Chapitre 11

Variables aléatoires discrètes

`https://www.jean-philippe-preaux.fr`

Table des matières

0. Préliminaire : Familles sommables

Tenter de sommer une famille dénombrables de nombres (réels ou complexes) n'est pas sans soulever certaines difficultés. Par exemple pour la famille $((-1)^n)_{n\in\mathbb{Z}}$:

$$\lim_{N\to+\infty}\sum_{n=-N}^{N}(-1)^n = \lim_{N\to+\infty}(-1)^N \quad \text{diverge}$$
$$\lim_{N\to+\infty}\sum_{n=-N}^{N+1}(-1)^n = \lim_{N\to+\infty}0 \quad \text{converge}$$

Pour une famille $(a_n)_{n\in\mathbb{N}}$ indicée par $\mathbb{N}$, la sommation sous forme d'une série numérique, $\sum_{n=0}^{+\infty}a_n$ impose l'ordre de sommation par l'ordre sur les entiers naturels. Mais pour une famille dénombrable, quelle est l'influence de l'ordre de sommation sur la convergence de la somme ?

D'ailleurs, c'est un résultat remarquable de Riemann, que pour toute série semi-convergente, sans ordre de sommation pré-établi tout peut survenir :

Théorème de réarrangement de Riemann
Soit $\sum_{n\in\mathbb{N}}a_n$ une série réelle semi-convergente (i.e. convergente et non absolument convergente), et soit $\ell\in\overline{\mathbb{R}}$. Il existe une bijection $\sigma:\mathbb{N}\longrightarrow\mathbb{N}$ telle que la série $\sum_{n\in\mathbb{N}}a_{\sigma(n)}$ ait pour limite ℓ.

Par contre il s'avère pour une série <u>absolument</u> convergente $\sum_{n\in\mathbb{N}}a_n$, quelque soit la bijection $\sigma:\mathbb{N}\longrightarrow\mathbb{N}$, toutes les séries $\sum_{n\in\mathbb{N}}a_{\sigma(n)}$ ont même nature et même limite : l'ordre de sommation n'a aucune influence.

Pour les variables aléatoires nous aurons besoin de disposer d'un minimum de connaissances pour pouvoir sommer des familles dénombrables de réels, ou calculer des séries de séries numériques. Nous les fournissons ici, sans preuve (elles ne présentent pas de difficulté particulière). Toute cette partie est à avoir à l'esprit mais "ne saurait faire l'objet d'une question d'un sujet de concours" selon le programme officiel.

Définition 1. (Sommabilité d'une famille dénombrable de réels positifs)
Soit $(a_i)_{i\in I}$ une famille dénombrable de réels positifs. On appelle <u>*somme*</u> *de cette famille :*

$$\sum_{i\in I} a_i = \sup\left\{ \sum_{i\in F} a_i \mid F \subset I \text{ et } F \text{ est fini} \right\} \in \mathbb{R}_+ \cup \{+\infty\}$$

La famille $(a_i)_{i\in I}$ est dite <u>*sommable*</u> *si cette somme est fini.*

Lorsque $I = \mathbb{N}$ la sommabilité d'une famille de réels positifs équivaut à la convergence de sa série.

Propriété 1. (Sommabilité d'une série de réels positifs)
Soit $(a_i)_{i\in\mathbb{N}}$ une suite de réels <u>*positifs*</u>*. La famille est sommable si et seulement si la série $\sum_{i\in\mathbb{N}} a_i$ converge, et dans ce cas :*

$$\sum_{i\in\mathbb{N}} a_i = \sum_{i=0}^{+\infty} a_i$$

Sans surprise le résultat sur la comparaison des séries à termes positifs demeure vrai :

Propriété 2. (Croissance)
Pour deux familles $(a_i)_{i\in I}$ et $(b_i)_{i\in I}$ de réels <u>*positifs*</u>*, si pour tout $i \in I$, $a_i \leqslant b_i$ alors si $(b_i)_{i\in I}$ est sommable, $(a_i)_{i\in I}$ l'est aussi et :*

$$\sum_{i\in I} a_i \leqslant \sum_{i\in I} b_i$$

Pour une famille dénombrable de réels quelconques, ou plus généralement de complexe, la sommabilité est définie comme étant celle de ses modules.

Notation. Pour tout réel x, on note :

$$x^+ = \max(x,0) \qquad ; \qquad x^- = \max(0,-x)$$

on a alors :

$$x = x^+ - x^- \qquad ; \qquad |x| = x^+ + x^-$$

Proposition-Définition 2. (Sommabilité d'une famille de complexes)
Soit $(a_i)_{i\in I}$ une famille dénombrable de réels, ou complexes. La famille est dite <u>*sommable*</u> *si la famille $(|a_i|)_{i\in I}$ est sommable. Sa somme est alors définie ainsi :*

- *Si $(a_i)_{i\in I}$ est une famille dénombrable de réels, la sommabilité de $(a_i)_{i\in I}$ équivaut à celle des familles $(a_i^+)_{i\in I}$ et $(a_i^-)_{i\in I}$ de réels positifs, et la somme est définie comme :*

$$\sum_{i\in I} a_i = \sum_{i\in I} a_i^+ - \sum_{i\in I} a_i^-$$

- *Si $(a_i)_{i\in I}$ est une famille dénombrable de complexes, la sommabilité de $(a_i)_{i\in I}$ équivaut à celle des familles $(\mathrm{Re}(a_i))_{i\in I}$ et $(\mathrm{Im}(a_i))_{i\in I}$ de réels, et la somme est définie comme :*

$$\sum_{i\in I} a_i = \sum_{i\in I} \mathrm{Re}(a_i) + \mathrm{i}\sum_{i\in I} \mathrm{Im}(a_i)$$

Exemple. $((-1)^n)_{n\in\mathbb{Z}}$ n'est pas sommable.

Lorsque $I = \mathbb{N}$, la sommabilité équivaut à la convergence absolue de la série.

Propriété 3. (Sommabilité d'une série de complexes)
Soit $(a_i)_{i\in\mathbb{N}}$ une suite de complexes. La famille est sommable si et seulement si la série $\sum_{i\in\mathbb{N}} a_i$ converge absolument, et dans ce cas :

$$\sum_{i\in\mathbb{N}} a_i = \sum_{i=0}^{+\infty} a_i$$

Pour une famille sommable, l'ordre de sommation n'a aucune influence sur sa sommabilité et sa somme. Remarquons que la preuve découle immédiatement des définitions (puisque les parties finies de I ne dépendent pas d'un ordre sur I).

Proposition 4. (Réarrangement)
Pour une famille sommable $(a_i)_{i\in I}$, pour tout bijection $\sigma : I \longrightarrow I$, la famille $(a_{\sigma(i)})_{i\in I}$ est sommable et toutes leur sommes sont égales.

La linéarité s'obtient facilement par linéarité finie puis passage à la borne supérieure.

Propriété 5. (Linéarité)
Pour toutes familles $(a_i)_{i\in I}$ et $(b_i)_{i\in I}$ sommables de complexes, et pour tout tout $(\lambda, \mu) \in \mathbb{C}^2$, la famille $(\lambda \cdot a_i + \mu \cdot b_i)_{i\in I}$ est sommable et :

$$\sum_{i\in I} \lambda \cdot a_i + \mu \cdot b_i = \lambda \sum_{i\in I} a_i + \mu \sum_{i\in I} b_i$$

Remarque. L'ensemble des familles sommables $(a_i)_{i\in I}$ à valeurs réelles ou complexes est noté $\ell^1(I, \mathbb{R})$ ou $\ell^1(I, \mathbb{C})$; c'est un espace vectoriel.

Pour le calcul, le théorème suivant, dit de sommation par paquet, est essentiel :

Théorème 6. (de sommation par paquet)
Si $(a_i)_{i\in I}$ est une famille sommable, et si $(I_j)_{j\in J}$ est une partition de I, alors :

$$\sum_{i\in I} a_i = \sum_{j\in J} \sum_{i\in I_j} a_i$$

En particulier son application aux sommes doublement indicées donne les théorèmes importants :

Théorème 7. (De Fubini)
Pour toute famille sommable $(u_{i,j})_{(i,j)\in I\times J}$:

$$\sum_{(i,j)\in I\times J} u_{i,j} = \sum_{i\in I} \sum_{j\in J} u_{i,j} = \sum_{j\in J} \sum_{i\in I} u_{i,j}$$

qui généralise la décomposition d'une somme double finie par ligne/colonne connue depuis la sup ; ainsi que :

Théorème 8. (Produit de sommes)
Si les deux familles $(a_i)_{i\in I}$ et $(b_j)_{j\in J}$ sont sommables, alors la famille $(a_i b_j)_{(i,j)\in I\times J}$ est sommable et :

$$\sum_{(i,j)\in I\times J} a_i b_j = \left(\sum_{i\in I} a_i\right) \times \left(\sum_{j\in J} b_j\right)$$

qui généralise le produit de Cauchy de séries absolument convergentes.

Exemple. Soit $q \in \mathbb{C}$ avec $|q| < 1$; la famille $(q^{i+j})_{(i,j)\in\mathbb{N}^2}$ est sommable puisque les familles $(q^i)_{i\in\mathbb{N}}$ et $(q^j)_{j\in\mathbb{N}}$ le sont et :

$$\sum_{(i,j)\in\mathbb{N}^2} q^{i+j} = \left(\sum_{i\in N} q^i\right) \times \left(\sum_{j\in\mathbb{N}} q^j\right) = \frac{1}{(1-q)^2}$$

On a pour les séries doubles :

Théorème 9. (De convergence des séries doubles)
Soit $(a_{n,m})_{(n,m)\in\mathbb{N}^2}$ une famille de complexes doublement indicée. Si :
– Pour tout $n \in \mathbb{N}$, la série $\sum_{m\in\mathbb{N}} a_{n,m}$ est absolument convergente, et
– la série $\sum_{n\in\mathbb{N}} \sum_{m=0}^{+\infty} |a_{n,m}|$ converge,
alors la famille $(a_{n,m})_{(n,m)\in\mathbb{N}^2}$ est sommable et :

$$\sum_{(n,m)\in\mathbb{N}^2} a_{n,m} = \sum_{n=0}^{+\infty}\sum_{m=0}^{+\infty} a_{n,m} = \sum_{m=0}^{+\infty}\sum_{n=0}^{+\infty} a_{n,m}$$

1. Variables aléatoires discrètes

1.1. Loi d'une variable aléatoire discrète.

Définition 3. (Variable aléatoire discrète)
Soit $(\Omega, \mathscr{A})$ un espace probabilisable et E un ensemble quelconque. Une variable aléatoire discrète sur $(\Omega, \mathscr{A})$ et à valeur dans E est une application :

$$X : \Omega \longrightarrow E$$

vérifiant :
- *L'image de X, noté $X(\Omega)$ ou $\operatorname{Im}\Omega$ est une partie au plus dénombrable de E.*
- *Pour tout $x \in X(\Omega)$, $X^{-1}(\{x\}) \in \mathscr{A}$.*

Lorsque $E = \mathbb{R}$ on parle de variable aléatoire réelle discrète.

Remarques.
• Souvent $E = \mathbb{R}^p$.
• Dans un univers probabilisable une variable aléatoire discrète est une *application* (et non une variable) qui associe à chaque issue de l'expérience aléatoire un élément de E telle que :
– $X(\Omega)$ est au plus dénombrable : c'est que ce que veut dire que la variable aléatoire est *discrète*.
– Tout élément $x \in X(\Omega)$ définit un *événement* $X^{-1}(\{x\}) \in \mathscr{A}$, noté aussi $(X = x)$.
Informellement elle sert à décrire des événements mesurés dans E lors d'une expérience aléatoire.
Exemple. d'une variable aléatoire discrète réelle : On lance deux dés 6 ; on prend comme univers $\Omega = [\![1, 6]\!]^2$ et comme tribu $\mathscr{P}(\Omega)$. Soit par exemple $X : \Omega \longrightarrow \mathbb{R}$ qui à $\omega \in \Omega$ associe la somme des 2 dés :

$$X((1,2)) = 3\ ; X((3,4)) = 7\ ; \text{etc.}\ ; X((a,b)) = a+b \qquad X(\Omega) = [\![2, 12]\!].$$
$$(X = 5) = X^{-1}(\{5\}) = \big\{(1,4),(2,3),(3,2),(4,1)\big\}$$

Proposition 10.
Soit X une variable aléatoire discrète sur $(\Omega, \mathscr{A})$ à valeur dans E. Pour toute partie $A \subset E$, $X^{-1}(A)$ est un événement, i.e. $X^{-1}(A) \in \mathscr{A}$.

Démonstration. Puisque $X(\Omega)$ est dénombrable, $A \cap X(\Omega)$ est au plus dénombrable. Ainsi :

$$X^{-1}(A) = X^{-1}(A \cap X(\Omega)) = \bigcup_{x\in A\cap X(\Omega)} X^{-1}(\{x\})$$

est une réunion dénombrable d'événements ; c'est donc un événement. ∎

Notation. – Pour tout $x \in E$, l'événement $X^{-1}(\{x\})$ est noté $(X = x)$.
– Pour toute partie $A \subset E$, l'événement $X^{-1}(A)$ est noté $(X \in A)$.
– Pour une variable aléatoire discrète réelle, et pour tout $(a, b) \in \mathbb{R}^2$ l'événement $(X \leqslant a)$ désigne $X^{-1}(]-\infty, a])$, $(a \leqslant X < b)$ désigne $X^{-1}([a, b[)$, etc.

Remarque. Puisque $X(\Omega)$ est au plus dénombrable, tout événement $(X \in A)$ est réunion au plus dénombrable d'événements de la forme $X^{-1}(\{x\})$.

Définition 4. (Loi d'une variable aléatoire discrète)
Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire discrète à valeurs dans E. L'application :

$$\begin{array}{rcl} P_X : \mathscr{P}(X(\Omega)) & \longrightarrow & [0,1] \\ A \subset X(\Omega) & \longmapsto & \mathbb{P}(X \in A) \end{array}$$

qui à toute partie A de l'image de X associe la probabilité de l'événement $(X \in A)$, est appelé la loi de X.

Proposition 11.
Sous les mêmes hypothèses, la loi de X est une probabilité sur l'espace probabilisable $(X(\Omega), \mathscr{P}(X(\Omega))$.

Démonstration. Puisque pour tout $A \subset X(\Omega)$, $(X \in A)$ est un événement de $\mathscr{A}$, l'application P_X est bien définie et à valeur dans $[0,1]$. Il y a deux conditions à vérifier :

$$P_X(X(\Omega)) = \mathbb{P}(X \in X(\Omega)) = \mathbb{P}(X^{-1}(X(\Omega)) = \mathbb{P}(\Omega) = 1.$$

Pour toute suite dénombrable $(A_n)_{n\in\mathbb{N}}$ de parties de $X(\Omega)$ deux à deux incompatibles,

$$P_X\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \mathbb{P}\left(X^{-1}(\bigcup_{n\in\mathbb{N}} A_n)\right) = \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} X^{-1}(A_n)\right) = \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} (X \in A_n)\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(X \in A_n) = \sum_{n\in\mathbb{N}} P_X(A_n)$$

par σ-additivité de $\mathbb{P}$ puisque les $(A_n)_n$ étant deux à deux incompatibles, il en est de même de $\left(X^{-1}(A_n)\right)_n$. ■

Remarques.

• Il est parfois délicat de déterminer $X(\Omega)$; mais on peut aussi considérer la loi de X comme une probabilité sur $(E, \mathscr{P}(E))$ pour un ensemble $E \supset X(\Omega)$ au plus dénombrable ; on aura alors :

$$\forall x \in E \setminus X(\Omega),\ P_X(\{x\}) = 0$$

• L'espace probabilisable $(X(\Omega), \mathscr{P}(X(\Omega)), P_X)$ a un univers $X(\Omega)$ au plus dénombrable. Ainsi ses événements sont réunions au plus dénombrable d'événements de la forme $X^{-1}(\{x\})$. Par σ-additivité, la loi de X est uniquement déterminée par la donnée de $\mathbb{P}(X = x)$ pour tout $x \in X(\Omega)$.

Proposition 12.
Sous les mêmes hypothèses, la loi de X est uniquement déterminée par la famille des $\mathbb{P}(X = x)$ pour tout $x \in X(\Omega)$; on a alors :

$$\mathbb{P}(X \in A) = \sum_{x\in A} \mathbb{P}(X = x).$$

La famille des événements $(X = x)_{x\in X(\Omega)}$ joue ainsi un rôle crucial pour la V.A.D. X ; notamment :

Proposition-Définition 5. (Système complet d'événements associé à X)
Sous les mêmes hypothèses, la famille d'événements $(X = x)$ pour x décrivant l'image $X(\Omega)$ de X, est un système complet d'événements de $\mathscr{A}$.
On l'appelle le système complet d'événements associé à X.

Démonstration. Clairement $\bigcup_{x\in X(\Omega)}(X = x) \subset \Omega$; pour l'inclusion inverse, en notant pour $\omega \in \Omega$ quelconque, $x_\omega = X(\omega)$, alors $\omega \in X^{-1}(\{x_\omega\}) = (X = x_\omega) \subset \bigcup_{x\in X(\Omega)}(X = x)$, ainsi $\Omega \subset \bigcup_{x\in X(\Omega)}(X = x)$.

D'autre part ces événements sont deux à deux incompatibles, puisque si $x \neq x'$, alors $(X = x) \cap (X = x') = X^{-1}(\{x\}) \cap X^{-1}(\{x'\}) = \varnothing$ par définition d'une application. ■

Notation. Soit X une variable aléatoire discrète sur $(\Omega, \mathscr{A}, \mathbb{P})$ et $\mathcal{L}$ une probabilité sur $(X(\Omega), \mathscr{P}(X(\Omega)))$; on dit que X suit la loi $\mathcal{L}$, que l'on note $X \hookrightarrow \mathcal{L}$ ou $X \sim \mathcal{L}$ pour signifier que $\mathcal{L}$ est la loi de X.

