Université d'Artois
Faculté des Sciences Jean Perrin
Mesure et Intégration (Licence 3 Mathématiques-Informatique)
Daniel Li

Introduction aux Probabilités

May 15, 2006

1 Généralités.

La Théorie des Probabilités a été formalisée par Kolmogorov en 1933, en se basant sur la Théorie de la mesure.

1.1 Espace de probabilité.

Définition 1 *Une* **probabilité** *(ou mesure de probabilité) sur un espace mesurable* $(\Omega, \mathcal{A})$ *est une mesure positive* $\mathbb{P}$ *telle que* $\boxed{\mathbb{P}(\Omega) = 1}$.

Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ s'appelle *espace de probabilité*.

Pour des raisons historiques, on a une *terminologie spécifique* :

- une partie mesurable $A \in \mathcal{A}$ s'appelle un *événement* ;
- un point $\omega \in \Omega$ s'appelle une *observation*, ou une expérience ;
  Quand $\omega \in A$, on dit que ω est une *réalisation* de l'événement A, ou que l'événement A *est réalisé* par ω ;
- $\emptyset$ est l'*événement impossible* et Ω est l'*événement certain*.
- Si $A, B \in \mathcal{A}$, et $\boxed{A \subseteq B}$, on dit que *l'événement A implique l'événement B*.
  $A \cup B$ est appelé l'événement *A ou B* ;
  $A \cap B$ est appelé l'événement *A et B*.
- Si $A_n \in \mathcal{A}$, $n \geqslant 1$, alors :
  - $\omega \in \bigcup_{n \geqslant 1} A_n$ si l'*un au moins* des événements A_n est réalisé par ω ;
  - $\omega \in \bigcap_{n \geqslant 1} A_n$ si *tous* les événements A_n sont réalisés par ω.
- Si A est un événement on dit que :

- A est <u>*presque impossible*</u> si $\mathbb{P}(A) = 0$;
- A est <u>*presque certain*</u> ou <u>*presque sûr*</u> si $\mathbb{P}(A) = 1$.

- Le terme "presque partout" (pour la mesure $\mathbb{P}$) est traduit par l'expression <u>**presque sûrement**</u>, abrégé souvent en *p.s.*.

1.2 Variables aléatoires.

On utilisera la convention suivante :

Convention. Pour $d = 1$, on convient que $\mathbb{R}^d$ désigne $\overline{\mathbb{R}}$.

Il sera aussi commode de convenir que $\mathbb{R}^d$ peut aussi désigner $\overline{\mathbb{R}}^d$.

Définition 2 *On apppelle <u>variable aléatoire</u> (en abrégé : v.a.) toute <u>application mesurable</u> $X : \Omega \to \mathbb{R}^d$.*

Pour $d = 1$, on dit <u>variable aléatoire réelle</u> (v.a.r.) ; pour $d \geqslant 2$, on dit aussi <u>vecteur aléatoire</u>.

L'usage est de noter les variables aléatoires avec des majuscules.

Terminologie. Soit X une *v.a.r.*.

- si $X \in \mathscr{L}^p(\Omega, \mathcal{A}, \mathbb{P})$, $1 \leq p < \infty$, on dit que X a un *moment d'ordre p*. Le moment absolu d'ordre p est alors $\int_\Omega |X|^p \, d\mathbb{P}$, et lorsque p est entier ou X est positive, le moment d'ordre p est $\int_\Omega X^p \, d\mathbb{P}$.
  Rappelons que, puisque $\mathbb{P}(\Omega) = 1$, on a :
  $$\mathscr{L}^p(\Omega, \mathcal{A}, \mathbb{P}) \subseteq \mathscr{L}^1(\Omega, \mathcal{A}, \mathbb{P}).$$

- Si $X \in \mathscr{L}^1(\Omega, \mathcal{A}, \mathbb{P})$, l'intégrale de X est appelée l'<u>*espérance*</u> de X (ou aussi <u>*moyenne*</u> de X) ; on note :
  $$\boxed{\mathbb{E}(X) = \int_\Omega X \, d\mathbb{P} = \int_\Omega X(\omega) \, d\mathbb{P}(\omega)}\,.$$
  C'est le moment d'ordre 1. Si $X \in \mathscr{L}^p(\Omega, \mathcal{A}, \mathbb{P})$, le moment absolu d'ordre p est donc $\exp(|X|^p)$; lorsqu'il est défini, le moment d'ordre p est $\mathbb{E}(X^p)$.

- Lorsque $X \in \mathscr{L}^2(\Omega, \mathcal{A}, \mathbb{P})$, on définit la <u>*variance*</u> de X par :
  $$\boxed{\operatorname{Var}(X) = \mathbb{E}\left[\big(X - \mathbb{E}(X)\big)^2\right]}\,,$$
  où $X - \mathbb{E}(X)$ est mis pour $X - \mathbb{E}(X)\,\mathbb{1}$. C'est un nombre $\geqslant 0$. On le calcule par la formule suivante.

Proposition 3 $\boxed{\operatorname{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2\,.}$

Remarque. Si $\mathbb{E}(X) = 0$, alors $\mathrm{Var}(X) = \mathbb{E}(X^2)$: c'est le moment d'ordre 2.

Preuve. On a, en développant et en utilisant la linéarité de l'espérance (= l'intégrale) :

$$\begin{aligned}\mathbb{E}\left[\left(X - \mathbb{E}(X)\right)^2\right] &= \mathbb{E}\left(X^2 - 2\mathbb{E}(X)X - [\mathbb{E}(X)]^2 \mathbb{1}\right) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) - [\mathbb{E}(X)]^2\mathbb{E}(\mathbb{1}) \\ &= \mathbb{E}(X^2) - [\mathbb{E}(X)]^2,\end{aligned}$$

car $\mathbb{E}(\mathbb{1}) = \int_\Omega \mathbb{1}\, d\mathbb{P} = \mathbb{P}(\Omega) = 1$. □

Remarque. Pour tout $a \in_R$: $\boxed{\mathrm{Var}(aX) = a^2\mathrm{Var}(X)}$.

- L'<u>*écart-type*</u> de X est :

$$\boxed{\sigma_X = \sqrt{\mathrm{Var}(X)}} = \|X - \mathbb{E}(X)\,\mathbb{1}\|_2.$$

1.3 Loi d'une variable aléatoire.

L'important lorsque l'on a une variable aléatoire est de savoir *comment se répartissent les valeurs* $X(\omega)$. Cette répartition est formalisée par la définition <u>fondamentale</u> suivante :

Définition 4 *On appelle <u>loi de la variable aléatoire</u> X (ou aussi <u>distribution</u> de X) la <u>mesure-image</u> de $\mathbb{P}$ par X.*

On la note $\mathbb{P}_X$. On dit que X <u>suit la loi</u> $\mathbb{P}_X$, et l'on note $\boxed{X \sim \mathbb{P}_X}$.

Par définition, on a donc :

$$\forall B \in \mathscr{B}or(\mathbb{R}^d) \qquad \boxed{\boxed{\mathbb{P}_X(B) = \mathbb{P}(X \in B)}},$$

où $\mathbb{P}(X \in B)$ est mis pour :

$$\boxed{\mathbb{P}(\{\omega \in \Omega\,;\ X(\omega) \in B\} = \mathbb{P}[X^{-1}(B)]}.$$

C'est une mesure positive sur $\mathbb{R}^d$, et :

$$\mathbb{P}(\mathbb{R}^d) = \mathbb{P}(X \in \mathbb{R}^d) = \mathbb{P}(\Omega) = 1\,;$$

c'est donc une <u>**probabilité**</u> sur $\mathbb{R}^d$.

