Variables aléatoires discrètes

1 Préliminaire : calcul de sommes doubles

Certains calculs de ce chapitre font intervenir une somme sur deux indices (i et j) : on parle de *somme double*, ou de *série double* si on a des sommes infinies.
On considère donc ici une famille de réels indexée par deux entiers $a_{i,j}$. On veut effectuer la somme $\sum\limits_{i,j} a_{i,j}$.

1.1 Sommes finies

Quand le domaine de sommation est fini, on peut organiser la sommation comme on veut. Deux cas particuliers sont importants :

Sommation sur un rectangle

Ici, $i \in [\![0,m]\!]$ et $j \in [\![0,n]\!]$. On a

$$\sum_{i,j\in[\![0,m]\!]\times[\![0,n]\!]} a_{i,j} = \sum_{i=0}^{m}\left(\sum_{j=0}^{n} a_{i,j}\right) = \sum_{j=0}^{n}\left(\sum_{i=0}^{m} a_{i,j}\right)$$

Sommation sur un triangle

Ici, $(i,j) \in [\![0,n]\!]^2$; en rajoutant la contrainte $i \leqslant j$. On a

$$\sum_{0\leqslant i\leqslant j\leqslant n} a_{i,j} = \sum_{j=0}^{n}\left(\sum_{i=0}^{j} a_{i,j}\right) = \sum_{i=0}^{n}\left(\sum_{j=i}^{n} a_{i,j}\right)$$

Essentiellement, il faut visualiser la contrainte entre les deux indices (ici : $i \leqslant j$). La somme extérieure englobe toutes les valeurs possibles d'un indice, alors que l'intérieure prend en compte la contrainte. On obtient :

$$\sum_{0\leqslant i\leqslant j\leqslant n} a_{i,j} = \underbrace{\sum_{i=0}^{n}}_{ext.}\left(\underbrace{\sum_{j=i}^{n}}_{int: j\geqslant i} a_{ij}\right) = \underbrace{\sum_{j=0}^{n}}_{ext.}\left(\underbrace{\sum_{i=0}^{j}}_{int: i\leqslant j} a_{ij}\right)$$

On peut envisager des variantes : condition $i < j$, variation dans $[\![1,n]\!]^2$, ...

Exemple 1. Calculer les sommes suivantes :

1. $\sum\limits_{i=0}^{n}\sum\limits_{j=0}^{n} 2^{2i-j}$
2. $\sum\limits_{1\leqslant i\leqslant j\leqslant n} i+j$

(NB : ne pas s'attendre à des expressions finales très simples...)

1.2 Cas infini dénombrable : absolue convergence et réarrangements

Pour calculer une somme infinie à deux indices, on est confrontés aux problèmes d'existence des sommes ; donc de convergence de séries. Les manipulations effectuées pour la calculer sont en fait légitimes si on dispose de **convergences absolues**.
NB : en fait le programme dispense officiellement d'invoquer la convergence absolue pour justifier l'existence et le calcul d'une somme double. Néanmoins vous aurez la plupart du temps à faire cela dans le cadre de calculs d'espérance et de variance, et on attend alors de vous que vous parliez de convergence absolue pour légitimer ce que vous faites.

Théorème 1 (Sommes rectangulaires infinies)**.**
Soient a_{ij} des nombres réels, avec des indices $i \in \mathbb{N}$ et $j \in \mathbb{N}$; on cherche à calculer $\sum\limits_{(i,j)\in\mathbb{N}^2} a_{ij}$. Alors si l'une des deux expressions :

$$\sum_{i=0}^{+\infty}\left(\sum_{j=0}^{+\infty}|a_{ij}|\right) \qquad \sum_{j=0}^{+\infty}\left(\sum_{i=0}^{+\infty}|a_{ij}|\right)$$

est correctement définie (c'est-à-dire que toutes les sommes infinies convergent), alors on dit que la famille (a_{ij}) est sommable*, et les deux procédés de sommation suivants convergent et donnent la même somme :*

$$\sum_{(i,j)\in\mathbb{N}^2} a_{ij} = \sum_{i=0}^{+\infty}\left(\sum_{j=0}^{+\infty}a_{ij}\right) = \sum_{j=0}^{+\infty}\left(\sum_{i=0}^{+\infty}a_{ij}\right)$$

Théorème 2 (Sommes triangulaires infinies)**.**
Soient a_{ij} des nombres réels, avec des indices $i \in \mathbb{N}$ et $j \in \mathbb{N}$; on cherche à calculer $\sum\limits_{\substack{(i,j)\in\mathbb{N}^2\\ i\leqslant j}} a_{ij}$. Alors si l'une des deux expressions :

$$\sum_{i=0}^{+\infty}\left(\sum_{j=i}^{+\infty}|a_{ij}|\right) \qquad \sum_{j=0}^{+\infty}\left(\sum_{i=0}^{j}|a_{ij}|\right)$$

est correctement définie (c'est-à-dire que toutes les sommes infinies convergent), alors on dit que la famille (a_{ij}) est sommable*, et les deux procédés de sommation suivants convergent et donnent la même somme :*

$$\sum_{\substack{(i,j)\in\mathbb{N}^2\\ i\leqslant j}} a_{ij} = \sum_{i=0}^{+\infty}\left(\sum_{j=i}^{+\infty}a_{ij}\right) = \sum_{j=0}^{+\infty}\left(\sum_{i=0}^{j}a_{ij}\right)$$

Pour une sommation triangulaire, isoler la « contrainte » permet de s'en sortir : par exemple,

$$\sum_{\substack{(i,j)\in\mathbb{N}^2\\ i<j}} a_{i,j} = \underbrace{\sum_{i=0}^{+\infty}}_{ext.}\left(\underbrace{\sum_{j=i+1}^{+\infty}}_{int:j>i} a_{ij}\right) = \underbrace{\sum_{j=1}^{+\infty}}_{ext.}\left(\underbrace{\sum_{i=0}^{j-1}}_{int:i<j} a_{ij}\right)$$

(NB : la somme sur j part de 1 car la contrainte $j > i$ empêche j de valoir 0...)

Remarque 1.
En pratique : pour montrer la sommabilité, on considère donc la somme double **des valeurs absolues des quantités à sommer** ; puis « sous réserve de convergence » , on effectue les manipulations sur la somme (changements d'indice, interversions, etc.) jusqu'à trouver une forme permettant de démontrer la sommabilité.
On justifie alors sur la copie que la famille est sommable ; puis on reprend les mêmes manipulations sur la somme **des quantités elles-mêmes (donc sans valeur absolue cette fois)** pour calculer la somme voulue.

Exemple 2. Calculer les sommes suivantes :

1. $\displaystyle\sum_{(k,n)\in\mathbb{N}^2} \frac{1}{2^{n+1}k!}$

2. $\displaystyle\sum_{\substack{n\geqslant 2\\ p\geqslant 2}} \frac{1}{n^p}$ (on remarquera : $\forall\, n \geqslant 2,\ \dfrac{1}{n(n-1)} = \dfrac{1}{n-1} - \dfrac{1}{n}$).

3. $\displaystyle\sum_{\substack{(k,n)\in\mathbb{N}^2\\ k\leqslant n}} \frac{1}{n!}\binom{n}{k}\frac{(-1)^k}{2^k}$

1.3 Interversion des signes de sommation

L'enjeu de ces techniques est souvent d'intervertir les signes de sommation : parmi les deux ordres exposés dans les sommes doubles qui précèdent, un des deux ne sera pas calculable (pas de formule explicite), alors que l'autre le sera. Il convient alors de savoir faire l'interversion, qui peut être subtile.
Voyons deux exemples :

Exemple 3. Montrer que $\displaystyle\sum_{i=1}^{\mathrm{N}}\left(\sum_{k=i}^{\mathrm{N}}\frac{i}{k}\right) = \sum_{k=1}^{\mathrm{N}}\left(\sum_{i=1}^{k}\frac{i}{k}\right)$. Calculer cette somme.

Exemple 4. Calculer $\displaystyle\sum_{i=1}^{+\infty}\left(\sum_{k=i}^{+\infty}\frac{1}{k}\left(\frac{1}{2}\right)^k\right)$.

2 Probabilités : Rappels rapides et pas très formalisés de première année

Une *expérience aléatoire* (**exemple : lancer d'un dé à six faces**) est une expérience qui peut produire plusieurs résultats (qu'on qualifiera hâtivement d'« aléatoires » , *ie.* imprévisibles), appelées *issues* (**ici : 1,2,3,4,5,6**). L'ensemble des issues est appelé *univers*, et noté traditionnellement Ω.

Un *événement* est le plus souvent formulé par une propriété vérifiée par le résultat de l'expérience (**exemple : obtenir un chiffre pair**) ; formellement il s'agit en fait d'une sous-partie de l'univers (**ici :** $\{$**2,4,6**$\}$).

Sur cet univers, on définit des *probabilités*. Outre leur signification héritée des statistiques, il faut voir les probabilités comme une *mesure* : chaque issue possède une probabilité (positive) ; la somme de toutes ces probabilités donne la mesure de l'univers, qui vaut par convention 1.
Ainsi dans notre exemple, on dispose des probabilités $\mathbb{P}(\{1\}) = p_1, \dots, \mathbb{P}(\{6\}) = p_6$ associées à chaque issue ; où les p_i :

- sont éléments de $[0,1]$
- et vérifient $p_1 + p_2 + \cdots + p_6 = 1$.

Dans le cas d'un dé équilibré tous les p_i sont égaux (à $1/6$) : on est en situation d'*équiprobabilité*. Mais cela n'a rien d'automatique : pour des dés déséquilibrés on peut avoir des p_i quelconques, tant qu'ils vérifient les deux conditions ci-dessus.

La donnée des p_i définit une probabilité. De manière formelle, on ne définit pas une probabilité sur les issues mais directement sur les événements.
Une probabilité est une *fonction* qui, à chaque événement A de l'univers associe sa probabilité $\mathbb{P}(\mathrm{A})$; et qui vérifie en outre :

- Pour tout événement A, $\mathbb{P}(\mathrm{A}) \in [0,1]$ (NB : en fait demander $\mathbb{P}(\mathrm{A}) \geqslant 0$ est suffisant) ;
- $\mathbb{P}(\Omega) = 1$;
- Si les événements A_i sont 2 à 2 disjoints, $\mathbb{P}\left(\bigsqcup_{i=0}^{+\infty} \mathrm{A}_i\right) = \sum_{i=0}^{+\infty} \mathbb{P}(\mathrm{A}_i)$ (propriété de « sigma-additivité »).

