$.

b) Loi de Pascal $\mathcal{P}(r,p)$:

$$X(\Omega) = \mathbb{N} \setminus \{0, \cdots, r-1\}\ ;\ P([X=k]) = C_{k-1}^{r-1} p^r (1-p)^{k-r} \text{ pour } k \geqslant r.$$

Schéma théorique : On considère toujours l'expérience $\mathcal{E}$ du 1)a) et l'événement A lié à $\mathcal{E}$ tel que $P(A) = p$. On répète maintenant $\mathcal{E}$ dans les mêmes conditions jusqu'à ce que A soit réalisé exactement r fois. On note Y_r le nombre d'épreuves effectuées, c'est-à-dire le rang de la r-ième réalisation de A. On a bien $Y_r(\Omega) = \mathbb{N} \setminus \{0, \cdots, r-1\}$ et $Y_r = k$ signifie qu'à l'épreuve k, l'événement A est réalisé pour la r-ième fois, c'est-à-dire que, jusqu'à la $(k-1)$-ième épreuve, il y a eu $r-1$ réalisations de A exactement et qu'à l'épreuve k, A est aussi réalisé. Le nombre d'éléments de $[Y_r = k]$ est donc le nombre de k-uplets se terminant par 1 et contenant exactement $r-1$ "1" dans les $k-1$ premières places (et donc $k-r$ "0").

Il y a C_{k-1}^{r-1} tels k-uplets, tous de même probabilité $p^r(1-p)^{k-r}$.

On a donc $Y_r(\Omega) = \{r, r+1, \cdots\}$ et pour tout $k \geqslant r$, $P([Y_r = k]) = C_{k-1}^{r-1} p^r (1-p)^{k-r}$.

c) Loi géométrique sur IN $\mathcal{G}_0(p)$ (ou $\mathcal{BN}(1,p)$) :

$$X(\Omega) = \mathbb{N}\ ;\ P([X = k]) = p(1-p)^k \text{ pour } k \in \mathbb{N}.$$

d) Loi binomiale négative $\mathcal{BN}(r,p)$:

$$X(\Omega) = \mathbb{N}\ ;\ P([X = k]) = C_{k+r-1}^{r-1} p^r (1-p)^k \text{ pour } k \in \mathbb{N}.$$

Schéma théorique : Dans les conditions du 2)b), on note Z_r le nombre d'échecs précédant la r-ième réalisation de A.

On a donc, si $Y_r = k$, $Z_r + r = k$, c'est-à-dire que $Z_r = Y_r - r$.

Ainsi, $Z_r(\Omega) = \mathbb{N}$ et $P([Z_r = k]) = P([Y_r = k + r]) = C_{k+r-1}^{r-1} p^r (1-p)^k$.

e) Loi de Poisson $\mathcal{P}(\lambda)$:

$$X(\Omega) = \mathbb{N}\ ;\ P([X = k]) = \exp(-\lambda)\frac{\lambda^k}{k!} \text{ pour } k \in \mathbb{N}.$$

Remarque : On a bien $\displaystyle\sum_{k=0}^{+\infty} P([X = k]) = 1$.

Exemple : Le nombre X d'événements aléatoires sur un intervalle T tels que des appels téléphoniques sur un central, où les arrivées des voitures à un péage d'autoroute suit une loi de Poisson.

On verra que λ représente le nombre moyen d'appels où d'arrivées pendant T.

3.4 Variables aléatoires réelles absolument continues

3.4.1 Définition, fonction de répartition, densité

On a vu que, si X est une v.a.r. sa fonction de répartition $F_X : x \mapsto P([X \leqslant x])$ est une fonction croissante, continue à droite et présente un saut en x si $P([X = x]) \neq 0$. Elle vérifie de plus $\lim_{x \to -\infty} F_X(x) = 0$ et $\lim_{x \to +\infty} F_X(x) = 1$.

On va s'intéresser ici à des v.a.r. X telles que $P([X = x]) = 0$ pour tout $x \in \mathbb{R}$, c'est-à-dire telles que F_X soit continue sur IR.

Remarque : Dans ce cas, $X(\Omega)$ ne peut pas être dénombrable. En effet, on aurait alors $P(\Omega) = \displaystyle\sum_{x \in X(\Omega)} P([X = x]) = 0$, ce qui contredit $P(\Omega) = 1$.

On peut commencer par rappeler un résultat d'analyse qui sera utile par la suite :

Rappel : Si f est intégrable sur $[a, b]$, alors, pour tout $x \in [a, b]$, f est intégrable sur $[a, x]$ et si on pose $F(x) = \int_a^x f(t)dt$, on a :
i) F est continue sur $[a, b]$;
ii) F est dérivable à gauche (resp. à droite) en tout point $x_0 \in]a, b]$, (resp. $[a, b[$) où f admet une limite à gauche (resp. à droite) et $F'_g(x_0) = f(x_{0-})$ (resp. $F'_d(x_0) = f(x_{0+})$).

Définition 3.8 : Une v.a.r. X de fonction de répartition F_X est dite *absolument continue* s'il existe une fonction numérique f définie sur $\mathbb{R}$, appelée *densité* de X telle que :
i) $f(x) \geqslant 0$ pour tout $x \in \mathbb{R}$;
ii) pour tout $B \in \mathcal{B}_{\mathbb{R}}$, $f^{-1}(B) = \{x \in \mathbb{R} \; ; \; f(x) \in B\} \in \mathcal{B}_{\mathbb{R}}$;
iii) $\int_{-\infty}^{+\infty} f(t)dt$ existe et vaut 1 ;
iv) $F_X(x) = \int_{-\infty}^{x} f(t)dt$.

Remarque : Une fonction vérifiant ii) est appelée *fonction borélienne*. En particulier, toute fonction continue ou "suffisamment continue" (par exemple, continue par morceaux) de $\mathbb{R}$ dans $\mathbb{R}$ est borélienne. Dans les exercices ou exemples, on ne considèrera que des fonctions f continues ou continues par morceaux : ainsi, le point ii) sera automatiquement vérifié.

Conséquence très importante de la définition : Si X est absolument continue, on a, si $a < b$:

$$\begin{aligned} P([a < X < b]) &= P([a < X \leqslant b]) = P([a \leqslant X < b]) = P([a \leqslant X \leqslant b]) \\ &= F_X(b) - F_X(a) = \int_a^b f(t)dt. \end{aligned}$$

En effet, $P([X = a]) = P([X = b]) = 0$ car F_X est continue, et $F_X(b) - F_X(a) = \int_{-\infty}^b f(t)dt - \int_{-\infty}^a f(t)dt = \int_a^b f(t)dt$ (par la relation de Chasles).

Remarques :
1) La connaissance de f détermine entièrement F_X mais la connaissance de F_X ne détermine pas f de façon unique : on peut modifier à son gré f sur un ensemble fini sans changer $\int_{-\infty}^x f(t)dt$. Dans la pratique, f sera choisie la moins discontinue possible. Par abus de langage, on appellera parfois "loi de X" la densité de X.

2) Si X est une v.a.r. absolument continue, alors F_X est continue, mais la réciproque est fausse.

3) Il existe des v.a.r. qui ne sont ni discrètes, ni absolument continues (celles, par exemple, dont la fonction de répartition est strictement croissante et présente des sauts en certains points).

Comment reconnaître la fonction de répartition d'une v.a.r. absolument continue ?

Une fonction de répartition F est celle d'une v.a.r. X absolument continue si F est continue sur $\mathbb{R}$, dérivable sur $\mathbb{R} \setminus I$, où I est un sous-ensemble fini (ou dénombrable) de $\mathbb{R}$, si F' est continue sur $\mathbb{R} \setminus I$ et si F' admet des limites à droite et à gauche en tout point de I.

Exemples :

1) Montrer que F définie par $F(x) = \dfrac{e^x}{2}\mathbb{I}_{]-\infty,0[}(x) + \left(\dfrac{x^2+8}{16}\right)\mathbb{I}_{[0,2[}(x) + \left(1 - \dfrac{e^{2-x}}{4}\right)\mathbb{I}_{[2,+\infty[}(x)$ est la fonction de répartition d'une v.a.r. absolument continue dont on déterminera une densité.

Démonstration. $\rightarrow$ F est continue sur $\mathbb{R} \setminus \{0,2\}$; d'autre part :

$$F(0_-) = \frac{e^0}{2} = \frac{1}{2} \; ; \; F(0_+) = \frac{8}{16} = \frac{1}{2} \; ; \; F(2_-) = \frac{4+8}{16} = \frac{3}{4} \; ; \; F(2_+) = 1 - \frac{e^0}{4} = \frac{3}{4}$$

donc F est continue sur $\mathbb{R}$.

$\rightarrow$ F est dérivable sur $\mathbb{R} \setminus \{0,2\}$ avec

$$F'(x) = \begin{cases} \dfrac{e^x}{2} > 0 & \text{si } x < 0 \\ \dfrac{x}{8} > 0 & \text{si } x \in]0,2[\\ \dfrac{e^{2-x}}{4} > 0 & \text{si } x > 2. \end{cases}$$

On a donc $F' \geqslant 0$ sur $\mathbb{R} \setminus \{0,2\}$ donc F est croissante sur les intervalles $]-\infty,0[$, $]0,2[$ et $]2,+\infty[$. Comme de plus, F est continue en 0 et en 2, F est croissante sur $\mathbb{R}$.

$\rightarrow$ $\lim\limits_{x\to-\infty} F(x) = \lim\limits_{x\to-\infty} \dfrac{e^x}{2} = 0$ et $\lim\limits_{x\to+\infty} F(x) = \lim\limits_{x\to+\infty} \left(1 - \dfrac{e^{2-x}}{4}\right) = 1$.

Ainsi, F est bien une fonction de répartition continue sur $\mathbb{R}$.

$\rightarrow$ De plus, F' est continue sur $\mathbb{R} \setminus \{0,2\}$ et

$$F'(0_-) = \frac{1}{2} \; ; \; F'(0_+) = 0 \; ; \; F'(2_-) = \frac{1}{4} \; ; \; F'(2_+) = \frac{1}{4}$$

donc F admet aussi des dérivées à droite et à gauche en 0 et en 2.

Finalement F est la fonction de répartition d'une v.a.r. absolument continue dont une densité est la fonction f définie par :

$$f(x) = \frac{e^x}{2}\mathbb{I}_{]-\infty,0[}(x) + \frac{x}{8}\mathbb{I}_{[0,2[}(x) + \frac{e^{2-x}}{4}\mathbb{I}_{]2,+\infty[}(x).$$

□

2) Montrer que la fonction f définie par $f(x) = \cos x \, \mathbb{I}_{[0,\frac{\pi}{2}]}(x)$ est la densité de probabilité d'une v.a.r. absolument continue dont on déterminera la fonction de répartition.

Démonstration. $\rightarrow$ f est une densité car :

i) $f(x) \geqslant 0$ pour tout $x \in \mathbb{R}$

ii) f est continue sur $\mathbb{R}^*$; $f(0_-) = 0$, $f(0_+) = 1$; $f\left(\dfrac{\pi}{2}_-\right) = f\left(\dfrac{\pi}{2}_+\right) = 0$;

iii) $\displaystyle\int_{-\infty}^{+\infty} f(t)dt = \int_{-\infty}^{0} f(t)dt + \int_{0}^{\pi/2} f(t)dt + \int_{\pi/2}^{+\infty} f(t)dt = \int_{0}^{\pi/2} \cos t \, dt = [\sin t]_0^{\pi/2} = 1.$

$\rightarrow$ Soit X une v.a.r. ayant f pour densité. On a alors $F_X(x) = \displaystyle\int_{-\infty}^{x} f(t)dt$;

- si $x < 0$, alors $F_X(x) = 0$;
- si $x \in \left[0, \dfrac{\pi}{2}\right]$, $F_X(x) = \displaystyle\int_{-\infty}^{0} f(t)dt + \int_{0}^{x} f(t)dt = \int_{0}^{x} \cos t \, dt = \sin x$;
- si $x > \dfrac{\pi}{2}$, alors $F_X(x) = \displaystyle\int_{-\infty}^{0} f(t)dt + \int_{0}^{\pi/2} f(t)dt + \int_{\pi/2}^{x} f(t)dt = 1.$

On a donc $F_X(x) = \sin x \, \mathbb{I}_{[0,\pi/2[}(x) + \mathbb{I}_{[\pi/2,+\infty[}(x)$.

F_X est continue sur $\mathbb{R}$, dérivable sur $\mathbb{R}^*$ (en particulier en $\pi/2$ aussi), non dérivable en 0 (point de discontinuité de f). □

3.4.2 Loi d'une fonction d'une v.a.r. absolument continue

Soit X une v.a.r. absolument continue (de densité f_X et de fonction de répartition F_X) et φ une fonction réelle définie au moins sur un un intervalle I contenant $X(\Omega)$. Pour déterminer la loi de $Y = \varphi(X)$, la méthode la plus générale consiste à exprimer sa fonction de répartition.

$$F_Y(y) = P([Y \leqslant y]) = P([\varphi(X) \leqslant y]) = P\left(\left[X \in \varphi^{-1}(]-\infty, y])\right]\right)$$

où $\varphi^{-1}(]-\infty, y]) = \{x \in \mathbb{R}\ ;\ \varphi(x) \leqslant y\}$ que l'on note ici $B(y)$.

Le plus souvent, $B(y)$ sera un intervalle ou bien une réunion d'intervalles disjoints et, comme $(X \in \bigcup\limits_k I_k) = \bigcup\limits_k [X \in I_k]$, il sera alors aisé d'exprimer F_Y à l'aide de F_X.

Exemples :

1) Si $Y = aX + b$, avec $a \neq 0$, $F_Y(y) = P([Y \leqslant y]) = P([aX + b \leqslant y]) = P([aX \leqslant y - b])$.

• Si $a > 0$, $F_Y(y) = P\left(\left[X \leqslant \dfrac{y-b}{a}\right]\right) = F_X\left(\dfrac{y-b}{a}\right)$. Si F_X est dérivable sauf éventuellement en quelques points, il en est de même pour F_Y par composition, et on a alors $f_Y(y) = F_Y'(y) = \dfrac{1}{a}F_X'\left(\dfrac{y-b}{a}\right) = \dfrac{1}{a}f_X\left(\dfrac{y-b}{a}\right)$.

• Si $a < 0$, $F_Y(y) = P\left(\left[X \geqslant \dfrac{y-b}{a}\right]\right) = 1 - F_X\left(\dfrac{y-b}{a}\right)$ car F_X et continue. Ici aussi, F_Y est dérivable (sauf en quelques points) et, en utilisant $(F \circ g)' = (F' \circ g) \times g'$, on a alors $f_Y(y) = F_Y'(y) = -\dfrac{1}{a}F_X'\left(\dfrac{y-b}{a}\right) = -\dfrac{1}{a}f_X\left(\dfrac{y-b}{a}\right)$.

Finalement, $f_{aX+b}(y) = \dfrac{1}{|a|}f_X\left(\dfrac{y-b}{a}\right)$.

2) Si $Y = X^2$, alors $F_Y(y) = P([Y \leqslant y]) = P([X^2 \leqslant y])$.

• Si $y < 0$, alors $F_Y(y) = 0$ et $f_Y = F_Y' = 0$ sur $\mathbb{R}_-^*$.

• Si $y > 0$, alors $F_Y(y) = P([-\sqrt{y} \leqslant X \leqslant \sqrt{y}]) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$ et $f_Y(y) = \dfrac{1}{2\sqrt{y}}\left(f_X(\sqrt{y}) + f_X(-\sqrt{y})\right)$ dans ce cas.

Finalement, $f_{X^2}(y) = \dfrac{1}{2\sqrt{y}}\left(f_X(\sqrt{y}) + f_X(-\sqrt{y})\right) \mathbb{1}_{]0,+\infty[}(y)$.

3) Si $Y = e^X$, alors $F_Y(y) = P([Y \leqslant y]) = P([e^X \leqslant y])$.

• Si $y \leqslant 0$, alors $F_Y(y) = 0$ et $f_Y = F_Y' = 0$.

• Si $y > 0$, alors $F_Y(y) = P([X \leqslant \ln y]) = F_X(\ln y)$ car $t \mapsto \ln t$ est une bijection strictement croissante de $\mathbb{R}_+^*$ sur $\mathbb{R}$, puis $f_Y(y) = F_Y'(y) = \dfrac{1}{y}f_X(\ln y)$.

Finalement, $f_{e^X}(y) = \dfrac{1}{y}f_X(\ln y)\, \mathbb{1}_{]0,+\infty[}(y)$.

De façon plus générale, on pourra utiliser, lorsque φ est bijective, le résultat suivant :

Théorème 3.3 : Soit X une v.a.r. absolument continue de densité f_X et φ une fonction numérique de classe C^1 (continue et dérivable de dérivée continue) sur un intervalle I contenant $X(\Omega)$, telle que φ' ne s'annule en aucun point de I ; alors $Y = \varphi(X)$ est une v.a.r. absolument continue et elle admet la densité f_Y définie par :

$$f_Y(y) = \frac{f_X(\varphi^{-1}(y))}{|\varphi'(\varphi^{-1}(y))|} \mathbb{1}_{\varphi(I)}(y).$$

Démonstration. φ' est continue et ne s'annule pas sur I ; elle y garde donc un signe constant.

On part de $F_Y(y) = P([Y \leqslant y]) = P([\varphi(X) \leqslant y])$.

- Si $y \geqslant \sup \varphi(X(\Omega))$, alors $P([\varphi(X) \leqslant y])$ est toujours vérifié et $F_Y(y) = 1$;
- Si $y < \inf \varphi(X(\Omega))$, alors $P([\varphi(X) \leqslant y])$ n'est jamais vérifié et $F_Y(y) = 0$;
- Si $y \in [\inf \varphi(X(\Omega)), \sup \varphi(X(\Omega))]$, alors $y \in \varphi(I)$ car $\varphi(X(\Omega)) \subset \varphi(I)$.

→ Si $\varphi' > 0$, alors φ et aussi φ^{-1} sont strictement croissantes sur I. Il vient donc

$$F_Y(y) = P([\varphi(X) \leqslant y]) = P([X \leqslant \varphi^{-1}(y)]) = F_X(\varphi^{-1}(y)).$$

F_X est continue, et si elle est dérivable sauf éventuellement en un nombre fini ou dénombrable de points, comme φ^{-1} est dérivable, par composition, $F_Y = F_X \circ \varphi^{-1}$ est dérivable sur $\varphi(I)$, sauf peut-être en un nombre fini ou dénombrable de points et :

$$F_Y'(y) = (F_X \circ \varphi^{-1})'(y) = F_X'(\varphi^{-1}(y))(\varphi^{-1})'(y) = \frac{f_X(\varphi^{-1}(y))}{\varphi'(\varphi^{-1}(y))}.$$

→ Si $\varphi' < 0$, alors φ et aussi φ^{-1} sont strictement décroissantes sur I. Il vient donc

$$F_Y(y) = P([\varphi(X) \leqslant y]) = P([X \geqslant \varphi^{-1}(y)]) = 1 - F_X(\varphi^{-1}(y)).$$

F_X est continue, dérivable sauf éventuellement en un nombre fini ou dénombrable de points, et φ^{-1} est dérivable. D'où $F_Y = 1 - F_X \circ \varphi^{-1}$ est dérivable sur $\varphi(I)$, sauf peut-être en un nombre fini ou dénombrable de points et :

$$F_Y'(y) = -(F_X \circ \varphi^{-1})'(y) = -F_X'(\varphi^{-1}(y))(\varphi^{-1})'(y) = -\frac{f_X(\varphi^{-1}(y))}{\varphi'(\varphi^{-1}(y))}.$$

Donc, dans les deux cas, $f_Y(y) = \dfrac{f_X(\varphi^{-1}(y))}{|\varphi'(\varphi^{-1}(y))|}$. □

Remarque : Si φ' s'annule en un nombre fini de points sans changer de signe, le résultat précédent reste valable en tout point où il a un sens.

<u>*Exemple d'application*</u> : Si $\varphi = \tan$ et $f_X(x) = \cos x \, \mathbb{1}_{]0,\frac{\pi}{2}[}(x)$, $X(\Omega) = I = \left]0, \dfrac{\pi}{2}\right[$.

L'application φ est une bijection croissante de $\left]0, \dfrac{\pi}{2}\right[$ sur $]0, +\infty[$, avec $\varphi^{-1}(y) = \text{Arctan}(y)$ et $(\varphi^{-1})'(y) = \dfrac{1}{1+y^2}$.

$$f_{\tan X}(y) = \frac{\cos(\text{Arctan} y)}{1+y^2} \mathbb{1}_{]0,+\infty[}(y) = (1+y^2)^{-3/2} \mathbb{1}_{]0,+\infty[}(y).$$

3.4.3 Lois absolument continues classiques

Rappel : Si Δ est un sous-ensemble de $\mathbb{R}$, alors $\mathbb{1}_\Delta(x) = \begin{cases} 1 \text{ si } x \in \Delta \\ 0 \text{ si } x \notin \Delta \end{cases}$. La fonction $\mathbb{1}_\Delta$ est appelée <u>*fonction indicatrice*</u> de Δ.

a) Loi uniforme $\mathcal{U}([a,b])$ **:**

loi de densité f définie par $f(x) = \dfrac{1}{b-a}\mathbb{1}_{[a,b]}(x)$.

On a dans ce cas, $F(x) = \displaystyle\int_{-\infty}^{x} f(t)\,dt = \begin{cases} 0 & \text{si } x < a \\ \dfrac{x-a}{b-a} & \text{si } x \in [a,b[\\ 1 & \text{si } x \geqslant b. \end{cases}$, c'est-à-dire

$$F(x) = \frac{x-a}{b-a}\,\mathbb{1}_{[a,b[}(x) + \mathbb{1}_{[b,+\infty[}(x).$$

b) Loi exponentielle $\mathcal{E}(\lambda)$ **:**

loi de densité f définie par $f(x) = \lambda\exp(-\lambda x)\mathbb{1}_{]0,+\infty[}(x)$.

On a dans ce cas, $F(x) = \displaystyle\int_{-\infty}^{x} f(t)dt = \begin{cases} 0 & \text{si } x < 0 \\ \displaystyle\int_0^x \lambda e^{-\lambda t}\,dt & \text{si } x \geqslant 0 \end{cases}$,

où $\displaystyle\int_0^x \lambda e^{-\lambda t}dt = \left[-e^{-\lambda t}\right]_0^x = 1 - e^{-\lambda x}$, c'est-à-dire

$$F(x) = \left(1 - e^{-\lambda x}\right)\,\mathbb{1}_{[0,+\infty[}(x).$$

c) Loi Gamma $\gamma(\lambda, a)$ **(ou bien** $\Gamma(a,\lambda)$**) :**

loi de densité f définie par $f(x) = \dfrac{\lambda^a}{\Gamma(a)}\exp(-\lambda x)x^{a-1}\mathbb{1}_{]0,+\infty[}(x)$ où $\Gamma(a) = \displaystyle\int_0^{+\infty} e^{-t}\,t^{a-1}dt$.

d) Loi Beta $B(a,b)$ **:**

loi de densité f définie par $f(x) = \dfrac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}\mathbb{1}_{]0,1[}(x)$ où $B(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

e) Loi de Cauchy $\mathcal{C}(\alpha)$ **:**

loi de densité f définie par $f(x) = \dfrac{\alpha}{\pi(\alpha^2 + x^2)}$.

f) Loi Normale $\mathcal{N}(m, \sigma^2)$ **:**

loi de densité f définie par $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}}\exp\left(-\dfrac{(x-m)^2}{2\sigma^2}\right)$.

On parle de loi normale lorsque l'on a affaire à une v.a.r. absolument continue, dépendant d'un grand nombre de causes indépendantes entre elles, dont les effets s'additionnent et dont aucune n'est prépondérante.

Exemple : Les dimensions de pièces fabriquées dépendent du réglage de l'appareil de fabrication, des vibrations auxquelles il est soumis, de l'homogénéité de la matière première, de la température, de l'humidité...

Lorsque tous ces facteurs sont indépendants et qu'aucun n'est prépondérant, on peut supposer que les dimensions suivent une loi normale.

m est appelé la moyenne et σ l'écart-type.

Attention : Certains ouvrages écrivent $\mathcal{N}(m,\sigma)$ au lieu de $\mathcal{N}(m,\sigma^2)$.

Formule à retenir : $\displaystyle\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{t^2}{2}} dt = \sqrt{2\pi}$ (intégrale impossible à calculer par les méthodes usuelles).
Pour s'en souvenir, on sait que $t \mapsto \dfrac{1}{\sqrt{2\pi}} e^{-t^2/2}$ est la densité de la loi $\mathcal{N}(0,1)$ donc son intégrale vaut 1.

Étude de f densité de la loi $\mathcal{N}(m,\sigma^2)$:
$f(m+u) = f(m-u)$ pour tout $u \in \mathbb{R}$; d'où symétrie par rapport à la droite d'équation $x = m$.
$f'(u) = -\dfrac{(u-m)}{\sigma^2} f(u)$ donc f est croissante jusqu'à m, puis décroissante ;
$f''(u) = \left(-\dfrac{1}{\sigma^2} + \dfrac{(u-m)^2}{\sigma^4}\right) f(u) = \left(\dfrac{(u-m)^2 - \sigma^2}{\sigma^4}\right) f(u)$ donc f est concave sur $[m - \sigma, m+\sigma]$, convexe sinon. On obtient une "courbe en cloche".
De plus, $f(m) = \dfrac{1}{\sqrt{2\pi}}$, $\displaystyle\lim_{u\to\pm\infty} f(u) = 0$.
Plus σ est petit, plus la courbe est "étroite" et "monte haut".

