7.2 Mesure produit

On rappelle ici qu'un espace mesuré (E, T, m) est σ-fini (on dit aussi que m est σ-finie) s'il existe une famille $(A_n)_{n\in\mathbb{N}} \subset T$ telle que $E = \bigcup_{n\in\mathbb{N}} A_n$ et $m(A_n) < +\infty$, pour tout $n \in \mathbb{N}$. L'espace mesuré $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ est σ-fini (prendre, par exemple, $A_n = [-n, n]$). Il existe par contre des mesures non σ-finies. L'exemple le plus simple sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ consiste à prendre $m(A) = +\infty$ pour tout $A \in \mathcal{B}(\mathbb{R})$, $A \neq \emptyset$. Un exemple plus intéressant (intervenant pour certains problèmes) consiste à se donner un borélien non vide B de $\mathbb{R}$ (B peut être, par exemple, réduit à un point) et à définir m_B par $m_B(A) = +\infty$ si $A \in \mathcal{B}(\mathbb{R})$ et $A \cap B \neq \emptyset$ et $m_B(A) = 0$ si $A \in \mathcal{B}(\mathbb{R})$ et $A \cap B = \emptyset$.

Définition 7.1 (Tribu produit) *Soient* (E_1, T_1) *et* (E_2, T_2) *des espaces mesurables. On pose* $E = E_1 \times E_2$. *On appelle tribu produit la tribu sur* E *engendrée par* $T_1 \times T_2 = \{A_1 \times A_2,\ A_1 \in T_1,\ A_2 \in T_2\}$. *Cette tribu produit est notée* $T_1 \otimes T_2$.

Un exemple fondamental est $(E_1, T_1) = (E_2, T_2) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On va montrer que, dans ce cas, $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^2)$.

Proposition 7.2 (Tribu $\mathcal{B}(\mathbb{R}^N)$) *Pour tout* $N \geq 2$, *on a* $\mathcal{B}(R^{N-1}) \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(R^N)$.

Démonstration – La démonstration est faite pour N = 2 dans l'exercice 2.6). Elle s'adapte facilement pour traiter aussi le cas N > 2 (exercice 7.1). ■

Théorème 7.3 (Mesure produit) *Soient* (E_1, T_1, m_1) *et* (E_2, T_2, m_2) *deux espaces mesurés σ-finis,* $E = E_1 \times E_2$ *et* $T = T_1 \otimes T_2$. *Alors, il existe une et une seule mesure* m *sur* T *vérifiant :*

$$m(A_1 \times A_2) = m_1(A_1)m_2(A_2),\ \forall (A_1, A_2) \in T_1 \times T_2; m_i(A_i) < \infty,\ i = 1, 2. \tag{7.1}$$

Cette mesure est notée $m = m_1 \otimes m_2$. *De plus,* m *est σ-finie.*

7.3 Théorèmes de Fubini-Tonelli et Fubini

Théorème 7.7 (Fubini-Tonelli) *Soient* (E_1, T_1, m_1) *et* (E_2, T_2, m_2) *des espaces mesurés* σ*-finis. On note* (E, T, m) *l'espace produit (donc,* $T = T_1 \otimes T_2$ *et* $m = m_1 \otimes m_2$*). Soit* $f : E \to \overline{\mathbb{R}}_+$ *une fonction mesurable positive (i.e.* T*-mesurable positive). Alors :*

1. $f(x_1, \cdot) \in \mathcal{M}_+(E_2, T_2)$ *pour tout* $x_1 \in E_1$,

   on pose

   $$\varphi_f(x_1) = \int f(x_1, \cdot) dm_2 = \int f(x_1, x_2) dm_2(x_2) \text{ pour tout } x_1 \in E_1,$$

   de sorte que $\varphi_f : E_1 \to \overline{\mathbb{R}}_+$,
2. $\varphi_f \in \mathcal{M}_+(E_1, T_1)$,
3. $\int f dm = \int \varphi_f dm_1 = \int \Big(\int f(x_1, x_2) dm_2(x_2) \Big) dm_1(x_1)$,
4. *les mêmes résultats sont vrais en inversant les rôles de* m_1 *et* m_2*, de sorte que :*

   $$\int \Big(\int f(x_1, x_2) dm_2(x_2) \Big) dm_1(x_1) = \int \Big(\int f(x_1, x_2) dm_1(x_1) \Big) dm_2(x_2).$$