Une variable aléatoire peut être obtenue par composition de X par une application f définie sur $\operatorname{Im} X$; sa loi se déduit alors de celle de X :

PROPOSITION 13. (Image d'une VAD par une application)
Soit X une variable aléatoire discrète sur $(\Omega, \mathscr{A}, \mathbb{P})$, et soit f une application définie sur $\operatorname{Im} X$. Alors $f \circ X$ est une variable aléatoire discrète notée $f(X)$, sur $(\Omega, \mathscr{A}, \mathbb{P})$ à valeurs dans $f(\operatorname{Im} X)$, et sa loi est donnée par :

$$\forall y \in f(\operatorname{Im} X),\ \mathbb{P}(f(X) = y) = \sum_{x \in f^{-1}(\{y\})} \mathbb{P}(X = x)$$

Démonstration. Clairement $f \circ X$ est une VAD puisque l'image par une application d'un ensemble au plus dénombrable est au plus dénombrable. Son image est $f(\operatorname{Im} X)$; quant à sa loi : soit $y \in f(\operatorname{Im} X)$ quelconque,

$$(f \circ X = y) = (f \circ X)^{-1}(\{y\}) = X^{-1}\left(f^{-1}(\{y\})\right) = \bigcup_{x \in f^{-1}(\{y\})} X^{-1}(\{x\}) = \bigcup_{x \in f^{-1}(\{y\})} (X = x)$$

Les événements $(X = x)$ pour $x \in f^{-1}(\{y\})$ étant deux à deux incompatibles, et $f^{-1}(\{y\})$ étant au plus dénombrable en tant que partie de l'ensemble au plus dénombrable $X(\Omega)$, on a par σ-additivité de $\mathbb{P}$:

$$\mathbb{P}(f \circ X = y) = \sum_{x \in f^{-1}(\{y\})} \mathbb{P}(X = x)$$

■

DÉFINITION 6. (Loi conditionnelle)
Soit X une variable aléatoire discrète sur $(\Omega, \mathscr{A}, \mathbb{P})$ et B un événement de probabilité non nulle. On appelle loi conditionnelle de X sachant B, la loi de X sur l'espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P}_B)$; pour tout $A \in \operatorname{Im} X$, on a :

$$\mathbb{P}(X \in A \mid B) = \mathbb{P}_B(X \in A) = \frac{\mathbb{P}((X \in A) \cap B)}{\mathbb{P}(B)}$$

1.2. Couple de variables aléatoires.

PROPOSITION-DÉFINITION 7. (Couple de variables aléatoires)
Étant donnés un espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$ et deux applications $X : \Omega \longrightarrow E_1$ et $Y : \Omega \longrightarrow E_2$, l'application :

$$\begin{array}{rcl} Z : \Omega & \longrightarrow & E_1 \times E_2 \\ \omega & \longmapsto & (X(\omega), Y(\omega)) \end{array}$$

est une variable aléatoire discrète si et seulement si X et Y le sont aussi.
On parle alors du couple de variables aléatoires discrètes (X, Y).

Démonstration. Soit $\pi_1 : E_1 \times E_2 \longrightarrow E_1$ et $\pi_2 : E_1 \times E_2 \longrightarrow E_2$ les deux projections : $\pi_1(x, y) = x$ et $\pi_2(x, y) = y$. D'après la proposition 13, si Z est une V.A.D. il en est de même de $\pi_1 \circ Z = X$ et $\pi_2 \circ Z = Y$.
Réciproquement : supposons que X et Y soient des VAD et vérifions que Z aussi. Il y a deux points à vérifier.
– Le produit cartésien $\operatorname{Im} X \times \operatorname{Im} Y$ est au plus dénombrable et puisque $\operatorname{Im} Z \subset \operatorname{Im} X \times \operatorname{Im} Y$, $\operatorname{Im} Z$ est au plus dénombrable.
– Soit $z = (x, y) \in \operatorname{Im} Z$; montrons que $Z^{-1}(\{z\}) \in \mathscr{A}$.

$$\omega \in Z^{-1}(\{z\}) \iff Z(\omega) = (x, y) \iff \omega \in X^{-1}(\{x\}) \cap Y^{-1}(\{y\})$$

et donc $Z^{-1}(\{z\}) = X^{-1}(\{x\}) \cap Y^{-1}(\{y\}) \in \mathscr{A}$ puisque c'est l'intersection de deux éléments de $\mathscr{A}$.
Ceci montre que Z est un V.A.D. ■

Remarque. On a toujours $\operatorname{Im}(X, Y) \subset \operatorname{Im} X \times \operatorname{Im} Y$, mais en général l'inégalité est stricte. Comme on le voit par exemple en considérant le couple (X, X) pour X ne suivant pas la loi certaine.

PROPOSITION 14.
Soit (X, Y) un couple de variables aléatoires discrètes. La famille d'événements :

$$((X = x) \cap (Y = y))_{(x,y) \in \operatorname{Im} X \times \operatorname{Im} Y}$$

est un système complet d'événements.

Démonstration. La famille est obtenue en complétant le système complet d'événements associé à la variable $Z = (X, Y)$ (cf. proposition-définition 5) :

$$((X = x) \cap (Y = y))_{(x,y) \in \operatorname{Im}(X,Y)}$$

en ajoutant tous les événements impossibles $((X = x) \cap (Y = y))$ pour $(x, y) \in \operatorname{Im} X \times \operatorname{Im} Y \setminus \operatorname{Im}(X, Y)$. La famille est donc encore un S.C.E. ■

La loi conjointe de X et Y est la loi du couple (X, Y) :

Définition 8. (Loi conjointe)
Étant données deux variables aléatoires X et Y sur l'espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$, on appelle loi conjointe des variables X et Y, la loi du couple (X, Y) :

$$\begin{array}{rcl} \mathbb{P}_{(X,Y)} : \mathscr{P}(\operatorname{Im}(X,Y)) & \longrightarrow & [0,1] \\ A & \longmapsto & \mathbb{P}((X,Y) \in A) \end{array}$$

Elle est uniquement déterminée par la famille des :

$$(\mathbb{P}((X = x) \cap (Y = y)))_{(x,y) \in \operatorname{Im} X \times \operatorname{Im} Y}$$

Pour tout $(x, y) \in \operatorname{Im} X \times \operatorname{Im} Y$, on note :

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}((X = x) \cap (Y = y))$$

Les lois marginales du couple (X, Y) sont les lois de X et de Y :

Définition 9. (Lois marginales)
Étant données deux variables aléatoires X et Y sur l'espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$, on appelle lois marginales du couple (X, Y) :
- *La première loi marginale est la loi de X,*
- *La seconde loi marginale est la loi de Y.*

Le point essentiel est que les lois marginales se déduisent de la loi conjointe. La réciproque est en général fausse.

Proposition 15.
Soit (X, Y) un couple de variables aléatoire discrètes.
- *Pour tout $x \in \operatorname{Im} X$:*
$$\mathbb{P}(X = x) = \sum_{y \in \operatorname{Im} Y} \mathbb{P}(X = x, Y = y)$$
- *Pour tout $y \in \operatorname{Im} Y$:*
$$\mathbb{P}(Y = y) = \sum_{x \in \operatorname{Im} X} \mathbb{P}(X = x, Y = y)$$

Démonstration. Puisque $(Y = y)_{y \in \operatorname{Im} Y}$ est un système complet d'événements, pour tout $x \in \operatorname{Im} X$:

$$\mathbb{P}(X = x) = \sum_{y \in \operatorname{Im} Y} \mathbb{P}((X = x) \cap (Y = y)) = \sum_{y \in \operatorname{Im} Y} \mathbb{P}(X = x, Y = y)$$

L'argument pour la seconde loi marginale est identique avec le SCE $(X = x)_{x \in \operatorname{Im} X}$. ∎

Remarque. En notant pour tout $(x_n, y_m) \in \operatorname{Im} X \times \operatorname{Im} Y$, $p_{n,m} = \mathbb{P}(X = x_n, Y = y_m)$, et en les représentant au sein d'un tableau, les lois marginales sont obtenues dans les marges en sommant les $(p_{n,m})$ par ligne ou par colonne.

$\mathbb{P}$	$(Y = y_1)$	$(Y = y_2)$	$\cdots$	$(Y = y_j)$	$\cdots$	loi de X
$(X = x_1)$	$p_{1,1}$	$p_{1,2}$	$\cdots$	$p_{1,j}$	$\cdots$	$\boldsymbol{p_{1,\cdot}}$
$(X = x_2)$	$p_{2,1}$	$p_{2,2}$	$\cdots$	$p_{2,j}$	$\cdots$	$\boldsymbol{p_{2,\cdot}}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
$(X = x_i)$	$p_{i,1}$	$p_{i,2}$	$\cdots$	$p_{i,j}$	$\cdots$	$\boldsymbol{p_{i,\cdot}}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
loi de Y	$\boldsymbol{p_{\cdot,1}}$	$\boldsymbol{p_{\cdot,2}}$	$\cdots$	$\boldsymbol{p_{\cdot,j}}$	$\cdots$	**1**

On note parfois pour tout (n, m), $p_{n,\cdot} = \mathbb{P}(X = x_n)$ et $p_{\cdot,m} = \mathbb{P}(Y = y_m)$.

Exemple. Les lois marginales ne déterminent pas la loi conjointe. En voici un contre-exemple : deux lois conjointes distinctes ayant mêmes loi marginales. Soit $0 \leqslant p \leqslant q \leqslant 1$,

	$Y=0$	$Y=1$	$\mathbb{P}_X$
$X=0$	p	$q-p$	q
$X=1$	0	$1-q$	$1-q$
$\mathbb{P}_Y$	p	$1-p$	1

et

	$Y=0$	$Y=1$	$\mathbb{P}_X$
$X=0$	pq	$q-pq$	q
$X=1$	$p-pq$	$1-p-q+pq$	$1-q$
$\mathbb{P}_Y$	p	$1-p$	1

Les variables aléatoires sont bien définies puisque la somme des probabilités fait 1 ; elles ont mêmes lois marginales mais des lois conjointes différentes.

Exercice 1. Soit $n \in \mathbb{N}^*$. On choisit au hasard un nombre X dans $[\![1, n]\!]$. On choisit ensuite un nombre Y au hasard dans $[\![1, X]\!]$.

1. Déterminer la loi conjointe de X et Y.
2. En déduire la loi (marginale) de Y.

Résolution.

On peut généraliser facilement au cas de n-uplets, ou vecteurs, de variables aléatoires :

PROPOSITION-DÉFINITION 10. (Vecteur de VAD)
Soient $X_1, X_2, \dots, X_n$ n variables aléatoires sur un même espace probabilisé, $(\Omega, \mathscr{A}, \mathbb{P})$, $X_i : \Omega \longrightarrow E_i$. Alors l'application :

$$\begin{array}{rcl} (X_1, \dots, X_n) : \Omega & \longrightarrow & E_1 \times \cdots \times E_n \\ \omega & \longmapsto & (X_1(\omega), \dots, X_n(\omega)) \end{array}$$

est une variable aléatoire discrète.
- *La loi conjointe est la loi de la variable $(X_1, \dots, X_n)$.*
- *Les lois marginales sont les lois des variables $X_1, \dots, X_n$.*

Démonstration. L'argument est semblable au cas des couples de VAD. ■

1.3. Indépendance de variables aléatoires.

DÉFINITION 11. (Variables aléatoires indépendantes)
Deux variables aléatoires discrètes X et Y définies sur un même espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$ sont dites indépendantes lorsque pour tout $(x, y) \in X(\Omega) \times Y(\Omega)$

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \times \mathbb{P}(Y = y).$$

On note dans ce cas $X \perp\!\!\!\perp Y$.

Autrement dit, pour des variables aléatoires indépendantes, la loi conjointe s'obtient par produit à l'aide des lois marginales. Par exemple pour (X, Y) donné par le tableau :

	$Y=0$	$Y=1$	$\mathbb{P}_X$
$X=0$	p	$q-p$	q
$X=1$	0	$1-q$	$1-q$
$\mathbb{P}_Y$	p	$1-p$	1

X et Y sont indépendantes si et seulement si $p = 0$ ou $q = 1$.

Remarque. Un cas fréquent de variables aléatoires indépendantes survient lorsque X et Y décrivent les résultats de deux épreuves menées de manières indépendantes ; par exemple : on lance deux fois un dé, soit X le premier chiffre obtenu, soit Y le second. On tire une boule dans une urne, soit X son numéro, on la remet, et on en retire une pour obtenir le numéro Y.

Proposition 16. (Caractérisation de l'indépendance)
Soient X et Y deux variables aléatoires sur $(\Omega, \mathscr{A}, \mathbb{P})$; X et Y sont indépendantes si et seulement si pour tout $A \subset X(\Omega)$ et tout $B \subset Y(\Omega)$,

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \times \mathbb{P}(Y \in B).$$

Démonstration. On montre deux implications ; la première est évidente : si pour tout $A \subset \operatorname{Im} X$ et $B \subset \operatorname{Im} Y$, $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \times \mathbb{P}(Y \in B)$, alors en particulier pour A et B des singletons quelconques $A = \{x\}$ et $B = \{y\}$ avec $x \in \operatorname{Im} X$ et $y \in \operatorname{Im} Y$; ainsi $\forall (x, y) \in \operatorname{Im} X \times \operatorname{Im} Y$, $\mathbb{P}((X = x) \cap (Y = y)) = \mathbb{P}(X = x) \times \mathbb{P}(Y = y)$; donc $X \perp\!\!\!\perp Y$.

Réciproquement ; la famille $((X = x) \cap (Y \in B))_{x \in \operatorname{Im} X}$ étant une famille au plus dénombrable d'événements deux à deux incompatibles, on a par σ-additivité :

$$\mathbb{P}((X \in A) \cap (Y \in B)) = \sum_{x \in A} \mathbb{P}((X = x) \cap (Y \in B))$$

et en appliquant encore la σ-additivité à toutes les familles $((X = x) \cap (Y = y))_{y \in \operatorname{Im} Y}$ pour tout $x \in \operatorname{Im} X$:

$$\mathbb{P}((X \in A) \cap (Y \in B)) = \sum_{x \in A} \sum_{y \in B} \mathbb{P}((X = x) \cap (Y = y))$$

par indépendance de X et Y :

$$\begin{aligned}
\mathbb{P}((X \in A) \cap (Y \in B)) &= \sum_{x \in A} \sum_{y \in B} \mathbb{P}(X = x) \times \mathbb{P}(Y = y) \\
&= \sum_{x \in A} \left(\mathbb{P}(X = x) \times \sum_{y \in B} \mathbb{P}(Y = y) \right) \\
&= \sum_{x \in A} (\mathbb{P}(X = x) \times \mathbb{P}(Y \in B)) \\
&= \left(\sum_{x \in A} \mathbb{P}(X = x) \right) \times \mathbb{P}(Y \in B) = \mathbb{P}(X \in A) \times \mathbb{P}(Y \in B)
\end{aligned}$$

■

Proposition 17. (Lemme des coalitions)
Soient $X : \Omega \longrightarrow E$ et $Y : \Omega \longrightarrow F$ deux variables aléatoires indépendantes. Pour toutes fonctions f et g définies respectivement sur $\operatorname{Im} X$ et $\operatorname{Im} Y$, les variables aléatoires $f(X)$ et $g(Y)$ sont indépendantes.

Démonstration. Soient $a \in f(\operatorname{Im} X)$ et $b \in g(\operatorname{Im} Y)$ quelconques :

$$\begin{aligned}
\mathbb{P}((f(X) = a) \cap (g(Y) = b)) &= \mathbb{P}((X \in f^{-1}(\{a\})) \cap (Y \in g^{-1}(\{b\}))) \\
&= \mathbb{P}(X \in f^{-1}(\{a\}) \times \mathbb{P}(Y \in g^{-1}(\{b\})) \qquad \text{par indépendance (prop. 16)} \\
&= \mathbb{P}(f(X) \in \{a\}) \times \mathbb{P}(g(Y) \in \{b\})
\end{aligned}$$

d'où l'indépendance de $f(X)$ et $g(Y)$. ■

On peut caractériser l'indépendance de deux variables aléatoires à l'aide des lois conditionnelles.

Proposition 18. (Caractérisation de l'indépendance)
Deux variables X et Y sur $(\Omega, \mathscr{A}, \mathbb{P})$ sont indépendantes si et seulement pour tout $y \in \operatorname{Im} Y$ tel que $\mathbb{P}(Y = y) \neq 0$ la loi de X sachant $(Y = y)$ est la même que la loi de X. Autrement dit si et seulement si :

$$\forall y \in \operatorname{Im} Y, \mathbb{P}(Y = y) \neq 0 \implies \forall x \in \operatorname{Im} X, \mathbb{P}((X = x) \mid (Y = y)) = \mathbb{P}(X = x).$$

Démonstration. Si X et Y sont indépendantes : soit $y \in \operatorname{Im} Y$ avec $\mathbb{P}(Y = y) \neq 0$, alors :

$$\forall x \in \operatorname{Im} X, \ \mathbb{P}((X = x) \mid (Y = y)) = \frac{\mathbb{P}((X = x) \cap (Y = y))}{\mathbb{P}(Y = y)} = \frac{\mathbb{P}(X = x) \times \mathbb{P}(Y = y)}{\mathbb{P}(Y = y)} = \mathbb{P}(X = x)$$

Réciproquement soient $y \in \operatorname{Im} Y$ et $x \in \operatorname{Im} X$. Si $\mathbb{P}(Y = y) \neq 0$ alors :

$$\mathbb{P}((X = x) \cap (Y = y)) = \mathbb{P}((X = x) \mid (Y = y)) \times \mathbb{P}(Y = y) = \mathbb{P}(X = x) \times \mathbb{P}(Y = y)$$

tandis que si $\mathbb{P}(Y = y) = 0$, alors pour tout $x \in \operatorname{Im} X$, $(X = x) \cap (Y = y) \subset (Y = y)$ et donc $\mathbb{P}((X = x) \cap (Y = y)) \leqslant \mathbb{P}(Y = y) = 0$; ainsi $\mathbb{P}((X = x) \cap (Y = y)) = 0 = \mathbb{P}(X = x) \times \mathbb{P}(Y = y)$. Dans tous les cas on a bien $\mathbb{P}((X = x) \cap (Y = y)) = \mathbb{P}(X = x) \times \mathbb{P}(Y = y)$: les variables X et Y sont donc indépendantes. ■

On étend la notion de couples de variables aléatoires indépendantes à une suite finie ou dénombrable de variables aléatoires (on parle de variables mutuellement indépendantes).

Définition 12. (Indépendance mutuelle de variables aléatoires)

- *Soit* $(X_k)_{k\in[\![1,n]\!]}$ *une famille finie de variables aléatoires sur un même espace probabilisé. Les variables aléatoires* $X_1,\cdots,X_n$ *sont dites (mutuellement) indépendantes lorsque : pour tout* $(x_1,\cdots,x_n)\in X_1(\Omega)\times\cdots\times X_n(\Omega)$ *:*
$$\mathbb{P}\left(X_1=x_1,\cdots,X_n=x_n\right)=\mathbb{P}\left(X_1=x_1\right)\times\cdots\times\mathbb{P}\left(X_n=x_n\right).$$
- *Soit* $(X_n)_{n\in\mathbb{N}}$ *une famille dénombrable de variables aléatoires sur un même espace probabilisé. Les variables aléatoires* $(X_n)_{n\in\mathbb{N}}$ *sont dites (mutuellement) indépendantes lorsque pour tout sous-ensemble fini* $I\subset\mathbb{N}$*, les variables aléatoires de la sous-famille finie* $(X_i)_{i\in I}$ *sont mutuellement indépendantes.*

Remarques.

• On dit souvent indépendantes, plutôt que mutuellement indépendantes.

• Bien sûr, par commutativité de l'intersection et de la multiplication l'indépendance ne dépend pas de l'ordre des variables.

• L'indépendance mutuelle peut être supposée dès que les $(X_n)_n$ décrivent des épreuves indépendantes : les résultats de plusieurs lancers de dés, de plusieurs tirages avec remise dans une urne, *etc.*

• Un cas typique apparaît lorsqu'on répète de manière indépendante une même épreuve, et que chaque X_n décrit le résultat de la n-ième épreuve. Les (X_n) sont alors indépendantes et toutes de même loi. On parle d'une suite de variables indépendantes identiquement distribuées, que l'on note en abrégé des variables *i.i.d.*.

Exemple. Jeu de pile ou face infini.

