

# Quelques rappels de cours pour le soutien de théorie des probabilités

Allan Merino

## Table des matières

1	Rappels généraux de mesure	1
2	Application aux probabilités	5

## 1 Rappels généraux de mesure

Pour commencer, nous allons rappeler quelques notions fondamentales de théorie de la mesure vue au premier semestre. Commençons par des rappels assez élémentaires.

**Définition 1.1.** Soit  $X$  un ensemble. On appelle tribu (ou  $\sigma$ -algèbre) sur  $X$  toute famille  $\mathcal{A}$  de parties de  $X$  vérifiant :

1.  $\emptyset \in \mathcal{A}$ ,
2. Si  $A \in \mathcal{A}$ , alors  $A^c \in \mathcal{A}$ ,
3. Si  $A_n \in \mathcal{A}$ ,  $n \geq 1$ , alors,  $\bigcup_{n \geq 1} A_n \in \mathcal{A}$ .

Un espace mesurable est un couple  $(X, \mathcal{A})$ , où  $X$  est un ensemble et  $\mathcal{A}$  une tribu sur  $X$ .

**Exemple 1.2.** 1. Soit  $X$  un ensemble et  $A \subseteq X$ ,  $A \neq \emptyset$ ,  $A \neq X$ . Alors,  $\mathcal{A} = \{\emptyset, X, A, A^c\}$  forme une tribu sur  $X$ .

2. Soit  $(\mathcal{A}_i)_{i \in I}$  une famille quelconque de tribus sur  $X$ ,  $I \neq \emptyset$ , alors,

$$\mathcal{A} := \bigcap_{i \in I} \mathcal{A}_i$$

forme une tribu sur  $X$ .

3. Soit  $\mathcal{E} \subseteq \mathcal{P}(X)$  une famille de parties de  $X$ . Il existe une plus petite tribu (au sens de l'inclusion) contenant  $\mathcal{E}$ . On la note  $\sigma(\mathcal{E})$  : c'est la tribu engendrée par  $\mathcal{E}$ . La tribu engendrée par  $\mathcal{E}$  est en fait l'intersection de toutes les tribus contenant  $\mathcal{E}$ .

4. Un cas particulier de tribu engendrée est la tribu borélienne. Soit  $(X, \mathcal{O}(X))$  un espace topologique. La tribu borélienne de  $X$ , également appelée tribu des boréliens de  $X$ , est définie par :

$$\mathcal{B}(X) = \sigma(\mathcal{O}(X)). \quad (1)$$

On s'intéressera par la suite plus précisément au cas  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .

5. Soit  $f : X \rightarrow Y$  et  $\mathcal{B}$  une tribu sur  $Y$ . Alors,

$$\mathcal{A} := \{f^{-1}(B), B \in \mathcal{B}\}$$

est une tribu sur  $X$ . Cette tribu est appelée tribu image-réciproque. On la note  $f^{-1}(\mathcal{B})$ .

6. Soit  $f : X \rightarrow Y$  et  $\mathcal{A}$  une tribu sur  $X$ . On appelle tribu image de  $\mathcal{A}$  par  $f$ , la tribu sur  $Y$  définie par :

$$\mathcal{B} := \{B \in \mathcal{P}(Y) / f^{-1}(B) \in \mathcal{A}\}.$$

**Lemme 1.3** (Lemme de transport). Soit  $f : X \rightarrow Y$  et  $\mathcal{E} \subseteq \mathcal{P}(Y)$ . Alors,

$$\sigma(f^{-1}(\mathcal{E})) = f^{-1}(\sigma(\mathcal{E})). \quad (2)$$

**Définition 1.4.** Soient  $(X, \mathcal{A})$  et  $(Y, \mathcal{B})$  deux espaces mesurables. Une fonction  $f : (X, \mathcal{A}) \rightarrow (Y, \mathcal{B})$  est dite mesurable si pour tout  $B \in \mathcal{B}$ , on a  $f^{-1}(B) \in \mathcal{A}$ .

