Chapitre 2

Probabilité et espérance

Dans ce qui suit, Ω désigne l'ensemble des "états du monde" ω et $\mathcal{B}$ l'algèbre des évènements considérés sur Ω. Si Ω est fini, on pourra supposer que $\mathcal{B} = \mathcal{P}(\Omega)$; si Ω est infini, il faudra, pour certaines propriétés énoncées, supposer que $\mathcal{B}$ soit une σ-algèbre, par exemple la σ-algèbre des boréliens. Nous ne nous intéressons pas ici à expliquer en détails pourquoi ces précautions sont nécessaires et renvoyons au cours de probabilités de L3 sur cet aspect.

2.1 Probabilité et espérance des v.a. élémentaires

Définition : On appelle *probabilité sur* $(\Omega, \mathcal{B})$ toute fonction $\mathbb{P}\ : \mathcal{B} \longrightarrow [0,1]$ telle que $\mathbb{P}(\Omega) = 1$ et $\mathbb{P}(A \dot\cup B) = \mathbb{P}(A) + \mathbb{P}(B)$, où $A \dot\cup B$ désigne[1] simplement l'union $A \cup B$ tout en exprimant qu'*on suppose* que $A \cap B = \emptyset$.

Une manière d'interpréter $\mathbb{P}$ est que $\mathbb{P}(A)$ "mesure l'espérance[2] qu'on peut avoir que l'évènement A se réalisera". Cette mesure n'a nullement besoin d'exprimer les "chances" qu'a A de se réaliser (même si, comme nous le verrons avec la *loi des grands-nombres*, cela peut servir à cela dans un sens à préciser). Nous verrons ci-dessous des exemple d'autres mesures utiles. En revanche, nous attendons que cette espérance soit linéaire dans le sens suivant : nous avons vu qu'à tout évènement A nous pouvons faire correspondre la v.a. $\mathbb{I}_A$, où $\mathbb{I}_A(\omega)$ vaut 1 ou 0 selon que "A est vrai ou non" c'est-à-dire selon que "l'état du monde" ω appartient à A ou non. Si l'on désigne par $\mathbb{E}$ l'espérance (mathématique), la linéarité évoquée ici signifie simplement que pour tous évènements A et B de $\mathcal{B}$ et tous réels a et b, on a $\mathbb{E}(a\mathbb{I}_A + b\mathbb{I}_B) = a\mathbb{E}(\mathbb{I}_A) + b\mathbb{E}(\mathbb{I}_B) = a\mathbb{P}(A) + b\mathbb{P}(B)$, et en particulier $\mathbb{E}(\mathbb{I}_A) = \mathbb{P}(A)$. Nous allons devoir expliquer, étant donné une probabilité $\mathbb{P}$ sur $\mathcal{B}$, comment l'étendre à une espérance $\mathbb{E}(X)$ définie pour toute v.a. $\mathcal{B}$-mesurable. Nous ne donnerons la construction que dans le cas élémentaire et indiquerons, sans preuve, à quelle formules on abouti lorsqu'on étend la notion d'espérance au moyen des constructions classiques de la théorie de la mesure.