On lance indéfiniment une pièce donnant pile avec probabilité $p\in]0,1[$; on note X_n la variable de Bernoulli (voir §2.1.2) qui donne 1 lorsque le n-ème lancer donne Pile, et 0 lorsqu'il donne Face. La suite $(X_n)_{n\in\mathbb{N}^*}$ est une suite de variables indépendantes toutes de même loi, (i.i.d.) sur l'espace probabilisé $(\{0,1\}^{\mathbb{N}^*},\mathscr{A},\mathbb{P})$ où la tribu $\mathscr{A}$ est la plus petite tribu contenant tous les événements P_n/F_n : "le n-ième lancer donne Pile/Face", pour $n\in\mathbb{N}$.

Propriété 19. (L'indépendance se transmet à toute sous-famille)
Toute sous-famille d'une famille de variables indépendantes, est aussi indépendante.

Démonstration. C'est évident par définition pour une famille dénombrable. Il suffit de montrer que si $X_1,X_2,\dots,X_n$ est une famille de variables indépendantes, $X_2,\dots,X_n$ est aussi indépendante. Soient $(x_2,\dots,x_n)\in\operatorname{Im}X_2\times\cdots\times\operatorname{Im}X_n$ fixés. Alors $((X_1=x)\cap(X_2=x_2)\cap\cdots\cap(X_n=x_n))_{x\in\operatorname{Im}X_1}$ est une famille dénombrable d'événements deux à deux incompatibles puisque les événements $(X_1=x)_{x\in\operatorname{Im}X_1}$ sont deux à deux incompatibles. Alors :

$$\begin{aligned}
\mathbb{P}(X_2=x_2,\dots,X_n=x_n)&=\mathbb{P}(X_1\in\operatorname{Im}X,X_2=x_2,\dots,X_n=x_n) && \text{car } (X_1\in\operatorname{Im}X)=\Omega\\
&=\mathbb{P}\Big(\bigcup_{x\in\operatorname{Im}X_1}(X_1=x,X_2=x_2,\dots,X_n=x_n)\Big)\\
&=\sum_{x\in\operatorname{Im}X_1}\mathbb{P}(X_1=x,X_2=x_2,\dots,X_n=x_n) && \text{par }\sigma\text{-additivité}\\
&=\sum_{x\in\operatorname{Im}X_1}\mathbb{P}(X_1=x)\times\mathbb{P}(X_2=x_2)\times\cdots\times\mathbb{P}(X_n=x_n) && \text{par indépendance}\\
&=\underbrace{\left(\sum_{x\in\operatorname{Im}X_1}\mathbb{P}(X_1=x)\right)}_{=1}\times\mathbb{P}(X_2=x_2)\times\cdots\times\mathbb{P}(X_n=x_n) && \text{par linéarité de }\textstyle\sum\\
&=\mathbb{P}(X_2=x_2)\times\cdots\times\mathbb{P}(X_n=x_n)
\end{aligned}$$

Les variables $X_2,\dots,X_n$ sont donc indépendantes. ∎

En appliquant le résultat aux sous-familles de deux variables :

Corollaire 20. (Mutuellement indépendants $\implies$ 2 à 2 indépendants)
Soit $I\subset\mathbb{N}$ *; si* $(X_n)_{n\in I}$ *est une famille de variables mutuellement indépendantes, alors les variables sont deux à deux indépendantes, i.e. :*
$$\forall(i,j)\in I^2,\ i\neq j\implies X_i\perp\!\!\!\perp X_j.$$

L'indépendance de variables aléatoires assure bien sûr de l'indépendance des événements décrits par ces variables aléatoires.

PROPOSITION 21.
Si les variables aléatoires $X_1, \dots, X_n$ sont indépendantes, alors pour tout $A_1 \subset \operatorname{Im} X_1, \dots, A_n \subset \operatorname{Im} X_n$:

$$\mathbb{P}((X_1 \in A_1) \cap \cdots \cap (X_n \in A_n)) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i)$$

autrement dit, les événements $(X_1 \in A_1), \dots, (X_n \in A_n)$ sont mutuellement indépendants.

Démonstration. Sous ces hypothèses :

$$\bigcap_{i=1}^n (X_i \in A_i) = \bigcup_{(x_1,\dots,x_n)\in A_1\times\cdots\times A_n} ((X_1 = x_1) \cap \cdots \cap (X_n = x_n))$$

Puisque les ensembles $A_1, \dots, A_n$ sont dénombrables, leur produit cartésien est aussi dénombrables. Ainsi $((X_1 = x_1) \cap \cdots \cap (X_n = x_n))_{(x_1,\dots,x_n)\in A_1\times\cdots\times A_n}$ est une famille dénombrable d'événements ; qui plus est deux à deux incompatibles, puisque $x_i \neq x_i' \implies (X_i = x_i) \cap (X_i = x_i') = \varnothing$. Donc par σ-additivité :

$$\begin{aligned}
\mathbb{P}\Big(\bigcap_{i=1}^n (X_i \in A_i)\Big) &= \sum_{(x_1,\dots,x_n)\in A_1\times\cdots\times A_n} \mathbb{P}((X_1 = x_1) \cap \cdots \cap (X_n = x_n)) \\
&= \sum_{(x_1,\dots,x_n)\in A_1\times\cdots\times A_n} \mathbb{P}(X_1 = x_1) \times \cdots \times \mathbb{P}(X_n = x_n) && \text{par ind. de } X_1, \dots, X_n \\
&= \sum_{x_1\in A_1} \sum_{(x_2,\dots,x_n)\in A_2\times\cdots\times A_n} \mathbb{P}(X_1 = x_1) \times \mathbb{P}(X_2 = x_2) \cdots \times \mathbb{P}(X_n = x_n) && \text{sommation par paquet} \\
&= \left(\sum_{x_1\in A_1} \mathbb{P}(X_1 = x_1)\right) \times \sum_{(x_2,\dots,x_n)\in A_2\times\cdots\times A_n} \mathbb{P}(X_2 = x_2) \cdots \times \mathbb{P}(X_n = x_n) \\
&= \mathbb{P}(X_1 \in A_1) \times \sum_{(x_2,\dots,x_n)\in A_2\times\cdots\times A_n} \mathbb{P}(X_2 = x_2) \cdots \times \mathbb{P}(X_n = x_n) \\
&= \quad \vdots && \text{de même} \dots \\
&= \mathbb{P}(X_1 \in A_1) \times \mathbb{P}(X_2 \in A_2) \times \cdots \times \mathbb{P}(X_n \in A_n)
\end{aligned}$$

■

Un résultat important pour établir l'indépendance est le lemme des coalitions.

PROPOSITION 22. (Lemme des coalitions)
Soit $X_1, \dots, X_n$ des variables aléatoires mutuellement indépendantes.

- *Si pour tout $i \in [\![1, n]\!]$, f_i est une fonction définie sur $\operatorname{Im} X_i$ alors les variables aléatoires $f_1(X_1), \dots, f_n(X_n)$ sont mutuellement indépendantes.*
- *Si f et g sont définies respectivement sur $\operatorname{Im} X_1 \times \cdots \times \operatorname{Im} X_p$ et $\operatorname{Im} X_{p+1} \times \cdots \times \operatorname{Im} X_n$ alors les variables $f(X_1, \dots, X_p)$ et $g(X_{p+1}, \dots, X_n)$ sont indépendantes.*

On peut aussi regrouper par paquets :

- *Soit $\tau : [\![1, p+1]\!] \longrightarrow [\![1, n]\!]$ une application strictement croissante, et $f_1, \dots, f_p$ des applications où pour tout $i \in [\![1, p]\!]$, f_i et définie sur $\operatorname{Im} X_{\tau(i)} \times \cdots \times \operatorname{Im} X_{\tau(i)-1}$, alors les variables de la famille $\big(f_i(X_{\tau(i)}, \dots, X_{\tau(i)-1})\big)_{i\in[\![1,p]\!]}$ sont mutuellement indépendantes.*

Démonstration. (Esquisse) En guise d'exemple on démontre le premier point ; les suivants ne présentent pas de difficulté supplémentaire en dehors du formalisme.
Soit pour tout $i \in [\![1, n]\!]$, $a_i \in f_i(\operatorname{Im} X_i)$.

$$\begin{aligned}
\mathbb{P}\Big(\bigcap_{i=1}^n f_i(X_i) = a_i\Big) &= \mathbb{P}\Big(\bigcap_{i=1}^n X_i \in f_i^{-1}\{a_i\}\Big) \\
&= \prod_{i=1}^n \mathbb{P}\left(X_i \in f_i^{-1}\{a_i\}\right) && \text{par indépendance avec la proposition 21} \\
&= \prod_{i=1}^n \mathbb{P}(f_i(X_i) = a_i)
\end{aligned}$$

d'où l'indépendance des $f_i(X_i)$. ■

Exemple. Soient Soit $(X_k)_{1\leqslant k\leqslant n}$ une famille de variables aléatoires réelles mutuellement indépendantes, avec $n \geqslant 5$. Alors $Y_1 = X_1 + X_2$, $Y_2 = X_3 + \cdots + X_{n-2}$ et $Y_3 = X_{n-1}X_n$ sont mutuellement indépendantes.

1.4. **Variables aléatoires réelles discrètes usuelles.**

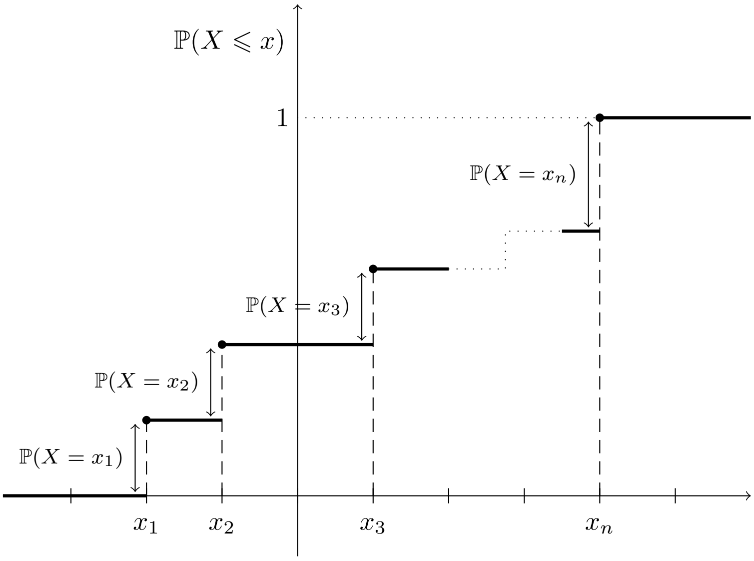
Rappelons qu'une variable aléatoire réelle discrète est dite réelle lorsqu'elle est à valeur dans $\mathbb{R}$.

Pour une variable aléatoire réelle discrète (en abrégé V.A.R.D ou plus simplement V.A.R.), on peut définir sa fonction de répartition :

DÉFINITION 13. (Fonction de répartition d'une V.A.R.D.)
Soit $X : \Omega \longrightarrow \mathbb{R}$ une variable aléatoire réelle discrète. Sa fonction de répartition est l'application :

$$\begin{array}{rcl} F_X : \mathbb{R} & \longrightarrow & [0,1] \\ x & \longmapsto & \mathbb{P}(X \leqslant x) \end{array}$$

Rappelons qu'elle a comme propriétés (triviales) :
- F_X est croissante,
- $\lim\limits_{x \to -\infty} F_X(x) = 0$, $\lim\limits_{x \to +\infty} F_X(x) = 1$,
- $\forall (a,b) \in \mathbb{R}^2$, avec $a \leqslant b$, $\mathbb{P}(a < X \leqslant b) = F_X(b) - F_X(a)$.



Somme et produit de variables aléatoires discrètes réelles sur un même espace probabilisé sont des variables discrètes réelles sur ce même espace probabilisé.

PROPOSITION 23. (Somme et produit de variables aléatoires discrètes réelles)
Soient X et Y deux variables aléatoires réelles sur un même espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$. Alors $X + Y$ est $X \times Y$ sont aussi des variables aléatoires discrètes réelles sur $(\Omega, \mathscr{A}, \mathbb{P})$.

Démonstration. Notons :

$$\begin{array}{rcl} S : \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ (x,y) & \longmapsto & x+y \end{array} \quad \text{et} \quad \begin{array}{rcl} P : \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ (x,y) & \longmapsto & x \times y \end{array}$$

Alors $X + Y = S \circ (X,Y)$ et $X \times Y = P \circ (X,Y)$ sont des variables aléatoires discrètes (réelles) d'après la proposition 13. ■

Commençons par rappeler les lois usuelles vues en première année ; elles sont toutes à valeurs finies.

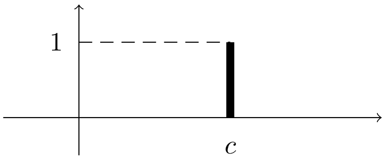
1.4.1. *Loi certaine.*

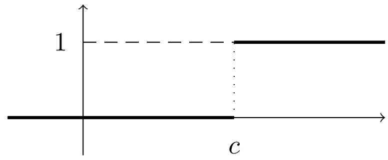
Loi certaine.
Une V.A.R. certaine est une V.A.R. dont l'univers image est un singleton : $X(\Omega) = \{c\}$.
On dit que X suit une loi certaine. L'événement $(X = c)$ est l'événement certain.

Remarque. Réciproquement toute application constante X définie sur Ω est clairement une variable aléatoire, et elle suit une loi certaine.

• Situation type : Lorsque l'issue est certaine : une seule issue.

• Histogramme et fonction de répartition.





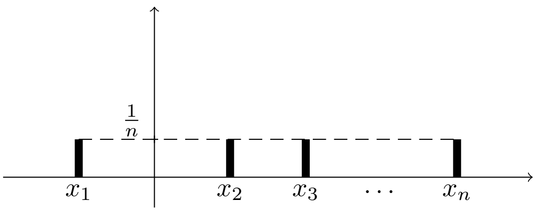
1.4.2. *Loi uniforme.*

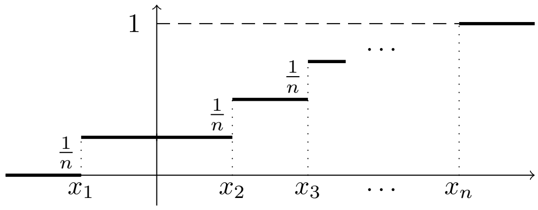
Loi uniforme.
Soit $E = \{x_1, x_2, \dots, x_n\}$ une partie finie de $\mathbb{R}$ de cardinal n. Une V.A.R. X suit la loi uniforme sur E si :
- $X(\Omega) = E$,
- $\forall k \in [\![1, n]\!]$, $\mathbb{P}(X = x_k) = 1/n$.

Dans ce cas on écrit : $X \hookrightarrow \mathscr{U}(E)$. Lorsque $E = [\![1, n]\!]$ on note plus simplement $X \hookrightarrow \mathscr{U}(n)$.

• Situation type
On choisit au hasard (de manière équiprobable) un objet parmi n. Par exemple : on lance une pièce équilibré ; on lance un dé 6 non pipé ; on tire au hasard une boule dans une urne...

• Histogramme et fonction de répartition





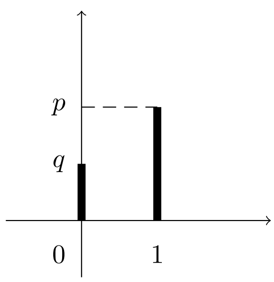
1.4.3. *Loi de Bernoulli.*

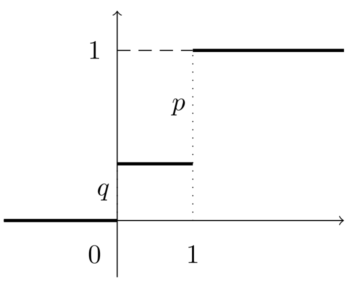
Loi de Bernoulli.
Une V.A.R. X est dite <u>de Bernoulli</u> lorsqu'elle ne prend que deux valeurs 0 et 1, et avec des probabilités non nulles. Ainsi X suit une loi de Bernoulli si :
- $X(\Omega) = \{0, 1\}$,
- $\mathbb{P}(X = 1) = p \in]0, 1[$
- $\mathbb{P}(X = 0) = 1 - p$.

On note alors $X \hookrightarrow \mathscr{B}(p)$; $p \in]0, 1[$ est la paramètre de la loi de Bernoulli.

• Situation type : un choix binaire, vrai/faux, codifié par 1/0.

• Histogramme et fonction de répartition





• Soit A un événement d'un espace probabilisé avec $p = \mathbb{P}(A) \in]0, 1[$. L'indicatrice $\mathbb{1}_A$ de A dans Ω :

$$\begin{array}{rcl} \mathbb{1}_A : \Omega & \longrightarrow & \mathbb{R} \\ \omega & \longmapsto & \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon} \end{cases} \end{array}$$

suit une loi de Bernoulli de paramètre p.

Réciproquement, si X suit une loi de Bernoulli de paramètre p, en posant $A = (X = 1)$, X est l'indicatrice $\mathbb{1}_A$ de A dans Ω.

Une variable de Bernoulli X, est l'indicatrice $\mathbb{1}_A$ de l'évènement $A = (X = 1)$.

1.4.4. *Loi binomiale.*

Loi binomiale.
On dit qu'une V.A.R. suit la loi binomiale de paramètres n et p si :
- $X(\Omega) = [\![0, n]\!]$,
- $\forall k \in [\![0, n]\!], \ \mathbb{P}(X = k) = \binom{n}{k} p^k q^{n-k}$

en notant $q = 1 - p$. On note alors $X \hookrightarrow \mathscr{B}(n, p)$.

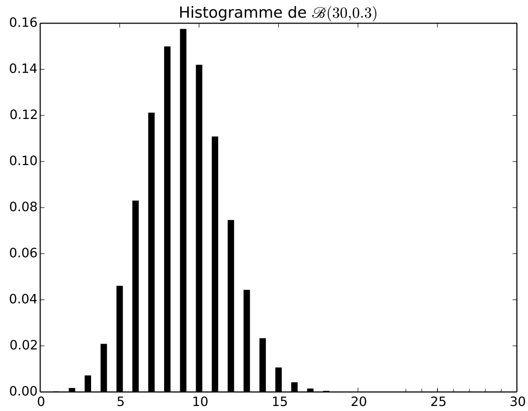
Remarque. Avec la formule du binôme, on retrouve bien que :

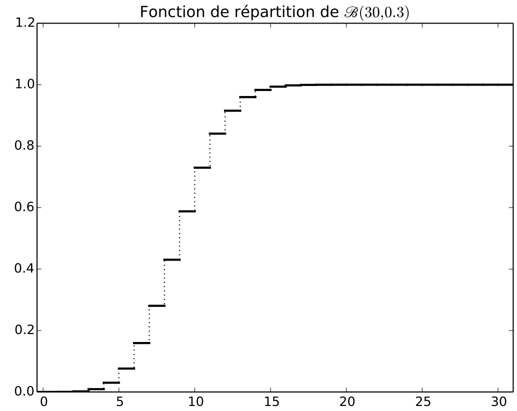
$$\sum_{k=0}^{n} \mathbb{P}(X = k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p + q)^n = 1^n = 1.$$

• Histogramme et fonction de répartition

L'histogramme d'une loi binomiale s'appuie sur une courbe en cloche, centrée lorsque $p = \frac{1}{2}$, et décalé vers la gauche (respectivement vers la droite) lorsque $p < \frac{1}{2}$ (respectivement $p > \frac{1}{2}$).