Théorème 3.4 : Une v.a.r. X a pour loi la loi normale $\mathcal{N}(m,\sigma^2)$ si et seulement si $X^* = \dfrac{X-m}{\sigma}$ est une v.a.r. de loi normale $\mathcal{N}(0,1)$.

Démonstration. On a vu que $f_{aX+b}(y) = \dfrac{1}{|a|} f_X\left(\dfrac{y-b}{a}\right)$.
Si X a pour loi la loi normale $\mathcal{N}(m,\sigma^2)$, alors $f_X(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left(-\dfrac{(x-m)^2}{2\sigma^2}\right)$ et alors, comme $X^* = \dfrac{1}{\sigma}X - \dfrac{m}{\sigma}$, on applique la formule avec $a = \dfrac{1}{\sigma}$ et $b = -\dfrac{m}{\sigma}$. On a alors :

$$f_{X^*}(y) = \sigma f_X\left(\sigma\left(y + \frac{m}{\sigma}\right)\right) = \sigma f_X(\sigma y + m) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2},$$

et X^* suit bien la loi $\mathcal{N}(0,1)$.
Réciproquement, si $f_{X^*}(x) = \dfrac{1}{\sqrt{2\pi}} e^{-x^2/2}$, comme $X = \sigma X^* + m$, on a alors :

$$f_X(y) = \frac{1}{\sigma} f_{X^*}\left(\frac{y-m}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-m)^2}{2\sigma^2}}$$

et X suit bien la loi $\mathcal{N}(m,\sigma^2)$. □

Conséquence : Tout calcul de probabilité faisant intervenir la loi normale $\mathcal{N}(m,\sigma^2)$ se ramène à un calcul faisant intervenir la loi normale $\mathcal{N}(0,1)$.

On note Φ la fonction de répartition d'une v.a.r. de loi $\mathcal{N}(0,1)$:

$$\boxed{\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt}$$

Comme on ne sait pas calculer Φ, il existe des tables qui donnent des valeurs approchées de $\Phi(x)$ pour tout $x \in \mathbb{R}$.

Théorème 3.5 : Soit X une v.a.r. de loi normale $\mathcal{N}(0,1)$, de fonction de répartition Φ. Alors :
1) pour tout $x \in \mathbb{R}$, $\Phi(x) = 1 - \Phi(-x)$; en particulier $\Phi(0) = \dfrac{1}{2}$;
2) pour tout $x \in \mathbb{R}$, $P([|X| \leqslant x]) = 2\Phi(x) - 1$ et $P([|X| \geqslant x]) = 2(1 - \Phi(x))$.

Démonstration. **1)** En faisant le changement de variable $u = -t$, on obtient :

$$\begin{aligned}\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt &= -\int_{+\infty}^{-x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du = \int_{-x}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du - \int_{-\infty}^{-x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du\end{aligned}$$

et on a bien $\Phi(x) = 1 - \Phi(-x)$. En particulier $\Phi(0) = 1 - \Phi(-0) = 1 - \Phi(0)$, d'où $2\Phi(0) = 1$.

2) $P([|X| \leqslant x]) = P([-x \leqslant X \leqslant x]) = \Phi(x) - \Phi(-x) = 2\Phi(x) - 1$.
De plus, $P([|X| \geqslant x]) = 1 - P([|X| < x]) = 1 - (2\Phi(x) - 1) = 2(1 - \Phi(x))$. □

Exemple : Le poids d'un bébé est une v.a.r. de loi $\mathcal{N}(3.2\,,\, 0.16)$. Calculer la probabilité que le bébé pèse 1) plus de 4 kg ? 2) moins de 3 kg ? 3) entre 2.8 kg et 3.6 kg ?

Soit X la v.a.r. poids du bébé : elle a pour loi $\mathcal{N}(m, \sigma^2)$ avec $m = 3.2$ et $\sigma = 0.4$ et la v.a.r. $X^* = \dfrac{X - 3.2}{0.4}$ suit donc la loi $\mathcal{N}(0,1)$.

1) $P([X \geqslant 4]) = P\left(\left[\dfrac{X - 3.2}{0.4} \geqslant \dfrac{4 - 3.2}{0.4}\right]\right) = P([X^* \geqslant 2]) = 1 - \Phi(2) \approx 0.023$ $(2,3\%)$

2) $P([X \leqslant 3]) = P\left(\left[\dfrac{X - 3.2}{0.4} \leqslant \dfrac{3 - 3.2}{0.4}\right]\right) = P([X^* \leqslant -0.5]) = \Phi(-0.5) = 1 - \Phi(0.5) \approx 0.309$ $(30,9\%)$

3) $P([2.8 \leqslant X \leqslant 3.6]) = P\left(\left[\dfrac{2.8 - 3.2}{0.4} \leqslant \dfrac{X - 3.2}{0.4} \leqslant \dfrac{3.6 - 3.2}{0.4}\right]\right) = P(|X^*| \leqslant 1) = 2\Phi(1) - 1 \approx 0.683$ $(68,3\%)$.

Chapitre 4

PROBABILITÉS CONDITIONNELLES

La notion de probabilité conditionnelle est la plus importante, mais aussi la plus délicate de la théorie des probabilités. Elle est introduite en particulier chaque fois que, pendant le déroulement d'une expérience aléatoire, une information partielle "de dernière minute" est fournie à l'expérimentateur.
Un événement en conditionne un autre, si la réalisation de ce dernier dépend de la réalisation du premier. Les notions d'indépendance et de conditionnement sont donc étroitement liées.

4.1 Introduction

Envisageons, par exemple, les deux problèmes suivants :

Problème 1 : **1)** On lance un dé. Quelle est la probabilité d'obtenir 3 ?

2) On lance un dé et on obtient un chiffre impair. Quelle est la probabilité que ce soit le chiffre 3 ?

Dans le cas 1), on répond 1/6, mais dans le cas 2),on répond 1/3.

Pour modéliser le cas 2), on peut prendre $\Omega' = \{1, 3, 5\}$, muni de l'équiprobabilité (probabilité P' telle que $P'(\{1\}) = P'(\{3\}) = P'(\{5\}) = \dfrac{1}{3}$).

Mais on peut prendre aussi $\Omega = \{1, 2, 3, 4, 5, 6\}$, muni de la probabilité Q telle que :

$$Q(\{1\}) = Q(\{3\}) = Q(\{5\}) = \frac{1}{3} \text{ et } Q(\{2\}) = Q(\{4\}) = Q(\{6\}) = 0.$$

Cette façon de procéder à l'avantage de conserver le même univers que dans le cas 1) et de montrer que la différence entre 1) et 2) est liée à un changement de probabilités.

Problème 2 : Au cours d'une fête foraine, on veut étudier le résultat d'une loterie où il n'y a qu'un numéro gagnant parmi 100. On prend $\Omega = \{1, \cdots, 100\}$ et P l'équiprobabilité sur Ω.

Supposons que l'astrologue de la fête nous assure que le numéro gagnant se termine par un 4 : cette information modifie la loi de probabilité sur Ω : il y a 10 numéros se terminant par 4, chacun ayant la même chance de sortir (1/10) ; les autres n'ont aucune chance de sortir.

Approche fréquentielle de la notion de probabilité conditionnelle :

Soit $\mathcal{E}$ une expérience aléatoire et soient A et B deux événements liés à $\mathcal{E}$. On répète N fois l'expérience. parmi les N_B fois où B est réalisé, il y a eu $N_{A\cap B}$ expériences où A et B ont eu lieu simultanément : la fréquence de A quand B a lieu est donc :

$$\frac{N_{A\cap B}}{N_B} = \frac{\frac{N_{A\cap B}}{N}}{\frac{N_B}{N}} = \frac{f_{A\cap B}}{f_B}.$$

Définition 4.1 : Soit $B \in \mathcal{A}$ tel que $P(B) > 0$. On appelle <u>*probabilité de A sachant B*</u> le nombre $P^B(A)$ (ou $P(A/B)$) défini par $P^B(A) = \dfrac{P(A\cap B)}{P(B)}$.

Remarque : On préfèrera la notation $P(A/B)$ lorsque l'expression de B est trop "longue" pour la mettre en exposant.

<u>*Exemple*</u> :
1) Dans le problème 1) on note A "on obtient le chiffre 3" et B "on obtient un chiffre impair". On a

$$A\cap B = A \ ; \ P(A\cap B) = \frac{\text{card}(A\cap B)}{\text{card}\Omega} = \frac{1}{6} \ ; \ P(B) = \frac{\text{card}B}{\text{card}\Omega} = \frac{3}{6} = \frac{1}{2}.$$

D'où $P^B(A) = \dfrac{P(A\cap B)}{P(B)} = \dfrac{\frac{1}{6}}{\frac{1}{2}} = \dfrac{1}{3}$.

2) Dans le problème 2), on note B"le numéro se termine par un 4".
On a card$B = 10$; $P(B) = \dfrac{10}{100} = \dfrac{1}{10}$ et $P^B(\{\omega\}) = \dfrac{P(\{\omega\}\cap B)}{P(B)}$.
Or $\{\omega\}\cap B = \begin{cases} \{\omega\} & \text{si } \omega \in B \\ \emptyset & \text{si } \omega \notin B \end{cases}$ donc $P(\{\omega\}\cap B) = \begin{cases} \dfrac{1}{100} & \text{si } \omega \in B \\ 0 & \text{si } \omega \notin B \end{cases}$ et

$$P^B(\{\omega\}) = \begin{cases} \dfrac{1}{10} & \text{si } \omega \in B \\ 0 & \text{si } \omega \notin B \end{cases}.$$

Propriété 4.1 : L'application $P^B : A \mapsto P^B(A)$ est une probabilité sur $(\Omega, \mathcal{A})$.

Démonstration. i) $P^B(\Omega) = \dfrac{P(\Omega\cap B)}{P(B)} = \dfrac{P(B)}{P(B)} = 1$ et $P^B(A) = \dfrac{P(A\cap B)}{P(B)} \in [0,1]$ car $A\cap B \subset B$.
ii) Pour toute suite d'événements A_n, incompatibles 2 à 2, les événements $A_n\cap B$ sont aussi incompatibles 2 à 2 et on a, en utilisant la distributivité de l'intersection par rapport à la réunion dénombrable, puis le fait que P est une probabilité :

$$\begin{aligned} P^B\left(\bigcup_n A_n\right) &= \frac{P\left(\left(\bigcup_n A_n\right)\cap B\right)}{P(B)} = \frac{P\left(\bigcup_n (A_n\cap B)\right)}{P(B)} \\ &= \frac{\sum_n P(A_n\cap B)}{P(B)} = \sum_n P^B(A_n) \end{aligned}$$

□

Conséquence importante : $P^B(\overline{A}) = 1 - P^B(A)$.

Probabilités conditionnelles et indépendance :

Théorème 4.1 : Si A et B sont deux événements indépendants de probabilité non nulle, alors

$$P^B(A) = P(A) \text{ et } P^A(B) = P(B).$$

Démonstration. Si A et B sont indépendants, alors $P(A \cap B) = P(A)P(B)$ donc $\dfrac{P(A \cap B)}{P(B)} = P(A)$ c'est-à-dire $P^B(A) = P(A)$ et de même $P^A(B) = \dfrac{P(A \cap B)}{P(A)} = P(B)$. □

Exemples :

---

a) Quelle est la probabilité que les 2 enfants d'une même famille soient des garçons sachant que l'aîné est un garçon ?

On note A l'événement “2 garçons” et B l'événement “l'aîné est un garçon”. On a $P(A) = \dfrac{1}{4}$ et $P(B) = \dfrac{2}{4} = \dfrac{1}{2}$ et on cherche $P^B(A)$.

Ici $A \cap B = A$ car $A \subset B$, donc $P^B(A) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P(A)}{P(B)} = \dfrac{\frac{1}{4}}{\frac{1}{2}} = \dfrac{1}{2}$.

---

b) Quelle est la probabilité que les 2 enfants d'une même famille soient des garçons sachant que l'un d'eux au moins est un garçon ?

On note C l'événement “au moins un garçon”. On a $P(C) = \dfrac{3}{4}$. De plus, $A \subset C$ donc $A \cap C = A$ et $P^C(A) = \dfrac{P(A \cap C)}{P(C)} = \dfrac{P(A)}{P(C)} = \dfrac{\frac{1}{4}}{\frac{3}{4}} = \dfrac{1}{3}$.

---

c) On a 3 pièces : l'une a les 2 faces rouges, l'autre les 2 faces blanches et la dernière 1 face rouge et 1 face blanche. On tire au hasard l'une des trois pièces et on constate que la face visible est rouge. Quelle est la probabilité pour que l'autre face soit rouge ?

On note A l'événement “les 2 faces sont rouges” : $P(A) = \dfrac{1}{3}$; et B l'événement “la face visible est rouge $P(B) = \dfrac{3}{6} = \dfrac{1}{2}$ (en effet, il y a 6 faces visibles possibles, dont 3 rouges).

On a encore ici $A \cap B = A$ car $A \subset B$, donc $P^B(A) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P(A)}{P(B)} = \dfrac{\frac{1}{3}}{\frac{1}{2}} = \dfrac{2}{3}$.

Attention ! Il ne faut pas confondre ici B “la face visible est rouge”, de probabilité $\dfrac{1}{2}$, avec C “au moins une face est rouge” de probabilité $\dfrac{2}{3}$.

4.2 Probabilités composées

On déduit facilement de la définition de la probabilité conditionnelle que, si A et B sont deux événements de probabilité non nulle, $P(A \cap B) = P^B(A)P(B) = P^A(B)P(A)$.

Plus généralement,

Théorème 4.2 : Soit $(A_i)_{1 \leqslant i \leqslant n}$ une famille d'événements telle que $P(A_1 \cap \cdots \cap A_{n-1}) \neq 0$; alors :

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)P(A_2/A_1)\cdots P(A_n/A_1 \cap \cdots \cap A_{n-1}).$$

Démonstration. On remarque d'abord que toutes ces probabilités conditionnelles sont définies car $A_1 \cap \cdots \cap A_{n-1} \subset A_1 \cap \cdots \cap A_k$ pour $k \leqslant n-1$ donc $0 < P(A_1 \cap \cdots \cap A_{n-1}) \leqslant P(A_1 \cap \cdots \cap A_k)$.

La démonstration se fait par récurrence sur n.

• La relation est vraie pour $n = 1$ et $n = 2$.

• Supposons la relation vraie jusqu'au rang n ; alors, en l'appliquant successivement au rang 2 et au rang n, on obtient

$$\begin{aligned} P(A_1 \cap \cdots \cap A_n \cap A_{n+1}) &= P(A_{n+1}/A_1 \cap \cdots \cap A_n)P(A_1 \cap \cdots \cap A_n) \\ &= P(A_1)\cdots P(A_n/A_1 \cap \cdots \cap A_{n-1})P(A_{n+1}/A_1 \cap \cdots \cap A_n) \end{aligned}$$

□

Exemples :

---

a) Soit 2 urnes U_1 et U_2 contenant initialement chacune 2 boules noires et 3 blanches. On tire dans U_1 une boule, on note sa couleur, on la remet dans U_2. On tire ensuite une boule dans U_2. Quelle est la probabilité d'obtenir 2 fois une boule noire ?

On note N_1 l'événement “la première boule tirée est noire” et N_2 l'événement “la deuxième boule tirée est noire” et on cherche $P(N_1 \cap N_2)$.

On a, d'après le théorème 4.2, $P(N_1 \cap N_2) = P(N_1) \times P^{N_1}(N_2)$. Or $P(N_1) = \frac{2}{5}$ et $P^{N_1}(N_2) = \frac{3}{6} = \frac{1}{2}$ (en effet, on sait que l'on a mis une boule noire en plus dans U_2 : il y a donc 6 boules dont 3 noires dans U_2 pour le deuxième tirage).

Finalement, $P(N_1 \cap N_2) = \frac{2}{5} \times \frac{3}{6} = \frac{1}{5}$.

---

b) Une urne contient 4 boules blanches et 3 noires. On tire une à une 3 boules sans remise. Quelle est la probabilité que la première boule soit blanche, la deuxième blanche et la troisième noire ?

On cherche $P(B_1 \cap B_2 \cap N_3)$ où B_i est l'événement “la i-ème boule tirée est blanche” et N_i l'événement “la i-ème boule tirée est noire”.

D'après le théorème 4.2, on a

$$P(B_1 \cap B_2 \cap N_3) = P(B_1)P(B_2/B_1)P(N_3/B_1 \cap B_2).$$

Les événements B_1, B_2 et N_3 ne sont pas indépendants car la composition de l'urne varie suivant les boules tirées.

On a $P(B_1) = \frac{4}{7}$; $P(B_2/B_1) = \frac{3}{6}$ car il reste 6 boules dont 3 blanches ;

de même, $P(N_3/B_1 \cap B_2) = \dfrac{3}{5}$ car il reste 5 boules dont 3 noires.

On a finalement $P(B_1 \cap B_2 \cap N_3) = \dfrac{4}{7} \times \dfrac{3}{6} \times \dfrac{3}{5} = \dfrac{6}{35}$.

---

c) Un conducteur sobre a 1 chance sur 1000 d'avoir un accident de voiture tout seul au cours d'une période donnée alors qu'un conducteur ivre a 1 chance sur 50. On admet qu'1 conducteur sur 100 est ivre. Quelle est la probabilité qu'il y ait un accident individuel et que le conducteur soit ivre ?

On note A l'événement "il y a un accident" et I l'événement "le conducteur est ivre". On cherche $P(A \cap I)$.

On a $P(A \cap I) = P(I)P^I(A)$ avec $P(I) = \dfrac{1}{100}$ et $P^I(A) = \dfrac{1}{50}$ donc $P(A \cap I) = \dfrac{1}{5000}$.

4.3 Formule des probabilités totales

Théorème 4.3 : Soit $(B_i)_{i \in I}$ un système complet d'événements de probabilités non nulles. Alors, pour tout $A \in \mathcal{A}$,

$$P(A) = \sum_{i \in I} P^{B_i}(A)P(B_i).$$

Démonstration. Les B_i sont 2 à 2 disjoints et $\Omega = \bigcup_{i \in I} B_i$ donc, par distributivité de l'intersection par rapport à l'union dénombrable :

$$A = A \cap \Omega = A \cap \left(\bigcup_{i \in I} B_i \right) = \bigcup_{i \in I} (A \cap B_i)$$

Or, comme les $A \cap B_i$ sont aussi 2 à 2 disjoints, on a $P(A) = \sum_{i \in I} P(A \cap B_i)$.

On applique alors la formule des probabilités composées : $P(A \cap B_i) = P^{B_i}(A)P(B_i)$. □

Cas particulier très utile : $P(A) = P^B(A)P(B) + P^{\overline{B}}(A)P(\overline{B})$.

En effet, $(B, \overline{B})$ forme un système complet d'événements de Ω.

Exemples :

---

a) Un individu est choisi au hasard dans une population possédant la proportion p de tricheurs. On fait tirer 1 carte d'un jeu de 52 à cet individu et on admet que, si c'est un tricheur, il est sûr de retourner un as. Quelle est la probabilité qu'il retourne un as ?

On note A l'événement "l'individu retourne un as" et T l'événement "l'individu est un tricheur" et on cherche $P(A)$.

On sait que $P(T) = p$; $P^T(A) = 1$ et $P^{\overline{T}}(A) = \dfrac{4}{52} = \dfrac{1}{13}$.

D'après le théorème 4.3, on a

$$\begin{aligned} P(A) &= P^T(A)P(T) + P^{\overline{T}}(A)P(\overline{T}) \\ &= p \times 1 + (1-p) \times \frac{1}{13} = \frac{1}{13} + p \times \left(1 - \frac{1}{13}\right) = \frac{1+12p}{13} \end{aligned}$$

---

b) Une urne contient n_1 boules blanches et n_2 boules noires. On tire une boule : si elle est blanche, on la remet ; si elle est noire, on la remplace par a boules blanches prises ailleurs. On tire une deuxième boule. Quelle est la probabilité qu'elle soit blanche ?

On note B_i l'événement “la i-ème boule tirée est blanche” et on cherche $P(B_2)$.
D'après le théorème 4.3, on a $P(B_2) = P^{B_1}(B_2)P(B_1) + P^{\overline{B_1}}(B_2)P(\overline{B_1})$.
Or $P(B_1) = \dfrac{n_1}{n_1+n_2}$; $P^{B_1}(B_2) = \dfrac{n_1}{n_1+n_2}$ et $P^{\overline{B_1}}(B_2) = \dfrac{n_1+a}{n_1+n_2+a-1}$.
Donc, finalement, $P(B_2) = \left(\dfrac{n_1}{n_1+n_2}\right)^2 + \dfrac{n_2}{n_1+n_2} \times \dfrac{n_1+a}{n_1+n_2+a-1}$.

4.4 Théorème de Bayes

Théorème 4.4 : Soit $(B_i)_{i\in I}$ un système complet d'événements de probabilités non nulles et A un événement de probabilité non nulle. Alors, pour tout $i_0 \in I$:

$$P^A(B_{i_0}) = \frac{P^{B_{i_0}}(A)P(B_{i_0})}{\sum\limits_{i\in I} P^{B_i}(A)P(B_i)}.$$

Démonstration. On a $P^A(B_{i_0}) = \dfrac{P(A\cap B_{i_0})}{P(A)}$. Or, d'après la formule des probabilités composées (théorème 4.2), $P(A\cap B_{i_0}) = P^{B_{i_0}}(A)P(B_{i_0})$ et d'après la formule des probabilités totales (théorème 4.3), $P(A) = \sum\limits_{i\in I} P^{B_i}(A)P(B_i)$. □

Cas particulier très important : $P^A(B) = \dfrac{P^B(A)P(B)}{P^B(A)P(B) + P^{\overline{B}}(A)P(\overline{B})}$.

Exemples :

---

a) On reprend les conditions de l'exemple 4.2.c). Quelle est la probabilité qu'un conducteur soit ivre sachant qu'il a eu un accident ?

On cherche ici $P^A(I)$. D'après le théorème de Bayes (théorème 4.4), on a

$$\begin{aligned} P^A(I) &= \frac{P^I(A)P(I)}{P^I(A)P(I) + P^{\overline{I}}(A)P(\overline{I})} \\ &= \frac{\frac{1}{50}\times\frac{1}{100}}{\frac{1}{50}\times\frac{1}{100} + \frac{1}{1000}\times\frac{99}{100}} = \frac{\frac{1}{5}}{\frac{1}{5}+\frac{99}{100}} = \frac{20}{119}. \end{aligned}$$

---

b) Un gardien doit ouvrir une porte et il a 10 clefs différentes. S'il est à jeun, après avoir essayé une mauvaise clef, il la retire du lot. S'il est ivre, il remélange les clefs.

Sachant qu'il est ivre 1 jour sur 3 et qu'il vient d'essayer au moins 9 clefs pour ouvrir la porte, quelle est la probabilité qu'il soit ivre ?

On note C l'événement "le gardien a essayé au moins 9 clefs" et I l'événement "le gardien est ivre". On cherche $P^C(I)$.

$$P^C(I) = \frac{P^I(C)P(I)}{P^I(C)P(I) + P^{\overline{I}}(C)P(\overline{I})}.$$

On sait que $P(I) = \dfrac{1}{3}$. Il reste donc à trouver $P^I(C)$ et $P^{\overline{I}}(C)$. Pour cela, on note X le nombre de clefs nécessaire pour ouvrir la porte. On a donc $C = [X \geqslant 9]$. Puis :

$$\begin{aligned} P^I(C) &= P^I([X \geqslant 9]) = 1 - P^I([X \leqslant 8]) = 1 - \sum_{k=1}^{8} P^I([X = k]) \\ &= 1 - \sum_{k=1}^{8} \left(\frac{9}{10}\right)^{k-1} \times \frac{1}{10} = 1 - \frac{1}{10} \times \frac{1 - \left(\frac{9}{10}\right)^8}{1 - \frac{9}{10}} = \left(\frac{9}{10}\right)^8 \end{aligned}$$

et $P^{\overline{I}}(C) = P^{\overline{I}}([X \geqslant 9]) = P^{\overline{I}}([X = 9]) + P^{\overline{I}}([X = 10]) = \dfrac{2}{10} = \dfrac{1}{5}$.

Donc finalement, $P^C(I) = \dfrac{\frac{1}{3} \times \left(\frac{9}{10}\right)^8}{\frac{1}{3} \times \left(\frac{9}{10}\right)^8 + \frac{2}{3}\frac{1}{5}} \approx 0,518$.

Chapitre 5

OPÉRATIONS SUR LES V.A.R. ET CALCULS DE LOIS

5.1 Variables aléatoires réelles discrètes

Soit X une v.a.r. discrète définie sur un espace probabilisé $(\Omega, \mathcal{A}, P)$.
On pose $X(\Omega) = \{x_i \ ; \ i \in \mathbb{N}\}$ (le cas fini est similaire au cas infini dénombrable et présente moins de difficultés ; on traitera donc uniquement le cas dénombrable).

