Probabilités

Chapitre 2 : Variables aléatoires

Lucie Le Briquer

On va tout faire pour ne plus regarder $(\Omega, \mathcal{A}, \mathbb{P})$ en détail. L'objet essentiel : les variables aléatoires (v.a.) et leur loi.

Définition 1 (variable aléatoire)

Une variable aléatoire X est une fonction mesurable

$$X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$$

$X(\omega)$ est un élément aléatoire de E.
Si $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on dit que X est une v.a. rééelle (v.a.r). X est une observation de l'aléa.

Exemple. On jette 2 dés. $X =$"la somme des 2 dés" est une v.a.

$$X : \left\{ \begin{array}{ccc} \{1, ..., 6\}^2 & \to & \mathbb{R} \\ (\omega_1, \omega_2) & \to & \omega_1 + \omega_2 \end{array} \right.$$

Exemple. Dans une population de 1000 personnes, N personnes veulent voter "Oui" à un referendum. On interroge une personne au hasard et on lui demande son opinion.
Modélisation. $\Omega = \{1, ..., 1000\}$ (de 1 à N ils votent "Oui"), $\mathcal{A} = \mathcal{P}(\Omega)$, $\mathbb{P}$ probabilité uniforme :

$$\mathbb{P}(\text{répond Oui}) = \frac{N}{1000}$$

Une autre modélisation. $\Omega = \{\text{“Oui”}, \text{“Non”}\}$, $\mathcal{A} = \mathcal{P}(\Omega)$, $\mathbb{P}$ tel que $\mathbb{P}(\{\text{“Oui”}\}) = \frac{N}{1000}$ et $\mathbb{P}(\{\text{“Non”}\}) = 1 - \frac{N}{1000}$. $\mathbb{P}(\text{répond Oui}) = \frac{N}{1000}$.

Remarque. Choix non unique de $(\Omega, \mathcal{A}, \mathbb{P})$.
Si on interroge une seconde personne, il faut changer Ω, il faudrait désormais prendre $\{1, ..., 1000\}^2$.

Modélisation par v.a. Soit $X : \Omega \to \{\text{“Oui”}, \text{“Non”}\}$ telle que $\mathbb{P}(X = \text{“Oui”}) = \frac{N}{1000}$. Et on raisonne de manière à ce que ça marche sur tout $(\Omega, \mathcal{A}, \mathbb{P})$ tel que X existe. Si on s'intéresse à une seconde personne on définit juste une seconde v.a. Y telle que :

$$\mathbb{P}(X = \text{“Oui”}, Y = \text{“Oui”}) = \left(\frac{N}{1000}\right)^2$$

1 Évènements décrits par des variables aléatoires

Si $X, Y, (X_i)_i$ sont des variables aléatoires sur Ω, on note :

- $(X \leqslant 4)$ pour $\{\omega \in \Omega \mid X(\omega) \leqslant 4\}$
- $\left(\limsup\limits_{n\to+\infty} \frac{1}{n}\sum_{i=1}^n X_i = +\infty\right)$ pour $\left\{\omega \in \Omega \mid \limsup\limits_{n\to+\infty} \frac{1}{n}\sum_{i=1}^n X_i(\omega) = +\infty\right\}$
- $(\sum_{i=1}^n X_i < +\infty)$ pour $\{\omega \in \Omega \mid X_i(\omega) < +\infty\}$

Exemple. On jette une infinité de dès et on note X le numéro du jet où apparaît pour la première fois 6. X est une variable aléatoire à valeurs dans $\mathbb{N}^* \cup \{+\infty\}$:

$$X : \left\{ \begin{array}{ccc} \Omega & \longrightarrow & \mathbb{N}^* \cup \{+\infty\} \\ \omega & \longmapsto & \inf\{i \mid \omega_i = 6\} \end{array} \right.$$

On prend $\Omega = [\![1, 6]\!]^{\mathbb{N}^*}$, $\mathcal{A}$ est la tribu cylindrique.
Regardons l'événement $X < +\infty$:

$$\mathbb{P}(X = +\infty) \leqslant \mathbb{P}(\forall i \leqslant n,\ \omega_i \neq 6) = \sum_{i_1,\ldots,i_k \in [\![1,5]\!]} \mathbb{P}(\forall j \leqslant n,\ \omega_j = i_j) \leqslant \left(\frac{5}{6}\right)^n \xrightarrow[n\to+\infty]{} 0$$

Donc $\mathbb{P}(X = +\infty) = 0$ et $\mathbb{P}(X < +\infty) = 1$.

Remarques.

- On note A p.s. (presque sûrement) si $\mathbb{P}(A) = 1$, ou A $\mathbb{P}-$p.p. (presque partout).
- Il faut différencier $X < +\infty$ (événement) et $X < +\infty$ p.s. (décrit un fait mathématique : $\mathbb{P}(X < +\infty) = 1$) !
- Il faut différencier $X < +\infty$ p.s. de X ne prend que des valeurs finies (ce qui est faux) !

Propriété 1 (événements p.s.)

- Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements tels que $\forall n \in \mathbb{N}$ on a A_n p.s. alors on a :

$$\left(\bigcap_{n\in\mathbb{N}} A_n\right) \quad \text{p.s.}$$

- Si on a A p.s., alors :

$$\forall B,\ \mathbb{P}(A \cap B) = \mathbb{P}(B)$$

Preuve.

- $\mathbb{P}\left(\bigcap_{n\in\mathbb{N}} A_n\right) = 1 - \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} \overline{A_n}\right) \geqslant 1 - \sum_{n\in\mathbb{N}} \underbrace{\mathbb{P}(\overline{A_n})}_{=1-\mathbb{P}(A_n)=0} = 1$
- $\mathbb{P}(A \cap B) + \underbrace{\mathbb{P}(\overline{A} \cap B)}_{\leqslant \mathbb{P}(\overline{A})=0} = 1$

$\square$

Remarque. Premier point fort pour une intersection non dénombrable :
Soit $(\Omega, \mathcal{A}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \mathrm{Leb})$, $A_x = \{x\}$. On a $\mathbb{P}(A_x) = \mathrm{Leb}(\{x\}) = 0$ donc $[0,1]\backslash\{x\}$ p.s. Mais :

$$\mathbb{P}\left(\bigcap_{x\in[0,1]} ([0,1]\backslash\{x\})\right) = \mathbb{P}(\emptyset) = 0$$

2 Tribu à partir d'une variable aléatoire

La construction de tribus définies par des variables aléatoires est cruciale !