Par exemple, pour $n = 30$ et $p = \frac{1}{3}$:





• Situation type : schéma de Bernoulli ; nombre de succés.

On considère une expérience qui conduit à deux issues possibles : le succès S avec probabilité $p \in]0, 1[$, et l'échec E avec probabilité $q = 1 - p \in]0, 1[$. On répète cette expérience n fois (où $n \in \mathbb{N}^*$), de manière indépendante ; soit X le nombre de succès obtenus.

Un tel contexte s'appelle un <u>schéma de Bernoulli</u> : répéter de manière indépendante une expérience menant à deux issues possibles succès/échec ; ici on répète l'expérience n fois et X compte le nombre de succès.

– Espace probabilisé : On prend comme univers $\Omega = \{S, E\}^n$, l'ensemble des n-listes de $\{S, E\}$. Chaque issue est un élément du produit cartésien de $\{S, E\}^n$; si elle comporte exactement $k \in [\![0, n]\!]$ succès, par indépendance des épreuves, sa probabilité est $p^k q^{n-k}$.

– Loi de X : X peut prendre toute valeur entière entre 0 (aucun succès) et n (que des succès). Son image est donc $X(\Omega) = [\![0, n]\!]$.

Soit $k \in [\![0, n]\!]$; calculons $\mathbb{P}(X = k)$:

L'événement $(X = k)$ se partitionne en tous les événement élémentaires ayant exactement k succès ; chacun de ces événements élémentaires a pour probabilité $p^k q^{n-k}$. Or il y en a exactement $\binom{n}{k}$; en effet une issue comportant exactement k succès est déterminée par les rangs d'apparition des succès (aux 1ère, 2ème, ..., ou n-ième épreuves). Il y en a donc autant que de k-combinaisons dans $[\![1, n]\!]$. Ainsi :

$$\forall k \in [\![0, n]\!],\ \mathbb{P}(X = k) = \binom{n}{k} p^k q^{n-k}$$

Cela détermine la loi de X.

Un résultat important est que toute loi binomiale est somme de variables de bernoulli i.i.d.

Proposition 24. Loi de la somme de V.A.R. de Bernoulli indépendantes et de même paramètre.
Soit $X_1, X_2, \ldots, X_n$ des V.A.R. mutuellement indépendantes et suivant toutes une même loi de Bernoulli $\mathscr{B}(p)$. Alors leur somme $X_1 + X_2 + \cdots + X_n$ suit la loi binomiale $\mathscr{B}(n, p)$.

Démonstration. Considérons $X_1, X_2, \ldots, X_n$ des V.A.R. sur Ω mutuellement indépendantes et suivant toutes une même loi de Bernoulli $\mathscr{B}(p)$. Déterminons la loi de leur somme $Y = X_1 + X_2 + \cdots + X_n$.
Puisque $\forall k \in [\![1, n]\!]$, $X_k(\Omega) = \{0, 1\}$, $Y(\Omega) \subset [\![0, n]\!]$. Soit $k \in [\![0, n]\!]$:

$$\begin{aligned}(Y = k) &= \text{"exactement } k \text{ des } X_i \text{ valent 1, } n-k \text{ valent 0"}\\ &= \bigcup_{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n} \left(\bigcap_{i \in \{i_1, \ldots, i_k\}} (X_i = 1) \cap \bigcap_{i \notin \{i_1, \ldots, i_k\}} (X_i = 0) \right)\end{aligned}$$

C'est une réunion d'événements 2 à 2 incompatibles, donc :

$$\mathbb{P}(Y = k) = \sum_{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n} \mathbb{P} \left(\bigcap_{i \in \{i_1, \ldots, i_k\}} (X_i = 1) \cap \bigcap_{i \notin \{i_1, \ldots, i_k\}} (X_i = 0) \right)$$

par indépendance mutuelle

$$\begin{aligned} &= \sum_{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n} \prod_{i \in \{i_1, \ldots, i_k\}} \mathbb{P}(X_i = 1) \times \prod_{i \notin \{i_1, \ldots, i_k\}} \mathbb{P}(X_i = 0)\\ &= \sum_{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n} p^k q^{n-k}\\ &= \binom{n}{k} p^k q^{n-k}\end{aligned}$$

puisqu'il y a exactement $\binom{n}{k}$ k-combinaisons $\{i_1, i_2, \ldots, i_k\}$ dans $[\![1, n]\!]$. Ainsi Y suit la loi binomiale $\mathscr{B}(n, p)$. ∎

Il faut savoir reconnaître les lois usuelles.

Exercice 2. Dans chacune des expériences suivantes, reconnaître la loi suivie par la V.A.R. X et préciser ses paramètres ainsi que son image $X(\Omega)$.

1. On range au hasard 20 objets dans 3 tiroirs. Soit X le nombre d'objets dans le premier tiroirs.
2. Un enclos contient 15 lamas, 15 dromadaires et 15 chameaux. On sort un animal au hasard de cet enclos. Soit X le nombre de bosses de cet animal.
3. On choisit au hasard un nombre n entre 0 et 100 ; soit X le reste de n dans la division euclidienne par 2.
4. On prend un jeu de 32 cartes mélangées. On retourne une par une les cartes jusqu'à l'apparition de l'as de coeur. Soit X le nombre de cartes que l'on a retournées.
5. On suppose que la probabilité de naissance d'une fille et d'un garçon sont identiques. Soit X le nombre de garçons dans une famille de 3 enfants.

Résolution.

1.4.5. *Loi géométrique.*

Loi géométrique.
On dit qu'une variable aléatoire X suit une loi de géométrique de paramètre $p \in]0,1[$ et on note $X \hookrightarrow \mathscr{G}(p)$ lorsque $X(\Omega) = \mathbb{N}^*$ et

$$\forall k \in \mathbb{N}^*,\ \mathbb{P}(X = k) = (1-p)^{k-1}p.$$

Remarques.
• On vérifie que :

$$\mathbb{P}(X \in \mathbb{N}^*) = \sum_{k=1}^{+\infty} \mathbb{P}(X = k) = \sum_{k=1}^{+\infty} q^{k-1}p = p\sum_{k=0}^{+\infty} q^k = \frac{p}{1-q} = \frac{p}{p} = 1$$

par convergence de la série géométrique $\sum q^k$; d'où le nom.
• On parlera aussi de loi géométrique lorsque $\mathbb{N}^* \subsetneq X(\Omega)$. Dans ce cas $\mathbb{P}(X \notin \mathbb{N}^*) = 1 - \mathbb{P}(X \in \mathbb{N}^*) = 0$; l'événement $X \in X(\Omega) \setminus \mathbb{N}^*$ est négligeable.

• Situation type : Schéma de Bernoulli ; rang d'apparition du premier succès.

On répète une même expérience conduisant à deux issues possibles : le succès S avec probabilité $p \in]0,1[$ et l'échec E avec probabilité $q = 1-p$. On répète cette expérience de manière indépendante jusqu'à l'apparition du premier succès ; soit X son rang d'apparition ; s'il n'apparaît jamais, X prend la valeur $+\infty$. Alors $\operatorname{Im} X = \mathbb{N}^* \cup \{+\infty\}$.

On prend comme univers $\Omega = \{S,E\}^{\mathbb{N}^*}$ l'ensemble des suites à valeur dans $\{S,E\}$, et comme tribu $\mathscr{A}$, la plus petite (au sens de l'inclusion) contenant pour tout $i \in \mathbb{N}^*$, les événements :
S_i (respectivement E_i) : "La i-ème expérience a conduit à un succès (respectivement un échec)"

$$S_i = \{(u_1,\dots,u_n,\dots) \in \{S,E\}^{\mathbb{N}^*} \mid u_i = S\}$$
$$E_i = \{(u_1,\dots,u_n,\dots) \in \{S,E\}^{\mathbb{N}^*} \mid u_i = E\}$$

Alors pour tout $k \in \mathbb{N}^*$:

$$(X = k) = \left(\bigcap_{i=1}^{k-1} E_i\right) \cap S_k$$

et par indépendance :

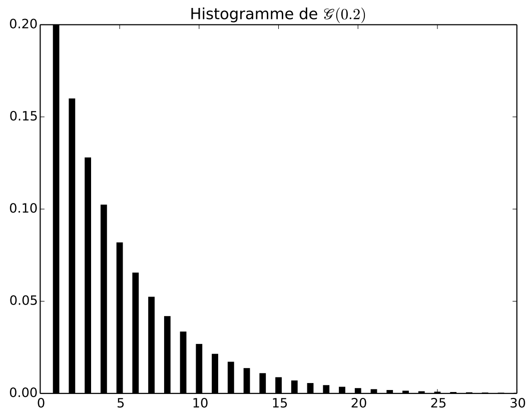
$$\mathbb{P}(X = k) = \left(\prod_{i=1}^{k-1} \mathbb{P}(E_i)\right) \times \mathbb{P}(S_k) = q^{k-1}p$$

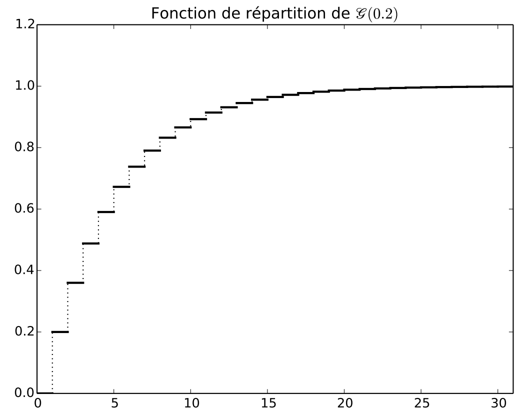
et donc $X \hookrightarrow \mathscr{G}(p)$.
Remarquons, s'il était besoin, que par continuité décroissante :

$$\mathbb{P}(X = \infty) = \mathbb{P}(X \notin \mathbb{N}^*) = \mathbb{P}\left(\bigcap_{i=1}^{+\infty} E_i\right) = \lim_{n\to+\infty} \mathbb{P}\left(\bigcap_{i=1}^{n} E_i\right) = \lim_{n\to+\infty} q^n = 0$$

• Histogramme et fonction de répartition

Par exemple, pour $p = \dfrac{2}{10}$:





Exercice 3.
On lance une pièce équilibrée jusqu'à l'obtention du premier pile au rang N ; on choisit alors aléatoirement une entier entre 1 et N. Quelle est la probabilité qu'on choisisse le 1 ?

Résolution.

Propriété 25.
Soit $X \hookrightarrow \mathscr{G}(p)$; alors pour tout $k \in \mathbb{N}$,

$$\mathbb{P}(X > k) = (1-p)^k$$

Démonstration. On a :

$$(X > k) = \left(\bigcup_{i=k+1}^{+\infty} (X = i)\right) \cup (X > k \cap X \notin \mathbb{N}^*)$$

C'est une réunion dénombrable d'événements deux à deux incompatibles (et le dernier est négligeable), et donc par σ-additivité :

$$\mathbb{P}(X > k) = \sum_{i=k+1}^{+\infty} \mathbb{P}(X = i) = \sum_{i=k+1}^{+\infty} (1-p)^{i-1} p = p(1-p)^k \times \frac{1}{1-(1-p)} = (1-p)^k$$

comme somme d'une série géométrique. ■

Remarque. Cette propriété est même caractéristique d'une loi géométrique puisque pour tout $k \in \mathbb{N}^*$:

$$\mathbb{P}(X = k) = \mathbb{P}(X > k-1) - \mathbb{P}(X > k) = (1-p)^{k-1} - (1-p)^k = (1-p)^{k-1} p$$

Exercice 4. Soient X_1 et X_2 deux variables aléatoires indépendantes suivant chacune une loi géométrique de paramètre respectif p_1, p_2. Déterminer la loi de la variable $\min(X_1, X_2)$.

Résolution.

Une propriété importante (bien que hors-programme) qu'il est bon de connaître, est que la loi géométrique est "sans mémoire" :

Proposition 26. (Loi sans mémoire)
Si X suit une loi géométrique, alors pour tout $k \geqslant 0$ et $n \geqslant 0$,

$$\mathbb{P}(X > n+k \mid X > k) = \mathbb{P}(X-k > n \mid X > k) = \mathbb{P}(X > n)$$

Informellement, sachant que $X > k$, $X-k$ suit la même loi que X : on peut oublier les résultats des k premières expériences et recommencer à zéro. (D'où l'expression de loi sans mémoire).

Démonstration. D'après la propriété précédente, pour tout $n \geqslant 0$, $\mathbb{P}(X > n) = q^n$ avec $q = 1-p$. Pour tout $k \geqslant 0$ et $n \geqslant 0$

$$\mathbb{P}(X > n+k \mid X > k) = \frac{\mathbb{P}((X > n+k) \cap (X > k))}{\mathbb{P}(X > k)} = \frac{\mathbb{P}(X > n+k)}{\mathbb{P}(X > k)} = \frac{q^{n+k}}{q^k} = q^n = \mathbb{P}(X > n).$$

∎

De plus cette propriété est caractéristique de la loi géométrique :

Remarque. Réciproquement, si $X : \Omega \to \mathbb{N}^*$ vérifie pour tout $k \geqslant 0$ et $n \geqslant 0$,

$$\mathbb{P}(X > k) \neq 0 \quad \text{et} \quad \mathbb{P}(X > n+k \mid X > k) = \mathbb{P}(X > n)$$

alors X suit une loi géométrique de paramètre $p = \mathbb{P}(X = 1)$. En effet :
Notons $q = 1-p = \mathbb{P}(X > 1) \neq 0$ et soit $n \geqslant 0$. Puisque $(X > (n+1))$ implique $(X > 1)$:

$$\begin{aligned}\mathbb{P}(X > n+1) &= \mathbb{P}((X > n+1) \cap (X > 1)) \\ &= \mathbb{P}(X > n+1 \mid X > 1) \times \mathbb{P}(X > 1) = \mathbb{P}(X > n) \times \mathbb{P}(X > 1) = q \times \mathbb{P}(X > n).\end{aligned}$$

On a donc immédiatement par récurrence $\mathbb{P}(X > n) = q^n$ pour $n \geqslant 1$, valable également pour $n = 0$.
Il vient pour tout $n \geqslant 1$,

$$\mathbb{P}(X = n) = \mathbb{P}(X > n-1) - \mathbb{P}(X > n) = q^{n-1} - q^n = q^{n-1}(1-q) = (1-p)^{n-1}p.$$

Donc $X \hookrightarrow \mathscr{G}(p)$.

1.4.6. *Loi de Poisson.*

Elle décrit le nombre d'événements se produisant dans un intervalle de temps fixé, avec une fréquence λ connue, et indépendamment du temps écoulé depuis l'événement précédent. On l'utilise notamment pour prédire le nombre de clients se présentant chaque jour à un guichet, la fréquentation moyenne étant connue, ou le nombre de vols, ou d'accidents, *etc.* survenant chaque jour.

Loi de Poisson.
On dit qu'un variable aléatoire X suit une loi de Poisson de paramètre $\lambda > 0$ que l'on note $X \hookrightarrow \mathscr{P}(\lambda)$ lorsque $\operatorname{Im} X = \mathbb{N}$ *et*

$$\forall k \in \mathbb{N},\ \mathbb{P}(X = k) = e^{-\lambda} \times \frac{\lambda^k}{k!}$$

Remarque. On a bien :

$$\mathbb{P}(X \in \mathbb{N}) = \sum_{k=0}^{+\infty} e^{-\lambda} \times \frac{\lambda^k}{k!} = e^{-\lambda} \times \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \times e^{\lambda} = 1$$

par convergence de la série exponentielle.

Afin de bien comprendre la situation type d'application, montrons le résultat suivant qui établit qu'une loi de Poisson est limite de lois binomiales :

Théorème 27. (Approximation de la loi binomiale par une loi de Poisson)
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires telles que $\forall n \in \mathbb{N}$, X_n suit la loi binomiale $\mathscr{B}(n, p_n)$. On suppose qu'il existe $\lambda > 0$ tel que :

$$p_n \underset{+\infty}{\sim} \frac{\lambda}{n}$$

alors pour tout $k \in \mathbb{N}$:

$$\lim_{n\to+\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!}$$

Remarque. Ainsi, la suite des lois $\mathscr{B}\left(n, \frac{\lambda}{n}\right)$ tend lorsque $n \to +\infty$ vers la loi de Poisson $\mathscr{P}(\lambda)$.

Démonstration. En effet, soit $k \in \mathbb{N}$, pour $n \geqslant k$:

$$\mathbb{P}(X = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{1}{k!}\,\frac{n!}{(n-k)!}\, p_n^k (1-p_n)^{n-k}$$

On va montrer que :

$$(1-p_n)^{n-k} \underset{n\to+\infty}{\longrightarrow} e^{-\lambda} \qquad \frac{n!}{(n-k)!}\, p_n^k \underset{n\to+\infty}{\longrightarrow} \lambda^k$$

on aura alors la limite souhaitée.
D'une part :

$$(n-k)\ln(1-p_n) \underset{+\infty}{\sim} -(n-k)\frac{\lambda}{n} \underset{+\infty}{\sim} -\lambda$$

donc $(1-p_n)^{n-k} = \exp\left((n-k)\ln(1-p_n)\right) \underset{n\to+\infty}{\longrightarrow} e^{-\lambda}$ par composition des limites.
D'autre part par produit et puissance d'équivalents :

$$p_n^k \underset{+\infty}{\sim} \frac{\lambda^k}{n^k} \quad \text{et} \quad \frac{n!}{(n-k)!} = \prod_{j=0}^{k-1}(n-j) \underset{n\to+\infty}{\sim} n^k$$

donc $\dfrac{n!}{(n-k)!}\, p_n^k \underset{n\to+\infty}{\longrightarrow} \lambda^k$. ■

La loi de Poisson $\mathscr{P}(\lambda)$ est une "bonne approximation" d'une loi binomiale $\mathscr{B}(n, p_n)$ associée :
– à un grand nombre de tentatives (n est grand),
– de probabilité de succès très faible (p_n est petit),
– de sorte que np_n soit proche de λ.
Elle permet donc de comptabiliser la survenue d'événements rares. On dit que c'est la "loi des événements rares".

Exemple. Une entreprise fabrique une grand nombre de pièces identiques, par exemple 10000, avec 2% de pièces défectueuses.
On choisit 100 pièces au hasard. Quelle est la probabilité $\mathbb{P}$ d'en avoir exactement 4 défectueuses ?
En théorie. $\mathbb{P} = \dfrac{\binom{200}{4}\binom{9800}{96}}{\binom{10000}{100}} \simeq 0,09018$ mais on peut approcher par un tirage avec remise (le nombre de pièces choisies étant faible, la probabilité de choisir plusieurs fois la même est très petite) et donc par une loi binomiale $\mathscr{B}(100; 0,02)$; $\mathbb{P} = \binom{100}{4} p^4(1-p)^{96} \simeq 0,09021$ avec $p = 0,02, n = 100$.
Comme p est faible, que n est assez grand et que $\lambda = np = 100 \times 0,02 = 2$, on peut approximer par une loi de Poisson, beaucoup plus simple à calculer : $\mathbb{P} = e^{-\lambda}\frac{\lambda^4}{4!} = e^{-2}\frac{2^4}{4!} \simeq 0,09022$.