5.1.1 Espérance

Définition 5.1 : On dit que X possède une *espérance* si la série $\sum_{n \geqslant 0} |x_n| P([X = x_n])$ converge ; on appelle alors espérance de X et on note $\mathbb{E}(X)$ le nombre défini par :

$$\mathbb{E}(X) = \sum_{n \geqslant 0} x_n P([X = x_n]).$$

Remarques :
1) Si $X(\Omega)$ est fini, X possède toujours une espérance (la somme d'un nombre fini de termes est toujours finie).

2) Une v.a.r. discrète peut ne pas avoir d'espérance :
Exemple : $\sum_{n \geqslant 1} \frac{1}{n^2}$ est une série convergente (par Riemann), de somme S (en fait, $S = \frac{\pi^2}{6}$). On pose, pour $n \in \mathbb{N}^*$, $p_n = \frac{6}{\pi^2 n^2}$: la famille (p_n) définit bien une probabilité. Une v.a.r. X de loi (p_n) n'admet pas d'espérance car $nP([X = n]) = np_n = \frac{6}{\pi^2 n}$ est le terme général d'une série divergente.

3) $\mathbb{E}(X) = \frac{\sum_n x_n P([X = x_n])}{\sum_n P([X = x_n])}$: $\mathbb{E}(X)$ apparaît ainsi comme le barycentre des x_n affectés des masses $P([X = x_n])$; c'est donc la moyenne des valeurs prises par X pondérées par leur probabilité.

4) La plupart du temps, on aura $X(\Omega) \subset \mathbb{R}_+$. La convergence absolue de la série $\sum_{n\geqslant 0} x_n P([X = x_n])$ équivaut alors à sa convergence et on peut chercher directement l'espérance.

5.1.2 Exemples d'espérances pour des lois discrètes classiques

On reprend ici les lois évoquées au Chapitre 3.

a) Loi de Bernoulli $\mathcal{B}(p)$ (ou $\mathcal{B}(1,p)$) :
$X(\Omega) = \{0, 1\}$; $P([X = 1]) = p$; $P([X = 0]) = 1 - p$;

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

En effet, $\mathbb{E}(X) = 0 \times P([X = 0]) + 1 \times P([X = 1]) = p$.

b) Loi binomiale $\mathcal{B}(n,p)$:
$X(\Omega) = \{0, \cdots, n\}$ et, pour $k \in X(\Omega)$, $P([X = k]) = C_n^k p^k (1-p)^{n-k}$;

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

voir le paragraphe 5.4 pour la démonstration.

c) Loi hypergéométrique $\mathcal{H}(n, M, N)$:
$X(\Omega) = \{\max(0, n - N + M), \cdots, \min(n, M)\}$; pour $k \in X(\Omega)$, $P([X = k]) = \dfrac{C_M^k C_{N-M}^{n-k}}{C_N^n}$.

$$\boxed{\mathbb{E}(X) = \frac{Mn}{N}.}$$

En effet, $\mathbb{E}(X) = \displaystyle\sum_{k\in X(\Omega)} k \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$ avec, pour $k \geqslant 1$, $kC_M^k = \dfrac{M!}{(k-1)!(M-k)!} = MC_{M-1}^{k-1}$,
donc $\mathbb{E}(X) = \dfrac{M}{C_N^n} \displaystyle\sum_{k\in X(\Omega)\backslash\{0\}} C_{M-1}^{k-1} C_{N-1-(M-1)}^{n-1-(k-1)} = \dfrac{M}{C_N^n} C_{N-1}^{n-1} = \dfrac{Mn!(N-n)!(N-1)!}{N!(n-1)!(N-n)!} = \dfrac{Mn}{N}$.

(pour trouver $\displaystyle\sum_{k\in X(\Omega)\backslash\{0\}} C_{M-1}^{k-1} C_{N-1-(M-1)}^{n-1-(k-1)} = C_{N-1}^{n-1}$, on s'est servi de $\displaystyle\sum_{k\in X(\Omega)} \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} = 1$,
soit $\displaystyle\sum_{k\in X(\Omega)} C_M^k C_{N-M}^{n-k} = C_N^n$ appliqué à $k' = k - 1$, $N' = N - 1$, $M' = M - 1$ et $n' = n - 1$.)

d) Loi géométrique sur $\mathbb{N}^*$ $\mathcal{G}(p)$ (ou $\mathcal{P}(1,p)$) :
$X(\Omega) = \mathbb{N}^*$ et, pour tout $k \in X(\Omega)$, $P([X = k]) = p(1-p)^{k-1}$.

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

En effet, en posant $q = 1 - p$,

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k\geqslant 1} kP([X=k]) = p\sum_{k\geqslant 1} kq^{k-1}\\ &= p\sum_{k\geqslant 1}\frac{d}{dq}(q^k) = p\frac{d}{dq}\left(\sum_{k\geqslant 0} q^k\right) = p\frac{d}{dq}\left(\frac{1}{1-q}\right) = \frac{p}{(1-q)^2} = \frac{1}{p}.\end{aligned}$$

e) Loi de Poisson $\mathcal{P}(\lambda)$:

$X(\Omega) = \mathbb{N}$, et, pour tout $k \in X(\Omega)$, $P([X=k]) = e^{-\lambda}\dfrac{\lambda^k}{k!}$.

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

En effet,

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k\geqslant 0} kP([X=k]) = \sum_{k\geqslant 1} ke^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k\geqslant 1}\frac{\lambda^k}{(k-1)!}\\ &= e^{-\lambda}\sum_{j\geqslant 0}\frac{\lambda^{j+1}}{j!} = \lambda e^{-\lambda}\sum_{j\geqslant 0}\frac{\lambda^j}{j!} = \lambda e^{-\lambda}e^{\lambda} = \lambda.\end{aligned}$$

5.1.3 Moments d'ordre r

Définition 5.2 : i) On appelle <u>*moment d'ordre r*</u> de X le nombre $m_r(X)$ défini par $m_r(X) = \sum_{n\geqslant 0} x_n^r P([X=x_n])$ pourvu que cette série converge absolument.
ii) On appelle <u>*moment centré d'ordre r*</u> de X le nombre

$$\mu_r(X) = \sum_{n\geqslant 0}(x_n - \mathbb{E}(X))^r P([X=x_n]),$$

pourvu que cette série converge absolument.
iii) On appelle <u>*variance*</u> de X et on note $\mathrm{var}(X)$ le moment centré d'ordre 2 : $\mu_2(X)$.
iv) Si X admet une variance, on appelle <u>*écart-type*</u> de X le nombre $\sigma(X)$ défini par $\sigma(X) = \sqrt{\mathrm{var}(X)}$.

Remarque : $\mathbb{E}(X) = m_1(X)$ et $\mu_1(X) = 0$.

Attention : Souvent, c'est le moment d'ordre r qui est noté $\mu_r(X)$ au lieu de $m_r(X)$ et dans ces cas, il n'y a pas de notations spécifiques pour le moment centré.

Propriété 5.1 : Si X est d'ordre r, les moments et les moments centrés d'ordre $k \leqslant r$ existent.

Démonstration. Si $|x_n| \leqslant 1$, alors $|x_n|^k \leqslant 1$ et si $|x_n| > 1$, alors $|x_n|^k \leqslant |x_n|^r$, donc, dans tous les cas, $|x_n|^k \leqslant 1 + |x_n|^r$. La série $\left(\sum P([X=x_n])\right)$ converge et, si X admet un moment d'ordre r, il en est de même pour la série $\left(\sum |x_n|^r P([X=x_n])\right)$. On a donc aussi, puisque $|x_n|^k P([X=x_n]) \leqslant P([X=x_n]) + |x_n|^r P([X=x_n])$, convergence de la série $\left(\sum |x_n|^k P([X=x_n])\right)$. Par conséquent, si $m_r(X)$

existe, alors $m_k(X)$ existe. D'autre part,

$$\begin{aligned}
\sum_n |x_n - \mathbb{E}(X)|^k P([X = x_n]) &\leqslant \sum_n (|x_n| + |\mathbb{E}(X)|)^k P([X = x_n]) \\
&= \sum_n \left(\sum_{i=0}^k C_k^i |x_n|^i |\mathbb{E}(X)|^{k-i} P([X = x_n]) \right) \\
&= \sum_{i=0}^k C_k^i |\mathbb{E}(X)|^{k-i} \left(\sum_n |x_n|^i P([X = x_n]) \right) < +\infty
\end{aligned}$$

car on a une somme finie de terme finis (les $k+1$ séries qui interviennent convergent). □

Exemple : Si X suit la loi de Bernoulli $\mathcal{B}(p)$,

$$m_k(X) = 0^k \times P([X = 0]) + 1^k \times P([X = 1]) = p.$$

5.1.4 Calcul de $\mathbb{E}(\varphi(X))$

Soit $\varphi : \mathbb{R} \to \mathbb{R}$ une fonction numérique quelconque. L'application $Y = \varphi(X) : \Omega \to \mathbb{R}$ est encore une v.a.r. discrète. En effet, $\varphi(X)(\Omega) = \varphi(X(\Omega))$ est dénombrable car c'est l'image par φ d'un ensemble dénombrable et, pour tout $y_n \in \varphi(X(\Omega))$, si $I_n = \{i \ ; \ \varphi(x_i) = y_n \text{ où } x_i \in X(\Omega)\}$, alors $[\varphi(X) = y_n] = \bigcup_{i \in I_n} [X = x_i]$ est dans la tribu $\mathcal{A}$, comme réunion dénombrable d'éléments de $\mathcal{A}$.

Théorème 5.1 : Soit φ une fonction de $\mathbb{R}$ dans $\mathbb{R}$ telle que $\varphi(X)$ admette une espérance. Alors :

$$\mathbb{E}(\varphi(X)) = \sum_{n \geqslant 0} \varphi(x_n) P([X = x_n]).$$

Remarque : Ce théorème est l'un des plus importants : il permet en effet de calculer $\mathbb{E}(\varphi(X))$ sans connaître la loi de $\varphi(X)$ mais en connaissant simplement la loi de X.

Démonstration. Par définition de $\mathbb{E}(\varphi(X))$, on a :

$$\begin{aligned}
\mathbb{E}(\varphi(X)) &= \sum_n y_n P([\varphi(X) = y_n]) = \sum_n y_n P\left(\bigcup_{i \in I_n} [X = x_i] \right) \\
&= \sum_n y_n \sum_{i \in I_n} P([X = x_i]) = \sum_n \sum_{i \in I_n} \varphi(x_i) P([X = x_i]) \\
&= \sum_i \varphi(x_i) P([X = x_i])
\end{aligned}$$

car $(I_n)_n$ est une partition de $\mathbb{N}$ et $\{x_i \ ; \ i \in \mathbb{N}\} = \bigcup_n \{x_i \ ; \ i \in I_n\}$. □

Conséquences :

1) $\boxed{\mathbb{E}(aX + b) = a\mathbb{E}(X) + b}$ si $\mathbb{E}(X)$ existe. En effet, si $X(\Omega) = \{x_n\}$,

$$\begin{aligned}
\mathbb{E}(aX+b) &= \sum_n (ax_n+b)P([X=x_n]) = a\sum_n x_n P([X=x_n]) + b\sum_n P([X=x_n]) \\
&= a\mathbb{E}(X)+b.
\end{aligned}$$

2) $\boxed{\mathrm{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2}$. En effet, en posant $m = \mathbb{E}(X)$,

$$\begin{aligned}
\mathrm{var}(X) &= \mathbb{E}((X-m)^2) = \sum_n (x_n-m)^2 P([X=x_n]) \\
&= \mathbb{E}(X^2 - 2mX + m^2) = \sum_n (x_n^2 - 2mx_n + m^2)P([X=x_n]) \\
&= \sum_n x_n^2 P([X=x_n]) - 2m\sum_n x_n P([X=x_n]) + m^2 \sum_n P([X=x_n]) \\
&= \mathbb{E}(X^2) - 2m\mathbb{E}(X) + m^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2
\end{aligned}$$

3) $\boxed{\mathrm{var}(aX+b) = a^2\mathrm{var}(X)}$. En effet,

$$\begin{aligned}
\mathrm{var}(aX+b) &= \mathbb{E}\left(((aX+b) - \mathbb{E}(aX+b))^2\right) \\
&= \mathbb{E}\left((aX+b-(a\mathbb{E}(X)+b))^2\right) = \mathbb{E}\left((aX+b-a\mathbb{E}(X)-b)^2\right) \\
&= \mathbb{E}\left(a^2(X-\mathbb{E}(X))^2\right) = a^2\mathbb{E}\left((X-\mathbb{E}(X))^2\right) = a^2\mathrm{var}(X)
\end{aligned}$$

5.1.5 Fonctions génératrices

La fonction génératrice est un outil de calcul permettant de calculer les moments d'une v.a.r. discrète à valeurs dans $\mathbb{N}$, et plus particulièrement l'espérance et la variance.

On pose $G_X(t) = \sum_{n\geqslant 0} t^n P([X=n])$. En appliquant le théorème 5.1 à $\varphi_t : x \mapsto t^x$, on a donc $G_X(t) = \mathbb{E}(t^X)$.

Pour $t=1$, $G_X(1)=1$, et, pour $|t| \leqslant 1$, $\sum_n |t|^n P([X=n]) \leqslant \sum_n P([X=n]) = 1$. La série $\sum_n t^n P([X=n])$ converge donc absolument pour $|t| \leqslant 1$ et $G_X(t)$ existe pour $t \in [-1,1]$.

Théorème 5.2 : Soit $G_X(t) = \mathbb{E}(t^X) = \sum_{n\geqslant 0} t^n P([X=n])$. Alors :

$$\mathbb{E}(X) = \lim_{t\to 1_-} G'_X(t) \text{ et } \mathrm{var}(X) = \lim_{t\to 1_-} \left(G''_X(t) + G'_X(t) - (G'_X(t))^2\right).$$

Démonstration. Si l'étude des séries entières n'a pas encore été faite en analyse, on pourra déjà admettre qu'à l'intérieur du domaine de convergence, on peut dériver terme à terme et qu'on a, pour $|t|<1$:

$$G'_X(t) = \sum_{n\geqslant 1} n t^{n-1} P([X=n])$$

$$G''_X(t) = \sum_{n\geqslant 2} n(n-1) t^{n-2} P([X=n])$$

$$\vdots$$

$$G_X^{(k)}(t) = \sum_{n \geqslant k} n(n-1)\cdots(n-k+1)t^{n-k}P([X=n])$$

$$\vdots$$

On peut aussi admettre que, sous réserve de convergence absolue, on peut passer à la limite dans la somme lorsque t tend vers 1 par valeurs inférieures (c'est-à-dire intervertir lim et $\sum$). On a alors, en appliquant le théorème 5.1 aux fonctions $\varphi_k : x \mapsto x(x-1)\cdots(x-k+1)$:

$$\lim_{t\to 1} G'_X(t) = \sum_{n\geqslant 1} nP([X=n]) = \mathbb{E}(X)$$

$$\lim_{t\to 1} G''_X(t) = \sum_{n\geqslant 2} n(n-1)P([X=n]) = \mathbb{E}(X(X-1))$$

$$\vdots$$

$$\lim_{t\to 1} G_X^{(k)}(t) = \sum_{n\geqslant k} n(n-1)\cdots(n-k+1)P([X=n]) = \mathbb{E}(X(X-1)\cdots(X-k+1))$$

$$\vdots$$

En particulier,

$$\begin{aligned} \mathrm{var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}(X(X-1)) + \mathbb{E}(X) - \mathbb{E}(X)^2 \\ &= G''_X(1_-) + G'_X(1_-) - G'_X(1_-)^2. \end{aligned}$$

□

Application aux lois classiques :

a) Loi binomiale $\mathcal{B}(n,p)$:
$X(\Omega) = \{0, \cdots, n\}$ et, pour $k \in X(\Omega)$, $P([X=k]) = C_n^k p^k (1-p)^{n-k}$.

$$\boxed{G_X(t) = (pt+1-p)^n\,;\ \mathbb{E}(X) = np\,;\ \mathrm{var}(X) = np(1-p).}$$

En effet,

$$G_X(t) = \sum_{k=0}^{n} t^k P([X=k]) = \sum_{k=0}^{n} t^k C_n^k p^k (1-p)^{n-k} = (pt+1-p)^n$$

$$G'_X(t) = np(pt+1-p)^{n-1} \quad ; \quad G''_X(t) = n(n-1)p^2(pt+1-p)^{n-2}.$$

$$G'_X(1_-) = np \quad ; \quad G''_X(1_-) = n(n-1)p^2$$

et donc $\mathrm{var}(X) = G''_X(1) + G'_X(1_-) - G'_X(1_-)^2 = n(n-1)p^2 + np - n^2p^2 = np - np^2 = np(1-p)$.

b) Loi géométrique sur $\mathbb{N}^*$ $\mathcal{G}(p)$:
$X(\Omega) = \mathbb{N}^*$ et, pour tout $k \in X(\Omega)$, $P([X=k]) = p(1-p)^{k-1}$.

$$\boxed{G_X(t) = \frac{p}{1-p}\left(\frac{1}{1-t(1-p)} - 1\right);\ \mathbb{E}(X) = \frac{1}{p}\,;\ \mathrm{var}(X) = \frac{1-p}{p^2}.}$$

En effet, en posant $q = 1 - p$, on a :

$$\begin{aligned} G_X(t) &= \sum_{k\geqslant 1} t^k P([X=k]) = \sum_{k\geqslant 1} t^k p q^{k-1} \\ &= \frac{p}{q}\left(\sum_{k\geqslant 1}(tq)^k\right) = \frac{p}{q}\left(\frac{1}{1-tq} - 1\right) \end{aligned}$$

$$G'_X(t) = \frac{p}{q}\frac{q}{(1-tq)^2} = \frac{p}{(1-tq)^2} \text{ et } G''_X(t) = \frac{2pq}{(1-tq)^3}.$$

On a donc $G'_X(1_-) = \dfrac{p}{(1-q)^2} = \dfrac{1}{p}$ et $G''_X(1_-) = \dfrac{2pq}{(1-q)^3} = \dfrac{2q}{p^2}$ puis finalement

$$G''_X(1_-) + G'_X(1_-) - G'_X(1_-)^2 = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2q+p-1}{p^2} = \frac{q}{p^2}.$$

c) Loi binomiale négative $\mathcal{BN}(r,p)$:

$Z_r(\Omega) = \mathbb{N}$ et, pour tout $k \in Z_r(\Omega)$, $P([Z_r = k]) = C_{k+r-1}^{r-1} p^r (1-p)^k$.

$$\boxed{G_X(t) = \left(\frac{p}{1-t(1-p)}\right)^r \text{ ; } \mathbb{E}(X) = \frac{r(1-p)}{p} \text{ ; } \mathrm{var}(X) = \frac{r(1-p)}{p^2}.}$$

En effet, en posant $q = 1 - p$,

$$G_X(t) = \sum_{k\geqslant 0} t^k P([Z_r = k]) = p^r \sum_{k\geqslant 0} q^k \frac{(k+r-1)\cdots(k+1)}{(r-1)!}.$$

On sait que $\dfrac{1}{1-x} = \displaystyle\sum_{k\geqslant 0} x^k$. En dérivant ce développement, on obtient successivement :

$$\frac{d}{dx}\left(\frac{1}{1-x}\right) = \frac{1}{(1-x)^2} = \sum_{k\geqslant 1} k x^{k-1} = \sum_{k\geqslant 0}(k+1)x^k,$$

$$\frac{d^2}{dx^2}\left(\frac{1}{1-x}\right) = \frac{2}{(1-x)^3} = \sum_{k\geqslant 2} k(k-1)x^{k-2} = \sum_{k\geqslant 0}(k+2)(k+1)x^k,$$

$$\vdots$$

$$\frac{d^{r-1}}{dx^{r-1}}\left(\frac{1}{1-x}\right) = \frac{(r-1)!}{(1-x)^r} = \sum_{k\geqslant 0}(k+r-1)\cdots(k+1)x^k$$

donc $G_{Z_r}(t) = \dfrac{p^r}{(1-qt)^r}$, $G'_{Z_r}(t) = \dfrac{rp^r q}{(1-qt)^{r+1}}$ et $G''_{Z_r}(t) = \dfrac{r(r+1)p^r q^2}{(1-qt)^{r+2}}$, puis $G'_{Z_r}(1_-) = \dfrac{rq}{p}$ $G''_{Z_r}(1_-) = \dfrac{r(r+1)q^2}{p^2}$ et finalement

$$G''_{Z_r}(1_-) + G'_{Z_r}(1_-) - G'_{Z_r}(1_-)^2 = \frac{r(r+1)q^2}{p^2} + \frac{rq}{p} - \frac{r^2q^2}{p^2} = \frac{rq}{p^2}((r+1)q + p - rq) = \frac{rq}{p^2}.$$

d) Loi de Pascal $\mathcal{P}(r,p)$: ($X = Y + r$ avec Y de loi $\mathcal{BN}(r,p)$)

$Y_r(\Omega) = \{r, r+1, \cdots\}$ et pour tout $k \in Y_r(\Omega)$, $P([Y_r = k]) = C_{k-1}^{r-1} p^r (1-p)^{k-r}$.

$$\boxed{\mathbb{E}(X) = \frac{r}{p}\,;\ \mathrm{var}(X) = \frac{r(1-p)}{p^2}}$$

En effet, on a vu que, si Y_r suit la loi de Pascal, $\mathcal{P}(r,p)$, alors $Y_r = Z_r + r$, où Z_r suit la loi binomiale négative $\mathcal{BN}(r,p)$.

On a alors $\mathbb{E}(Y_r) = \mathbb{E}(Z_r + r) = \mathbb{E}(Z_r) + r = \dfrac{rq}{p} + r = \dfrac{r}{p}$ et

$$\mathrm{var}(Y_r) = \mathrm{var}(Z_r + r) = \mathrm{var}(Z_r) = \frac{rq}{p^2}.$$

e) Loi de Poisson $\mathcal{P}(\lambda)$:

$X(\Omega) = \mathbb{N}$, et, pour tout $k \in X(\Omega)$, $P([X = k]) = e^{-\lambda}\dfrac{\lambda^k}{k!}$.

$$\boxed{G_X(t) = e^{-\lambda}e^{\lambda t}\,;\ \mathbb{E}(X) = \lambda\,;\ \mathrm{var}(X) = \lambda.}$$

En effet, $G_X(t) = \displaystyle\sum_{k\geqslant 0} t^k P([X=k]) = e^{-\lambda}\sum_{k\geqslant 0} t^k \frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda t}$.

$G'_X(t) = e^{-\lambda}\lambda e^{\lambda t}$, $G''_X(t) = e^{-\lambda}\lambda^2 e^{\lambda t}$, donc $G'_X(1_-) = \lambda$, $G''_X(1_-) = \lambda^2$ et

$$G''_X(1_-) + G'_X(1) - G'_X(1)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

5.2 Variables aléatoires réelles absolument continues

5.2.1 Moments d'une v.a.r. absolument continue

Définition 5.3 : Soit X une v.a.r. absolument continue, de densité f_X.
i) Soit $r \in \mathbb{N}^*$. On appelle *moment d'ordre r* de X et on note $m_r(X)$ le réel défini par :
$m_r(X) = \displaystyle\int_{-\infty}^{+\infty} t^r f_X(t)dt$; pourvu que cette intégrale converge absolument.
ii) En particulier, on appelle *espérance* de X, le moment d'ordre 1 de X, s'il existe.
iii) Si X admet une espérance, on appelle *moment centré d'ordre r* de X le moment d'ordre r, s'il existe, de $X - \mathbb{E}(X)$. On le note $\mu_r(X)$.
iv) En particulier, si $\mu_2(X)$ existe, $\mu_2(X)$ est appelé *variance* de X et noté $\mathrm{var}(X)$.
Si X admet une variance, on appelle *écart-type* de X le nombre $\sigma(X)$ défini par

$$\sigma(X) = \sqrt{\mathrm{var}(X)}.$$

Exemples :
1) Soit X une v.a.r. de densité f définie par $f(x) = \cos x\, \mathbb{1}_{[0,\frac{\pi}{2}]}(x)$.

$$\int xf(x)dx = \int_0^{\pi/2} x\cos x\,dx = [x\sin x]_0^{\pi/2} - \int_0^{\pi/2}\sin x\,dx = \frac{\pi}{2} + [\cos x]_0^{\pi/2} = \frac{\pi}{2} - 1.$$

(intégration par parties avec $u = x$ et $v' = \cos x$).

2) Soit X une v.a.r. de loi de Cauchy de densité f définie par $f(x) = \dfrac{1}{\pi(1+x^2)}$.

$$\int_a^A xf(x)dx = \frac{1}{2\pi}\int_a^A \frac{2x}{(1+x^2)}\,dx = \frac{1}{2\pi}\left[\ln(1+x^2)\right]_a^A = \frac{1}{2\pi}[\ln(1+A^2) - \ln(1+a^2)].$$

On a donc $\lim\limits_{A\to+\infty}\int_a^A xf(x)dx = +\infty$ et $\int_{-\infty}^{+\infty} xf(x)dx$ diverge.

Propriétés 5.2 :

1) Soit X une v.a.r. de densité f_X admettant une espérance et soit $(a,b) \in \mathbb{R}^* \times \mathbb{R}$. Alors $aX+b$ admet une espérance et $\mathbb{E}(aX+b) = a\mathbb{E}(X)+b$.

