

Probabilités

Sommaire

23.1 Événements, probabilités	486
23.1.1 Espace probabilisé	486
23.1.2 Conditionnement et indépendance	489
23.2 Variables aléatoires	492
23.2.1 Variable aléatoire sur Ω	492
23.2.2 Loi et fonction de répartition d'une variable aléatoire	492
23.2.3 Trois lois usuelles	494
23.3 Couples de variables aléatoires	495
23.3.1 Couple de variables aléatoires	495
23.3.2 Loi conjointe, loi marginale, loi conditionnelle	496
23.3.3 Couples de variables aléatoires indépendantes	498
23.3.4 Variables mutuellement indépendantes	499
23.4 Espérance et variance	500
23.4.1 Espérance d'une variable aléatoire réelle	500
23.4.2 Variance, écart-type	503

23.1 Événements, probabilités

23.1.1 Espace probabilisé

Dans tout ce qui suit, on va étudier des situations dans le déroulement desquelles intervient le hasard (on parlera d'« expériences aléatoires ») et à l'issue desquelles on pourra dire si tel ou tel événement particulier s'est, ou ne s'est pas, produit. Pour estimer les « chances » que possède un événement de se produire effectivement (et avant même que l'expérience aléatoire n'ait débuté), on cherchera à lui attribuer une probabilité.

Il est souvent possible (mais pas toujours nécessaire, ni même utile) de décrire l'ensemble de toutes les « issues » de l'expérience (on parle aussi de « résultats élémentaires »). Cet ensemble Ω est appelé « univers » ou « ensemble des possibles ». Avec ce vocabulaire, un événement peut être vu comme une collection particulière d'issues de l'expérience.

Le programme de première année se limite au cas où l'univers Ω est fini.

Notre objectif, dans un premier temps, est de fournir un modèle mathématique qui rende compte du concept d'expérience aléatoire, d'événement au sein de cette expérience, et qui permette de définir précisément ce qu'est la probabilité qu'un tel événement « se produise ».

▷ Terminologie des expériences aléatoires

- On fait le choix d'un ensemble fini Ω , appelé *univers*, et qui décrit de façon assez informelle l'ensemble des *issues possibles* (ou *résultats possibles* ou *réalisations*) de l'expérience aléatoire étudiée. Il n'est souvent ni nécessaire ni même utile d'en donner une définition précise.
- Les éléments A de $\mathcal{P}(\Omega)$ (ce sont donc des parties de Ω) sont appelés *événements*. Les événements sont donc des collections d'issues élémentaires de l'expérience.
- Les singletons $\{\omega\}$, avec ω dans Ω , sont appelés *événements élémentaires*.
- Quand l'expérience aléatoire « se déroule », elle se conclut par une issue ω (un élément de Ω).
Si A est un événement et si cette issue ω est dans A , on dit alors que l'« événement A est réalisé ».
- L'ensemble \emptyset est appelé *événement impossible*, et l'ensemble Ω est appelé *événement certain*.
Si A est un événement, on dit que \bar{A} est l'*événement contraire* de A .
- Soit A et B deux événements :
 - si $A \subset B$, on dit que A *implique* B ;
 - si $A \cap B = \emptyset$, on dit que A et B sont *incompatibles* (ou tout simplement : disjoints) ;
 - l'ensemble $A \cap B$ est appelé « événement A et B » ;
 - l'ensemble $A \cup B$ est appelé « événement A ou B ».

Définition 23.1.1 (Système complet d'événements)

On dit qu'une famille $(A_i)_{0 \leq i \leq n}$ de Ω est un *système complet d'événements* si elle vérifie les conditions :

- (facultatif) les événements A_0, \dots, A_n sont tous non vides : $\forall i \in \llbracket 0, n \rrbracket, A_i \neq \emptyset$.
- les événements A_0, \dots, A_n sont deux à deux incompatibles : $\forall (i, j) \in \llbracket 0, n \rrbracket^2, i \neq j \Rightarrow A_i \cap A_j = \emptyset$.
- la réunion des événements A_0, \dots, A_n est l'événement certain : $\bigcup_{n=0}^m A_n = \Omega$.

Avec cette définition, et pour toute issue ω de l'expérience, un et un seul des événements A_i est réalisé.

Définition 23.1.2 (probabilité sur Ω ; espace probabilisé)

Soit Ω un univers fini.

On appelle *probabilité* sur Ω toute application \mathbb{P} définie sur $\mathcal{P}(\Omega)$, à valeurs dans $[0, 1]$ telle que :

- a) $\mathbb{P}(\Omega) = 1$ (probabilité de l'événement certain)
- b) pour tous événements A, B incompatibles, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Le couple (Ω, \mathbb{P}) est alors appelé un *espace probabilisé fini*.

▷ Quelques propriétés immédiates

Soit (Ω, \mathbb{P}) un espace probabilisé fini.

Dans ce qui suit, A, B, \dots désignent des événements (des parties de Ω) quelconques.

- On a $\boxed{\mathbb{P}(\emptyset) = 0}$ (choisir $A = B = \emptyset$ dans la définition précédente).

- Si $(A_n)_{0 \leq n \leq m}$ sont deux à deux incompatibles, alors $\mathbb{P}\left(\bigcup_{n=0}^m A_n\right) = \sum_{n=0}^m \mathbb{P}(A_n)$ (récurrence facile)
- Pour tout événement A , on a $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$ (choisir $B = \bar{A}$ dans la définition)
- Si $A \subset B$ alors $\mathbb{P}(A) \leq \mathbb{P}(B)$ (croissance de la probabilité).
En effet $B = A \cup (B \setminus A)$, avec $A \cap (B \setminus A) = \emptyset$, donc $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$.
- Si A et B sont deux événements, alors $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

On a $\begin{cases} A \cup B = A \cup (B \setminus A) \\ B = (A \cap B) \cup (B \setminus A) \end{cases}$ (union d'événements incompatibles)

Il en résulte $\begin{cases} \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \\ \mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A) \end{cases}$

Par différence, on trouve $\mathbb{P}(A \cup B) - \mathbb{P}(B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$, ce qu'il fallait démontrer.

- Si A, B, C sont trois événements, alors :

$$\begin{aligned} \mathbb{P}(A \cup B \cup C) &= \mathbb{P}((A \cup B) \cup C) = \mathbb{P}(A \cup B) + \mathbb{P}(C) - \mathbb{P}((A \cup B) \cap C) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) + \mathbb{P}(C) - \mathbb{P}((A \cap C) \cup (B \cap C)) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) + \mathbb{P}(C) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C) \\ &= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C) \end{aligned}$$

▷ Remarques

- Si $(A_i)_{0 \leq i \leq n}$ est un système complet d'événements, alors $\sum_{i=0}^n \mathbb{P}(A_i) = 1$.
- Une probabilité \mathbb{P} sur Ω est caractérisée par la donnée des $(\mathbb{P}(\{\omega\}))_{\omega \in \Omega}$, c'est-à-dire par la probabilité des événements élémentaires, avec bien sûr $\sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = 1$.
Pour tout événement A , on a : $\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$.
- Une situation courante est l'*hypothèse d'équiprobabilité*, où on pose : $\forall \omega \in \Omega, \mathbb{P}(\{\omega\}) = \frac{1}{\text{Card}(\Omega)}$.

On parle aussi de *probabilité uniforme*.

Pour toute partie A de Ω , on a alors $\mathbb{P}(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)}$.

Définition 23.1.3 (événements négligeables, événements presque sûrs)

Soit (Ω, \mathbb{P}) un espace probabilisé fini.

On dit qu'un événement A est *négligeable* si sa probabilité est nulle.

On dit qu'un événement A est *presque sûr* si sa probabilité vaut 1.

Remarques

- L'événement impossible \emptyset est évidemment négligeable, de même que l'événement certain Ω est presque sûr. Réciproquement, il peut exister des événements qui sont négligeables sans être impossibles ou presque sûrs sans être certains.

- Tout événement qui contient un événement presque sûr est lui-même presque sûr.
Tout événement inclus dans un événement négligeable est lui-même négligeable.
- Si A est presque sûr, et si B est un événement quelconque, alors $\mathbb{P}(B \cap A) = \mathbb{P}(B)$.

L'événement $B \cup A$ contient A , donc est presque sûr. Ainsi $\mathbb{P}(A) = \mathbb{P}(B \cup A) = 1$.
L'égalité $\mathbb{P}(B \cup A) = \mathbb{P}(B) + \mathbb{P}(A) - \mathbb{P}(B \cap A)$ donne alors $\mathbb{P}(B) = \mathbb{P}(B \cap A)$.

23.1.2 Conditionnement et indépendance

Il est sous-entendu que tous les énoncés de cette sous-section (définitions et propositions) commencent par la même phrase « Soit (Ω, \mathbb{P}) un espace probabilisé fini ».

Proposition 23.1.1 (probabilité conditionnelle de A sachant B)

Soit B un événement tel que $\mathbb{P}(B) > 0$. Pour tout événement A , on pose $\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.

Alors \mathbb{P}_B est une probabilité sur Ω , dite « application probabilité conditionnée à B ».

