La terminologie probabiliste utilise les termes “variable aléatoire” ou “vecteur aléatoire” (selon l’espace d’arrivée) au lieu de “fonction mesurable” (ou “application mesurable”).

Définition 3.12 (Variable aléatoire, vecteur aléatoire)

1. *Soit* (E, T) *un espace probabilisable, on appelle variable aléatoire réelle (v.a.r.) une fonction* X *définie de* E *dans* $\mathbb{R}$ *et* T*-mesurable,i.e. telle que* $X^{-1}(A) \in T$, *pour tout* $A \in \mathcal{B}(\mathbb{R})$.
2. *Plus généralement, soient* (E, T) *et* $(F, \mathcal{T})$ *deux espaces probabilisables. Une fonction* X, *définie de* E *dans* F, *est une variable aléatoire si c’est une fonction* $(T, \mathcal{T})$*-mesurable (c’est-à-dire si* $X^{-1}(A) \in T$, *pour tout* $A \in \mathcal{T}$*). Lorsque* F *est un espace vectoriel, on dit que* X *est une variable aléatoire vectorielle ou un “vecteur aléatoire”.*

Remarque 3.13 Comme cela a été dit dans la proposition 3.10, on dit, en l’absence d’ambiguïté, “mesurable” au lieu de “T-mesurable”. On remarque d’ailleurs que le terme probabiliste “variable aléatoire” ne mentionne pas la dépendance par rapport à la tribu. Dans la définition 3.12, on a noté X la variable aléatoire plutôt que f car c’est l’usage dans la littérature probabiliste.

Définition 3.14 (Tribu engendrée par une fonction mesurable) *Soient* (E, T) *un espace mesurable (resp. probabilisable) et* f *(resp.* X*) une fonction mesurable de* E *dans* $\mathbb{R}$ *(resp. une variable aléatoire) alors l’ensemble* $\{f^{-1}(A), A \in \mathcal{B}(\mathbb{R})\}$ *(resp.* $\{X^{-1}(A), A \in \mathcal{B}(\mathbb{R})\}$*) est une tribu sur* E *qu’on appelle tribu engendrée par la fonction mesurable* f *(resp. la variable aléatoire* X*). Cette tribu est aussi la tribu image réciproque de* $\mathcal{B}(\mathbb{R})$ *par* f *(resp.* X*).*

3.4 Mesure image, loi d'une v.a., v.a. indépendantes

Soit (E, T) et $(F, \mathcal{T})$ deux espaces mesurables (l'exemple fondamental est $(F, \mathcal{T}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$) et f une fonction mesurable de E vers F. Si m est une mesure sur T, alors on peut définir, à partir de f et m, une mesure sur $\mathcal{T}$ de la manière suivante :

Proposition 3.24 (Mesure image) *Soient* (E, T, m) *un espace mesuré,* $(F, \mathcal{T})$ *un espace mesurable et* f *une fonction mesurable de* E *vers* F *(c'est-à-dire* $(T, \mathcal{T})$*-mesurable). Alors, l'application* m_f *définie de* $\mathcal{T}$ *dans* $\mathbb{R}_+$ *par :* $m_f(A) = m(f^{-1}(A))$*, pour tout* $A \in \mathcal{T}$*, est une mesure sur* $\mathcal{T}$*, appelée mesure image par* f*.*

DÉMONSTRATION – Il suffit de remarquer que m_f est bien définie, que $m_f(\emptyset) = 0$ et que m_f est σ−additive, ce qui découle naturellement des propriétés de m. ■

Définition 3.25 (Loi de probabilité et fonction de répartition d'une v.a.)
Soient (E, T, p) *un espace probabilisé,* X *une variable aléatoire réelle (c'est-à-dire une fonction mesurable de* E*, muni de la tribu* T*, dans* $\mathbb{R}$*, muni de la tribu borélienne). On appelle loi de probabilité de la variable aléatoire* X *la probabilité* p_X *image de* p *par* X *(cette probabilité est donc définie sur* $\mathcal{B}(\mathbb{R})$*). On appelle fonction de répartition de la variable aléatoire* X *la fonction de répartition de la probabilité* p_X.