2) Si $\int_{-\infty}^{+\infty} x^2 f_X(x)dx$ converge, alors X^2 admet une espérance et

$$\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x)dx.$$

Démonstration. **1)** D'après l'exemple 1) qui précéde le théorème 3.3 , si X a pour densité f_X, alors $aX+b$ a pour densité f_{aX+b} définie par $f_{aX+b}(t) = \frac{1}{|a|}f_X\left(\frac{t-b}{a}\right)$.

Donc, sous réserve d'existence, $\mathbb{E}(aX+b) = \int_{-\infty}^{+\infty} tf_{aX+b}(t)dt = \frac{1}{|a|}\int_{-\infty}^{+\infty} tf_X\left(\frac{t-b}{a}\right)\,dt.$

On fait, dans la dernière intégrale, le changement de variable $u = \frac{t-b}{a}$.

- Si $a > 0$, $\mathbb{E}(aX+b) = \frac{1}{a}\int_{-\infty}^{+\infty}(au+b)f_X(u)\,adu = a\int_{-\infty}^{+\infty} uf_X(u)du + b\int_{-\infty}^{+\infty} f_X(u)du = a\mathbb{E}(X)+b\,;$
- Si $a < 0$, $\mathbb{E}(aX+b) = \frac{1}{-a}\int_{+\infty}^{-\infty}(au+b)f_X(u)\,adu = a\int_{-\infty}^{+\infty} uf_X(u)du + b\int_{-\infty}^{+\infty} f_X(u)du = a\mathbb{E}(X)+b.$

Comme $\mathbb{E}(X)$ existe, $\mathbb{E}(aX+b)$ existe aussi et $\mathbb{E}(aX+b) = a\mathbb{E}(X)+b$.

2) D'après l'exemple 2) qui précède le théorème 3.3, si X a pour densité f_X, alors X^2 a pour densité f_{X^2} définie par $f_{X^2}(t) = \frac{1}{2\sqrt{t}}\left(f_X(\sqrt{t}) + f_X(-\sqrt{t})\right)\,\mathbb{1}_{]0,+\infty[}(t)$.

Donc, sous réserve d'existence,

$$\begin{aligned}\mathbb{E}(X^2) &= \int_{-\infty}^{+\infty} tf_{X^2}(t)\,dt = \int_0^{+\infty}\frac{\sqrt{t}}{2}\left(f_X(\sqrt{t}) + f_X(-\sqrt{t})\right)\,dt \\ &= \int_0^{+\infty}\frac{\sqrt{t}}{2}f_X(\sqrt{t})dt + \int_0^{+\infty}\frac{\sqrt{t}}{2}f_X(-\sqrt{t})\,dt.\end{aligned}$$

Dans la première intégrale, on fait le changement de variable $u = \sqrt{t}$ et dans la deuxième intégrale, on fait le changement de variable $u = -\sqrt{t}$ ($t = u^2$, $dt = 2udu$).

$$\mathbb{E}(X^2) = \int_0^{+\infty} u^2 f_X(u)du + \int_0^{-\infty}(-u^2)f_X(u)du = \int_{-\infty}^{+\infty} u^2 f_X(u)du.$$

□

Plus généralement, on admet le théorème suivant :

Théorème 5.3 : Soit X une v.a.r. absolument continue de densité f_X et φ une fonction numérique continue et dérivable sur $X(\Omega)$, alors $\varphi(X)$ est une v.a.r. absolument continue et, si elle admet une espérance, alors

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

Conséquences :

1) $m_r(X) = \mathbb{E}(X^r)$ (appliquer le théorème 5.3. à $\varphi_r : t \mapsto t^r$)

2) $F_X(x) = \mathbb{E}(\mathbb{1}_{]\infty,x]}(X))$ (en effet, $\mathbb{E}(\mathbb{1}_{]-\infty,x]}(X)) = \int \mathbb{1}_{]-\infty,x]}(t) f_X(t)\, dt = \int_{-\infty}^{x} f_X(t)\, dt$).

3) On retrouve $\mathbb{E}(aX+b) = \int_{-\infty}^{+\infty} (at+b) f_X(t) dt = a\mathbb{E}(X) + b$.

4) $\mathrm{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.
En effet, en posant $\mathbb{E}(X) = m$, on a

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

5) $\mathrm{var}(aX+b) = a^2 \mathrm{var}(X)$.
En effet,

$$\begin{aligned} \mathrm{var}(aX+b) &= \mathbb{E}\left((aX+b-\mathbb{E}(aX+b))^2\right) = \mathbb{E}\left((aX+b-a\mathbb{E}(X)-b)^2\right) \\ &= \mathbb{E}\left(a^2(X-\mathbb{E}(X))^2\right) = a^2\mathbb{E}((X-\mathbb{E}(X))^2) = a^2\mathrm{var}(X). \end{aligned}$$

Application aux lois classiques :

a) Loi uniforme $\mathcal{U}([a,b])$:
Loi de densité f_X définie par $f_X(x) = \dfrac{1}{b-a}\,\mathbb{1}_{[a,b]}(x)$.

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} t f_X(t) dt = \int_a^b \frac{1}{b-a} t dt = \frac{1}{b-a}\left[\frac{t^2}{2}\right]_a^b = \frac{1}{2}\frac{b^2-a^2}{b-a} = \frac{b+a}{2}.$$

$$\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} t^2 f_X(t) dt = \int_a^b \frac{1}{b-a} t^2 dt = \frac{1}{b-a}\left[\frac{t^3}{3}\right]_a^b = \frac{1}{3}\frac{b^3-a^3}{b-a} = \frac{b^2+ab+a^2}{3}.$$

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

$$\boxed{\mathbb{E}(X) = \frac{a+b}{2} \text{ et } \mathrm{var}(X) = \frac{(a-b)^2}{12}.}$$

b) Loi exponentielle $\mathcal{E}(\lambda)$:
Loi de densité f_X définie par $f_X(x) = \lambda e^{-\lambda x}\,\mathbb{1}_{]0,+\infty[}(x)$.

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} t f_X(t) dt = \int_0^{+\infty} t\lambda e^{-\lambda t}\, dt = \frac{1}{\lambda}\int_0^{+\infty} u e^{-u} du = \frac{1}{\lambda}\Gamma(2) = \frac{1}{\lambda}\,;$$

$$\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} t^2 f_X(t) dt = \int_0^{+\infty} t^2\lambda e^{-\lambda t}\, dt = \frac{1}{\lambda^2}\int_0^{+\infty} u^2 e^{-u} du = \frac{1}{\lambda^2}\Gamma(3) = \frac{2}{\lambda^2}\,;$$

$$\text{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

$$\boxed{\mathbb{E}(X) = \frac{1}{\lambda} \text{ et } \text{var}(X) = \frac{1}{\lambda^2}.}$$

c) Loi Gamma $\gamma(\lambda, a)$:

Loi de densité f_X définie par $f_X(x) = \dfrac{\lambda^a}{\Gamma(a)} e^{-\lambda x} x^{a-1}\, \mathbb{1}_{]0,+\infty[}(x)$, où $\Gamma(a) = \displaystyle\int_0^{+\infty} e^{-t} t^{a-1}\, dt$.

X admet des moments de tous ordres car $\displaystyle\lim_{t\to+\infty} t^2(t^r f_X(t)) = 0$ donc $t^r f_X(t) \leqslant \dfrac{1}{t^2}$ pour $t \geqslant t_0$.

$\mathbb{E}(X) = \displaystyle\int_{-\infty}^{+\infty} t f_X(t) dt = \int_0^{+\infty} t \frac{\lambda^a}{\Gamma(a)} e^{-\lambda t} t^{a-1}\, dt = \frac{1}{\lambda\Gamma(a)} \int_0^{+\infty} u^a e^{-u}\, du = \frac{1}{\lambda}\frac{\Gamma(a+1)}{\Gamma(a)} = \frac{a}{\lambda}$;

$\mathbb{E}(X^2) = \displaystyle\int_{-\infty}^{+\infty} t^2 f_X(t) dt = \int_0^{+\infty} t^2 \frac{\lambda^a}{\Gamma(a)} e^{-\lambda t} t^{a-1}\, dt = \frac{1}{\lambda^2\Gamma(a)} \int_0^{+\infty} u^{a+1} e^{-u}\, du = \frac{1}{\lambda^2}\frac{\Gamma(a+2)}{\Gamma(a)} = \frac{a(a+1)}{\lambda}$;

$$\text{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{a(a+1)}{\lambda^2} - \frac{a^2}{\lambda^2} = \frac{a}{\lambda^2}.$$

$$\boxed{\mathbb{E}(X) = \frac{a}{\lambda} \text{ et } \text{var}(X) = \frac{a}{\lambda^2}.}$$

d) Loi Normale $\mathcal{N}(m, \sigma^2)$:

Loi de densité f_X définie par $f_X(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left(-\dfrac{(x-m)^2}{\sigma^2}\right)$.

X admet des moments de tous ordres car $\displaystyle\lim_{t\to+\infty} t^2(t^r f_X(t)) = 0$ donc $t^r f_X(t) \leqslant \dfrac{1}{t^2}$ pour $t \geqslant t_0$.

On a vu au Chapitre 3 que, si une v.a.r. X a pour loi $\mathcal{N}(m, \sigma^2)$, alors $X^* = \dfrac{X-m}{\sigma}$ est une v.a.r. de loi normale $\mathcal{N}(0,1)$.

$\mathbb{E}(X^*) = \displaystyle\int_{-\infty}^{+\infty} t f_{X^*}(t)\, dt = \int_{-\infty}^{+\infty} t \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 0$ (intégrale d'une fonction impaire sur $\mathbb{R}$)

$\mathbb{E}(X^{*2}) = \displaystyle\int_{-\infty}^{+\infty} t^2 f_{X^*}(t)\, dt = 2\int_0^{+\infty} t^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ (intégrale d'une fonction paire sur $\mathbb{R}$).

On fait une intégration par parties en posant $u' = te^{-\frac{t^2}{2}}$ et $v = t$ ($u = -e^{-\frac{t^2}{2}}$ et $v' = 1$).

$2\displaystyle\int_0^A t^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt = \frac{2}{\sqrt{2\pi}} \left(\left[-te^{-\frac{t^2}{2}} \right]_0^A + \int_0^A e^{-\frac{t^2}{2}}\, dt \right) = -\frac{2}{\sqrt{2\pi}} A e^{-\frac{A^2}{2}} + \frac{2}{\sqrt{2\pi}} \int_0^A e^{-\frac{t^2}{2}} dt \to 1$ quand $A \to +\infty$ car $te^{-\frac{t^2}{2}} \to 0$ et $\dfrac{2}{\sqrt{2\pi}} \displaystyle\int_0^A e^{-\frac{t^2}{2}}\, dt = \frac{1}{\sqrt{2\pi}} \int_{-A}^A e^{-\frac{t^2}{2}}\, dt \to 1$.

On a donc $\mathbb{E}(X^*) = 0$, $\mathbb{E}(X^{*2}) = 1$ et $\text{var}(X^*) = \mathbb{E}(X^{*2}) - \mathbb{E}(X^*)^2 = 1$.

Or $X = \sigma X^* + m$, donc $\mathbb{E}(X) = \sigma\mathbb{E}(X^*) + m$ et $\text{var}(X) = \sigma^2 \text{var}(X^*)$ d'où :

$$\boxed{\mathbb{E}(X) = m \text{ et } \text{var}(X) = \sigma^2.}$$

Parallèle avec les v.a.r. discrètes :

v.a.r. discrète	v.a.r. absolument continue
$X(\Omega) = \{x_n\}$	$X(\Omega) \subset \mathbb{R}$
$F_X(x) = P([X \leqslant x]) = \sum_{x_n \leqslant x} P([X = x_n])$	$F_X(x) = P([X \leqslant x]) = \int_{-\infty}^{x} f_X(t)dt$.
$\mathbb{E}(\varphi(X)) = \sum_n \varphi(x_n) P([X = x_n])$	$\mathbb{E}(\varphi(X)) = \int_{-\infty}^{+\infty} \varphi(t) f_X(t)dt$

Chapitre 6

COUPLES ALÉATOIRES DISCRETS

6.1 Généralités

Soit X et Y deux v.a.r. discrètes, définies sur un même espace probabilisé $(\Omega, \mathcal{A}, P)$.
On pose $X(\Omega) = \{x_i \ ; \ i \in \mathbb{N}\}$ et $Y(\Omega) = \{y_j \ ; \ j \in \mathbb{N}\}$.

6.1.1 Loi de probabilité d'un couple (X, Y)

Définition 6.1 : On appelle *loi de probabilité du couple (X, Y)* l'ensemble des triplets (x_i, y_j, p_{ij}) avec

$$p_{ij} = P([X = x_i] \cap [Y = y_j]) \text{ pour } i \in \mathbb{N} \text{ et } j \in \mathbb{N}.$$

On a alors $\displaystyle\sum_i \sum_j p_{ij} = \sum_j \sum_i p_{ij} = 1$.

Dans le cas où $X(\Omega)$ et $Y(\Omega)$ sont finis avec peu d'éléments, les résultats peuvent être donnés dans un tableau à doubles entrées.

Exemple : Une urne contient 3 boules blanches et 4 boules rouges. On tire successivement 2 boules. On envisage les deux modes de tirages avec ou sans remise et, dans les deux cas, X_k vaut 1 si la k-ième boule tirée est blanche, 0 sinon ($k = 1$ ou 2). On pose $Y = \max(X_1, X_2)$. Déterminer, dans les deux cas, la loi de (X_1, X_2), puis la loi de (X_1, Y).

$$X_1(\Omega) = X_2(\Omega) = Y(\Omega) = \{0, 1\}.$$

Si on pose $q_{ij} = P([X_1 = i] \cap [Y = j])$, on a :
$q_{00} = P([X_1 = 0] \cap [Y = 0]) = P([X_1 = 0] \cap [X_2 = 0])$;
$q_{10} = P([X_1 = 1] \cap [Y = 0]) = 0$;
$q_{01} = P([X_1 = 0] \cap [Y = 1]) = P([X_1 = 0] \cap [X_2 = 1])$;
$q_{11} = P([X_1 = 1] \cap [Y = 1]) = P([X_1 = 1])$ car $[X_1 = 1] \subset [Y = 1]$.

a) Cas du tirage avec remise

Les événements $[X_1 = i]$ et $[X_2 = j]$ sont indépendants car les tirages se font avec remise et sont donc indépendants les uns des autres (la composition de l'urne ne varie pas).

On a donc $p_{ij} = P([X_1 = i] \cap [X_2 = j]) = P([X_1 = i])P([X_2 = j])$ pour $i, j \in \{0, 1\}$, avec $P([X_k = 1]) = \dfrac{3}{7}$ et $P([X_k = 0]) = \dfrac{4}{7}$ pour $k \in \{1, 2\}$. Ainsi, on obtient :

$$p_{00} = \frac{4}{7} \times \frac{4}{7} = \frac{16}{49} \ ; \ p_{10} = p_{01} = \frac{4}{7} \times \frac{3}{7} = \frac{12}{49} \ ; \ p_{11} = \frac{3}{7} \times \frac{3}{7} = \frac{9}{49} \ ;$$

$$q_{00} = p_{00} = \frac{16}{49} \ ; \ q_{10} = 0 \ ; \ q_{01} = p_{01} = \frac{12}{49} \ ; \ q_{11} = P([X_1 = 1]) = \frac{3}{7} = \frac{21}{49}.$$

b) Cas du tirage sans remise

IL n'y a pas ici indépendance des tirages car la composition de l'urne varie. On a ici

$$p_{ij} = P([X_1 = i] \cap [X_2 = j]) = P([X_1 = i])P^{[X_1=i]}([X_2 = j]).$$

Il vient alors :

$$p_{00} = \frac{4}{7} \times \frac{3}{6} = \frac{2}{7} \ ; \ p_{10} = \frac{3}{7} \times \frac{4}{6} = \frac{2}{7} \ ; \ p_{01} = \frac{4}{7} \times \frac{3}{6} = \frac{2}{7} \ ; \ p_{11} = \frac{3}{7} \times \frac{2}{6} = \frac{1}{7} \ ;$$

$$q_{00} = p_{00} = \frac{2}{7} \ ; \ q_{10} = 0 \ ; \ q_{01} = p_{01} = \frac{2}{7} \ ; \ q_{11} = P([X_1 = 1]) = \frac{3}{7}.$$

6.1.2 Lois marginales

Définition 6.2 : Les v.a.r. X et Y sont appelées *v.a.r. marginales* du couple (X, Y) ; les lois des v.a.r. X et Y sont appelées *lois marginales* du couple (X, Y).

On pose $p_{i.} = \sum_j p_{ij}$ et $p_{.j} = \sum_i p_{ij}$.

Théorème 6.1 : La loi de X est définie par l'ensemble des couples $(x_i, p_{i.})$ pour $i \in \mathbb{N}$ et la loi de Y est définie par l'ensemble des couples $(y_j, p_{.j})$ pour $j \in \mathbb{N}$.

Démonstration. $Y(\Omega) = \{y_j \ ; \ j \in \mathbb{N}\}$ donc $\Omega = \bigcup_j [Y = y_j]$ réunion dénombrable d'événements disjoints. Les événements $[Y = y_j]$ forment un système complet d'événements, donc :

$$\begin{aligned} [X = x_i] &= [X = x_i] \cap \Omega = [X = x_i] \cap \left(\bigcup_j [Y = y_j] \right) \\ &= \bigcup_j ([X = x_i] \cap [Y = y_j]) \end{aligned}$$

d'après la distributivité de $\cap$ par rapport à $\bigcup_j$. On a alors

$$P([X = x_i]) = \sum_j P\left([X = x_i] \cap [Y = y_j]\right) = \sum_j p_{ij} = p_{i.}$$

et de même

$$P([X = y_j]) = \sum_i P\left([X = x_i] \cap [Y = y_j]\right) = \sum_i p_{ij} = p_{.j}.$$

□

Lorsque l'on représente la loi du couple dans un tableau, la loi de X est obtenue en faisant la somme sur les lignes (c'est-à-dire que $p_{i.}$ est la somme des p_{ij} sur la i-ème ligne) et la loi de Y est obtenue en faisant la somme sur les colonnes (c'est-à-dire que $p_{.j}$ est la somme des p_{ij} sur la j-ème colonne).

Sur l'exemple de 6.1.1,

- dans le cas a), $p_{0.} = p_{00} + p_{01} = \dfrac{16+12}{49} = \dfrac{28}{49} = \dfrac{4}{7}$ et $p_{1.} = p_{10} + p_{11} = \dfrac{12+9}{49} = \dfrac{21}{49} = \dfrac{3}{7}$; de même $p_{.0} = \dfrac{4}{7}$ et $p_{.1} = \dfrac{3}{7}$;
- dans le cas b), $p_{0.} = p_{00} + p_{01} = \dfrac{2+2}{7} = \dfrac{4}{7}$ et $p_{1.} = p_{10} + p_{11} = \dfrac{2+1}{7} = \dfrac{3}{7}$; de même $p_{.0} = \dfrac{4}{7}$ et $p_{.1} = \dfrac{3}{7}$.

On constate que les lois marginales du couple (X_1, X_2) sont les mêmes dans les 2 cas, c'est-à-dire quelles que soient les modalités du tirage, alors que la loi du couple (X_1, X_2) est différente dans les 2 cas.

Ceci montre bien que la donnée des lois marginales est une information trop pauvre pour reconstituer la loi du couple.

On peut passer de la loi du couple aux lois marginales, mais on ne peut pas faire l'inverse en général.

6.1.3 Lois conditionnelles

Si $p_{i.} \neq 0$, la loi conditionnelle de Y sachant $[X = x_i]$ est définie par l'ensemble des couples $\left(y_j, \dfrac{p_{ij}}{p_{i.}}\right)$, pour $j \in \mathbb{N}$. On peut de même définir la loi conditionnelle de X sachant $[Y = y_j]$.

Lorsque l'on représente la loi du couple dans un tableau, la loi conditionnelle de X sachant $[Y = y_j]$ est obtenue en prenant la j-ème colonne du tableau, divisée par la somme des p_{ij} sur la j-ème colonne et la loi conditionnelle de Y sachant $[X = x_i]$ est obtenue en prenant la i-ème ligne, divisée par la somme des p_{ij} sur la i-ème ligne.

Sur l'exemple 6.1.1, loi conditionnelle de :

- X_1 sachant $[X_2 = 0]$: $\dfrac{p_{00}}{p_{.0}} = \dfrac{4}{7}$, $\dfrac{p_{10}}{p_{.0}} = \dfrac{3}{7}$ pour a) ; $\dfrac{p_{00}}{p_{.0}} = \dfrac{1}{2}$; $\dfrac{p_{10}}{p_{.0}} = \dfrac{1}{2}$ pour b) ;
- X_1 sachant $[X_2 = 1]$: $\dfrac{p_{01}}{p_{.1}} = \dfrac{4}{7}$, $\dfrac{p_{11}}{p_{.1}} = \dfrac{3}{7}$ pour a) ; $\dfrac{p_{01}}{p_{.1}} = \dfrac{2}{3}$; $\dfrac{p_{11}}{p_{.1}} = \dfrac{1}{3}$ pour b) ;
- X_2 sachant $[X_1 = 0]$: $\dfrac{p_{00}}{p_{0.}} = \dfrac{4}{7}$, $\dfrac{p_{01}}{p_{0.}} = \dfrac{3}{7}$ pour a) ; $\dfrac{p_{00}}{p_{0.}} = \dfrac{1}{2}$; $\dfrac{p_{01}}{p_{0.}} = \dfrac{1}{2}$ pour b) ;
- X_2 sachant $[X_1 = 1]$: $\dfrac{p_{10}}{p_{1.}} = \dfrac{4}{7}$, $\dfrac{p_{11}}{p_{1.}} = \dfrac{3}{7}$ pour a) ; $\dfrac{p_{10}}{p_{1.}} = \dfrac{2}{3}$; $\dfrac{p_{11}}{p_{1.}} = \dfrac{1}{3}$ pour b) ;

- X_1 sachant $[Y = 0]$: $\dfrac{q_{00}}{q_{.0}} = 1$, $\dfrac{q_{10}}{q_{.0}} = 0$ pour a) ; $\dfrac{q_{00}}{q_{.0}} = 1$; $\dfrac{q_{10}}{q_{.0}} = 0$ pour b) ;
- X_1 sachant $[Y = 1]$: $\dfrac{q_{01}}{q_{.1}} = \dfrac{12}{33}$, $\dfrac{q_{11}}{q_{.1}} = \dfrac{21}{33}$ pour a) ; $\dfrac{q_{01}}{q_{.1}} = \dfrac{2}{5}$; $\dfrac{q_{11}}{q_{.1}} = \dfrac{3}{5}$ pour b) ;
- Y sachant $[X_1 = 0]$: $\dfrac{q_{00}}{q_{0.}} = \dfrac{4}{7}$, $\dfrac{q_{01}}{q_{0.}} = \dfrac{3}{7}$ pour a) ; $\dfrac{q_{00}}{q_{0.}} = \dfrac{1}{2}$; $\dfrac{q_{01}}{q_{0.}} = \dfrac{1}{2}$ pour b) ;

- Y sachant $[X_1 = 1]$: $\dfrac{q_{10}}{q_{1.}} = 0$, $\dfrac{q_{11}}{q_{1.}} = 1$ pour a) ; $\dfrac{q_{10}}{q_{1.}} = 0$; $\dfrac{q_{11}}{q_{1.}} = 1$ pour b).

On remarque que, dans le cas a), la loi conditionnelle de X_k sachant $[X_l = 0]$ et la loi conditionnelle de X_k sachant $[X_l = 1]$ sont les mêmes que la loi de X_k, ce qui n'est pas le cas dans le cas b).

6.1.4 Indépendance de deux v.a.r. discrètes

Définition 6.3 : Deux v.a.r. discrètes X et Y sont dites *indépendantes* si, pour tout $(i, j) \in \mathbb{N}^2$, on a :

$$P([X = x_i] \cap [Y = y_j]) = P([X = x_i])P([Y = y_j]) \text{ (c'est-à-dire } p_{ij} = p_{i.}p_{.j}).$$

Remarque : C'est le seul cas où la loi du couple est entièrement déterminée par les lois marginales.

Propriété 6.1(admise) : Les v.a.r. X et Y sont indépendantes si et seulement si, pour toutes fonctions numériques f et g, les v.a.r. $f(X)$ et $g(Y)$ sont indépendantes.

6.1.5 Somme de deux v.a.r. discrètes

Théorème 6.2 : Soit X et Y deux v.a.r. discrètes et soit $Z = X + Y$.
Pour $z \in Z(\Omega)$, on pose $I_z = \{(i, j) \in \mathbb{N}^2 \ ; \ x_i + y_j = z\}$; alors $P([Z = z]) = \displaystyle\sum_{(i,j)\in I_z} p_{ij}$.

Démonstration. Pour calculer $P([Z = z])$, on considère l'ensemble $I_z = \{(i, j) \in \mathbb{N}^2 \ ; \ x_i + y_j = z\}$ qui décrit toutes les situations possibles pour lesquelles $X + Y$ prend la valeur z.