Il est clair que \mathbb{P}_B est à valeurs dans $[0, 1]$, et que $\mathbb{P}_B(\Omega) = 1$.

On suppose que A, A' sont incompatibles.

Il en est donc de même de $A \cap B$ et de $A' \cap B$. Ainsi :

$$\begin{aligned} \mathbb{P}_B(A \cup A') &= \frac{1}{\mathbb{P}(B)} \mathbb{P}((A \cup A') \cap B) = \frac{1}{\mathbb{P}(B)} \mathbb{P}((A \cap B) \cup (A' \cap B)) \\ &= \frac{1}{\mathbb{P}(B)} (\mathbb{P}(A \cap B) + \mathbb{P}(A' \cap B)) = \mathbb{P}_B(A) + \mathbb{P}_B(A') \end{aligned}$$

Le réel $\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ (aussi noté $\mathbb{P}(A | B)$) est la « probabilité conditionnelle de A sachant B ».

Avec cette définition, on a l'égalité évidente (mais importante) : $\mathbb{P}(A \cap B) = \mathbb{P}(B) \mathbb{P}_B(A)$.

Convention d'écriture

Si $\mathbb{P}(B) = 0$, qui implique $\mathbb{P}(A \cap B) = 0$, on convient que : $\mathbb{P}(A \cap B) = \mathbb{P}(B) \mathbb{P}(A | B) = 0$.

Cependant, dans ce cas, on évitera d'utiliser la notation \mathbb{P}_B .

Proposition 23.1.2 (formule des probabilités composées)

Soit $(A_n)_{n \geq 0}$ une suite d'événements. Pour tout n de \mathbb{N} , posons $B_n = \bigcap_{0 \leq k \leq n} A_k$.

Il est clair que la suite $n \mapsto \mathbb{P}(B_n)$ est décroissante.

Alors, pour tout entier n tel que $\mathbb{P}(B_n) > 0$, on a l'égalité $\mathbb{P}(B_{n+1}) = \mathbb{P}(A_0) \prod_{k=0}^n \mathbb{P}_{B_k}(A_{k+1})$.

Pour $n = 0$, la formule précédente s'écrit : $\mathbb{P}(A_0 \cap A_1) = \mathbb{P}(A_0) \mathbb{P}_{A_0}(A_1)$ (à condition que $\mathbb{P}(A_0) > 0$).

Pour $n = 1$, elle s'écrit : $\mathbb{P}(A_0 \cap A_1 \cap A_2) = \mathbb{P}(A_0) \mathbb{P}_{A_0}(A_1) \mathbb{P}_{A_0 \cap A_1}(A_2)$ (à condition que $\mathbb{P}(A_0 \cap A_1) > 0$).

Pour $n = 2$, elle s'écrit (sous la condition que $\mathbb{P}(A_0 \cap A_1 \cap A_2) > 0$) :

$$\mathbb{P}(A_0 \cap A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_0) \mathbb{P}_{A_0}(A_1) \mathbb{P}_{A_0 \cap A_1}(A_2) \mathbb{P}_{A_0 \cap A_1 \cap A_2}(A_3)$$

En utilisant l'autre notation pour les probabilités conditionnelles, cela s'écrit :

$$\mathbb{P}(A_0 \cap A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_0) \mathbb{P}(A_1 | A_0) \mathbb{P}(A_2 | A_0 \cap A_1) \mathbb{P}(A_3 | A_0 \cap A_1 \cap A_2)$$

Pour tout n de \mathbb{N} , on a $B_{n+1} = B_n \cap A_{n+1} \subset B_n$, ce qui justifie la décroissance de la suite $(\mathbb{P}(B_n))_{n \geq 0}$.

On se donne un entier m tel que $\mathbb{P}(B_m) > 0$. On a alors $\mathbb{P}(B_n) \geq \mathbb{P}(B_m) > 0$ pour tout n de $\llbracket 0, m \rrbracket$.

On a prouver l'égalité $(E_n) : \mathbb{P}(B_{n+1}) = \mathbb{P}(A_0) \prod_{k=0}^n \mathbb{P}(A_{k+1} | B_k)$ par récurrence finie sur n , avec $0 \leq n \leq m$.

On sait que $B_0 = A_0$ et $B_1 = A_0 \cap A_1$.

(E_0) est vraie car elle se réduit à $\mathbb{P}(B_1) = \mathbb{P}(A_0) \mathbb{P}(A_1 | B_0)$, c'est-à-dire $\mathbb{P}(A_0 \cap A_1) = \mathbb{P}(A_0) \mathbb{P}(A_1 | A_0)$.

On se donne n dans $\llbracket 1, m \rrbracket$ et on suppose que (E_{n-1}) est vraie.

On a $B_{n+1} = B_n \cap A_{n+1}$ et $\mathbb{P}(B_n > 0)$. On en déduit :

$$\mathbb{P}(B_{n+1}) = \mathbb{P}(B_n) \mathbb{P}(A_{n+1} | B_n) = \left(\mathbb{P}(A_0) \prod_{k=0}^{n-1} \mathbb{P}(A_{k+1} | B_k) \right) \mathbb{P}(A_{n+1} | B_n) = \mathbb{P}(A_0) \prod_{k=0}^n \mathbb{P}(A_{k+1} | B_k)$$

ce qui prouve l'égalité (E_n) et achève la récurrence.

Proposition 23.1.3 (formule des probabilités totales)

Soit $(A_i)_{0 \leq i \leq n}$ un système complet d'événements. Soit B un événement.

$$\text{Alors } \mathbb{P}(B) = \sum_{i=0}^n \mathbb{P}(B \cap A_i) = \sum_{i=0}^n \mathbb{P}(A_i) \mathbb{P}(B | A_i).$$

Rappel : si $\mathbb{P}(A_i) = 0$, on convient que $\mathbb{P}(A_i) \mathbb{P}(B | A_i) = 0$.

On a $\bigcup_{i=0}^n A_i = \Omega$ donc $B = B \cap \left(\bigcup_{i=0}^n A_i \right) = \bigcup_{i=0}^n (B \cap A_i)$.

De plus, les événements $B \cap A_i$ sont deux à deux incompatibles. On en déduit :

$$\mathbb{P}(B) = \mathbb{P}\left(\bigcup_{i=0}^n (B \cap A_i)\right) = \sum_{i=0}^n \mathbb{P}(B \cap A_i) = \sum_{i=0}^n \mathbb{P}(A_i) \mathbb{P}(B | A_i).$$

Proposition 23.1.4 (formule de Bayes, première version)

Soit A et B deux événements tels que $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$. Alors $\mathbb{P}(A | B) = \frac{\mathbb{P}(A) \mathbb{P}(B | A)}{\mathbb{P}(B)}$.

C'est une simple réécriture de : $\mathbb{P}(B) \mathbb{P}(A | B) = \mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B | A)$.

Proposition 23.1.5 (formule de Bayes, deuxième version)

Soit $(A_i)_{0 \leq i \leq n}$ un système complet fini d'événements.

Soit B un événement tel que $\mathbb{P}(B) > 0$.

$$\text{Pour tout } i \text{ de } \llbracket 0, n \rrbracket, \text{ on a : } \mathbb{P}(A_i | B) = \frac{\mathbb{P}(A_i) \mathbb{P}(B | A_i)}{\sum_{0 \leq j \leq n} \mathbb{P}(A_j) \mathbb{P}(B | A_j)}$$

On applique la formule précédente, et on utilise $\mathbb{P}(B) = \sum_{0 \leq j \leq n} \mathbb{P}(A_j) \mathbb{P}(B | A_j)$.

Définition 23.1.4 (indépendance de deux événements)

On dit que deux événements A et B sont *indépendants* si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Interprétation importante : si $\mathbb{P}(B) > 0$, l'indépendance de A et B équivaut à l'égalité $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Définition 23.1.5 (indépendance mutuelle d'une famille finie d'événements)

Soit $(A_i)_{0 \leq i \leq n}$ une famille finie d'événements.

On dit que A_0, \dots, A_n sont *mutuellement indépendants* si : $\forall J \subset \llbracket 0, n \rrbracket, \mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j)$.

Il est clair que toute sous-famille d'événements mutuellement indépendants est à son tour constituée d'événements mutuellement indépendants.

▷ **Rapport entre « indépendance mutuelle » et « indépendance deux à deux »**

– Soit $(A_i)_{0 \leq i \leq n}$ une famille finie d'événements.

S'ils sont mutuellement indépendants, ils sont indépendants deux à deux.

– À partir de trois événements la réciproque est fautive.

On retiendra donc que « l'indépendance deux à deux n'entraîne pas l'indépendance mutuelle ».

Il suffit pour s'en convaincre de connaître l'exemple de référence suivant :

On lance deux fois de suite une pièce équilibrée. On considère les événements A : « le 1^{er} lancer a renvoyé Pile », B : « le 2nd lancer a renvoyé Face » et C : « les deux lancers ont renvoyé le même résultat ».

Chacun est de probabilité $\frac{1}{2}$, et leurs intersections deux à deux sont toutes de probabilité $\frac{1}{4}$.

Mais $A \cap B \cap C = A \cap B$ donc $\mathbb{P}(A \cap B \cap C) = \frac{1}{4}$ est différent de $\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \frac{1}{8}$.