Théorème 7.12 (Fubini) *Soient* (E_1, T_1, m_1) *et* (E_2, T_2, m_2) *des espaces mesurés* σ*-finis. On note* (E, T, m) *l'espace produit. Soit* f *une fonction* T*-mesurable de* E *dans* $\mathbb{R}$ *(c'est-à-dire* $f \in \mathcal{M}(E, T)$*) et intégrable pour la mesure m, c'est-à-dire* $f \in \mathcal{L}^1_{\mathbb{R}}(E, T, m)$*. Alors :*

1. $f(x_1, \cdot) \in \mathcal{L}^1_{\mathbb{R}}(E_2, T_2, m_2)$ *pour presque tout* $x_1 \in E_1$,

   on pose $\varphi_f(x_1) = \int f(x_1, \cdot) dm_2$ *pour* $x_1 \in E_1$ *t.q.* $f(x_1, \cdot) \in \mathcal{L}^1_{\mathbb{R}}(E_2, T_2, m_2)$*. La fonction* φ_f *est donc définie p.p. sur* E_1 *(et à valeurs dans* $\mathbb{R}$*).*
2. $\varphi_f \in L^1_{\mathbb{R}}(E_1, T_1, m_1)$ *(au sens : il existe* $g \in \mathcal{L}^1_{\mathbb{R}}(E_1, T_1, m_1)$ *t.q.* $f = g$ *p.p.).*
3. $\displaystyle\int f dm = \int \varphi_f dm_1 = \int \Big(\int f(x_1, x_2) dm_2(x_2) \Big) dm_1(x_1)$,
4. *les mêmes résultats sont vrais en inversant les rôles de* m_1 *et* m_2*, de sorte que :*

$$\int \Big(\int f(x_1, x_2) dm_2(x_2) \Big) dm_1(x_1) = \int \Big(\int f(x_1, x_2) dm_1(x_1) \Big) dm_2(x_2).$$

Chapitre 9

Vecteurs aléatoires

9.1 Définition, propriétés élémentaires

Définition 9.1 (Vecteur aléatoire) *Soit $d \geq 1$ et $(\Omega, \mathcal{A})$ un espace probabilisable. On appelle vecteur aléatoire (souvent noté v.a.) de dimension d une application mesurable de Ω dans $\mathbb{R}^d$ (où $\mathbb{R}^d$ est muni de la tribu borélienne).*

Noter que la notation "v.a." signifie indifféremment "variable(s) aléatoire(s)" ou "vecteur(s) aléatoire(s)".

Proposition 9.2 *Soit $d > 1$ et $(\Omega, \mathcal{A})$ un espace probabilisable. Soit X une application de Ω dans $\mathbb{R}^d$. On note $X_1, \ldots, X_d$ les composantes de X (de sorte de $X = (X_1, \ldots, X_d)^t$). On note $\sigma(X)$ la tribu engendrée par X (et $\sigma(X_i)$, pour $i = 1, \ldots, d$ la tribu engendrée par X_i). On a alors :*

1. $\sigma(X)$ est la plus petite tribu contenant les tribus $\sigma(X_1), \ldots, \sigma(X_d)$.