On a alors $[Z = z] = \displaystyle\bigcup_{(i,j)\in I_z} ([X = x_i] \cap [Y = y_j])$ réunion d'événements disjoints, donc

$$P([Z = z]) = \sum_{(i,j)\in I_z} P([X = x_i] \cap [Y = y_j]) = \sum_{(i,j)\in I_z} p_{ij}.$$

□

Exemple : On lance un dé deux fois de suite et on s'intéresse à la somme des points.

$$Z(\Omega) = \{2, \cdots, 12\},$$
$$I_6 = \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}, \ I_7 = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}.$$

$P\left([X = x_i] \cap [Y = y_j]\right) = \dfrac{1}{36}$ pour tout couple $(x_i, y_j) \in \{1, \cdots, 6\}^2$ donc

$$P([Z = 6]) = \frac{5}{36} \text{ et } P([Z = 7]) = \frac{6}{36}.$$

Cas particulier important si X et Y sont indépendantes et à valeurs dans $\mathbb{N}$:

Théorème 6.3 : Si X et Y sont indépendantes à valeurs dans $\mathbb{N}$, de fonctions génératrices respectives G_X et G_Y, alors $P([X + Y = n]) = \displaystyle\sum_{i+j=n} p_{i.}p_{.j}$ et $G_{X+Y}(t) = G_X(t)G_Y(t)$ pour tout $t \in [0, 1]$.

Démonstration.

$$G_X(t)G_Y(t) = \left(\sum_{i\geqslant 0} t^i P([X=i])\right)\left(\sum_{j\geqslant 0} t^j P([Y=j])\right) = \sum_{i\geqslant 0}\sum_{j\geqslant 0} t^{i+j} P([X=i])P([Y=j])$$

(On peut échanger l'ordre des termes car les deux séries sont à termes positifs.)

On a, d'après l'indépendance de X et de Y, $P([X=i])P([Y=j]) = P([X=i]\cap[Y=j])$ et, si $I_k = \{(i,j)\in\mathbb{N}^2\ ;\ i+j=k\}$, pour $k\in\mathbb{N}$, les I_k forment une partition de $\mathbb{N}^2$. D'où :

$$G_X(t)G_Y(t) = \sum_{k\geqslant 0}\sum_{(i,j)\in I_k} t^{i+j}P([X=i]\cap[Y=j]) = \sum_{k\geqslant 0} t^k P([X+Y=k]) = G_{X+Y}(t).$$

□

Application :

Propriétés 6.2 : Si X et Y sont deux v.a.r. indépendantes :
1) si X suit la loi binomiale $\mathcal{B}(n,p)$ et Y la loi $\mathcal{B}(m,p)$, alors $X+Y$ suit la loi $\mathcal{B}(n+m,p)$;
2) si X suit la loi de Poisson $\mathcal{P}(\lambda)$ et Y la loi $\mathcal{P}(\mu)$, alors $X+Y$ suit la loi $\mathcal{P}(\lambda+\mu)$.

Démonstration. **1)** $G_X(t) = \displaystyle\sum_{k=0}^{n} t^k C_n^k p^k(1-p)^{n-k} = (pt+1-p)^n$ et de même $G_Y(t) = (pt+1-p)^m$ donc

$$G_{X+Y}(t) = G_X(t)G_Y(t) = (pt+1-p)^{n+m}.$$

On reconnait là la fonction génératrice de la loi binomiale $\mathcal{B}(n+m,p)$ donc $X+Y$ suit la loi $\mathcal{B}(n+m,p)$.

2) De même, $G_X(t) = \displaystyle\sum_{k=0}^{+\infty} t^k e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda t}$ et $G_Y(t) = e^{-\mu}e^{\mu t}$ donc

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

On reconnait là la fonction génératrice de la loi de Poisson $\mathcal{P}(\lambda+\mu)$ donc $X+Y$ suit la loi $\mathcal{P}(\lambda+\mu)$. □

Remarque : On peut aussi démontrer directement ces résultats :

1) $P([X+Y=k]) = \displaystyle\sum_{(i,j)\in I_k} C_n^i p^i(1-p)^{n-i}C_m^j p^j(1-p)^{m-j} = p^k(1-p)^{n+m-k}\sum_{(i,j)\in I_k} C_n^i C_m^j.$

Or $\displaystyle\sum_{(i,j)\in I_k} C_n^i C_m^j = C_{n+m}^k$ (formule combinatoire que l'on peut montrer en utilisant, par exemple, les chemins). Donc $P([X+Y=k]) = C_{n+m}^k p^k(1-p)^{n+m-k}$.

2) $P([X+Y=k]) = \displaystyle\sum_{(i,j)\in I_k}\frac{e^{-\lambda}\lambda^i}{i!}\frac{e^{-\mu}\mu^j}{j!} = e^{-(\lambda+\mu)}\sum_{(i,j)\in I_k}\frac{\lambda^i\mu^j}{i!j!} = \frac{e^{-(\lambda+\mu)}}{k!}\sum_{i=0}^{k} C_k^i\lambda^i\mu^{k-i}$, d'où

$$P([X+Y=k]) = e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^k}{k!}.$$

□

6.1.6 Espérance conditionnelle

Soit X une v.a.r discrète définie sur un espace probabilisé $(\Omega, \mathcal{A}, P)$ et Y une v.a.r. discrète définie sur le même espace probabilisé et admettant une espérance ($\mathbb{E}(|Y|)$ existe et est fini). Soit $x_i \in X(\Omega)$ tel que $P([X = x_i]) > 0$. Montrons que l'espérance d'une v.a.r. discrète Z ayant pour loi la loi conditionnelle de Y à l'événement $[X = x_i]$ (c'est-à-dire $P([Z = y_j] = P^{[X=x_i]}([Y = y_j])$ pour tout $y_j \in Y(\Omega)$) existe. On a en effet :

$$\begin{aligned}
\sum_{y_j \in Y(\Omega)} |y_j|\, P^{[X=x_i]}([Y = y_j]) &= \sum_{y_j \in Y(\Omega)} |y_j|\, \frac{P([X = x_i] \cap [Y = y_j])}{P([X = x_i])} \\
&= \frac{1}{P([X = x_i])} \sum_{y_j \in Y(\Omega)} |y_j| P([X = x_i] \cap [Y = y_j])
\end{aligned}$$

avec $\displaystyle\sum_{y_j \in Y(\Omega)} |y_j| P([X = x_i] \cap [Y = y_j]) \leqslant \sum_{y_j \in Y(\Omega)} |y_j| P([Y = y_j]) < +\infty$. Ainsi,

Définition 6.4 : L'espérance d'une v.a.r. dont la loi est la loi conditionnelle de Y à l'événement $[X = x_i]$ est appelée *espérance conditionnelle de Y à l'événement $[X = x_i]$*. Elle est notée $\mathbb{E}^{[X=x_i]}(Y)$. On a donc

$$\mathbb{E}^{[X=x_i]}(Y) = \sum_j y_j P^{[X=x_i]}([Y = y_j]).$$

Exemple : Soit X une v.a.r. de Poisson de paramètre $\lambda > 0$ et Y une v.a.r. de Poisson de paramètre $\mu > 0$ avec X et Y indépendantes. On a vu, en propriété 6.2 2), que $X + Y$ est de loi de Poisson de paramètre $\lambda + \mu$. On a alors, si $0 \leqslant i \leqslant n$,

$$\begin{aligned}
P^{[X+Y=n]}([X = i]) &= \frac{P([X = i] \cap [X + Y = n])}{P([X + Y = n])} = \frac{P([X = i] \cap [Y = n - i])}{P([X + Y = n])} \\
&= \frac{P([X = i])P([Y = n - i])}{P([X + Y = n])} = \frac{e^{-\lambda} \frac{\lambda^i}{i!}\, e^{-\mu} \frac{\mu^{n-i}}{(n-i)!}}{e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^n}{n!}} \\
&= \frac{n!}{i!(n-i)!} \left(\frac{\lambda}{\lambda + \mu}\right)^i \left(\frac{\mu}{\lambda + \mu}\right)^{n-i} = C_n^i \left(\frac{\lambda}{\lambda + \mu}\right)^i \left(1 - \frac{\lambda}{\lambda + \mu}\right)^{n-i} = P_Z(\{i\})
\end{aligned}$$

où Z est de loi binomiale $\mathcal{B}\left(n, \dfrac{\lambda}{\lambda + \mu}\right)$. Ainsi, la loi conditionnelle de X à $[X + Y = n]$ est la loi binomiale $\mathcal{B}\left(n, \dfrac{\lambda}{\lambda + \mu}\right)$, d'espérance $\dfrac{n\lambda}{\lambda + \mu}$ et on a donc

$$\mathbb{E}^{[X+Y=n]}(X) = \frac{n\mu}{\lambda + \mu}.$$

Définition 6.5 : Soit X une v.a.r. discrète sur $(\Omega, \mathcal{A}, P)$ telle que, pour tout $x \in X(\Omega)$, $P([X = x]) \neq 0$ et soit Y une v.a.r discrète sur $(\Omega, \mathcal{A}, P)$ admettant une espérance. On appelle *espérance conditionnelle de Y à X*, la v.a.r. discrète $h(X)$ avec $h(x) = \mathbb{E}^{[X=x]}(Y)$ pour tout x. On peut donc écrire :

$$\mathbb{E}^X(Y) = \sum_{x \in X(\Omega)} \mathbb{E}^{[X=x]}(Y)\, \mathbb{I}_{[X=x]}.$$

Exemple : Dans l'exemple précédent ($P_X = \mathcal{P}(\lambda)$, $P_Y = \mathcal{P}(\mu)$ avec X et Y indépendantes), on a

$$\mathbb{E}^{X+Y}(X) = \sum_{n=0}^{+\infty} \frac{n\lambda}{\lambda+\mu} \mathbb{1}_{[X+Y=n]} = \frac{(X+Y)\lambda}{\lambda+\mu}$$

puisque la somme porte sur toutes les valeurs possibles de $X+Y$.

Théorème 6.4 : (Théorème de l'espérance totale)
Soit X et Y deux v.a.r. discrètes définies sur le même espace telles que $\mathbb{E}(Y)$ existe. Alors la v.a.r. discrète $\mathbb{E}^X(Y)$ admet une espérance et

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

Démonstration. On rappelle que $\mathbb{E}(\varphi(X)) = \sum_{x\in X(\Omega)} \varphi(x)P([X=x])$ (sous réserve d'existence, c'est-à-dire de convergence absolue) que l'on applique ici à $\varphi(X) = \mathbb{E}^X(Y) = \sum_{x\in X(\Omega)} \mathbb{E}^{[X=x]}(Y)\,\mathbb{1}_{[X=x]}$, c'est-à-dire que $\varphi(x) = \mathbb{E}^{[X=x]}(Y) = \sum_{y\in Y(\Omega)} yP^{[X=x]}([Y=y])$ pour tout $x\in X(\Omega)$. On a alors

$$\begin{aligned}
\mathbb{E}(\mathbb{E}^X(Y)) &= \sum_{x\in X(\Omega)}\sum_{y\in Y(\Omega)} yP^{[X=x]}([Y=y])P([x=x]) \\
&= \sum_{x\in X(\Omega)}\sum_{y\in Y(\Omega)} yP([X=x]\cap[Y=y]) \\
&= \sum_{y\in Y(\Omega)} y\left(\sum_{x\in X(\Omega)} P([X=x]\cap[Y=y]\right) = \sum_{y\in Y(\Omega)} yP([Y=y]) = \mathbb{E}(Y)
\end{aligned}$$

les interversions de sommes étant légitimes puisque

$$\sum_{y\in Y(\Omega)} |y|P([Y=y]) = \sum_{y\in Y(\Omega)} |y| \sum_{x\in X(\Omega)} P([X=x]\cap[Y=y]) < +\infty.$$

□

6.2 Opérateurs classiques

6.2.1 Espérance

Propriétés 6.3 :
1) Si X et Y possèdent une espérance, alors $\mathbb{E}(X+Y)$ existe et $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.
2) Si X et Y possèdent un moment d'ordre 2, alors $\mathbb{E}(XY)$ existe.

Démonstration. **1)** Soit $Z = X+Y$.
Alors $[Z=z] = \bigcup_{(i,j)\in I_z} ([X=x_i]\cap[Y=y_j])$ où $I_z = \{(i,j)\in\mathbb{N}^2 \;;\; x_i+y_j = z\}$.

$$\begin{aligned}
\sum_{z\in Z(\Omega)} zP([Z=z]) &= \sum_{z\in Z(\Omega)} z \sum_{(i,j)\in I_z} P([X=x_i]\cap[Y=y_j]) \\
&= \sum_{z\in Z(\Omega)} \sum_{(i,j)\in I_z} (x_i+y_j)P([X=x_i]\cap[Y=y_j]) = \sum_{(i,j)}(x_i+y_j)p_{ij} \\
&= \sum_{(i,j)} x_i p_{ij} + \sum_j y_j p_{ij} = \sum_i x_i \sum_j p_{ij} + \sum_j y_j \sum_i p_{ij} \\
&= \sum_i x_i p_{i.} + \sum_j y_j p_{.j}
\end{aligned}$$

Les deux séries obtenues sont absolument convergentes car $\mathbb{E}(X)$ et $\mathbb{E}(Y)$ existent donc, en reprenant la démarche précédente avec $\sum_{z\in Z(\Omega)} |z|P([Z=z])$ et en utilisant les inégalités du type $|z| \leqslant |x_i| + |y_j|$, on obtient l'absolue convergence de $\sum_{z\in Z(\Omega)} zP([Z=z])$, ce qui légitime ce qui a été fait ci-dessus (*i.e.* intervertir les sommations).

On a donc bien $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.

2) Si $Z = XY$ et si, pour $z \in Z(\Omega)$, on pose $J_z = \{(i,j) \in \mathbb{N}^2 \ ; \ x_i y_j = z\}$, alors $[Z=z] = \bigcup_{(i,j)\in J_z} ([X=x_i]\cap[Y=y_j])$ et, sous réserve d'absolue convergence pour pouvoir intervertir les sommes, on a :

$$\begin{aligned}
\sum_{z\in Z(\Omega)} zP([Z=z]) &= \sum_{z\in Z(\Omega)} z \sum_{(i,j)\in I_z} P([X=x_i]\cap[Y=y_j]) \\
&= \sum_{z\in Z(\Omega)} \sum_{(i,j)\in J_z} x_i y_j P([X=x_i]\cap[Y=y_j]) \\
&= \sum_{z\in Z(\Omega)} \sum_{(i,j)\in J_z} x_i y_j p_{ij} = \sum_{(i,j)} x_i y_j p_{ij}.
\end{aligned}$$

Or $|x_i y_j| \leqslant \frac{1}{2}(x_i^2 + y_j^2)$ car $\frac{1}{2}(x_i^2+y_j^2) - |x_i y_j| = \frac{1}{2}\left(|x_i| - |y_j|\right)^2 \geqslant 0$ et

$$\begin{aligned}
\sum_{(i,j)} |x_i y_j| p_{ij} &\leqslant \frac{1}{2}\sum_{(i,j)}(x_i^2+y_j^2)p_{ij} = \frac{1}{2}\left(\sum_{(i,j)} x_i^2 p_{ij} + \sum_{(i,j)} y_j^2 p_{ij}\right) \\
&= \frac{1}{2}\sum_i x_i^2 \sum_j p_{ij} + \frac{1}{2}\sum_j y_j^2 \sum_i p_{ij} = \frac{1}{2}\left(\sum_i x_i^2 p_{i.} + \sum_j y_j^2 p_{.j}\right) \\
&= \frac{1}{2}\mathbb{E}(X^2) + \frac{1}{2}\mathbb{E}(Y^2) < +\infty
\end{aligned}$$

car X et Y sont d'ordre 2. La convergence absolue est donc bien vérifiée et on a alors

$$\mathbb{E}(XY) = \sum_{z\in Z(\Omega)} zP([Z=z]) = \sum_{(i,j)} x_i y_j p_{ij} \text{ et } \mathbb{E}(XY) \leqslant \frac{1}{2}\mathbb{E}(X^2) + \frac{1}{2}\mathbb{E}(Y^2).$$

□

Théorème 6.5 : Soit φ une application de $\mathbb{R}^2$ dans $\mathbb{R}$, telle que $\varphi(X,Y)$ admette une espérance. Alors :

$$\mathbb{E}(\varphi(X,Y)) = \sum_{(i,j)} \varphi(x_i,y_j)P([X=x_i]\cap[Y=y_j]) = \sum_{i,j} \varphi(x_i,y_j)p_{ij}.$$

Démonstration. La démonstration est identique à celle des propriétés précédentes.

On pose $Z = \varphi(X, Y)$ et, pour $z \in Z(\Omega)$, $K_z = \{(i, j) \in \mathbb{N}^2 \ ; \ \varphi(x_i, y_j) = z\}$

$Z(\Omega)$ est fini ou dénombrable car $X(\Omega)$ et $Y(\Omega)$ le sont. On a alors

$$\begin{aligned}
\sum_{z \in Z(\Omega)} zP([Z = z]) &= \sum_{z \in Z(\Omega)} z \sum_{(i,j) \in K_z} P([X = x_i] \cap [Y = y_j]) \\
&= \sum_{z \in Z(\Omega)} \sum_{(i,j) \in K_z} \varphi(x_i, y_j) P([X = x_i] \cap [Y = y_j]) \\
&= \sum_{z \in Z(\Omega)} \sum_{(i,j) \in K_z} \varphi(x_i, y_j) p_{ij} = \sum_{(i,j)} \varphi(x_i, y_j) p_{ij}.
\end{aligned}$$

Les interversions des sommes sont légitimes car Z est d'ordre 1 donc $\sum_{z \in Z(\Omega)} zP([Z = z])$ est absolument convergente. On a donc

$$\mathbb{E}(\varphi(X, Y)) = \sum_{(i,j)} \varphi(x_i, y_j) p_{ij}$$

□

Application :

Propriété 6.4 : Si X et Y sont indépendantes, alors $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Démonstration. En effet, on a alors $p_{ij} = p_{i.}p_{.j}$ et, d'après ce qui précède,

$$\mathbb{E}(XY) = \sum_{(i,j)} x_i y_j p_{ij} = \sum_{(i,j)} x_i y_j p_{i.} p_{.j} = \left(\sum_i x_i p_{i.} \right) \left(\sum_j y_j p_{.j} \right) = \mathbb{E}(X)\mathbb{E}(Y).$$

□

Remarque : La réciproque est en général fausse.

6.2.2 Variance et covariance

Définition 6.4 : Si X et Y sont d'ordre 2, on appelle :

1) <u>*covariance*</u> de X et de Y, le réel $\text{cov}(X, Y)$ défini par

$$\text{cov}(X, Y) = \mathbb{E}\left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)) \right) \ ;$$

2) <u>*coefficient de corrélation linéaire*</u> de X et de Y, le réel $\rho(X, Y)$ défini par

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)} \ ;$$

3) <u>*matrice de covariance*</u> de X et de Y, la matrice $\Gamma(X, Y)$ définie par :

$$\Gamma(X, Y) = \begin{pmatrix} \text{var}(X) & \text{cov}(X, Y) \\ \text{cov}(Y, X) & \text{var}(Y) \end{pmatrix} = \begin{pmatrix} \text{cov}(X, X) & \text{cov}(X, Y) \\ \text{cov}(Y, X) & \text{cov}(Y, Y) \end{pmatrix}.$$

Propriétés 6.5 :
1)i) $\text{cov}(X,Y) = \text{cov}(Y,X) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$;
1)ii) $\text{var}(X+Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X,Y)$;
1)iii) si X et Y sont indépendantes, alors $\text{cov}(X,Y) = 0$ et $\text{var}(X+Y) = \text{var}(X) + \text{var}(Y)$;
1)iv) $\text{cov}(aX+b, cY+d) = ac\,\text{cov}(X,Y)$.

2) $\rho(X,Y) \in [-1,1]$ et $\rho(aX+b, cY+d) = \dfrac{ac}{|ac|}\rho(X,Y)$.

3) $\Gamma(X,Y)$ est une matrice réelle symétrique et, pour tout $(u,v) \in \mathbb{R}^2$,

$$(u,v)\Gamma(X,Y)\begin{pmatrix} u \\ v \end{pmatrix} \geqslant 0.$$

Démonstration. **1)i)** $\text{cov}(X,Y) = \mathbb{E}\left((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\right) = \mathbb{E}(XY - \mathbb{E}(X)Y - X\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y))$ puis, en utilisant la linéarité de $\mathbb{E}$,

$$\text{cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(\mathbb{E}(X)Y) - \mathbb{E}(X\mathbb{E}(Y)) + \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

1)ii) Toujours en utilisant la linéarité de $\mathbb{E}$, on a :

$$\begin{aligned}
\text{var}(X+Y) &= \text{cov}(X+Y, X+Y) = \mathbb{E}((X+Y)^2) - \mathbb{E}(X+Y)^2 \\
&= \mathbb{E}(X^2 + 2XY + Y^2) - (\mathbb{E}(X) + \mathbb{E}(Y))^2 \\
&= \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) - \mathbb{E}(X)^2 - \mathbb{E}(Y)^2 - 2\mathbb{E}(X)\mathbb{E}(Y) \\
&= \text{var}(X) + \text{var}(Y) + 2\text{cov}(X,Y)
\end{aligned}$$

1)iii) Découle de 1)i) et de la propriété 6.4.
1)iv)

$$\begin{aligned}
\text{cov}(aX+b, cY+d) &= \mathbb{E}\left((aX+b)(cY+d)\right) - \mathbb{E}(aX+b)\mathbb{E}(cY+d) \\
&= \mathbb{E}(acXY + adX + bcY + bd) - (a\mathbb{E}(X)+b)\left(c\mathbb{E}(Y)+d\right) \\
&= ac\mathbb{E}(XY) + ad\mathbb{E}(X) + bc\mathbb{E}(Y) + bd \\
&\quad -ac\mathbb{E}(X)\mathbb{E}(Y) - ad\mathbb{E}(X) - bc\mathbb{E}(Y) - bd \\
&= ac\text{cov}(X,Y)
\end{aligned}$$

2)i) Pour tout $\lambda \in \mathbb{R}$, on a $\text{var}(X+\lambda Y) \geqslant 0$.
Or $\text{var}(X+\lambda Y) = \text{var}(X) + 2\lambda\text{cov}(X,Y) + \lambda^2\text{var}(Y)$, qui est un trinôme du second degré en λ toujours positif : son discriminant réduit est donc négatif, c'est-à-dire que l'on a :

$$(\text{cov}(X,Y))^2 - \text{var}(X)\text{var}(Y) \leqslant 0.$$

Il s'ensuit alors que $\rho^2(X,Y) - 1 \leqslant 0$.

3)i) Immédiat car $\text{cov}(X,Y) = \text{cov}(Y,X)$.

3)ii)

$$\begin{aligned}
\begin{pmatrix} u & v \end{pmatrix}\Gamma(X,Y)\begin{pmatrix} u \\ v \end{pmatrix} &= \begin{pmatrix} u & v \end{pmatrix}\begin{pmatrix} \text{var}(X) & \text{cov}(X,Y) \\ \text{cov}(X,Y) & \text{var}(Y) \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} \\
&= \begin{pmatrix} u & v \end{pmatrix}\begin{pmatrix} u\text{var}(X) + v\text{cov}(X,Y) \\ u\text{cov}(X,Y) + v\text{var}(Y) \end{pmatrix} \\
&= u^2\text{var}(X) + 2uv\text{cov}(X,Y) + v^2\text{var}(Y) = \text{var}(uX+vY) \geqslant 0.
\end{aligned}$$

□

<u>*Exercice*</u> : Une urne contient N boules dont 2 blanches et $N-2$ noires. On tire toutes les boules une par une sans remise et on note R_1 (resp. R_2) le rang de la première (resp. de la deuxième) boule blanche tirée.
1) Déterminer la loi du couple (R_1, R_2) et en déduire les lois marginales de R_1 et de R_2.
2) Montrer que R_1, $N+1-R_2$ et R_2-R_1 ont même loi et en déduire, sans calculs compliqués $\mathbb{E}(R_1)$, $\mathbb{E}(R_2)$ et $\rho(R_1, R_2)$.

Éléments de réponses : $\mathbb{E}(R_1) = \dfrac{N+1}{3}$, $\mathbb{E}(R_2) = \dfrac{2(N+1)}{3}$ et $\rho(R_1, R_2) = \dfrac{1}{2}$.

Chapitre 7

COUPLES ALÉATOIRES À DENSITÉ

7.1 Loi de probabilité d'un couple de variables aléatoires réelles

Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité et X et Y deux v.a.r. définies sur $(\Omega, \mathcal{A}, P)$.