Les événements A, B, C sont donc indépendants deux à deux, mais pas mutuellement indépendants.

Remarques

– Si l'événement A est négligeable (c'est-à-dire si $\mathbb{P}(A) = 0$) ou s'il est presque sûr (c'est-à-dire si $\mathbb{P}(A) = 1$) alors A est indépendant de tout événement B (réciproque vraie).

Si $\mathbb{P}(A) = 0$, alors $\mathbb{P}(A \cap B) = 0$ et on a bien l'égalité $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Si $\mathbb{P}(A) = 1$, on sait que $\mathbb{P}(A \cap B) = \mathbb{P}(B)$ et on a bien l'égalité $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Si l'événement A est indépendant de tout événement B , il est indépendant de lui-même : on a donc les égalités $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$ et il en résulte $\mathbb{P}(A) = 0$ ou $\mathbb{P}(A) = 1$.

– Si A et B sont indépendants, alors A et \overline{B} le sont également.

En effet $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B})$ donc $\mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(\overline{B})$.

– Soit $(A_i)_{0 \leq i \leq n}$ une famille finie d'événements, mutuellement indépendants.

Pour tout $i \in \llbracket 0, n \rrbracket$, soit $B_i = A_i$ ou $B_i = \overline{A_i}$. Alors B_0, \dots, B_n sont mutuellement indépendants.

L'ordre dans lequel on considère les événements A_i n'a évidemment aucune importance.

Si dans la famille $(A_i)_{0 \leq i \leq n}$ on change *un* des A_i en son contraire, l'indépendance mutuelle est conservée

(définition de l'indépendance mutuelle et remarque précédente). Il suffit ensuite de procéder par une récurrence finie sur le nombre de A_i qu'on a remplacé par leur contraire.

23.2 Variables aléatoires

23.2.1 Variable aléatoire sur Ω

Définition 23.2.1 (variable aléatoire sur Ω)

Soit Ω un univers fini. Soit E un ensemble.

Toute application $X: \Omega \rightarrow E$ est appelée une *variable aléatoire*.

Si $E = \mathbb{R}$, on dit que X est une *variable aléatoire réelle*.

Remarque

La définition d'une variable aléatoire ne nécessite pas la connaissance d'une probabilité \mathbb{P} .

En revanche, si X est une variable aléatoire sur un espace probablisé fini (Ω, \mathbb{P}) , on sera intéressé à la connaissance des probabilités des événements $(X = x)$ (à suivre...).

Définition 23.2.2 (variable indicatrice d'un événement)

Soit Ω un univers fini. Soit A un événement.

On appelle *indicatrice de l'événement A* la variable aléatoire réelle définie par
$$\begin{cases} \forall \omega \in A, X(\omega) = 1 \\ \forall \omega \notin A, X(\omega) = 0 \end{cases}$$

Elle est souvent notée $\mathbb{1}_A$.

Proposition 23.2.1 (événements liés à une variable aléatoire)

Soit $X: \Omega \rightarrow E$ une variable aléatoire sur Ω .

Soit x un élément de E , et soit U une partie de E .

L'événement $X^{-1}(\{x\}) = \{\omega \in \Omega, X(\omega) = x\}$ est noté $(X = x)$.

L'événement $X^{-1}(U) = \{\omega \in \Omega, X(\omega) \in U\}$ est noté $(X \in U)$.

Remarque

Soit X une variable aléatoire réelle sur Ω .

On considère souvent les événements notés $(X \leq a)$, $(a \leq X \leq b)$, $(X > a)$, etc.

23.2.2 Loi et fonction de répartition d'une variable aléatoire

Définition 23.2.3 (loi d'une variable aléatoire, première définition)

Soit (Ω, \mathbb{P}) un espace probablisé fini. Soit $X: \Omega \rightarrow E$ une variable aléatoire.

L'ensemble fini $X(\Omega)$ peut-être considéré comme un nouvel univers.

L'application \mathbb{P}_X définie par : $\forall A \subset X(\Omega), \mathbb{P}_X(A) = \mathbb{P}(X \in A)$ est une probabilité sur $X(\Omega)$.

Cette probabilité est appelée *loi de la variable aléatoire X* .

Définition 23.2.4 (loi d'une variable aléatoire, deuxième définition)

Soit (Ω, \mathbb{P}) un espace probabilisé fini. Soit $X : \Omega \rightarrow E$ une variable aléatoire.

Posons $X(\Omega) = \{x_i, 0 \leq i \leq n\}$, les x_i étant ici distincts deux à deux.

Les événements $(X = x_i)$ constituent alors un système complet d'événements.

On appelle loi de X la donnée des probabilités $p_i = \mathbb{P}(X = x_i)$. On a bien sûr $\sum_{i=0}^n p_i = 1$.

Équivalence des deux définitions

Les deux définitions précédentes sont deux façons équivalentes de se donner la loi de X .

- En effet, si on se donne la loi de X au sens de la première définition, on se donne également les probabilités $p_i = \mathbb{P}(X = x_i)$ (considérer les singletons $A = \{x_i\}$ de $X(\Omega)$).
- Réciproquement, si on se donne les probabilités $p_i = \mathbb{P}(X = x_i)$ pour chacun des x_i de $X(\Omega)$, alors on connaît la probabilité $\mathbb{P}(X \in A)$ pour chaque partie A de $X(\Omega)$. Il suffit en effet d'écrire $\mathbb{P}(A) = \sum_{x_i \in A} \mathbb{P}(X = x_i)$ (somme finie).

Définition 23.2.5 (fonction de répartition d'une variable aléatoire réelle)

Soit (Ω, \mathbb{P}) un espace probabilisé fini, et soit $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire réelle.

La fonction $F_X : \mathbb{R} \rightarrow [0, 1]$ définie par $F_X(x) = \mathbb{P}(X \leq x)$ est appelée *fonction de répartition* de X .

Proposition 23.2.2 (croissance et limites en $\pm\infty$ d'une fonction de répartition)

Soit F_X la fonction de répartition d'une variable aléatoire réelle X .

Alors la fonction F_X est croissante sur \mathbb{R} .

Pour tout $x < \min(X(\Omega))$, on a $F_X(x) = 0$. Pour tout $x \geq \max(X(\Omega))$, on a $F_X(x) = 1$.

Si $x \leq y$, on a $(X \leq x) \subset (X \leq y)$ donc $\mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y)$: c'est la croissance de l'application F_X .

Si $x < \min(X(\Omega))$, l'événement $X \leq x$ est impossible, donc $F_X(x) = 0$.

Si $x \geq \max(X(\Omega))$, l'événement $X \leq x$ est certain, donc $F_X(x) = 1$.

► Remarques

Soit X une variable aléatoire réelle sur un espace probabilisé fini (Ω, \mathbb{P}) .

L'ensemble $X(\Omega)$ est fini. On note $X(\Omega) = \{x_0, x_1, \dots, x_n\}$, avec $x_0 < x_1 < \dots < x_n$.

- La fonction de répartition F_X est en escaliers sur \mathbb{R} .

Plus précisément : $\begin{cases} \forall x < x_0, F_X(x) = 0 \\ \forall x \geq x_n, F_X(x) = 1 \end{cases}$, et $F_X(x) = \sum_{j=0}^k \mathbb{P}(X = x_j)$ sur $[x_k, x_{k+1}[$ (avec $k \in \llbracket 0, n \rrbracket$).

Inversement : $\mathbb{P}(x_0) = F_X(x_0)$ et $\forall k \in \llbracket 1, n \rrbracket$, $\mathbb{P}(X = x_k) = F_X(x_k) - F_X(x_{k-1})$.

Cette égalité permet donc de retrouver la loi de X à partir de sa fonction de répartition.

La fonction F_X est discontinue en chaque point x_k où $\mathbb{P}(X = x_k) > 0$, et la valeur de $\mathbb{P}(X = x_k)$ est alors le « saut de discontinuité » de la fonction F_X en ce point.

- Pour $a < b$ on a : $\mathbb{P}(X > a) = 1 - F_X(a)$ et $F_X(b) - F_X(a) = \mathbb{P}(a < X \leq b)$.

C'est évident car $(X > a) = \overline{(X \leq a)}$ donc $\mathbb{P}(X > a) = 1 - \mathbb{P}(X \leq a) = 1 - F_X(a)$.

Ensuite $(X \leq b) = (X \leq a) \cup (a < X \leq b)$ (union disjointe).

Il en résulte : $\mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) = \mathbb{P}(a < X \leq b)$.

▷ Image d'une variable aléatoire par une fonction, loi associée

Proposition 23.2.3 (fonction d'une variable aléatoire)

Soit $X : \Omega \rightarrow E$ une variable aléatoire sur Ω .

Soit φ une fonction définie sur E (en tout cas sur $X(\Omega)$) et à valeurs dans un ensemble F .

L'application $Y = \varphi \circ X$ (par convention notée $\varphi(X)$) est donc une variable aléatoire sur Ω .

On sait que la loi \mathbb{P}_Y de Y est une probabilité sur $Y(\Omega)$.