On a vu comment intégrer par rapport à une mesure-image :

Théorème 5 (Théorème de transfert) *Soit $X\colon \Omega \to \mathbb{R}^d$ une v.a.. Pour toute fonction $\varphi\colon \mathbb{R}^d \to \overline{\mathbb{R}}$ borélienne positive, ou qui est $\mathbb{P}_X$-intégrable, on a :*

$$\boxed{\int_{\mathbb{R}^d} \varphi(x)\, d\mathbb{P}_X(x) = \int_{\Omega} \varphi[X(\omega)]\, d\mathbb{P}\,(\omega)}\,.$$

Cette formule est d'une importance capitale car elle permet de calculer les intégrales des fonctions de X en ne connaissant que la loi de X ; on peut donc très souvent "oublier" l'espace abstrait $(\Omega, \mathcal{A}, \mathbb{P})$ et l'on ne travaille plus que sur $\mathbb{R}^d$ muni de la loi de probabilité (mesure positive) $\mathbb{P}_X$.

Notons que, lorsque $\varphi \circ X \in \mathscr{L}^1(\mathbb{P})$, on a noté :

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

On note $\boxed{\mathbb{E}\,[\varphi(X)] = \mathbb{E}\,(\varphi \circ X)}$; on a donc :

$$\boxed{\boxed{\mathbb{E}\,[\varphi(X)] = \int_{\mathbb{R}^d} \varphi(x)\, d\mathbb{P}\,(x)}}\,.$$

En particulier, si X est une *v.a.r.*, X a un moment d'ordre p si et seulement si :

$$\int_{\mathbb{R}} |x|^p\, d\mathbb{P}\,(x) < +\infty\,,$$

et :

- pour $p = 1$: $\boxed{\mathbb{E}\,(X) = \displaystyle\int_{\mathbb{R}} x\, d\mathbb{P}\,(x)}$
- pour $p = 2$: $\boxed{\mathbb{E}\,(X^2) = \displaystyle\int_{\mathbb{R}} x^2\, d\mathbb{P}\,(x)}$.

Remarque. On peut montrer que pour toute mesure de probabilité Q sur $\mathbb{R}$ (ou sur $\mathbb{R}^d$), il existe une *v.a.r.* $X\colon \Omega \to \mathbb{R}$ telle que $\mathbb{P}_X = Q$.

Parmi toutes les lois de probabilité sur $\mathbb{R}$, on en distingue *deux classes particulières* :

① Celles qui sont **discrètes** : elles s'écrivent :

$$\boxed{\mathbb{P}_X = \sum_{n=1}^{\infty} c_n \delta_{a_n}}\,,$$

où δ_{a_n} est la mesure de Dirac en $a_n \in \mathbb{R}$ et $c_n \geqslant 0$, $\sum_{n=1}^{\infty} c_n = 1$.

On dit alors que la <u>*v.a.r.*</u> est <u>discrète</u>.

Cela signifie que X prend (presque sûrement) uniquement les valeurs a_n (et seulement celles pour lesquelles $c_n \neq 0$), et que :

$$\boxed{\boxed{\mathbb{P}\left(X = a_n\right) = c_n}}, \quad \forall n \geqslant 1.$$

Lorsque X a une telle loi, son espérance (lorsqu'elle existe) s'écrit :

$$\boxed{\boxed{\mathbb{E}\left(X\right) = \sum_{n=1}^{\infty} c_n a_n = \sum_{n=1}^{\infty} a_n \, \mathbb{P}\left(X = a_n\right)}},$$

et si $X \in \mathscr{L}^2(\mathbb{P})$:

$$\boxed{\mathbb{E}\left(X^2\right) = \sum_{n=1}^{\infty} a_n^2 \, \mathbb{P}\left(X = a_n\right)}.$$

② Les *v.a.r.* dont la loi **possède une densité** (par rapport à la mesure de Lebesgue) : il existe une fonction $f_X : \mathbb{R} \to \mathbb{R}_+$ mesurable positive telle que :

$$\boxed{\mathbb{P}_X = f_X . \lambda},$$

c'est-'a-dire que pour tout borélien $B \subseteq \mathbb{R}$, on a :

$$\boxed{\mathbb{P}\left(X \in B\right) = \mathbb{P}_X(B) = \int_B f_X(x) \, dx}.$$

On dit que la <u>*v.a.r.*</u> X est <u>à densité</u>.

Lorsque $X \in \mathscr{L}^1(\mathbb{P})$, on a :

$$\boxed{\boxed{\mathbb{E}\left(X\right) = \int_{\mathbb{R}} x f_X(x) \, dx}}$$

et plus généralement si $\varphi : \mathbb{R} \to \mathbb{R}$ est borélienne positive ou si $\varphi \circ X \in \mathscr{L}^1(\mathbb{P})$:

$$\boxed{\boxed{\mathbb{E}\left[\varphi(X)\right] = \int_{\mathbb{R}} \varphi(x) f_X(x) \, dx}}.$$

1.4 Exemples usuels de lois.

a) **Lois discrètes.**

• **La loi de Bernoulli**, dite aussi " *loi du pile ou face*" :

$$\mathbb{P}_X = (1-p)\,\delta_0 + p\,\delta_1\,, \quad 0 < 1.$$

C'est la loi d'une *v.a.r.* prenant *p.s.* les valeurs 0 et 1 :

$$\boxed{\mathbb{P}\,(X=1) = p \quad ; \quad \mathbb{P}\,(X=0) = 1-p}\,.$$

Exemple. On lance une pièce de monnaie ; on définit la *v.a.r.* X par $X = 1$ si on a "*face*" et $X = 0$ si on a "*pile*". Si la pièce n'est pas truquée, on a : $\mathbb{P}\,(X=0) = \mathbb{P}\,(X=1) = 1/2$.

Pour une telle *v.a.r.* X, on a $\boxed{\mathbb{E}\,(X) = p}$; en effet :

$$\mathbb{E}\,(X) = (1-p)\times 0 + p\times 1 = p.$$

Calculons la variance :

$$\mathbb{E}\,(X^2) = (1-p)\times 0^2 + p\times 1^2 = p\,;$$

donc

$$\mathrm{Var}\,(X) = \mathbb{E}\,(X^2) - [\mathbb{E}\,(X)]^2 = p - p^2 = p(1-p).$$

• **La loi binômiale $\mathscr{B}(n,p)$.**

On se donne un entier $n \geqslant 1$ et un nombre réel p tel que $0 < p < 1$. On dit que X suit la loi binômiale $\mathscr{B}(n,p)$ de paramètres n et p si X prend *p.s.* les valeurs $0, 1, 2, \ldots, n$ et que :

$$\boxed{\mathbb{P}\,(X=k) = C_n^k\, p^k\,(1-p)^{n-k}} \quad 0 \leqslant k \leqslant n.$$

Notons que l'on a bien $\sum_{k=0}^n \mathbb{P}\,(X=k) = 1$, par la formule du binôme.