Dans de nombreux cas, les modèles probabilistes seront déterminés par une loi de probabilité d'une variable aléatoire. Une conséquence immédiate du théorème 2.61 est que la loi de probabilité d'une variable aléatoire réelle est entièrement déterminée par sa fonction de répartition. Ceci est énoncé dans la proposition suivante.

Proposition 3.26 (Égalité de deux lois) *Soient* (E, T, p) *et* (E', T', p') *des espaces probabilisés,* X *une variable aléatoire réelle sur* E *(c'est-à-dire une fonction mesurable de* E*, muni de* T*, dans* $\mathbb{R}$ *muni de* $\mathcal{B}(\mathbb{R})$*) et* X′ *une variable aléatoire réelle sur* E′*. On a alors* $p_X = p_Y$ *si et seulement si* $p(\{X \le t\}) = p'(\{Y \le t\})$ *pour tout* $t \in \mathbb{R}$*. On a aussi* $p_X = p_Y$ *si et seulement si* $p(\{s \le X \le t\}) = p'(\{s \le Y \le t\})$ *pour tout* $s, t \in \mathbb{R}$*,* $s < t$*. (Les inégalités strictes peuvent être remplacées par des inégalités larges.)*

DÉMONSTRATION – Cette proposition est une conséquence des théorèmes 2.61 et 2.62. Il suffit de remarquer que $p(\{X \le t\}) = p_X(]-\infty, t])$ et $p(\{s \le X \le t\}) = p_X([s, t])$ (et les mêmes égalités avec Y au lieu de X). ■

On rappelle que la notation $p(\{X \le t\})$ (si X est une v. a. réelle sur l'espace probabilisé (E, T, p)) signifie $p(\{\omega \in E\,;\, X(\omega) \le t\})$. Cette notation sera parfois abrégée sous la forme $p(X \le t)$.

Définition 3.27 (Variables aléatoires équidistribuées)
Soient (E, T, p) *et* (E', T', p') *des espaces probabilisés,* X *(resp.* X′*) une variable aléatoire de* (E, T, p) *(resp.* (E', T', p')*) dans* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$*, on dit que les variables aléatoires* X *et* X′ *sont équidistribuées si elles ont même loi de probabilité.*

Définition 3.28 (Variable aléatoire discrète, entière, continue) *Soient* (E, T, p) *un espace probabilisé,* X *une variable aléatoire réelle sur* (E, T, p)*,* p_X *la loi de la variable aléatoire* X *et* F_X *sa fonction de répartition ;*

1. Si $X(E)$ *est dénombrable, on dit que la variable aléatoire* X *est discrète.*

2. Si $X(E) \subset \mathbb{N}$*, on dit que la variable aléatoire* X *est entière.*

3. Si la fonction de répartition F_X *définie de* $\mathbb{R}$ *dans* $[0,1]$ *est continue, on dit que la variable aléatoire est continue.*

Définition 3.29 (Variables aléatoires indépendantes) *Soit* (E,T,p) *un espace probabilisé.*

1. Soit $N > 1$ *et* $X_1,\ldots,X_N$ *une famille de variables aléatoires réelles. On dit que* $X_1,\ldots,X_N$ *sont indépendantes (ou que la famille (* $X_1,\ldots,X_N$*) est indépendante) si les tribus engendrées par* $X_1,\ldots,X_N$ *(on notera souvent* $\tau(X)$ *ou* $\sigma(X)$ *la tribu engendrée par la variable aléatoire* X*) sont indépendantes.*

2. Soit $(X_n)_{n\in\mathbb{N}^*}$ *une suite de variables aléatoires réelles. On dit cette suite est indépendante (ou que les v.a.* $X_1,\ldots,$ $X_n,\ldots$ *sont indépendantes) si, pour tout* $N > 1$*, les v.a.* $X_1,\ldots,X_N$ *sont indépendantes.*

On appellera “suite de v.a.r.i.i.d. ” une suite de variables aléatoires réelles indépendantes identiquement distribuées (ce dernier point signifiant que toutes les v.a. de la suite ont même loi).