Notation

Je noterai parfois « $\bigsqcup$ » l'union disjointe :

$$\text{On a } \mathrm{E} = \bigsqcup_{i \in \mathrm{I}} \mathrm{A}_i \text{ ssi } \left(\mathrm{E} = \bigcup_{i \in \mathrm{I}} \mathrm{A}_i \text{ et } \forall (i,j) \in \mathrm{I}^2,\ i \neq j \Rightarrow \mathrm{A}_i \cap \mathrm{A}_j = \varnothing\right)$$

C'est plus ou moins officiel : sur une copie, expliquez cette notation à la première utilisation, ou écrivez « union disjointe » .

Les *probabilités conditionnelles* se définissent ainsi : si A et B sont deux événements, avec $\mathbb{P}(\mathrm{B}) > 0$, on définit la probabilité de A sachant B :

$$\mathbb{P}_{\mathrm{B}}(\mathrm{A}) = \frac{\mathbb{P}(\mathrm{A} \cap \mathrm{B})}{\mathbb{P}(\mathrm{B})}$$

(NB : un événement peut être de proba nulle sans être vide pour autant !)
On voit donc que **« A sachant B » n'est pas un événement**. La probabilité conditionnelle est en fait *une autre mesure de probabilité*, qui à tout événement A associe sa « probabilité sachant B » (on peut vérifier que la fonction $\mathbb{P}_{\mathrm{B}}$ introduite ci-dessus vérifie les trois points précédents qui caractérisent une probabilité).

Formules à revoir :

- *Formule des probabilités composées.*
  Si $B_1, \ldots, B_n$ *sont des événements tels que* $\mathbb{P}(B_1 \cap \ldots \cap B_{n-1}) \neq 0$, *alors*

$$\mathbb{P}(B_1 \cap \ldots \cap B_{n-1} \cap B_n) = \mathbb{P}(B_1)\mathbb{P}_{B_1}(B_2)\mathbb{P}_{B_1 \cap B_2}(B_3)\ldots\ldots P_{B_1 \cap \ldots \cap B_{n-1}}(B_n)$$

- *Formule des probabilités totales : Si* $(B_i)_{i \in I}$ *est un système complet d'événements, et A un événement, alors*

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A \cap B_i)$$

  Si de plus les $\mathbb{P}(B_i)$ *sont non nulles, on a aussi*

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}_{B_i}(A)\,\mathbb{P}(B_i)$$

- *Formule de Bayes : si A et B sont de probabilités non nulles,*

$$\mathbb{P}_B(A) = \frac{\mathbb{P}_A(B)\,\mathbb{P}(A)}{\mathbb{P}(B)}$$

Une *variable aléatoire* X est une quantité associée à une issue de l'expérience. C'est donc, formellement, une *fonction définie sur l'univers* : à une issue ω, on associe le nombre $X(\omega)$.
Dans le cas de notre lancer de dé, on peut par exemple définir la variable X qui vaut 1 ssi le chiffre obtenu est impair ; et 0 ssi ce chiffre est pair. Ainsi $X(3) = 1$, $X(6) = 0$.

L'ensemble des valeurs prises par X se note $X(\Omega)$ (**ici :** $X(\Omega) = \{0, 1\}$) et est appelé *univers-image* de X.
Si on note $X(\Omega) = I$ (dans ce chapitre ce sera un ensemble dénombrable de réels), on peut considérer, pour $i \in I$ quelconque, l'événement regroupant toutes les issues ω pour lesquelles $X(\omega) = i$: on note cet événement $(X = i)$.
Ainsi :

$$(X = i) = \{\omega \in \Omega \mid X(\omega) = i\}$$

(tâchez de bien comprendre cette dernière ligne, c'est une bonne gymnastique pour se familiariser avec les objets et concepts manipulés dans ce cours). Dans notre dernier exemple on a donc $(X = 1) = \{1, 3, 5\}$

Si on considère l'ensemble des événements $(X = i)$, on obtient des événements 2 à 2 disjoints :

$$i \neq j \Rightarrow (X = i) \cap (X = j) = \varnothing$$

et recouvrant tout l'univers :

$$\bigcup_{i \in I}(X = i) = \Omega \quad \text{(qu'on peut donc noter : } \bigsqcup_{i \in I}(X = i))$$

On rappelle qu'un ensemble d'événements vérifiant les deux lignes précédentes est appelé *système complet d'événements.*

On voit que pour faire des calculs se rapportant à X il n'est pas forcément nécessaire d'aller fouiller dans les issues ω ; il suffit de manipuler les événements $(X = i)$ et leurs probabilités.
Avec les propriétés générales d'une fonction probabilité, on a les propriétés suivantes pour les $\mathbb{P}(X = i)$:

- $\forall\, i \in I,\ \mathbb{P}(X = i) \in [0, 1]$;
- $\sum_{i \in I} \mathbb{P}(X = i) = 1$.

La donnée de ces probabilités constitue la *loi de probabilité* de la variable X.

Réciproquement, toute collection $(a_i)_{i \in I}$ *de nombres vérifiant :*

- $\forall\, i \in I,\ a_n \geqslant 0$
- $\sum_{i \in I} a_i = 1$

peut légitimement être vue comme la loi de probabilité d'une variable aléatoire, indépendamment d'un contexte expérimental ; c'est à dire qu'il existe une variable aléatoire X telle que $X(\Omega) = I$, *et* $\forall\, i \in I$, $\mathbb{P}(X = i) = a_i$.

Formules à revoir : *lois de probabilité discrètes usuelles : Bernoulli, binomiale, géométrique, Poisson.*

Par exemple, si on considère une urne contenant 2 boules portant le numéro « 1 » et 3 boules portant le numéro « 2 » ; qu'on tire une boule au hasard (de manière équiprobable) dans cette urne ; et qu'on appelle X le numéro de la boule tirée, alors X suit la loi de probabilité suivante :

i	1	2
$\mathbb{P}(X = i)$	$2/5$	$3/5$

Attention : la loi de probabilité de X permet d'obtenir beaucoup d'informations sur X, mais elle ne définit pas à elle seule la variable : **deux variables aléatoires suivant la même loi ne sont pas forcément égales**.
Considérons en effet un tirage à Pile ou Face équilibré ; on note X la v.a. qui vaut 1 si on obtient Pile et 0 si on obtient Face ; et Y la v.a. qui vaut 0 si on obtient Pile et 1 si on obtient Face.
On voit alors que X et Y suivant la même loi :

i	0	1
$\mathbb{P}(X = i)$	$1/2$	$1/2$

i	0	1
$\mathbb{P}(Y = i)$	$1/2$	$1/2$

mais on n'a clairement pas $X = Y$.

On définit ensuite, *sous réserve de convergence <u>absolue</u> des séries en jeu* (car l'ensemble $X(\Omega)$ peut être infini, auquel cas les sommes qui suivent sont des sommes de séries, donc convergence non garantie !) :

- *L'espérance* $\mathbb{E}(X) = \sum_{x \in X(\Omega)} x\,\mathbb{P}(X = x)$ (s'interprète comme la valeur moyenne observée pour X lors d'un grand nombre d'expériences).
  Attention : $\mathbb{E}(X)$ *existe ssi* $\sum_{x \in X(\Omega)} |x|\,\mathbb{P}(X = x)$ *converge ; mais on a bel et bien* $\mathbb{E}(X) = \sum_{x \in X(\Omega)} x\,\mathbb{P}(X = x)$ *(sans les valeurs absolues).*
- Le *théorème de transfert* étend cela à une fonction quelconque de X : $\mathbb{E}\big(f(X)\big) = \sum_{x \in X(\Omega)} f(x)\,\mathbb{P}(X = x)$.
- Variance, etc. : cf plus bas.

Et pour les nouveautés....

3 Couples de variables aléatoires

On étend dans ce chapitre les propriétés rencontrées pour une variable aléatoire aux situations pour lesquelles on a à manipuler simultanément *plusieurs* variables aléatoires, définies pour la même expérience aléatoire (donc sur le même univers).

L'enjeu principal dans le traitement de deux variables est celui de leur dépendance : on aura à manipuler des événements de la forme $\big((X = i) \cap (Y = j)\big)$, dont la probabilité NE SERA PAS, en général $\mathbb{P}(X = i) \times \mathbb{P}(Y = j)$.

On considère dans toute cette section un *espace probabilisé*[1] $(\Omega, \mathscr{A}, \mathbb{P})$. Les variables aléatoires considérées seront à valeurs réelles.

3.1 Loi d'un couple de VAD

On rappelle qu'une variable aléatoire X est dite *discrète* ssi $X(\Omega)$ est fini ou dénombrable. Dans tout le cours, on abrégera « variable aléatoire discrète » en VAD.

On commence par définir la loi du couple (X, Y) :

Définition 1. *Soient X et Y deux VAD définies sur* Ω.
On appelle loi conjointe *du couple (X,Y) (parfois désignée comme la « loi du couple (X,Y) ») la donnée, pour tous* $(x, y) \in X(\Omega) \times Y(\Omega)$, *des probabilités* $\mathbb{P}\big((X = x) \cap (Y = y)\big)$.

Dans le cas où $X(\Omega)$ et $Y(\Omega)$ sont finis, la loi conjointe se représente naturellement dans un tableau.

Exemple 5. On considère trois lancers successifs d'une pièce équilibrée. On appelle X la VAD qui vaut 1 si le premier lancer donne P, et 0 s'il donne F ; on note Y le nombre de Pile obtenu sur les trois lancers. Donner la loi conjointe du couple (X, Y).

Comme dans le cas d'une seule VAD, on obtient directement un résultat : « la somme des probabilités vaut 1 » .