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit $X : (\Omega, \mathcal{A}) \longrightarrow (E, \mathcal{E})$ une variable aléatoire mesurable. *Quelle est la plus petite tribu ($\subseteq \mathcal{A}$) qui rend X mesurable ?*

Appelons $\mathcal{B}$ cette tribu. $\mathcal{B}$ doit contenir $X^{-1}(C)$ pour $C \in \mathcal{E}$ donc $\{X^{-1}(C) \mid C \in \mathcal{E}\} = X^{-1}(\mathcal{E}) \subseteq \mathcal{B}$. Or, on vérifie facilement que $\{X^{-1}(C) \mid C \in \mathcal{E}\}$ est une tribu, et donc $\mathcal{B} = X^{-1}(\mathcal{E})$. On la note $\sigma(X) = \mathcal{B}$ et on l'appelle *tribu engendrée par X*.

Définition 2 (tribu engendrée par une variable aléatoire)

Soit $X : (\Omega, \mathcal{A}) \longrightarrow (E, \mathcal{E})$ une variable aléatoire. On note $\sigma(X)$ la plus petite tribu de $\mathcal{A}$ qui rend X mesurable. On a :

$$\sigma(X) = X^{-1}(\mathcal{E})$$

Remarque. C'est la bonne manière de penser les tribus : une tribu est l'ensemble des propriétés qu'on peut évaluer sur l'aléa au travers de l'observation de X.

Propriété 2

Si $\mathcal{B} \subseteq \mathcal{A}$, on a :

$$X : (\Omega, \mathcal{B}) \to (E, \mathcal{E}) \text{ est mesurable} \quad \Leftrightarrow \quad \sigma(X) \subseteq \mathcal{B}$$

Remarque.

- Toute tribu apparaît comme cela. Si $\mathcal{B}$ est une sous-tribu de $\mathcal{A}$, alors si $X : (\Omega, \mathcal{A}) \longrightarrow (E, \mathcal{E})$ est l'identité, on a $\sigma(X) = \mathcal{B}$.
- σ est utilisé dans des situations très différentes. On est contraint dans le cadre d'un espace de probabilités $(\Omega, \mathcal{A}, \mathbb{P})$ de définir des tribus pas trop grosses (pas d'informations inutiles), et $\sigma(.)$ est toujours une sous-tribu de $\mathcal{A}$ de "la bonne taille" (assez pour contenir l'information nécessaire de la variable aléatoire, mais pas plus).
- Si A est un événement : $\sigma(A) = \{\emptyset, A, \overline{A}, \Omega\}$

3 Indépendance

On sort de la théorie de la mesure pour faire des probabilités avec la notion d'indépendance, dont on va rappeler brièvement les définitions :

Définition 3 (indépendance de tribus)

Soient $(\mathcal{B}_i)_{i\in I}$ des sous-tribus de $\mathcal{A}$. On dit que les $(\mathcal{B}_i)_{i\in I}$ sont indépendantes si :

$$\forall k, \forall i_1 < ... < i_k \in I, \forall B_1 \in \mathcal{B}_{i_1}, ..., B_k \in \mathcal{B}_{i_k}, \quad \mathbb{P}\left(\bigcap B_j\right) = \prod_{j=1}^{k} \mathbb{P}(B_j)$$

Définition 4 (indépendance d'événements)

Soient $(B_i)_{i\in I}$ des événements $\mathcal{A}$. On dit que les $(B_i)_{i\in I}$ sont indépendants si :

$$\forall k, \forall i_1 < ... < i_k \in I, \quad \mathbb{P}\left(\bigcap B_{i_j}\right) = \prod_{j=1}^{k} \mathbb{P}(B_{i_j})$$

Définition 5 (indépendance de variables aléatoires)

Soient $(X_i)_{i\in I}$ des variables aléatoires. On dit que les $(X_i)_{i\in I}$ sont indépendantes si $(\sigma(X_i))_{i\in I}$ sont des tribus indépendantes.

Remarques.

1. Les $(X_i : (\Omega, \mathcal{A}) \longrightarrow (E_i, \mathcal{E}_i))_{i\in I}$ sont indépendantes si et seulement si :

$$\forall k, \forall i_1 < ... < i_k \in I, \forall C_1 \in \mathcal{E}_{i_1}, ..., C_k \in \mathcal{E}_{i_k}, \quad \mathbb{P}\left(X \in \bigcap_{j\in[\![1,k]\!]} C_j\right) = \prod_{i=1}^{k} \mathbb{P}(X \in C_j)$$

2. $(A_i)_{i\in I}$ indépendantes $\Leftrightarrow$ $(\sigma(1_i))_{i\in I}$ indépendantes $\Leftrightarrow$ $(\mathbb{1}_{A_i})_{i\in I}$ indépendantes
3. L'indépendance est une propriété sur un nombre fini d'objets.
4. Si I est infini, $(B_i)_{i\in I}$ sont indépendantes si et seulement si toute sous-famille finie l'est.

Exemple. On jette 2 dès. On pose $A =$ "le premier est pair$''$ et $B =$ "le second est pair$''$ et $C =$ "ils sont égaux$''$. On prend $\Omega = [\![1, 6]\!]^2$, $\mathcal{A} = \mathcal{P}(\Omega)$ et $\mathbb{P}$ la probabilité uniforme. Alors on a :

$$\begin{cases} \mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{2} \\ \mathbb{P}(C) = \frac{1}{6} \\ \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \frac{1}{12} = \mathbb{P}(A)\mathbb{P}(C) = \mathbb{P}(B)\mathbb{P}(C) \\ \mathbb{P}(A \cap B) = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(B) \\ \mathbb{P}(A \cap B \cap C) = \frac{1}{12} \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) \end{cases}$$

Donc $\{A, B\}, \{A, C\}$ et $\{B, C\}$ sont des systèmes d'événements indépendants, mais $\{A, B, C\}$ ne l'est pas.

Remarques.

- L'indépendance a lieu lorsque l'on ne peut rien déduire sur l'un des objets à partir des autres. A indépendant de C car savoir que le premier dès est pair n'influe pas sur les chances qu'ils soient égaux.
  Par contre savoir A et C implique B, donc on n'a pas indépendance des trois.

- Pour 2 événements avec $\mathbb{P}(B) \neq 0$:

$$A \text{ et } B \text{ indépendants} \quad \Leftrightarrow \quad \mathbb{P}(A|B) = \mathbb{P}(A)$$

  Le fait que B soit réalisé n'influe pas sur A.

4 Loi d'une variable aléatoire : on substitue $\mathbb{P}$

La dernière étape avant de se débarrasser de notre espace de probabilité est de trouver un substitut à $\mathbb{P}$.

Définition 6 (loi d'une variable aléatoire)

Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$ une variable aléatoire. Sa loi μ_X est la mesure de probabilité sur $(E, \mathcal{E})$ définie par :

$$\forall A \in \mathcal{E},\ \mu_X(A) = \mathbb{P}(X \in A)$$

Remarque. μ_X est la mesure image de $\mathbb{P}$ par X.
On note $X \sim \mu$ lorsque X a pour loi la mesure de probabilité μ.