2. X est un v.a. si et seulement si X_i est, pour tout $i \in \{1, \ldots, d\}$, une v.a.r..

DÉMONSTRATION – On rappelle que $\sigma(X) = \{X^{-1}(B), B \in \mathcal{B}(\mathbb{R}^d)\}$ et que, pour tout i, $\sigma(X_i) = \{X_i^{-1}(A), A \in \mathcal{B}(\mathbb{R})\}$. On note $\mathcal{C} = \{\prod_{i=1}^d A_i, A_i \in \mathcal{B}(\mathbb{R})$ pour tout $i\}$. On rappelle que $\mathcal{C} \subset \mathcal{B}(\mathbb{R}^d)$ (voir l'exercice 7.10) et que $\mathcal{C}$ engendre $\mathcal{B}(\mathbb{R}^d)$ (c'est même encore vrai si on se limite à prendre pour A_i des intervalles, ouverts, voir l'exercice 2.7).

On note T la plus petite tribu contenant les tribus $\sigma(X_1), \ldots, \sigma(X_d)$. Soit $i \in \{1, \ldots, d\}$ et $A \in \mathcal{B}(\mathbb{R})$. En prenant $B = \prod_{j=1}^d A_j$ avec $A_j = \mathbb{R}$ si $j \neq i$ et $A_i = A$, on a $X^{-1}(B) \in \sigma(X)$ (car $B \in \mathcal{C} \subset \mathcal{B}(\mathbb{R}^d)$) et $X^{-1}(B) = X_i^{-1}(A)$. On en déduit $X_i^{-1}(A) \in \sigma(X)$ et donc $\sigma(X_i) \subset \sigma(X)$ pour tout $i \in \{1, \ldots, d\}$. Comme $\sigma(X)$ est une tribu, on a donc $T \subset \sigma(X)$.

On montre maintenant l'inclusion inverse (c'est-à-dire $T \supset \sigma(X)$). On remarque que, si $B \in \mathcal{C}$ on a $B = \prod_{i=1}^d A_i$ avec des A_i appartenant à $\mathcal{B}(\mathbb{R})$. On a donc $X^{-1}(B) = \bigcap_{i=1}^d X_i^{-1}(A_i) \in T$ (car $X_i^{-1}(A_i) \in \sigma(X_i) \subset T$). Or, il est facile de voir que $\{B \in \mathcal{B}(\mathbb{R}^d)$, t.q. $X^{-1}(B) \in T\}$ est une tribu. Cette tribu contient $\mathcal{C}$, elle contient donc la tribu engendrée par $\mathcal{C}$, c'est-à-dire $\mathcal{B}(\mathbb{R}^d)$. On vient donc de montrer que $\{B \in \mathcal{B}(\mathbb{R}^d)$, t.q. $X^{-1}(B) \in T\} = \mathcal{B}(\mathbb{R}^d)$, c'est-à-dire que $X^{-1}(B) \in T$ pour tout $B \in \mathcal{B}(\mathbb{R}^d)$, ou encore que $\sigma(X) \subset T$. On a bien, finalement, $\sigma(X) = T$, ce qui donne le premier item de la proposition.

Le deuxième item est un conséquence facile du premier. En effet, si X est un v.a., on a pour tout i, $\sigma(X_i) \subset \sigma(X) \subset \mathcal{A}$ et donc X_i est une v.a.r.. Réciproquement ; si X_i est, pour tout i, une v.a.r., on a $\sigma(X_i) \subset \mathcal{A}$ pour tout i. Comme $\mathcal{A}$ est une tribu, on a donc $\mathcal{A} \supset T$ et comme $T = \sigma(X)$ on en déduit que X est un v.a.. ■

La démonstration de la proposition 9.2 n'utilise pas vraiment le fait que les X_i soient des applications à valeurs dans $\mathbb{R}$. Elle se généralise donc facilement au cas où les X_i sont des applications à valeurs dans $\mathbb{R}^{d_i}$.