Proposition 3. *Les événements* $\big((X = x) \cap (Y = y)\big)$, *pour* $(x, y) \in X(\Omega) \times Y(\Omega)$ *forment un SCE. On en déduit*

$$\sum_{(x,y)\in X(\Omega)\times Y(\Omega)} \mathbb{P}\big((X = x) \cap (Y = y)\big) = \sum_{y\in Y(\Omega)} \left(\sum_{x\in X(\Omega)} \mathbb{P}\big((X = x) \cap (Y = y)\big) \right) = \sum_{x\in X(\Omega)} \left(\sum_{y\in Y(\Omega)} \mathbb{P}\big((X = x) \cap (Y = y)\big) \right) = 1$$

[1] En gros, un univers sur lequel on peut faire des calculs de probabilité. Inutile d'en savoir plus.

Exemple 6. Soient X et Y deux VAD suivant la loi conjointe :

$$\forall (n,k) \in \mathbb{N}^2,\ \mathbb{P}\big((X=n)\cap(Y=k)\big) = \frac{a}{2^{n+1}k!}$$

Déterminer la valeur de a.

3.2 Lois marginales

La loi conjointe du couple (X, Y) donne *toutes* les informations sur X et Y nécessaires au calcul de probabilités. On peut notamment reconstruire la loi d'une des deux variables. Dans ce contexte, on appelle ces lois les *lois marginales* ; mais ce ne sont jamais que la loi de X et la loi de Y, au sens usuel.

Définition 2.
Si (X,Y) est un couple de VAD défini sur Ω*, on appelle* «loi marginale de *X*» *la loi de X (et de même pour Y).*

Le passage de la loi conjointe aux lois marginales s'effectue en utilisant la formule des probabilités totales. Rappelons à cet effet que :

Si X est une VAD, la famille d'événements $\big((X=x)\big)_{x\in X(\Omega)}$ est un SCE, appelé *système complet d'événements associé à* X.

(peut être utilisé sans démonstration).

On obtient :

Théorème 4 (Loi conjointe ⇒ lois marginales)**.**
Soit (X,Y) un couple de VAD défini sur Ω*. Pour un* $x \in X(\Omega)$ *fixé, on a :*

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

et de même, pour un $y \in Y(\Omega)$ *fixé :*

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

Démonstration. C'est une simple application de la formule des probabilités totales, avec les SCE associés aux variables Y (première formule) et X (seconde formule). □

Si on reprend la représentation de la loi conjointe dans un tableau, on obtient les lois marginales en sommant selon les lignes (ou les colonnes) de ce tableau.

Exemple 7. Retrouver la loi du nombre de Pile dans une série de 3 lancers P/F à l'aide de la loi conjointe de l'exemple 5.

X\ Y	0	1	2	3	loi de X
0					
1					
loi de Y					

Remarque 2. Par contre, la donnée des lois marginales ne permet pas, sans hypothèses supplémentaires, de reconstituer la loi conjointe.

Exemple 8. Les deux lois conjointes :

X\ Y	0	1
0	0.3	0.1
1	0.1	0.5

et

X\ Y	0	1
0	0.4	0
1	0	0.6

donnent les mêmes lois marginales pour les variables X et Y.

Ceci vient du fait que les variables aléatoires X et Y ne sont pas forcément « indépendantes » (à comprendre dans un sens intuitif pour l'instant ; nous préciserons ça un peu plus loin). Cette intrication est précisément ce qui est encodé dans la loi conjointe.

On peut aussi manipuler les probabilités conditionnelles, en conditionnant par des valeurs de Y ou des valeurs de X. Là aussi, ce ne sont jamais que les définitions usuelles :

Définition 3. *Soit (X,Y) est un couple de VAD défini sur* Ω, *et* $y \in \mathrm{Y}(\Omega)$ *tel que* $\mathbb{P}(\mathrm{Y}=y) \neq 0$.
On appelle loi conditionnelle de X sachant $(\mathrm{Y}=y)$ *la donnée, pour tout* $x \in \mathrm{X}(\Omega)$, *des probabilités conditionnelles*

$$\mathbb{P}_{(\mathrm{Y}=y)}(\mathrm{X}=x) = \frac{\mathbb{P}\big((\mathrm{X}=x)\cap(\mathrm{Y}=y)\big)}{\mathbb{P}(\mathrm{Y}=y)}$$

De même, si $\mathbb{P}(\mathrm{X}=x) \neq 0$, *on appelle loi conditionnelle de Y sachant* $(\mathrm{X}=x)$ *la donnée, pour tout* $y \in \mathrm{Y}(\Omega)$, *des probabilités conditionnelles*

$$\mathbb{P}_{(\mathrm{X}=x)}(\mathrm{Y}=y) = \frac{\mathbb{P}\big((\mathrm{X}=x)\cap(\mathrm{Y}=y)\big)}{\mathbb{P}(\mathrm{X}=x)}$$

À nouveau en utilisant les probabilités totales, on déduit des lois conditionnelles et marginales permettent de retrouver la loi conjointe :

Théorème 5. $\forall\,(x,y) \in \mathrm{X}(\Omega)\times\mathrm{Y}(\Omega)$,

$$\mathbb{P}\big((\mathrm{X}=x)\cap(\mathrm{Y}=y)\big) = \mathbb{P}_{(\mathrm{X}=x)}(\mathrm{Y}=y)\times\mathbb{P}(\mathrm{X}=x)$$

$$\mathbb{P}\big((\mathrm{X}=x)\cap(\mathrm{Y}=y)\big) = \mathbb{P}_{(\mathrm{Y}=y)}(\mathrm{X}=x)\times\mathbb{P}(\mathrm{Y}=y)$$

Remarque 3. On dispose donc de plusieurs moyens de dire la même chose. Si toute l'information est contenue dans la loi conjointe, quel besoin a-t-on de parler de lois conditionnelles ?
Ceci se joue en fait dans la présentation d'une expérience aléatoire « concrète » : les lois conditionnelles seront parfois ce qui apparaît naturellement dans la modélisation. Un schéma fréquent est celui où Y désigne le résultat d'une expérience aléatoire dont les conditions sont définies par la valeur de X ; l'interprétation de l'expérience fournira alors les valeurs des probabilités conditionnelles $\mathbb{P}_{(X=x)}(Y = y)$; desquelles on pourra déduire lois conjointe et marginales.

Le résultat suivant est très important en pratique : il permet de déterminer la loi marginale d'une des deux variables en disposant de la loi marginale de l'autre, et de lois conditionnelles :

Théorème 6.

$$\forall\, x \in X(\Omega),\ \mathbb{P}(X = x) = \sum_{y\in Y(\Omega)} \mathbb{P}_{(Y=y)}(X = x) \times \mathbb{P}(Y = y)$$

$$\forall\, y \in Y(\Omega),\ \mathbb{P}(Y = y) = \sum_{x\in X(\Omega)} \mathbb{P}_{(X=x)}(Y = y) \times \mathbb{P}(X = x)$$

Démonstration. Ce n'est qu'une reformulation du théorème 4 ! □

Remarque 4. Il est inutile d'apprendre ces formules : elles ne sont que des conséquences immédiates de la formule des probabilités totales, appliquée aux SCE donnés par les variables X et Y.
Il suffit de retenir la démarche (à l'œuvre dans le théorème 4 comme dans le théorème 6) :

pour obtenir la loi de Y, on applique la formule des probas totales avec le SCE associé à la variable X.

3.3 Cas d'école d'utilisation de ces formules : une « expérience en deux étapes »

Exemple 9. On considère l'expérience aléatoire suivante (avec $n \in \mathbb{N}^*$) :

- On dispose d'une urne U contenant n boules numérotées de 1 à n. On tire au hasard, de manière équiprobable, une boule dans cette urne ; X est la variable aléatoire égale à ce numéro.
- Si $X = k$, on tire au hasard une boule dans une urne qui contient cette fois des boules numérotées de 1 à k. On note Y le numéro de cette seconde boule.

1. Déterminer $Y(\Omega)$.

Méthode : détermination de $Y(\Omega)$

Ici, pour chaque valeur de X, la variable Y peut prendre un certain nombre de valeurs. Il faut « collecter » toutes ces valeurs. Un moyen est de le formaliser par une union indexée par $X(\Omega)$.

2. Calculer, pour $(k,\ell) \in X(\Omega) \times Y(\Omega)$, la proba conditionnelle $\mathbb{P}_{(X=k)}(Y = \ell)$.

3. En déduire la loi de Y (on obtiendra une somme qu'on ne cherchera pas à calculer explicitement).

4 Indépendance

4.1 Définition

Définition 4. *Soient* X *et* Y *deux VAD définies sur un même univers* Ω. *On dit que X et Y sont indépendantes ssi :*

$$\forall (x,y) \in X(\Omega) \times Y(\Omega),\ \mathbb{P}\big((X = x) \cap (Y = y)\big) = \mathbb{P}(X = x) \times \mathbb{P}(Y = y)$$

Ceci signifie que, pour tous (x, y), les événements $(X = x)$ et $(Y = y)$ sont indépendants.
On déduit en fait de cela que (on respire un grand coup) :

Si X et Y sont indépendantes, alors « tout événement dépendant uniquement de X est indépendant de tout événement dépendant uniquement de Y » .

Par exemple, si X et Y sont indépendantes, et si $x \in \mathbb{R}$:

- $(X \leqslant x)$ et (Y est pair) sont des événements indépendants ;
- $(X \geqslant 0)$ et $(Y \leqslant 1)$ sont des événements indépendants.
- Par contre, on ne peut rien affirmer sur l'indépendance de $(X \geqslant 0)$ et $(X \leqslant Y)$ car le second événement fait intervenir *à la fois* X et Y.

Remarque 5. Si X est une variable aléatoire quelconque, et a une variable aléatoire constante, alors X et a sont indépendantes.

Remarque 6. L'indépendance peut se lire dans le tableau de la loi conjointe de X et Y : deux VAD sont indépendantes ssi toute proba de la loi conjointe est obtenue comme produit des deux probas des lois marginales correspondantes.
Par exemple, X et Y de loi conjointe

X\ Y	0	1	loi de X
0	0.3	0.1	0.4
1	0.1	**0.5**	0.6
loi de Y	0.4	0.6	

ne sont pas indépendantes car 0.5 (en gras) n'est pas égal à 0.6×0.6.

Par contre X et Y de loi conjointe

X\ Y	0	1	loi de X
0	0.16	0.24	0.4
1	0.24	0.36	0.6
loi de Y	0.4	0.6	

sont indépendantes.

On remarque d'ailleurs (et c'est évident avec la définition) que : *pour des variables indépendantes, les lois marginales fournissent la loi conjointe.* On a vu que c'était faux en général.