Voici ce cas élémentaire : considérons une v.a. X sur $(\Omega, \mathcal{B})$ et notons $\mathcal{X} := X(\Omega)$. Supposons que cet ensemble des valeurs de X soit fini : $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$. Nous dirons qu'une telle v.a. est *élémentaire*. Pour $i = 1..n$, soit $A_i := \{X = x_i\}$, l'ensemble des $\omega \in \Omega$ tels que $X(\omega) = x_i$. Il est facile de voir que $\Omega = \dot\bigcup_{i=1..n} A_i$ et que les A_i sont bien dans $\mathcal{B}$ puisque la v.a. X est par hypothèse $\mathcal{B}$ mesurables (expliquez cela !) ; donc $\mathbb{P}(\{X = x_i\}) = \mathbb{P}(A_i)$ est bien défini pour tout $x_i \in \mathcal{X}$. Cette petite construction nous permet d'écrire $X = x_1\mathbb{I}_{A_1} + x_2\mathbb{I}_{A_2} + \ldots + x_n\mathbb{I}_{A_n}$ (vérifiez cela, en envisageant la valeurs de chacun des deux membres de l'égalité lorsque $\omega \in A_i$). Par linéarité de l'espérance postulée ci-dessus, nous devons donc poser $\mathbb{E}(X) = \mathbb{E}(x_1\mathbb{I}_{A_1} + x_2\mathbb{I}_{A_2} + \ldots + x_n\mathbb{I}_{A_n}) = x_1\mathbb{E}(\mathbb{I}_{A_1}) + x_2\mathbb{E}(\mathbb{I}_{A_2}) + \ldots + x_n\mathbb{E}(\mathbb{I}_{A_n})$, et donc

$$\mathbb{E}(X) = \sum_{i=1}^{n} x_i p_i \text{ où } p_i := \mathbb{P}(A_i) \ \ (= \mathbb{E}(\mathbb{I}_{A_i})). \tag{2.1}$$

[1] De façon analogue, on notera $\dot\bigcup_{i\in I} A_i$ la réunion $\bigcup_{i\in I} A_i$ tout en exprimant qu'*on suppose* que $A_i \neq A_j$ si $i \neq j$

[2] le sens de la locution "mesurer l'espérance" n'est pas précisé ici et nous laissons au lecteur la liberté de choisir une intuition qui lui convienne. Avec l'exemple de paris sur une course de chevaux, nous lui proposons un peu plus loin un exemple où cette mesure de l'espérance est peut-être différente de ce qui lui vient à l'esprit. Nous figerons bientôt le sens du mot espérance dans un sens mathématique, d'ailleurs déduit de celui de probabilité, le mot mesure admettant lui aussi un sens mathématique, qui sera, lui, précisé en L3

Exemple : Voici un exemple de construction d'une probabilité associée à une situation à l'issue inconnue, et où la probabilité considérée ne prétend pas refléter les chances des diverses issues possibles. Considérons un cafetier qui souhaite organiser un pari sur une course de chevaux, pour le seul agrément de ses clients, et sans avoir l'intention de s'enrichir mais avec l'exigence de ne pas s'appauvrir non plus. La course comporte n chevaux numérotés $i = 1..n$. Voici une formulation probabiliste de cette situation :

Soit Ω l'ensemble des états du monde, et notons A_i l'évènement "le cheval i est le gagnant" (nous n'envisagerons pas ici le cas de gagnants ex aequo). Notons X_i le gain par euro misé sur le cheval i ; on a donc $X_i = x_i \mathbb{I}_{A_i}$, où x_i est un nombre à préciser. Notons s_i la somme des mises sur le cheval i, et $s := \sum_{i=1}^n s_i$ la mise totale. La somme totale à payer aux parieurs par le cafetier s'écrit donc

$$S := \sum_{i=1}^{n} s_i x_i \mathbb{I}_{A_i}.$$

Comme de cafetier ne veut pas s'appauvrir, il faut donc que $S \leq s$, c'est-à-dire que $S(\omega) \leq s$ pour tout état du monde $\omega \in \Omega$. En choisissant $\omega \in A_{i_0}$ on voit immédiatement que le contrainte du cafetier implique que $x_{i_0} \leq \frac{s}{s_{i_0}}$ pour n'importe quel i_0 (faites le calcul en supposant que $\omega \in A_{i_0}$).

Notons que jusqu'ici nous n'avons pas fait appel à la notion de probabilité. Postulons maintenant qu'une même probabilité $\mathbb{P}^*$ permette de prendre en compte les mises de tous les joueurs ; ceci signifie que chaque euro misé sur le cheval i se fait avec l'espoir de gagner au moins un euro, où le sens donné au mot "espoir" s'accorde tout-à-fait avec le fait que dans certains états du monde on perde sa mise, bien entendu. Nous pouvons à présent revenir à un discours exclusivement mathématique.