Exemple. Dans un jeu de 32 cartes, on tire n cartes. On définit X par :

$$X = k \text{ si on a tiré } k \text{ cœurs.}$$

La probabilité de tirer un œur étant $1/4$, on a :

$$\mathbb{P}\,(X=k) = C_n^k\,\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^{n-k}.$$

On a $\boxed{\mathbb{E}\,(X) = np}$; en effet,

$$\mathbb{E}\,(X) = \sum_{k=0}^n k\; C_n^k\; p^k\,(1-p)^{n-k}\,,$$

et comme, pour $k \geqslant 1$, on a :

$$k\ C_n^k = n\ C_{n-1}^{k-1}\,,$$

on obtient :

$$\begin{aligned}\mathbb{E}\,(X) &= \sum_{k=1}^{n} n\ C_{n-1}^{k-1}\ p\,p^{k-1}\,(1-p)^{n-1-(k-1)} \\ &= np\sum_{l=0}^{n-1} C_{n-1}^{l}\ p^{l}\,(1-p)^{n-1-l} = np.\end{aligned}$$

De même, on peut montrer que $\mathrm{Var}\,(X) = np(1-p)$, mais on retrouvera ces résultats plus tard, sans calculs.

• **La loi de Poisson.**

Une *v.a.r.* X suit la loi de Poisson $\mathscr{P}(\lambda)$ de paramètre $\lambda > 0$ (c'est la notation habituelle, mais il faut faire attention à ne pas confondre ce nombre positif λ avec la mesure de Lebesgue !!!) si X prend *p.s.* les valeurs $0, 1, 2, 3, \ldots$ et que :

$$\boxed{\mathbb{P}\,(X = k) = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}} \qquad k \in \mathbb{N}.$$

On a bien $\sum_{k=0}^{\infty} \mathbb{P}\,(X = k) = 1$.

Cette loi apparaît comme une approximation, plus facile à utiliser, de la *loi binômiale* $\mathscr{B}(n,p)$, quand n est grand (et p petit). Plus précisément :

Proposition 6 *Soit $\lambda > 0$. Si la v.a.r. suit la loi $\mathscr{B}\big(n, \frac{\lambda}{n}\big)$, alors, pour tout $k \in \mathbb{N}$:*

$$\mathbb{P}\,(X_n = k) \underset{n\to+\infty}{\longrightarrow} \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}\cdot$$

Preuve. Pour $n \geqslant k$:

$$\begin{aligned}\mathbb{P}\,(X_n = k) &= \frac{n(n-1)\cdots(n-k+1)}{k!}\Big(\frac{\lambda}{n}\Big)^k\Big(1-\frac{\lambda}{n}\Big)^{n-k} \\ &= \frac{1}{k!}\Big[1\times\Big(1-\frac{1}{n}\Big)\times\cdots\times\Big(1-\frac{k-1}{n}\Big)\Big]\,\lambda^k\Big(1-\frac{\lambda}{n}\Big)^{n-k} \\ &\underset{n\to+\infty}{\longrightarrow} \frac{1}{k!}\times 1\times\lambda^k\times\mathrm{e}^{-\lambda}\,.\end{aligned}$$

□

On a $\boxed{\mathbb{E}\,(X) = \lambda}$:

$$\mathbb{E}\,(X) = \sum_{k=0}^{\infty} k\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k=1}^{\infty}\mathrm{e}^{-\lambda}\frac{\lambda^{k-1}.\,\lambda}{(k-1)!} = \lambda\,,$$

et $\boxed{\mathrm{Var}\,(X) = \lambda}$; en effet, en écrivant $k^2 = k(k-1) + k$, on obtient :

$$\begin{aligned}
\mathbb{E}\,(X^2) &= \sum_{k=0}^{\infty} k^2 \, \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} \\
&= \sum_{k=2}^{\infty} \mathrm{e}^{-\lambda} \lambda^2 \frac{\lambda^{k-2}}{(k-2)!} + \sum_{k=1}^{\infty} \mathrm{e}^{-\lambda} \lambda \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda^2 + \lambda\,;
\end{aligned}$$

d'où :

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

b) **Lois à densité.**

• **La loi uniforme** sur un intervalle $[a, b]$.

Sa densité est :

$$\boxed{f_X(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x)}\,.$$

Cela signifie que pour tout borélien $B \subseteq \mathbb{R}$:

$$\mathbb{P}\,(X \in B) = \frac{1}{b-a} \lambda(B \cap [a, b]).$$

En particulier, X prend presque sûrement ses valeurs dans $[a, b]$, et si B est un borélien contenu dans $[a, b]$, alors

$$\mathbb{P}\,(X \in B) = \frac{\lambda(B)}{b-a}\cdot$$

est proportionnelle à la mesure de B : la *v.a.r.* X modélise le fait de "*choisir un point au hasard*" dans $[a, b]$.

Pour une telle *v.a.r.*, on a $\mathbb{P}\,(\mathbb{Q}) = 0$: cela signifie qu'obtenir un rationnel en choisissant au hasard un nombre dans $[a, b]$ est un événement *presque impossible* (mais pas impossible).

On a $\boxed{\mathbb{E}\,(X) = \frac{b+a}{2}}$ (le milieu de $[a, b]$) :

$$\mathbb{E}\,(X) = \int_{\mathbb{R}} x\, f_X(x)\, dx = \frac{1}{b-a} \int_a^b x\, dx = \frac{1}{b-a} \left[\frac{x^2}{2}\right]_a^b = \frac{b+a}{2}\cdot$$

• **La loi exponentielle** $\mathscr{E}(\lambda)$, de paramètre $\lambda > 0$.

(Là aussi, ne pas confondre le paramètre $\lambda > 0$ avec la mesure de Lebesgue !)

Sa densité est :

$$\boxed{f_X(x) = \lambda\, \mathrm{e}^{-\lambda x} \mathbb{1}_{x>0}}$$

(on note habituellement $\mathbb{1}_{x>0} = \mathbb{1}_{]0,+\infty[}(x)$). On a bien $\int_{\mathbb{R}} f_X(x)\,dx = 1$. Pour tout borélien $B \subseteq \mathbb{R}$, on a :

$$\mathbb{P}\,(X \in B) = \int_{B\cap]0,+\infty[} \lambda\,\mathrm{e}^{-\lambda x}\,dx.$$

En particulier, X est <u>*p.s.*. à valeurs positives</u>.

Calculons l'espérance et la variance. On a, dune part :

$$\mathbb{E}\,(X) = \int_0^{+\infty} x\,\lambda\,\mathrm{e}^{-\lambda x}\,dx = \Big[-x\,\mathrm{e}^{-\lambda x}\Big]_0^{+\infty} + \int_0^{+\infty} \mathrm{e}^{-\lambda x}\,dx = \frac{1}{\lambda}\,;$$

et, d'autre part :

$$\mathbb{E}\,(X^2) = \int_0^{+\infty} x^2\,\lambda\,\mathrm{e}^{-\lambda x}\,dx = \Big[-x^2\,\mathrm{e}^{-\lambda x}\Big]_0^{+\infty} + 2\int_0^{+\infty} x\,\mathrm{e}^{-\lambda x}\,dx = \frac{2}{\lambda^2}\,,$$

d'où :

$$\mathrm{Var}\,(X) = \frac{2}{\lambda^2} - \Big(\frac{1}{\lambda}\Big)^2 = \frac{1}{\lambda^2}.$$

• **<u>La loi normale</u>** ou **<u>loi gaussienne</u>** $\mathscr{N}(m,\sigma^2)$ de paramètres m et σ^2 ($m \in \mathbb{R}$, $\sigma > 0$).