**Exemple 1.5.** Soit  $(X, \mathcal{A})$  un espace mesurable. Soit  $A \subseteq X$ . On définit  $\mathbb{1}_A : (X, \mathcal{A}) \rightarrow (\{0, 1\}, \mathcal{P}(\{0, 1\}))$  par :

$$\mathbb{1}_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon} \end{cases}$$

On constate que la fonction  $\mathbb{1}_A$  est mesurable si et seulement si  $A \in \mathcal{A}$ .

**Définition 1.6.** Soit  $(X, \mathcal{A})$  un espace mesurable. On appelle mesure (positive) sur  $(X, \mathcal{A})$  toute application  $\mu : \mathcal{A} \rightarrow \bar{\mathbb{R}}_+$  vérifiant :

1.  $\mu(\emptyset) = 0$ ,
2. Si  $(A_n)_{n \geq 1}$  est une suite d'éléments de  $\mathcal{A}$ , deux à deux disjoints (éventuellement vides) :

$$\mu\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n \geq 1} \mu(A_n). \quad (3)$$

Si  $\mu(X) < +\infty$ , la mesure est dite finie ou bornée, si  $\mu(X) = 1$ ,  $\mu$  est une mesure de probabilité.

On utilisera par la suite principalement deux mesures, à savoir la mesure de Lebesgue et la mesure de comptage.

**Définition 1.7.** Un espace mesuré est un triplet  $(\Omega, \mathcal{F}, \mu)$ , où  $(\Omega, \mathcal{F})$  est un espace mesurable et  $\mu$  une mesure sur  $(\Omega, \mathcal{F})$ .

**Lemme 1.8** (La mesure image). Soient  $(X, \mathcal{A})$  et  $(Y, \mathcal{B})$  deux espaces mesurables et une fonction mesurable  $h : (X, \mathcal{A}) \rightarrow (Y, \mathcal{B})$ . Si  $\mu$  est une mesure sur  $(X, \mathcal{A})$ , l'application  $\nu : \mathcal{B} \rightarrow \bar{\mathbb{R}}_+$  définie par :

$$\nu(B) := \mu(h^{-1}(B)) \quad (4)$$

est une mesure sur  $(Y, \mathcal{B})$ , appelée mesure image de  $\mu$  par  $h$ , de même masse que  $\mu$ .

Ainsi,  $(Y, \mathcal{B}, \nu)$  forme un espace mesuré.

*Démonstration.* On vérifie directement les deux axiomes d'une mesure. Clairement, on a  $\nu(\emptyset) = \mu(h^{-1}(\emptyset)) = \mu(\emptyset) = 0$ . A présent, considérons  $(B_n)_{n \geq 1}$  une suite d'éléments de  $\mathcal{B}$  qui sont deux à deux disjoints. Pour tout  $i, j / i \neq j$ , on a  $h^{-1}(B_i) \cap h^{-1}(B_j) = h^{-1}(B_i \cap B_j) = h^{-1}(\emptyset) = \emptyset$ . On a alors que  $(h^{-1}(B_n))_{n \geq 1}$  forme une suite d'éléments de  $\mathcal{A}$  deux à deux disjoints. De plus, on a  $\bigcup_{n \geq 1} h^{-1}(B_n) = h^{-1}(\bigcup_{n \geq 1} B_n)$ . Ainsi, on a :

$$\nu\left(\bigcup_{n \geq 1} B_n\right) = \mu\left(h^{-1}\left(\bigcup_{n \geq 1} B_n\right)\right) = \mu\left(\bigcup_{n \geq 1} h^{-1}(B_n)\right) = \sum_{n \geq 1} \mu(h^{-1}(B_n)) = \sum_{n \geq 1} \nu(B_n).$$