La contrainte des parieurs sur l'évènement A_i s'écrit donc

$1 \leq \mathbb{E}^*(X_i) = \mathbb{E}^*(x_i \mathbb{I}_{A_i}) = x_i \mathbb{E}^*(\mathbb{I}_{A_i}) = x_i \mathbb{P}^*(A_i)$, où $\mathbb{E}^*$ désigne l'espérance au sens de la probabilité $\mathbb{P}^*$.

Donc $\mathbb{P}^*(A_i) \geq \frac{1}{x_i} \geq \frac{s_i}{s}$ pour tout $i = 1..n$. A présent, comme $\Omega = \dot{\bigcup}_{i=1..n} A_i$, nous obtenons

$$1 = \mathbb{P}^*(\Omega) = \mathbb{P}^*\left(\dot{\bigcup}_{i=1..n} A_i\right) = \sum_{i=1..n} \mathbb{P}^*(A_i) \geq \sum_{i=1..n} \frac{s_i}{s} = 1,$$

qui n'est possible que si $\sum_{i=1..n} \mathbb{P}^*(A_i) = \sum_{i=1..n} \frac{s_i}{s}$, qui n'est, à son tour, possible que si

$$\mathbb{P}^*(A_i) = \frac{s_i}{s}.$$

Nous voyons donc qu'ici la mesure de l'espérance que le cheval i gagne qui se dégage de ce pari est égale à la proportion des mises qui se sont portées sur ce cheval : c'est une "mesure de marché", le marché des clients prêts à parier un euro dans le seul "espoir" d'en gagner autant mais avec la possiblité, selon l'état de monde, d'en gagner plus.

A noter qu'il est à présent possible de vérifier que toutes les inégalités que nous avons dégagées pour la modélisation doivent donc être des égalité : en particulier $S = s$, qui exprime que le cafetier ne gagne rien à cela (sauf si certains parieurs heureux décident de payer une tournée générale, mais ça, c'est une autre affaire !). Nous voyons aussi que pour $\mathbb{P}^*$, l'espérance de gain par euro misé sur le cheval i_0 est égale à $\mathbb{E}^*(X_{i_0}) = \mathbb{E}^*(x_{i_0} \mathbb{I}_{A_{i_0}}) = x_{i_0} \mathbb{P}^*(\mathbb{I}_{A_{i_0}}) = \frac{s}{s_{i_0}} \times \frac{s_{i_0}}{s} = 1$. En d'autres termes, l'espérance est égale à la mise : c'est l'idée que l'on se fait d'une loterie équitable.

Notons que généralement les cafetiers n'ont pas l'esprit frondeur et ne se permettent pas d'organiser des jeux d'argent, qui ne sont, rappelelons-le, autorisés que dans un cadre très strict : ils préfèrent passer un accord avec le PMU ou la Française des Jeux (FdJ), qui partagent avec l'état (40%) et les cafetiers, une part importante des mises, et ne redistribue que le solde aux parieurs. Ce faisant, on remplace la contraite $S \leq s$ par $S < 40\% s$. On en déduit alors que les client de la FdJ acceptent allègrement que "l'espoir de gain" pour la probabilité de marché $\mathbb{P}^*$ soit strictement inférieur à la mise ; cela peut avoir un sens si l'on pense que pour un cheval i_0 on a $\mathbb{P}(A_{i_0}) > \mathbb{P}^*(A_{i_0})$ pour une autre probabilité $\mathbb{P}$ qui leur semble plus pertinente que la probabilité de marché $\mathbb{P}^*$.