Définition 7 (espérance)

- Soit $X : (\Omega, \mathcal{A}) \to (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ une variable aléatoire. Alors on définit l'espérance de X comme :

$$\mathbb{E}[X] = \int_\Omega X(\omega) d\mathbb{P}(\omega)$$

- Soit $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une variable aléatoire telle que $\mathbb{E}[|X|] < +\infty$. Alors on définit l'espérance de X comme :

$$\mathbb{E}[X] = \int_\Omega X(\omega) d\mathbb{P}(\omega)$$

- Soit $X : (\Omega, \mathcal{A}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ une variable aléatoire. On note $X = (X_1, ..., X_d)$ et on définit l'espérance de X lorsque $\forall i, \mathbb{E}[|X_i|] < +\infty$, ou lorsque $\mathbb{E}[\|X\|] < +\infty$, par :

$$\mathbb{E}[X] = (\mathbb{E}[X_1], ..., \mathbb{E}[X_d])$$

Remarque. Si μ est une probabilité sur $(E, \mathcal{E})$ alors il existe une variable aléatoire de loi μ. Considérer en effet $(\Omega, \mathcal{A}, \mathbb{P}) = (E, \mathcal{E}, \mu)$ et X l'identité sur Ω.

Exemple. (lois de variables aléatoires)
On distingue 3 catégories :

- le cas discret : cas où il existe A dénombrable (ou fini) tel que $X \in A$ presque sûrement
- le cas continu : cas où X est à valeurs dans $\mathbb{R}^d$ et sa loi a une densité par rapport à la mesure de Lebesgue
- autre cas ?

1. [*Lois discrètes*] $X \in A$ presque sûrement avec A fini ou dénombrable.

   (a) $\exists c \in \mathcal{E}$ tq $X = c$ p.s. :

   $$\mathbb{P}(X \in A)\mathbb{P}(X_c)\mathbb{1}_{c\in A} + \underbrace{\mathbb{P}(X \neq c \cap X \in A)}_{\leqslant \mathbb{P}(X\neq c)=0} = \delta_c(A)$$

   où $\delta_c(A) : A \longmapsto \mathbb{1}_{c\in A}$. On a $X = c$ p.s. $\Leftrightarrow$ $X \sim \delta_c$.

   (b) Plus généralement, si $A = \{a_1, ..., a_i, ...\}$, on note $p_i = \mathbb{P}(X = a_i)$. Automatiquement, $\sum_i p_i = \sum_i \mathbb{P}(X = a_i) \underset{\text{union disjointe}}{=} \mathbb{P}(A) = 1$.

   On a alors :

   $$\begin{aligned}\mu_X(B) = \mathbb{P}(X \in B) &\underset{\text{par (prop1)}}{=} \mathbb{P}(X \in B, X \in A) \\ &= \mathbb{P}\left(\bigsqcup_{a_i\in B} X = a_i\right) = \sum_{i \ \mid \ a_i\in B} \mathbb{P}(X = a_i) \\ &= \sum_i p_i\delta_{a_i}(B) = \left(\sum_i \delta_{a_i}\right)(B)\end{aligned}$$

   Donc $\mu_X = \sum_i \delta_{a_i}$.

   Remarques.
   - La loi d'une variable aléatoire discrète est donnée par l'ensemble $\{a_1, ..., a_i, ...\}$, où cette variable aléatoire vit presque sûrement, et la suite des $p_i = \mathbb{P}(X = a_i)$.
   - Pour une mesure μ, on dit que x est un **atome** si $\mu(\{x\}) > 0$. Donc une mesure discrète a pour atomes tous les a_i tels que $p_i > 0$ et donc est "portée" par ses atomes. On appelle ces mesures des **mesures purement atomiques**.
   - Il y a des mesures sans atomes (comme la mesure de Lebesgue).

   Soit $A \subseteq \mathbb{R}^d$. Soit X une variable aléatoire discrète et $X \geqslant 0$ p.s. (ou si $|X|$ est intégrable) :

   $$\begin{aligned}\mathbb{E}[X] &= \int_{X^{-1}(A)} X(\omega)d\mathbb{P}(\omega) + \underbrace{\int_{X^{-1}(\overline{A})} X(\omega)d\mathbb{P}(\omega)}_{=0} \\ &= \sum_i \int_{X^{-1}(a_i)} \underbrace{X(\omega)}_{a_i}d\mathbb{P}(\omega) = \sum_i a_ip_i\end{aligned}$$

(c) Si $X \in \{0,1\}$ p.s., soit $p = \mathbb{P}(X=1)$. Alors $\mu_X = p\delta_1 + (1-p)\delta_0$. On appelle cette loi **la loi de Bernoulli de paramètre** p **notée** $\mathcal{B}(p)$. Elle modélise un événement qui a une probabilité p de succès. On a $\mathbb{E}[X] = p$.

(d) Si $X \in \{a_1, ..., a_n\}$ p.s. (les $(a_i)_i$ sont deux à deux disjoints) avec $\mathbb{P}(a_i) = \frac{1}{n}$ on a $\mu_X = \frac{1}{n}\sum_{1 \leqslant i \leqslant n} \delta_{a_i}$ et on dit que X suit **la loi uniforme sur** $\{a_1, ..., a_n\}$. On a $\mathbb{E}[X] = \frac{1}{n}\sum_{1 \leqslant i \leqslant n} a_i$.

(e) Si $X_1, ..., X_n$ sont des variables aléatoires indépendantes de même loi $\mathcal{B}(p)$. On pose $X = \sum_{1 \leqslant i \leqslant n} X_i$ qui modélise le nombre de succés dans une répétition de n expériences indépendantes ayant une probabilité p de réussir.

On a $X \in \{0, ..., n\}$ p.s. Sa loi est déterminée par $\mathbb{P}(X = k)$ pour $k \in [\![0, n]\!]$:

$$\begin{aligned}
\mathbb{P}(X=k) &= \mathbb{P}(\text{“exactement } k\ X_i \text{ valent } 1'') \\
&= \mathbb{P}\left(\bigsqcup_{i_1 < \ldots < i_k} \big(\forall j,\ X_{i_j} = 1 \text{ et } \forall i \notin \{i_1, ..., i_k\},\ X_i = 0\big) \right) \\
&= \sum_{i_1 < \ldots < i_k} \mathbb{P}\left(\forall j,\ X_{i_j} = 1 \text{ et } \forall i \notin \{i_1, ..., i_k\},\ X_i = 0\right) \\
&\underset{\text{indép}}{=} \sum_{i_1 < \ldots < i_k} \left(\prod_{j=1}^{k} \mathbb{P}(X_{i_j} = 1) \prod_{\notin \{i_1, ..., i_k\}} \mathbb{P}(X_j = 0) \right) \\
&= \binom{n}{k} p^k (1-p)^{n-k}
\end{aligned}$$

On appelle cette loi la loi **binomiale de paramètre** p**, notée** $\mathcal{B}(n,p)$.