Pour tout $B \subset Y(\Omega)$, on a : $\mathbb{P}_Y(B) = \mathbb{P}(Y \in B) = \mathbb{P}(\varphi(X) \in B) = \mathbb{P}(X \in \varphi^{-1}(B)) = \mathbb{P}_X(\varphi^{-1}(B))$.

En particulier : $\forall y_n \in Y(\Omega), \mathbb{P}(Y=y_n) = \sum_{\varphi(x_k)=y_n} \mathbb{P}(X=x_k)$.

Les cas usuels de fonctions d'une variable aléatoire sont $Y = aX + b$, $Y = X^2$, $Y = |X|$, etc.

Si $Y = X^2$, par exemple, on écrira : $\forall a \in \mathbb{R}^{+*}, \mathbb{P}(Y = a) = \mathbb{P}(X = \sqrt{a}) + \mathbb{P}(X = -\sqrt{a})$.

23.2.3 Trois lois usuelles

Dans les exemples qui suivent, on considère une variable aléatoire réelle X attachée à une expérience aléatoire dont on note Ω l'univers (il est fini).

▷ Loi uniforme sur un intervalle de n entiers consécutifs

Soit $\llbracket a, b \rrbracket$ un intervalle de n entiers consécutifs (donc $b = a + n - 1$)

Le plus souvent, il s'agit de $\llbracket 0, n-1 \rrbracket$ ou de $\llbracket 1, n \rrbracket$.

On dit que X suit la loi uniforme sur $\llbracket a, b \rrbracket$, et on note $X \hookrightarrow \mathcal{U}(\llbracket a, b \rrbracket)$ si $\begin{cases} X(\Omega) = \llbracket a, b \rrbracket = \llbracket a, a+n-1 \rrbracket \\ \forall k \in \llbracket a, b \rrbracket, \mathbb{P}(X=k) = \frac{1}{n} \end{cases}$

Remarque : si $X \hookrightarrow \mathcal{U}(\llbracket a, b \rrbracket)$, alors $Y = X - a + 1 \hookrightarrow \mathcal{U}(\llbracket 1, n \rrbracket)$.

▷ Loi de Bernoulli $\mathcal{B}(p)$ avec $0 \leq p \leq 1$

Soit p dans $[0, 1]$. Dans ce contexte, on note souvent $q = 1 - p$.

On dit que X suit la loi de Bernoulli de paramètre p , et on note $X \hookrightarrow \mathcal{B}(p)$ si $\begin{cases} X(\Omega) = \{0, 1\} \\ \mathbb{P}(X=1) = p \\ \mathbb{P}(X=0) = 1 - p \end{cases}$

On modélise ainsi une « épreuve de Bernoulli de paramètre p », c'est-à-dire une expérience aléatoire qui ne possède deux issues possibles : le « succès » (ici la valeur 1) avec la probabilité p , et l'« échec » (ici la valeur 0) avec la probabilité $q = 1 - p$.

Avec ce modèle, on peut dire que X est la variable indicatrice de l'événement « l'expérience s'est soldée par un succès ».

Remarque : si $X \hookrightarrow \mathcal{B}(p)$, alors $Y = 1 - X \hookrightarrow \mathcal{B}(q)$.

▷ **Loi binomiale** $\mathcal{B}(n, p)$ de paramètres $n \in \mathbb{N}^*$ et $p \in [0, 1]$

Soit n dans \mathbb{N}^* . Soit p dans $[0, 1]$ (dans ce contexte, on note souvent $q = 1 - p$).

On dit que X suit la loi binomiale de paramètres n, p , et on note $X \hookrightarrow \mathcal{B}(n, p)$ si :

$$X(\Omega) = \llbracket 0, n \rrbracket \quad \text{et} \quad \forall k \in \llbracket 0, n \rrbracket, \mathbb{P}(X=k) = \binom{n}{k} p^k q^{n-k}$$

On modélise ainsi une expérience aléatoire consistant en la répétition de n épreuves de Bernoulli indépendantes de même paramètre p (l'indépendance est une notion qui sera évoquée plus loin et on la prend ici dans son sens intuitif : les conditions de l'expérience sont telles que le résultat d'une de ces épreuves de Bernoulli « n'influe » pas sur celui des autres épreuves).

Avec ce modèle, on peut dire que X est la variable qui compte le nombre de succès dans cette répétition de n épreuves de Bernoulli indépendantes de même paramètre p .

Un autre modèle où intervient la loi $\mathcal{B}(n, p)$ est l'expérience aléatoire consistant à effectuer n tirages successifs « avec remise » d'une boule dans une urne contenant une proportion p de boules blanches, et de compter le nombre de boules blanches obtenues après ces n tirages. Le fait que le tirage s'effectue avec remise est important pour garantir l'indépendance des résultats des tirages successifs.

Remarque : si $X \hookrightarrow \mathcal{B}(n, p)$, alors $Y = n - X \hookrightarrow \mathcal{B}(n, q)$ (on rappelle que $q = 1 - p$).

23.3 Couples de variables aléatoires

23.3.1 Couple de variables aléatoires

Proposition 23.3.1 (couple formé par deux variables aléatoires)

Soit Ω un univers fini.

Soit $X : \Omega \rightarrow E$ et $Y : \Omega \rightarrow F$ deux variables aléatoires.

En posant : $\forall \omega \in \Omega, Z(\omega) = (X(\omega), Y(\omega))$, on définit une variable aléatoire $Z : \Omega \rightarrow E \times F$.

La proposition précédente admet une réciproque :

Proposition 23.3.2 (composantes d'une variable aléatoire à valeurs dans $E \times F$)

Soit Ω un univers fini.

Soit $Z : \Omega \rightarrow E \times F$ une variable aléatoire.

Soit $X : \Omega \rightarrow E$ et $Y : \Omega \rightarrow F$ les composantes de Z , définies par : $\forall \omega \in \Omega, Z(\omega) = (X(\omega), Y(\omega))$.

Alors les applications X et Y sont des variables aléatoires.

Définition 23.3.1 (couple de variables aléatoires)

Soit Ω un univers fini.

Dans les propositions précédentes, on dit que $Z = (X, Y)$ est un *couple de variables aléatoires*.

On retiendra qu'il importe peu que X, Y préexistent à Z ou au contraire s'en déduisent.

On remarquera également que $Z(\Omega)$ est une partie (souvent stricte) de $X(\Omega) \times Y(\Omega)$.

Pour tout (x, y) de $X(\Omega) \times Y(\Omega)$, l'événement $(Z = (x, y))$ sera noté $(X=x, Y=y)$.

▷ Combinaisons linéaires de variables aléatoires réelles

Soit $Z = (X, Y)$ un couple de variables aléatoires réelles sur Ω .

Pour tous scalaires λ, μ , l'application $\lambda X + \mu Y$ est une variable aléatoire réelle.

Il suffit en effet d'appliquer à $Z = (X, Y)$ la fonction φ définie sur \mathbb{R}^2 par $\varphi(x, y) = \lambda x + \mu y$.

L'ensemble des variables aléatoires réelles sur Ω est donc un \mathbb{R} -espace vectoriel.

Remarque : pour des raisons similaires, l'application XY est une variable aléatoire réelle sur Ω .

▷ Vecteurs aléatoires

On peut généraliser ce qui précède à une application de Ω dans un produit cartésien $F_1 \times \cdots \times F_n$.

Si on note $Z = (X_1, \dots, X_n)$, dire que Z est une variable aléatoire sur Ω équivaut à dire que chacune des applications $X_i : \Omega \rightarrow F_i$ est elle-même une variable aléatoire.

On dit alors que Z est un vecteur de variables aléatoires.

On peut aussi définir une variable aléatoire fonction $\varphi(X_1, X_2, \dots, X_n)$ des n variables aléatoires X_i .

23.3.2 Loi conjointe, loi marginale, loi conditionnelle

Définition 23.3.2 (loi conjointe d'un couple de variables aléatoires)

Soit X et Y deux variables aléatoires sur un espace probabilisé fini (Ω, \mathbb{P}) .

On appelle *loi conjointe* de X et Y la loi de la variable aléatoire $Z = (X, Y)$.

C'est donc la probabilité \mathbb{P}_Z définie sur $Z(\Omega)$ par : $\forall A \subset Z(\Omega), \mathbb{P}_Z(A) = \mathbb{P}(Z^{-1}(A))$.

▷ Autres présentations de la loi conjointe

En termes équivalents : *La loi conjointe de X et Y est la donnée des probabilités $\mathbb{P}(X=x, Y=y)$ pour tout (x, y) .*

On se souvient que $Z(\Omega)$ n'est qu'une partie de $X(\Omega) \times Y(\Omega)$.

Par ailleurs, il est clair que si (x, y) est dans le complémentaire de $Z(\Omega)$ dans $X(\Omega) \times Y(\Omega)$, l'événement $(X=x, Y=y)$ est impossible donc de probabilité nulle.

On peut donc également adopter la définition suivante :

La loi conjointe de X et Y est la donnée des $\mathbb{P}(X=x, Y=y)$ pour tout (x, y) de $X(\Omega) \times Y(\Omega)$.

Définition 23.3.3 (lois marginales d'un couple de variables aléatoires)

Soit $Z = (X, Y)$ un couple de variables aléatoires sur un espace probabilisé fini (Ω, \mathbb{P}) .