2.2 Extension du domaine de l'espérance

Une fois définie l'espérance $\mathbb{E}$ par (2.1) pour les v.a. élémentaires, on l'étend, "par approximation", aux v.a. quelconques en posant $\mathbb{E}(X) = \lim_k \mathbb{E}(X_k)$ où $(X_k)_{k=1,2,...}$ est une suite de v.a. élémentaires "tendant" vers X. Esquissons comment cela sera fait en L3. Par exemple, si X ne prend qu'un ensemble dénombrable de valeurs $(x_i)_{i\in\mathbb{N}}$, si $\sum_{i=0}^{+\infty}|x_i|p_i < +\infty$, $p_i := \mathbb{P}(\{X = x_i\})$, il suffit de poser $X_k(\omega) = X(\omega)$ si $X(\omega) = x_i$ pour $i \le k$, et $X_k(\omega) = 0$ sinon ; nous voyons que le choix des X_k dépend de l'ordre dans lequel on a numéroté les valeurs de X, mais on peut montrer que l'hypothèse que $\sum_{i=0}^{+\infty}|x_i|p_i < +\infty$ assure que ceci est sans conséquence sur la valeur de

$$\mathbb{E}(X) := \lim_{k\to+\infty}\mathbb{E}(X_k) = \lim_{k\to+\infty}\sum_{i=0}^{k} x_i\mathbb{P}(\{X = x_i\}) = \sum_{i=0}^{+\infty} x_i\mathbb{P}(\{X = x_i\}) = \sum_{i=0}^{+\infty} x_i p_i \text{ où } p_i := \mathbb{P}(\{X = x_i\}).$$

Si l'ensemble des valeurs de X n'est pas dénombrable, il convient de "bricoler" un peu plus : on commence par scinder X en différence de deux v.a. positives $X = X_+ - X_-$, avec $X_+(\omega) = \mathrm{Max}\{X(\omega), 0\}$ et $X_-(\omega) = \mathrm{Max}\{-X(\omega), 0\}$, puis, pour chaque v.a. positive Y, on l'approche (de façon monotone) par des v.a. Y_k, avec $Y_k(\omega) = y_i^{(k)}$ si $0 = y_0^{(k)} \le y_i^{(k)} \le Y(\omega) < y_{i+1}^{(k)} \le k$, et $Y_k(\omega) = 0$ si $Y(\omega) > k$, les $y_i^{(k)}$ constituant une discrétisation $[0..k]_{\frac{1}{2^k}}$ de l'intervalle $[0, k]$ par des point espacés de $\frac{1}{2^k}$. Chaque v.a. aléatoire $X_+^{(k)}$ et $X_-^{(k)}$ ainsi confectionnée est alors élémentaire. La théorie de la mesure montre alors que si $\lim_k \mathbb{E}(X_+^{(k)}) < +\infty$ et $\lim_k \mathbb{E}(X_-^{(k)}) < +\infty$, alors $\lim_{k\to+\infty} X_+^{(k)} - X_-^{(k)} = X$ et on peut poser $\mathbb{E}(X) := \lim_k \mathbb{E}(X_+^{(k)}) - \lim_k \mathbb{E}(X_-^{(k)})$, toute autre manière d'approcher X_- et X_+ de façon monotone par des v.a. élémentaires conduisant nécessairement au même résultat (la définition précise et la preuve élégante de toutes ces affirmations est précisément l'objet de la théorie de Lebesgue).

On définit ainsi $\mathbb{E}(X)$ pour une large classe de v.a., dites "intégrables" (voir ci-dessous un exemple de v.a. non intégrable). Nous allons indiquer ici, sans démonstration, comment se calcule $\mathbb{E}(X)$ dans un certain nombre de cas de v.a. non élémentaires, en particulier lorsque la *loi* de X admet une *densité*.