Définition 7.1 : On appelle *fonction de répartition du couple* (X, Y) la fonction $F_{X,Y}$ définie sur $\mathbb{R}^2$ par :

$$F_{X,Y}(x, y) = P([X \leqslant x] \cap [Y \leqslant y]).$$

Propriété 7.1 : Les fonctions de répartition des v.a.r. X et Y vérifient

$$F_X(x) = \lim_{y \to +\infty} F_{X,Y}(x, y) \text{ et } F_Y(y) = \lim_{x \to +\infty} F_{X,Y}(x, y).$$

Démonstration. $\Omega = \bigcup_n [Y \leqslant n]$ réunion croissante d'événements, donc, par distributivité,

$$[X \leqslant x] = [X \leqslant x] \cap \left(\bigcup_n [Y \leqslant n] \right) = \bigcup_n ([X \leqslant x] \cap [Y \leqslant n])$$

réunion croissante d'événements. On a donc, d'après la propriété 1.6 5),

$$P([X \leqslant x]) = \lim_{n \to +\infty} P([X \leqslant x] \cap [Y \leqslant n]).$$

Il vient alors

$$F_X(x) = \lim_{n \to +\infty} F_{X,Y}(x, n) = \lim_{y \to +\infty} F_{X,Y}(x, y)$$

et, X et Y jouant des rôles symétriques, la deuxième assertion en découle de la même façon. $\square$

7.1.1 Définition d'un couple absolument continu

Définition 7.2 : La loi du couple (X,Y) est dite *absolument continue* s'il existe une application $f_{X,Y}$ de $\mathbb{R}^2$ sur $\mathbb{R}$, positive et borélienne (c'est-à-dire que, pour tout $B \in \mathcal{B}_{\mathbb{R}}$, $f_{X,Y}^{-1}(B) \in \mathcal{B}_{\mathbb{R}^2}$), appelée *densité* du couple (X,Y), telle que, pour tout $(x,y) \in \mathbb{R}^2$:

$$F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)dudv.$$

Remarques :

1) Toutes les fonctions "suffisamment continues" sont boréliennes. En pratique, on prendra souvent $f_{X,Y}$ continue sur l'intérieur d'un sous-ensemble D de $\mathbb{R}^2$ et nulle sur son complémentaire.

2) Il existe des couples (X,Y) non discrets n'admettant pas non plus de densité. C'est le cas si, par exemple, X est une v.a.r. discrète et Y une v.a.r. absolument continue.

On n'étudiera pas ces couples dans ce cours.

Propriétés 7.2 :

1) $\displaystyle\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} f_{X,Y}(u,v)du\right)dv = \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} f_{X,Y}(u,v)dv\right)du = 1.$

2) Les lois marginales de X et de Y admettent les densités f_X et f_Y définies par

$$f_X(u) = \int_{-\infty}^{+\infty} f_{X,Y}(u,v)dv \text{ et } f_Y(v) = \int_{-\infty}^{+\infty} f_{X,Y}(u,v)du.$$

3) En tout (x_0,y_0) où $f_{X,Y}$ est continue, on a $f_{X,Y}(x_0,y_0) = \dfrac{\partial^2 F_{X,Y}}{\partial x \partial y}(x_0,y_0)$.

Démonstration. **3)** Découle de la définition de $f_{X,Y}$ (définition 7.2).

2) $F_X(x) = \lim\limits_{y\to+\infty} F_{X,Y}(x,y)$ avec

$$F_{X,Y}(x,y) = \int_{-\infty}^{x}\left(\int_{-\infty}^{y} f_{X,Y}(u,v)dv\right)du = \int_{-\infty}^{y}\left(\int_{-\infty}^{x} f_{X,Y}(u,v)du\right)dv$$

donc $F_X(x) = \displaystyle\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{x} f_{X,Y}(u,v)du\right)dv = \int_{-\infty}^{x}\left(\int_{-\infty}^{+\infty} f_{X,Y}(u,v)dv\right)du$ et

$$f_X(x) = F_X'(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,v)dv.$$

Comme X et Y jouent des rôles symétriques, la deuxième assertion en découle de la même façon.

1) Comme $\displaystyle\int_{-\infty}^{+\infty} f_X(x)dx = 1$, on déduit de 2) que $\displaystyle\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} f_{X,Y}(x,v)dv\right)dx = 1.$ □

Parallèle avec les couples discrets :

1) est l'analogue de $\displaystyle\sum_i\left(\sum_j p_{ij}\right) = \sum_j\left(\sum_i p_{ij}\right) = 1$;

2) est l'analogue de $\displaystyle p_{i.} = \sum_j p_{ij}$ et $\displaystyle p_{.j} = \sum_i p_{ij}$.

Exemple : Soit (X, Y) un couple de loi uniforme sur le disque $D(O, R)$. Déterminer sa densité ainsi que les lois marginales de X et de Y.

(X, Y) est un couple de loi uniforme sur le disque $D(O, R)$ donc sa densité est constante sur $D(O, R)$ et nulle ailleurs.

$f_{X,Y}(x, y) = C\, \mathbb{1}_{D(O,R)}(x, y)$ donc $C \displaystyle\int\int_{D(O,R)} dx\, dy = 1$. Or $\displaystyle\int\int_{D(O,R)} dx\, dy$ est l'aire de $D(O, R)$, c'est-à-dire πR^2 et $C\pi R^2 = 1$ donne $C = \dfrac{1}{\pi R^2}$ d'où

$$f_{X,Y}(x, y) = \frac{1}{\pi R^2}\, \mathbb{1}_{D(O,R)}(x, y).$$

Ainsi, $f_{X,Y}(x, y) = \dfrac{1}{\pi R^2}$ si $x^2 + y^2 \leqslant R^2$ et 0 sinon.

- Si $|x| > R$, alors $x^2 + y^2 > R^2$ et $f_{X,Y}(x, y) = 0$ pour tout $y \in \mathbb{R}$, d'où $f_X(x) = 0$.
- Si $|x| \leqslant R$, alors $x^2 + y^2 \leqslant R^2$ pour $y^2 \leqslant R^2 - x^2$, c'est-à-dire pour $y \in \left[-\sqrt{R^2 - x^2}, \sqrt{R^2 - x^2}\right]$

et

$$f_X(x) = \int_{-\infty}^{+\infty} \frac{1}{\pi R^2}\, \mathbb{1}_{[-\sqrt{R^2-x^2}, \sqrt{R^2-x^2}]}(y)\, dy = \frac{2\sqrt{R^2 - x^2}}{\pi R^2}.$$

Finalement, $f_X(x) = \dfrac{2\sqrt{R^2 - x^2}}{\pi R^2}\, \mathbb{1}_{[-R,R]}(x)$ et $f_Y(y) = \dfrac{2\sqrt{R^2 - y^2}}{\pi R^2}\, \mathbb{1}_{[-R,R]}(y)$ car X et Y jouent le même rôle.

7.1.2 Lois conditionnelles dans le cas continu

Définition 7.3 : Pour tout $x \in \mathbb{R}$ tel que $f_X(x) \neq 0$, la <u>*densité conditionnelle*</u> de Y sachant $[X = x]$ est la fonction g_x, notée $f_Y^{X=x}$ définie par $g_x(y) = \dfrac{f_{X,Y}(x, y)}{f_X(x)}$.

Remarque : la fonction g_x est bien une densité car elle est positive et $\displaystyle\int_{-\infty}^{+\infty} g_x(y)\, dy = 1$.

<u>Parallèle avec les couples discrets :</u>

Ceci est l'analogue de $\dfrac{p_{ij}}{p_{i.}} = \dfrac{p_{ij}}{\sum\limits_k p_{ik}}$.

7.1.3 Espérance conditionnelle dans le cas continu

Soit (X, Y) un couple de v.a.r. absolument continu, de densité $f_{X,Y}$ et de densités marginales f_X et f_Y tel que $\mathbb{E}(Y)$ existe ($\mathbb{E}(|Y|)$ existe et est fini). On a vu que la loi conditionnelle $P_Y^{[X=x]}$ admet dans ce cas la densité définie par $f_Y^{X=x}(y) = \dfrac{f_{X,Y}(x, y)}{f_X(x)}$ pour tout $x \in \mathbb{R}$ tel que $f_X(x) > 0$. On a alors :

$$\int_{\mathbb{R}} |y| f_{X,Y}(x, y)\, dy < +\infty.$$

Définition 7.4 : L'*espérance conditionnelle* de Y à l'événement $[X = x]$ est le réel :

$$\mathbb{E}^{X=x}(Y) = \int_{\mathbb{R}} y f_Y^{X=x}(y)\, dy.$$

L'*espérance conditionnelle* de Y à X est la v.a.r. $\mathbb{E}^X(Y) = h(X)$ avec :

$$h(x) = \mathbb{E}^{X=x}(Y).$$

On a ici, tout comme dans le cas discret, le théorème de l'espérance totale si $\mathbb{E}(Y)$ existe :

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

(La démonstration est la même en remplaçant les sommes par des intégrales et les probabilités par des densités).
Exemple : Soit (X, Y) de densité $\lambda^2 e^{-\lambda y}$ et de support $D = \{(x, y)\ ;\ 0 \leqslant x \leqslant y\}$.

On a alors $f_X(x) = \displaystyle\int_x^{+\infty} \lambda^2 e^{-\lambda y}\, dy\, \mathbb{1}_{[0,+\infty[}(x) = \lambda e^{-\lambda x} \mathbb{1}_{[0,+\infty[}(x)$ (X est de loi exponentielle $\mathcal{E}(\lambda)$ et

$$f_Y^{X=x}(y) = \lambda\, e^{-\lambda(y-x)}\, \mathbb{1}_{[x,+\infty[}(y)\ ;$$

l'espérance de Y conditionnée par $X = x$ est égale à :

$$\mathbb{E}^{X=x}(Y) = \int_x^{+\infty} y\lambda\, e^{-\lambda(y-x)}\, dy = \int_0^{+\infty} (x + u)\lambda e^{-\lambda u}\, du = x + \frac{1}{\lambda}$$

(en faisant le changement de variable $u = y - x$). D'où il vient :

$$\mathbb{E}^X(Y) = X + \frac{1}{\lambda}.$$

7.2 Opérateurs

Si φ est une fonction de $\mathbb{R}^2$ sur $\mathbb{R}$ telle que $\varphi(X, Y)$ admette une espérance, alors :

$$\mathbb{E}(\varphi(X, Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \varphi(x, y) f_{X,Y}(x, y) dx dy.$$

En particulier, $\mathbb{E}(XY) = \displaystyle\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_{X,Y}(x, y) dx dy$ si cette espérance existe.

Les résultats démontrés pour les couples discrets au sujet des opérateurs restent valables : il suffit ici de remplacer les sommes par des intégrales et les probabilités par des densités.

7.3 Indépendance

Définition 7.5 : Deux v.a.r. X et Y sont dites *indépendantes* si, pour tout $(x, y) \in \mathbb{R}^2$, on a :

$$F_{X,Y}(x, y) = F_X(x) F_Y(y).$$

On admet que ceci équivaut à $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ pour tout $(x, y) \in \mathbb{R}^2$ ou bien à $\mathbb{E}(g(X)h(Y)) = \mathbb{E}(g(X))\mathbb{E}(h(Y))$ pour toutes fonctions g et h pourvu que ces espérances existent.

Attention : $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ n'est pas suffisant pour conclure sur l'indépendance de X et de Y.

Propriété 7.3 : Si X et Y sont indépendantes, alors $f_Y^{X=x} = f_Y$ pour tout x tel que $f_X(x) \neq 0$ et $f_X^{Y=y} = f_X$ pour tout y tel que $f_Y(y) \neq 0$.

Démonstration. $f_Y^{X=x}(y) = \dfrac{f_{X,Y}(x,y)}{f_X(x)}$ donc, si X et Y sont indépendantes, $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ et

$$f_Y^{X=x}(y) = \frac{f_X(x)f_Y(y)}{f_X(x)} = f_Y(y).$$

La deuxième assertion se démontre de la même façon, X et Y jouant des rôles symétriques. □

7.4 Changement de variables

Soit (X, Y) un couple aléatoire de densité $f_{X,Y}$ et ψ une fonction (borélienne) de $\mathbb{R}^2$ sur $\mathbb{R}^2$. On souhaite connaître la loi du couple $(U, V) = \psi(X, Y)$.

On va d'abord énoncer l'analogue de $f_{\varphi(X)}(y) = f_X(\varphi^{-1}(y))|(\varphi^{-1})'(y)|$ sur $\mathbb{R}$.

Théorème 7.1 : Soit (X, Y) un couple aléatoire de densité $f_{X,Y}$ et ψ une fonction de $\mathbb{R}^2$ sur $\mathbb{R}^2$. Si $f_{X,Y}$ est continue sur l'intérieur d'un ensemble D et nulle sur son complémentaire, si ψ est une bijection de D sur $E = \psi(D)$ telle que les dérivées partielles de ψ et de ψ^{-1} existent et soient continues, et si, de plus, $J(\psi^{-1}) = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} \neq 0$ sur E, alors le couple aléatoire $(U, V) = \psi(X, Y)$ admet pour densité la fonction $f_{U,V}$ définie par :

$$f_{U,V}(u, v) = f_{X,Y}(\psi^{-1}(u, v)) \left|J(\psi^{-1})(u, v)\right| \text{ si } (u, v) \in E.$$

Remarque : Comme dans le cas réel, il peut arriver que ψ ne soit pas bijective sur D tout entier mais que sa restriction ψ_i à chacun des sous-ensembles D_i, $i \in \{1, 2, \cdots, q\}$ d'une partition de D, le soit. On a alors

$$f_{U,V}(u, v) = \sum_{i=1}^{q} f_{X,Y}(\psi^{-1}(u, v))|J(\psi_i^{-1})(u, v)|\, \mathbb{1}_{\psi(D_i)}(u, v).$$

<u>Exemple</u> : Soit (X, Y) un couple aléatoire de densité $f_{X,Y}$ où $f_{X,Y}(x, y) = \dfrac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$. Quelle est la loi de $U = \dfrac{X}{Y}$?

Méthode : On calcule d'abord la loi de $(U, V) = \left(\dfrac{X}{Y}, Y\right)$, puis on déduit la loi de U comme loi marginale du couple (U, V).

On pose $\psi(x, y) = (u, v) = \left(\dfrac{x}{y}, y\right)$: $y = v$ et $x = uv$ conduit à

$$\psi^{-1}(u, v) = (x, y) = (uv, v).$$

Le domaine $\mathbb{R} \times \mathbb{R}^*$ se transforme en $\mathbb{R} \times \mathbb{R}^*$. Pour $(u,v) \in \mathbb{R} \times \mathbb{R}^*$:

$$J(\psi^{-1})(u,v) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v \neq 0,$$

donc $J(\psi^{-1}) \neq 0$ sur $\mathbb{R}\times]0,+\infty[$ et sur $\mathbb{R}\times]-\infty,0[$.

On a alors, pour $(u,v) \in \mathbb{R} \times \mathbb{R}^*$:

$$f_{U,V}(u,v) = f_{X,Y}(\psi^{-1}(u,v))|J(\psi^{-1}(u,v))| = \frac{1}{2\pi} e^{-\frac{1}{2}((uv)^2+v^2)}|v|.$$

Il vient alors

$$\begin{aligned} f_U(u) &= \int_{-\infty}^{+\infty} f_{U,V}(u,v)dv = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}v^2(u^2+1)}|v|\,dv = 2\times\frac{1}{2\pi}\int_0^{+\infty} e^{-\frac{1}{2}v^2(u^2+1)}v\,dv \\ &= \frac{1}{\pi}\frac{1}{u^2+1}\left[-e^{-\frac{1}{2}v^2(u^2+1)}\right]_0^{+\infty} = \frac{1}{\pi}\frac{1}{1+u^2}. \end{aligned}$$

Donc, si (X,Y) suit la loi de densité $f_{X,Y}$ définie par $f_{X,Y}(x,y) = \frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)}$ alors la loi de $U = \frac{X}{Y}$ est la loi de Cauchy $\mathcal{C}(1)$.

7.5 Sommes de deux v.a.r. absolument continues

Théorème 7.2 : Soit (X,Y) un couple aléatoire de densité $f_{X,Y}$. Alors :
1) la v.a.r. $X+Y$ a pour densité la fonction f_{X+Y} définie par :

$$f_{X+Y}(w) = \int_{-\infty}^{+\infty} f_{X,Y}(w-v,v)dv = \int_{-\infty}^{+\infty} f_{X,Y}(u,w-u)du.$$

2) Si X et Y sont indépendantes, alors :

$$f_{X+Y}(w) = \int_{-\infty}^{+\infty} f_X(w-v)f_Y(v)dv = \int_{-\infty}^{+\infty} f_X(u)f_Y(w-u)du.$$

Démonstration. On pose $(U,V) = (X+Y,Y)$ et $\psi(x,y) = (u,v) = (x+y,y)$. Alors ψ est une bijection de $\mathbb{R}^2$ sur lui-même et $\psi^{-1}(u,v) = (x,y) = (u-v,v)$.

$J(\psi^{-1})(u,v) = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1$, donc $f_{U,V}(u,v) = f_{X,Y}(u-v,v)$ et

$$f_U(u) = \int_{-\infty}^{+\infty} f_{U,V}(u,v)\,dv = \int_{-\infty}^{+\infty} f_{X,Y}(u-v,v)\,dv.$$

De plus, si X et Y sont indépendantes, $f_U(u) = \int_{-\infty}^{+\infty} f_X(u-v)f_Y(v)dv.$ □

Remarque : Si on prend $(U,V) = (X,X+Y)$, on obtient

$$f_{X+Y}(v) = \int_{-\infty}^{+\infty} f_{X,Y}(u,v-u)\,du.$$

Exemple : Soit (X,Y) un couple de densité f définie par $f(x,y) = k(x^2+y^2)\,\mathbb{1}_{[-1,1]^2}(x,y)$. Déterminer k et calculer $\mathrm{cov}(X,Y)$. Les v.a.r. X et Y sont-elles indépendantes ? Déterminer la loi de $X+Y$.

Réponse :

- $k = \dfrac{3}{8}$, $f_X(x) = \dfrac{3}{4}(x^2+1)\,\mathbb{1}_{[-1,1]}(x) = f_Y(x)$.
- $\mathrm{cov}(X,Y) = 0$ mais X et Y ne sont pas indépendantes.
- $f_{X+Y}(u) = \dfrac{1}{4}\left(1+(1-|u|)^3\right)\,\mathbb{1}_{[-2,2]}(u)$.

Application aux lois normales :

Théorème 7.3 : Si X et Y sont deux v.a.r. indépendantes de lois respectives $\mathcal{N}(m_1,\sigma_1^2)$ et $\mathcal{N}(m_2,\sigma_2^2)$, alors la v.a.r. $X+Y$ suit la loi normale $\mathcal{N}(m_1+m_2,\sigma_1^2+\sigma_2^2)$.

Démonstration. On a $f_{X+Y}(u) = \dfrac{1}{2\pi\sigma_1\sigma_2}\displaystyle\int_{-\infty}^{+\infty}\exp\left(-\frac{1}{2}\left(\frac{(u-v-m_1)^2}{\sigma_1^2}+\frac{(v-m_2)^2}{\sigma_2^2}\right)\right)dv$.

On va mettre $\dfrac{(u-v-m_1)^2}{\sigma_1^2}+\dfrac{(v-m_2)^2}{\sigma_2^2}$ sous la forme $\dfrac{(v-M)^2}{A}+B$, où A, B et M sont dépendantes de u. On aura alors :

$$
\begin{aligned}
f_{X+Y}(u) &= \frac{1}{2\pi\sigma_1\sigma_2}\exp\left(-\frac{1}{2}B\right)\int_{-\infty}^{+\infty}\exp\left(-\frac{1}{2}\left(\frac{(v-M)^2}{A}\right)\right)dv \\
&= \frac{1}{2\pi\sigma_1\sigma_2}\exp\left(-\frac{1}{2}B\right)\sqrt{2\pi A}.
\end{aligned}
$$

Il reste donc à déterminer A, B et M.

$$
\begin{aligned}
\frac{(u-v-m_1)^2}{\sigma_1^2}+\frac{(v-m_2)^2}{\sigma_2^2} &= \frac{1}{\sigma_1^2\sigma_2^2}\left(\sigma_2^2(v-(u-m_1))^2+\sigma_1^2(v-m_2)^2\right) \\
&= \frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}\left(v^2-2v\frac{(u-m_1)\sigma_2^2+m_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}+\frac{(u-m_1)^2\sigma_2^2+m_2^2\sigma_1^2}{\sigma_1^2+\sigma_2^2}\right)
\end{aligned}
$$

Or, on a :

$$
\begin{aligned}
&\frac{(u-m_1)^2\sigma_2^2+m_2^2\sigma_1^2}{\sigma_1^2+\sigma_2^2}-\left(\frac{(u-m_1)\sigma_2^2+m_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}\right)^2 \\
=\;& \frac{((u-m_1)^2\sigma_2^2+m_2^2\sigma_1^2)(\sigma_1^2+\sigma_2^2)-((u-m_1)\sigma_2^2+m_2\sigma_1^2)^2}{(\sigma_1^2+\sigma_2^2)^2} \\
=\;& \frac{(u-m_1)^2\sigma_1^2\sigma_2^2+m_2^2\sigma_1^2\sigma_2^2-2m_2(u-m_1)\sigma_1^2\sigma_2^2}{(\sigma_1^2+\sigma_2^2)^2} \\
=\;& \frac{\sigma_1^2\sigma_2^2(u-m_1-m_2)^2}{(\sigma_1^2+\sigma_2^2)^2}
\end{aligned}
$$

donc $A = \dfrac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}$ et $B = \dfrac{(u-m_1-m_2)^2}{\sigma_1^2+\sigma_2^2}$ et, comme $f_{X+Y}(u) = \dfrac{\sqrt{A}}{\sqrt{2\pi}\sigma_1\sigma_2}\exp\left(-\dfrac{1}{2}B\right)$,

$$
f_{X_1+X_2}(u) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}}\exp\left(-\frac{1}{2}\frac{(u-m_1-m_2)^2}{\sigma_1^2+\sigma_2^2}\right).
$$

On reconnait là la loi normale $\mathcal{N}(m_1+m_2,\sigma_1^2+\sigma_2^2)$. □

Chapitre 8

COMPLÉMENTS SUR LE CONDITIONNEMENT

8.1 Complément sur les lois conditionnelles

Dans les chapitres précédents, on a défini la loi conditionnelle dans le cas d'un couple de v.a.r. discret (chapitre 6) et dans le cas d'un couple de v.a.r. absolument continu (chapitre 7). On va donner ici rapidement les analogues dans le cas où l'une des deux v.a.r. est discrète et l'autre absolument continue.

8.1.1 Loi d'une variable absolument continue Y conditionnée par une variable discrète X

On considére des couples aléatoires mixtes dont l'une des composantes X est discrète tandis que l'autre Y est continue.

Définition 8.1 : Un couple aléatoire (X, Y) défini sur $D \times \mathbb{R}$ où D est un espace au plus dénombrable est dit <u>*mixte*</u> si sa loi $P_{X,Y}$ est définie par :

$$\text{pour tout } (x_k, B) \in D \times \mathcal{B}_{\mathbb{R}},\; P_{X,Y}(\{x_k\}, B) = \int_B f_{X,Y}(x_k, y)\, dy$$

où $f_{X,Y}$ est une fonction borélienne positive vérifiant $\displaystyle\sum_k \int_{\mathbb{R}} f_{X,Y}(x_k, y)\, dy = 1$.

Les lois marginales sont ici :

$$f_Y(y) = \sum_k f_{X,Y}(x_k, y) \text{ et } P_X(\{x_k\}) = \int_{\mathbb{R}} f_{X,Y}(x_k, y)\, dy.$$

Le support de (X, Y) est un ensemble d'intervalles portés par les droites parallèles d'abscisses $(x_k)_k$.

<u>*Exemple :*</u> Soit le couple (X, Y) défini sur $\mathbb{N} \times \mathbb{R}_+$ pour lequel :

$$f_{X,Y}(k, y) = e^{-2y} \frac{y^k}{k!}.$$

Il est facile de montrer que : $\displaystyle\int_{\mathbb{R}_+} \sum_k f_{X,Y}(k,y)\,dy = 1.$

Rappel : $B(u,v) = \displaystyle\int_0^1 t^{u-1}(1-t)^{v-1}\,dt = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}$ où $\Gamma(z) = \displaystyle\int_0^{+\infty} e^{-t}\,t^{z-1}\,dt$ et $\Gamma(n) = (n-1)!$ si $n \in \mathbb{N}^*$.

$f_Y(y) = \displaystyle\sum_{k=0}^{+\infty} f_{X,Y}(k,y) = \sum_{k=0}^{+\infty} e^{-2y}\frac{y^k}{k!}\,\mathbb{1}_{[0,+\infty[}(y) = e^{-y}\,\mathbb{1}_{[0,+\infty[}(y)$: Y est une v.a. exponentielle $\mathcal{E}(1)$.

$P_X(\{k\}) = \displaystyle\int_{\mathbb{R}_+} f_{X,Y}(k,y)\,dy = \int_0^{+\infty} e^{-2y}\frac{y^k}{k!}\,dy = \left(\frac{1}{2}\right)^{k+1}$; X est de loi géométrique $\mathcal{G})$ sur $\mathbb{N}$de paramètre $\frac{1}{2}$.

Théorème 8.1 : Étant donné le couple mixte (X,Y), la densité conditionnelle de Y sachant $[X = x_k]$ est définie par :

$$f_Y^{[X=x_k]}(y) = \frac{f_{X,Y}(x_k,y)}{\int_{\mathbb{R}} f_{X,Y}(x_k,y)\,dy}.$$

Démonstration. On recherche la loi conditionnelle de Y sachant X sous la forme d'une densité g vérifiant :

$$\text{pour tout } k \in \mathbb{N} \text{ et tout borélien } B,\ P^{[X=x_k]}([Y \in B]) = \int_B g(x_k,y)\,dy \qquad (*)$$

$$P^{[X=x_k]}([Y \in B]) = \frac{P([X = x_k] \cap [Y \in B])}{P([X = x_k])} = \frac{\int_B f_{X,Y}(x_k,y)\,dy}{\int_{\mathbb{R}} f_{X,Y}(x_k,y)\,dy} = \int_B \frac{f_{X,Y}(x_k,y)\,dy}{\int_{\mathbb{R}} f_{X,Y}(x_k,y)\,dy} \qquad (**)$$

L'égalité de $(*)$ et $(**)$, pour tout B et k, permet de déduire :

$$g(x_k,y) = \frac{f_{X,Y}(x_k,y)}{\int_{\mathbb{R}} f_{X,Y}(x_k,y)\,dy} =_{(notée)} f_Y^{[X=x_k]}(y).$$

□

8.1.2 Loi d'une variable discrète conditionnée par une variable absolument continue

On se place sous les mêmes hypothèses qu'en 8.1.1.