Sa densité est :

$$\boxed{\boxed{f_X(x) = \gamma_{m,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}(\frac{x-m}{\sigma})^2}}}.$$

Rappel. On a vu que :

$$\int_{\mathbb{R}} \mathrm{e}^{-u^2}\,du = 2\int_0^{+\infty} \mathrm{e}^{-u^2}\,du = \sqrt{\pi}\,;$$

on a donc, en posant $u = \frac{x-m}{\sigma}$:

$$\int_{\mathbb{R}} \gamma_{m,\sigma}(x)\,dx = 1.$$

On dit que la loi $\mathscr{N}(0,1)$ de paramètres 0 et 1 est la **<u>loi normale</u> <u>centrée réduite</u>** ; sa densité est :

$$\boxed{\gamma(x) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-x^2/2}}.$$

Proposition 7 *Si $X \sim \mathscr{N}(m,\sigma^2)$, alors $X_0 = \frac{X-m}{\sigma} \sim \mathscr{N}(0,1)$.*
Réciproquement, si $X_0 \sim \mathscr{N}(0,1)$, alors $X = \sigma X_0 + m \sim \mathscr{N}(m,\sigma^2)$.

Preuve. Posons $\varphi(x) = \frac{x-m}{\sigma}\cdot$ Pour tout borélien $B \subseteq \mathbb{R}$, on a :

$$\begin{aligned}\mathbb{P}_{\varphi(X)}(B) &= \mathbb{P}\big(\varphi(X) \in B\big) = \mathbb{P}_X\big(\varphi^{-1}(B)\big) \\ &= \int_{\varphi^{-1}(B)} \frac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}(\frac{x-m}{\sigma})^2}\,dx \underset{\xi=\varphi(x)}{=} \int_B \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\xi^2/2}\,d\xi\,;\end{aligned}$$

donc $\varphi(X) = \frac{X-m}{\sigma} \sim \mathscr{N}(0,1)$.

La réciproque se montre de la même façon. $\square$

Proposition 8 *Si $X \sim \mathscr{N}(m,\sigma^2)$, alors :*

$$\boxed{\mathbb{E}\,(X) = m} \quad \textit{et} \quad \boxed{\mathrm{Var}\,(X) = \sigma^2}\,.$$

Preuve. 1) Si $X_0 \sim \mathscr{N}(0,1)$, alors :

$$\mathbb{E}\,(X_0) = \int_{\mathbb{R}} x\,\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-x^2/2}\,dx = 0$$

(car on intègre sur $\mathbb{R}$ une fonction impaire). Alors :

$$\begin{aligned}\mathrm{Var}\,(X_0) = \mathbb{E}\,(X_0^2) &= \int_{\mathbb{R}} x^2\,\frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-x^2/2}\,dx \\ &= \int_{\mathbb{R}} x\big(x\,e^{-x^2/2}\big)\,\frac{dx}{\sqrt{2\pi}} \\ &= \frac{1}{\sqrt{2\pi}}\Big[-x\,\mathrm{e}^{-x^2/2}\Big]_{-\infty}^{+\infty} + \int_{\mathbb{R}} e^{-x^2/2}\,\frac{dx}{\sqrt{2\pi}} = 1.\end{aligned}$$

2) Si $X \sim \mathscr{N}(m,\sigma^2)$, alors $X_0 = \frac{X-m}{\sigma} \sim \mathscr{N}(0,1)$; on a $0 = \mathbb{E}\,(X_0) = \frac{1}{\sigma}\big(\mathbb{E}\,(X) - m\big)$, donc $\mathbb{E}\,(X) = m$, et $1 = \mathrm{Var}\,(X_0) = \frac{1}{\sigma^2}\mathrm{Var}\,(X-m) = \frac{1}{\sigma^2}\mathrm{Var}\,(X)$, d'où $\mathrm{Var}\,(X) = \sigma^2$. $\square$

2 Indépendance.

C'est la notion d'indépendance qui fait de la Théorie des Probabilités autre chose qu'une simple traduction de la Théorie de la mesure.

2.1 Événements indépendants.

Définition 9 *On dit que deux événements A et B sont <u>indépendants</u> (et l'on note $A \perp\!\!\!\perp B$) si*

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

L'exemple suivant est **<u>fondamental</u>**, en ce sens qu'il décrit la situation générique : *on peut toujours se ramener à ce cas*, dans lequel $(\Omega,\mathbb{P})$ est un <u>espace mesuré produit</u>.

Exemple. Soit $\Omega = [0,1] \times [0,1]$, $\mathcal{A}$ sa tribu borélienne et $\mathbb{P}$ la mesure de Lebesgue $\lambda_2 = \lambda \otimes \lambda$.

Soit S_1 et S_2 deux borélien de $[0,1]$ et :

$$A = S_1 \times [0,1]\,, \qquad B = [0,1] \times S_2$$

(le point essentiel est que **A ne dépend que de la première variable** tandis que **B ne dépend que de la deuxième variable**). Alors A et B sont **indépendants** :

$$\begin{aligned} \mathbb{P}\,(A \cap B) &= \lambda_2(A \cap B) \\ &= \boxed{\lambda_2(S_1 \times S_2) = \lambda \otimes \lambda(S_1 \times S_2)} \\ &= \lambda_2(S_1 \times [0,1])\,\lambda_2([0,1] \times S_2) = \mathbb{P}\,(A)\,\mathbb{P}\,(B). \end{aligned}$$

Pour plus de deux événements, la définition est un peu plus compliquée.

Définition 10 *On dit que n événements $A_1, \ldots, A_n$ sont indépendants si :*

$$\mathbb{P}\Big(\bigcap_{j\in J} A_j\Big) = \prod_{j\in J} \mathbb{P}\,(A_j)$$

pour tout ensemble d'indices $J \subseteq \{1, 2, \ldots, n\}$.

C'est plus fort que de dire qu'ils sont deux-à-deux indépendants (qui correspond au cas où J est un ensemble de deux indices). De plus, $A_{k_1}, \ldots, A_{k_r}$ sont indépendants quel que soit le choix des entiers distincts $k_1, \ldots, k_r$ dans $\{1, \ldots, n\}$.

2.2 Variables aléatoires indépendantes.

Définition 11 *On dit que n v.a.r. $X_1, \ldots, X_n$ sont indépendantes si quels que soit les boréliens $B_1, \ldots, B_n$ de $\overline{\mathbb{R}}$, les événements :*

$$(X_1 \in B_1), \ldots, (X_n \in B_n)$$

sont indépendants.

Si X et Y sont indépendantes, on note $X \perp\!\!\!\perp Y$.