Définition : On appelle *fonction de répartition* de la v.a. X sur $\mathbb{R}^d$ $(d \ge 1)$ la fonction $F_X : \mathbb{R}^d \longrightarrow [0, 1]$ définie par

$$F_X(x) := \mathbb{P}(\{X \le x\}) \text{ (où } X(\omega) \le x \text{ signifie } X_i(\omega) \le x_i \text{ pour tout i=1..d)};$$

on l'appelle aussi la *loi* de X. On dit que la loi de X admet une *densité* f_X si la fonction $f_X : \mathbb{R}^d \longrightarrow \mathbb{R}^+$ est telle que pour tout $x_0 \in \mathbb{R}^d$ on a $F_X(x_0) = \int_{x\le x_0} f_X(x)dx$.

Remarques : La fonction de répartition est croissante puisque si $x' < x''$ on a $\{X \le x'\} \subseteq \{X \le x''\}$ et donc

$$F_X(x') = \mathbb{P}(\{X \le x'\}) \le \mathbb{P}(\{X \le x'\}) + \mathbb{P}(\{x' < X \le x"\}) = \mathbb{P}(\{X \le x''\}) = F_X(x'').$$

Par monotonie, comme F_X est bornée (par 0 et 1), les limites $\lim_{x\to-\infty} F_X(x)$ et $\lim_{x\to+\infty} F_X(x)$ existent, et on montre que nécessairement

$$\lim_{x\to-\infty} F_X(x) = 0 \text{ et } \lim_{x\to+\infty} F_X(x) = 1$$

Si F_X est dérivable de dérivée F'_X continue, la dérivée $f_X := F'_X$ est nécessairement positive puisque F_X est croissante, et X admet $f_X := F'_X$ pour densité. Observez que l'hypothèse de mesurabilité des v.a. assure précisément que $\{X \le x\} \in \mathcal{B}$ pour tout $x \in \mathbb{R}$; les nombres $\mathbb{P}(\{X \le x\})$ sont donc tous bien définis.

Exercice : Montrer que $\lim_{x\to-\infty} F_X(x) = 0$ et $\lim_{x\to+\infty} F_X(x) = 1$ dans le cas où la v.a. X ne prend qu'un nombre fini de valeurs $x_1 < x_2 < \ldots < x_n$.

Exemples : On dit que la v.a. X suit une *loi uniforme* sur $[a, b]$, et on note $X \rightsquigarrow \mathcal{U}([a, b])$ si et seulement si $a < b$, et si $x \in [a, b]$ alors $\mathbb{P}(\{X \le x\}) = \frac{x-a}{b-a}$; si $x \le a$ alors $\mathbb{P}(\{X \le x\}) = 0$; si $x \ge b$ alors $\mathbb{P}(\{X \le x\}) = 1$.

On dit que la v.a. suit une *loi normale centrée réduite* et on note $X \rightsquigarrow \mathcal{N}(0, 1)$ si elle admet pour densité la fonction $f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, dite *fonction gaussienne*[3] ; le théorème central-limit nous expliquera pourquoi cette loi est capitale, notamment en statistiques.

Les observations faites en début de section conduisent au théorème suivant dont la preuve sera donnée en cours de Probabilités de L3 :

[3] en l'honneur de Karl Friedrich Gauss

Théorème 2.1 *Soit X une v.a. à valeurs dans $\mathbb{R}^d$. Si X ne prends qu'un nombre fini de valeurs $x_1, \dots, x_n$ alors*

$$\mathbb{E}(X) = \sum_{i=1}^{n} x_i p_i \ , \text{ où } p_i := \mathbb{P}(\{X = x_i\}). \tag{2.2}$$

Si X ne prend qu'une suite dénombrable de valeurs $(x_i)_{i\in\mathbb{N}}$ et si $\sum_{i=0}^{+\infty} |x_i| p_i < +\infty$, alors

$$\mathbb{E}(X) = \sum_{i=1}^{+\infty} x_i p_i \ , \text{ où } p_i := \mathbb{P}(\{X = x_i\}). \tag{2.3}$$