Théorème 8.2 : La loi de probabilité conditionnelle de la v.a.r. discrète X sachant $[Y = y]$ est définie, pour tout y tel que $f_Y(y) > 0$ par :

$$\text{pour tout } x_k,\ P^{Y=y}([X = x_k]) = \frac{f_{X,Y}(x_k,y)}{f_Y(y)} = \frac{f_{X,Y}(x_k,y)}{\sum_k f_{X,Y}(x_k,y)}.$$

Démonstration. On va donner ici une ébauche de preuve plus intuitive que la démonstration basée sur le théorème d'identification fonctionnelle : on suppose que $f_Y(y) > 0$ pour tout y ; on a alors

$$P^{[Y\in[y,y+\Delta y]]}([X = x_k]) = \frac{P([X = x_k] \cap [Y \in [y, y+\Delta y]])}{f_Y(y)\Delta y}$$

égale d'aprés l'hypothèse à :

$$\frac{f_{X,Y}(x_k,y)\,\Delta y}{f_Y(y)\,\Delta y} = \frac{f_{X,Y}(x_k,y)}{f_Y(y)}.$$

$\square$

Exercice : Soit la v.a.r. X de loi $\mathcal{B}(n,Y)$ où le paramètre Y est une v.a.r. $\mathcal{U}([0,1])$.

$$P^{Y=y}([X=k]) = C_n^k y^k (1-y)^{n-k} \text{ pour } 0 \leqslant k \leqslant n$$

Déterminer les lois de X et de Y conditionnée par X.

---

réponse : On a, si $0 \leqslant k \leqslant n$, $f_{X,Y}(k,y) = P^{Y=y}([X=k])f_Y(y) = C_n^k y^k (1-y)^{n-k}\, \mathbb{1}_{[0,1]}(y)$

donc $P([X=k]) = \displaystyle\int f_{X,Y}(k,y)\,dy = C_n^k B(k+1, n-k+1) = C_n^k \frac{k!(n-k)!}{(n+1)!} = \frac{1}{n+1}$; la loi de X est donc l'équiprobabilité sur $\{0,1,\cdots,n\}$.

On a alors $f_Y^{[X=k]}(y) = \dfrac{f_{X,Y}(k,y)}{P([X=k])} = (n+1)\, C_n^k\, y^k (1-y)^{n-k}\, \mathbb{1}_{[0,1]}(y)$.

8.2 Compléments sur l'espérance conditionnelle

Dans cette partie, on va donner une autre approche de l'espérance conditionnelle, très utile dans le cas des variables ayant des moments d'ordre 2.

Lorsqu'une v.a.r. Y dépend d'une v.a.r. X, on détermine une v.a.r. $\Phi(X)$ qui donne le maximum d'information sur Y sans qu'on ait à l'observer.

Définition 8.2 : L'*espérance conditionnelle* de Y sachant X, notée $\mathbb{E}^X(Y)$, est la meilleure approximation quadratique que l'on puisse avoir de Y par une v.a.r. de la forme $\Phi(X)$:

$$\Phi(X) = \mathbb{E}^X(Y) \text{ minimise } \mathbb{E}((Y - \Phi(X))^2)$$

Le cadre de la géométrie hilbertienne est particulièrement adapté à la modélisation et à la compréhension des propriétés de l'espérance conditionnelle. Considérons l'espace $L^2(\Omega,\mathcal{A},P)$ des variables aléatoires X de carré intégrable (*i.e* : $\mathbb{E}(X^2)$ fini) et définies sur $(\Omega,\mathcal{A},P)$, dans lequel on définit le produit scalaire $\langle X,Y\rangle = \mathbb{E}(XY)$ des v.a.r. X et Y. En analyse hilbertienne, on définit la norme quadratique $\|X\|_2 = \sqrt{\mathbb{E}(X^2)}$, pour laquelle l'espace $L^2(\Omega,\mathcal{A},P)$ est un espace de Hilbert. Le coefficient de corrélation ρ de X et Y, égal à $\text{cov}(X,Y)/\sigma_X\sigma_Y$, s'identifie au cosinus de l'angle déterminé par les v.a.r. X et Y ; une v.a. X est donc orthogonale à la v.a. Y si $\text{cov}(X,Y)=0$ (c'est-à-dire si X et Y sont décorrélées).

Le théorème suivant est la caractérisation hilbertienne de l'espérance conditionnelle.

Théorème 8.3 : Étant données deux v.a. quelconques X et Y de $L^2(\Omega,\mathcal{A},P)$:

$$(1)\ \mathbb{E}^X(Y) \text{ espérance conditionnelle de } Y \text{ sachant } X$$

$$\Updownarrow$$

$$(2)\ \mathbb{E}^X(Y) \text{ est la projection orthogonale de } Y \text{ sur } L^2(\Omega,\mathcal{B}_X,P)).$$

Propriétés 8.1 :
(1) Linéarité : $\mathbb{E}^Z(aX+bY) = a\mathbb{E}^Z(X) + b\mathbb{E}^Z(Y)$ avec a et b réels, X, Y, Z des v.a.r. quelconques.
(2) Croissance : si $Y \geqslant X$ alors pour toute v.a.r. Z, $\mathbb{E}^Z(Y) \geqslant \mathbb{E}^Z(X)$
(3) Si Y est indépendante de X, alors $\mathbb{E}^X(Y) = \mathbb{E}(Y)$
(4) Pour toute v.a.r. Y intégrable et toute fonction ϕ bornée :

$$\mathbb{E}^X(\phi(X)Y) = \phi(X)\mathbb{E}^X(Y)$$

(5) Propriété de l'espérance totale : pour tout couple aléatoire (X, Y) :

$$\mathbb{E}(\mathbb{E}^X(Y)) = \mathbb{E}(Y)$$

(6) $\mathbb{E}(\phi(X)Y) = \mathbb{E}(\phi(X)\mathbb{E}^X(Y))$ qui généralise la propriété précédente.

(7) Majoration de la variance :

$$\mathrm{var}(\mathbb{E}^X(Y)) \leqslant \mathrm{var}(Y)$$

(Si X conditionne Y, toute observation de X apporte une information sur Y et en réduit donc la variance).

Démonstration. (1) (2) sont triviales.

(3) La v.a.r. $\mathbb{E}^X(Y)$ est de la forme $\varphi(X)$, laquelle est indépendante de Y : elle est nécessairement constante et égale à $\mathbb{E}(Y)$ (*cf* théorème précédent).

(4) Pour toute fonction-test ψ, $P_{X,Y}$-intégrable,

$$\mathbb{E}^{X=x}(\psi(X, Y)) = E^{X=x}(\psi(x, Y))\ ;$$

prenons $\psi(x, Y) = \varphi(x)Y$. Alors, pour tout x,

$$\mathbb{E}^{X=x}(\varphi(X)Y) = \mathbb{E}^{X=x}(\varphi(x)Y) = \varphi(x)\mathbb{E}^{X=x}(Y),$$

d'où l'égalité attendue.

(5) Démonstration géométrique : $\mathbb{E}^X(Y)$ est la projection orthogonale de Y sur le sous-espace L^2_X, que l'on projette sur le sous-espace des v.a.r. constantes pour obtenir $\mathbb{E}(\mathbb{E}^X(Y))$ (théorème précédent) ; le théorème des trois perpendiculaires de la géométrie hilbertienne ou euclidienne permet de conclure.

(6)

$$\begin{aligned}
\mathbb{E}(\phi(X)Y) &= \mathbb{E}(\mathbb{E}^X(\phi(X)Y)) = \int_{\mathbb{R}} \mathbb{E}^{[X=x]}(\phi(x)Y) f_X(x)\,dx \\
&= \int_{\mathbb{R}} \phi(x)\mathbb{E}^{[X=x]}(Y) f_X(x)\,dx = \mathbb{E}(\phi(X)\mathbb{E}^X(Y))
\end{aligned}$$

(7) Démonstration qui s'apparente à (5) : il suffit de considérer dans L^2, le triangle rectangle de côtés Y, $\mathbb{E}^X(Y)$ et $Y - \mathbb{E}^X(Y)$, et de lui appliquer le théorème de Pythagore : l'inégalité est alors évidente. □

Attention ! L'indépendance de deux v.a.r. n'implique pas nécessairement l'indépendance de leur espérance conditionnelle. Montrons cela sur un contre-exemple classique :

Soit (X, Y) un vecteur gaussien centré, réduit et de covariance nulle ; on définit $S = X + Y$ et $D = X - Y$.

(S, D) est un vecteur gaussien centré, de covariance $\mathbb{E}((X+Y)(X-Y))$ nulle, donc de composantes indépendantes et pourtant $\mathbb{E}^X(S) = X$ et $\mathbb{E}^X(D) = X$, donc S et D ne sont pas conditionnellement indépendantes par rapport à X.

Le théorème suivant établit un résultat pressenti dans les chapitres précédents : $\mathbb{E}(X)$ est la meilleure approximation (quadratique) de X par une constante.

Théorème 8.4 : Pour toute v.a.r. X de $L^2(\Omega, \mathcal{A}, P)$, $\mathbb{E}(X)$ s'identifie à la projection de X sur le sous-espace des v.a.r. constantes.

Démonstration. D'aprés le (2) du théorème de caractérisation de l'espérance conditionnelle, il s'agit de montrer que $\mathbb{E}(X)$ minimise $\mathbb{E}((X-c)^2)$; or $\mathbb{E}((X-c)^2) = \mathrm{var}(X) + (\mathbb{E}(X)-c)^2$ est minimale pour $c = \mathbb{E}(X)$. □

Régression linéaire de Y par rapport à X

Théorème 8.5 : La meilleure approximation quadratique de Y par une fonction affine de X (supposée de variance non nulle) est la variable aléatoire $\hat{Y}$ définie par :

$$\hat{Y} = \left(\rho_{X,Y}\frac{\sigma_Y}{\sigma_X}\right) X + \left(\mathbb{E}(Y) - \rho_{X,Y}\frac{\sigma_Y}{\sigma_X}\,\mathbb{E}(X)\right)$$

La *droite de régression* de Y par rapport à X, est définie par l'équation :

$$y = ax + b \text{ où } a = \rho_{X,Y}\,\frac{\sigma_Y}{\sigma_X} \text{ et } b = \mathbb{E}(Y) - a\mathbb{E}(X).$$

Démonstration. Il s'agit de minimiser $\phi(a,b) = \mathbb{E}\left((Y-(aX+b))^2\right)_{(a,b)\in\mathbb{R}^2}$.

La fonction $\phi(a,b)$ est convexe ; son minimum est atteint pour les valeurs de a et b qui annulent ses dérivées partielles :

$$\left\{\begin{array}{ll} \dfrac{\partial\phi}{\partial a} = 0 & \Leftrightarrow \mathbb{E}[X(Y-(aX+b))] = 0 \\ \dfrac{\partial\phi}{\partial b} = 0 & \Leftrightarrow \mathbb{E}[Y-(aX+b)] = 0 \end{array}\right. \longrightarrow \left\{\begin{array}{ll} \mathbb{E}(X^2)a + \mathbb{E}(X)b & = \mathbb{E}(XY) \\ \mathbb{E}(X)a + b & = \mathbb{E}(Y) \end{array}\right.$$

d'où, sachant que $\sigma_X \neq 0$: $\left\{\begin{array}{l} a = \dfrac{\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}{\sigma_X^2} = \rho_{X,Y}\,\dfrac{\sigma_Y}{\sigma_X} \\ b = \mathbb{E}(Y) - a\mathbb{E}(X) \end{array}\right.$. □

De façon symétrique, on détermine la meilleure approximation affine de X par Y :

$$\hat{X} = \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}\,Y + \mathbb{E}(X) - \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}\,\mathbb{E}(Y)$$

Dans ce cas l'équation de la régression est : $x = \alpha y + \beta$, avec $\alpha = \rho_{X,Y}\dfrac{\sigma_X}{\sigma_Y}$ et $\beta = \mathbb{E}(X) - \alpha\mathbb{E}(Y)$.

Dans le cadre de la théorie de la régression linéaire, le couple de v.a.r. (X,Y) fait l'objet d'observations donnant lieu à un échantillon $(x_i, y_i)_{i=1,\ldots,n}$, à partir duquel on construit une estimation de la pente a et de l'ordonnée à l'origine b.

Remarques :

(a) Les droites $y = ax+b$ et $x = \alpha y + \beta$ ne sont généralement pas confondues, sauf dans le cas où $\rho_{X,Y} = \pm 1$.

(b) La relation fonctionnelle entre Y et X peut évidemment relever d'un modèle non affine, par rapport aux paramètres inconnus, de forme analytique quelconque (polynomiale, exponentielle ou autre). Bien que de structure identique, le problème qualifié alors de régression non-linéaire, plus délicat à résoudre, nécessite la mise en oeuvre de méthodes numériques.

(c) Le problème de régression linéaire est généralisable au cas d'une v.a.r. Y conditionnée par un ensemble de n v.a.r. $X_1, X_2, \ldots, X_n$; il sera alors question de régression multiple ou multilinéaire.

On rassemble ici les deux types d'approximation quadratique d'une v.a.r. Y en fonction d'une v.a.r. X :

Théorème 8.6 : Si (X, Y) est un vecteur aléatoire de carrés intégrables, les propriétés suivantes sont vérifiées :
(1) La meilleure approximation quadratique de Y par une fonction affine dépendant de X est la v.a. $\hat{Y}$:

$$\hat{Y} = \left(\rho_{X,Y}\frac{\sigma_Y}{\sigma_X}\right) X + \left(\mathbb{E}(Y) - \rho_{X,Y}\frac{\sigma_Y}{\sigma_X}\,\mathbb{E}(X)\right)$$

(2) $\mathbb{E}^X(Y)$ est la meilleure approximation quadratique de Y par une fonction quelconque $\varphi(X)$:

$$\text{Pour toute fonction } \varphi : \ \mathbb{E}((Y - \mathbb{E}^X(Y))^2) \leqslant \mathbb{E}((Y - \varphi(X))^2)$$

8.3 Variance conditionnelle

Définition 8.3 : Étant données deux v.a.r. X et Y, la <u>*variance conditionnelle*</u> de Y sachant X est la variable aléatoire :

$$\mathbb{E}^X\left((Y - \mathbb{E}^X(Y))^2\right) =_{(\textit{noté})} \mathrm{var}^X(Y).$$

$\mathrm{var}^{X=x}(Y)$ exprime la dispersion quadratique de Y autour de l'espérance conditionnelle $E^{X=x}(Y)$.

Théorème 8.7 : (Propriété de la variance totale)
Pour tout couple de variables aléatoires (X, Y) :

$$\mathrm{var}(Y) = \mathbb{E}(\mathrm{var}^X(Y)) + \mathrm{var}(\mathbb{E}^X(Y))$$

Dans le cadre de la géométrie hilbertienne de $L^2(\Omega, \mathcal{A}, P)$, la propriété de la variance totale équivaut au théorème de Pythagore appliqué au triangle rectangle de côtés Y, $\mathbb{E}(Y)$, $\mathbb{E}^X(Y)$.

Démonstration.

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

Le dernier terme est égal à $\mathrm{var}(\mathbb{E}^X(Y))$, le premier terme est égal à $\mathbb{E}(\mathrm{var}^X(Y))$; quant au terme médian, il est nul. □

<u>*Exemple*</u> : Variance de la taille d'une population constituée d'un nombre aléatoire de sous-populations.

Soit un nombre aléatoire N de populations de tailles décrites par les variables aléatoires $(X_i)_i$ indépendantes entre elles, de même loi que X et indépendantes de N : $S = \sum_{i=1}^{N} X_i$ étant la taille de la population totale.

On établit :

$$\mathrm{var}(S) = \mathbb{E}(N)\mathrm{var}(X) + \mathrm{var}(N)(\mathbb{E}(X))^2.$$

Chapitre 9

CONVERGENCE DE SUITES DE VARIABLES ALÉATOIRES

Les variables aléatoires X_n, X utilisées dans ce chapitre sont toutes définies sur le même espace probabilisé $(\Omega, \mathcal{A}, P)$, à valeurs dans $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, de fonctions de répartition respectives F_n et F. Dans ce chapitre, on va aborder les notions de convergence en moyenne, en moyenne quadratique et en probabilité mais c'est surtout la convergence en loi qui retiendra notre attention, ainsi que les questions d'approximations de lois.

9.1 Inégalités

Propriété 9.1 : (Inégalité de Markov)
Soit ε et α deux réels strictement positifs. Si X admet un moment d'ordre α ($\mathbb{E}(|X|^r)$ existe et est fini), alors

$$P([|X| \geqslant \varepsilon]) \leqslant \frac{1}{\varepsilon^\alpha}\mathbb{E}(|X|^\alpha).$$

Démonstration. On a :

$$|X|^\alpha = |X|^\alpha\, \mathbb{1}_{[|X|\geqslant\varepsilon]} + |X|^\alpha\, \mathbb{1}_{[|X|<\varepsilon]}.$$

Il est clair que, puisque $|X|^\alpha$ possède une espérance, les v.a.r. $|X|^\alpha\, \mathbb{1}_{[|X|\geqslant\varepsilon]}$ et $|X|^\alpha\, \mathbb{1}_{[|X|<\varepsilon]}$ possèdent également une espérance. On a, par ailleurs :

$$|X|^\alpha\, \mathbb{1}_{[|X|\geqslant\varepsilon]} \geqslant \varepsilon^\alpha\, \mathbb{1}_{[|X|\geqslant\varepsilon]}.$$

Cette inégalité implique que :

$$\begin{array}{rcl}\mathbb{E}(|X|^\alpha) & = & \mathbb{E}\left(|X|^\alpha\, \mathbb{1}_{[|X|\geqslant\varepsilon]}\right) + \mathbb{E}\left(|X|^\alpha\, \mathbb{1}_{[|X|>\varepsilon]}\right) \\ & \geqslant & \mathbb{E}\left(|X|^\alpha\, \mathbb{1}_{[|X|\geqslant\varepsilon]}\right) \geqslant \varepsilon^\alpha\, \mathbb{E}\left(\mathbb{1}_{[|X|\geqslant\varepsilon]}\right) = \varepsilon^\alpha\, P\left([|X| \geqslant \varepsilon]\right)\end{array}$$

ce qui donne le résultat. □

Corollaire 9.1 : (Inégalité de Bienaymé-Tchebicheff)
Soit ε un réel strictement positif. Si X admet un moment d'ordre 2, alors

$$P([|X - \mathbb{E}(X)|| \geqslant \varepsilon]) \leqslant \frac{1}{\varepsilon^2}\mathrm{var}(X).$$

Démonstration. On a $\text{var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) < +\infty$, c'est-à-dire la v.a.r. $Y = X - \mathbb{E}(X)$ admet un moment d'ordre 2. Il suffit d'appliquer l'inégalité de Markov à Y avec $\alpha = 2$. □

On termine cette partie par deux autres inégalités classiques :

Inégalité de Jensen : Soit X une v.a.r. et φ une fonction convexe de $\mathbb{R}$ dans $\mathbb{R}$. Alors, si X et $\varphi(X)$ possèdent une espérance, on a :

$$\varphi(\mathbb{E}(X)) \leqslant \mathbb{E}(\varphi(X)).$$

Inégalité de Cauchy-Schwarz : Si X et Y sont des v.a.r. ayant un moment d'ordre 2, alors :

$$\mathbb{E}(|XY|) \leqslant \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}.$$

9.2 Convergence en moyenne et en moyenne quadratique

Définition 9.1 : Soit $(X_n)_{n\in\mathbb{N}}$ et X des variables aléatoires réelles admettant une espérance. On dit que $(X_n)_{n\in\mathbb{N}}$ *converge en moyenne* vers X si et seulement si $\lim\limits_{n\to+\infty} \mathbb{E}(|X_n - X|) = 0$. On note alors $X_n \overset{M.}{\to} X$.

Définition 9.2 : Soit $(X_n)_{n\in\mathbb{N}}$ et X des variables aléatoires réelles admettant un moment d'ordre 2. On dit que $(X_n)_{n\in\mathbb{N}}$ *converge en moyenne quadratique* vers X si et seulement si $\lim\limits_{n\to+\infty} \mathbb{E}((X_n - X)^2) = 0$. On note alors $X_n \overset{M.Q.}{\to} X$.

On va maintenant énoncer quelques propriétés relatives à ces modes de convergence.

Propriétés 9.2 : Soit $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ sont deux suites de v.a.r. et X et Y deux v.a.r.
1) si $X_n \overset{M.}{\to} X$ et $Y_n \overset{M.}{\to} Y$, alors, pour tout $(\lambda, \mu) \in \mathbb{R}^2$, $(\lambda X_n + \mu Y_n) \overset{M.}{\to} \lambda X + \mu Y$.
2) si $X_n \overset{M.Q.}{\to} X$ et $Y_n \overset{M.Q.}{\to} Y$, alors, pour tout $(\lambda, \mu) \in \mathbb{R}^2$, $(\lambda X_n + \mu Y_n) \overset{M.Q.}{\to} \lambda X + \mu Y$.
3) si $X_n \overset{M.Q.}{\to} X$, alors $X_n \overset{M.}{\to} X$.

Démonstration. 1) On a :

$$0 \leqslant \mathbb{E}(|\lambda X_n + \mu Y_n - (\lambda X + \mu Y)|) \leqslant |\lambda|\mathbb{E}(|X_n - X|) + |\mu|\mathbb{E}(|Y_n - Y|) \underset{n\to+\infty}{\to} 0$$

d'où le résultat.

2) On a

$$0 \leqslant \mathbb{E}((\lambda X_n + \mu Y_n - (\lambda X + \mu Y))^2) \leqslant 2\left(\lambda^2\mathbb{E}((X_n - X)^2) + \mu^2\mathbb{E}((Y_n - Y)^2)\right) \underset{n\to+\infty}{\to} 0$$

3) L'inégalité de Cauchy-Schwarz donne :

$$\mathbb{E}(|X_n - X|) \leqslant \sqrt{\mathbb{E}((X_n - X)^2)} \underset{n\to+\infty}{\to} 0.$$

□

Propriétés 9.3 : Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a.r. admettant une variance. Si $\mathbb{E}(X_n) \underset{n\to+\infty}{\to} \mu$ et $\text{var}(X_n) \underset{n\to+\infty}{\to} 0$, alors $X_n \overset{M.Q.}{\to} \mu$.

Démonstration. On a

$$\begin{aligned}\mathbb{E}((X_n-\mu)^2 &= \mathbb{E}\left((X_n-\mathbb{E}(X_n)+\mathbb{E}(X_n)-\mu)^2\right)\\ &= \mathbb{E}((X_n-\mathbb{E}(X_n))^2)+\mathbb{E}((\mathbb{E}(X_n)-\mu)^2)=\mathrm{var}(X_n)+\mathbb{E}((\mathbb{E}(X_n-\mu))^2)\underset{n\to+\infty}{\to}0\end{aligned}$$

□

Propriétés 9.4 : Soit $(X_n)_{n\in\mathbb{N}}$ et X des v.a.r.
1) Si $X_n \overset{M.}{\to} X$, alors $\lim\limits_{n\to+\infty}\mathbb{E}(X_n)=\mathbb{E}(X)$.
2) Si $X_n \overset{M.Q.}{\to} X$, alors $\lim\limits_{n\to+\infty}\mathbb{E}(X_n)=\mathbb{E}(X)$ et $\lim\limits_{n\to+\infty}\mathrm{var}(X_n)=\mathrm{var}(X)$.

Démonstration. Les résultats se déduisent directement de l'inégalité de Jensen appliquée à la fonction convexe $\varphi(x)=|x|$:

$$|\mathbb{E}(X_n)-\mathbb{E}(X)|=|\mathbb{E}(X_n-X)|\leqslant\mathbb{E}(|X_n-X|)$$

et d'autre part de l'inégalité :

$$\left|\sqrt{\mathbb{E}(X_n^2)}-\sqrt{\mathbb{E}(X^2)}\right|\leqslant\sqrt{\mathbb{E}((X_n-X)^2)}.$$

□

9.3 Convergence en probabilité

Définition 9.3 : Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a.r. et soit X une v.a.r. On dit que $(X_n)_{n\in\mathbb{N}}$ *converge en probabilité* vers X si et seulement si, pour tout $\varepsilon>0$, $\lim\limits_{n\to+\infty}P([|X_n-X|<\varepsilon])=1$.
On note alors $X_n\overset{P}{\to}X$.

On a alors les propriétés suivantes :

Propriété 9.5 : Soit $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ deux suites de v.a.r. et X et Y deux v.a.r. Si $X_n\overset{P}{\to}X$ et $Y_n\overset{P}{\to}Y$ et si f est une fonction continue de $\mathbb{R}$ dans $\mathbb{R}$ et g une fonction continue de $\mathbb{R}^2$ dans $\mathbb{R}$, alors :
1) $f(X_n)\overset{P}{\to}f(X)$
2) $g(X_n,Y_n)\overset{P}{\to}g(X,Y)$.