Ainsi,  $\nu$  est bien une mesure sur  $(Y, \mathcal{B})$ . De plus, on a :

$$\nu(Y) = \mu(h^{-1}(Y)) = \mu(X).$$

□

**Théorème 1.9** (Théorème de transfert). Soit  $(X, \mathcal{A}, \mu)$  un espace mesuré,  $(Y, \mathcal{B})$  un espace mesurable et  $h : (X, \mathcal{A}) \rightarrow (Y, \mathcal{B})$  une application mesurable. Notons  $\nu$  la mesure image de  $\mu$  par  $h$ . Soit  $f : (Y, \mathcal{B}) \rightarrow \mathbb{C}$  une fonction mesurable. Alors,  $f$  est  $\nu$ -intégrable (i.e.  $f \in L^1(Y, \nu)$ ) si et seulement si  $f \circ h$  est  $\mu$ -intégrable, et dans ce cas, on a :

$$\int_Y f d\nu = \int_X f \circ h d\mu. \quad (5)$$

*Idée de preuve.* On commence par montrer cela pour des fonctions simples. On prend  $f = \mathbb{1}_B$  avec  $B \in \mathcal{B}$ . On a :

$$\int_Y f d\nu = \int_Y \mathbb{1}_B(x) d\nu(x) = \nu(B) = \mu(h^{-1}(B)) = \int_X \mathbb{1}_{h^{-1}(B)}(x) d\mu(x).$$

Or, on a  $\mathbb{1}_{h^{-1}(B)} = \mathbb{1}_B \circ h$ . Ainsi, on a :

$$\int_X \mathbb{1}_{h^{-1}(B)}(x) d\mu(x) = \int_X (\mathbb{1}_B \circ h)(x) d\mu(x),$$

ce qui implique

$$\int_Y f d\nu = \int_X f \circ h d\mu.$$

Cela est donc vrai pour les fonctions étagées, et en utilisant des résultats de densité puis le théorème de Beppo-Lévi, on obtient le résultat.

□

Nous allons à présent parler de la notion de mesure à densité, qui nous donnera par la suite la possibilité de construire des mesures de probabilités.

Soit  $(X, \mathcal{A}, \mu)$  un espace mesuré. Soit  $f : (X, \mathcal{A}) \rightarrow \mathbb{R}_+$ . On pose, pour tout  $A \in \mathcal{A}$ ,

$$\nu(A) := \int_A f d\mu. \quad (6)$$

**Proposition 1.10.** *L'application  $\nu$  est une mesure sur  $\mathcal{A}$ . De plus, cette dernière vérifie :*

$$(\forall A \in \mathcal{A}), \mu(A) = 0 \Rightarrow \nu(A) = 0. \quad (7)$$

*Démonstration.* Commençons par montrer que  $\nu$  est une mesure. Remarquons tout d'abord que vu que  $f$  est à valeurs positives,  $\nu : \mathcal{A} \rightarrow \overline{\mathbb{R}}_+$ . Clairement, on a :

$$\nu(\emptyset) = \int_{\emptyset} f(x) d\mu(x) = \int_X \mathbb{1}_{\emptyset}(x) f(x) d\mu(x) = \int_X 0 d\mu(x) = 0.$$

A présent, considérons  $(A_n)_{n \geq 1}$  une famille d'éléments de  $\mathcal{A}$  deux à deux disjoints. On a :

$$\nu\left(\bigcup_{n \geq 1} A_n\right) = \int_{\bigcup_{n \geq 1} A_n} f d\mu = \int_X \mathbb{1}_{\bigcup_{n \geq 1} A_n} f d\mu.$$

Or, vu que les  $A_n$  sont deux à deux disjoints, on a  $\mathbb{1}_{\bigcup_{n \geq 1} A_n} = \sum_{n \geq 1} \mathbb{1}_{A_n} = \lim_{n \rightarrow +\infty} \sum_{i=1}^n \mathbb{1}_{A_i}$ . Ainsi, vu que  $f$  est à valeurs positives, en utilisant le théorème de Beppo-Levi, on conclue :

$$\nu\left(\bigcup_{n \geq 1} A_n\right) = \int_X \lim_{n \rightarrow +\infty} \sum_{i=1}^n \mathbb{1}_{A_i} f d\mu = \lim_{n \rightarrow +\infty} \sum_{i=1}^n \int_X \mathbb{1}_{A_i} f d\mu = \sum_{n \geq 1} \nu(A_n).$$

□

**Notation 1.11.** En reprenant les notations précédentes, on note parfois, par abus de notation,  $\nu = f \cdot \mu$ .

**Définition 1.12.** 1. Si deux mesures  $\mu$  et  $\nu$  vérifient la relation (7), alors, la mesure  $\nu$  est dite absolument continue par rapport à  $\mu$  et on note  $\nu \ll \mu$ .