Remarque 7. Méthodologie : montrer la (non-)indépendance de deux variables aléatoires.
On rappelle la définition : X et Y sont indépendantes ssi

$$\forall (x,y) \in X(\Omega) \times Y(\Omega),\ \mathbb{P}\big((X = x) \cap (Y = y)\big) = \mathbb{P}(X = x) \times \mathbb{P}(Y = y)$$

Pour montrer que X et Y sont indépendantes, il sera donc nécessaire de vérifier cette propriété, en traitant effectivement toutes les valeurs de x et y (d'où : gestion du quantificateur dans la rédaction).

NB : éviter les considérations philosophiques proclamant que les variables sont indépendantes parce que le déroulement de l'expérience laisse penser que leurs valeurs ne semblent pas corrélées, blablabla... C'est parfois trompeur, et de toute façon ça ne montre rien mathématiquement, et par conséquent ça ne rapporte pas de point.

Pour montrer que X et Y *ne sont pas* indépendantes, on considère la négation de la propriété d'indépendance : il suffit de trouver **un** couple (x, y) tel que $\mathbb{P}\big((X = x) \cap (Y = y)\big) \neq \mathbb{P}(X = x) \times P(Y = y)$ (un contre-exemple).
À vous de voir où on peut trouver cela... Une situation assez fréquente est celle où $\mathbb{P}\big((X = x) \cap (Y = y)\big) = 0$ alors que $\mathbb{P}(X = x) \neq 0$ et $\mathbb{P}(Y = y) \neq 0$: deux événements possibles, mais qui ne peuvent pas avoir lieu simultanément.
Par exemple, sur 3 lancers P/F, si X est le nombre de Pile obtenus, et F le nombre de Face, $\mathbb{P}(X = 3)$ et $\mathbb{P}(Y = 3)$ sont non nulles ; mais $\mathbb{P}\big((X = 3) \cap (Y = 3)\big)$ est nulle (impossible d'obtenir 3 Pile *et* 3 Face en 3 lancers...). X et Y ne sont donc pas indépendantes.

On peut généraliser la notion d'indépendance à n variables aléatoires : on parle alors d'*indépendance mutuelle*.

Définition 5. *Soient* $X_1, ..., X_n$ *des VAD définies sur un même univers* Ω. *On dit qu'elles sont mutuellement indépendantes ssi :*

$$\forall\,(x_1, \dots, x_n) \in X_1(\Omega) \times \cdots \times X_n(\Omega),\ \mathbb{P}\left(\bigcap_{i=1}^{n} (X_i = x_i)\right) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i)$$

On montre alors les propriétés suivantes :

Proposition 7.

- *On peut écrire une égalité similaire à celle de la définition sur des événements quelconques dépendant des* X_i. *Si* $X_1, ..., X_n$ *sont mutuellement indépendantes, on aura par exemple :*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} (X_i \geqslant i)\right) = \prod_{i=1}^{n} \mathbb{P}(X_i \geqslant i)$$

- *Si* $(X_1, ..., X_n)$ *sont mutuellement indépendantes, alors toute sous-famille de* $(X_1, ..., X_n)$ *est constituée de VAD mutuellement indépendantes.*
- *En particulier,* $\forall\, i \neq j$, *les VAD* X_i *et* X_j *sont indépendantes (on dit que les* X_i *sont* deux à deux indépendantes*)*

Remarque 8. La réciproque de ce dernier résultat est fausse : si $(X_1, ..., X_n)$ sont deux à deux indépendantes, elles ne sont pas forcément mutuellement indépendantes.

Considérons par exemple deux variables indépendantes X et Y suivant $\mathscr{B}(1/2)$; et Z la VAD valant 0 si X = Y, et 1 sinon. On peut vérifier que X et Z sont indépendantes en vérifiant $\mathbb{P}\big((X = i) \cap (Z = j)\big) = \mathbb{P}(X = i) \times \mathbb{P}(Z = j)$ pour tous (i, j) ; de même Y et Z le seront. X,Y,Z sont donc deux à deux indépendantes. Par contre X,Y,Z ne sont pas mutuellement indépendantes : par exemple $\mathbb{P}\big((X = 1) \cap (Y = 1) \cap (Z = 1)\big) = 0$.

Par convention, quand on parle de variables indépendantes, elles seront mutuellement indépendantes.

4.2 Fonctions de variables aléatoires indépendantes ; lemme des coalitions

Le principe du résultat ci-dessous est le suivant : des fonctions de VAD indépendantes sont encore indépendantes. Ceci se traduit par le *lemme des coalitions* :

Théorème 8 (Lemme des coalitions)**.** *Soient* $X_1, ..., X_n$ *des VAD indépendantes. Soit* $k \in [\![1, n-1]\!]$, $f : \mathbb{R}^k \to \mathbb{R}$, *et* $g : \mathbb{R}^{n-k} \to \mathbb{R}$; *on définit*

$$Z_1 = f(X_1, ..., X_k) \qquad \text{et} \qquad Z_2 = g(X_{k+1}, ..., X_n)$$

Alors Z_1 *et* Z_2 *sont indépendantes.*

Démonstration. Admis. □

Exemple 10. Si X,Y,Z,T sont des VAD indépendantes, alors

- $XY + Z$ et $T^2 + 1$ sont deux variables indépendantes ;
- Y^2 et $X + Z$ sont indépendantes ;
- e^X et $\sqrt{1+Z^2}$ sont indépendantes ;
- par contre XY et YT ne le sont pas forcément, car Y « intervient » dans les deux variables.

Remarque 9. Ce résultat est d'utilisation plus fréquente que ce qu'on pourrait penser, et repérer les endroits où il faut l'utiliser est un exercice subtil. Nous le rencontrerons à plusieurs reprises ; voir plus bas la démonstration concernant l'espérance du produit de n variables indépendantes.

On a un corollaire simple, et d'usage fréquent, dans le cas de deux variables :

Corollaire 9. *Si X et Y sont deux VAD indépendantes, et f et g deux fonctions, alors $f(X)$ et $g(Y)$ sont indépendantes.*

5 Étude d'une variable aléatoire de la forme $g(X, Y)$

5.1 Somme de deux variables aléatoires

Soient X et Y deux VAD sur Ω. On considère la VAD $Z = X + Y$; on va déterminer la loi de Z.
La méthode générale est de scinder cet événement selon les valeurs possibles de X :

$$(X+Y=n) = \bigsqcup_{k\in X(\Omega)} \big((X=k)\cap(Y=n-k)\big)$$

ou selon celles de Y :

$$(X+Y=n) = \bigsqcup_{j\in Y(\Omega)} \big((X=n-j)\cap(Y=j)\big)$$

Par incompatibilité (union disjointe $\sqcup$) on en déduit :

Proposition 10. *Soit $n \in Z(\Omega)$. On a*

$$\mathbb{P}(Z=n) = \sum_{k\in X(\Omega)} \mathbb{P}\big((X=k)\cap(Y=n-k)\big) = \sum_{j\in Y(\Omega)} \mathbb{P}\big((X=n-j)\cap(Y=j)\big)$$

Remarque 10. Certains événements de la forme $\big((X=k)\cap(Y=n-k)\big)$ peuvent être impossibles : la probabilité associée sera nulle. Typiquement, ceci aura pour effet de « tronquer » la somme $\sum_k$ (cf. exemple ci-dessous).

Exemple 11. On lance deux dés équilibrés à 6 faces. Donner la probabilité que la somme des deux dés donne 9.

Exemple 12. Soient X et Y deux variables indépendantes suivant $\mathscr{G}(p)$. Donner la loi de $X+Y$.

5.2 Cas particuliers de somme : théorèmes de stabilité

Pour des variables suivant les lois usuelles, il existe des résultats donnant simplement la loi de la somme. **Ils sont valables dans le cas de variables indépendantes**.

Théorème 11 (Stabilité de la loi binomiale)**.**

- *Soient* $X \hookrightarrow \mathscr{B}(n_1, p)$ *et* $Y \hookrightarrow \mathscr{B}(n_2, p)$ *deux variables aléatoires indépendantes.*
  Alors $X+Y \hookrightarrow \mathscr{B}(n_1+n_2, p)$.
- *Plus généralement, si les* X_i *sont mutuellement indépendantes et suivent des lois binômiales* $\mathscr{B}(n_i, p)$, *alors la somme* $X_1+\cdots+X_n$ *suit* $\mathscr{B}(\sum n_i, p)$.

Démonstration. En annexe. □

Remarque 11. On retrouve le fait que la somme de n variables de Bernoulli indépendantes de même loi $\mathscr{B}(p)$ suit $\mathscr{B}(n, p)$: en effet $\mathscr{B}(p) = \mathscr{B}(1, p)$.

Il existe un résultat similaire pour des VAD suivant une loi de Poisson :

Théorème 12 (Stabilité de la loi de Poisson)**.**

- *Soient* $X \hookrightarrow \mathscr{P}(\lambda)$ *et* $Y \hookrightarrow \mathscr{P}(\mu)$ *deux variables aléatoires indépendantes. Alors* $X+Y \hookrightarrow \mathscr{P}(\lambda+\mu)$.
- *Plus généralement, si les* X_i *sont mutuellement indépendantes et suivent des lois de Poisson* $\mathscr{P}(\lambda_i)$, *alors la somme* $X_1+\cdots+X_n$ *suit* $\mathscr{P}(\sum \lambda_i)$.

Démonstration. En annexe. □

5.3 Loi et théorème de transfert

Théorème 13. *Soient X et Y deux VAD sur Ω, et g une fonction définie sur l'ensemble*

$$X(\Omega) \times Y(\Omega) = \big\{(x, y) \mid x \in X(\Omega), y \in Y(\Omega)\big\}$$

des couples constitués d'une valeur de X et d'une valeur de Y.
Alors $Z = g(X, Y)$ *est aussi une VAD. On a, pour tout* $z \in Z(\Omega)$ *:*

$$\mathbb{P}(Z = z) = \sum_{(x,y) \mid g(x,y)=z} \mathbb{P}\big((X = x) \cap (Y = y)\big)$$

Démonstration. L'événement $(Z = z)$ s'écrit comme l'union disjointe $\displaystyle\bigsqcup_{(x,y) \mid g(x,y)=z} \big((X = x) \cap (Y = y)\big)$. □

Il est difficile de donner plus de résultats ayant une portée générale. Pour une application pratique, voir des exemples de loi de la somme dans la section suivante.