Proposition 9.3 *Soit $p \in \mathbb{N}$, $p \geq 2$, $d_1, \ldots, d_p \in \mathbb{N}^*$. soit $(\Omega, \mathcal{A})$ un espace probabilisable et, pour tout $i \in \{1, \ldots, p\}$, X_i une application de Ω dans $\mathbb{R}^{d_i}$. On note $X = (X_1, \ldots, X_d)^t$, de sorte que X est une application de Ω dans $\mathbb{R}^d$ avec $d = d_1 + \ldots + d_p$. Soit $d \geq 1$ et $(\Omega, \mathcal{A})$ un espace probabilisable. Soit X une application de Ω dans $\mathbb{R}^d$. On note $\sigma(X)$ la tribu engendrée par X (et $\sigma(X_i)$, pour $i = 1, \ldots, p$ la tribu engendrée par X_i). On a alors :*

1. $\sigma(X)$ est la plus petite tribu contenant les tribus $\sigma(X_1), \ldots, \sigma(X_p)$.

2. X est un v.a. (de dimension d) si et seulement si X_i est, pour tout $i \in \{1, \ldots, p\}$, un v.a. (de dimension d_i).

Définition 9.4 (Loi d'un v.a.) *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé, $d \geq 1$ et X un v.a. de dimension d. On appelle loi de probabilité de X, et on note cette loi P_X, la probabilité sur $\mathcal{B}(\mathbb{R}^d)$ image par X de P (c'est-à-dire $P_X(A) = P(X^{-1}(A)) = P(\{X \in A\})$ pour tout $A \in \mathcal{B}(\mathbb{R}^d)$). Si $(X_1, \ldots, X_d)$ sont les composantes de X, La probabilité P_X est aussi appelée loi conjointe de $(X_1, \ldots, X_d)$.*

Un exemple important est la loi normale multidimensionnelle.

Définition 9.5 (Loi normale multidimensionnelle) *Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé, $d \geq 1$ et X un v.a. de dimension d. Soit $m \in \mathbb{R}^d$ et D une matrice (à coefficients réels, de taille $d \times d$) s.d.p. (c'est-à-dire symétrique définie positive). Le v.a. X a pour loi $\mathcal{N}(m, D)$ (loi normale de paramètre m et D) si $P_X = f\lambda_d$ avec :*

$$f(x) = \frac{1}{(2\pi)^{d/2}\sqrt{det(D)}} \exp(-\frac{1}{2}(x-m)^t D^{-1}(x-m)) \text{ pour } x \in \mathbb{R}^d.$$

Si D est seulement semi–définie positive (c'est-à-dire $Du \cdot u \geq 0$ pour tout $u \in \mathbb{R}^d$) au lieu d'être définie positive (c'est-à-dire $Du \cdot u > 0$ pour tout $u \in \mathbb{R}^d$, $u \neq 0$), la loi normale $\mathcal{N}(0, D)$ est aussi définie, voir la proposition 9.33, mais ce n'est pas une loi de densité (par rapport à la mesure de Lebesgue), voir l'exercice 10.13 (et l'exercice 9.15 qui donne un exemple de loi normale bidimensionnelle qui n'est pas de densité).

Nous verrons dans la section 9.3 que si X est un v.a. suivant une loi normale multidimensionnelle, X est un v.a. gaussien. L'extension de la définition de la loi normale multidimensionnelle au cas D non inversible permettra alors de dire que les v.a. gaussiens sont exactement ceux qui suivent une loi normale multidimensionnelle (proposition 9.33).

Le fait que les composantes du v.a. X suivent une loi normale ne donne pas que X suit une loi normale multidimensionnelle (voir, par exemple, l'exercice 10.14). Mais, si les composantes du v.a. X suivent une loi normale et sont indépendantes, le v.a. X suit alors une loi normale multidimensionnelle (cf. l'exercice 9.9 ou l'exercice 10.14).