2. Si  $\nu = f \cdot \mu$ ,  $f$  est appelée selon les cas densité de  $\nu$  par rapport à  $\mu$  ou dérivée de Radon-Nikodym de  $\nu$  par rapport à  $\mu$ . On la note souvent  $\frac{d\nu}{d\mu}$ .

Pour conclure cette partie de rappels, nous allons voir un théorème qui peut être vu, sous certaines hypothèses, comme une réciproque de la proposition précédente. C'est le théorème de Radon-Nikodym.

**Théorème 1.13** (Théorème de Radon-Nikodym). *Soit  $\mu$  une mesure positive  $\sigma$ -finie et  $\nu$  une mesure réelle telle que  $\nu \ll \mu$ , alors, il existe une fonction  $f$  (unique à équivalence près) de  $L^1(\Omega, \mathcal{F}, \mu)$  telle que :*

$$(\forall A \in \mathcal{F}), \nu(A) = \int_A f(x) d\mu(x). \quad (8)$$

*De plus, si  $\nu$  est positive, alors,  $f$  est positive.*

*Remarque 1.14.* La fonction  $f$  est appelée dérivée de Radon-Nikodym de  $\nu$  par rapport à  $\mu$  et est notée  $f = \frac{d\nu}{d\mu}$ . De plus, si  $\nu$  et  $\mu$  sont des mesures de probabilité, on a  $\int_{\Omega} f d\mu = 1$ .

Les hypothèses du théorème sont nécessaires. En effet, en toute généralité, si  $\mu$  et  $\nu$  sont deux mesures quelconques telles que  $\nu \ll \mu$ , alors  $\nu$  n'admet pas nécessairement une densité par rapport à  $\mu$ .

**Contre-exemple 1.15.** Pour un contre-exemple, on pourra, par exemple, consulter [4], page 200.

Pour terminer cette partie, nous allons rappeler le théorème de changement de variables. Mais avant cela, introduisons quelques notations. L'idée de ce théorème est de transporter le domaine d'intégration  $D$  sur un autre ouvert  $\Delta$  homéomorphe à  $D$ .

Si  $D$  est un ouvert de  $\mathbb{R}^d$ , alors  $\mathcal{B}(D) := \{B \in \mathcal{B}(\mathbb{R}^d) / B \subseteq D\}$  (car  $D$  est un borélien en tant qu'ouvert). Par définition, on notera  $\lambda_D := 1_D \cdot \lambda_d$  la restriction à  $D$  de la mesure de Lebesgue  $\lambda_d$  sur  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ , appelée mesure de Lebesgue sur  $D$ . Soit  $\phi : \Delta \rightarrow D$  une application différentiable d'un ouvert  $\Delta$  de  $\mathbb{R}^d$  à valeurs dans  $D$ . En tout point  $u \in \Delta$ , la dérivée  $\phi'(u)$  est une application linéaire de  $\mathbb{R}^d$  dans  $\mathbb{R}^d$ . On appelle Jacobien de  $\phi$  au point  $u$  la quantité  $J_{\phi}(u) = \det(\phi'(u))$ .

Par définition, une application  $\phi : \Delta \rightarrow D$  est un  $\mathcal{C}^1$ -difféomorphisme si  $\phi$  est bijective, de classe  $\mathcal{C}^1$  sur  $\Delta$  (i.e. continûment différentiable) et si  $\phi^{-1}$  est de classe  $\mathcal{C}^1$  sur  $D$ .

Il existe plusieurs manières équivalentes pour formuler le résultat suivant. En voici probablement la forme la plus simple.