On a $\mathbb{E}[X] = \mathbb{E}[\sum X_i] = \sum \mathbb{E}[X_i] = np$.

(f) On se donne $k \geqslant 2$, $p_1, ..., p_k \in [0,1]$ tels que $\sum p_i = 1$. On se donne $X_1, ..., X_n$ indépendants de même loi $\mathbb{P}(X_i = j) = p_j$.
On note, pour $j \in [\![1, k]\!]$, $N_j = \mathrm{Card}(\{1 \leqslant i \leqslant n\ |X_i = j\})$ (“le nombre de particules tombées dans chacune des k boîtes”).

On note $N = \begin{pmatrix} N_1 \\ \vdots \\ N_k \end{pmatrix}$. On a $N \in \{(x_1, ..., x_k) \in \mathbb{N}^k \mid \sum x_i = n\}$.

Généralisation de la loi précédente : on sait que $N_j \sim \mathcal{B}(n, p_j)$. Quelle est la loi de N ?

Soit $(x_1, ..., x_k) \in \mathbb{N}^k$ tels que $\sum x_i = n$.

$$\begin{aligned}
\mathbb{P}\left(N = \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}\right) &= \mathbb{P}\left(\bigsqcup_{I_1,...,I_k \text{ partition de } [\![1,n]\!] \ | \ |I_j|=x_j} (\forall j, \forall i \in I_j, X_i = j)\right) \\
&= \sum_{I_1,...,I_k \text{ partition de } [\![1,n]\!] \ | \ |I_j|=x_j} \mathbb{P}(\forall j, \forall i \in I_j, X_i = j) \\
&= \sum_{I_1,...,I_k \text{ partition de } [\![1,n]\!] \ | \ |I_j|=x_j} \prod_j \prod_{i \in I_j} p_j \qquad \text{(indépendance)} \\
&= \sum_{I_1,...,I_k \text{ partition de } [\![1,n]\!] \ | \ |I_j|=x_j} \prod_j p_j^{x_j} \\
&= \binom{n}{x_1}\binom{n-x_1}{x_2}\binom{n-x_1-x_2}{x_3}...\binom{n-x_1-...-x_{k-1}}{x_k} \times \prod_{j=1}^k p_j^{x_j} \\
&= \underbrace{\frac{n!}{x_1!...x_k!}}_{\binom{n}{x_1...x_k}} \prod_{j=1}^k p_j^{x_j}
\end{aligned}$$

C'est la **loi multinomiale de paramètre** $(n, (p_1, ..., p_k))$.

(g) Si $X \in \mathbb{N}$ p.s. et s'il existe $\lambda > 0$ tel que :

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

On dit que X suit une **loi de Poisson de paramètre** λ **notée** $\mathcal{P}(\lambda)$.

(h) Si $(X_i)_{i \geqslant 1}$ sont des variables aléatoires indépendantes de loi $\mathcal{B}(p)$ (on montrera l'existence d'une telle suite plus tard), on pose $X = \inf\{i \mid X_i = 1\}$ (avec la convention $\inf \emptyset = +\infty$).

On modélise le nombre d'expériences à faire pour obtenir un premier succès dans une suite d'expériences indépendantes ayant une probabilité $p \in]0, 1[$ de réussir. On a $X \in \mathbb{N} \cup \{+\infty\}$ p.s. On a :

$$\begin{aligned}
\mathbb{P}(X = k) &= \mathbb{P}(\forall i < k, X_i = 0 \text{ et } X_k = 1) \\
&\underset{\text{indép}}{=} \prod_{i=1}^{k-1} \mathbb{P}(X_i = 0)\mathbb{P}(X_k = 1) \\
&= (1-p)^{k-1} p
\end{aligned}$$

Et $\mathbb{P}(X = +\infty) \leqslant \mathbb{P}(\forall i \leqslant n, X_i = 0) = (1-p)^k \xrightarrow[n \to +\infty]{} 0$ donc $\mathbb{P}(X = +\infty) = 0$. Donc $X \in \mathbb{N}$ p.s. Cette loi est la **loi géométrique de paramètre** p **notée** $\mathcal{G}(p)$.

2. [*Lois continues*] On dit que X a une loi de densité $f \colon \mathbb{R}^d \longrightarrow \mathbb{R}_+$ mesurable par rapport à Lebesgue si pour tout A mesurable sur $\mathbb{R}^d$, on a $\mathbb{P}(X \in A) = \int_A f dx$ (intégration par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$). Forcément, $\int_{\mathbb{R}^d} f dx = 1$. f est définie Leb$-$p.p. Dans ce cas-là, la loi de X est déterminée par f.

(a) Soit x à valeurs dans $\mathbb{R}$ ayant pour densité $f(x) = \lambda e^{-\lambda x}\mathbb{1}_{x>0}$ avec $\lambda > 0$. Alors on dit que X suit la **loi exponentielle de paramètre λ noté $\mathcal{E}(\lambda)$**.

(b) Si X à valeurs dans $\mathbb{R}^d$ a une loi de densité $\frac{\mathbb{1}_{x\in D}}{\text{Leb}(D)}$ pour D mesurable tel que $0 < \text{Leb}(D) < +\infty$, alors on dit que la loi de X **est uniforme sur** D.

(c) Sur $\mathbb{R}$, si X a pour densité $f(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$, on dit que X a pour loi la **gaussienne centrée réduite notée** $\mathcal{N}(0,1)$.

Si X a cette loi, si $m \in \mathbb{R}$ et $\sigma \in \mathbb{R}^*$, soit $Y = m + \sigma X$. Quelle est la loi de Y ? On a $Y \in \mathbb{R}$ et :

$$\mathbb{P}(Y \in A) = \mathbb{P}(m+\sigma X \in A) = \mathbb{P}\left(X \in \frac{A-m}{\sigma}\right) = \int_{\frac{A-m}{\sigma}} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}dx \underset{x=\frac{y-m}{\sigma}}{=} \int_A \frac{e^{-\frac{(y-m)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}dy$$

Donc Y suit la loi de densité $\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y-m)^2}{2\sigma^2}\right)$, appelée **gaussienne de moyenne m et de variance σ^2 et notée** $\mathcal{G}(m,\sigma^2)$.