Les lois de X et de Y sont appelées *lois marginales* de Z .

▷ La loi conjointe détermine les lois marginales

Soit $Z = (X, Y)$ un couple de variables aléatoires sur (Ω, \mathbb{P}) .

Proposition 23.3.3 (de la loi conjointe aux lois marginales)

Soit $Z = (X, Y)$ un couple de variables aléatoires sur (Ω, \mathbb{P}) .

On se donne la loi conjointe de X et Y , donc les probabilités $\mathbb{P}(X=x, Y=y)$ pour $(x, y) \in X(\Omega) \times Y(\Omega)$.

Pour tout x de $X(\Omega)$, on a alors : $\mathbb{P}(X=x) = \sum_{y \in Y(\Omega)} \mathbb{P}(X=x, Y=y)$ (première loi marginale de Z).

De même : $\forall y \in Y(\Omega)$, $\mathbb{P}(Y=y) = \sum_{x \in X(\Omega)} \mathbb{P}(X=x, Y=y)$ (deuxième loi marginale de Z).

Les événements $(X=x)$ (quand x décrit $X(\Omega)$) forment un système complet.

Il en est de même des événements $(Y=y)$ quand y décrit $Y(\Omega)$.

Avec la formule des probabilités totales, on obtient effectivement :

$$\forall x \in X(\Omega), \mathbb{P}(X=x) = \sum_{y \in Y(\Omega)} \mathbb{P}(X=x, Y=y) \quad \text{et} \quad \forall y \in Y(\Omega), \mathbb{P}(Y=y) = \sum_{x \in X(\Omega)} \mathbb{P}(X=x, Y=y).$$

On voit donc qu'on obtient les deux lois marginales de $Z = (X, Y)$ à partir de la loi conjointe de X et Y , par « sommation dans les marges » (d'où le nom).

▷ Les lois marginales ne déterminent pas la loi conjointe

Soit $Z = (X, Y)$ un couple de variables aléatoires sur (Ω, \mathbb{P}) .

On suppose connues les lois marginales de Z , c'est-à-dire les lois de X et de Y , c'est-à-dire les probabilités $\mathbb{P}(X=x)$ pour tout x de $X(\Omega)$ et $\mathbb{P}(Y=y)$ pour tout y de $Y(\Omega)$.

Ces informations ne sont pas suffisantes pour retrouver la loi conjointe de X et de Y .

Un exemple simple : on pose $Z = (X, Y)$ avec $X(\Omega) = Y(\Omega) = \{0, 1\}$ et la loi conjointe :

$$\mathbb{P}(X=0, Y=0) = \mathbb{P}(X=1, Y=1) = p \quad \text{et} \quad \mathbb{P}(X=1, Y=0) = \mathbb{P}(X=0, Y=1) = \frac{1}{2} - p$$

On voit que $\mathbb{P}(X=0) = \mathbb{P}(X=1) = \frac{1}{2}$ et $\mathbb{P}(Y=0) = \mathbb{P}(Y=1) = \frac{1}{2}$.

Ici la simple connaissance des lois marginales du couple (X, Y) (toutes les deux uniformes sur $\{0, 1\}$) ne permet pas de reconstituer la loi conjointe de X et Y (car le paramètre p est quelconque dans $[0, \frac{1}{2}]$).

Définition 23.3.4 (loi conditionnelle de Y sachant $(X=x)$)

Soit $Z = (X, Y)$ un couple de variables aléatoires sur (Ω, \mathbb{P}) .

Soit x un élément de $X(\Omega)$ tel que $\mathbb{P}(X=x) > 0$.

Pour tout élément y de $Y(\Omega)$, on pose $\mathbb{P}_{(X=x)}(Y=y) = \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(X=x)}$.

Alors $\mathbb{P}_{(X=x)}$ est une probabilité sur $Y(\Omega)$, dite « probabilité conditionnelle de Y sachant $(X=x)$ ».

Remarques

– On peut bien sûr utiliser la notation : $\mathbb{P}(Y=y | X=x) = \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(X=x)}$

– Soit $y \in Y(\Omega)$ tel que $\mathbb{P}(Y=y) > 0$. La probabilité conditionnelle de X sachant $(Y=y)$ est :

$$\forall x \in X(\Omega), \mathbb{P}_{(Y=y)}(X=x) = \mathbb{P}(X=x | Y=y) = \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(Y=y)}.$$

Proposition 23.3.4 (des lois conditionnelles aux lois marginales)

Soit $Z = (X, Y)$ un couple de variables aléatoires sur (Ω, \mathbb{P}) .

Pour tout x de $X(\Omega)$, on a : $\mathbb{P}(X=x) = \sum_{y \in Y(\Omega)} \mathbb{P}(Y=y) \mathbb{P}(X=x | Y=y)$.

Pour tout y de $Y(\Omega)$, on a : $\mathbb{P}(Y=y) = \sum_{x \in X(\Omega)} \mathbb{P}(X=x) \mathbb{P}(Y=y | X=x)$.

On a les égalités $\mathbb{P}(X=x, Y=y) = \mathbb{P}(Y=y) \mathbb{P}(X=x | Y=y) = \mathbb{P}(X=x) \mathbb{P}(Y=y | X=x)$.

Il s'agit donc ici d'une simple réécriture d'une proposition précédente.

Rappel : si $\mathbb{P}(Y=y) = 0$ on convient que $\mathbb{P}(Y=y) \mathbb{P}(X=x | Y=y) = 0$.

Cette convention évite d'avoir à supposer que toutes les probabilités $\mathbb{P}(Y=y)$ sont strictement positives dans la première des égalités de la proposition précédente.

C'est bien sûr la même chose avec la seconde égalité à chaque fois que $\mathbb{P}(X=x) = 0$.

23.3.3 Couples de variables aléatoires indépendantes

Définition 23.3.5 (indépendance de deux variables aléatoires)

Soit $Z = (X, Y)$ un couple de variables aléatoires sur un espace probabilisé fini (Ω, \mathbb{P}) .

On dit que X et Y sont *indépendantes* si, $\forall (x, y) \in X(\Omega) \times Y(\Omega)$, $\mathbb{P}(X=x, Y=y) = \mathbb{P}(X=x) \mathbb{P}(Y=y)$.

Remarques

- L'indépendance de X et Y implique que, pour tout x de $X(\Omega)$ tel que $\mathbb{P}(X=x) > 0$, la loi conditionnelle de Y sachant $(X=x)$ est égale à la loi de Y .
Autrement dit : $\forall (x, y) \in X(\omega) \times Y(\omega)$ avec $\mathbb{P}(X=x) > 0$, $\forall y \in Y(\omega)$, $\mathbb{P}_{(X=x)}(Y=y) = \mathbb{P}(Y=y)$.
- L'indépendance de X et Y signifie que, pour tout (x, y) de $X(\Omega) \times Y(\Omega)$, les événements $(X=x)$ et $(Y=y)$ sont indépendants.
- En général, l'indépendance de deux variables aléatoires résulte du modèle décrivant l'expérience. C'est plus une hypothèse a priori qu'une conséquence d'un calcul.
- Si une variable aléatoire X est constante (ou seulement « presque constante ») elle est indépendante de toute autre variable aléatoire.

Proposition 23.3.5 (caractérisation de l'indépendance de deux variables aléatoires)

Soit $Z = (X, Y)$ un couple de variables aléatoires indépendantes sur (Ω, \mathbb{P}) .

Alors X, Y sont indépendantes $\Leftrightarrow \forall A \subset X(\Omega), \forall B \subset Y(\Omega), \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B)$.

- On suppose que $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B)$ pour tout $A \subset X(\omega)$ et tout $B \subset Y(\omega)$.
C'est vrai en particulier pour les singletons $A = \{x\}$ et $B = \{y\}$ pour tout (x, y) de $X(\Omega) \times Y(\Omega)$.
On a donc : $\forall (x, y) \in X(\Omega) \times Y(\Omega), \mathbb{P}(X=x, Y=y) = \mathbb{P}(X=x) \mathbb{P}(Y=y)$, ce qui est la définition de l'indépendance de X et Y .
- Réciproquement, on suppose que X et Y sont indépendantes.
Soit A une partie de $X(\Omega)$ et soit B une partie de $Y(\Omega)$.
Tout comme $X(\Omega)$ et $Y(\Omega)$, les ensembles A et B sont finis.
On pose $A = \{x_j, j \in J \subset \mathbb{N}\}$ (les x_j distincts) et $B = \{y_k, k \in K \subset \mathbb{N}\}$ (les y_k distincts).