**Théorème 1.16.** Soit  $\phi$  un  $\mathcal{C}^1$ -difféomorphisme entre deux ouverts  $\Delta$  et  $D$  de  $\mathbb{R}^d$ . Alors, pour toute fonction borélienne  $f : D \rightarrow \mathbb{K}$  ( $\mathbb{K} = \mathbb{R}$  ou  $\mathbb{C}$ ),  $f$  est  $\lambda_D$ -intégrable sur  $D$  si et seulement si  $(f \circ \phi)|_{J_{\phi}}$  est  $\lambda_{\Delta}$ -intégrable sur  $\Delta$  et, dans ce cas, on a :

$$\int_D f(x) d\lambda_D(x) = \int_{\Delta} f(\phi(u)) |J_{\phi}(u)| d\lambda_{\Delta}(u). \quad (9)$$

## 2 Application aux probabilités

Nous allons reprendre à présent les idées précédentes. Nous allons juste à présent donner du vocabulaire. Considérons  $(\Omega, \mathcal{F})$  un espace probabilisable (mesurable). L'ensemble  $\Omega$  est appelé univers. Ses éléments  $\omega$  sont appelés des épreuves. Tout élément de  $\mathcal{F}$  est appelé évènement.

**Définition 2.1.** Soit  $(\Omega, \mathcal{F})$  un espace probabilisable. On dit que  $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$  est une mesure de probabilité si  $\mathbb{P}$  est une mesure sur  $(\Omega, \mathcal{F})$  telle que  $\mathbb{P}(\Omega) = 1$ .

Un espace probabilisé est donc un triplet  $(\Omega, \mathcal{F}, \mathbb{P})$ , où  $\mathbb{P}$  est une mesure sur  $(\Omega, \mathcal{F})$ .

**Définition 2.2** (La notion de variables aléatoires). Soient  $(\Omega, \mathcal{F})$  et  $(E, \mathcal{B})$  deux espaces probabilisables. On appelle variable aléatoire définie sur  $(\Omega, \mathcal{F})$ , à valeur dans  $(E, \mathcal{B})$ , toute application mesurable de  $(\Omega, \mathcal{F})$  dans  $(E, \mathcal{B})$ , i.e. toute application  $X : \Omega \rightarrow E$  telle que

$$(\forall B \in \mathcal{B}), X^{-1}(B) \in \mathcal{F}. \quad (10)$$

**Définition 2.3** (Loi de probabilité). Soit  $X$  une variable aléatoire sur  $(\Omega, \mathcal{F}, \mathbb{P})$  à valeurs dans  $(E, \mathcal{B})$ . On appelle loi de probabilité de la variable aléatoire  $X$  la probabilité  $\mathbb{P}_X$  sur  $(E, \mathcal{B})$  image de  $\mathbb{P}$  par  $X$ .

**Définition 2.4.** Une variable aléatoire  $X$  définie sur  $(\Omega, \mathcal{F}, \mathbb{P})$  est discrète si et seulement si  $X(\Omega)$  est dénombrable presque sûrement.

**Proposition 2.5.** Une variable aléatoire  $X$  est discrète si et seulement si sa loi de probabilité est atomique, i.e.

$$\mathbb{P}_X = \sum_{x \in X(\Omega)} \mathbb{P}(X = x) \delta_x. \quad (11)$$

**Définition 2.6.** Une variable aléatoire  $X$  à valeurs dans  $\mathbb{R}^n$  est dite absolument continue (par rapport à la mesure de Lebesgue) si et seulement si sa loi de probabilité  $\mathbb{P}_X$  est absolument continue par rapport à la mesure de Lebesgue  $\lambda_n$  sur  $\mathbb{R}^n$ , i.e. il existe une fonction  $f_X : \mathbb{R}^n \rightarrow \mathbb{R}^+$  telle que

1.  $\frac{d\mathbb{P}_X}{d\lambda_n} = f_X$ ,
2.  $\int_{\mathbb{R}^n} f_X(x) dx = 1$ .

*Remarque 2.7.* Contrairement au cas des variables aléatoires discrètes, la loi de probabilité d'une variable aléatoire absolument continue ne possède pas d'atomes, i.e.  $(\forall x \in \mathbb{R}^n), \mathbb{P}_X(\{x\}) = 0$ .

Nous allons à présent donner des exemples de probabilité dans le cas discret et continu.