3. [Autre loi ?] Sur $\mathbb{R}$: $\mu = \frac{1}{2}\sigma_0 + \frac{1}{2}\mathcal{E}(\lambda)$ est une mesure de probabilité.

$$\mu(A) = \frac{1}{2}\mathbb{1}_{0\in A} + \frac{1}{2}\int_A \lambda e^{-\lambda x}\mathbb{1}_{x>0}dx$$

Existe-t-il d'autres mesures de probabilité ?

Théorème 1 (décomposition de Lebesgue)

Si μ est une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, alors μ se décompose sous la forme :

$$\mu = \alpha\mu_d + \beta\mu_c + \gamma\mu_s$$

avec $\alpha + \beta + \gamma = 1$, et :

- μ_d est une mesure discrète,
- μ_c est une mesure continue,
- μ_s est une mesure singulière, i.e. qui n'a pas d'atomes et telle que :

$$\exists B \in \mathcal{B}(\mathbb{R}) \mid \mu_s(B) = 1 \text{ et } \text{Leb}(B) = 0$$

Cette décomposition est essentiellement unique.

Remarque. Pour une mesure à densité f, si $\text{Leb}(B) = 0$, alors $\mu(B) = 0$. Si de plus $\mu(\bar{B}) = 0$ alors $f = 0$ Leb-p.p. Cette condition s'oppose au fait d'avoir une densité !

Exemple. (mesure singulière)

- Si $b \geqslant 2$ et $x \in [0,1[$, x a une unique décomposition b-addique :

$$x = \sum_{i\geqslant 1} \frac{x_i}{b^i}$$

où les $x_i \in \{0,...,b-1\}$ sont non tous égaux à $b-1$ à partir d'un certain rang.

- On regarde le cas $b = 3$, posons :

$$K_0 = [0,1] = \left\{ \sum_{i>0} \frac{x_i}{3^i} \mid x_i \in [\![0,2]\!] \text{ non tous égaux à 2 à partir d'un certain rang} \right\}$$

Puis :

$$K_1 = \frac{1}{3}K_0 \cup \left(\frac{2}{3} + \frac{1}{3}K_0\right) = \left\{ \sum_{i>0} \frac{x_i}{3^i} \mid x_1 \in \{0,2\} \text{ et } x_{i>1} \in [\![0,2]\!] \text{ non tous ...} \right\}$$

Et définissons par récurrence :

$$K_n = \frac{1}{3}K_{n-1} \cup \left(\frac{2}{3} + \frac{1}{3}K_{n-1}\right) = \left\{ \sum_{i>0} \frac{x_i}{3^i} \mid x_1, ..., x_n \in \{0,2\} \text{ et } x_{i>n} \in [\![0,2]\!] \text{ non tous ...} \right\}$$

On obtient alors l'ensemble triadique de Cantor par :

$$K = \bigcap_{n\in\mathbb{N}} K_n = \left\{ \sum_{i>0} \frac{x_i}{3^i} \mid x_i \in \{0,2\} \right\}$$

Il est facile de vérifier que K est compact dans $[0,1]$. Pour autant, K est de mesure nulle. En effet :

$$\text{Leb}(K_n) = \text{Leb}\left(\frac{1}{3}K_{n-1}\right) + \text{Leb}\left(\frac{2}{3} + \frac{1}{3}K_{n-1}\right) = \frac{2}{3}\text{Leb}(K_{n-1}) \underset{\text{réc}}{=} \left(\frac{2}{3}\right)^n \xrightarrow[n\to+\infty]{} 0$$

On a $\text{Leb}(K) \leqslant \text{Leb}(K_n) \xrightarrow[n\to+\infty]{} 0$ donc $\text{Leb}(K) = 0$.

Soit $X \sim \mathcal{U}([0,1])$ et :

$$\varphi \colon \begin{cases} [0,1[& \longrightarrow & K \\ \sum_{i>0} \frac{x_i}{3^i} & \longmapsto & \sum_{i>0} \frac{2x_i}{3^i} \end{cases}$$

φ est injective. Soit $Y = \varphi(X)$. Alors μ_Y est singulière. En effet :

$$\mu_{\{x\}} = \mathbb{P}(Y = x) = \mathbb{P}(\varphi(X) = x) = \mathbb{P}\Big(X \in \underbrace{\varphi^{-1}(\{x\})}_{\text{singleton ou vide par inj.}}\Big) = \text{Leb}(\varphi^{-1}(\{x\})) = 0$$

Il n'y a donc pas d'atomes. On n'a pas non plus de densité, puisque $\mu_Y(K) = 1$ et $\text{Leb}(K) = 0$. *On a bien construit une mesure singulière.*

5 Espérance

Théorème 2 (Fatou)

Si $(X_n)_{n\geqslant 1}$ est une suite de v.a. $\geqslant 0$ p.s. ($\forall n, X_n \geqslant 0$ p.s.), alors :

$$\mathbb{E}[\liminf X_n] \leqslant \liminf \mathbb{E}[X_n]$$

Théorème 3 (de convergence monotone)

Si $(X_n)_{n\geqslant 1}$ est une suite croissante p.s. de v.a. positives p.s. ($\forall n, 0 \leqslant X_n \leqslant X_{n+1}$ p.s.), alors :

$$\mathbb{E}\left[\lim_{n\to+\infty}\uparrow X_n\right] = \lim \uparrow \mathbb{E}[X_n]$$

($\lim \uparrow X_n = \lim \uparrow X_n$ lorsqu'elle existe, 0 sinon)

Théorème 4 (de convergence dominée)

Si $(X_n)_{n\geqslant 1}$ est une suite de v.a. réelles telle que :

- $X_n \xrightarrow[n\to+\infty]{} X$ p.s..
- $\forall n,\ |X_n| \leqslant Z$ p.s. où $\mathbb{E}[Z] < +\infty$

alors,

$$\mathbb{E}[X_n] \xrightarrow[n\to+\infty]{} \mathbb{E}[X]$$

On veut calculer des espérances sans faire de calcul dans Ω.