On alors, successivement :

$$\begin{aligned}
 \mathbb{P}(X \in A, Y \in B) &= \mathbb{P}\left(\bigcup_{x_j \in A} (X = x_j) \cap \bigcup_{x_k \in B} (Y = y_k)\right) = \mathbb{P}\left(\bigcup_{x_j \in A} \left(\bigcup_{x_k \in B} (X = x_j) \cap (Y = y_k)\right)\right) \\
 &= \sum_{x_j \in A} \mathbb{P}\left(\bigcup_{x_k \in B} (X = x_j) \cap (Y = y_k)\right) \\
 &= \sum_{x_j \in A} \sum_{x_k \in B} \mathbb{P}(X = x_j, Y = y_k) \\
 &= \sum_{x_j \in A} \sum_{x_k \in B} \mathbb{P}(X = x_j) \mathbb{P}(Y = y_k) \quad (\text{indépendance de } X \text{ et } Y) \\
 &= \sum_{x_j \in A} \mathbb{P}(X = x_j) \sum_{x_k \in B} \mathbb{P}(Y = y_k) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B)
 \end{aligned}$$

Proposition 23.3.6 (fonctions de deux variables aléatoires indépendantes)

Soit $Z = (X, Y)$ un couple de variables aléatoires indépendantes sur (Ω, \mathbb{P}) .

Soit $X' = \varphi(X)$ une fonction de X et $Y' = \psi(Y)$ une fonction de Y .

Alors les variables aléatoires X' et Y' sont indépendantes.

On se donne deux parties A' de $X'(\Omega)$ et B' de $Y'(\Omega)$. On a successivement :

$$\begin{aligned}
 \mathbb{P}(X' \in A', Y' \in B') &= \mathbb{P}(X \in \varphi^{-1}(A'), Y \in \varphi^{-1}(B')) \\
 &= \mathbb{P}(X \in \varphi^{-1}(A')) \mathbb{P}(Y \in \varphi^{-1}(B')) = \mathbb{P}(X' \in A') \mathbb{P}(Y' \in B')
 \end{aligned}$$

□

23.3.4 Variables mutuellement indépendantes

Définition 23.3.6 (variables aléatoires mutuellement indépendantes)

Soit $(X_i)_{1 \leq i \leq n}$ une suite finie de variables aléatoires sur (Ω, \mathbb{P}) .

On dit que ces variables sont *mutuellement indépendantes* si :

$\forall (x_1, x_2, \dots, x_n) \in X_1(\Omega) \times X_2(\Omega) \times \dots \times X_n(\Omega)$, les événements $(X_1 = x_1), (X_2 = x_2), \dots, (X_n = x_n)$ sont mutuellement indépendants.

Rappelons que cela signifie : pour toute partie J de $\llbracket 1, n \rrbracket$, $\mathbb{P}\left(\bigcap_{j \in J} (X_j = x_j)\right) = \prod_{j \in J} \mathbb{P}(X_j = x_j)$.

Proposition 23.3.7 (caractérisation de l'indépendance mutuelle de n variables aléatoires)

Soit $(X_i)_{1 \leq i \leq n}$ une suite finie de variables aléatoires sur (Ω, \mathbb{P}) .

Elles sont mutuellement indépendantes si et seulement si :

$$\left\{ \begin{array}{l} \forall A_1 \subset X_1(\Omega), \forall A_2 \subset X_2(\Omega), \dots, \forall A_n \subset X_n(\Omega) \\ \text{pour toute partie } J \text{ de l'intervalle } \llbracket 1, n \rrbracket \end{array} \right. \quad \text{on a l'égalité : } \mathbb{P}\left(\bigcap_{j \in J} (X_j \in A_j)\right) = \prod_{j \in J} \mathbb{P}(X_j \in A_j).$$

▷ La loi binomiale comme somme de variables de Bernoulli indépendantes

Soit X une variable aléatoire réelle suivant la loi binomiale $\mathcal{B}(n, p)$ de paramètres $n \in \mathbb{N}^*$ et $p \in [0, 1]$.
On sait que X représente le nombre de succès dans une répétition de n épreuves de Bernoulli indépendantes de même paramètre p .

Soit X_i la variable de Bernoulli de paramètre p associée à l'épreuve n° i .

Avec ces notations, on a $X = X_1 + X_2 + \dots + X_n$.

Remarques

- Dans la pratique, si on réalise n expériences aléatoires mutuellement indépendantes (les résultats d'une ou plusieurs d'entre elles n'affectent pas le résultat des autres : ce sont les conditions générales de ces expériences qui permettent d'émettre cette hypothèse) et si on attache à chacune de ces expériences une variable aléatoire X_i , alors les X_i sont mutuellement indépendantes.
- Si X_1, X_2, \dots, X_n sont mutuellement indépendantes, il est de même des $(X_j)_{j \in J}$, avec $J \subset \llbracket 1, n \rrbracket$. En particulier les variables aléatoires X_i sont indépendantes deux à deux.
- Dire que deux événements A et B sont indépendants, c'est dire que leurs variables indicatrices $\mathbb{1}_A$ et $\mathbb{1}_B$ sont indépendantes.

Écrire que $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ c'est écrire $\mathbb{P}(\mathbb{1}_A = 1, \mathbb{1}_B = 1) = \mathbb{P}(\mathbb{1}_A = 1)\mathbb{P}(\mathbb{1}_B = 1)$.

Mais cela implique $\begin{cases} \mathbb{P}(\mathbb{1}_A = 1, \mathbb{1}_B = 0) = \mathbb{P}(\mathbb{1}_A = 1)\mathbb{P}(\mathbb{1}_B = 0) \\ \mathbb{P}(\mathbb{1}_A = 0, \mathbb{1}_B = 1) = \mathbb{P}(\mathbb{1}_A = 0)\mathbb{P}(\mathbb{1}_B = 1) \\ \mathbb{P}(\mathbb{1}_A = 0, \mathbb{1}_B = 0) = \mathbb{P}(\mathbb{1}_A = 0)\mathbb{P}(\mathbb{1}_B = 0) \end{cases}$ d'où l'indépendance de $\mathbb{1}_A$ et $\mathbb{1}_B$.

Réciproquement, si les deux variables indicatrices $\mathbb{1}_A$ et $\mathbb{1}_B$ sont indépendantes, alors on a l'égalité $\mathbb{P}(\mathbb{1}_A = 1, \mathbb{1}_B = 1) = \mathbb{P}(\mathbb{1}_A = 1)\mathbb{P}(\mathbb{1}_B = 1)$ c'est-à-dire $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

De même, dire que les événements $(A_i)_{1 \leq i \leq n}$ sont mutuellement indépendants c'est dire que leurs variables indicatrices $\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_n}$ sont mutuellement indépendantes.

- De même que l'indépendance « deux à deux » des événements A_1, \dots, A_n (où $n \geq 3$) n'implique pas leur indépendance mutuelle, l'indépendance « deux à deux » de n variables X_1, \dots, X_n n'implique pas leur indépendance mutuelle (prendre les variables indicatrices $\mathbb{1}_{A_i}$).

À partir de variables aléatoires mutuellement indépendantes, on peut en former d'autres :

Proposition 23.3.8 (fonctions de variables mutuellement indépendantes)

Soit $(X_i)_{1 \leq i \leq n}$ une suite finie de variables aléatoires sur (Ω, \mathbb{P}) , mutuellement indépendantes.

Pour tout i de $\llbracket 1, n \rrbracket$, soit $Y_i = \varphi_i(X_i)$ une fonction de la variable aléatoire X_i .

Alors les variables aléatoires $(Y_i)_{1 \leq i \leq n}$ sont mutuellement indépendantes.

Autre exemple : si X_1, X_2, \dots, X_n sont mutuellement indépendantes, et si J et K sont deux parties disjointes de $\llbracket 1, n \rrbracket$, toute fonction Y des $(X_j)_{j \in J}$ est indépendante de toute fonction Z des $(X_k)_{k \in K}$.

23.4 Espérance et variance

23.4.1 Espérance d'une variable aléatoire réelle

- Soit $X : \Omega \rightarrow E$ une variable aléatoire réelle sur (Ω, \mathbb{P}) , avec Ω un ensemble fini.

La quantité $E(X) = \sum_{x \in X(\Omega)} \mathbb{P}(X=x) x$ est appelée *espérance* de la variable aléatoire réelle X .

Cette somme finie est la *moyenne pondérée* des valeurs x que la variable X est susceptible de prendre, le poids affecté à chacune de ces valeurs x étant la probabilité de l'événement $(X=x)$.

– Plutôt que de sommer sur les valeurs x de $X(\Omega)$, on peut sommer sur les résultats élémentaires ω .

On obtient alors la relation : $E(X) = \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) X(\omega)$.

L'événement $(X = x)$ est la réunion disjointe des singletons $\{\omega\}$ tels que $X(\omega) = x$.

Ainsi $\mathbb{P}(X = x) = \sum_{\substack{\omega \in \Omega \\ X(\omega) = x}} \mathbb{P}(\{\omega\})$ et on en déduit :

$$\sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) X(\omega) = \sum_{x \in X(\Omega)} \left(\sum_{\substack{\omega \in \Omega \\ X(\omega) = x}} \mathbb{P}(\{\omega\}) X(\omega) \right) = \sum_{x \in X(\Omega)} \left(x \sum_{\substack{\omega \in \Omega \\ X(\omega) = x}} \mathbb{P}(\{\omega\}) \right) = \sum_{x \in X(\Omega)} x \mathbb{P}(X = x) = E(X)$$

▷ Espérance de lois usuelles

– Espérance d'une loi constante : si X est constante, de valeur a , alors $E(X) = a$.