Définition 9.6 (I-ème projecteur) *Soit $d \geq 1$. On appelle i-ème projecteur de $\mathbb{R}^d$ l'application π_i, de $\mathbb{R}^d$ dans $\mathbb{R}$, qui a un vecteur de $\mathbb{R}^d$ fait correspondre sa i-ème composante dans la base canonique de $\mathbb{R}^d$.*

Définition 9.7 (Probabilité marginale) *Soit $d > 1$.*

1. *Soit p une probabilité sur les boréliens de $\mathbb{R}^d$ et $i \in \{1,\ldots,d\}$. On appelle i-ème probabilité marginale de p, la probabilité sur les boréliens de $\mathbb{R}$ défini par $p_i(\mathrm{B}) = p(\pi_i^{-1}(\mathrm{B}))$ pour tout $\mathrm{B} \in \mathcal{B}(\mathbb{R})$. (On a donc, par exemple, $p_1(\mathrm{B}) = p(\mathrm{B}\times\mathbb{R}^{d-1})$.)*
2. *Soit $(\Omega, \mathcal{A}, \mathrm{P})$ un espace probabilisé et X un vecteur aléatoire de dimension d. On appelle i-ème loi de probabilité marginale du vecteur aléatoire X (ou i-ème probabilité marginale du vecteur aléatoire X) la probabilité marginale de $\mathrm{P_X}$ (c'est donc aussi la mesure image de $\mathrm{P_X}$ par le i-ème projecteur π_i). La remarque suivante montre que cette probabilité marginale est en fait la loi de X_i notée $\mathrm{P}_{\mathrm{X}_i}$.*

Remarque 9.8 Soit $(\Omega, \mathcal{A}, \mathrm{P})$ espace probabilisé, $d \geq 1$ et $\mathrm{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_d)^t$ un vecteur aléatoire de dimension d. On note q_i la i-ème probabilité marginale du vecteur aléatoire X. Soit $\mathrm{B} \in \mathrm{T}$, par définition de la loi marginale, on a :

$$q_i(\mathrm{B}) = \mathrm{P_X}(\pi_i^{-1}(\mathrm{B})) = \mathrm{P}(\mathrm{X}^{-1}(\pi_i^{-1}(\mathrm{B}))).$$

Comme $\mathrm{X}_i = \pi_i \circ \mathrm{X}$, on a donc :

$$q_i(\mathrm{B}) = \mathrm{P}(\mathrm{X}_i^{-1}(\mathrm{B})).$$

La probabilité q_i est donc aussi la loi de la variable aléatoire X_i.

Remarque 9.9 Soit $(\Omega, \mathcal{A}, \mathrm{P})$ espace probabilisé, $d > 1$ et $\mathrm{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_d)^t$ un vecteur aléatoire de dimension d. La connaissance de $\mathrm{P_X}$ entraîne la connaissance des $\mathrm{P}_{\mathrm{X}_i}$. La réciproque est en général fausse.

On définit la densité d'une loi de manière analogue au cas scalaire.

Définition 9.10 (Loi de densité) *Soit p une probabilité sur $\mathcal{B}(\mathbb{R}^d)$ ($d \geq 1$), on dit que p est une probabilité de densité (par rapport à la mesure de Lebesgue) s'il existe une application borélienne f de $\mathbb{R}^d$ dans $\mathbb{R}_+$ t.q. $p = f\lambda_d$.*

De même que dans le cas scalaire, on a un théorème qui donne la correspondance entre l'intégrale par rapport à la probabilité P d'une fonction de la v.a.r. X et l'intégrale de cette fonction par rapport à la probabilité $\mathrm{P_X}$:

Théorème 9.11 (Loi image) *Soit $d \geq 1$, $(\Omega,\mathcal{A},\mathrm{P})$ un espace probabilisé, X un v.a. de dimension d et $\mathrm{P_X}$ la loi de X. Soit φ une application borélienne de $\mathbb{R}^d$ dans $\mathbb{R}$. On a alors (On rappelle que $\varphi(\mathrm{X})$ désigne $\varphi \circ \mathrm{X}$) :*

1. *$\varphi(\mathrm{X}) \in \mathrm{L}^1_{\mathbb{R}}(\Omega,\mathcal{A},\mathrm{P})$ si et seulement si $\varphi \in \mathrm{L}^1_{\mathbb{R}}(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d),\mathrm{P_X})$,*
2. *L'égalité*

$$\int_\Omega \varphi(\mathrm{X})d\mathrm{P} = \int_{\mathbb{R}^d} \varphi d\mathrm{P_X},$$

est vraie si φ prend ses valeurs dans $\mathbb{R}_+$, ou si φ est bornée ou encore si $\varphi(\mathrm{X}) \in \mathrm{L}^1_{\mathbb{R}}(\Omega,\mathcal{A},\mathrm{P})$.