Corollaire : Soit $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ deux suites de v.a.r. et X et Y deux v.a.r. Alors, si $X_n\overset{P}{\to}X$ et $Y_n\overset{P}{\to}Y$, on a :
1) $|X_n|\overset{P}{\to}|X|$
2) Pour tout $(\lambda,\mu)\in\mathbb{R}^2$, $\lambda X_+\mu Y_n\overset{P}{\to}\lambda X+\mu Y$
3) $X_nY_n\overset{P}{\to}XY$.

Théorème 9.1 : Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a.r. admettant un moment d'ordre 2 et X une v.a.r. admettant un moment d'ordre 2. Si $X_n\overset{M.Q.}{\to}X$, alors $X_n\overset{P}{\to}X$.

Démonstration. L'implication découle de l'inégalité de Markov :

$$\text{pour tout } \varepsilon>0,\ P([|X_n-X|>\varepsilon])\leqslant\frac{1}{\varepsilon^2}\mathbb{E}((X_n-X)^2)\underset{n\to+\infty}{\to}0.$$

□

9.4 Convergence en loi

Souvent, en statistique, on ne connait pas les lois : on observe seulement une certaine distribution. Les théorèmes de convergence en loi permettent de justifier certaines approximations de distributions observées par des lois théoriques connues.

Définition 9.4 : On dit que la suite de variables aléatoires (X_n) converge en loi vers X si, pour tout x en lequel F est continue, $\lim\limits_{n\to+\infty} F_n(x) = F(x)$. On écrit $X_n \xrightarrow{L} X$.

Théorème 9.2 : Si X_n suit la loi binomiale $\mathcal{B}\left(n, \dfrac{\lambda}{n}\right)$ et si X suit la loi de Poisson $\mathcal{P}(\lambda)$, alors $X_n \xrightarrow{L} X$.

Démonstration. La fonction de répartition F de X est continue sur $\mathbb{R}\setminus\mathbb{N}$. Pour tout $x \in \mathbb{R}\setminus\mathbb{N}$, on a :

$$F_n(x) = \sum_{k\leqslant x} P(X_n = k) = \sum_{k=0}^{\min(E(x),n)} C_n^k p_n^k (1-p_n)^{n-k}$$

où $E(x)$ désigne la partie entière de x. Pour $n \geqslant E(x)$, on a alors

$$F_n(x) = \sum_{k=0}^{E(x)} \frac{(np_n)^k}{k!} \frac{n(n-1)\cdots(n-k+1)}{n^k}(1-p_n)^{-k}(1-p_n)^n.$$

Or, pour tout $k \in \{0, \cdots, E(x)\}$,

$$\lim_{n\to+\infty} \frac{n(n-1)\cdots(n-k+1)}{n^k}(1-p_n)^{-k} = 1$$

car $p_n \sim \dfrac{\lambda}{n} \to 0$ quand $n \to +\infty$, et

$$\frac{n(n-1)\cdots(n-k+1)}{n^k} = \left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right) \to 1.$$

D'autre part, $\lim\limits_{n\to+\infty}(1-p_n)^n = \lim\limits_{n\to+\infty} e^{n\ln(1-p_n)} = \lim\limits_{n\to+\infty} e^{-np_n} = e^{-\lambda}$ donc finalement

$$\lim_{n\to+\infty} F_n(x) = \sum_{k=0}^{E(x)} \frac{\lambda^k}{k!}e^{-\lambda} = F(x).$$

□

Loi faible des grands nombres : (admise)

Théorème 9.3 : Soit $(X_n)_n$ une suite de variables aléatoires d'ordre deux, deux à deux indépendantes, de même espérance m et de même variance $\sigma^2 > 0$.
Alors, si $S_n = X_1 + \cdots + X_n$ et si $Z_n = \dfrac{S_n}{n}$, on a $Z_n \xrightarrow{L} m$.

9.5 Convergence vers la loi normale

Le **théorème central limite**, partie fondamentale de ce chapitre, sera admis.

Théorème 9.4 : Soit $(X_n)_n$ une suite de v.a.r. indépendantes d'ordre deux, de même loi, d'espérance m et de variance σ^2.
Alors, si $S_n = X_1 + \cdots + X_n$ et si $T_n = \dfrac{S_n - nm}{\sqrt{n}\sigma}$, on a $T_n \xrightarrow{L} T$ de loi normale $\mathcal{N}(0,1)$.

Remarque : $\mathbb{E}(S_n) = nm$, $\mathrm{var}(S_n) = n\sigma^2$ et $T_n = \dfrac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{var}(S_n)}}$.

Application : Quand n est grand, c'est-à-dire en pratique, lorsque $n \geqslant 30$, S_n suit approximativement la loi normale $\mathcal{N}(\mathbb{E}(S_n), \mathrm{var}(S_n))$, c'est-à-dire la loi $\mathcal{N}(nm, n\sigma^2)$ et $Z_n = \dfrac{S_n}{n}$ la loi normale $\mathcal{N}(\mathbb{E}(Z_n), \mathrm{var}(Z_n))$, c'est-à-dire la loi $\mathcal{N}(m, \dfrac{\sigma^2}{n})$.

Correction de continuité : Lorsque l'on approxime la loi d'une v.a.r. entière X à valeurs dans $\mathbb{N}$ par la loi normale $\mathcal{N}(\mathbb{E}(X), \mathrm{var}(X))$, de fonction de répartition F, cela revient à considérer X comme une variable aléatoire qui prend toutes les valeurs réelles, et l'intervalle $]k - 0.5, k + 0.5]$ est l'ensemble de ces valeurs qui "s'arrondissent" à k.
On remplacera donc $P([X = 0])$ par $F(0.5)$ et $P([X = k])$ par $F(k + 0.5) - F(k - 0.5)$.
(Si X est à valeur dans $\{0, \cdots, n\}$, on procède de même, sauf pour $P([X = n])$ que l'on remplace par $1 - F(n - 0.5)$.)

9.6 Approximations

9.6.1 Approximation de la loi hypergéométrique par la loi binomiale

Soit E un ensemble de N éléments, dont une proportion p est de type 1. On effectue dans E une série de n tirages successifs sans remise. Soit X le nombre d'éléments de type 1 obtenus : X suit la loi hypergéométrique $\mathcal{H}(n, Np, N)$.

Théorème 9.5 : Si n et p sont fixés, pour tout $k \in \{0, \cdots, n\}$,

$$\lim_{N \to +\infty} \frac{C_{Np}^k C_{N(1-p)}^{n-k}}{C_N^n} = C_n^k p^k (1-p)^{n-k}.$$

Démonstration.

$$\begin{aligned}
\frac{C_M^k C_{N-M}^{n-k}}{C_N^n} &= C_n^k \frac{M(M-1)\cdots(M-k+1)(N-M)(N-M-1)\cdots(N-M-(n-k)+1)}{N(N-1)\cdots(N-n+1)} \\
&= C_n^k \frac{N^k N^{n-k}}{N^n} \frac{p\left(p-\frac{1}{N}\right)\cdots\left(p-\frac{k-1}{N}\right)(1-p)\left(1-p-\frac{1}{N}\right)\cdots\left(1-p-\frac{n-k-1}{N}\right)}{\left(1-\frac{1}{N}\right)\cdots\left(1-\frac{n-1}{N}\right)}.
\end{aligned}$$

Or, comme $\displaystyle\lim_{N \to +\infty} \frac{1}{N} = 0$, $\displaystyle\lim_{N \to +\infty} \frac{C_{Np}^k C_{N(1-p)}^{n-k}}{C_N^n} = C_n^k p^k (1-p)^{n-k}$. □

Application : Quand N est grand, c'est-à-dire en pratique lorsque $N \geqslant 10n$, on peut approximer la loi hypergéométrique $\mathcal{H}(n, Np, N)$ par la loi binomiale $\mathcal{B}(n, p)$.

(Dans le modèle précédent, ceci revient à considérer un tirage avec remise, car on a peu de chances de tirer 2 fois le même élément).

9.6.2 Approximation de la loi binomiale par la loi de Poisson

Théorème 9.6 : Si $(p_n)_n$ est une suite de réels de $[0, 1]$ telle que $\lim_{n\to+\infty} np_n = \lambda$, alors, pour tout k fixé,

$$\lim_{n\to+\infty} C_n^k p_n^k (1-p_n)^{n-k} = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}.$$

Démonstration. C'est une partie de la démonstration du théorème 8.1. □

Application : En pratique, lorsque $p \leqslant 0.1$, $n \geqslant 30$ et $np < 15$, on peut approximer la loi binomiale $\mathcal{B}(n, p)$ par la loi de Poisson $\mathcal{P}(np)$.

Exemple : Si X suit la loi binomiale $\mathcal{B}(40, 0.03)$, $P([X = 2]) = C_{40}^2 (0.03)^2 (0.97)^{38} \approx 0.2206$.

Or $40 \times 0.03 = 1.2 < 15$, donc $\mathcal{B}(40, 0.03)$ peut être approximée par $\mathcal{P}(1.2)$ et $P([X = 2]) \approx e^{-1.2}\frac{(1.2)^2}{2!} \approx 0.2169$.

9.6.3 Approximation de la loi de Poisson par la loi normale

Théorème 9.7 : Si $(X_n)_n$ est une suite de variables aléatoires indépendantes, de même loi de Poisson $\mathcal{P}(\lambda)$ et si $S_n = X_1 + \cdots + X_n$, alors S_n suit la loi de Poisson $\mathcal{P}(n\lambda)$.

Démonstration. La démonstration se fait par récurrence sur n.

- Pour $n = 1$, X_1 suit bien la loi de Poisson $\mathcal{P}(\lambda)$.
- Si $S_n = X_1 + \cdots + X_n$ suit la loi de Poisson $\mathcal{P}(n\lambda)$, $S_{n+1} = S_n + X_{n+1}$ et, vu que S_n et X_{n+1} sont indépendantes, que S_n suit la loi de Poisson $\mathcal{P}(n\lambda)$ et X_{n+1} la loi de Poisson $\mathcal{P}(\lambda)$, d'après la propriété 6.2.2), $S_n + X_{n+1} = S_{n+1}$ suit la loi de Poisson de paramètre $n\lambda + \lambda = (n+1)\lambda$. □

Application : S_n suit la loi de Poisson $\mathcal{P}(n\lambda)$ donc $\mathbb{E}(S_n) = n\lambda$ et $\mathrm{var}(S_n) = n\lambda$. D'après le théorème central limite, la loi de S_n peut être approximée par la loi normale $\mathcal{N}(\mathbb{E}(S_n), \mathrm{var}(S_n))$, c'est-à-dire $\mathcal{N}(n\lambda, n\lambda)$.

En pratique, lorsque $\lambda \geqslant 15$, on peut approximer la loi de Poisson $\mathcal{P}(\lambda)$ par la loi normale $\mathcal{N}(\lambda, \lambda)$.

Exemple : Si X suit la loi de Poisson $\mathcal{P}(16)$, $P([X = 16]) = e^{-16}\frac{16^{16}}{16!} \approx 0.0992$.

En approximant la loi $\mathcal{P}(16)$ par la loi $\mathcal{N}(16, 16)$ de fonction de répartition F, on obtient :

$$P([X = 16]) \approx F(16.5) - F(15.5) = \Phi\left(\frac{0.5}{4}\right) - \Phi\left(-\frac{0.5}{4}\right) = 2\Phi(0.125) - 1 \approx 0.0995.$$

Le gain de temps est surtout sensible pour le calcul des valeurs de la fonction de répartition :

$$P([X \leqslant 20]) \approx F(20.5) = \Phi\left(\frac{20.5 - 16}{4}\right) = \Phi(1.125) \approx 0.8697$$

(alors qu'en gardant la loi de Poisson, il faudrait faire la somme de 21 termes !)

9.6.4 Approximation de la loi binomiale par la loi normale

Théorème 9.8 : Si $(X_n)_n$ est une suite de variables aléatoires indépendantes de même loi de Bernoulli $\mathcal{B}(p)$ et si $S_n = X_1 + \cdots + X_n$, alors S_n suit la loi binomiale $\mathcal{B}(n, p)$.

Démonstration. La démonstration se fait par récurrence sur n.

- Pour $n = 1$, X_1 suit bien la loi $\mathcal{B}(p)$, c'est-à-dire la loi binomiale $\mathcal{B}(1, p)$.
- Si $S_n = X_1+\cdots+X_n$ suit la loi binomiale $\mathcal{B}(n, p)$, $S_{n+1} = S_n+X_{n+1}$ et, vu que S_n et X_{n+1} sont indépendantes, que S_n suit la loi binomiale $\mathcal{B}(n, p)$ et X_{n+1} la loi binomiale $\mathcal{B}(1, p)$, d'après la propriété 6.2 1), $S_n + X_{n+1} = S_{n+1}$ suit la loi binomiale $\mathcal{B}(n + 1, p)$. □

Application : S_n suit la loi binomiale $\mathcal{B}(n, p)$ donc $\mathbb{E}(S_n) = np$ et $\mathrm{var}(S_n) = np(1-p)$. D'après le théorème central limite, la loi de S_n peut être approximée par la loi normale $\mathcal{N}(\mathbb{E}(S_n), \mathrm{var}(S_n))$, c'est-à-dire la loi $\mathcal{N}(np, np(1 - p))$.

En pratique, lorsque $n \geqslant 30$, $np \geqslant 15$ et $np(1 - p) > 5$, la loi binomiale $\mathcal{B}(n, p)$ peut être approximée par la loi normale $\mathcal{N}(np, np(1 - p))$.

Si X suit la loi $\mathcal{B}(n, p)$, et si F est la fonction de répartition de la loi $\mathcal{N}(np, np(1 - p))$, on remplacera donc $P([X = 0])$ par $F(0.5)$; $P([X = k])$ par $F(k + 0.5) - F(k - 0.5)$ pour $1 \leqslant k \leqslant n - 1$ et $P([X = n])$ par $1 - F(n - 0.5)$.

<u>Exemple</u> : Combien de lancers d'un dé faut-il effectuer pour pouvoir affirmer avec un risque d'erreur inférieur à 5% que la fréquence d'apparition de l'as au cours de ces lancers diffère de $\frac{1}{6}$ d'au plus $\frac{1}{100}$?

Soit S_n le nombre d'as obtenus après n lancers : S_n suit la loi binomiale $\mathcal{B}\left(n, \frac{1}{6}\right)$ et la fréquence d'apparition de l'as au cours de ces lancers est $\frac{S_n}{n}$.

On cherche n tel que $P\left(\left|\frac{S_n}{n} - \frac{1}{6}\right| \geqslant \frac{1}{100}\right) \leqslant \frac{5}{100}$.

Si $n \geqslant 30$, $\frac{n}{6} \geqslant 15$ et $\frac{5n}{36} > 5$, c'est-à-dire $n \geqslant 90$, on peut approximer la loi de $\frac{S_n}{n}$ par la loi normale $\mathcal{N}\left(\frac{1}{6}, \frac{5}{36n}\right)$ et la loi de $\frac{S_n}{n} - \frac{1}{6}$ par la loi normale $\mathcal{N}\left(0, \frac{5}{36n}\right)$. On a alors

$$P\left(\left|\frac{S_n}{n} - \frac{1}{6}\right| \geqslant \frac{1}{100}\right) = P\left(\left|\frac{6\sqrt{n}}{\sqrt{5}}\left(\frac{S_n}{n} - \frac{1}{6}\right)\right| \geqslant \frac{6\sqrt{n}}{\sqrt{5}}\frac{1}{100}\right) \approx 2\left(1 - \Phi\left(\frac{6\sqrt{n}}{\sqrt{5}}\frac{1}{100}\right)\right) \approx 0.05$$

d'après le théorème 3.4. On a donc $\Phi\left(\frac{6\sqrt{n}}{\sqrt{5}}\frac{1}{100}\right) \approx 0.975 \approx \Phi(1.96)$ et

$$n \approx 1.96^2 \times \frac{50000}{36} \approx 5536.$$

On a bien $5536 \geqslant 90$, ce qui légitime l'approximation.

Exercice : Sur une autoroute, la proportion des camions par rapport à l'ensemble des véhicules est 0.07.
1) Soit X le nombre de camions parmi 100 véhicules choisis au hasard.
Calculer $P([X \geqslant 5])$.
2) Soit Y le nombre de camions parmi 1000 véhicules choisis au hasard.
Calculer $P([65 \leqslant Y \leqslant 75])$.
3) On choisit n véhicules au hasard. Pour quelles valeurs de n peut-on affirmer que la proportion de camions est entre 0.06 et 0.08 avec un risque d'erreur inférieur à 5% ?

Solution :
1) Soit X une v.a.r. de loi binomiale $\mathcal{B}(100, 0.07)$.

$100 \geqslant 30$, $100 \times 0.07 = 7 < 15$, $0.07 \leqslant 0.1$ donc l'approximation à utiliser est celle par la loi de Poisson $\mathcal{P}(7)$ et $P([X \geqslant 5]) \approx 1 - e^{-7} \displaystyle\sum_{k=0}^{4} \frac{7^k}{k!} \approx 0.827$.

2) Y suit la loi binomiale $\mathcal{B}(1000, 0.07)$.

$1000 \geqslant 30$, $1000 \times 0.07 = 70 \geqslant 15$, $70 \times 0.93 = 65.1 > 5$ donc l'approximation à utiliser est celle par la loi normale $\mathcal{N}(70, 65.1)$ et si F désigne la fonction de répartition de la loi $\mathcal{N}(70, 65.1)$,

$$\begin{aligned} P([65 \leqslant Y \leqslant 75]) &\approx F(75.5) - F(64.5) = \Phi\left(\frac{5.5}{\sqrt{65.1}}\right) - \Phi\left(-\frac{5.5}{\sqrt{65.1}}\right) \\ &= 2\Phi\left(\frac{5.5}{\sqrt{65.1}}\right) - 1 \approx 2\Phi(0.68) - 1 \approx 0.5 \end{aligned}$$

3) On choisit n véhicules au hasard. Le nombre S_n des camions parmi ces n véhicules suit la loi binomiale $\mathcal{B}(n, 0.07)$ et la proportion des camions est $\dfrac{S_n}{n}$.

On cherche n tel que $P\left(\left|\dfrac{S_n}{n} - 0.07\right| \geqslant 0.01\right) = 0.05$.

Si $n \geqslant 30$, $0.07n \geqslant 15$ et $0.07 \times 0.93 \times n > 5$, c'est-à-dire $n \geqslant 215$, on peut approximer la loi de $\dfrac{S_n}{n}$ par la loi normale $\mathcal{N}\left(0.07, \dfrac{0.0651}{n}\right)$ et la loi de $\dfrac{S_n}{n} - 0.07$ par la loi normale $\mathcal{N}\left(0, \dfrac{0.065}{n}\right)$. On a alors

$$\begin{aligned} P\left(\left|\frac{S_n}{n} - 0.07\right| \geqslant 0.01\right) &= P\left(\left|\frac{\sqrt{n}}{\sqrt{0.0651}}\left(\frac{S_n}{n} - 0.07\right)\right| \geqslant \frac{\sqrt{n}}{\sqrt{0.0651}}\frac{1}{100}\right) \\ &\approx 2\left(1 - \Phi\left(\frac{\sqrt{n}}{\sqrt{651}}\right)\right) \approx 0.05 \end{aligned}$$

d'après le théorème 3.4. On a donc $\Phi\left(\dfrac{\sqrt{n}}{\sqrt{651}}\right) \approx 0.975 \approx \Phi(1.96)$ et $n \approx 1.96^2 \times 651 \approx 2501$.

$2501 \geqslant 90$, ce qui légitime l'approximation.

RAPPELS DES LOIS CLASSIQUES

1- Lois discrètes finies

a) Loi de Bernoulli $\mathcal{B}(p)$ (ou $\mathcal{B}(1,p)$) :

$$P([X=1]) = p\ ;\ P([X=0]) = 1-p$$

$$G_X(t) = pt + 1 - p\ ;\ \mathbb{E}(X) = p\ ;\ \mathrm{var}(X) = p(1-p)$$

b) Loi binomiale $\mathcal{B}(n,p)$:

$$P([X=k]) = C_n^k p^k (1-p)^{n-k} \text{ pour } 0 \leqslant k \leqslant n$$

$$G_X(t) = (pt+1-p)^n\ ;\ \mathbb{E}(X) = np\ ;\ \mathrm{var}(X) = np(1-p)$$

c) Loi hypergéométrique $\mathcal{H}(n,M,N)$:

$$P([X=k]) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \text{ pour } \max(0,\ n-N+M) \leqslant k \leqslant \min(n,\ M)$$

$$\mathbb{E}(X) = \frac{nM}{N}\ ;\ \mathrm{var}(X) = n\frac{M(N-M)(N-n)}{N^2(N-1)}$$

d) Loi équiprobable $\mathcal{U}([\![1,n]\!])$:

$$P([X=k]) = \frac{1}{n} \text{ pour } 1 \leqslant k \leqslant n$$

$$\mathbb{E}(X) = \frac{n+1}{2}\ ;\ \mathrm{var}(X) = \frac{n^2-1}{12}$$

2- Lois discrètes infinies

a) Loi géométrique sur $\mathbb{N}^*$ $\mathcal{G}(p)$ (ou $\mathcal{P}(1,p)$) :

$$P([X=k]) = p(1-p)^{k-1} \text{ pour } k \in \mathbb{N}^*$$

$$G_X(t) = \frac{p}{1-p}\left(\frac{1}{1-t(1-p)} - 1\right)\ ;\ \mathbb{E}(X) = \frac{1}{p}\ ;\ \mathrm{var}(X) = \frac{1-p}{p^2}$$

b) Loi de Pascal $\mathcal{P}(r,p)$:

$$P([X=k]) = C_{k-1}^{r-1} p^r (1-p)^{k-r} \text{ pour } k \geqslant r$$

$$\mathbb{E}(X) = \frac{r}{p}\ ;\ \mathrm{var}(X) = \frac{r(1-p)}{p^2}$$

c) Loi géométrique sur $\mathbb{N}$ $\mathcal{G}_0(p)$ (ou $\mathcal{BN}(1,p)$) :

$$P([X=k]) = p(1-p)^k \text{ pour } k \in \mathbb{N}$$

$$G_X(t) = \left(\frac{p}{1-t(1-p)}\right)\ ;\ \mathbb{E}(X) = \frac{1-p}{p}\ ;\ \mathrm{var}(X) = \frac{1-p}{p^2}$$

d) Loi binomiale négative $\mathcal{BN}(r,p)$:

$$P([X=k]) = C_{k+r-1}^{r-1} p^r (1-p)^k \text{ pour } k \in \mathbb{N}$$

$$G_X(t) = \left(\frac{p}{1-t(1-p)}\right)^r \ ; \ \mathbb{E}(X) = \frac{r(1-p)}{p} \ ; \ \operatorname{var}(X) = \frac{r(1-p)}{p^2}$$

e) de Poisson $\mathcal{P}(\lambda)$:

$$P([X=k]) = \exp(-\lambda)\frac{\lambda^k}{k!} \text{ pour } k \in \mathbb{N}$$

$$G_X(t) = \exp(\lambda(t-1)) \ ; \ \mathbb{E}(X) = \lambda \ ; \ \operatorname{var}(X) = \lambda$$

3- Lois absolument continues

a) Loi uniforme $\mathcal{U}([a,b])$:

$$f(x) = \frac{1}{b-a}\mathbb{1}_{[a,b]}(x) \ ; \ \mathbb{E}(X) = \frac{a+b}{2} \ ; \ \operatorname{var}(X) = \frac{(a-b)^2}{12}$$

b) Loi exponentielle $\mathcal{E}(\lambda)$:

$$f(x) = \lambda \exp(-\lambda x)\mathbb{1}_{]0,+\infty[}(x) \ ; \ \mathbb{E}(X) = \frac{1}{\lambda} \ ; \ \operatorname{var}(X) = \frac{1}{\lambda^2}$$

c) Loi Gamma $\gamma(\lambda, a)$ **(ou** $\Gamma(a,\lambda)$**)** :

$$f(x) = \frac{\lambda^a}{\Gamma(a)}\exp(-\lambda x)x^{a-1}\mathbb{1}_{]0,+\infty[}(x) \ ; \ \mathbb{E}(X) = \frac{a}{\lambda} \ ; \ \operatorname{var}(X) = \frac{a}{\lambda^2}$$

d) Loi Beta $B(a,b)$:

$$f(x) = \frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}\mathbb{1}_{]0,1[}(x) \ ; \ \mathbb{E}(X) = \frac{a}{a+b} \ ; \ \operatorname{var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

e) Loi Normale $\mathcal{N}(m,\sigma^2)$:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \ ; \ \mathbb{E}(X) = m \ ; \ \operatorname{var}(X) = \sigma^2$$

f) Loi de Cauchy $\mathcal{C}(\alpha)$:

$$f(x) = \frac{\alpha}{\pi(\alpha^2+x^2)}$$

Pas d'espérance ni de variance.

On rappelle que $\Gamma(a) = \int_0^{+\infty} \exp(-t)t^{a-1}dt$ et que $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.
On rappelle également que, si E est un sous-ensemble de $\mathbb{R}$, alors $\mathbb{1}_E(x) = 1$ si $x \in E$ et $\mathbb{1}_E(x) = 0$ si $x \notin E$.