Évidentissime : si $X(\Omega) = \{a\}$, alors $\mathbb{P}(X = a) = 1$ donc $E(X) = \sum_{x \in X(\Omega)} \mathbb{P}(X = x) x = \mathbb{P}(X = a) a = a$.

– Espérance d'une variable indicatrice : si X est l'indicatrice d'un événement A , alors $E(X) = \mathbb{P}(A)$.

Évident : on a $(X = 1) = A$ et $(X = 0) = \bar{A}$ donc $E(X) = \mathbb{P}(X = 0) 0 + \mathbb{P}(X = 1) 1 = \mathbb{P}(X = 1) = \mathbb{P}(A)$.

– Espérance d'une loi de Bernoulli :

Si X suit la loi de Bernoulli $\mathcal{B}(p)$ de paramètre p , alors $E(X) = p$.

Trop facile : $E(X) = \mathbb{P}(X = 0) \cdot 0 + \mathbb{P}(X = 1) \cdot 1 = (1 - p) \cdot 0 + p \cdot 1 = p$.

– Espérance d'une loi uniforme :

Si X suit la loi uniforme $\mathcal{U}([a, b])$ sur $[a, b]$, alors $E(X) = \frac{a + b}{2}$.

Très facile : on a $E(X) = \sum_{k=a}^b k \mathbb{P}(X = k) = \frac{1}{n} \sum_{k=a}^b k = \frac{1}{n} \left(n \frac{a+b}{2} \right) = \frac{a+b}{2}$.

– Espérance de la loi binomiale :

Si X suit la loi binomiale $\mathcal{B}(n, p)$ de paramètres n, p , alors $E(X) = np$.

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

▷ Théorème du transfert

La propriété suivante est très importante.

On considère une fonction $Y = f(X)$ d'une variable aléatoire X .

Pour calculer $E(Y)$, il faut théoriquement connaître la loi de Y , c'est-à-dire les probabilités $\mathbb{P}(Y = y)$.

En fait il suffit de connaître, et d'utiliser, la loi de X :

Proposition 23.4.1 (théorème du transfert)

Soit $X : \Omega \rightarrow E$ une variable aléatoire sur (Ω, \mathbb{P}) , avec Ω fini.

Soit $f : E \rightarrow \mathbb{R}$ une application. Alors $E(f(X)) = \sum_{x \in X(\Omega)} \mathbb{P}(X=x) f(x)$.

L'événement $(Y=y)$ est la réunion disjointe des événements $(X=x)$ pour les x de $X(\Omega)$ tels que $f(x) = y$.

$$E(Y) = \sum_{y \in Y(\Omega)} y \mathbb{P}(Y=y) = \sum_{y \in Y(\Omega)} \left(\sum_{\substack{x \in X(\Omega) \\ f(x)=y}} y \mathbb{P}(X=x) \right) = \sum_{x \in X(\Omega)} \left(\sum_{\substack{y \in Y(\Omega) \\ y=f(x)}} y \mathbb{P}(X=x) \right) = \sum_{x \in X(\Omega)} f(x) \mathbb{P}(X=x).$$

L'interversion $\sum_{y \in Y(\Omega)} \sum_{\substack{x \in X(\Omega) \\ f(x)=y}} = \sum_{x \in X(\Omega)} \sum_{\substack{y \in Y(\Omega) \\ y=f(x)}}$ est justifiée (sommées finies), et $\sum_{\substack{y \in Y(\Omega) \\ y=f(x)}}$ se réduit à un seul terme.

▷ **Croissance et linéarité de l'espérance**

Soit X et Y deux variables aléatoires réelles sur (Ω, \mathbb{P}) , avec Ω un ensemble fini.

– On suppose que $X(\omega) \leq Y(\omega)$ pour tout ω de Ω . Alors $E(X) \leq E(Y)$.

$$\text{Évident : } E(X) = \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) X(\omega) \leq \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) Y(\omega) = E(Y).$$

– Si $X \geq 0$ alors on ne peut avoir $E(X) = 0$ que si $\mathbb{P}(X=0) = 1$ (c'est-à-dire : X est « presque nulle »).

La somme $E(X) = \sum_{x \in X(\Omega)} \mathbb{P}(X=x) x$ (à termes positifs ou nuls) est nulle donc tous ses termes sont nuls.

Cela implique $\mathbb{P}(X=x) = 0$ si $x > 0$, donc $\mathbb{P}(X=0) = 1$.

– Pour tous réels a, b , on a : $E(aX + b) = aE(X) + b$.

On applique le théorème du transfert à $Y = aX + b$:

$$E(aX + b) = \sum_{x \in X(\Omega)} \mathbb{P}(X=x)(ax + b) = a \sum_{x \in X(\Omega)} \mathbb{P}(X=x)x + \sum_{x \in X(\Omega)} \mathbb{P}(X=x) = aE(X) + b.$$

– Plus généralement, on a l'égalité : $E(aX + bY) = aE(X) + bE(Y)$.

$S = aX + bY$ est une fonction du couple $Z = (X, Y)$ (en fait $S = \varphi(Z)$, où $\varphi(x, y) = ax + by$).

$$\text{On en déduit : } E(S) = \sum_{(x,y)} (ax + by) \mathbb{P}(X=x, Y=y) = a \sum_{(x,y)} x \mathbb{P}(X=x, Y=y) + b \sum_{(x,y)} y \mathbb{P}(X=x, Y=y).$$

$$\text{Mais } \sum_{(x,y)} x \mathbb{P}(X=x, Y=y) = \sum_x x \left(\sum_y \mathbb{P}(X=x, Y=y) \right) = \sum_x x \mathbb{P}(X=x) = E(X).$$

$$\text{De même } \sum_{(x,y)} y \mathbb{P}(X=x, Y=y) = \sum_y y \left(\sum_x \mathbb{P}(X=x, Y=y) \right) = \sum_y y \mathbb{P}(Y=y) = E(Y).$$

On a finalement obtenu : $E(aX + bY) = aE(X) + bE(Y)$.

– Par récurrence évidente : pour toutes variables X_1, \dots, X_n , on a : $E\left(\sum_{i=1}^n \lambda_i X_i\right) = \sum_{i=1}^n \lambda_i E(X_i)$.

▷ **Où l'on retrouve l'espérance de la loi binomiale** $\mathcal{B}(n, p)$

On sait que si $X \leftrightarrow \mathcal{B}(n, p)$ alors $X = \sum_{i=1}^n X_i$ où les $X_i \leftrightarrow \mathcal{B}(p)$ sont mutuellement indépendantes.

Cette représentation de X permet de retrouver : $E(X) = \sum_{i=1}^n \underbrace{E(X_i)}_{=p} = np$.

Remarque : l'indépendance des X_i n'est pas utilisée ici.

▷ **Espérance du produit de deux variables indépendantes**

Soit X et Y deux variables aléatoires réelles sur (Ω, \mathbb{P}) , avec Ω un ensemble fini.

On suppose que X et Y sont indépendantes. Alors : $E(XY) = E(X)E(Y)$.

On utilise le théorème du transfert, appliqué à XY en tant que fonction du couple (X, Y) .

$$E(XY) = \sum_{(x,y)} xy \mathbb{P}(X=x, Y=y) = \sum_{(x,y)} xy \mathbb{P}(X=x) \mathbb{P}(Y=y) = \sum_x x \mathbb{P}(X=x) \sum_y y \mathbb{P}(Y=y) = E(X)E(Y).$$

Inversement, l'égalité $E(XY) = E(X)E(Y)$ n'implique pas l'indépendance de X et Y en général.

23.4.2 Variance, écart-type

Soit X une variable aléatoire réelle sur (Ω, \mathbb{P}) , avec Ω fini. Posons $m = E(X)$.

La variable positive $Y = (X - m)^2$ mesure la « dispersion » de X autour de son espérance m .

L'espérance de Y est appelé *variance* de X et est notée $V(X)$. Ainsi $V(X) = E((X - E(X))^2)$.

La quantité $\sigma(X) = \sqrt{V(X)}$ est appelée *écart-type* de la variable aléatoire X .

L'utilisation de l'écart-type est justifiée pour des raisons d'homogénéité par rapport aux valeurs de X .

▷ **Propriétés de la variance et de l'écart-type**

– On a la *formule de König-Huyghens* : $V(X) = E(X^2) - E^2(X)$.

Posons $m = E(X)$.

On développe : $V(X) = E((X - m)^2) = E(X^2 - 2mX + m^2) = E(X^2) - 2m \overbrace{E(X)}^{=m} + m^2 = E(X^2) - m^2$.

– Pour tous réels a, b on a : $V(aX + b) = a^2V(X)$, donc $\sigma(aX + b) = |a|\sigma(X)$.

Si $Y = aX + b$, alors $E(Y) = aE(X) + b$ donc $Y - E(Y) = a(X - E(X))$.

Ainsi $V(Y) = E((Y - E(Y))^2) = E(a^2(X - E(X))^2) = a^2E((X - E(X))^2) = a^2V(X)$.

– On a toujours $V(X) \geq 0$ et on ne peut avoir $V(X) = 0$ que si X est « presque constante ».

Posons $m = E(X)$. La variable $(X - m)^2 \geq 0$ est à valeurs positives.