**Exemple 2.8.** Commençons naturellement par le cas discret. Commençons par les lois dites uniformes. Soit  $n \in \mathbb{N}^*$  et considérons  $\Omega = \{1, \dots, n\}$ . On a que  $(\Omega, \mathcal{P}(\Omega))$  forme un espace probabilisable. Posons  $\mathbb{P} : \Omega \rightarrow [0, 1]$  définie par :

$$\mathbb{P}(\{i\}) = \frac{1}{n}. \quad (12)$$

Pour reprendre les notations précédentes, on a  $X : (\Omega, \mathcal{P}(\Omega)) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  la variable aléatoire en question. On a une mesure  $\mathbb{P}$  sur  $(\Omega, \mathcal{P}(\Omega))$ . On définit alors la mesure image  $\mathbb{P}_X$  sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  par  $\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)), B \in \mathcal{B}(\mathbb{R})$ . Plus précisément, pour tout  $B \in \mathcal{B}(\mathbb{R})$ , on a :

$$\mathbb{P}_X(B) = \mathbb{P}(B \cap \{1, \dots, n\}).$$

Nous allons à présent donner un exemple où l'espace  $\Omega$  est dénombrable, mais infini : la loi de Poisson. Fixons  $\Omega = \mathbb{N}$  et soit  $\lambda > 0$ . Sur l'espace probabilisé  $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ , on définit la mesure de probabilité  $\mathbb{P} : \Omega \rightarrow [0, 1]$  par :

$$\mathbb{P}(\{i\}) = e^{-\lambda} \frac{\lambda^i}{i!}. \quad (13)$$

On a clairement :

$$\sum_{k=0}^{+\infty} \mathbb{P}(\{k\}) = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^\lambda = 1.$$

Si on prend l'application  $X : (\mathbb{N}, \mathcal{P}(\mathbb{N}), \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  définie par  $X(i) = i$ , alors  $X$  est une application mesurable. On définit alors la loi  $\mathbb{P}_X$  sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  par :

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)), B \in \mathcal{B}(\mathbb{R}). \quad (14)$$

Plus précisément, on a :

$$\mathbb{P}_X(B) = \mathbb{P}(B \cap \mathbb{N}) = \sum_{i \in \mathbb{N} \cap B} \mathbb{P}(X = i) = e^{-\lambda} \sum_{i \in \mathbb{N} \cap B} \frac{\lambda^i}{i!}. \quad (15)$$

Il existe évidemment d'autres lois discrètes ; parmi les lois "connues", on peut citer la loi de Bernoulli, la loi binomiale et la loi géométrique. Passons à présent au cas continu. Nous allons ici en citer 3.

### 1. La loi uniforme.

Une variable aléatoire  $X$  est de loi uniforme sur  $[a, b]$  ( $a < b$ ) si et seulement si  $X$  est absolument continue de dérivée de Radon-Nikodym :

$$\frac{d\mathbb{P}_X}{dx} = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x). \quad (16)$$

### 2. La loi exponentielle.

Une variable aléatoire  $X$  est dite de loi exponentielle de paramètre  $\lambda$  ( $\lambda > 0$ ) si et seulement si  $X$  est absolument continue de dérivée de Radon-Nikodym :

$$\frac{d\mathbb{P}_X}{dx} = \lambda e^{-\lambda x} \mathbb{1}_{[0,+\infty[}(x). \quad (17)$$

### 3. La loi normale.

Une variable aléatoire  $X$  est de loi normale de paramètre 0 et 1 si et seulement si elle est absolument continue de dérivée de Radon-Nikodym :

$$\frac{d\mathbb{P}_X}{dx} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \quad (18)$$

Une variable aléatoire  $X$  est de loi normale de paramètre  $m$  et  $\sigma^2$  ( $m \in \mathbb{R}, \sigma > 0$ ) si et seulement si elle est absolument continue de dérivée de Radon-Nikodym :

$$\frac{d\mathbb{P}_X}{dx} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}. \quad (19)$$

Pour tout ce qui est espérance et variance, on regardera cela en exercice.

Il existe différentes méthodes pour construire des mesures de probabilité. La proposition suivante donne une première solution.