Soit[4] *$X : \Omega \longrightarrow \mathbb{R}$ (donc $d = 1$). Si la loi de X admet une densité f_X et si $\int |x| f_X(x)dx < +\infty$, alors*

$$\mathbb{E}(X) = \int_{x\in\mathbb{R}} x f_X(x)dx = \int_{-\infty}^{+\infty} x f_X(x)dx. \tag{2.4}$$

Si la loi de X admet une densité f_X, si la v.a. Y est définie par $Y = g(X)$, et si $\int |g(x)| f_X(x)dx < +\infty$, alors

$$\mathbb{E}(Y) = \int g(x) f_X(x)dx. \tag{2.5}$$

Notons qu'il convient encore de s'assurer que les fonctions f_X et g mentionnées dans ce paragraphe sont "Borel-mesurables" c'est-à-dire suffisamment régulières pour que les intégrales mentionnées aient bien un sens, une limitation guère contraignante dans un bonne théorie de l'intégration.

Exercice : Montrer que si $X \rightsquigarrow \mathcal{U}([a, b])$, alors $\mathbb{E}(X) = \frac{a+b}{2}$; déterminer la loi de $Y = \alpha X + \beta$. Montrer que si $X \rightsquigarrow \mathcal{N}(0, 1)$, alors $\mathbb{E}(X) = 0$; déterminer la loi de $Y = \alpha X + \beta$.

Exercice : On dit que X suit une *loi de Cauchy* si et seulement si la loi de X admet pour densité la fonction $f_X(x) = \frac{C}{1+x^2}$. Determiner la valeur de la constante C. Montrer que $\int |x| f_X(x)dx \left(:= \int_{-\infty}^{+\infty} |x| f_X(x)dx\right) = +\infty$: on dit que la v.a. X *n'est pas intégrable*. En particulier l'espérance $\mathbb{E}(x)$ n'est pas définie. **Rep :** $C = \pi$.

2.3 Quelques propriétés des probabilités et de leur espérance

Proposition 2.2 *Soit $\mathbb{P}$ une probabilité sur $(\Omega, \mathcal{B})$. Soient A, B, $(A_i)_{i=1..n}$ des évènement. On a les relations suivantes :*

1. $\mathbb{P}(\emptyset) = 0$
2. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
4. *Si les A_i sont deux-à-deux disjoints, alors $\mathbb{P}\left(\dot{\bigcup}_{i=1..n} A_i\right) = \sum_{i=1..n} \mathbb{P}(A_i)$, et en particulier, si $A = \{\omega_1, \dots, \omega_n\}$ est fini et $\{\omega_i\} \in \mathcal{B}$, alors $\mathbb{P}(A) = \sum_{i=1..n} \mathbb{P}(\{\omega_i\})$.*
5. *Si $\Omega = \dot{\bigcup}_{i=1..n} A_i$, alors $\mathbb{P}(B) = \sum_{i=1..n} \mathbb{P}(B \cap A_i)$.*
6. $\mathbb{P}(A) = \mathbb{E}(\mathbb{I}_A)$.

Théorème 2.3 *Pour toutes v.a. intégrables X et Y, et tous réels a et b on a*

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

si $X \leq Y$ (au sens où $X(\omega) \leq Y(\omega)$ pour tout $\omega \in \Omega$), alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$

[4]Si d est quelconque, on a $X = (X_1, \dots, X_d)$ et $\mathbb{E}(X) := \big(\mathbb{E}(X_1), \dots, \mathbb{E}(X_d)\big)$, avec $\mathbb{E}(X_i) = \int_{x\in\mathbb{R}^d} x_i f_X(x)dx$.