Ainsi : $V(X) = 0 \Leftrightarrow E((X - m)^2) = 0 \Leftrightarrow \mathbb{P}((X - m)^2 = 0) = 1 \Leftrightarrow \mathbb{P}(X = m) = 1$.

▷ **Variance de lois usuelles**

– Variance d'une loi constante : si X est constante, alors $V(X) = 0$.

Si X est constante de valeur a , alors X^2 est constante de valeur a^2 .

Ainsi $E(X^2) = a^2$ et $E(X) = a$, donc $V(X) = E(X^2) - E(X)^2 = 0$.

– Variance d'une variable indicatrice : si $X = \mathbb{1}_A$, alors $V(X) = \mathbb{P}(A) \mathbb{P}(\bar{A})$.

En effet, on a encore $X^2 = \mathbb{1}_A$, donc $E(X) = E(X^2) = \mathbb{P}(A)$.

Ainsi $V(X) = E(X^2) - E(X)^2 = \mathbb{P}(A)^2 - \mathbb{P}(A) = \mathbb{P}(A)(1 - \mathbb{P}(A)) = \mathbb{P}(A) \mathbb{P}(\bar{A})$.

– Variance d'une loi de Bernoulli :

Si X suit la loi de Bernoulli $\mathcal{B}(p)$ de paramètre p , alors $V(X) = pq$.

Ici X et X^2 suivent la même loi, donc $V(X) = E(X) - E(X)^2 = p - p^2 = pq$.

– Variance d'une loi uniforme :

Si X suit la loi uniforme $\mathcal{U}(\llbracket a, b \rrbracket)$ (n entiers successifs), alors $V(X) = \frac{n^2 - 1}{12}$.

Par translation, on se ramène à $X \leftrightarrow \mathcal{U}(\llbracket 1, n \rrbracket)$.

$E(X^2) = \sum_{k=1}^n k^2 \mathbb{P}(X = k) = \frac{1}{n} \sum_{k=1}^n k^2 = \frac{(n+1)(2n+1)}{6}$. D'autre part $E(X) = \frac{n+1}{2}$.

Ainsi : $V(X) = E(X^2) - E^2(X) = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n+1}{12}(2(2n+1) - 3(n+1)) = \frac{n^2 - 1}{12}$

– Variance de la loi binomiale :

Si X suit la loi binomiale $\mathcal{B}(n, p)$ de paramètres n, p , alors $V(X) = npq$.

On sait déjà que $E(X) = np$. Pour calculer $V(X)$, on calcule $E(X(X-1))$.

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

D'autre part $E(X) = np$.

Ainsi $V(X) = E(X(X-1)) + E(X) - E^2(X) = n(n-1)p^2 + np - n^2p^2 = np(1-p) = npq$.

▷ Covariance de deux variables aléatoires réelles (univers fini, MPSI)

La *covariance* de X et Y est l'espérance, notée $\text{Cov}(X, Y)$, de la variable $(X - E(X))(Y - E(Y))$.

On remarque que $V(X) = \text{Cov}(X, X)$.

On a l'égalité $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$.

$$\text{Soit } m = E(X) \text{ et } m' = E(Y) : \begin{cases} \text{Cov}(X, Y) = E(X - m)(Y - m') = E(XY - mY - m'X + mm') \\ \quad = E(XY) - mE(Y) - m'E(X) + mm' = E(XY) - mm' \end{cases}$$

Si X et Y sont indépendantes, elles sont de covariance nulle (réciproque fautive).

On a $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ (symétrie) et $\text{Cov}(aX + bX', Y) = a \text{Cov}(X, Y) + b \text{Cov}(X', Y)$ (bilinearité).

La symétrie est évidente. Quant à la linéarité à gauche (qui devient bilinéarité par symétrie) :

$$\begin{aligned} \text{Cov}(aX + bX', Y) &= E((aX + bX')Y) - E(aX + bX')E(Y) = E(aXY + bX'Y) - (aE(X) + bE(X'))E(Y) \\ &= aE(XY) + bE(X'Y) - aE(X)E(Y) - bE(X')E(Y) \\ &= a(E(XY) - E(X)E(Y)) + b(E(X'Y) - E(X')E(Y)) = a \text{Cov}(X, Y) + b \text{Cov}(X', Y). \quad \square \end{aligned}$$

On vérifie que $V(X + Y) = V(X) + 2 \text{Cov}(X, Y) + V(Y)$.

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

□

On en déduit que si X et Y sont indépendantes, alors $V(X + Y) = V(X) + V(Y)$.

Plus généralement si X_1, \dots, X_n sont deux à deux indépendantes, on a : $V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i)$.

$$\text{Par bilinéarité : } V\left(\sum_{i=1}^n X_i\right) = \text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j\right) = \sum_{i=1}^n \overbrace{\text{Cov}(X_i, X_i)}^{=V(X_i)} + 2 \sum_{i < j} \overbrace{\text{Cov}(X_i, X_j)}^{=0} = \sum_{i=1}^n V(X_i)$$

On remarque que dans le résultat précédent, il n'est pas nécessaire que les variables E_i soient mutuellement indépendantes : il suffit qu'elles soient indépendantes deux à deux.

▷ Où l'on retrouve la variance de la loi binomiale $\mathcal{B}(n, p)$

On sait que si $X \leftrightarrow \mathcal{B}(n, p)$ alors $X = \sum_{i=1}^n X_i$ où les $X_i \leftrightarrow \mathcal{B}(p)$ sont mutuellement indépendantes.

Cette représentation de X permet de retrouver : $V(X) = \sum_{i=1}^n \overbrace{V(X_i)}^{=pq} = npq$.

Remarque : seule l'indépendance deux à deux des X_i est utilisée ici, pas leur indépendance mutuelle.

▷ Coefficient de corrélation linéaire

Soit X, Y deux variables aléatoires réelles de variance non nulle.

On appelle coefficient de corrélation linéaire de X et Y la quantité $\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$.

Si X et Y sont indépendantes, on a bien sûr $\rho(X, Y) = 0$.

On a toujours $|\rho(X, Y)| \leq 1$.

$$\text{D'autre part : } \begin{cases} \rho(X, Y) = 1 & \Leftrightarrow (\exists a > 0, \exists b \in \mathbb{R}, \mathbb{P}(Y = aX + b) = 1) \\ \rho(X, Y) = -1 & \Leftrightarrow (\exists a < 0, \exists b \in \mathbb{R}, \mathbb{P}(Y = aX + b) = 1) \end{cases}$$

En termes plus suggestifs, dire que $\rho(X, Y)$ vaut ± 1 (ou en est très proche) c'est dire que chacune des deux variables X et Y dépend « quasiment » de l'autre de façon affine.

C'est une inégalité de Cauchy-Schwarz...

Pour tout réel λ , on a $A(\lambda) = V(\lambda X - Y) = \lambda^2 V(X) - 2\lambda \text{Cov}(X, Y) + V(Y)$.

Ce trinôme $A(\lambda)$ reste positif, donc son discriminant $\Delta = \text{Cov}(X, Y)^2 - V(X)V(Y)$ est négatif ou nul.

Mais écrire $\Delta \leq 0$, c'est écrire $|\rho(X, Y)| \leq 1$.

Ensuite : $|\rho(X, Y)| = 1 \Leftrightarrow \Delta = 0 \Leftrightarrow \exists a \in \mathbb{R}, V(aX - Y) = 0$.

Avec ces notations, la racine double de $A(\lambda)$ est $a = \frac{\text{Cov}(X, Y)}{V(X)}$ donc a et $\rho(X, Y)$ sont de même signe.

Mais $V(aX - Y) = 0$ signifie que $aX - Y$ est « presque sûre » càd : $\exists b \in \mathbb{R}, \mathbb{P}(Y = aX + b) = 1$.

▷ Inégalités de Markov et de Bienaymé-Tchebychev

Soit X une variable aléatoire réelle positive sur (Ω, \mathbb{P}) , avec Ω un ensemble fini.

– Pour tout $a > 0$, on a l'inégalité de Markov : $\mathbb{P}(X \geq a) \leq \frac{E(X)}{a}$.

On décompose $E(X) = \sum_{x \in X(\Omega)} \mathbb{P}(X=x) x$ en $E(X) = \sum_{x < a} \mathbb{P}(X=x) x + \sum_{x \geq a} \mathbb{P}(X=x) x$.

La première somme ne contient que des termes positifs.

Ainsi $E(X) \geq \sum_{x \geq a} \mathbb{P}(X=x) x \geq a \sum_{x \geq a} \mathbb{P}(X=x)$, c'est-à-dire $E(X) \geq a \mathbb{P}(X \geq a)$.

– Pour tout $\varepsilon > 0$, on a : $\mathbb{P}(|X - E(X)| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}$ (inégalité de Bienaymé-Chebychev).

On applique l'inégalité de Markov à la variable positive $Y = (X - E(X))^2$, dont l'espérance est $V(X)$.

Pour tout $\varepsilon > 0$, on a : $\mathbb{P}(Y \geq \varepsilon^2) \leq \frac{V(X)}{\varepsilon^2}$, c'est-à-dire : $\mathbb{P}(|X - E(X)| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}$.