Le théorème de transfert donne, sous réserve d'existence, une formule pour le calcul de l'espérance d'une fonction de X et Y. Ce théorème est très utile : il permet de calculer $\mathbb{E}(g(X, Y))$ *sans connaître la loi de* $g(X, Y)$.

Théorème 14 (Théorème de transfert)**.**
$Z = g(X, Y)$ *admet une espérance si et seulement si la série double*

$$\sum_{(x,y)\in X(\Omega)\times Y(\Omega)} g(x, y)\mathbb{P}\big((X = x) \cap (Y = y)\big)$$

converge ***absolument****. Dans ce cas on a :*

$$\mathbb{E}\big(g(X, Y)\big) = \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} g(x, y)\mathbb{P}\big((X = x) \cap (Y = y)\big)$$

Démonstration. Admis. □

On obtient alors une propriété très importante de l'espérance :

Théorème 15 (Linéarité de l'espérance)**.**
Soient X et Y deux VAD sur un univers Ω *admettant une espérance, et* $\lambda \in \mathbb{R}$*. Alors :*

- $\lambda X + Y$ *admet une espérance ;*
- *on a* $\mathbb{E}(\lambda X + Y) = \lambda \mathbb{E}(X) + \mathbb{E}(Y)$

En particulier, si $\beta \in \mathbb{R}$*, on peut le traiter comme une VAD constante (donc d'espérance* β*) ; et on obtient :*
Si X admet une espérance, et si $(\alpha, \beta) \in \mathbb{R}^2$*, alors* $\alpha X + \beta$ *admet une espérance et*

$$\mathbb{E}(\alpha X + \beta) = \alpha \mathbb{E}(X) + \beta$$

On peut aussi généraliser à une combinaison linéaire quelconque : si les variables X_i *admettent des espérances, et si les* λ_i *sont des réels, alors*

- $\sum_{i=1}^{n} \lambda_i X_i$ *admet une espérance ;*
- *on a* $\mathbb{E}\left(\sum_{i=1}^{n} \lambda_i X_i\right) = \sum_{i=1}^{n} \lambda_i \mathbb{E}(X_i)$.

Démonstration. En annexe. □

Remarque 12. Contrairement à plusieurs résultats qui suivront, **la linéarité de l'espérance est valable même si les variables ne sont pas indépendantes.**

5.4 Espérance d'un produit de VAD indépendantes

Ce résultat est très important en pratique :

Théorème 16 (Espérance du produit de VAD indépendantes)**.** *Soient X et Y deux VAD* ***indépendantes*** *admettant une espérance. Alors*

- XY *admet une espérance ;*
- *on a :* $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Démonstration. En annexe. □

Remarque 13. L'hypothèse d'indépendance intervient aussi dans la première partie du résultat énoncé par le théorème : si X et Y sont deux VAD quelconques admettant des espérances, on ne peut pas affirmer que XY admet une espérance.

Corollaire 17.

- *Soit* $n \geqslant 2$*. Si* $X_1, \ldots, X_n$ *sont mutuellement indépendantes et admettent des espérances, alors* $\prod_{i=1}^{n} X_i$ *admet également une espérance, et*
$$\mathbb{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} \mathbb{E}(X_i)$$
- *Si X et Y sont indépendantes, toutes VAD de la forme* $f(X)$ *et* $g(Y)$ *le sont également : on aura donc, pour toutes fonctions* f *et* g *telles que* $\mathbb{E}\big(f(X)\big)$ *et* $\mathbb{E}\big(g(Y)\big)$ *existent :*
$$\mathbb{E}\big(f(X)g(Y)\big) \textit{ existe, et } \mathbb{E}\big(f(X)g(Y)\big) = \mathbb{E}\big(f(X)\big)\mathbb{E}\big(g(Y)\big)$$

Démonstration. On montre le premier point par récurrence.
Pour $n = 2$ on cherche à montrer que si X_1 et X_2 sont indépendantes, alors $\mathbb{E}(X_1X_2) = \mathbb{E}(X_1)\mathbb{E}(X_2)$: c'est le théorème précédent.
Soit $n \geqslant 2$; on suppose le théorème acquis dans le cas de n variables indépendantes.
Soient $X_1, X_2, \dots, X_{n+1}$ des variables aléatoires indépendantes. On peut écrire :

$$\begin{aligned}
\mathbb{E}(X_1X_2\dots X_{n+1}) &= \mathbb{E}\big((X_1X_2\dots X_n)X_{n+1}\big) \\
&= \mathbb{E}\big((X_1X_2\dots X_n)\big)\mathbb{E}(X_{n+1}) && \text{car } X_1X_2\dots X_n \text{ et } X_{n+1} \text{ sont indépendantes} \\
& && \text{d'après le lemme des coalitions} \\
&= \left(\prod_{i=1}^{n}\mathbb{E}(X_i)\right)\times\mathbb{E}(X_{n+1}) && \text{d'après l'hypothèse de récurrence au rang } n \\
&= \prod_{i=1}^{n+1}\mathbb{E}(X_i)
\end{aligned}$$

ce qui montre l'hérédité et achève la récurrence. □

Remarque 14. La réciproque du théorème 16 est fausse : si $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, on ne peut pas conclure que X et Y sont indépendantes.
Considérer par exemple X et Y de loi conjointe :

X\ Y	0	1	2
0	1/4	0	1/4
1	0	1/2	0

Montrer que $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, mais que X et Y ne sont pas indépendantes.

5.5 Maximum et minimum de variables aléatoires indépendantes

Si X et Y sont deux VAD, un exercice ultra-classique est de s'intéresser aux variables aléatoires $\min(X,Y)$ et $\max(X,Y)$ (ce n'est pas juste de la perversion ; nous verrons des applications concrètes de cela en fin d'année). La méthode permettant d'obtenir les lois de ces variables est de **passer par leur fonction de répartition**.

La chose à observer est la suivante :

Proposition 18. *Soient X et Y deux VAD, et $k \in \mathbb{R}$. On a :*

- $\big(\max(X,Y) \leqslant k\big) = (X \leqslant k) \cap (Y \leqslant k)$
- $\big(\min(X,Y) \geqslant k\big) = (X \geqslant k) \cap (Y \geqslant k)$

et les propriétés similaires si on remplace les inégalités larges par des inégalités strictes.

Remarque 15. Ceci se généralise facilement à n variables aléatoires. On écrira alors :

- $\big(\max(X_1,\dots,X_n) \leqslant k\big) = (X_1 \leqslant k) \cap \dots \cap (X_n \leqslant k)$
- $\big(\min(X_1,\dots,X_n) \geqslant k\big) = (X_1 \geqslant k) \cap \dots \cap (X_n \geqslant k)$

NB : ceci reste valable en l'absence de l'hypothèse d'indépendance.

On utilise maintenant l'indépendance de X et Y. En notant $Z_1 = \max(X,Y)$ et $Z_2 = \min(X,Y)$:

- $\mathbb{P}(Z_1 \leqslant k) = \mathbb{P}(X \leqslant k) \times \mathbb{P}(Y \leqslant k)$
- $\mathbb{P}(Z_2 \geqslant k) = \mathbb{P}(X \geqslant k) \times \mathbb{P}(Y \geqslant k)$

Si on suppose de plus que X et Y sont à valeurs dans $\mathbb{Z}$, on a : $\forall\, k \in \mathbb{Z},\ (Z \geqslant k) = (Z = k) \sqcup (Z \geqslant k+1)$, ce qui implique :

$$\forall\, k \in \mathbb{Z}, \mathbb{P}(Z = k) = \mathbb{P}(Z \geqslant k) - \mathbb{P}(Z \geqslant k+1)$$

De manière similaire, on montre :

$$\forall\, k \in \mathbb{Z}, \mathbb{P}(Z = k) = \mathbb{P}(Z \leqslant k) - \mathbb{P}(Z \leqslant k-1)$$

ce qui permet de revenir à l'expression de la loi.

Exemple 13. Soient X et Y indépendantes, suivant $\mathscr{U}(\llbracket 1, n \rrbracket)$. Soit $Z = \max(X, Y)$. Donner la loi de Z.

6 Variance, covariance et corrélation

6.1 Moments d'une VAD

Plusieurs démonstrations des résultats de cette section sont reportés en annexe.

On récapitule ici les propriétés d'existence des quantités considérées dans ce qui suit.
On rappelle :

Définition 6.
Soit X une VAD. On dit que X admet un moment d'ordre *k ($k \in \mathbb{N}^*$) ssi la série* $\sum_{x \in X(\Omega)} x^k \mathbb{P}(X = x)$ *converge absolument.*

On voit notamment que :

- *X admet un moment d'ordre 1 ssi elle admet une espérance ;*
- *X admet un moment d'ordre k ssi X^k admet une espérance.*

On a alors :

Théorème 19 (Existence de moments)**.**

- *Si X admet un moment d'ordre $k \in \mathbb{N}^*$, alors elle admet des moments d'ordre p pour tout $p \in \llbracket 0, k \rrbracket$ (HP dans son énoncé général)*
- *Corollaire : si X admet un moment d'ordre 2, elle admet une espérance.*
- *Corollaire : si X admet un moment d'ordre 2, elle admet une variance ; et en fait X admet un moment d'ordre 2 ssi elle admet une variance.*
- *Si X, Y admettent un moment d'ordre 2, alors $X + Y$ admet un moment d'ordre 2.*
- *Si X, Y admettent un moment d'ordre 2, alors XY admet une espérance.*

Démonstration. En annexe. □

Remarque 16.

- Le premier point dit qu'admettre des moments d'ordre élevé est une propriété « forte » . Par exemple une variable admettant un moment d'ordre 4 admet automatiquement une espérance, une variance et un moment d'ordre 3.
- Si $X(\Omega)$ est fini, X admet des moments de tous ordres ; notamment une espérance et une variance.
- Des variables suivant les lois usuelles (Bernoulli, binomiale, Poisson, géométrique) admettent des variances ; **apprenez les formules pour les espérances et les variances !!!**

6.2 Variance et covariance

Soit X une VAD admettant un moment d'ordre 2. Alors X admet une espérance, et on définit alors

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

On appelle V(X) la variance de X. On récapitule ici les propriétés de la variance :

- $V(X) \geqslant 0$.
- $V(X) = 0$ ssi X est presque sûrement constante.