• Situation type.

Supposons que l'on souhaite modéliser le nombre de personnes se présentant à un guichet dans une période d'une heure. On sait que la fréquentation moyenne est de 30 personnes/heure.
– On peut aborder le problème en décomposant minute par minute et en supposant qu'à chaque minute, avec probabilité 1/2 une personne se présente, avec probabilité 1/2 aucune ne se présente. Le nombre de personnes se présentant par heure pour ce modèle suit alors la loi binomiale $\mathscr{B}(60, 1/2)$ qui donne bien une fréquentation moyenne de 30 personnes/heure. Mais ce modèle est très simpliste puisqu'il impose :
— qu'au plus une personne se présente chaque minute,
— au plus 60 personnes se présentent chaque heure.

– Pour rendre le modèle plus réaliste on peut songer à faire un découpage seconde par seconde, une personne se présentant chaque seconde avec probabilité $\frac{1}{120}$. On aboutirait à une loi binomiale $\mathscr{B}(3600, 1/120)$, avec encore une fréquentation moyenne de 30 personnes/heure, et les mêmes limitations, bien que beaucoup moins restrictives.

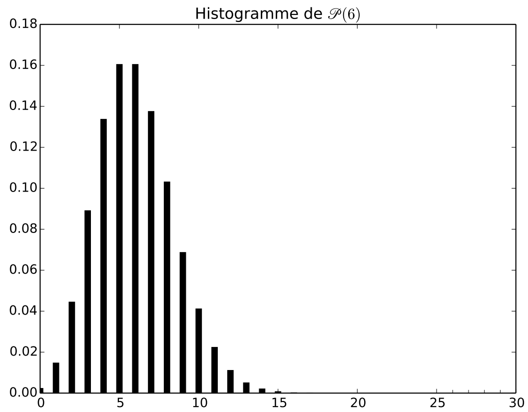
– Afin de rendre le modèle plus réaliste, il suffit d'affiner encore le découpage. Avec le théorème précédent on se rapproche alors de la loi de Poisson $\mathscr{P}(30)$, qui s'avère donc le choix le plus pertinent, puisqu'il ne présente plus les limitations précédentes.

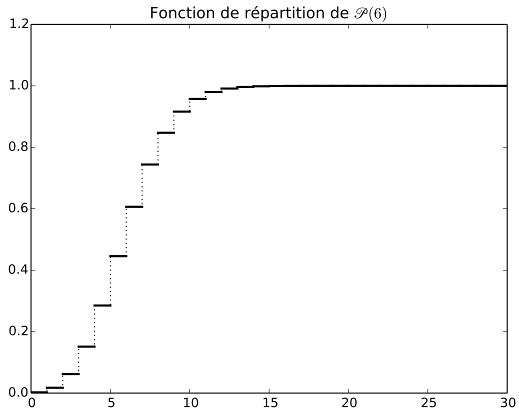
Plus généralement, la loi de Poisson $\mathscr{P}(\lambda)$ s'utilise pour modéliser un flux d'arrivée durant une période T donnée, homogène dans le temps (une arrivée ne dépendant pas du délai depuis la dernière arrivée), et dont le flux moyen λ sur la période T est connu, comme par exemple :

– Nombre d'appels quotidien à un standard téléphonique,
– Fréquentation quotidienne d'un magasin,
– Nombre mensuel de naissances,
– Nombre mensuel d'accidents routiers, *etc.*

• Histogramme et fonction de répartition

Par exemple, pour $\lambda = 6$:





Exercice 5.
Montrer que si X_1 et X_2 sont des variables aléatoires indépendantes suivant des lois de Poisson de paramètre respectif λ, μ, alors $X_1 + X_2$ suit la loi de Poisson de paramètre $\lambda + \mu$.

Résolution.

2. Espérance et variance

Dans toute cette partie, les variables sont à valeurs réelles ou complexes, voire à valeurs uniquement réelles dans la seconde partie. Afin de ne pas restreindre le champ d'application, les variables à valeurs réelles positives pourront prendre leur valeur dans $\overline{\mathbb{R}_+} = \mathbb{R}_+ \cup \{+\infty\}$ (penser par exemple au rang d'apparition du premier pile de la loi géométrique).

2.1. Espérance (VAD réelle ou complexe).

Définition 14. (Espérance d'une VAD à valeur dans $\mathbb{R}$ ou $\mathbb{C}$)
• *Une variable aléatoire discrète* X *à valeurs dans* $\mathbb{R}$ *ou* $\mathbb{C}$ *est dite d'espérance finie lorsque la famille* $(x \times \mathbb{P}(X = x))_{x \in \operatorname{Im} X}$ *est sommable.*
• *En se donnant* $\operatorname{Im} X$ *sous la forme* $\operatorname{Im} X = \{x_n \in \mathbb{C} \mid n \in \mathbb{N}\}$, X *est d'espérance finie lorsque la série* $\sum_{n \geqslant 0} x_n \mathbb{P}(\mathrm{X} = x_n)$ *est absolument convergente.*
• *Si tel est le cas, on appelle espérance de* X*, noté* $\mathbb{E}(\mathrm{X})$*, le réel ou complexe :*

$$\mathbb{E}(\mathrm{X}) = \sum_{n=0}^{+\infty} x_n \underbrace{\mathbb{P}(\mathrm{X} = x_n)}_{p_n} = \sum_{n=0}^{+\infty} p_n x_n.$$

Remarques.
• L'espérance de la variable aléatoire X est donc la moyenne des valeurs prises par X pondérée par leur probabilité d'apparition.
• Lorsque $\mathbb{E}(X) = 0$, on parle de variable aléatoire centrée.
• Lorsque X est à valeurs dans $\mathbb{R}_+$, il arrive que l'on note $\mathbb{E}(X) = +\infty$ lorsque la variable aléatoire n'admet pas d'espérance (et que, donc, la série diverge vers $+\infty$). Plus précisément :

Définition 15. (Espérance d'une VAD à valeurs dans $\overline{\mathbb{R}_+}$)
Soit X *une VAD à valeur dans* $\overline{\mathbb{R}_+} = \mathbb{R}_+ \cup \{+\infty\}$*. On appelle espérance de* X*, notée* $\mathbb{E}(X)$ *:*

$$\mathbb{E}(X) = \sum_{x \in \operatorname{Im} X} x \times \mathbb{P}(X = x)$$

avec les conventions :
• *si* $\mathbb{P}(X = +\infty) = 0$*, alors* $x \times \mathbb{P}(X = +\infty) = 0$*,*
• $\mathbb{E}(X) = +\infty$ *lorsque la famille* $(x \times \mathbb{P}(X = x))_{x \in \operatorname{Im} X}$ *n'est pas sommable.*

Pour les variables à valeurs dans $\overline{\mathbb{N}}$, et seulement pour celles là, l'espérance admet l'écriture suivante.

Propriété 28. (Espérance d'une variable à valeurs dans $\overline{\mathbb{N}}$)
Soit X *une VAD à valeurs dans* $\overline{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$*, i.e.* $\operatorname{Im} X \subset \mathbb{N} \cup \{+\infty\}$*. Alors :*

$$\mathbb{E}(X) = \sum_{n \in \mathbb{N}} \mathbb{P}(X > n)$$

Remarque. La famille peut ne pas être sommable, dans quel cas $\mathbb{E}(X) = +\infty$.

Démonstration. Montrons d'abord que lorsque X n'est pas d'espérance finie, alors $\sum \mathbb{P}(X \geqslant n)$ diverge (vers $+\infty$). Pour tout $N \in \mathbb{N}^*$, par σ-additivité :

$$\forall n \in [\![1, N]\!],\ \mathbb{P}(X \geqslant n) = \Big(\sum_{k=n}^{+\infty} \mathbb{P}(X = k)\Big) + \mathbb{P}(X = +\infty) \geqslant \sum_{k=n}^{N} \mathbb{P}(X = k)$$

d'où :

$$\sum_{n=1}^{N} \mathbb{P}(X \geqslant n) \geqslant \sum_{n=1}^{N} \sum_{k=n}^{N} \mathbb{P}(X = k) = \sum_{k=1}^{N} \sum_{n=1}^{k} \mathbb{P}(X = k) = \sum_{k=1}^{N} k\mathbb{P}(X = k)$$

Ainsi, si X n'est pas d'espérance finie, alors $\sum_{k=1}^{N} k\mathbb{P}(X = k) \underset{N \to +\infty}{\longrightarrow} +\infty$ et donc par comparaison de séries à termes positifs, $\sum_{n=1}^{N} \mathbb{P}(X \geqslant n) \underset{N \to +\infty}{\longrightarrow} +\infty$.

Supposons désormais que X est d'espérance finie ; en particulier $\mathbb{P}(X = +\infty) = 0$. Pour $N \in \mathbb{N}^*$ on a :

$$\sum_{n=0}^{N} n\,\mathbb{P}(X = n) = \sum_{n=0}^{N} n\left(\mathbb{P}(X \geqslant n) - \mathbb{P}(X \geqslant n+1)\right)$$

$$
\begin{aligned}
&= \sum_{n=0}^{N} n\,\mathbb{P}(X \geqslant n) - \sum_{n=0}^{N} n\,\mathbb{P}(X \geqslant n+1) \\
&= \sum_{n=0}^{N} n\,\mathbb{P}(X \geqslant n) - \sum_{n=1}^{N+1} (n-1)\,\mathbb{P}(X \geqslant n) \\
&= \sum_{n=0}^{N} n\,\mathbb{P}(X \geqslant n) - \sum_{n=1}^{N+1} n\,\mathbb{P}(X \geqslant n) + \sum_{n=1}^{N+1} \mathbb{P}(X \geqslant n) \\
&= -(N+1)\,\mathbb{P}(X \geqslant N+1) + \sum_{n=1}^{N+1} \mathbb{P}(X \geqslant n)
\end{aligned}
$$

$$
\Longrightarrow \sum_{n=1}^{N+1} \mathbb{P}(X \geqslant n) = \underbrace{\sum_{n=0}^{N} n\,\mathbb{P}(X = n)}_{\xrightarrow[N\to+\infty]{} \mathbb{E}(X)} + \underbrace{(N+1)\,\mathbb{P}(X \geqslant N+1)}_{\leqslant \sum_{n=N+1}^{+\infty} n\,\mathbb{P}(X=n) \xrightarrow[N\to+\infty]{} 0}
$$

puisque la série $\sum_{n\geqslant 0} n\,\mathbb{P}(X = n)$ converge, et que le second terme ci-dessus est majoré par son reste d'ordre N. Ainsi :

$$
\sum_{n=1}^{+\infty} \mathbb{P}(X \geqslant n) = \sum_{n=0}^{+\infty} n\,\mathbb{P}(X = n) = \mathbb{E}(X).
$$

■

Exemples. À connaître.

Bien sûr toute VADR à valeurs finies admet une espérance finie.

• Si X suit une loi certaine alors son espérance est $\mathbb{E}(X) = c$ où $\operatorname{Im} X = \{c\}$.

• Si $X \hookrightarrow \mathscr{U}(E)$ suit une loi uniforme alors son espérance est la moyenne de ses valeurs prises : en notant $E = \{x_1, \dots, x_n\}$:

$$
\mathbb{E}(X) = \frac{1}{n}\sum_{k=1}^{n} x_k
$$

• Si $X \hookrightarrow \mathscr{B}(p)$ suit la loi de Bernoulli de paramètre p alors $\mathbb{E}(X) = 1 \times p + 0 \times (1-p) = p$.

• Si $X \hookrightarrow \mathscr{B}(n,p)$ suit la loi binomiale de paramètres n, p alors $\mathbb{E}(X) = np$; en effet :

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{n} k \times \mathbb{P}(X = k) \\
&= \sum_{k=0}^{n} k\binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k\binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} n\binom{n-1}{k-1} p^k (1-p)^{n-k} && \text{car } \binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1} \text{ si } 1 \leqslant k \leqslant n \\
&= n\sum_{k=0}^{n-1} \binom{n-1}{k} p^{k+1} (1-p)^{n-1-k} \\
&= np\sum_{k=0}^{n-1} \binom{n-1}{k} p^{k} (1-p)^{n-1-k} \\
&= np \times (p + (1-p))^{n-1} && \text{formule du binôme} \\
&= np
\end{aligned}
$$

• Si $X \hookrightarrow \mathscr{G}(p)$ suit la loi géométrique de paramètre p, alors X admet une espérance finie et $\mathbb{E}(X) = \frac{1}{p}$; en effet (sous réserve de convergence) :

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{n=0}^{+\infty} \mathbb{P}(X > n) && \text{Proposition 28} \\
&= \sum_{n=0}^{+\infty} (1-p)^n && \text{Proposition 25}
\end{aligned}
$$

$$= \frac{1}{1-(1-p)} \qquad \text{série géométrique convergente}$$
$$= \frac{1}{p}$$

Espérance d'une loi géométrique :

$X \hookrightarrow \mathscr{G}(p)$ *admet une espérance finie :* $\mathbb{E}(X) = \dfrac{1}{p}$.

• Si $X \hookrightarrow \mathscr{P}(\lambda)$ suit la loi de Poisson de paramètre λ, alors X admet une espérance fini et $\mathbb{E}(X) = \lambda$; en effet (sous réserve de convergence) :

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{n=0}^{+\infty} ne^{-\lambda}\frac{\lambda^n}{n!} \\
&= \sum_{n=1}^{+\infty} e^{-\lambda}\frac{\lambda^n}{(n-1)!} \\
&= e^{-\lambda}\sum_{n=0}^{+\infty} \frac{\lambda^{n+1}}{n!} \\
&= \lambda e^{-\lambda}\sum_{n=0}^{+\infty} \frac{\lambda^n}{n!} \\
&= \lambda e^{-\lambda}e^{\lambda} && \text{série exponentielle} \\
&= \lambda
\end{aligned}$$

Espérance d'une loi de Poisson :

$X \hookrightarrow \mathscr{P}(\lambda)$ *admet une espérance finie :* $\mathbb{E}(X) = \lambda$.

On rappelle que si X est une variable aléatoire discrète, alors pour toute fonction f définie sur $\operatorname{Im} X$, $f(X)$ est aussi une variable aléatoire discrète. On dispose dans ce cas de la formule de transfert qui permet de donner (lorsqu'elle a un sens) son espérance sans connaitre la loi de $f(X)$, mais juste celle de X :

Théorème 29. (Formule de transfert)
Soit X une V.A.D. à valeurs réelles, et soit f une fonction définie sur $\operatorname{Im} X$ et à valeurs réelles ou complexes. Alors :
• *la variable $f(X)$ est d'espérance finie si et seulement si la famille $(f(x) \times \mathbb{P}(X = x))_{x\in\operatorname{Im} X}$ est sommable et dans ce cas :*

$$\mathbb{E}(f(X)) = \sum_{x\in\operatorname{Im} X} f(x) \times \mathbb{P}(X = x)$$

• *En particulier si $\operatorname{Im} X = \{x_n \mid n \in \mathbb{N}\}$, $f(X)$ est d'espérance finie si et seulement si la série $\sum f(x_n) \times \mathbb{P}(X = x_n)$ est absolument convergente, et dans ce cas :*

$$\mathbb{E}(f(X)) = \sum_{n=0}^{+\infty} f(x_n) \times \mathbb{P}(X = x_n)$$

Démonstration. Par définition $f(X)$ admet une espérance finie si et seulement si la famille $(y \times \mathbb{P}(f(X) = y))_{y\in\operatorname{Im} f(X)}$ est sommable, et dans ce cas $\mathbb{E}(f(X)) = \sum_{y\in\operatorname{Im} f(X)} y \times \mathbb{P}(f(X) = y)$.
Pour tout $y \in f(\operatorname{Im} X)$,

$$(f(X) = y) = \bigcup_{x\in f^{-1}(\{y\})} (X = x)$$

c'est une réunion dénombrable, puisque X est discrète, d'événements deux à deux incompatibles, et donc par σ-additivité :

(*) $$\mathbb{P}(f(X) = y) = \sum_{x\in f^{-1}(\{y\})} \mathbb{P}(X = x)$$

on partitionne $\operatorname{Im} X$ en la famille $f^{-1}(\{y\})$ lorsque y décrit $\operatorname{Im} f(X)$. Lorsque la famille $(f(x) \times \mathbb{P}(X=x))_{x\in\operatorname{Im} X}$ est sommable, d'après le théorème de sommation par paquets :

$$\begin{aligned}\sum_{x\in\operatorname{Im} X} f(x)\times\mathbb{P}(X=x) &= \sum_{y\in\operatorname{Im} f(X)}\ \sum_{x\in f^{-1}(\{y\})} f(x)\times\mathbb{P}(X=x)\\ &= \sum_{y\in\operatorname{Im} f(X)} y \sum_{x\in f^{-1}(\{y\})} \mathbb{P}(X=x)\\ &= \sum_{y\in\operatorname{Im} f(X)} y\times\mathbb{P}(f(X)=y)\\ &= \mathbb{E}(f(X))\end{aligned}$$

et dans ce cas $f(X)$ admet une espérance finie.

Réciproquement supposons que la famille $(f(x)\times\mathbb{P}(X=x))_{x\in\operatorname{Im} X}$ n'est pas sommable, et montrons que la famille $(y\times\mathbb{P}(f(X)=y))_{y\in\operatorname{Im} f(X)}$ ne l'est pas non plus (et donc que $f(Y)$ n'a pas une espérance finie).

Par définition :

$$\sup\left\{\sum_{x\in A}|f(x)|\times\mathbb{P}(X=x)\ \middle|\ A\subset\operatorname{Im} X \text{ fini}\right\}=+\infty.$$

Or d'après $(*)$, pour tout $A\subset\operatorname{Im} X$ (fini) :

$$\sum_{x\in A}|f(x)|\times\mathbb{P}(X=x)\leqslant\sum_{y\in f(A)}|y|\times\mathbb{P}(f(X)=y)$$

et bien sur lorsque $A\subset\operatorname{Im} X$ est fini, $f(A)\subset\operatorname{Im} f(X)$ est aussi fini. Ainsi :

$$\sup\left\{\sum_{y\in B}|y|\times\mathbb{P}(f(X)=y)\ \middle|\ B\subset\operatorname{Im} f(X) \text{ fini}\right\}=+\infty.$$

et donc par définition la famille $(y\times\mathbb{P}(f(X)=y))_{y\in\operatorname{Im} f(X)}$ n'est pas sommable. ∎

Remarque. La formule s'applique aussi aux couples ou vecteurs de variables aléatoires. Si (X,Y) est un couple de VAD et $f:\operatorname{Im}(X,Y)\longrightarrow\mathbb{C}$ alors si $f(X,Y)$ admet une espérance, on a la formule de transfert :

$$\mathbb{E}(f(X,Y))=\sum_{(x,y)\in\operatorname{Im}(X,Y)} f(x,y)\times\mathbb{P}(X=x,Y=y)$$

Exercice 6. Soit $X\hookrightarrow\mathscr{G}(p)$; montrer que la VAD $\frac{1}{X}$ est d'espérance finie et calculer son espérance.

Résolution.