Théorème 5 (de transfert)

$X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$, μ une probabilité sur $(E, \mathcal{E})$. Alors :

$$X \sim \mu \quad \Leftrightarrow \quad \forall f : (E,\mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})) \text{ mesurable positive, } \mathbb{E}[f(X)] = \int_E f(x) d\mu(x)$$

Intérêt 1. Calculer des espérances sans passer par Ω. Si $X \sim \mathcal{E}(2)$, que vaut $\mathbb{E}[e^X]$?

$$\mathbb{E}[e^X] = \int_{\mathbb{R}} e^x d\mu_X(x) = \int_{\mathbb{R}} e^x \mathbb{1}_{x>0} 2e^{-2x} dx = 2\int_0^{+\infty} e^{-x} dx = 2$$

Intérêt 2. (crucial)
Trouver la loi d'une v.a. Soit $X \sim \mathcal{E}(\lambda)$ et $\theta > 0$, quelle est la loi de $Y = \theta X$? Soit f mesurable positive, calculons :

$$
\begin{aligned}
\mathbb{E}[f(Y)] &= \mathbb{E}[f(\theta X)] \\
&= \int_{\mathbb{R}} f(\theta x) d\mu_X(x) \qquad \text{théorème de transfert appliqué à } x \to f(\theta x) \\
&= \int_{\mathbb{R}} f(\theta x) \lambda e^{-\lambda x} \mathbb{1}_{x>0} dx \\
&= \int_{\mathbb{R}} f(y) \left(\frac{\lambda}{\theta}\right) e^{-\frac{\lambda}{\theta} y} \mathbb{1}_{y>0} dy \\
&= \int_{\mathbb{R}} f(y) d\mu_{\xi(\frac{\lambda}{\theta})}(y)
\end{aligned}
$$

Vrai pour toute f mesurable positive, donc on en déduit $Y \sim \xi(\frac{\lambda}{\theta})$.

Preuve.

$\Leftarrow$: Soit $B \in \xi$, posons $f = \mathbb{1}_B$. On a :

$$
\begin{aligned}
\mathbb{E}[f(X)] &= \mathbb{E}[\mathbb{1}_{X \in B}] \\
&= \int_{\Omega} \mathbb{1}_{\{\omega | X(\omega) \in B\}} d\mathbb{P}(\omega) \\
&= \mathbb{P}(X \in B) \\
&= \mu_X(B)
\end{aligned}
$$

Or :

$$
\begin{aligned}
\mathbb{E}[f(X)] &= \int_E f(x) d\mu(x) \\
&= \int_E \mathbb{1}_{x \in B} d\mu(x) \\
&= \mu(B)
\end{aligned}
$$

Vrai pour tout B donc $\mu_X = \mu$, d'où $X \sim \mu$.

$\Rightarrow$: Si $X \sim \mu$ pour quels f a-t-on $\mathbb{E}[f(X)] = \int f(x) d\mu(x)$?

- c'est vrai pour $f = \mathbb{1}_B$ puisqu'alors cela se réécrit $\mu_X(B) = \mu(B)$
- par linéarité c'est vrai pour f étagée
- si f est mesurable positive, il existe une suite f_n de fonctions étagées positives telle que $f = \lim \uparrow f_n$.

Alors :

$$
\mathbb{E}[f(X)] \underset{\text{TCM}}{=} \lim \uparrow \mathbb{E}[f_n(X)] = \lim \uparrow \int f_n(x) d\mu(x) \underset{TCM}{=} \int f(x) d\mu(x)
$$

□

Remarque. Dans le théorème on peut remplacer "Pour toute f mesurable positive" par :

- pour toute f indicatrice
- pour tout f mesurable bornée

Théorème 6 (cas particulier de $\mathbb{R}^d$)

X v.a. à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ et μ une probabilité sur $\mathbb{R}^d$. On a équivalence entre :

1. $X \sim \mu$
2. $\forall f$ mesurable positive $\mathbb{R}^d \to \mathbb{R}$, $\mathbb{E}[f(X)] = \int_{\mathbb{R}^d} f(x)d\mu(x)$
3. $\forall f$ mesurable bornée $\mathbb{R}^d \to \mathbb{R}$, $\mathbb{E}[f(X)] = \int_{\mathbb{R}^d} f(x)d\mu(x)$
4. $\forall f$ continue bornée $\mathbb{R}^d \to \mathbb{R}$, $\mathbb{E}[f(X)] = \int_{\mathbb{R}^d} f(x)d\mu(x)$
5. $\forall f$ continue à support compact $\mathbb{R}^d \to \mathbb{R}$, $\mathbb{E}[f(X)] = \int_{\mathbb{R}^d} f(x)d\mu(x)$
6. $\forall \lambda \in \mathbb{R}^d$, $\mathbb{E}[e^{i<\lambda,X>}] = \int_{\mathbb{R}^d} e^{i<\lambda,x>}d\mu(x)$

Preuve.
On a déjà :

$$1 \Leftrightarrow 2 \Leftrightarrow 3 \Rightarrow 4 \Rightarrow 5$$

$5 \Rightarrow 1$: Montrons que $\forall a_1 < b_1, ..., a_d < b_d$, si C est le pavé $[a_1, b_1] \times ... \times [a_d, b_d]$ alors $\mathbb{P}(X \in C) = \mu(C)$.
Fixons C. Posons pour $n \in \mathbb{N}$:

$$g_n(x) = (1 - nd(x, C))_+$$

($x_+ = max(x, 0)$). g_n est continue à support compact. $\|g_n\|_\infty \leqslant 1$ donc g_n est intégrable pour μ_X et μ puisque ce sont des probabilités. Et $g_n(x) \xrightarrow[n\to+\infty]{} \mathbb{1}_{x\in C}$. Ainsi :

$$\mu_X(C) = \mathbb{E}[\mathbb{1}_{X\in C}] \underset{\text{TCD}}{=} \lim \mathbb{E}[g_n(X)] = \lim \int g_n d\mu \underset{\text{TCD}}{=} \int \mathbb{1}ux \in C d\mu = \mu(C)$$

μ_X et μ coïncident sur $\mathcal{X} = \{\text{pavés}\}$ classe stable par intersection finie, donc elles coïncident sur la tribu engendrée par le lemme des classes monotones. Or $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R}^d)$.

$6 \Rightarrow 5$: Soit f à support compact, on veut approcher f par une fonction périodique. Si $T > 0$ tel que Support$(f) \subseteq \left[-\frac{T}{2}; \frac{T}{2}\right]^d$, on définit le périodisé de f par :

$$f_T(x) = \begin{cases} f(x) \text{ si } x \in \left[-\frac{T}{2}; \frac{T}{2}\right]^d \\ f\left(x + \begin{pmatrix} k_1 T \\ \vdots \\ k_d T \end{pmatrix}\right) \text{ sinon} \end{cases}$$

où k_i est l'unique entier relatif tel que $-\frac{T}{2} \leqslant x_i + k_i T < \frac{T}{2}$.

f_T vérifie : f_T est continue, T-périodique dans les d directions et $\|f_T\| = \|f\|_\infty$. De plus,

$$f_T|_{[-\frac{T}{2};\frac{T}{2}]^d} = f|_{[-\frac{T}{2};\frac{T}{2}]^d}$$

On a :

$$\mu\left(\overline{\left[-\frac{T}{2};\frac{T}{2}\right]}\right) \xrightarrow[T\to+\infty]{} 0 \quad \text{et} \quad \mu_X\left(\overline{\left[-\frac{T}{2};\frac{T}{2}\right]}\right) \xrightarrow[T\to+\infty]{} 0$$