Soit (E,T,p) un espace probabilisé et X_1,X_2,X_3 trois v.a.r. (c’est-à-dire variables aléatoires réelles). Le fait que X_1 soit indépendante de X_2 et X_3 n’implique pas que X_1 soit indépendante de (par exemple) X_2+X_3, même si X_2 et X_3 sont indépendantes. Mais, on a bien X_1 indépendante de X_2+X_3 si la famille (X_1,X_2,X_3) est indépendante. Ceci est une conséquence de la proposition suivante.

Proposition 3.30 (Indépendance et composition) *Soit* (E,T,p) *un espace probabilisé,* $n \geq 1$*,* $m \geq 1$ *et* $X_1,\ldots,$ X_n*,* $Y_1,\ldots,Y_m$ *des v.a.r. indépendantes. Soit* φ *une fonction borélienne de* $\mathbb{R}^n$ *dans* $\mathbb{R}$ *et* ψ *une fonction borélienne de* $\mathbb{R}^m$ *dans* $\mathbb{R}$*. Alors, les v.a.r.* $\varphi(X_1,\ldots,X_n)$ *et* $\psi(Y_1,\ldots,Y_m)$ *sont indépendantes. Nous avons ici décomposé la famille initiale de v.a.r. indépendantes en 2 groupes. la proposition peut se généraliser à une décomposition en un nombre quelconque de groupes.*

DÉMONSTRATION – La notation $\varphi(X_1,\ldots,X_n)$ est un peu incorrecte (mais est toujours utilisée). Elle désigne (comme on le devine facilement) la composition de φ (qui va de $\mathbb{R}^n$ dans $\mathbb{R}$) avec l’application de E dans $\mathbb{R}^n$ donnée par les X_i, $i=1,\ldots,n$.

La démonstration de cette proposition (et de sa généralisation à un nombre quelconque de groupes) est une conséquence simple de la proposition 2.59 dés que l’on remarque que la tribu engendrée par $\varphi(X_1,\ldots,X_n)$ est incluse dans la tribu engendrée par $X_1,\ldots,X_n$, ce que nous démontrons maintenant.

On note τ la tribu engendrée par $X_1,\ldots,X_n$ et X l’application de E dans $\mathbb{R}^n$ qui à $\omega \in E$ associe $(X_1(\omega),\ldots,X_n(\omega))^t$. Il est facile de voir que $\{A \in \mathcal{B}(\mathbb{R}^n)$ t.q. $X^{-1}(A) \in \tau\}$ est une tribu (sur $\mathbb{R}^n$). Si $A = \prod_{i=1}^n A_i$ avec $A_i \in \mathcal{B}(\mathbb{R})$ pour tout $i=1,\ldots,n$, on a

$$X^{-1}(A) = \bigcap_{i=1}^{n} X_i^{-1}(A_i) \in \tau$$

(car $X_i^{-1}(A_i)$ appartient à $\tau(X_i)$ et donc à τ). Comme $\mathcal{B}(\mathbb{R}^n)$ est engendrée par l’ensemble des produits de boréliens de $\mathbb{R}$ (et même par l’ensemble des produits d’intervalles ouverts de $\mathbb{R}$, voir l’exercice 2.7), on en déduit que

$$\{A \in \mathcal{B}(\mathbb{R}^n) \text{ t.q. } X^{-1}(A) \in \tau\} \supset \mathcal{B}(\mathbb{R}^n).$$

Pour tout $B \in \mathcal{B}(\mathbb{R})$, on a donc $(\varphi(X))^{-1}(B) = X^{-1}(\varphi^{-1}(B)) \in \tau$ car $\varphi^{-1}(B) \in \mathcal{B}(\mathbb{R}^n)$ (puisque φ est borélienne), ce qui prouve bien que la tribu engendrée par $\varphi(X_1,\ldots,X_n)$ est incluse dans la tribu engendrée par $X_1,\ldots,X_n$. ■

Nous verrons au chapitre 7 la conséquence principale de l'indépendance. Cette conséquence est que, si X, Y sont des v.a.r. indépendantes, la loi du couple (X, Y) est le produit des lois P_X et P_Y (c'est-à-dire , avec les notations du Chapitre 7, $P_{(X,Y)} = P_X \otimes P_Y$). Une propriété analogue est vraie pour une famille $(X_1, \ldots, X_n)$ de v.a.r. indépendantes.