Par définition, $X_1, \ldots, X_n$ sont indépendantes si et seulement si pour tous boréliens $B_1, \ldots, B_n$, on a, pour toute partie $J \subseteq \{1, 2, \ldots, n\}$:

$$\mathbb{P}\Big(\bigcap_{j\in J}(X_j \in B_j)\Big) = \prod_{j\in J} \mathbb{P}\,(X_j \in B_j).$$

Mais cela équivaut à dire que, pour tous boréliens $B_1, \ldots, B_n$, on a :

$$(I) \qquad \mathbb{P}\Big(\bigcap_{j=1}^{n}(X_j \in B_j)\Big) = \prod_{j=1}^{n} \mathbb{P}\,(X_j \in B_j).$$

En effet, l'indépendance entraîne (I) : il suffit de prendre $J = \{1, 2, \ldots, n\}$; mais inversement, si l'on a (I), alors, pour toute partie J de $\{1, 2, \ldots, n\}$, on a, en posant $B'_j = B_j$ si $j \in J$ et $B'_j = \overline{\mathbb{R}}$ pour $j \in \{1, 2, \ldots, n\} \setminus J$:

$$\begin{aligned}\mathbb{P}\Big(\bigcap_{j\in J}(X_j \in B_j)\Big) &= \mathbb{P}\Big(\bigcap_{j=1}^{n}(X_j \in B'_j)\Big)\\ &= \prod_{j=1}^{n}\mathbb{P}(X_j \in B'_j) = \prod_{j\in J}\mathbb{P}(X_j \in B_j),\end{aligned}$$

car, pour $j \notin J$, $\mathbb{P}(X_j \in B'_j) = \mathbb{P}(X_j \in \overline{\mathbb{R}}) = 1$.

Mais, par définition, d'une part :

$$\mathbb{P}(X_j \in B_j) = \mathbb{P}_{X_j}(B_j)\,, \quad 1 \leqslant j \leqslant n,$$

et, d'autre part :

$$\begin{aligned}\mathbb{P}\Big(\bigcap_{j=1}^{n}(X_j \in B_j)\Big) &= \mathbb{P}\big((X_1, \ldots, X_n) \in B_1 \times \cdots \times B_n\big)\\ &= \mathbb{P}_{(X_1,\ldots,X_n)}(B_1 \times \cdots \times B_n);\end{aligned}$$

donc (I) équivaut à :

$$\boxed{\mathbb{P}_{(X_1,\ldots,X_n)}(B_1 \times \cdots \times B_n) = \prod_{j=1}^{n}\mathbb{P}(X_j \in B_j)}\,.$$

Mais on sait que la **mesure-produit** $\mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$ est la **seule** mesure donnant la valeur :

$$\prod_{j=1}^{n}\mathbb{P}(X_j \in B_j)$$

pour $B_1 \times \cdots \times B_n$.

On a donc obtenu :

Théorème 12 *Les v.a.r. $X_1, \ldots, X_n$ sont* **indépendantes** *si et seulement si la loi $\mathbb{P}_{(X_1,\ldots,X_n)}$ du vecteur aléatoire $(X_1, \ldots, X_n)$ est égale au* **<u>produit</u>** *$\mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$ des lois $\mathbb{P}_{X_1}, \ldots, \mathbb{P}_{X_n}$ des lois de $X_1, \ldots, X_n$.*

En particulier, deux *v.a.r.* X et Y sont indépendantes si et seulement si $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$.

Pour savoir si des *v.a.r.* $X_1, \ldots, X_n$ sont indépendantes, on a **besoin de connaître la loi du vecteur aléatoire** $(X_1, \ldots, X_n)$. Cette loi s'appelle la *loi conjointe* des *v.a.r.* $X_1, \ldots, X_n$. Les lois $\mathbb{P}_{X_1}, \ldots, \mathbb{P}_{X_n}$ s'appellent les *lois marginales* du vecteur aléatoire $(X_1, \ldots, X_n)$.

Les lois marginales **ne suffisent pas** pour connaître la loi conjointe.

Par contre, à partir de la loi conjointe, on peut trouver les lois marginales. Par exemple :

$$\begin{aligned}\mathbb{P}_{X_1} &= \mathbb{P}(X_1 \in B_1) = \mathbb{P}(X_1 \in B_1, X_j \in \overline{\mathbb{R}}, j \geqslant 2) \\ &= \mathbb{P}_{(X_1,\ldots,X_n)}(B_1 \times \overline{\mathbb{R}} \times \cdots \times \overline{\mathbb{R}}).\end{aligned}$$

Exemple. supposons que X et Y sont deux *v.a.r.* discrètes et que l'on connaisse :

$$\mathbb{P}_{(X,Y)}(\{j,k\}) = \mathbb{P}(X = j, Y = k) = a_{j,k}\,, \quad 1 \leqslant j \leqslant r, 1 \leqslant k \leqslant s.$$

Alors :

$$\begin{aligned}\mathbb{P}_X(\{j\}) &= \mathbb{P}(X = j) = \mathbb{P}(X = j, Y = 1, 2, \ldots, s) \\ &= \sum_{k=1}^{s} \mathbb{P}(X = j, Y = k) \quad \text{(les événements sont deux-à-deux disjoints)} \\ &= \sum_{k=1}^{s} a_{j,k} = b_j,\end{aligned}$$

et

$$\mathbb{P}_Y(\{k\}) = \mathbb{P}(Y = k) = \sum_{j=1}^{r} \mathbb{P}(X = j, Y = k) = \sum_{j=1}^{r} a_{j,k} = c_k.$$

b_1	b_2	b_3	b_4	b_5	b_6	X / Y
$a_{1,1}$	$a_{2,1}$	$a_{3,1}$	$a_{4,1}$	$a_{5,1}$	$a_{6,1}$	c_1
$a_{1,2}$	$a_{2,2}$	$a_{3,2}$	$a_{4,2}$	$a_{5,2}$	$a_{6,2}$	c_2
$a_{1,3}$	$a_{2,3}$	$a_{3,3}$	$a_{4,3}$	$a_{5,3}$	$a_{6,3}$	c_3
$a_{1,4}$	$a_{2,4}$	$a_{3,4}$	$a_{4,4}$	$a_{5,4}$	$a_{6,4}$	c_4

Les lois de X de de Y se lisent *dans les marges*, ce qui explique le nom de *lois marginales*.

Proposition 13 *Si la loi du couple (X,Y) possède une densité $f_{(X,Y)}$, par rapport à la mesure de Lebesgue λ_2, alors les lois marginales ont des densités données par :*

$$\begin{cases} f_X(x) &= \displaystyle\int_{\mathbb{R}} f_{(X,Y)}(x,y)\,dy \\ f_Y(y) &= \displaystyle\int_{\mathbb{R}} f_{(X,Y)}(x,y)\,dx. \end{cases}$$

De plus, X et Y sont inépendantes si et seulement si :

$$f_{(X,Y)}(x,y) = f_X(x)\,f_Y(y)$$

pour presque tous $(x,y) \in \mathbb{R}^2$.

Preuve. 1) Pour tout borélien B, on a :

$$\begin{aligned}\mathbb{P}_X(B) = \mathbb{P}_{(X,Y)}(B \times \mathbb{R}) &= \int_{B\times\mathbb{R}} f_{(X,Y)}(x,y)\,dxdy \\ &\underset{\text{Fubini-Tonelli}}{=} \int_B \left[\int_{\mathbb{R}} f_{(X,Y)}(x,y)\,dy\right] dx = \int_B f_X(x)\,dx\,;\end{aligned}$$

donc $\mathbb{P}_X = f_X.\lambda$.

De même $\mathbb{P}_Y = f_Y.\lambda$.