  Démonstration. S'il existe $a \in \mathbb{R}$ tel que $\mathbb{P}(X = a) = 1$, alors $\mathbb{E}(X) = a$ et $(X - \mathbb{E}(X))^2 = 0$ avec une probabilité 1, d'où $\mathbb{E}\left((X - \mathbb{E}(X))^2\right) = 0$, et la variance est bien nulle.
  Si $V(X) = 0$, alors la VAD $(X - \mathbb{E}(X))^2$ est à valeurs positives, d'espérance nulle, donc est égale à 0 avec une probabilité 1. On en déduit que $X = \mathbb{E}(X)$ avec une probabilité égale à 1 : X est presque sûrement constante. □

- $\forall (a, b) \in \mathbb{R}^2$, $aX + b$ admet une variance, et $V(aX + b) = a^2 V(X)$.

On définit également l'écart-type de la variable X : $\sigma(X) = \sqrt{V(X)}$.

Le théorème suivant fournit une expression simple de la variance :

Théorème 20 (Koenig-Huygens)**.** *Soit X admettant une variance. Alors* $V(X) = \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2$.

Démonstration. X admet donc une espérance. En utilisant la linéarité de l'espérance :

$$V(X) = E\left((X - \mathbb{E}(X))^2\right) = E\left(X^2 - 2\underbrace{\mathbb{E}(X)}_{=\text{cste}}X + \underbrace{\mathbb{E}(X)^2}_{=\text{cste}}\right) = \mathbb{E}\left(X^2\right) - 2\mathbb{E}(X)\mathbb{E}(X) + \mathbb{E}(X)^2 = E\left(X^2\right) - \mathbb{E}(X)^2$$

□

Dans le cas de VAD indépendantes, on a le résultat important suivant :

Proposition 21. *Si X et Y sont deux VAD indépendantes admettant une variance, alors* $V(X + Y) = V(X) + V(Y)$.

Démonstration. On utilise Koenig-Huygens et la linéarité de l'espérance :

$$\begin{aligned} V(X + Y) = \mathbb{E}\left((X + Y)^2\right) - (\mathbb{E}(X + Y))^2 &= \mathbb{E}\left(X^2\right) + 2\mathbb{E}(XY) + \mathbb{E}\left(Y^2\right) - (\mathbb{E}(X)^2 + 2\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(Y)^2) \\ &= \left[\mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2\right] + \left[\mathbb{E}\left(Y^2\right) - \mathbb{E}(Y)^2\right] + 2\left[\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\right] \\ &= V(X) + V(Y) + 2\left[\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\right] \\ &= V(X) + V(Y) \quad \text{par indépendance de X et Y} \end{aligned}$$

□

Le calcul précédent fait apparaître une quantité qui est la différence entre $V(X + Y)$ et $V(X) + V(Y)$. On peut songer à l'utiliser pour « mesurer la dépendance » entre ces deux VAD.

Définition 7. *Soient X et Y deux VAD admettant des variances. On appelle covariance de X et Y (qui est bien définie d'après le théorème 19) :*

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Le résultat sur l'espérance d'un produit de VAD indépendantes donne alors :

Théorème 22. *Si* X, Y *sont deux VAD indépendantes, alors* $\text{Cov}(X, Y) = 0$.

Remarque 17. Il découle de la remarque 14 que la réciproque de ce résultat est fausse. Deux VAD de covariance nulle sont dites *décorrélées* ; la décorrélation n'équivaut pas à l'indépendance.

La covariance mesure la manière dont deux variables aléatoires « varient ensemble » : si une expérience aléatoire fournit une l'issue ω, qui mène à deux valeurs $X(\omega)$ et $Y(\omega)$, on peut se demander si « plus X est grand, plus Y est grand » ; ou si « plus X est grand, plus Y est petit » (NB : tout en sachant que cela ne sera pas une règle absolue, mais seulement une tendance).
Le premier cas correspondra à une covariance positive ; et le second à une covariance négative.

Voir les exercices pour des exemples de chaque comportement. À noter que cette interprétation fait régulièrement l'objet d'une question de concours. Une phrase de « bon sens » fournira (une fois n'est pas coutume) des points facilement gagnés.

6.3 Coefficient de corrélation linéaire

La covariance sert à quantifier l'interdépendance entre deux VAD ; son signe donne une tendance croissante ou décroissante. On peut penser que sa valeur absolue augmentera si cette corrélation est plus affirmée.

Une bonne idée pour exploiter cette valeur absolue peut être de se débarrasser, dans ce coefficient, de *l'échelle* des VAD considérées : il suffit pour cela de diviser par les ordres de grandeur typiques de X et Y, qu'on caractérise à l'aide de l'écart-type. On introduit alors le *coefficient de corrélation* :

Définition 8. *Si X et Y sont deux VAD de variance non nulle, on appelle coefficient de corrélation entre X et Y la quantité :*

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

Ce coefficient appartient à $[-1,1]$:

Théorème 23. *Pour toutes VAD X et Y de variance non nulle,* $\left|\rho(X,Y)\right| \leqslant 1$.

Démonstration. En annexe. □

Remarque 18. On a donc bien éliminé les ordres de grandeur de X et Y, en procédant à une *normalisation*.

On dispose également d'une propriété lorsque les valeurs extrêmes sont atteintes :

Théorème 24. *Soient X et Y deux VAD de variances non nulles.*
$\rho(X,Y) = \pm 1$ *si et seulement si il existe* $a \in \mathbb{R}^*$, $b \in \mathbb{R}$ *tels que* $Y = aX + b$ *presque sûrement.*
Dans ce cas, le signe de $\rho(X,Y)$ *est celui de* a.

Démonstration. En annexe. □

Remarque 19. Ceci permet de préciser l'intuition vue en introduction : une covariance grande en valeur absolue signale que les points représentatifs $(X(\omega), Y(\omega))$, pour les différentes issues ω, se situeront approximativement sur une droite.

6.4 Propriétés algébriques

Les propriétés suivantes sont assez importantes en pratique. Elles se démontrent en revenant à la définition, principalement à l'aide de la linéarité de l'espérance.

Proposition 25 (Règles de calcul sur les variances/covariances)**.**
Dans ce qui suit, X, X', Y, Y' *sont des VAD admettant des variances.*

- $\mathrm{Cov}(X,Y) = \mathbb{E}\Big(\big(X-\mathbb{E}(X)\big)\big(Y-\mathbb{E}(Y)\big)\Big)$.
- $\mathrm{Cov}(X,X) = V(X)$.
- $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$ *(la covariance est* symétrique*).*
- *Le calcul de la proposition 21 montre que, pour toutes variables X et Y admettant des variances :* $V(X+Y) = V(X)+V(Y)+2\mathrm{Cov}(X,Y)$.
- *La covariance est* bilinéaire*, c'est-à-dire linéaire en chacune des deux variables. Pour tous* a, b *réels :*

$$\mathrm{Cov}(aX+X', Y) = a\,\mathrm{Cov}(X,Y) + \mathrm{Cov}(X',Y)$$

$$\mathrm{Cov}(X, bY+Y') = b\,\mathrm{Cov}(X,Y) + \mathrm{Cov}(X,Y')$$

- *Si* $Y = a$ *est une VAD presque sûrement constante (c'est-à-dire* $\mathbb{P}(Y=a)=1$*, et* $\forall\, y \neq a$, $\mathbb{P}(Y=y)=0$*, alors* $\mathrm{Cov}(X,Y)=0$*. On écrit plus simplement*

$$\mathrm{Cov}(X,a) = 0$$

On peut ainsi écrire, par exemple :

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

ou

$$\begin{aligned}\mathrm{Cov}(X-Y, X+Y) &= \mathrm{Cov}(X,X)+\mathrm{Cov}(X,Y)-\mathrm{Cov}(Y,X)-\mathrm{Cov}(Y,Y)\\ &= V(X)+\mathrm{Cov}(X,Y)-\mathrm{Cov}(X,Y)-V(Y) \qquad \text{(symétrie)}\\ \mathrm{Cov}(X-Y,X+Y) &= V(X)-V(Y)\end{aligned}$$

(NB : et constater une certaine ressemblance avec une identité remarquable...)

7 Inégalités de concentration : Markov et Bienaymé-Tchebychev

En probas on appelle inégalité de concentration toute inégalité qui permet de contrôler la probabilité qu'une variable aléatoire « dévie de sa valeur moyenne » (à prendre dans un sens assez large ; voir les exemples).
Il y en a deux à votre programme[2] : l'inégalité de Markov et l'inégalité de Bienaymé-Tchebychev.

7.1 Inégalité de Markov

Théorème 26 (Inégalité de Markov)**.** *Soit X une variable aléatoire* ***positive****, admettant une espérance. On a :*

$$\forall\, t>0,\ \mathbb{P}(X \geqslant t) \leqslant \frac{\mathbb{E}(X)}{t}$$

Démonstration. En annexe. □

[2] L'inégalité de Markov est inscrite dans le programme, tout en étant « non exigible » : comprenne qui pourra.

On voit que l'inégalité de Markov permet de majorer la probabilité que X devienne « très grande » .
On sait déjà que cette probabilité tend vers 0 : si on se place dans le cas $X(\Omega) = \mathbb{N}$, $\mathbb{P}(X > n)$ est le reste partiel de la série convergente $\sum \mathbb{P}(X = n)$.
Ici on a plus fort : $\mathbb{P}(X > n)$ tend vers 0 *plus vite que* $\frac{1}{n}$.
On verra dans un exercice comment exploiter cette inégalité quand X admet des moments d'ordre supérieur.

Remarque 20. Comme $X > t \Rightarrow X \geqslant t$, on a l'inclusion entre événements $(X > t) \subset (X \geqslant t)$ et donc on peut aussi écrire $\mathbb{P}(X > t) \leqslant \mathbb{P}(X \geqslant t) \leqslant \frac{\mathbb{E}(X)}{t}$.