Nous terminons ce paragraphe par un théorème très utile en probabilités sur la représentation d'une v.a. mesurable par rapport à une autre v.a..

Théorème 3.31 (V.a. mesurable par rapport à une autre v.a.)
Soient X et Y deux v.a. réelles définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$. Alors, la v.a. Y est mesurable par rapport à la tribu engendrée par X (notée $\tau(X)$ ou $\sigma(X)$) si et seulement si il existe une fonction borélienne f de $\mathbb{R}$ dans $\mathbb{R}$ telle que $Y = f(X)$.

DÉMONSTRATION – La démonstration de ce résultat fait l'objet de l'exercice 3.17. Il est intéressant de remarquer que la démonstration de ce théorème effectuée dans l'exercice 3.17 donne les informations complémentaires suivantes :

- Y est $\tau(X)$-mesurable bornée si et seulement s'il existe f borélienne bornée t.q. $Y = f(X)$,
- Y est $\tau(X)$-mesurable positive si et seulement s'il existe f borélienne positive t.q. $Y = f(X)$.

La partie "si" de ces deux résultats est immédiate. Pour la partie "seulement si", il suffit de remarquer que la démonstration faite dans l'exercice 3.17 donne f t.q.

$$\mathrm{Im}(f) = \{f(t), t \in \mathbb{R}\} \subset \mathrm{Im}(Y) \cup \{0\}, \text{ avec } \mathrm{Im}(Y) = \{Y(\omega), \omega \in \Omega\}.$$

■

3.5 Convergence p.p., p.s., en mesure, en probabilité

On introduit ici plusieurs notions de convergence de fonctions définies sur un espace mesuré à valeurs dans $\mathbb{R}$ (ou $\overline{\mathbb{R}}_+$) et on donne des liens entre ces différentes convergences. On introduit les notions équivalentes pour les variables aléatoires en langage probabiliste.

Définition 3.32 (Égalité presque partout) *Soient (E, T, m) un espace mesuré, F un ensemble et f et g des fonctions définies de E dans F ($F = \mathbb{R}$ ou $F = \overline{\mathbb{R}}_+$, par exemple) ; on dit que $f = g$ m-presque partout (et on note $f = g$ m-p.p.) si l'ensemble $\{x \in E ; f(x) \neq g(x)\}$ est négligeable, c'est-à-dire qu'il existe $A \in T$ tel que $m(A) = 0$ et $f(x) = g(x)$ pour tout $x \in A^c$.*

On peut remarquer que si f et g sont des fonctions mesurables de E (muni de la tribu T et de la mesure m) dans $\mathbb{R}$ (ou $\overline{\mathbb{R}}_+$), l'ensemble $\{x \in E ; f(x) \neq g(x)\}$ (noté aussi $\{f \neq g\}$) appartient à T. Le fait que $f = g$ m p.p. revient donc à dire que $m(\{f \neq g\}) = 0$. Dans la cas où f ou g n'est pas mesurable, l'ensemble $\{f \neq g\}$ peut être négligeable sans appartenir à T (il appartient nécessairement à T si la mesure est complète, voir la définition 2.26).

En l'absence de confusion possible, on remplace m-p.p. par p.p.. Cette définition se traduit en langage probabiliste par :

Définition 3.33 (Égalité presque sûre) *Soient (E, T, p) un espace probabilisé, X et Y des variables aléatoires réelles. On dit que $X = Y$ presque sûrement (et on note $X = Y$ p.s.), si l'ensemble $\{x \in E ; X(x) \neq Y(x)\}$ est négligeable.*

Définition 3.34 (Convergence presque partout) *Soient* (E, T, m) *un espace mesuré,* F *un ensemble,* $(f_n)_{n\in\mathbb{N}}$ *une suite de fonctions de* E *dans* F *et* f *une fonction de* E *dans* F *(*$F = \mathbb{R}$ *ou* $F = \overline{\mathbb{R}}_+$*, par exemple) ; on dit que* f_n *converge presque partout vers* f *(*$f_n \to f$ p.p.*) s'il existe une partie* A *de* E*, négligeable, t.q., pour tout élément* x *de* A^c*, la suite* $(f_n(x))_{n\in\mathbb{N}}$ *converge vers* $f(x)$*.*

Noter que la convergence simple entraîne la convergence presque partout.