Donc si on fixe $\varepsilon > 0$, on peut trouver T tel que $\int \mathbb{1}_{\|x\|_\infty \geqslant \frac{T}{2}} d\mu \leqslant \varepsilon$, $\mathbb{P}(\|X\|_\infty \geqslant \frac{T}{2}) \leqslant \varepsilon$. Et alors :

$$\begin{aligned}\mathbb{E}[f(X)] - \int f(x)d\mu(x) &= \mathbb{E}\left[f_T(X)\mathbb{1}_{\|X\|_\infty \leqslant \frac{T}{2}}\right] - \int f_T(x)\mathbb{1}_{\|x\|_\infty \leqslant \frac{T}{2}} d\mu(x) \\ &= \mathbb{E}[f_T(X)] - \int f_T(x)d\mu(x) \\ &\quad - \left(\mathbb{E}\left[f_T(X)\mathbb{1}_{\|X\|_\infty \geqslant \frac{T}{2}}\right] - \int f(x)\mathbb{1}_{\|x\|_\infty \geqslant \frac{T}{2}} d\mu(x)\right)\end{aligned}$$

Or,

$$\left|\mathbb{E}\left[f_T(X)\mathbb{1}_{\|X\|_\infty > \frac{T}{2}}\right]\right| \leqslant \|f_T\|_\infty \mathbb{P}\left(\|X\|_\infty > \frac{T}{2}\right) \leqslant \|f\|_\infty \varepsilon$$

Il existe un polynôme trigonométrique $P_{\varepsilon,T}$ tel que $\|P_{\varepsilon,T}\|_\infty \leqslant \varepsilon$. Donc :

$$\begin{aligned}E[f(X)] - \int f(x)d\mu(x) \leqslant &\underbrace{\left|E[P(X)] - \int P(x)d\mu(x)\right|}_{=0} + \left|E[\underbrace{(f_T - P)}_{\leqslant \varepsilon}(X)]\right| \\ &+ \left|\int \underbrace{(f_T - P)}_{\leqslant \varepsilon}(x)d\mu(x)\right| + 2\|f\|_\infty \varepsilon\end{aligned}$$

□

Définition 8 (fonction caractéristique)

Si X est à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ on définit sa fonction caractéristique par :

$$\phi_X : \left\{\begin{array}{ccc} \mathbb{R} & \longrightarrow & \mathbb{C} \\ \lambda & \longmapsto & \mathbb{E}[e^{i<\lambda|X>}]\end{array}\right.$$

Remarque. On a prouvé deux résultats :

- ϕ_X ne dépend que de μ_X. Donc on peut définir pour μ probabilité sur $\mathbb{R}^d$, $\phi_\mu(x) = \int e^{i<\lambda|x>} d\mu(x)$
- propriété : $\phi_X = \phi_Y \Rightarrow X$ et Y ont la même loi (idem pour ϕ_μ, ϕ_ν)

Application. Si X est dans $\mathbb{R}^d$ a une loi de densité $f(x_1, ..., x_d)$ par rapport à Lebesgue$_{\mathbb{R}^d}$, alors $\forall i,\ X_i$ a une loi de densité $\int f(x_1, ..., x_d)dx_1...dx_{i-1}dx_{i+1}...dx_d$ par rapport à Lebesgue$_{\mathbb{R}}$

Preuve. La loi de X_i ? Soit g mesurable positive :

$$\begin{aligned}
\mathbb{E}[g(X_i)] &= \int_{\mathbb{R}^d} g(x_i) d\mu_{X_1...X_d}(x_1...x_d) \qquad \text{(transfert)}\\
&= \int_{\mathbb{R}^d} g(x_i) f(x_1...x_d) dx_1...dx_d\\
&= \int_{\mathbb{R}} g(x_i) \left(\int_{\mathbb{R}^{d-1}} f(x_1...x_d) dx_1....dx_{i-1} dx_{i+1}...dx_d \right) dx_i \quad \text{(Fubini)}\\
&= \int_{\mathbb{R}} g(x_i) d\mu_{X_i}(x_i) \qquad \text{Vrai pour tout } g
\end{aligned}$$

□

Remarque. Si X est à valeurs dans un espace produit $(E_1 \times ... \times E_d, \xi_1 \otimes ... \otimes \xi_d)$, μ_X détermine $\mu_{X_1}, ..., \mu_{X_d}$. En effet :

$\mu_{X_i}(B) = \mathbb{P}(X_i \in B) = \mathbb{P}(X \in E_1 \times ... E_{i-1} \times B \times E_{i+1} ... \times E_d) = \mu_X(E_1 \times ... \times E_{i-1} \times B \times E_{i+1} ... \times E_d)$

Mais la réciproque est fausse :

- $(X, Y) \sim \mathcal{U}([0,1]^2)$ alors $\mu_X = \mu_Y = \mathcal{U}([0,1])$
- $Z \sim \mathcal{U}([0,1])$, $\tilde{X} = Z$, $\tilde{Y} = Z$ alors $\mu_{\tilde{X}} = \mu_{\tilde{Y}} = \mathcal{U}([0,1])$

mais $\mu_{(X,Y)} \neq \mu_{(\tilde{X},\tilde{Y})}$ par exemple pour $\Delta = \{(Z,Z) \mid 0 \leqslant Z \leqslant 1\}$ on a $\mu_{(X,Y)}(\Delta) = 0 \neq 1 = \mu_{(\tilde{X},\tilde{Y})}(\Delta)$

Définition 9 (espace L^p)

Pour $p \geqslant 1$ on définit :

$$L^p(\mathbb{R}^d) = \{X \text{ v.a. à valeurs dans } (R^d, \mathcal{B}(\mathbb{R}^d)) \mid \mathbb{E}[\|X\|^p] < +\infty\}$$

Remarque.

- ne dépend pas du choix de la norme
- si $p < q$, $L^q \subseteq L^p$, $|x|^p \leqslant |x|^q$ dès que $|x| \geqslant 1$ donc $|x|^p \leqslant 1 + |x|^q$ $\forall x \in \mathbb{R} \Rightarrow \mathbb{E}[\|X\|^p] \leqslant 1 + \mathbb{E}[\|X\|^q]$ (spécifique aux espaces de probabilité où 1 est intégrable)

5.1 Rappels

- $L^p(\mathbb{R}^d)$ muni de $\|.\|_p$ où $\|X\|_p = \mathbb{E}[\|X\|^p]^{\frac{1}{p}}$ est un espace normé complet
- on a l'inégalité de Hölder : si $\frac{1}{p} + \frac{1}{q} = 1$ alors $\mathbb{E}[\|XY\|] \leqslant \|X\|_p \|Y\|_q$
- on dispose des 3 inégalités classiques suivantes

Proposition 1 (inégalité de Markov)

Si $X \geqslant 0$ p.s.