PROPOSITION 30.
Soient X et Y deux VAD définies sur un même espace probabilisé et à valeurs complexes. Si $|X|\leqslant Y$ et si Y est d'espérance finie, alors X est aussi d'espérance finie.

Démonstration. On peut supposer que $\operatorname{Im} X$ est infini (dénombrable) sinon la conclusion est claire. Considérons le couple $Z=(X,Y)$; c'est une VAD. Notons $\operatorname{Im} Z=\{(x_n,y_n)\mid n\in\mathbb{N}\}$ où les (x_n,y_n) sont deux à deux distincts.

Notons $f:(x,y)\longmapsto x$ et $g:(x,y)\longmapsto y$. Alors $Y=g(Z)$ et $X=f(Z)$. Comme Y est d'espérance finie, d'après le formule de transfert :

$$\mathbb{E}(Y)=\sum_{n\in\mathbb{N}} g(x_n,y_n)\times\mathbb{P}(Z=(x_n,y_n))=\sum_{n\in\mathbb{N}} y_n\times\mathbb{P}(Z=(x_n,y_n))$$

Puisque par hypothèse, pour tout $n\in\mathbb{N}$, $|x_n|\leqslant y_n$, par comparaison, la série $\sum_{n\in\mathbb{N}} x_n\times\mathbb{P}(Z=(x_n,y_n))$ converge absolument, et d'après le théorème de transfert :

$$\sum_{n\in\mathbb{N}} x_n\times\mathbb{P}(Z=(x_n,y_n))=\sum_{n\in\mathbb{N}} f(x_n,y_n)\times\mathbb{P}(Z=(x_n,y_n))=\mathbb{E}(X)$$

et donc X est d'espérance finie. ∎

Établissons les propriétés de l'espérance.

Propriété 31. (Positivité de l'espérance)
Si X est une VAD réelle à valeurs dans $\mathbb{R}_+$ et d'espérance finie, alors $\mathbb{E}(X) \geqslant 0$.
De plus $E(X) = 0$ si et seulement si X est nulle presque surement, c'est à dire si $(X = 0)$ est presque sur.

Démonstration. On a $\mathbb{E}(X) = \sum_{x \in \operatorname{Im} X} x\mathbb{P}(X = x)$; puisque $\operatorname{Im} X \subset \mathbb{R}_+$, c'est une somme de termes positifs, et donc, en cas de convergence, $\mathbb{E}(X) \geqslant 0$.

Si $E(X) = 0$ alors pour tout $x \in \operatorname{Im} X$, $x\mathbb{P}(X = x) = 0$, et donc $x \neq 0 \implies \mathbb{P}(X = x) = 0$. Or

$$1 = \sum_{x \in \operatorname{Im}(X)} \mathbb{P}(X = x) = \mathbb{P}(X = 0) + \underbrace{\sum_{x \in \operatorname{Im}(X)^*} \mathbb{P}(X = x)}_{=0} \implies \mathbb{P}(X = 0) = 1.$$

■

Propriété 32. (Linéarité de l'espérance)
Soient X et Y deux V.A.D. sur un même espace probabilisé et à valeurs dans $\mathbb{C}$; si X et Y sont d'espérance finie, alors pour tout $(a, b) \in \mathbb{C}^2$, $aX + bY$ est d'espérance finie et :

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$$

Démonstration. On utilise le lemme suivant qui donne une expression de l'espérance qui n'utilise pas la loi de la VAD.

Lemme 33.
X une VAD à valeurs complexes définie sur une espace probabilisé $(\Omega, \mathscr{P}(\Omega), \mathbb{P})$ avec Ω au plus dénombrable. Alors X admet une espérance finie si et seulement si la famille $(X(\omega) \times \mathbb{P}(\{\omega\})_{\omega \in \Omega}$ est sommable et dans ce cas :

$$E(X) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\})$$

Démonstration. (Du Lemme). Sous ces hypothèses, l'application $Id_\Omega : \omega \longmapsto \omega$ est une VAD puisque $Id_\Omega(\Omega) = \Omega$ est dénombrable et que pour tout $x \in \Omega$, $Id_\Omega^{-1}(\{x\}) = \{x\}$ appartient à la tribu $\mathscr{P}(\Omega)$. Appliquons la formule de transfert à la VAD $X \circ Id_\Omega = X$:

$X \circ Id_\Omega = X$ admet un espérance finie si et seulement si la famille $(X(\omega) \times \mathbb{P}(Id_\Omega = \omega))_{\omega \in \operatorname{Im} Id_\Omega}$ est sommable, c'est à dire si et seulement si la famille $(X(\omega) \times \mathbb{P}(\{\omega\}))_{\omega \in \Omega}$ est sommable, et dans ce cas :

$$\mathbb{E}(X \circ Id_\Omega) = \sum_{\omega \in Id_\Omega(\Omega)} X(\omega) \times \mathbb{P}(Id_\Omega = \omega)$$

$$\text{c'est à dire } \; \mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) \times \mathbb{P}(\{\omega\})$$

■

Revenons à la preuve de la propriété. On considère d'abord le cas où l'univers Ω est au plus dénombrable ; notons $\Omega = \{\omega_n \mid n \in I\}$ où $I \subset \mathbb{N}$ est une partie finie de $\mathbb{N}$ ou tout $\mathbb{N}$. D'après le lemme, puisque X et Y admettent un espérance finie, les familles $(X(\omega_n) \times \mathbb{P}(\{\omega_n\}))_{n \in I}$ et $(Y(\omega_n) \times \mathbb{P}(\{\omega_n\}))_{n \in I}$ sont sommables, c'est à dire que lorsque $I = \mathbb{N}$, les séries :

$$\sum_{n \in I} X(\omega_n) \times \mathbb{P}(\{\omega_n\}) \qquad \text{et} \qquad \sum_{n \in I} Y(\omega_n) \times \mathbb{P}(\{\omega_n\})$$

sont absolument convergentes. La famille $((aX(\omega_n) + bY(\omega_n)) \times \mathbb{P}(\{\omega_n\}))_{n \in I}$ est donc sommable, c'est à dire que la somme :

$$\sum_{n \in I} (aX(\omega_n) + bY(\omega_n)) \times \mathbb{P}(\{\omega_n\})$$

est bien définie et à valeur dans $\mathbb{C}$ (c'est clair quand I est fini, et lorsque $I = \mathbb{N}$ la série converge absolument, comme combinaison linéaire de séries absolument convergentes). Or en appliquant le lemme (3 fois) et par linéarité de la somme, on obtient :

$$\mathbb{E}(aX + bY) = \sum_{n \in I} (aX(\omega_n) + bY(\omega_n)) \times \mathbb{P}(\{\omega_n\}) = a \times \sum_{n \in I} X(\omega_n) \times \mathbb{P}(\{\omega_n\}) + b \times \sum_{n \in I} Y(\omega_n) \times \mathbb{P}(\{\omega_n\}) = a \times \mathbb{E}(X) + b \times \mathbb{E}(Y)$$

Ce qui conclut la cas au plus dénombrable.

Il reste à prouver le cas où Ω est infini indénombrable. Nous allons voir qu'il se ramène au cas dénombrable en construisant des VAD $\tilde{X}$ et $\tilde{Y}$ sur un univers au plus dénombrable, de mêmes lois que X et Y. Supposons que X et Y sont d'espérance finie et considérons le couple de VAD (X, Y) ; c'est une VAD et donc son image

$$(X, Y)(\Omega) = \{(X(\omega), Y(\omega)) \mid \omega \in \Omega\} \subset \mathbb{C}^2$$

est au plus dénombrable. Posons $\tilde{\Omega} = (X, Y)(\Omega)$ son image et munissons le de la tribu $\mathscr{P}(\tilde{\Omega})$; on définit une probabilité sur $(\tilde{\Omega}, \mathscr{P}(\tilde{\Omega}))$ en posant pour tout $(x, y) \in \tilde{\Omega}$:

$$\mathbb{P}(\{(x, y)\}) = \mathbb{P}(X = x, Y = y)$$

et $\mathbb{P}$ est σ-additive, puisque $\sum_{(x,y) \in (X,Y)(\Omega)} \mathbb{P}(X = x, Y = y) = 1$. C'est un fait déjà remarqué (Chapitre "Espaces probabilisés"), résumé dans le lemme suivant.

Lemme 34.
Soit $\tilde{\Omega} = \{\omega_n \mid n \in I\}$ (avec $I \subset \mathbb{N}$) un ensemble au plus dénombrable et $(p_n)_{n \in I}$ une famille de réels positifs vérifiant $\sum_{i \in I} p_i = 1$. Alors il existe une unique probabilité sur $(\tilde{\Omega}, \mathscr{P}(\Omega))$ vérifiant :

$$\forall i \in I,\ \mathbb{P}(\{\omega_i\}) = p_i.$$

On a alors pour tout $A \subset \tilde{\Omega}$:

$$\mathbb{P}(A) = \sum_{\omega_i \in A} \mathbb{P}(\{\omega_i\}) = \sum_{\omega_i \in A} p_i.$$

Démonstration. (Du lemme).
Par analyse et synthèse. Soit $A \in \mathscr{P}(\Omega)$; alors A est au plus dénombrable et par σ-additivité si $\mathbb{P}$ s'étend sur $\mathscr{P}(\Omega)$ en une probabilité on a nécessairement $\mathbb{P}(A) = \sum_{i \in A} p_i \in [0,1]$. Cela définit une application $\mathbb{P} : \mathscr{P}(\Omega) \longrightarrow [0,1]$ qui est σ-additive. De plus :

$$\mathbb{P}(\Omega) = \sum_{\omega_i \in \Omega} \mathbb{P}(\{\omega_i\}) = \sum_{i \in I} p_i = 1$$

Donc $\mathbb{P}$ ainsi définie est bien une probabilité, et c'est la seule vérifiant $\forall i \in I,\ \mathbb{P}(\{\omega_i\}) = p_i$. ■

Revenons à la preuve. On a défini un espace probabilisé $(\tilde{\Omega}, \mathscr{P}(\tilde{\Omega}), \mathbb{P})$ d'univers $\tilde{\Omega}$ au plus dénombrable. En notant :

$$\begin{array}{rcl} \tilde{X} : \tilde{\Omega} & \longrightarrow & \mathbb{C} \\ (x,y) & \longmapsto & x \end{array} \quad \text{et} \quad \begin{array}{rcl} \tilde{Y} : \tilde{\Omega} & \longrightarrow & \mathbb{C} \\ (x,y) & \longmapsto & y \end{array}$$

$\tilde{X}$ et $\tilde{Y}$ sont des VAD, puisque :

$$\tilde{X}(\tilde{\Omega}) = X(\Omega) \quad \tilde{Y}(\tilde{\Omega}) = Y(\Omega)$$

sont au plus dénombrables et $\forall x \in \tilde{X}(\tilde{\Omega}),\ \tilde{X}^{-1}(\{x\}) \in \mathscr{P}(\tilde{\Omega})$ et $\forall y \in \tilde{Y}(\tilde{\Omega}),\ \tilde{Y}^{-1}(\{y\}) \in \mathscr{P}(\tilde{\Omega})$ sont bien des événements.

Le couple de VAD $(\tilde{X}, \tilde{Y})$ suit la même loi conjointe que le couple (X,Y), puisque $(\tilde{X}, \tilde{Y})(\Omega') = \Omega' = (X,Y)(\Omega)$ et

$$\forall (x,y) \in \Omega',\ \mathbb{P}(\tilde{X} = x, \tilde{Y} = y) = \mathbb{P}(\{(x,y)\}) = \mathbb{P}(X = x, Y = y)$$

et donc $\tilde{X}$ et X suivent la même première loi marginale, $\tilde{Y}$ et Y suivent la même seconde loi marginale, et $a\tilde{X} + b\tilde{Y}$ suit la même loi que $aX + bY$ (la loi de la somme étant uniquement déterminée par la loi conjointe).

Les VAD $\tilde{X}$ et $\tilde{Y}$ étant définie sur un univers au plus dénombrable, la première partie de la preuve s'applique pour montrer que $a\tilde{X} + b\tilde{Y}$ est d'espérance finie et :

$$\mathbb{E}(a\tilde{X} + b\tilde{Y}) = a\mathbb{E}(\tilde{X}) + b\mathbb{E}(\tilde{y})$$

Donc $aX + bY$ est aussi d'espérance finie et

$$\mathbb{E}(aX + bY) = \mathbb{E}(a\tilde{X} + b\tilde{Y}) = a\mathbb{E}(\tilde{X}) + b\mathbb{E}(\tilde{y}) = a\mathbb{E}(X) + b\mathbb{E}(Y)$$

ce qui conclut la preuve. ■

PROPRIÉTÉ 35. (Croissance de l'espérance)
Soient X et Y deux VAD sur un même espace probabilisé et à valeurs réelles. Si X et Y sont d'espérance finie, et si $X \leqslant Y$ alors :

$$\mathbb{E}(X) \leqslant \mathbb{E}(Y)$$

avec égalité si et seulement si $X = Y$ presque surement.

Démonstration. On applique la positivité de l'espérance à $Y - X$ pour obtenir $\mathbb{E}(Y - X) \geqslant 0$ avec égalité si et seulement si $Y = X$ presque surement, puis la linéarité de l'espérance pour en déduire $\mathbb{E}(Y) \geqslant \mathbb{E}(X)$ avec égalité si et seulement si $Y = X$ presque surement. ■

PROPRIÉTÉ 36. (Espérance du produit de VAD indépendantes)
Soient X et Y deux VAD d'espérance finie sur un même espace probabilisé. Si X et Y sont indépendantes, alors la VAD XY est d'espérance finie et :

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

Démonstration. Les deux familles $(x \times \mathbb{P}(X = x))_{x \in \mathrm{Im}\, X}$ et $(y \times \mathbb{P}(Y = y))_{y \in \mathrm{Im}\, Y}$ étant sommable, la famille $(xy \times \mathbb{P}(X = x)\mathbb{P}(Y = y))_{(x,y) \in \mathrm{Im}\, X \times \mathrm{Im}\, Y}$ est sommable (Théorème 8) et :

$$(1) \qquad \sum_{(x,y) \in \mathrm{Im}\, X \times \mathrm{Im}\, Y} xy \times \mathbb{P}(X = x)\mathbb{P}(Y = y) = \sum_{x \in \mathrm{Im}\, X} x \times \mathbb{P}(X = x) \times \sum_{y \in \mathrm{Im}\, Y} y \times \mathbb{P}(Y = y)$$

Or X et Y étant indépendantes :

$$(2) \qquad \sum_{(x,y) \in \mathrm{Im}\, X \times \mathrm{Im}\, Y} xy \times \mathbb{P}(X = x)\mathbb{P}(Y = y) = \sum_{(x,y) \in \mathrm{Im}\,(X,Y)} xy \times \mathbb{P}(X = x, Y = y)$$

On considère $f : (x,y) \longmapsto x \times y$ et on applique le théorème de transfert à $f \circ (X,Y) = XY$; on obtient que XY est d'espérance finie et :

$$\begin{aligned} \mathbb{E}(XY) &= \sum_{(x,y) \in \mathrm{Im}\,(X,Y)} xy \times \mathbb{P}(X = x, Y = y) \\ &\underset{(1)}{=} \sum_{(x,y) \in \mathrm{Im}\, X \times \mathrm{Im}\, Y} xy \times \mathbb{P}(X = x)\mathbb{P}(Y = y) \\ &\underset{(2)}{=} \sum_{x \in \mathrm{Im}\, X} x \times \mathbb{P}(X = x) \times \sum_{y \in \mathrm{Im}\, Y} y \times \mathbb{P}(Y = y) \\ &= \mathbb{E}(X)\mathbb{E}(Y) \end{aligned}$$

■

Remarque. Attention l'indépendance est nécessaire pour que XY admette une espérance finie. Il est facile de construire un exemple où X et Y sont d'espérance finie sans que XY ne le soit :

Soit $X = Y : \Omega \longrightarrow \mathbb{N}^*$ de loi :

$$\forall n \in \mathbb{N}^*,\ \mathbb{P}(X = n) = \frac{a}{n^3} \qquad \text{avec } a = \left(\sum_{k\in\mathbb{N}^*} \frac{1}{n^3}\right)^{-1}$$

Alors X et Y sont d'espérance finie puisque $\mathbb{E}(X) = \mathbb{E}(Y) = \sum \frac{an}{n^3} = \sum \frac{a}{n^2}$ converge, tandis que $XY = X^2$ ne l'est pas puisque $\mathbb{E}(XY) = \mathbb{E}(X^2) = \sum \frac{an^2}{n^3} = \sum \frac{a}{n}$ diverge.

Ce résultat se généralise aux vecteurs de variables aléatoires indépendantes :

Propriété 37. (Espérance du produit de VAD indépendantes)
Soient $X_1, \ldots, X_n$ des VAD d'espérance finie sur un même espace probabilisé. Si $X_1, \ldots, X_n$ sont indépendantes, alors la VAD $X_1 \times \cdots \times X_n$ est d'espérance finie et :

$$\mathbb{E}(X_1 \times \cdots \times X_n) = \mathbb{E}(X_1) \times \cdots \times \mathbb{E}(X_n)$$

Démonstration. Par récurrence sur $n \geqslant 2$.

(I) Pour $n = 2$ c'est le résultat précédent.

(H) Supposons l'assertion vraie au rang n et soient $X_1, \ldots, X_n, X_{n+1}$ $(n+1)$ variables indépendantes. Par hypothèse de récurrence $X_1 \times \cdots \times X_n$ est d'espérance finie et :

$$\mathbb{E}(X_1 \times \cdots \times X_n) = \mathbb{E}(X_1) \times \cdots \times \mathbb{E}(X_n)$$

D'après le lemme de coalition $X_1 \times \cdots \times X_n$ et X_{n+1} sont indépendantes. Alors d'après la propriété précédente on a que $X_1 \times \cdots \times X_n \times X_{n+1}$ est d'espérance finie et :

$$\mathbb{E}(X_1 \times \cdots \times X_n \times X_{n+1}) = \mathbb{E}(X_1 \times \cdots \times X_n) \times \mathbb{E}(X_{n+1}) = \mathbb{E}(X_1) \times \cdots \times \mathbb{E}(X_n) \times \mathbb{E}(X_{n+1})$$

■

2.2. Variance et covariance (VAD réelle).

Dans cette partie toutes les VAD sont à valeurs réelles.

Définition 16. (Moments d'une VARD)
Soit X une variable aléatoire réelle et soit $p \in \mathbb{N}^*$; on dit que X admet un moment d'ordre p si X^p est d'espérance finie. Le moment d'ordre p de X est alors $\mathbb{E}(X^p)$; c'est un réel.

Proposition 38.
Si X admet un moment d'ordre p, alors X admet un moment d'ordre q pour tout $q \leqslant p$.