La définition 3.34 se traduit en langage probabiliste par :

Définition 3.35 (Convergence presque sûre) *Soient* (E, T, p) *un espace probabilisé,* $(X_n)_{n\in\mathbb{N}}$ *une suite de variables aléatoires réelles et* X *une variable aléatoire réelle. On dit que* X_n *converge presque sûrement vers* X *(*$X_n \to X$ *p.s.) s'il existe une partie* A *de* E*, négligeable, t.q., pour tout élément* x *de* A^c*, la suite* $(X_n(x))_{n\in\mathbb{N}}$ *converge vers* $X(x)$*.*

Définition 3.36 (Convergence presque uniforme) *Soient* (E, T, m) *un espace mesuré,* $(f_n)_{n\in\mathbb{N}} \subset \mathcal{M}$ *et* $f \in \mathcal{M}$*. On dit que* f_n *converge presque uniformément vers* f *(*$f_n \to f$ *p.unif.) si, pour tout* $\varepsilon > 0$*, il existe* $A \in T$ *tel que* $m(A) \le \varepsilon$ *et* f_n *converge uniformément vers* f *sur* A^c*.*

La convergence presque uniforme entraîne la convergence presque partout (voir exercice 3.26).
Attention, la convergence presque uniforme ne donne pas la convergence uniforme en dehors d'un ensemble de mesure nulle. La convergence uniforme en dehors d'un ensemble de mesure nulle est reliée à la convergence essentiellement uniforme, c'est-à-dire la convergence pour le sup essentiel, défini ci-après, ou pour la norme $\|\cdot\|_\infty$ que nous verrons dans la section 6.1.2.

Définition 3.37 (Sup essentiel) *Soient* (E, T, m) *un espace mesuré et* $f \in \mathcal{M}$*. On dit que* f *est essentiellement bornée si il existe* $C \in \mathbb{R}_+$ *tel que* $|f| \le C$ *p.p.. On appelle alors sup essentiel de* $|f|$*, et on le note* $\|f\|_\infty$*, l'infimum des valeurs* C *telles que* $|f| \le C$ *p.p.. Si* f *n'est pas essentiellement bornée, on pose* $\|f\|_\infty = \infty$*.*

Remarquons que dans le cas où $(E, T, m) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, le sup essentiel d'une fonction continue est la borne supérieure de sa valeur absolue (ceci fait l'objet de la proposition 6.18).

Définition 3.38 (Convergence essentiellement uniforme) *Soit* (E, T, m) *un espace mesuré,* $(f_n)_{n\in\mathbb{N}}$ *une suite de* $\mathcal{M}$ *et* $f \in \mathcal{M}$*. On dit que* f_n *converge essentiellement uniformément vers* f *(*$f_n \to f$ *ess. unif.) si* $\|f_n - f\|_\infty \to 0$ *lorsque* $n \to +\infty$*.*

Il est facile de voir que la convergence essentiellement uniforme entraîne la convergence presque uniforme, mais la réciproque est fausse (voir l'exercice 3.27). Le théorème suivant donne, dans le cas où la mesure est finie, un résultat très important qui fait le lien entre la convergence presque partout et la convergence presque uniforme.