2) Comme :

$$\begin{aligned}(\mathbb{P}_X \otimes \mathbb{P}_Y)(B_1 \times B_2) = \mathbb{P}_X(B_1)\,\mathbb{P}_Y(B_2) &= \Big(\int_{B_1} f_X(x)\,dx\Big)\Big(\int_{B_2} f_Y(y)\,dy\Big) \\ &\underset{\text{Fubini-Tonelli}}{=} \int_{B_1\times B_2} f_X(x)\,f_Y(y)\,dxdy,\end{aligned}$$

la mesure-produit $\mathbb{P}_X \otimes \mathbb{P}_Y$ a pour densité :

$$(x,y) \longmapsto f_X(x)\,f_Y(y).$$

Mais $X \perp\!\!\!\perp Y$ si et seulement si $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$; par conséquent, $X \perp\!\!\!\perp Y$ si et seulement si $f_{(X,Y)}(x,y) = f_X(x)\,f_Y(y)$ pour presque tous $(x,y) \in \mathbb{R}^2$. □

Exemple.

2.3 Propriétés des *v.a.r.* indépendantes.

Théorème 14 *Soit $X_1, \ldots, X_n \colon \Omega \to \mathbb{R}$ des v.a.r. indépendantes. Alors, quelles que soient les fonctions mesurables $f_1, \ldots, f_n \colon \mathbb{R} \to \mathbb{R}$, les v.a.r.*

$$f_1(X_1), \ldots, f_n(X_n)$$

sont indépendantes.

Preuve. Pour tous boréliens $B_1, \ldots, B_n \subseteq \mathbb{R}$, les parties

$$f_1^{-1}(B_1), \ldots, f_n^{-1}(B_n)$$

sont boréliennes ; donc les événements

$$\big[f_1(X_1)\big]^{-1}(B_1) = X_1^{-1}[f_1^{-1}(B_1)], \ldots, \big[f_n(X_n)\big]^{-1}(B_n) = X_n^{-1}[f_n^{-1}(B_n)]$$

sont indépendants. □

Exemple. Si X et Y sont indépendantes, alors X^2 est indépendante de e^Y.

On peut généraliser un peu.

Théorème 15 *Soit* $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ *n v.a.r. indépendantes et* $1 < k < n$. *Soit* $f : \mathbb{R}^k \to \mathbb{R}$ *et* $g : \mathbb{R}^{n-k} \to \mathbb{R}$ *deux applications mesurables. Alors :*

$$Y = f(X_1, \ldots, X_k) \quad \textit{et} \quad Z = g(X_{k+1}, \ldots, X_n)$$

sont indépendantes.

Preuve. Soit $B_1 \in \mathscr{B}or(\mathbb{R}^k)$ et $B_2 \in \mathscr{B}or(\mathbb{R}^{n-k})$. Alors $B_1 \times B_2 \in \mathscr{B}or(\mathbb{R}^n)$, et :

$$\begin{aligned}
\mathbb{P}_{(Y,Z)}(B_1 \times B_2) &= \mathbb{P}\big((X_1, \ldots, X_k) \in f^{-1}(B_1), (X_{k+1}, \ldots, X_n) \in g^{-1}(B_2)\big) \\
&= \mathbb{P}\big((X_1, \ldots, X_k, X_{k+1}, \ldots, X_n) \in f^{-1}(B_1) \times g^{-1}(B_2)\big) \\
&= \mathbb{P}_{(X_1, \ldots, X_n)}\big(f^{-1}(B_1) \times g^{-1}(B_2)\big) \\
&= \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}\big(f^{-1}(B_1) \times g^{-1}(B_2)\big) \\
&= \big[(\mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_k}) \otimes (\mathbb{P}_{X_{k+1}} \otimes \cdots \otimes \mathbb{P}_{X_n})\big]\big(f^{-1}(B_1) \times g^{-1}(B_2)\big) \\
&= (\mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_k})\big(f^{-1}(B_1)\big)\,(\mathbb{P}_{X_{k+1}} \otimes \cdots \otimes \mathbb{P}_{X_n})\big(g^{-1}(B_2)\big)
\end{aligned}$$

(on a utilisé le fait que si $X_1, \ldots, X_n$ sont indépendantes, alors $X_1, \ldots, X_k$ sont indépendantes et $X_{k+1}, \ldots, X_n$ sont indépendantes

$$\begin{aligned}
&= \mathbb{P}\big(f(X_1, \ldots, X_k) \in B_1)\big)\,\mathbb{P}\big(g(X_{k+1}, \ldots, X_n) \in B_2)\big) \\
&= \mathbb{P}_Y(B_1)\,\mathbb{P}_Z(B_2).
\end{aligned}$$

□

Exemple. Si X_1, X_2, X_3, X_4 sont indépendantes, alors $Y = X_1 + X_3$ est indépendante de $Z = X_2X_4$ (noter que l'indépendance ne dépend pas de l'ordre dans lequel on écrit les *v.a.r.* ; on applique donc le thórème en utilisant l'indépendance de X_1, X_3, X_2, X_4).

Théorème 16 *Si* $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ *sont des v.a.r.* *indépendantes* *et intégrables, alors le produit* $X_1X_2 \cdots X_n$ *est intégrable et :*

$$\boxed{\mathbb{E}(X_1X_2 \cdots X_n) = \mathbb{E}(X_1)\,\mathbb{E}(X_2) \cdots \mathbb{E}(X_n)}\,.$$

Preuve. Comme $X_1 \perp\!\!\!\perp (X_2 \cdots X_n)$, il suffit de le faire pour $n = 2$, puis de raisonner par récurrence.

Considérons les deux *v.a.r.* $X, Y : \Omega \to \mathbb{R}$ indépendantes et intégrables ; on a :

$$\begin{aligned}
\mathbb{E}(|XY|) &= \int_{\mathbb{R}^2} |xy|\, d\mathbb{P}_{(X,Y)}(x,y) \underset{\perp\!\!\!\perp}{=} \int_{\mathbb{R}^2} |xy|\, d(\mathbb{P}_X \otimes \mathbb{P}_Y)(x,y) \\
&\underset{\text{Fubini-Tonelli}}{=} \left(\int_{\mathbb{R}} |x|\, d\mathbb{P}_X(x)\right)\left(\int_{\mathbb{R}} |y|\, d\mathbb{P}_Y(y)\right) < +\infty\,;
\end{aligned}$$

donc $XY \in \mathscr{L}^1(\Omega, \mathbb{P})$.

Maintenant :

$$\begin{aligned}\mathbb{E}(XY) &= \int_{\mathbb{R}^2} xy\, d\mathbb{P}_{(X,Y)}(x,y) \underset{\perp\!\!\!\perp}{=} \int_{\mathbb{R}^2} xy\, d(\mathbb{P}_X \otimes \mathbb{P}_Y)(x,y) \\ &\underset{\text{Fubini}}{=} \left(\int_{\mathbb{R}} x\, d\mathbb{P}_X(x)\right)\left(\int_{\mathbb{R}} y\, d\mathbb{P}_Y(y)\right) \\ &= \mathbb{E}(X)\,\mathbb{E}(Y)\end{aligned}$$

(notons que l'on a pu utiliser le Théorème de Fubini car on vient de voir que $\int_{\mathbb{R}^2} |xy|\, d\mathbb{P}_{(X,Y)}(x,y) < +\infty$). □

2.4 Somme de *v.a.r.* indépendantes.

Théorème 17 *Si $X_1, \dots, X_n \colon \Omega \to \mathbb{R}$ sont des v.a.r. <u>indépendantes</u>, de carré intégrable, alors :*

$$\boxed{\operatorname{Var}(X_1 + \cdots + X_n) = \operatorname{Var}(X_1) + \cdots + \operatorname{Var}(X_n)}\,.$$

Preuve. Comme $X_1 \perp\!\!\!\perp (X_2 + \cdots + X_n)$,, il suffit de montrer que pour deux *v.a.r.* indépendantes $X, Y \in \mathscr{L}^2(\Omega, \mathbb{P})$, on a :

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)\,,$$

puis de raisonner par récurrence.