Preuve : Comme d'habitude, nous ne prouvons ce théorème que dans le cas de v.a. élémentaires. Soient X et Y deux v.a. élémentaires, $\mathcal{X} := X(\Omega)$ et $\mathcal{Y} := Y(\Omega)$ les ensembles (fini) de leurs valeurs. Montrons tout d'abord que $\mathbb{E}(aX) = a\mathbb{E}(X)$; notons $a\mathcal{X} := \{ax, x \in \mathcal{X}\} = ax(\Omega)$. Comme $\{aX = ax\} = \{X = x\}$, on a

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

Il reste à montrer que $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$; à cette fin, posons $Z := X + Y$ et $\mathcal{Z} := Z(\Omega)$. notons $\mathcal{X} + \mathcal{Y} := \{x + y | x \in \mathcal{X}, y \in \mathcal{Y}\}$. On a bien-entendu $\mathcal{Z} \subseteq \mathcal{X} + \mathcal{Y}$. On a de plus $\{X = x\} = \dot{\bigcup}_{y \in \mathcal{Y}} \{X = x\} \cap \{Y = y\} =: \dot{\bigcup}_{y \in \mathcal{Y}} \{X = x, Y = y\}$, et donc

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{x \in \mathcal{X}} x\mathbb{P}(\{X = x\}) = \sum_{x \in \mathcal{X}} x\mathbb{P}\left(\dot{\bigcup}_{y \in \mathcal{Y}} \{X = x, Y = y\}\right) \\
&= \sum_{x \in \mathcal{X}} x \sum_{y \in \mathcal{Y}} \mathbb{P}(\{X = x, Y = y\}) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} x\mathbb{P}(\{X = x, Y = y\}) \\
&= \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} x\mathbb{P}(\{X = x, Y = y\}) \qquad \text{; de même a-t-on} \\
\mathbb{E}(Y) &= \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} y\mathbb{P}(\{X = x, Y = y\}) \qquad \text{et donc} \\
\mathbb{E}(X) + \mathbb{E}(Y) &= \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} x\mathbb{P}(\{X = x, Y = y\}) + \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} y\mathbb{P}(\{X = x, Y = y\}) \\
&= \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} (x + y)\mathbb{P}(\{X = x, Y = y\}) \text{ , puis en regroupant par même valeur } z \text{ de } x + y \\
&= \sum_{z \in \mathcal{X} + \mathcal{Y}} \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y},\ x+y=z} z\mathbb{P}(\{X = x, Y = y\}) \\
&= \sum_{z \in \mathcal{Z}} \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y},\ x+y=z} z\mathbb{P}(\{X = x, Y = y\}) \\
&\qquad \text{car, pour } (x, y) \in \mathcal{X} \times \mathcal{Y} \text{ et } (z =)x + y \in \mathcal{Z}^c \text{, on a } \{X = x, Y = y\} = \emptyset \\
&= \sum_{z \in \mathcal{Z}} z\mathbb{P}\left(\dot{\bigcup}_{(x,y) \in \mathcal{X} \times \mathcal{Y},\ x+y=z} \{X = x, Y = y\}\right) \\
&= \sum_{z \in \mathcal{Z}} z\mathbb{P}(\{Z = z\}) = \mathbb{E}(Z) = \mathbb{E}(X + Y).
\end{aligned}$$

Finalement, supposons que $X \le Y$, et montrons que $\mathbb{E}(X) \le \mathbb{E}(Y)$, ou encore, que $0 \le \mathbb{E}(Y) - \mathbb{E}(X) = \mathbb{E}(Y - X) =: \mathbb{E}(U)$, où $U := Y - X$. Soit $\mathcal{U} := U(\Omega)$; pour tout $u \in \mathcal{U}$ on a donc $u = U(\omega) = Y(\omega) - X(\omega) \ge 0$; comme $\mathbb{P}(\{U = u\}) \ge 0$, on a bien $\mathbb{E}(U) = \sum_{u \in \mathcal{U}} u\mathbb{P}(\{U = u\}) \ge 0$.

□

Kolmogorov[5] :   : Gauss[6]

[5] Andrey Nikolaevich Kolmogorov (1903-1987)
[6] Karl Friedrich Gauss (1777-1855)

.