Théorème 3.39 (Egorov) *Soient* (E, T, m) *un espace mesuré, tel que* $m(E) < +\infty$, $(f_n)_{n\in\mathbb{N}} \subset \mathcal{M}$ *et* $f \in \mathcal{M}$. *On suppose que* $f_n \to f$ *p.p.. Alors, pour tout* $\varepsilon > 0$, *il existe* $A \in T$ *tel que* $m(A) \leq \varepsilon$ *et* f_n *converge uniformément vers* f *sur* A^c. *(Autrement dit, la suite* $(f_n)_{n\in\mathbb{N}}$ *converge presque uniformément vers* f.*)*

La démonstration de ce théorème fait l'objet de l'exercice 3.27. Attention, lorsque $m(E) = +\infty$, on peut trouver des suites de fonctions qui convergent presque partout et non presque uniformément.

Définition 3.40 (Convergence en mesure) *Soient* (E, T, m) *un espace mesuré,* $(f_n)_{n\in\mathbb{N}} \subset \mathcal{M}$ *et* $f \in \mathcal{M}$. *On dit que* f_n *converge en mesure vers* f *si :*

$$\forall \varepsilon > 0, \lim_{n\to+\infty} m(\{x \in E \ ; |f(x) - f_n(x)| \geq \varepsilon\}) = 0.$$

Cette définition se traduit en langage probabiliste par :

Définition 3.41 (Convergence en probabilité) *Soient* (E, T, p) *un espace probabilisé,* $(X_n)_{n\in\mathbb{N}}$ *une suite de variables aléatoires réelles et* X *une variable aléatoire réelle . On dit que* X_n *converge en probabilité vers* X *si :*

$$\forall \varepsilon > 0, \lim_{n\to+\infty} p(\{x \in X_n \ ; |X(x) - X_n(x)| \geq \varepsilon\}) = 0.$$

On peut montrer (cf exercice 3.25) que si $(f_n)_{n\in\mathbb{N}} \subset \mathcal{M}$ converge en mesure vers $f \in \mathcal{M}$ et $(f_n)_{n\in\mathbb{N}}$ converge en mesure vers $g \in \mathcal{M}$, alors $f = g$ p.p.. On montre aussi que si $(f_n)_{n\in\mathbb{N}} \subset \mathcal{M}$ converge en mesure vers $f \in \mathcal{M}$ et $(g_n)_{n\in\mathbb{N}}$ converge en mesure vers $g \in \mathcal{M}$, alors $(f_n + g_n)_{n\in\mathbb{N}} \subset \mathcal{M}$ converge en mesure vers $f + g \in \mathcal{M}$, et, si m est une mesure finie, $(f_n g_n)_{n\in\mathbb{N}} \subset \mathcal{M}$ converge en mesure vers $f g \in \mathcal{M}$.
On montre à l'aide du théorème d'Egorov que si f_n converge vers f presque partout, et si $m(E) < +\infty$, alors f_n converge vers f en mesure. Réciproquement, si f_n converge vers f en mesure, alors il existe une sous-suite de $(f_n)_{n\in\mathbb{N}}$ qui converge vers f presque uniformément (et donc presque partout). Ce second résultat est vrai même si $m(E) = +\infty$ (voir exercice 3.28).

On donne maintenant un résumé des différents types de convergence connus jusqu'à présent avec les relations existantes entre eux. Les relations entre convergence presque partout et convergence en mesure (resp. convergence presque sûre et convergence en probabilité) sont étudiées dans l'exercice 3.28. (On en introduira bientôt encore quelques-unes)

Terminologie analyste	**Terminologie probabiliste**
convergence simple (cs)	
convergence uniforme (cu)	
convergence presque partout (cpp)	convergence presque sûre (cps)
convergence presque uniforme (cpu)	
convergence en mesure (cm)	convergence en probabilité (cp)

On a les implications suivantes :

Terminologie analyste	**Terminologie probabiliste**
(cu) $\Rightarrow$ (cs) $\Rightarrow$ (cpp)	
(cu) $\Rightarrow$ (cpu) $\Rightarrow$ (cpp)	
(cpp) $\Rightarrow$ (cpu) si la mesure est finie	
(cm) $\Rightarrow$ (cpu) pour une sous-suite	(cp) $\Rightarrow$ (cps) pour une sous-suite
(cpp) $\Rightarrow$ (cm) si la mesure est finie	(cps) $\Rightarrow$ (cp)
(cpu) $\Rightarrow$ (cm)	