Démonstration. Soit $p, q \in \mathbb{N}^*$ avec $q \leqslant p$; alors $|X^q| \leqslant 1 + |X^p|$. Puisque X admet un moment d'ordre p, d'après le théorème de transfert, la famille $(|x|^p \times \mathbb{P}(X = x))_{x\in \mathrm{Im}\, X}$ est sommable. La famille $(\mathbb{P}(X = x))_{x\in \mathrm{Im}\, X}$ est aussi sommable, puisque $\sum_{x\in \mathrm{Im}\, X} \mathbb{P}(X = x) = 1$. Ainsi par linéarité de la sommation et par comparaison, la famille $(|x|^q \times \mathbb{P}(X = x))_{x\in \mathrm{Im}\, X}$ est sommable ; d'après le théorème de transfert, X^q est d'espérance finie. ■

Proposition-Définition 17. (Variance d'une VADR)
Soit X une variable aléatoire réelle discrète.
Si X admet un moment d'ordre 2, alors X est d'espérance finie et on définit la variance de X par :

$$\mathrm{V}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right)$$

qui peut aussi s'écrire (formule de Koenig-Hugens) :

$$V(X) = \mathbb{E}\left(X^2\right) - \left(\mathbb{E}(X)\right)^2$$

La variance de X est un réel positif, et on appelle écart-type le réel $\sigma(X) = \sqrt{\mathrm{V}(X)}$.

Démonstration. Puisque X admet un moment d'ordre 2, d'après la proposition précédente, X^2 et X sont d'espérance finie, et par linéarité de l'espérance $(X - \mathbb{E}(X))^2 = X^2 - 2\mathbb{E}(X)X + \mathbb{E}(X)^2$ est aussi d'espérance finie. Ainsi $\mathrm{V}(X)$ est bien définie, et par positivité de l'espérance est à valeurs dans $\mathbb{R}_+$. De plus par linéarité de l'espérance :

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

ce qui prouve la formule de Koenig-Huygens. ■

Remarques.

- Par positivité de l'espérance :

$$V(X) = 0 \quad \text{si et seulement si} \quad X = \mathbb{E}(X) \text{ presque surement}$$

Ainsi une VARD X est de variance nulle si et seulement si elle suit une loi quasi-certaine, c'est à dire $\exists c \in \mathbb{R}$, $\mathbb{P}(X = c) = 1$. (Mais pour autant on n'a pas nécessairement $\operatorname{Im} X = \{c\}$.)

• Si X admet un moment d'ordre 2 alors X admet une variance finie. Réciproquement, si X admet une variance finie, alors par définition X admet une espérance finie, et par linéarité de l'espérance X^2 admet une espérance finie, c'est à dire X admet un moment d'ordre 2. Ainsi une VARD admet un moment d'ordre 2 si et seulement si elle admet (une espérance et) une variance finie(s).

Exemples. Variance de lois usuelles.

• La variance d'une loi quasi-certaine est nulle :

$$V(X) = \mathbb{E}((X - \mathbb{E}(X))^2 = (\mathbb{E}(X) - \mathbb{E}(X))^2 \times 1 + \sum_{x \neq \mathbb{E}(X)} (x - \mathbb{E}(X))^2 \times 0 = 0.$$

• Variance d'une loi de Bernoulli.

Soit $X \hookrightarrow \mathscr{B}(p)$ alors $\mathrm{V}(X) = p(1-p)$. En effet : $\operatorname{Im} X = \{0, 1\}$ et $\mathbb{E}(X) = p$, donc d'après la formule de transfert :

$$\begin{aligned}\mathrm{V}(X) &= \mathbb{E}((X-p)^2) = (0-p)^2\mathbb{P}(X=0) + (1-p)^2\mathbb{P}(X=1)\\ &= p^2 \times (1-p) + (1-p)^2 \times p = p(1-p)(p+1-p)\\ &= p(1-p)\end{aligned}$$

• Variance d'une loi géométrique.

Soit $X \hookrightarrow \mathscr{G}(p)$ avec $p \in]0,1[$; alors avec la formule de transfert, X admet un moment d'ordre 2 puisque la série

$$\mathbb{E}(X^2) = \sum_{n \in \mathbb{N}^*} n^2 (1-p)^{n-1} p$$

converge (par exemple en appliquant d'Alembert). Calculons sa variance : par linéarité de l'espérance, $X(X-1)$ admet une espérance finie, et avec la formule de transfert :

$$\begin{aligned}\mathbb{E}(X(X-1)) &= \sum_{n \in \mathbb{N}^*} n(n-1)(1-p)^{n-1} p\\ &= p(1-p) \sum_{n \geqslant 2} n(n-1)(1-p)^{n-2}\end{aligned}$$

on reconnaît la série géométrique dérivée seconde :

$$\begin{aligned}&= p(1-p) \times \frac{d^2}{dp^2}\left(\frac{1}{1-(1-p)}\right)\\ &= p(1-p) \times \frac{2}{p^3}\\ &= \frac{2(1-p)}{p^2}\end{aligned}$$

Donc par linéarité de l'espérance :

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

Et d'après Koenig-Huygens :

$$\mathrm{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \boxed{\frac{1-p}{p^2}}$$

Exercice 7. Montrer que la variance d'une loi de Poisson de paramètre $\lambda > 0$ est égale à λ.

Résolution.

PROPRIÉTÉ 39.
Soit $(a, b) \in \mathbb{R}^2$ et X une variable aléatoire admettant un moment d'ordre 2 ; alors
$$\mathrm{V}(aX + b) = a^2\mathrm{V}(X)$$

Démonstration. Par linéarité de l'espérance :
$$aX + b - \mathbb{E}(aX + b) = a(X - \mathbb{E}(X))$$
et X et X^2 admettant une espérance finie, il en va de même de $(a(X - \mathbb{E}(X)))^2$; on a alors :
$$\mathrm{V}(aX + b) = \mathbb{E}((a(X - \mathbb{E}(X)))^2) = a^2\mathbb{E}((X - \mathbb{E}(X))^2) = a^2\mathrm{V}(X)$$
∎

Remarque. Une VADR est dite <u>centrée</u> lorsque son espérance est nulle ; elle est dite <u>réduite</u> lorsque sa variance vaut 1. Pour toute variable X admettant un moment d'ordre 2, la variable aléatoire
$$Y = \frac{X - m}{\sigma} \qquad \text{avec } m = \mathbb{E}(X) \text{ et } \sigma = \sqrt{\mathrm{V}(X)}$$
vérifie : $\mathbb{E}(Y) = 0$ et $V(Y) = 1$, c'est une variable aléatoire <u>centrée réduite</u>.

PROPOSITION 40. (Condition suffisante d'espérance finie du produit)
Soient X et Y deux VADR sur un même espace probabilisé et admettant un moment d'ordre 2 ; alors XY est d'espérance finie.

Démonstration. Puisque $X^2 + Y^2 - 2|XY| = (|X| - |Y|)^2 \geqslant 0$, on a alors $|XY| \leqslant \frac{1}{2}(X^2 + Y^2)$. Par linéarité de l'espérance $\frac{1}{2}(X^2 + Y^2)$ admet un espérance finie, et avec la proposition 30, XY aussi. ∎

Cette propriété nous assure en corollaire qu'une combinaison linéaire de VARD admettant un moment d'ordre 2, admet aussi un moment d'ordre 2. Autrement dit :

COROLLAIRE 41. (Structure de $\mathbb{R}$-espace vectoriel)
L'ensemble des variables aléatoires sur un même espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$ et admettant un moment d'ordre 2 est un $\mathbb{R}$-espace vectoriel noté $\mathcal{L}^2(\Omega, \mathscr{A}, \mathbb{P})$.

Démonstration. Soient X et Y deux VADR sur $(\Omega, \mathscr{A}, \mathbb{P})$ admettant un moment d'ordre 2. Alors pour tout $\lambda \in \mathbb{R}$, $\lambda \cdot X + Y$ est une VADR sur $(\Omega, \mathscr{A}, \mathbb{P})$ et puisque :
$$(\lambda \cdot X + Y)^2 = \lambda^2 \cdot X^2 + Y^2 + 2\lambda \cdot XY$$
X^2, Y^2 et XY admettant une espérance finie, il en va de même, par linéarité de l'espérance, de $(\lambda \cdot X + Y)^2$; la VADR $(\lambda \cdot X + Y)$ admet donc un moment d'ordre 2. ∎

L'espérance finie du produit XY de deux VARD X et Y admettant un moment d'ordre 2 permet surtout de définir la notion de covariance de X et Y.

PROPOSITION-DÉFINITION 18. (Covariance de X et Y)
Soient X et Y deux variables aléatoires sur un même espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$ et admettant un moment d'ordre 2. On définit la <u>covariance de X et Y</u> comme le réel :
$$\mathrm{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))).$$
En particulier, $\mathrm{V}(X) = \mathrm{Cov}(X, X)$.
La covariance peut aussi s'écrire (Formule de Koenig-Huygens) :
$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Démonstration. Puisque X et Y admettent un moment d'ordre 2, alors X et Y admettent une espérance finie, ainsi que XY (proposition 40). Par linéarité de l'espérance, $(X - \mathbb{E}(X))(Y - \mathbb{E}(Y)) = XY - \mathbb{E}(Y)X - \mathbb{E}(X)Y + \mathbb{E}(X)\mathbb{E}(Y)$ admet donc aussi une espérance finie. La covariance de X, Y est donc bien définie. De plus :
$$\begin{aligned}\mathrm{Cov}(X, Y) &= \mathbb{E}\big[XY - \mathbb{E}(Y)X - \mathbb{E}(X)Y + \mathbb{E}(X)\mathbb{E}(Y)\big] \\ &= \mathbb{E}(XY) - \mathbb{E}(Y)\mathbb{E}(X) - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\end{aligned}$$
ce qui établit la formule de Koenig-Huygens. ∎

Propriété 42. (Forme bilinéaire symétrique positive)
Sur $\mathcal{L}^2(\Omega, \mathscr{A}, \mathbb{P})$ la covariance est une forme bilinéaire, symétrique et positive. Autrement dit pour tout $X, Y, Z \in \mathcal{L}^2(\Omega, \mathscr{A}, \mathbb{P})$, et pour tout $(a,b) \in \mathbb{R}^2$:
- $\mathrm{Cov}(aX + bY, Z) = a\,\mathrm{Cov}(X, Z) + b\,\mathrm{Cov}(Y, Z)$,
- $\mathrm{Cov}(X, aY + bZ) = a\,\mathrm{Cov}(X, Y) + b\,\mathrm{Cov}(X, Z)$,
- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$,
- $\mathrm{Cov}(X, X) = \mathrm{V}(X) \geqslant 0$.

Démonstration. Elle est triviale ; la bilinéarité découle de la linéarité de l'espérance, la symétrie de la commutativité de la multiplication, et la positivité découle de la définition. ■

Remarque. La covariance n'est pas un produit scalaire, par absence de définie positivité ; en effet $\mathrm{Cov}(X, X) = 0$ implique que X suit une loi quasi-certaine.

Une propriété très utile de la covariance est la suivante :

Propriété 43. Deux variables indépendantes ont une covariance nulle
Soient X et Y dans $\mathcal{L}^2(\Omega, \mathscr{A}, \mathbb{P})$; si X et Y sont indépendantes, alors $\mathrm{Cov}(X, Y) = 0$.

Démonstration. Pour deux variables indépendantes $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ (propriété 36), d'où la nullité de leur covariance, avec Koenig-Huygens. ■

Remarque. Ainsi la covariance peut être utile pour montrer que deux variables ne sont pas indépendantes. (Car la réciproque est fausse : $\mathrm{Cov}(X, Y)$ peut être nulle sans X et Y ne soient indépendants.).

La covariance n'est pas très loin d'être un produit scalaire, de "norme associée" l'écart-type. L'indépendance implique alors 'l'orthogonalité" pour cette analogie. Et on dispose encore d'une inégalité de Cauchy-Schwarz :

Théorème 44. (Inégalité de Cauchy-Schwarz)
Soient X et Y deux variables aléatoires dans $\mathcal{L}^2(\Omega, \mathscr{A}, \mathbb{P})$, alors :

$$\mathbb{E}(XY)^2 \leqslant \mathbb{E}\left(X^2\right) \times \mathbb{E}\left(Y^2\right).$$

avec égalité si et seulement si X et Y sont proportionnels presque surement.
Et :

$$|\mathrm{Cov}(X, Y)| \leqslant \sigma(X) \times \sigma(Y)$$

avec égalité si et seulement si il existe $(a,b) \in \mathbb{R}^2$ tel que $Y = aX + b$ p.s. ou $X = aY + b$ p.s..

Démonstration. On reprend la preuve de l'inégalité de Cauchy-Schwarz avec le pseudo-produit scalaire $(X, Y) \mapsto \mathbb{E}(XY)$ (on n'a pas le caractère défini, car $\mathbb{E}\left(X^2\right) = 0 \implies X = 0$ p.s. (et pas partout) mais la preuve est la même). Supposons que $\mathbb{E}(X^2) \neq 0$; soit $\lambda \in \mathbb{R}$; par positivité de l'espérance :

$$0 \leqslant \mathbb{E}((\lambda \cdot X + Y)^2) = \lambda^2 \mathbb{E}(X^2) + 2\lambda \cdot \mathbb{E}(XY) + \mathbb{E}(Y^2)$$

C'est un trinôme en λ, dont le discriminant est négatif ou nul :

$$\Delta = 4\mathbb{E}(XY)^2 - 4\mathbb{E}(X^2)\mathbb{E}(Y^2) \leqslant 0 \implies \mathbb{E}(XY)^2 \leqslant \mathbb{E}(X^2)\mathbb{E}(Y^2)$$

De plus en cas d'égalité, $\Delta = 0$, et donc il existe $\lambda \in \mathbb{R}$, tel que $\mathbb{E}((\lambda \cdot X + Y)^2) = 0$, c'est à dire $\lambda \cdot X + Y = 0$ presque surement (par positivité, propriété 31).

Si $\mathbb{E}(X^2) = 0$ alors, toujours d'après la propriété 31, $X = 0$ presque surement et là encore $X = 0 \cdot Y$ presque surement et $0 = \mathbb{E}(XY)^2 = \mathbb{E}(X^2) \times \mathbb{E}(Y^2)$. Réciproquement si $X = aY$ (ou $Y = aX$) presque surement, il est immédiat de vérifier que l'inégalité est une égalité.

En particulier, en appliquant le résultat à $(X - E(X))$ et $(Y - \mathbb{E}(Y)$:

$$\begin{aligned} &\mathrm{Cov}(X, Y)^2 = \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y)\right]^2 \leqslant \mathbb{E}((X - \mathbb{E}(X))^2) \times \mathbb{E}((Y - \mathbb{E}(Y))^2) \\ \iff &|\mathrm{Cov}(X, Y)| \leqslant \sigma(X) \times \sigma(Y) \end{aligned}$$

avec égalité si et seulement si il existe $a \in \mathbb{R}$ tel que presque surement $X - \mathbb{E}(X) = aY - a\mathbb{E}(Y)$ (ou $Y - \mathbb{E}(Y) = aX - a\mathbb{E}(X)$), et en particulier il existe $(a,b) \in \mathbb{R}^2$ tels que presque surement $X = aY + b$ (ou $Y = aX + b$). Réciproquement, si (par exemple) $X = aY + b$ presque surement, alors $\mathbb{E}(X) = a\mathbb{E}(Y) + b$ et donc $X - \mathbb{E}(X) = aY + b - a\mathbb{E}(Y) - b = aY - a\mathbb{E}(Y)$ et il y a donc égalité. ■

Remarque. Soient $X, Y \in \mathcal{L}^2(\Omega, \mathscr{A}, \mathbb{P})$ avec $\mathrm{V}(X) \neq 0$ et $\mathrm{V}(Y) \neq 0$; on définit leur coefficient de corrélation par la formule :

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

c'est un réel dans $[-1,1]$ qui vérifie :

$$\rho(X,Y) = \begin{cases} 0 & \text{lorsque } X \text{ et } Y \text{ sont indépendantes} \\ \pm 1 & \iff Y = aX+b \text{ p.s. ou } X = aY+b \text{ p.s.} \end{cases}$$

Il est utilisé notamment en statistique pour la régression linéaire. Lorsqu'il est proche de ± 1, les séries statistiques, que l'on peut voir comme des VADR, sont proches de suivre une relation affine.

Propriété 45. (Variance d'une somme)
Pour des VARD dans $\mathcal{L}^2(\Omega, \mathscr{A}, \mathbb{P})$:

$$\mathrm{V}(X+Y) = \mathrm{V}(X) + \mathrm{V}(Y) + 2\,\mathrm{Cov}(X,Y)$$

et plus généralement :

$$\mathrm{V}\Big(\sum_{k=1}^n X_k\Big) = \sum_{k=1}^n \mathrm{V}(X_k) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j) = \sum_{k=1}^n \mathrm{V}(X_k) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j)$$

Démonstration. On démontre la deuxième formule, plus générale ; pour cela on rappelle que :

$$\left(\sum_{i=1}^n x_i\right)^2 = \sum_{i=1}^n x_i^2 + 2 \times \sum_{1 \leqslant i<j \leqslant n} x_i x_j \qquad (*)$$

d'après le formule de Koenig-Huygens :

$$\begin{aligned}
\mathrm{V}\left(\sum_{i=1}^n X_i\right) &= \mathbb{E}\left(\left(\sum_{i=1}^n X_i\right)^2\right) - \left[\mathbb{E}\left(\sum_{i=1}^n X_i\right)\right]^2 \\
&= \mathbb{E}\left(\sum_{i=1}^n X_i^2 + 2 \times \sum_{1 \leqslant i<j \leqslant n} X_i X_j\right) - \left(\sum_{i=1}^n \mathbb{E}(X_i)\right)^2 && \text{avec (*) et par linéarité de } \mathbb{E} \\
&= \sum_{i=1}^n \mathbb{E}(X_i^2) + 2 \times \sum_{1 \leqslant i<j \leqslant n} \mathbb{E}(X_i X_j) - \sum_{i=1}^n \mathbb{E}(X_i)^2 - 2 \sum_{1 \leqslant i<j \leqslant n} \mathbb{E}(X_i)\mathbb{E}(X_j) && \text{idem} \\
&= \sum_{i=1}^n \left(\mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2\right) + 2 \times \sum_{1 \leqslant i<j \leqslant n} \left(\mathbb{E}(X_i X_j) - \mathbb{E}(X_i) E(X_j)\right) \\
&= \sum_{i=1}^n \mathrm{V}(X_i) + 2 \times \sum_{1 \leqslant i<j \leqslant n} \mathrm{Cov}(X_i, X_j) && \text{d'après Koenig-Huygens}
\end{aligned}$$

■

Corollaire 46. (Variance d'une somme de variables indépendantes)
Pour des VARD dans $\mathcal{L}^2(\Omega, \mathscr{A}, \mathbb{P})$ deux à deux indépendantes :

$$\mathrm{V}(X+Y) = \mathrm{V}(X) + \mathrm{V}(Y)$$

et plus généralement :

$$\mathrm{V}\Big(\sum_{k=1}^n X_k\Big) = \sum_{k=1}^n \mathrm{V}(X_k)$$

Remarque. Noter que l'indépendance mutuelle n'est pas nécessaire, l'indépendance 2 à 2 suffit.

Exemple. Variance d'une loi binomiale.

Soit $X \hookrightarrow \mathscr{B}(n,p)$; alors $X = X_1 + \cdots + X_n$ avec $X_1, \ldots, X_n$ des variables i.i.d. suivant la même loi de Bernoulli $\mathscr{B}(p)$. Alors :

$$\mathrm{V}(X) = \sum_{k=1}^n \mathrm{V}(X_k) = n \times p \times (1-p)$$

Exercice 8. *Navale 2015.* On distribue, aléatoirement, uniformément et indépendamment r boules dans n boîtes $B_1, \cdots, B_n$. On note Y_i la variable aléatoire prenant la valeur 1 si B_i est vide et 0 sinon.