$$\mathbb{P}(X \geqslant t) \leqslant \frac{1}{t}\mathbb{E}[X] \quad \forall t > 0$$

Preuve.

$$t\mathbb{1}_{X \geqslant t} \leqslant X \quad \text{p.s.}$$

$$\mathbb{E}[t\mathbb{1}_{X \geqslant t}] \leqslant \mathbb{E}[X] \quad \Rightarrow \quad \mathbb{P}(X \geqslant t) \leqslant \frac{1}{t}\mathbb{E}[X]$$

□

Définition 10 (variance, covariance)

Pour $X \in L^2$, on définit :

$$\mathrm{Var} X = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

Si $X, Y \in L^2$ on définit :

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

Proposition 2 (inégalité de Tchebychev)

$$\forall t \geqslant 0 \quad \mathbb{P}(|X - \mathbb{E}[X]| \geqslant t) \leqslant \frac{1}{t^2}\mathrm{Var} X$$

Preuve.

$$\mathbb{P}(|X - \mathbb{E}[X]| \geqslant t) = \mathbb{P}((X - \mathbb{E}[X])^2 \geqslant t^2) \underset{\text{Markov}}{\leqslant} \frac{\mathrm{Var} X}{t^2}$$

□

Proposition 3 (inégalité de Jensen)

Si φ est convexe et $X \in L^1$ alors $\varphi(\mathbb{E}[X]) \leqslant \mathbb{E}[\varphi(X)]$.

Remarque. Cette inégalité contient :

- $|\mathbb{E}[X]| \leqslant \mathbb{E}[|X|]$ ($\varphi(X) = |X|$)
- $\mathbb{E}[X]^2 \leqslant \mathbb{E}[X^2]$

Preuve.
Une fonction convexe est l'enveloppe supérieure des droites qui la minore.

$$\varphi(x) = \sup\{ax + b \mid (a, b) \text{ tq } \forall y : ay + b \leqslant \varphi(y)\}$$

En effet :

- si a, b sont tels que $ay + b \leqslant \varphi(y)$ $\forall y$ alors $ax + b \leqslant \varphi(x)$ donc $\varphi(x) \geqslant \sup\{...\}$
- inversement pour x fixé la pente entre x et y est $\frac{\varphi(y)-\varphi(x)}{y-x}$, décroissante en y, minorée par les pentes à gauche de x donc converge vers α lorsque $y \to x$. $\varphi(x) + \alpha(y - x)$ est la tangente à droite de φ en x et ne recroise jamais φ par convexité. Cette fonction affine apparaît dans le sup et vaut $\varphi(x)$ en x donc $\varphi(x) \leqslant \sup\{...\}$

alors :

$$\begin{aligned}
\varphi(\mathbb{E}[X]) &= \sup_{\forall y, ay+b\leqslant\varphi(y)} [a\mathbb{E}[X]+b] \\
&= \sup_{\forall y, ay+b\leqslant\varphi(y)} \mathbb{E}[\underbrace{aX+b}_{\leqslant\varphi(X)}] \\
&\leqslant \sup_{\forall y, ay+b\leqslant\varphi(y)} \mathbb{E}[\varphi(X)] \\
&\leqslant \mathbb{E}[\varphi(X)]
\end{aligned}$$

□

Exemples.
Si $Z \sim \mathcal{N}(0,1)$

$$\mathbb{E}[Z] = \int_{\mathbb{R}} y d\mu_Z(y) = \int_{\mathbb{R}} y e^{-\frac{y^2}{2}} \frac{dy}{\sqrt{2\pi}} = 0$$

$$\begin{aligned}
\mathrm{Var} Z = \mathbb{E}[Z^2] &= \int_{\mathbb{R}} y^2 d\mu_Z(y) = \int_{\mathbb{R}} y \times \underbrace{y e^{-\frac{y^2}{2}}}\frac{dy}{\sqrt{2\pi}} \\
&\underset{\text{IPP}}{=} \left[y\left(-e^{-\frac{y^2}{2}}\right)\right]_{-\infty}^{+\infty} + \int_{\mathbb{R}} e^{-\frac{y^2}{2}} \frac{dy}{\sqrt{2\pi}} = 1
\end{aligned}$$

Si $X \sim \mathcal{N}(m, \sigma^2)$, X a la même loi que $m + \sigma Z$ et donc $\mathbb{E}[X] = \mathbb{E}[m+\sigma Z] = m + \sigma \mathbb{E}[Z] = m$ et $\mathrm{Var} X = \mathrm{Var}(m+\sigma Z) = \sigma^2 \mathrm{Var} Z = 1$.

Lemme 1

$$\phi_{\mathcal{N}(m,\sigma^2)}(\lambda) = e^{i\lambda m} e^{-\frac{\lambda^2\sigma^2}{2}}$$

Preuve.

1. On a

   $$\phi_{\mathcal{N}(0,1)}(\lambda) = \int_{\mathbb{R}} e^{i\lambda x} e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}$$

   or

   $$\frac{\partial}{\partial\lambda}\left(e^{i\lambda x} e^{-\frac{x^2}{2}}\right) = ixe^{i\lambda x}e^{-\frac{x^2}{2}} \qquad \text{intégrable}$$

   Théorème de dérivation :

   $$\begin{aligned}
   \phi'_{\mathcal{N}(0,1)}(\lambda) &= \int_{\mathbb{R}} ie^{i\lambda x} x e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \\
   &\underset{\text{IPP}}{=} \left[ie^{i\lambda x}\left(-e^{-\frac{x^2}{2}}\right)\right]_{-\infty}^{+\infty} + \int_{\mathbb{R}} i^2\lambda e^{i\lambda x} e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}
   \end{aligned}$$

   Donc :

   $$\phi'_{\mathcal{N}(0,1)}(\lambda) = -\lambda \phi_{\mathcal{N}(0,1)}(\lambda)$$

   D'où finalement,

   $$\phi_{\mathcal{N}(0,1)}(\lambda) = \phi_{\mathcal{N}(0,1)}(0) e^{-\frac{\lambda^2}{2}}$$

2. Fait général :
$$\phi_{aX+b}(\lambda) = e^{ibx}\phi_X(a\lambda) = \mathbb{E}[i(aX+b)\lambda]$$
Donc si $X \sim \mathcal{N}(m, \sigma^2)$, X a la même loi que $m + \sigma Z$ avec $Z \sim \mathcal{N}(0,1)$. Ainsi :
$$\phi_{\mathcal{N}(m,\sigma^2)}(\lambda) = \phi_{m+\sigma Z}(\lambda) = e^{i\lambda m}\phi_Z(\sigma\lambda) = e^{i\lambda m}e^{-\sigma^2\frac{\lambda^2}{2}}$$

$\square$