Mais :

$$\begin{aligned}\mathbb{E}\big((X+Y)^2\big) &= \mathbb{E}(X^2 + 2XY + Y^2) = \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) \\ &\underset{\perp\!\!\!\perp}{=} \mathbb{E}(X^2) + 2\mathbb{E}(X)\,\mathbb{E}(Y) + \mathbb{E}(Y^2)\,;\end{aligned}$$

donc :

$$\begin{aligned}\operatorname{Var}(X+Y) &= \mathbb{E}\big((X+Y)^2\big) - \big[\mathbb{E}(X+Y)\big]^2 \\ &= \mathbb{E}(X^2) + 2\mathbb{E}(X)\,\mathbb{E}(Y) + \mathbb{E}(Y^2) - \big[\mathbb{E}(X) + \mathbb{E}(Y)\big]^2 \\ &= \mathbb{E}(X^2) + 2\mathbb{E}(X)\,\mathbb{E}(Y) + \mathbb{E}(Y^2) - \big[\big(\mathbb{E}(X)\big)^2 \\ &\qquad\qquad + 2\mathbb{E}(X)\,\mathbb{E}(Y) + \big(\mathbb{E}(Y)\big)^2\big] \\ &= \operatorname{Var}(X) + \operatorname{Var}(Y).\end{aligned}$$

□

Pour voir comment appliquer cela, nous allons prouver :

Théorème 18 *Soit $X_1, \dots, X_n$ n v.a.r. <u>indépendantes</u> suivant la loi de Bernoulli de paramètre p :*

$$\mathbb{P}(X_k = 1) = p \quad ; \quad \mathbb{P}(X_k = 0) = 1 - p.$$

Alors $S_n = X_1 + \cdots + X_n$ suit la loi binômiale $\mathscr{B}(n,p)$ de paramètres n et p.

Preuve. Il est clair que les valeurs prises par S_n sont $0, 1, 2, \ldots, n$.

Pour $0 \leqslant k \leqslant n$, $S_n = k$ si et seulement si k des n *v.a.r.* $X_1, \ldots, X_n$ prennent la valeur 1 et les autres la valeur 0. Si $J \subseteq \{1, \ldots, n\}$ et $\operatorname{card}(J) = k$, on a :

$$\begin{aligned}
&\mathbb{P}\left(X_j = 1, \forall j \in J \text{ et } X_j = 0, \forall j \notin J\right) \\
&\qquad \underset{\perp\!\!\!\perp}{=} \prod_{j \in J} \mathbb{P}\left(X_j = 1\right) \times \prod_{j \notin J} \mathbb{P}\left(X_j = 0\right) = p^k (1-p)^{n-k}.
\end{aligned}$$

Comme il y a C_n^k choix possibles de k *v.a.r.* parmi les n *v.a.r.* $X_1, \ldots, X_n$, on a :

$$\mathbb{P}\left(S_n = k\right) = C_n^k \, p^k (1-p)^{n-k}. \qquad \square$$

Corollaire 19 *Si $Y \sim \mathscr{B}(n,p)$, alors $\mathbb{E}(Y) = np$ et $\operatorname{Var}(Y) = np\,(1-p)$.*

Preuve. Y a la même loi que S_n ; donc :

$$\mathbb{E}(Y) = \mathbb{E}(S_n) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) = np$$

et :

$$\operatorname{Var}(Y) = \operatorname{Var}(S_n).$$

Mais, grâce à l'indépendance de $X_1, \ldots, X_n$:

$$\operatorname{Var}(S_n) = \operatorname{Var}(X_1) + \cdots + \operatorname{Var}(X_n) = np\,(1-p),$$

puisque l'on a vu que $\operatorname{Var}(X_j) = p\,(1-p)$ pour tout j. $\square$

Théorème 20 *Soit X et $Y : \Omega \to \mathbb{R}$ deux v.a.r. ayant une densité $f_X = f$ et $g_Y = g$.*

Si X et Y sont <u>indépendantes</u>, alors la loi conjointe du couple $\mathbb{P}_{(X,Y)}$ possède la densité :

$$f \otimes g : (x, y) \longmapsto f(x)\, g(y)$$

et la v.a.r. $S = X + Y$ possède la densité $f * g$ définie, pour presque tout $x \in \mathbb{R}$, par :

$$(f * g)(x) = \int_{\mathbb{R}} f(x-t)\, g(t)\, dt.$$

Preuve. 1) Par indépendance, pour tout borélien $B \in \mathscr{B}or(\mathbb{R}^2)$, on a :

$$\begin{aligned}
\mathbb{P}_{(X,Y)}(B) = (\mathbb{P}_X \otimes \mathbb{P}_Y)(B) &= \int_{\mathbb{R}^2} \mathbb{I}_B(x,y)\, d(\mathbb{P}_X \otimes \mathbb{P}_Y)(x,y) \\
&\underset{\text{Fubini-Tonelli}}{=} \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \mathbb{I}_B(x,y)\, d\mathbb{P}_X(x) \right] d\mathbb{P}_Y(y) \\
&= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \mathbb{I}_B(x,y)\, f(x)\, dx \right] g(y)\, dy \\
&\underset{\text{Fubini-Tonelli}}{=} \int_{\mathbb{R}^2} \mathbb{I}_B(x,y)\, f(x)\, g(y)\, dxdy.
\end{aligned}$$

2) Soit $h\colon \mathbb{R} \to \mathbb{R}_+$ une fonction mesurable positive ; on a :

$$\begin{aligned}
\mathbb{E}\,[h(S)] &= \int_{\mathbb{R}^2} h(x+y)\, d\mathbb{P}_{(X,Y)}(x,y) \\
&\underset{1)}{=} \int_{\mathbb{R}^2} h(x+y)\, f(x)\, g(y)\, dxdy \\
&\underset{\text{Fubini-Tonelli}}{=} \int_{\mathbb{R}} \left[\int_{\mathbb{R}} h(x+y)\, f(x)\, dx \right] g(y)\, dy \\
&\underset{u=x+y}{=} \int_{\mathbb{R}} \left[\int_{\mathbb{R}} h(u)\, f(u-y)\, du \right] g(y)\, dy \\
&\underset{\text{Fubini-Tonelli}}{=} \int_{\mathbb{R}} h(u) \left[\int_{\mathbb{R}} f(u-y)\, g(y)\, dy \right] du \\
&= \int_{\mathbb{R}} h(u)\, (f * g)(u)\, du.
\end{aligned}$$

En particulier, pour $h = \mathbb{1}_B$, on a :

$$\mathbb{P}_S(B) = \mathbb{E}\,[\mathbb{1}_B(S)] = \int_B (f*g)(u)\, du,$$

et donc S a la densité $f * g$. □

Comme corollaire, on obtient :

Théorème 21 *Si $X_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ et $X_2 \sim \mathcal{N}(m_2, \sigma_2^2)$ sont deux v.a.r.* ***gaussiennes*** **indépendantes***, alors $X_1 + X_2$ est encore* ***gaussienne*** *et, plus précisément : $X_1 + X_2 \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.*