1. Déterminer la loi de Y_i.
2. Calculer $\mathrm{Cov}\,(Y_i, Y_j)$ pour $i \neq j$.
3. Calculer l'espérance et la variance de la variable aléatoire donnant le nombre de boîtes vides.

Résolution.

2.3. Fonction génératrice d'une VAD à valeurs dans $\mathbb{N}$.

Dans cette partie toutes les VAD sont à valeurs dans $\mathbb{N}$.

Proposition 47. (Série entière de T.G. la loi de $X : \Omega \longrightarrow \mathbb{N}$)
Soit X une VAD à valeurs dans $\mathbb{N}$. Alors la série entière :
$$\sum \mathbb{P}(X = n) z^n$$
- *a un rayon de convergence $R \geqslant 1$,*
- *converge normalement sur le disque fermé $\overline{D(0,1)}$.*

Démonstration. Puisque $\sum \mathbb{P}(X = n) = 1$, la série converge pour $z = 1$, donc $R \geqslant 1$. De plus pour tout $z \in \overline{D(0,1)}$:
$$|z| \leqslant 1 \implies |\mathbb{P}(X = n) z^n| \leqslant \mathbb{P}(X = n)$$
et la série $\sum \mathbb{P}(X = n) = 1$ converge ; il y a donc convergence normale sur $\overline{D(0,1)}$. ∎

Définition 19. (Fonction génératrice)
Soit X une variable aléatoire discrète à valeurs dans $\mathbb{N}$. On pose en tout $t \in \mathbb{R}$ où c'est possible,
$$G_X(t) = \mathbb{E}\left(t^X\right) = \sum_{n=0}^{+\infty} \mathbb{P}(X = n) \cdot t^n$$
Cette fonction est appelée fonction génératrice de la variable aléatoire X.

Remarques.

• L'égalité entre $\mathbb{E}(t^X)$ et la somme découle de la formule de transfert.

• On a toujours $G_X(1) = 1$ puisque $G_X(1) = \sum \mathbb{P}(X = n) = 1$.

• Si X est à valeurs finies, G_X est une fonction polynomiale.

Exemples. Il faut connaître et savoir retrouver la fonction génératrice des lois usuelles.

• Fonction génératrice d'une loi de Bernoulli.

Soit $X \hookrightarrow \mathscr{B}(p)$. Alors :

$$G_X(t) = \mathbb{P}(X = 0) \cdot t^0 + \mathbb{P}(X = 1) \cdot t^1 = (1 - p) + tp = q + tp$$

la fonction génératrice est affine. Réciproquement si une fonction génératrice G_X est affine, $G_X(t) = q + pt$, alors X suit presque surement une loi de Bernoulli, *i.e.* $\operatorname{Im} X \supset \{0, 1\}$ et $\mathbb{P}(X = 1) = p$, $\mathbb{P}(X = 0) = q$.

• Fonction génératrice d'une loi binomiale.

Soit $X \hookrightarrow \mathscr{B}(n, p)$. Alors :

$$G_X(t) = \sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k} t^k = (1 - p + tp)^n = (q + tp)^n$$

• Fonction génératrice de $\mathscr{U}(\llbracket 1, n \rrbracket)$.

Soit $X \hookrightarrow \mathscr{U}(\llbracket 1, n \rrbracket)$, alors :

$$G_X(t) = \sum_{k=1}^{n} \frac{1}{n} \times t^k = \frac{t}{n} \times \frac{1 - t^n}{1 - t} \quad \text{si } t \neq 1 \text{ et } G_X(1) = 1$$

Propriété 48. (Analytique)
Sous les mêmes hypothèses :

- *Le domaine de définition de G_X contient $[-1, 1]$,*
- *La fonction G_X est continue sur le segment $[-1, 1]$,*
- *La fonction G_X est de classe $\mathscr{C}^\infty$ sur l'intervalle ouvert $]-1, 1[$.*

Démonstration. Elle découle de faits élémentaires sur les séries entières. Comme vu plus haut il y a convergence normale sur le segment $[-1, 1]$, et donc existence et continuité sur tout le segment. Le rayon de convergence est $R \geqslant 1$, la somme d'une série entière étant de classe $\mathscr{C}^\infty$ sur l'intervalle ouvert de convergence, elle l'est au moins sur l'ouvert $]-1, 1[$. ∎

L'intérêt de la fonction génératrice G_X réside notamment dans le fait qu'elle caractérise la loi de X :

Proposition 49. (La fonction génératrice caractérise la loi)
Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. La loi de X est entièrement déterminée par sa fonction génératrice G_X. Plus précisément :

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

Remarque. En particulier deux variables aléatoires à valeurs dans $\mathbb{N}$ ont même loi si et seulement si elles ont même fonction génératrice.

Démonstration. La fonction G_X est infiniment dérivable sur $]-1, 1[$. D'après le théorème de dérivation terme à terme, $\forall t \in]-1, 1[$, $\forall n \in \mathbb{N}$:

$$G_X^{(n)}(t) = \sum_{k=n}^{+\infty} \frac{k!}{(k - n)!} \mathbb{P}(X = k) t^{k-n}$$

en particulier $\forall n \in \mathbb{N}$:

$$G_X^{(n)}(0) = n! \times \mathbb{P}(X = n) \implies \mathbb{P}(X = n) = \frac{G_X^{(n)}(0)}{n!}$$

∎

Espérance et variance d'une variable aléatoire à valeurs entières sont aussi facilement déterminés à l'aide de la fonction génératrice.

THÉORÈME 50. (Calcul d'espérance et de variance)
La fonction $t \in [-1,1] \longmapsto G_X(t)$ est :
– dérivable en 1 ssi X est d'espérance finie, on a alors :
$$G'_X(1) = \mathbb{E}(X)$$
– deux fois dérivable en 1 ssi X admet une variance et :
$$G''_X(1) = \mathbb{E}(X(X-1))$$

Remarques.
• Pour retrouver la formule, on retiendra
$$G'_X(t) = \mathbb{E}\left(Xt^{X-1}\right) \quad \text{donc} \quad G'_X(1) = \mathbb{E}(X).$$
$$G''_X(t) = \mathbb{E}\left(X(X-1)t^{X-2}\right) \quad \text{donc} \quad G''_X(1) = \mathbb{E}(X(X-1)).$$
• Lorsque G_X est deux fois dérivable en 1, la variance s'obtient alors par :
$$\mathrm{V}(X) = G''_X(1) + G'_X(1) - G'^2_X(1)$$
puisque : $G''_X(1) = \mathbb{E}\left(X^2\right) - \mathbb{E}(X) = V(X) + (\mathbb{E}(X))^2 - \mathbb{E}(X) = \mathrm{V}(X) + \left(G'_X(1)\right)^2 - G'_X(1)$.

Démonstration. (La démonstration du sens réciproque est non exigible.)

On note $p_n = \mathbb{P}(X = n)$.

Pour l'espérance : Supposons que X est d'espérance finie, $\mathbb{E}(X) = \sum_{n=0}^{+\infty} n \times p_n$. Alors la série $\sum_{n\geqslant 1} nt^{n-1} \times p_n$ converge normalement sur $[0,1]$. Puisque la série $\sum_{n\geqslant 0} t^n \times p_n$ converge simplement sur $[0,1]$, d'après le théorème de dérivation $\mathscr{C}^1$, $G'_X(1)$ existe et vaut $G'_X(1) = \sum_{n=0}^{+\infty} n \times p_n = \mathbb{E}(X)$.

Réciproquement, supposons que G_X est dérivable en 1 : on a $\lim_{t\to 1} \frac{G_X(1)-G_X(t)}{1-t} = \ell \in \mathbb{R}$. Remarquons que pour $t < 1$:
$$\frac{G_X(1) - G_X(t)}{1-t} = \frac{\sum\limits_{n=0}^{+\infty} p_n(1-t^n)}{1-t} = \sum_{n=0}^{+\infty} p_n \left(1 + t + \cdots + t^{n-1}\right) \underset{t\to 1}{\longrightarrow} \ell$$
Puisque $t \mapsto \left(1 + t + \cdots + t^{n-1}\right)$ est croissante, on a pour tout $t \in [0,1[^2$
$$\sum_{n=0}^{+\infty} p_n \left(1 + t + \cdots + t^{n-1}\right) \leqslant \ell.$$
Donc, pour tout $N \in \mathbb{N}$, en faisant tendre t vers 1 :
$$\sum_{n=0}^{N} np_n \leqslant \ell$$
La série étant à terme positifs, d'après le théorème de la limite monotone, $\sum_{n\geqslant 0} np_n$ converge, ainsi X admet une espérance.

Pour la variance : En utilisant l'argument précédent mais sur la série dérivée, on établit que la fonction $G'_X(t) = \sum_{n=1}^{+\infty} np_n t^{n-1}$ est dérivable en 1 ssi $\sum_{n=1}^{+\infty} n(n-1)p_n < +\infty$, et qu'en cas de convergence, $\lim_{t\to 1} G''_X(t) = G''_X(1) = \sum_{n=1}^{+\infty} n(n-1)p_n$.

En particulier $\mathbb{E}(X(X-1)) < +\infty$. De plus, puisque G_X est dérivable en $1, \mathbb{E}(X) < +\infty$, donc $\mathbb{E}\left(X^2\right) < +\infty$ et X admet une variance. ■

THÉORÈME 51. (Fonction génératrice de la somme de variables indépendantes)
Soient X et Y deux variables aléatoires à valeurs dans $\mathbb{N}$ indépendantes.
Alors pour tout $t \in [-1,1]$
$$G_{X+Y}(t) = G_X(t) \times G_Y(t).$$

Exemple. Une loi binomiale $\mathscr{B}(n,p)$ étant somme de lois de Bernoulli $\mathscr{B}(p)$ i.i.d. on retrouve que puisque la fonction génératrice de $\mathscr{B}(p)$ est $pt+q$, celle de $\mathscr{B}(n,p)$ est $(pt+q)^n$.

Démonstration. Les variables X et Y étant indépendantes, d'après le lemme de coalition t^{X} et t^{Y} sont indépendantes donc pour $|t| \leqslant 1$, avec la propriété 36 :
$$G_{X+Y}(t) = \mathbb{E}\left(t^{X+Y}\right) = \mathbb{E}\left(t^X t^Y\right) = \mathbb{E}\left(t^X\right)\mathbb{E}\left(t^Y\right) = G_X(t) \times G_Y(t).$$
■

Exercice 9. Calculer les fonctions génératrices :
– d'une loi géométrique $\mathscr{G}(p)$
– d'une loi de Poisson $\mathscr{P}(\lambda)$.

Résolution.

• Formulaire à connaître

Loi de X	$\operatorname{Im} X$	$\mathbb{P}(X=k)$	$\mathbb{E}(X)$	$\mathrm{V}(X)$	$G_X(t)$
$\mathscr{B}(p)$	$\{0,1\}$	$\begin{cases} p & \text{si } k=1 \\ q & \text{si } k=0 \end{cases}$	p	pq	$pt+q$
$\mathscr{B}(n,p)$	$[\![0,n]\!]$	$\binom{n}{k} p^k q^{n-k}$	np	npq	$(pt+q)^n$
$\mathscr{U}([\![1,n]\!])$	$[\![1,n]\!]$	$\dfrac{1}{n}$	$\dfrac{n+1}{2}$	$\dfrac{n^2-1}{12}$	$\dfrac{t}{n}\,\dfrac{1-t^n}{1-t}\ (t \neq 1)$
$\mathscr{G}(p)$	$\mathbb{N}^*$	$q^{k-1}p$	$\dfrac{1}{p}$	$\dfrac{q}{p^2}$	$\dfrac{pt}{1-qt}$
$\mathscr{P}(\lambda)$	$\mathbb{N}$	$e^{-\lambda}\dfrac{\lambda^k}{k!}$	λ	λ	$e^{\lambda(t-1)}$

en notant $q = 1-p$. Il faut savoir retrouver rapidement les fonctions génératrices des lois usuelles.

2.4. Résultats asymptotiques.

Dans cette partie toutes les variables aléatoires sont à valeurs réelles.

Théorème 52. (Inégalité de Markov)
Soit X une VADR positive d'espérance finie ; alors pour tout $\varepsilon > 0$,

$$\mathbb{P}(X \geqslant \varepsilon) \leqslant \frac{1}{\varepsilon}\mathbb{E}(X).$$

Remarque. Si on ne suppose pas X positive, on peut écrire $\mathbb{P}(|X| \geqslant \varepsilon) \leqslant \dfrac{1}{\varepsilon}\mathbb{E}(|X|)$.

Démonstration. On écrit :

$$\begin{aligned} X &= X \times \mathbb{1}_{X \geqslant \varepsilon} + X \times \mathbb{1}_{X<\varepsilon} \\ &\geqslant \varepsilon \times \mathbb{1}_{X \geqslant \varepsilon} \quad \text{car } X \geqslant 0 \end{aligned}$$

donc par croissance de l'espérance :

$$\mathbb{E}(X) \geqslant \mathbb{E}\left(\varepsilon \times \mathbb{1}_{X \geqslant \varepsilon}\right) = \varepsilon \times \mathbb{P}(X \geqslant \varepsilon).$$

■

Exemple. Les salaires étant positifs, la proportion de la population recevant plus de 3 fois le salaire moyen est au plus d'un tiers. (Poser X qui à un salarié associe le rapport entre son salaire et le salaire moyen ; alors $\mathbb{E}(X) = 1$).

L'inégalité de Bienaymé-Tchebychev se déduit de l'inégalité de Markov ; elle est plus fine, faisant intervenir aussi la variance. Elle donne notamment un sens à la variance, comme mesure de la dispersion autour de la valeur moyenne (= espérance).

Théorème 53. (Inégalité de Bienaymé-Tchebychev)
Soit X une VADR admettant un moment d'ordre 2 ; alors pour tout $\varepsilon > 0$:

$$\mathbb{P}(|X - \mathbb{E}(X)| \geqslant \varepsilon) \leqslant \frac{1}{\varepsilon^2}\ \mathrm{V}(X).$$

Démonstration. On applique l'inégalité de Markov à la VADR positive $|X - \mathbb{E}(X)|^2$:

$$\begin{aligned}\mathbb{P}(|X - \mathbb{E}(X)| \geqslant \varepsilon) &= \mathbb{P}\left(|X - \mathbb{E}(X)|^2 \geqslant \varepsilon^2\right)\\ &\leqslant \frac{1}{\varepsilon^2}\mathbb{E}\left(|X - \mathbb{E}(X)|^2\right) = \frac{1}{\varepsilon^2}\ \mathrm{V}(X).\end{aligned}$$

■

Exemple. Soit X est une VADR vérifiant $\mathbb{E}(X) = 0$ et $\mathrm{V}(X) = 2$; déterminons $\alpha > 0$ tel que $\mathbb{P}(|X| \geqslant \alpha) \leqslant \frac{3}{4}$.

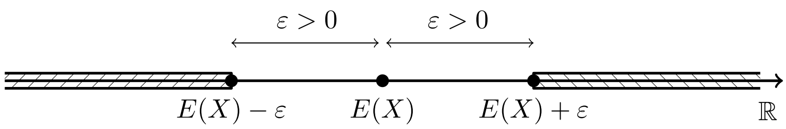
$$\mathbb{P}(|X| \geqslant \alpha) \leqslant \frac{3}{4} \iff \mathbb{P}(|X - \mathbb{E}(X)| \geqslant \alpha) \leqslant \frac{3}{4}$$

D'après Bienaymé-Tchebychev, il suffit de prendre α tel que $\dfrac{2}{\alpha^2} \leqslant \dfrac{3}{4}$ soit $\alpha \geqslant \dfrac{2\sqrt{6}}{3}$.

Remarques.

• La condition $|X - \mathbb{E}(X)| \geqslant \varepsilon$ s'interprète :

"X s'éloigne de $\mathbb{E}(X)$ d'au moins ε"



• L'écart-type (ainsi que la variance) mesurent la dispersion des valeurs prises par X autour de $\mathbb{E}(X)$. Il peu probable que X s'écarte de $\mathbb{E}(X)$ d'un écart bien supérieur à son écart-type. Un faible écart-type signifie donc une V.A.R. dont les valeurs sont concentrées autour de son espérance.

Avec $\varepsilon = 2 \times \sigma(X)$, la probabilité que X s'écarte de $\mathbb{E}(X)$ d'au moins $2 \times \sigma(X)$ est $\leqslant \frac{1}{4}$.

Avec $\varepsilon = 5 \times \sigma(X)$, la probabilité que X s'écarte de $\mathbb{E}(X)$ d'au moins $5 \times \sigma(X)$ est $\leqslant \frac{1}{25}$.

Avec $\varepsilon = 10 \times \sigma(X)$, la probabilité que X s'écarte de $\mathbb{E}(X)$ d'au moins $10 \times \sigma(X)$ est $\leqslant \frac{1}{100}$.

• La formule de Bienaymé-Tchebychev est très générale : elle s'applique à toute V.A.R. ; en contre-partie la majoration obtenue est relativement faible. Par exemple pour $\varepsilon \leqslant \sigma(X)$ cette formule ne donne aucune information, puisque dans ce cas le majorant $\left(\frac{\sigma(X)}{\varepsilon}\right)^2$ est supérieur à 1 tandis qu'une probabilité est toujours inférieure à 1. Lorsqu'on connaît la loi de X (et pas seulement $\mathbb{E}(X)$ et $\sigma(X)$) on peut souvent obtenir une majoration plus fine de $\mathbb{P}(|X - \mathbb{E}(X)| \geqslant \varepsilon)$.

Théorème 54. (Loi faible des grands nombres)
Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires réelles deux à deux indépendantes et de même loi, admettant un moment d'ordre 2. On pose $S_n = \sum_{k=1}^n X_k$ et $m = \mathbb{E}(X_1)$. Alors :

$$\forall \varepsilon > 0, \quad \mathbb{P}\left(\left|\frac{S_n}{n} - m\right| \geqslant \varepsilon\right) \underset{n\to+\infty}{\longrightarrow} 0.$$

Remarque. Ce résultat exprime que la moyenne empirique des résultats d'une expérience aléatoire tend "en probabilité" vers son espérance mathématique.

Démonstration. Par linéarité de l'espérance : $\mathbb{E}(S_n) = n\mathbb{E}(X_1) = nm$ et par indépendance : $\mathrm{V}(S_n) = n\ \mathrm{V}(X_1)$, toutes les variables étant de même loi. Ainsi :

$$\mathbb{E}\left(\frac{S_n}{n}\right) = m \quad \text{et } \mathrm{V}\left(\frac{S_n}{n}\right) = \frac{\mathrm{V}(X_1)}{n}$$

En appliquant l'inégalité de Bienaymé-Tchebychev :

$$\mathbb{P}\left(\left|\frac{S_n}{n} - m\right| \geqslant \varepsilon\right) \leqslant \frac{1}{\varepsilon^2}\ \mathrm{V}\left(\frac{S_n}{n}\right) = \frac{\mathrm{V}(X_1)}{n\varepsilon^2} \underset{n\to+\infty}{\longrightarrow} 0.$$

■