7.2 Inégalité de Bienaymé-Tchebychev

L'inégalité de Bienaymé-Tchebychev permet, pour une variable admettant une variance, de quantifier la probabilité que la variable *s'écarte de sa valeur moyenne.* La variance mesurant la propension d'une variable aléatoire à s'écarter de son espérance, il est raisonnable de la voir intervenir ici.

Théorème 27 (Inégalité de Bienaymé-Tchebychev)**.** *Soit $t > 0$ et X une variable admettant une variance. Alors X admet aussi une espérance, et on a :*

$$\mathbb{P}\big(|X - \mathbb{E}(X)| \geqslant t\big) \leqslant \frac{V(X)}{t^2}$$

Démonstration. En annexe. □

Remarque 21. Avec les mêmes arguments que pour Markov, on peut aussi écrire $\mathbb{P}\big(|X - \mathbb{E}(X)| > t\big) \leqslant \frac{V(X)}{t^2}$.

8 (Hors programme :)Théorèmes de limite monotone

Ces deux théorèmes sont sortis du programme mais sont très importants : ils permettent de calculer les probabilités d'unions ou d'intersections infinies d'événements.

Théorème 28.

- *Soit* (A_n) *une suite croissante d'événements : alors*

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

- *Soit* (B_n) *une suite décroissante d'événements : alors*

$$\mathbb{P}\left(\bigcap_{n=0}^{+\infty} B_n\right) = \lim_{n\to+\infty} \mathbb{P}(B_n)$$

Démonstration. On rappelle qu'on dit qu'une suite d'événements (A_n) est croissante (pour l'inclusion) ssi :

$$\forall n \in \mathbb{N},\ A_n \subset A_{n+1}$$

Soit une telle suite (A_n) : pour $n \geqslant 1$, on pose $B_n = A_n \setminus A_{n-1}$ et $B_0 = A_0$. On vérifie alors que les B_i sont 2 à 2 disjoints, et que

$$\forall n \in \mathbb{N},\ A_n = \bigsqcup_{k=0}^{n} B_k \quad \text{et} \quad \bigcup_{n=0}^{+\infty} A_n = \bigsqcup_{k=0}^{+\infty} B_k$$

Par propriété d'une probabilité, $\mathbb{P}\left(\bigcup_{n=0}^{+\infty} A_n\right) = \mathbb{P}\left(\bigsqcup_{k=0}^{+\infty} B_k\right) = \sum_{k=0}^{+\infty} \mathbb{P}(B_k)$; et $\mathbb{P}(A_N) = \mathbb{P}\left(\bigsqcup_{k=0}^{N} B_k\right) = \sum_{k=0}^{N} \mathbb{P}(B_k)$. En faisant tendre N vers $+\infty$,

$$\lim_{N\to+\infty} \mathbb{P}(A_N) = \sum_{k=0}^{+\infty} \mathbb{P}(B_k) = \mathbb{P}(\bigcup_{n=0}^{+\infty} A_n)$$

Pour la seconde propriété : si (B_n) est décroissante pour l'inclusion, la suite des complémentaires $(\overline{\mathrm{B}_n})$ est croissante pour l'inclusion et on peut donc lui appliquer le premier résultat :

$$\mathbb{P}\left(\bigcup_{n=0}^{+\infty} \overline{\mathrm{B}_n}\right) = \lim_{n\to+\infty} \mathbb{P}(\overline{\mathrm{B}_n})$$

et donc

$$\mathbb{P}\left(\overline{\bigcup_{n=0}^{+\infty} \overline{\mathrm{B}_n}}\right) = 1 - \lim_{n\to+\infty} \mathbb{P}(\overline{\mathrm{B}_n}) = \lim_{n\to+\infty} 1 - \mathbb{P}(\overline{\mathrm{B}_n}) = \lim_{n\to+\infty} \mathbb{P}(\mathrm{B}_n)$$

et avec de Morgan

$$\mathbb{P}\left(\bigcap_{n=0}^{+\infty} \mathrm{B}_n\right) = \lim_{n\to+\infty} \mathbb{P}(\mathrm{B}_n)$$

□

En conséquence :

Proposition 29. *Pour toute suite d'événements* (A_n)*, on a :*

$$\mathbb{P}\left(\bigcup_{k=0}^{+\infty} \mathrm{A}_k\right) = \lim_{n\to+\infty} \mathbb{P}\left(\bigcup_{k=0}^{n} \mathrm{A}_k\right)$$

$$\mathbb{P}\left(\bigcap_{k=0}^{+\infty} \mathrm{A}_k\right) = \lim_{n\to+\infty} \mathbb{P}\left(\bigcap_{k=0}^{n} \mathrm{A}_k\right)$$

Démonstration. On applique les propositions précédentes :

- à la suite croissante d'événements $\left(\bigcup_{k=0}^{n} \mathrm{A}_k\right)_{n\in\mathbb{N}}$.
- à la suite décroissante d'événements $\left(\bigcap_{k=0}^{n} \mathrm{A}_k\right)_{n\in\mathbb{N}}$.

□

9 Démonstrations

Théorème 15 (linéarité de l'espérance)

Soient X et Y deux VAD sur un univers Ω admettant une espérance, et $\lambda \in \mathbb{R}$. On a

$$\mathbb{E}(\lambda X + Y) = \lambda \mathbb{E}(X) + \mathbb{E}(Y)$$

Démonstration. On souhaite appliquer le théorème de transfert à $Z = \lambda X + Y$. Il faut donc vérifier que la série $\sum_{(x,y)\in X(\Omega)\times Y(\Omega)} (\lambda x + y)\mathbb{P}\big((X = x)\cap(Y = y)\big)$ converge.
$\sum_{x\in X(\Omega)} \sum_{y\in Y(\Omega)} x\mathbb{P}\big((X = x)\cap(Y = y)\big) = \sum_{x\in X(\Omega)} x \sum_{y\in Y(\Omega)} \mathbb{P}\big((X = x)\cap(Y = y)\big)$. La somme sur y est absolument convergente (à termes positifs ; de somme $\mathbb{P}(X = x)$ par probabilités totales). En sommant on obtient alors $\sum_{x\in X(\Omega)} x\mathbb{P}(X = x)$ qui est aussi absolument convergente car X admet une espérance. Cette somme vaut $\mathbb{E}(X)$.
De même, la série double $\sum\sum y\mathbb{P}\big((X = x)\cap(Y = y)\big)$ converge absolument, et a pour somme $\mathbb{E}(Y)$. On peut finalement appliquer la formule de transfert, et on obtient le résultat annoncé. □

Théorème 11 (Stabilité de la loi binomiale)

- **Soient** $X \hookrightarrow \mathscr{B}(n_1, p)$ **et** $Y \hookrightarrow \mathscr{B}(n_2, p)$ **deux variables aléatoires indépendantes.**
  Alors $X + Y \hookrightarrow \mathscr{B}(n_1 + n_2, p)$**.**
- **Plus généralement, si les** X_i **sont mutuellement indépendantes et suivent des lois binômiales** $\mathscr{B}(n_i, p)$**, alors la somme** $X_1 + \cdots + X_n$ **suit** $\mathscr{B}(\sum n_i, p)$**.**

Démonstration.

Préliminaire : formule de Vandermonde
On démontre :

$$\forall\, (n_1, n_2) \in \mathbb{N}^2,\ \forall\, k \in [\![0, n_1 + n_2]\!],\ \sum_{i=0}^{k} \binom{n_1}{i}\binom{n_2}{k-i} = \binom{n_1+n_2}{k}$$

(en posant $\binom{n}{p} = 0$ si $p > n$).
On va dénombrer les parties à k éléments d'un ensemble à $n_1 + n_2$ éléments.
Si E est un ensemble et $k \in [\![0, \operatorname{Card}(E)]\!]$, on note $\mathscr{P}_k(E)$ le nombre d eparties de E à k éléments. Soit E un ensemble à $n_1 + n_2$ éléments ; on peut écrire $E = A \cup B$ avec A à n_1 éléments, B à n_2 éléments, et $A \cap B = \varnothing$.

Une partie de E est à k élements ssi on peut l'écrire comme l'union d'une partie de A à i éléments, et d'une partie de B à $k - i$ éléments (où $i \in [\![0, k]\!]$). Cette écriture est unique.
On peut donc écrire :

$$\mathscr{P}_k(E) = \bigcup_{i=0}^{k} \big\{C_1 \cup C_2 \mid (C_1, C_2) \in \mathscr{P}_i(A) \times \mathscr{P}_{k-i}(B)\big\}$$

et donc en prenant les cardinaux :

$$\binom{n_1+n_2}{k} = \sum_{i=0}^{k} \binom{n_1}{i}\binom{n_2}{k-i}$$

NB : si $i > n_1$, $\mathscr{P}_i(A) = \varnothing$ et donc $\operatorname{Card}(\mathscr{P}_i(A)) = 0 = \binom{n_1}{i}$ donc les formules restent vraies avec la convention choisie.
Montrons maintenant le résultat de stabilité. On commence par montrer le premier point.
$X + Y$ est bien à valeurs dans $[\![0, n_1 + n_2]\!]$.