Définition 9.12 (Fonction de répartition)
Soit $d \geq 1$, $(\Omega,\mathcal{A},\mathrm{P})$ un espace probabilisé, $\mathrm{X} = (\mathrm{X}_1,\ldots,\mathrm{X}_d)^t$ un v.a. de dimension d et $\mathrm{P_X}$ la loi de X. On appelle fonction de répartition du vecteur aléatoire X la fonction définie de $\mathbb{R}^d$ dans $[0,1]$ par : $\mathrm{F_X}(t_1,\ldots,t_d) = \mathrm{P}(\{\mathrm{X}_1 \leq t_1,\ldots,\mathrm{X}_d \leq t_d\})$.

Proposition 9.13 *Soit $d \geq 1$, $(\Omega, \mathcal{A}, \mathrm{P})$ un espace probabilisé et X un v.a. de dimension d de fonction de répartition $\mathrm{F_X}$. Alors :*

1. $0 \leq \mathrm{F_X}(t) \leq 1$ *pour tout* $t \in \mathbb{R}^d$ *;*
2. *Si* $t, t' \in \mathbb{R}^d$, $t \leq t'$ *(i.e.* $t_i \leq t'_i$*, pour tout* $i = 1, \ldots, d$*),* $\mathrm{F_X}(t) \leq \mathrm{F_X}(t')$ *;*
3. $\mathrm{F_X}$ *est continue à droite en tout point ;*
4. $\mathrm{F_X}(t_1, \ldots, t_d) \to 1$ *lorsque* $(t_1, \ldots, t_d) \to (+\infty, \ldots, +\infty)$ *;*
5. $\mathrm{F_X}(t_1, \ldots, t_d) \to 0$ *lorsque* $t_i \to -\infty$ *(à* i *fixé).*

La démonstration découle facilement des propriétés d'une mesure (monotonie, continuité croissante et continuité décroissante).

Proposition 9.14 *Soit $d \geq 1$, $(\Omega, \mathcal{A}, \mathrm{P})$ un espace probabilisé et X un v.a. de dimension d de fonction de répartition $\mathrm{F_X}$. Si $\mathrm{F_X} \in \mathrm{C}^d(\mathbb{R}^d, [0,1])$, alors p_{X} est une probabilité de densité (par rapport à Lebesgue) et cette densité, notée f_{X}, vérifie*

$$f_{\mathrm{X}} = \frac{\partial^d \mathrm{F_X}}{\partial x_1 \ldots \partial x_d} \quad \lambda_d\text{-p.p..} \tag{9.1}$$

Démonstration – On suppose que $d = 1$. Pour tout $a, b \in \mathbb{R}$, $a < b$, la définition de $\mathrm{F_X}$ donne $\mathrm{P_X}(]a,b]) = \mathrm{F_X}(b) - \mathrm{F_X}(a)$. Mais, comme $\mathrm{F_X}$ est de classe C^1, on a $\mathrm{F_X}(b) - \mathrm{F_X}(a) = \int_a^b \mathrm{F'_X}(t)dt$. En notant m la mesure de densité $\mathrm{F'_X}$ par rapport à λ, on a donc $\mathrm{P_X}(]a,b]) = m(]a,b])$ pour tout $a, b \in \mathbb{R}$, $a < b$, ce qui est suffisant pour dire que $\mathrm{P_X} = m$ (en utilisant, par exemple, la proposition 2.31).