Preuve. X_1 et X_2 ont respectivement les densités :

$$\begin{cases} f(x) &= \frac{1}{\sigma_1\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-m_1}{\sigma_1}\right)^2 \right] \\ g(x) &= \frac{1}{\sigma_2\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-m_2}{\sigma_2}\right)^2 \right]. \end{cases}$$

D'après le théorème précédent, $X_1 + X_2$ a la densité $f * g$. Calculons-la.

$$\begin{aligned}
(f*g)(x) &= \int_{\mathbb{R}} f(x-t)\, g(t)\, dt \\
&= \int_{\mathbb{R}} \frac{1}{\sigma_1\sqrt{2\pi}} \exp\left[-\frac{1}{2}\Big(\frac{x-t-m_1}{\sigma_1}\Big)^2 \right] \frac{1}{\sigma_2\sqrt{2\pi}} \exp\left[-\frac{1}{2}\Big(\frac{t-m_2}{\sigma_2}\Big)^2 \right] dt \\
&= \frac{1}{2\pi\sigma_1\sigma_2} \int_{\mathbb{R}} \Big(\frac{x-t-m_1}{\sigma_1}\Big)^2 + \Big(\frac{t-m_2}{\sigma_2}\Big)^2 \Big]\, dt.
\end{aligned}$$

Posons :

$$\begin{cases} u &= t - m_2 \\ v &= x - (m_1 + m_2); \end{cases}$$

alors :

$$\begin{aligned}
\Big(\frac{x-t-m_1}{\sigma_1}\Big)^2+\Big(\frac{t-m_2}{\sigma_2}\Big)^2 &= \Big(\frac{v-u}{\sigma_1}\Big)^2+\Big(\frac{u}{\sigma_2}\Big)^2\\
&= \frac{\sigma_2^2 v^2-2\sigma_2^2 uv+(\sigma_1^2+\sigma_2^2)u^2}{\sigma_1^2\sigma_2^2}\\
&= \frac{1}{\sigma_1^2\sigma_2^2}\Big[(\sigma_1^2+\sigma_2^2)\Big(u-\frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\,v\Big)^2\\
&\qquad\qquad -\frac{\sigma_2^4}{\sigma_1^2+\sigma_2^2}\,v^2+\sigma_2^2 v^2\Big]\\
&= \frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}\Big(u-\frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\,v\Big)^2-\frac{v^2}{\sigma_1^2+\sigma_2^2}.
\end{aligned}$$

Donc :

$$\begin{aligned}
(f*g)(x) &= \frac{1}{2\pi\sigma_1\sigma_2}\,\mathrm{e}^{-\frac12\frac{v^2}{\sigma_1^2+\sigma_2^2}}\int_{\mathbb{R}}\mathrm{e}^{-\frac12\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}\big(u-\frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\big)^2}\,du\\
&= \frac{1}{2\pi\sigma_1\sigma_2}\,\mathrm{e}^{-\frac12\frac{v^2}{\sigma_1^2+\sigma_2^2}}\int_{\mathbb{R}}\mathrm{e}^{-\frac12 w^2}\,\frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}}\,dw\\
&= \frac{1}{2\pi\sqrt{\sigma_1^2+\sigma_2^2}}\,\mathrm{e}^{-\frac12\frac{v^2}{\sigma_1^2+\sigma_2^2}}\int_{\mathbb{R}}\mathrm{e}^{-\frac12 w^2}\,dw\\
&= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}}\,\mathrm{e}^{-\frac12\frac{v^2}{\sigma_1^2+\sigma_2^2}}\ . \qquad\Box
\end{aligned}$$

2.5 Loi des grands nombres.

Définition 22 *On dit qu'une suite $(X_n)_{n\geqslant 1}$ de v.a.r. converge en probabilité vers la v.a.r. X si pour tout $a>0$:*

$$\boxed{\mathbb{P}\left(|X_n-X|\geqslant a\right)\underset{n\to\infty}{\longrightarrow}0}\,.$$

On note $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$.

On a :

Proposition 23 *Soit $1\leqslant r<\infty$ et $X_n, X\in\mathscr{L}^r(\Omega,\mathbb{P})$. On a :*

$$\|X_n-X\|_r\underset{n\to\infty}{\longrightarrow}0 \qquad\Longrightarrow\qquad X_n\xrightarrow[n\to\infty]{\mathbb{P}}X.$$

Preuve. Cela résulte immédiatement du résultat facile, mais **très utile**, suivant :

Théorème 24 (inégalité de Markov) *Pour toute v.a.r. Y et tout $a > 0$, on a :*

$$\boxed{\boxed{\mathbb{P}(|Y| \geqslant a) \leqslant \frac{1}{a^r} \int_\Omega |Y|^r \, d\mathbb{P}}}.$$

Preuve.

$$\int_\Omega |Y|^r \, d\mathbb{P} \geqslant \int_{\{|Y| \geqslant a\}} |Y|^r \, d\mathbb{P} \geqslant \int_{\{|Y| \geqslant a\}} a^r \, d\mathbb{P} = a^r \mathbb{P}(|Y| \geqslant a). \qquad \square$$

On peut alors énoncer :

Théorème 25 (loi faible des grands nombres) *Si les v.a.r. $X_1, \ldots, X_n$ sont indépendantes, de même loi, et de carré intégrable, alors :*

$$\boxed{\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n \to \infty]{\mathbb{P}} m},$$

où m est l'espérance commune de $X_1, \ldots, X_n$.

Preuve. On a :

$$\begin{aligned}
\left\| \frac{X_1 + \cdots + X_n}{n} - m \right\|^2 &= \frac{1}{n^2} \|(X_1 + \cdots + X_n) - nm\|^2 \\
&= \frac{1}{n^2} \|(X_1 + \cdots + X_n) - \mathbb{E}(X_1 + \cdots + X_n)\|^2 \\
&= \frac{1}{n^2} \operatorname{Var}(X_1 + \cdots + X_n) \\
&\underset{\perp\!\!\!\perp}{=} \frac{1}{n^2} [\operatorname{Var}(X_1) + \cdots + \operatorname{Var}(X_n)] = \frac{1}{n^2} [n \operatorname{Var}(X_1)] \\
&= \frac{1}{n} \operatorname{Var}(X_1) \xrightarrow[n \to \infty]{} 0 \, ;
\end{aligned}$$

d'où le résultat par la Proposition 23. $\square$

Exemple. On lance une pièce de monnaie non truquée et l'on gagne 1€ si l'on obtient "*pile*" et l'on perd 1€ si l'on obtient "*face*". Les différents lancers sont supposés indépendants les uns des autres.

Pour chaque lancer, la *v.a.r.* X_k correspondant au $k^{ième}$ lancer possède la loi donnée par :

$$\mathbb{P}(X_k = -1) = \mathbb{P}(X_k = 1) = 1/2.$$

On a $\mathbb{E}(X_k) = 0$. Les *v.a.r.* $X_1, \ldots, X_n$ sont indépendantes.

Au bout de n lancers, la somme moyenne gagnée (ou perdue, si la quantité est négative) est $\frac{X_1 + \cdots + X_n}{n}$; elle tend vers 0 en probabilité : pour tout $a > 0$, la probabilité que $|X_1 + \cdots + X_n|$ dépasse na tend vers 0.