Pour tout $k \in [\![0, n_1 + n_2]\!]$, on a :

$$\begin{aligned}
\mathbb{P}(X_1 + X_2 = k) &= \sum_{i=0}^{k} \mathbb{P}(X_1 = i, X_2 = k - i) \\
&= \sum_{i=0}^{k} \mathbb{P}(X_1 = i)\mathbb{P}(X_2 = k - i) \\
&= \sum_{i=0}^{k} \binom{n_1}{i} p^i q^{n_1 - i} \binom{n_2}{k-i} p^{k-i} q^{n_2 - k + i} \\
&= p^k q^{n_1 + n_2 - k} \underbrace{\sum_{i=0}^{k} \binom{n_1}{i}\binom{n_2}{k-i}} \\
\mathbb{P}(X_1 + X_2 = k) &= \binom{n_1 + n_2}{k} p^k q^{n_1 + n_2 - k} \quad \text{(d'après le préliminaire)}
\end{aligned}$$

Le second point se montre par récurrence sur le nombre $p \geqslant 2$ de variables. L'hérédité utilise que $X_1 + \ldots + X_k$ et X_{k+1} sont indépendantes ; c'est une conséquence du lemme des coalitions. □

Théorème 12 (Stabilité de la loi de Poisson)

- **Soient** $X \hookrightarrow \mathscr{P}(\lambda)$ **et** $Y \hookrightarrow \mathscr{P}(\mu)$ **deux VA indépendantes. Alors** $X + Y \hookrightarrow \mathscr{P}(\lambda + \mu)$**.**
- **Plus généralement, si les** X_i **sont mutuellement indépendantes et suivent des lois de Poisson** $\mathscr{P}(\lambda_i)$**, alors la somme** $X_1 + \cdots + X_n$ **suit** $\mathscr{P}(\sum \lambda_i)$**.**

Démonstration. On a par indépendance, et en utilisant la formule du binôme de Newton :

$$\begin{aligned}
\mathbb{P}(X + Y = n) &= \sum_{k=0}^{n} \mathbb{P}(X = k)\mathbb{P}(Y = n - k) \\
&= e^{-(\lambda+\mu)} \sum_{k=0}^{n} \frac{\lambda^k}{k!} \frac{\mu^{n-k}}{(n-k)!} \\
&= \frac{e^{-(\lambda+\mu)}}{n!} \sum_{k=0}^{n} \binom{n}{k} \lambda^k \mu^{n-k} \\
\mathbb{P}(X + Y = n) &= \frac{e^{-(\lambda+\mu)}}{n!} (\lambda + \mu)^n
\end{aligned}$$

Le second point se montre par récurrence, avec le lemme des coalitions. □

Théorème 16 (espérance du produit de deux variables indépendantes)

Soient X et Y deux VAD indépendantes admettant une espérance. Alors XY **admet une espérance, et on a** $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

Démonstration.
D'après le théorème de transfert, on s'intéresse à la convergence absolue de la série double $\displaystyle\sum_{\substack{x \in X(\Omega) \\ y \in Y(\Omega)}} xy\mathbb{P}\big((X = x) \cap (Y = y)\big)$.

On peut écrire, sous réserve de convergence :

$$\sum_{\substack{x \in X(\Omega) \\ y \in Y(\Omega)}} |xy|\mathbb{P}\big((X = x) \cap (Y = y)\big) = \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} |xy|\mathbb{P}(X = x)\mathbb{P}(Y = y) = \sum_{x \in X(\Omega)} |x|\mathbb{P}(X = x) \sum_{y \in Y(\Omega)} |y|\mathbb{P}(Y = y)$$

La dernière expression montre qu'il y a bien cv absolue, car X et Y admettent une espérance. Le calcul donnant la formule s'effectue avec la même succession d'opérations, mais en enlevant les valeurs absolues. □

Théorème 19 (existence de moments)

Si X **admet un moment d'ordre** $k \in \mathbb{N}^*$**, alors elle admet des moments d'ordre** p **pour tout** $p \in [\![0,k]\!]$**.**

Démonstration. Soient $k \in \mathbb{N}^*$ et $p \in [\![0,k]\!]$.
Soit $x \in \mathbb{R}$. Si $|x| \geqslant 1$ alors $|x|^p \leqslant |x|^k \leqslant 1+|x|^k$; et si $|x|<1$ alors $|x|^p \leqslant 1 \leqslant 1+|x|^k$.
On en déduit, au niveau des variables aléatoires :

$$0 \leqslant |X|^p \leqslant 1+|X|^k$$

Les variables 1 et $|X|^k$ admettant une espérance, la variable $|X|^p$ en admet une également par domination.
Attention cependant : ce dernier résultat de domination est à la frontière du programme ; il convient de savoir le redémontrer. Partons donc de :

$$\forall x \in \mathbb{R},\ 0 \leqslant |x|^p \leqslant 1+|x|^k$$

On a alors :

$$\forall x \in X(\Omega),\ 0 \leqslant |x|^p \mathbb{P}(X=x) \leqslant \mathbb{P}(X=x)+|x|^k \mathbb{P}(X=x)$$

Les séries de terme général $\mathbb{P}(X=x)$ et $|x|^k\mathbb{P}(X=x)$ convergent (par existence de $\mathbb{E}(X^k)$) ; donc $\sum |x|^p \mathbb{P}(X=x)$ converge et on a bien l'existence de $\mathbb{E}(X^p)$.

□

Si X, Y **admettent un moment d'ordre 2, alors** $X+Y$ **admet un moment d'ordre 2.**

Démonstration. On doit démontrer la cv absolue de $\sum_{x,y}(x+y)^2\mathbb{P}\big((X=x)\cap(Y=y)\big)$. On découpe en :

- $\sum_{x,y} x^2 \mathbb{P}\big((X=x)\cap(Y=y)\big) = \sum_x x^2 \sum_y \mathbb{P}\big((X=x)\cap(Y=y)\big) = \sum_x x^2 \mathbb{P}(X=x)$ série sommable car X admet un moment d'ordre 2 ;
- $\sum y^2 \mathbb{P}\big((X=x)\cap(Y=y)\big)$: idem
- $\sum |xy| \mathbb{P}\big((X=x)\cap(Y=y)\big) \leqslant \dfrac{1}{2}\sum (x^2+y^2)\mathbb{P}\big((X=x)\cap(Y=y)\big)$, et on se ramène aux points précédents.

□

Si X, Y **admettent un moment d'ordre 2, alors** XY **admet une espérance.**

Démonstration. C'est le troisième point de la démo précédente. □

Théorème 23 (propriété du coefficient de corrélation)

Pour toutes VAD X et Y de variance non nulle, $\big|\rho(X,Y)\big| \leqslant 1$**.**

Démonstration. Pour tout $\lambda \in \mathbb{R}$, on a $V(X+\lambda Y) \geqslant 0$. En utilisant les propriétés de la covariance, on a par ailleurs :

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

Le trinôme $\lambda \mapsto \lambda^2 V(X) + 2\lambda \mathrm{Cov}(X,Y) + V(Y)$ est donc de signe constant sur $\mathbb{R}$: on en déduit que son discriminant est négatif, ce qui donne $\mathrm{Cov}(X,Y)^2 \leqslant V(X)V(Y)$; soit en prenant la racine : $\big|\mathrm{Cov}(X,Y)\big| \leqslant \sigma(X)\sigma(Y)$; ou encore $|\rho(X,Y)| \leqslant 1$. □

Théorème 24 (Caractérisation de $\rho(X,Y) = \pm 1$)

Soient X et Y deux VAD de variances non nulles.
$\rho(X,Y) = \pm 1$ **si et seulement si il existe** $a \in \mathbb{R}^*,\ b \in \mathbb{R}$ **tels que** $Y = aX+b$ **presque sûrement.**
Dans ce cas, le signe de $\rho(X,Y)$ **est celui de** a**.**

Démonstration. Par ailleurs, si $|\rho(X,Y)| = 1$, alors le discriminant étudié dans la démonstration précédente est nul ; et le trinôme a donc une racine réelle, qu'on note a. Comme $V(Y) \neq 0$, on a $a \neq 0$.
En remontant les calculs on trouve $V(aX+Y) = 0$; ce qui signifie que la VAD $aX+Y$ est presque sûrement constante. On a donc

$$\rho(X,Y) = \pm 1 \ \Rightarrow\ Y = aX+b \text{ presque sûrement, avec } (a,b) \in \mathbb{R}^* \times \mathbb{R}$$

Réciproquement, supposons que $Y = aX+b$ presque sûrement. On a déjà $V(Y) = a^2 V(X)$.
De plus la VAD $Y-(aX+b)$ est presque sûrement constante, et $\mathrm{Cov}(X, Y-aX+b) = 0$. Par bilinéarité de la covariance, on trouve

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(X, aX+b) = a\mathrm{Cov}(X,X) = aV(X)$$

On a donc

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)} = \frac{aV(X)}{\sigma(X) \times \sqrt{a^2}\,\sigma(X)} = \frac{a}{|a|} = \mathrm{signe}(a)$$

ce qui donne le sens réciproque, et la dernière partie du théorème. □

Théorème 26 (Inégalité de Markov)

Soit X une variable aléatoire positive, admettant une espérance. On a :

$$\forall\, t > 0,\ \mathbb{P}(X \geqslant t) \leqslant \frac{\mathbb{E}(X)}{t}$$

Démonstration. Soit $t > 0$. Comme $\mathbb{E}(X)$ existe, on a

$$\begin{aligned}
\mathbb{E}(X) = \sum_{x \in X(\Omega)} x\,\mathbb{P}(X = x) &= \sum_{\substack{x \in X(\Omega) \\ x < t}} x\,\mathbb{P}(X = x) + \sum_{\substack{x \in X(\Omega) \\ x \geqslant t}} x\,\mathbb{P}(X = x) \\
&\geqslant \sum_{\substack{x \in X(\Omega) \\ x \geqslant t}} x\,\mathbb{P}(X = x) && \text{(par positivité de X)} \\
&\geqslant t \sum_{\substack{x \in X(\Omega) \\ x \geqslant t}} \mathbb{P}(X = x) && (x \geqslant t \text{ sur le domaine de sommation}) \\
&= t\,\mathbb{P}(X \geqslant t)
\end{aligned}$$

Finalement, on a $\mathbb{E}(X) \geqslant t\,\mathbb{P}(X \geqslant t)$, ce qui permet de conclure en divisant par $t > 0$.

□

Théorème 27 (inégalité de Bienaymé-Tchebychev)

Soit X **une variable admettant une variance. Alors elle admet aussi une espérance, et on a :**

$$\mathbb{P}\big(|X - \mathbb{E}(X)| \geqslant t\big) \leqslant \frac{V(X)}{t^2}$$

Démonstration. L'existence d'une espérance a été vue précédemment.
Définissons $Y = (X - \mathbb{E}(X))^2$. Comme X admet une variance, Y admet une espérance et on a par définition $\mathbb{E}(Y) = V(X)$.
On a ensuite : $\mathbb{P}\big(|X - \mathbb{E}(X)| \geqslant t\big) = \mathbb{P}\big((X - \mathbb{E}(X))^2 \geqslant t^2\big) = \mathbb{P}\big(Y \geqslant t^2\big)$.
Par Markov (Y est bien à valeurs positives), on a : $\mathbb{P}\big(Y \geqslant t^2\big) \leqslant \dfrac{\mathbb{E}(Y)}{t^2}$.
On a finalement montré : $\mathbb{P}\big(|X - \mathbb{E}(X)| \geqslant t\big) \leqslant \dfrac{V(X)}{t^2}$. □