Pour $d > 1$, on note m la mesure de densité $\partial^d \mathrm{F_X}/\partial x_1 \ldots \partial x_d$ par rapport à λ_d. Un raisonnement voisin du précédent donne $m(\prod_{i=1}^d]a_i, b_i]) = \mathrm{P_X}(\prod_{i=1}^d]a_i, b_i])$ pour tout $a_i, b_i \in \mathbb{R}$, $a_i < b_i$, $i = 1, \ldots, d$. On en déduit $m = \mathrm{P_X}$ avec la proposition 2.31. ■

Définition 9.15 (Espérance) *Soit $(\Omega, \mathcal{A}, \mathrm{P})$ un espace probabilisé, $d > 1$ et $\mathrm{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_d)^t$ un vecteur aléatoire de dimension d. On suppose que $\mathrm{E}(|\mathrm{X}_i|) < \infty$ pour tout $i \in \{1, \ldots, d\}$. L'espérance de X, notée $\mathrm{E}(\mathrm{X})$, est alors le vecteur de $\mathbb{R}^d$ dont les composantes sont les espérances des X_i, c'est-à-dire $\mathrm{E}(\mathrm{X}) = (\mathrm{E}(\mathrm{X}_1), \ldots, \mathrm{E}(\mathrm{X}_d))^t$.*

Remarque 9.16 Soit $(\Omega, \mathcal{A}, \mathrm{P})$ un espace probabilisé, $d > 1$ et $\mathrm{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_d)^t$ un vecteur aléatoire de dimension d t.q. $\mathrm{E}(|\mathrm{X}_i|) < \infty$ pour tout $i \in \{1, \ldots, d\}$. Soit $u \in \mathbb{R}^d$. L'application $u \cdot \mathrm{X}$ (qui à $\omega \in \Omega$ associe $u \cdot \mathrm{X}(\omega)$, on rappelle que $\xi \cdot \eta$ est le produit scalaire canonique de ξ et η dans $\mathbb{R}^d$) est une v.a.r. intégrable et il est clair que $\mathrm{E}(u \cdot \mathrm{X}) = u \cdot \mathrm{E}(\mathrm{X})$. L'application $u \mapsto \mathrm{E}(u \cdot \mathrm{X})$ est donc l'application linéaire sur $\mathbb{R}^d$ représentée par $\mathrm{E}(\mathrm{X})$.

Définition 9.17 (Variance, covariance) *Soit $(\Omega, \mathcal{A}, \mathrm{P})$ un espace probabilisé, $d > 1$ et $\mathrm{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_d)^t$ un vecteur aléatoire de dimension d. On suppose que $\mathrm{E}(\mathrm{X}_i^2) < \infty$ pour tout $i \in \{1, \ldots, d\}$. On définit alors la matrice de covariance de X, notée $Cov(\mathrm{X})$, comme la matrice dont le coefficient i, j est donné par :*

$$Cov(\mathrm{X})_{i,j} = \mathrm{E}((\mathrm{X}_i - \mathrm{E}(\mathrm{X}_i))(\mathrm{X}_j - \mathrm{E}(\mathrm{X}_j))) \text{ pour tout } i, j \in \{1, \ldots, d\}.$$

On a donc $\mathrm{C}_{i,i} = Var(\mathrm{X}_i)$ et $\mathrm{C}_{i,j} = Cov(\mathrm{X}_i, \mathrm{X}_j)$ (noter d'ailleurs que $Cov(\mathrm{X}_i, \mathrm{X}_i) = Var(\mathrm{X}_i)$). Enfin, on peut aussi noter que $Cov(\mathrm{X})$ est l'espérance de la matrice $(\mathrm{X} - \mathrm{E}(\mathrm{X}))(\mathrm{X} - \mathrm{E}(\mathrm{X}))^t$.

Remarque 9.18 Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé, $d > 1$ et $X = (X_1, \ldots, X_d)^t$ un vecteur aléatoire de dimension d t.q. $E(X_i^2) < \infty$ pour tout $i \in \{1, \ldots, d\}$. Pour tout $u \in \mathbb{R}^d$, on a donc $E((u \cdot X)^2) < \infty$ et il est facile de voir (voir l'exercice 9.5) que $\mathrm{Var}(u \cdot X) = u^t \mathrm{Cov}(X) u$. La matrice $\mathrm{Cov}(X)$ est donc la matrice de la forme quadratique $u \mapsto \mathrm{Var}(u \cdot X)$, définie sur $\mathbb{R}^d$. Cette matrice est donc symétrique et semi–définie positive. On peut aussi noter que, pour tout $u, v \in \mathbb{R}^d$, on a $u^t \mathrm{Cov}(X) v = E((u \cdot X)(v \cdot X))$.

Comme dans le cas des v.a.r., on définit la convergence en loi de v.a. à partir de la convergence étroite (ou vague, puisque c'est équivalent pour des probabilités) des lois des v.a.. La convergence étroite est définie dans la définition 6.86.

Définition 9.19 (Convergence en loi de v.a.) *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé,* $d \geq 1$, $(X_n)_{n \in \mathbb{N}}$ *une suite de v.a. de dimension* d *et* X *un v.a. de dimension* d. *On dit que* $X_n \to X$ *en loi, quand* $n \to +\infty$, *si :*

$$\int_\Omega \varphi(X_n) dP \to \int_\Omega \varphi(X) dP \text{ pour tout } \varphi \in C_b(\mathbb{R}^d, \mathbb{R}).$$

(Ce que est équivalent à dire que $P_{X_n} \to P_X$ *étroitement.)*
Comme pour les v.a.r., il est possible de définir la convergence en loi pour une suite de v.a. définies sur des espaces probabilisés différents (c'est-à-dire que X_n *est définie sur l'espace probabilisé* $(\Omega_n, \mathcal{A}_n, P_n)$ *dépendant de n, et* X *est définie sur* $(\Omega, \mathcal{A}, P)$*). Il suffit que tous les v.a. aient même dimension. Nous utiliserons parfois implicitement cette définition plus générale.*

Ici, comme dans le cas des v.a.r., on remarque que la convergence en loi donne la tension de la suite des lois. Ceci est donné dans la proposition 9.20.

Proposition 9.20 *Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé,* $d \geq 1$, $(X_n)_{n \in \mathbb{N}}$ *une suite de v.a. de dimension* d *et* X *un v.a. de dimension* d. *On suppose que* $X_n \to X$ *en loi. La suite* $(P_{X_n})_{n \in \mathbb{N}}$ *est alors tendue, c'est-à-dire*

$$\lim_{a \to +\infty} P(\{|X_n| > a\}) = 0 \text{ uniformément par rapport à } n.$$

Démonstration – Le résultat est donnée par proposition 6.93 en prenant $m_n = P_{X_n}$. ■

La proposition 9.20 admet une réciproque partielle qui est due à la proposition 8.22. Nous donnons maintenant cette réciproque (partielle).

Proposition 9.21 *Soit* $d \geq 1$, $(\Omega, \mathcal{A}, P)$ *un espace probabilisé et* $(X_n)_{n \in \mathbb{N}}$ *une suite de v.a. de dimension* d. *On suppose que la suite des lois des* X_n *est tendue. Il existe alors un v.a.* X *(de dimension* d*) et une sous-suite de la suite* $(X_n)_{n \in \mathbb{N}}$ *telle que cette sous-suite converge en loi vers* X.

Démonstration – Pour $n \in \mathbb{N}$, on note m_n la loi de X_n. Comme $(m_n)_{n \in \mathbb{N}}$ est une suite tendue de probabilités sur $\mathcal{B}(\mathbb{R}^d)$, la proposition 8.22 donne l'existence d'une probabilité m sur $\mathcal{B}(\mathbb{R}^d)$ t.q. $m_n \to m$ étroitement. En prenant un v.a. X t.q. $P_X = m$ (X n'est pas nécessairement défini sur le même espace probabilisé), on a donc la convergence en loi de X_n vers X. ■