**Proposition 2.9.** *Toute combinaison convexe de mesures de probabilités sur  $(\Omega, \mathcal{A})$  est encore une mesure de probabilité. Autrement dit, si  $p \in [0, 1]$  et  $\mathbb{P}_1, \mathbb{P}_2$  sont deux mesures de probabilités sur  $(\Omega, \mathcal{A})$ , alors l'application  $\mathbb{P} = p\mathbb{P}_1 + (1 - p)\mathbb{P}_2$  définie par  $\mathbb{P}(A) = p\mathbb{P}_1(A) + (1 - p)\mathbb{P}_2(A)$ ,  $A \in \mathcal{A}$ , est une mesure de probabilité sur  $\mathcal{A}$ .*

Cette proposition montre en particulier que l'on peut construire des mesures de probabilité qui ont une partie discrète et une partie à densité. Ces mesures de probabilités sont dites mixtes. Citons à présent quelques résultats concernant les mesures de probabilités.

**Proposition 2.10.** *Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé. On a :*

1. Pour tout  $A \in \mathcal{A}$ , on a  $\mathbb{P}(A) + \mathbb{P}(A^c) = 1$ ,
2. Si  $A$  et  $B$  sont deux évènements tels que  $A \subseteq B$ , alors  $\mathbb{P}(A) \leq \mathbb{P}(B)$  et  $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$ ,
3. Si  $A$  et  $B$  sont deux évènements, alors

$$\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B), \quad (20)$$

4. Si  $I = \mathbb{N}$  ou  $[[1, N]]$  et  $(A_n)_{n \in I}$  est un système complet d'évènements (i.e. disjoints deux à deux et de réunion  $\Omega$ ), alors :

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A \cap A_i) \quad (21)$$

*Idée de preuve.* 1. On a  $\Omega = A \cup A^c$ . Donc,  $1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c)$ .

2. On écrit  $B = A \cup (B \setminus A)$ , ainsi,  $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$ , et on conclue en remarquant que  $\mathbb{P}(B \setminus A) \geq 0$ .
3. On a  $A \cup B = (A \cap B^c) \cup (A \cap B) \cup (B \cap A^c)$ . De même, on a :

$$A = (A \cap B^c) \cup (A \cap B) \quad \text{et} \quad B = (B \cap A^c) \cup (A \cap B).$$

Ainsi, on obtient :

$$\mathbb{P}(A) = \mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap B) \quad \text{et} \quad \mathbb{P}(B) = \mathbb{P}(A^c \cap B) + \mathbb{P}(A \cap B).$$

Finalement, on obtient :

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap B) + \mathbb{P}(B \cap A^c) \\ &= (\mathbb{P}(A) - \mathbb{P}(A \cap B)) + \mathbb{P}(A \cap B) + (\mathbb{P}(B) - \mathbb{P}(A \cap B)) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \end{aligned}$$

4. On remarque que  $A = A \cap \Omega$  et  $\Omega = \bigcup_{i \in I} A_i$ . Ainsi, on a :

$$\mathbb{P}(A) = \mathbb{P}(A \cap \Omega) = \mathbb{P}\left(\bigcup_{i \in I} (A \cap A_i)\right) = \sum_{i \in I} \mathbb{P}(A \cap A_i).$$

□



Nous allons à présent voir la notion d'espérance et de variance d'une variable aléatoire. Nous allons d'abord rappeler la définition dans le cas discret et dans le cas continu, et donner par la suite une réécriture de cette dernière à l'aide des résultats de mesure vu précédemment.

**Définition 2.11** (Pour le cas discret). Soit  $X$  un variable aléatoire discrète à valeur dans  $\mathbb{N}$ . Alors, l'espérance de  $X$  est défini par :

$$\mathbb{E}(X) = \sum_{n \in \mathbb{N}} n \mathbb{P}(X = n). \quad (22)$$

**Définition 2.12** (Pour le cas continu). Soit  $X$  une variable aléatoire qui admet une densité de probabilité, notée  $f_X$ , alors, on définit son espérance, notée  $\mathbb{E}(X)$ , comme

$$\mathbb{E}(X) = \int_{\mathbb{R}} x f_X(x) dx \quad (23)$$

*Remarque 2.13.* Il existe des variables aléatoires où l'espérance n'est pas définie. Prenons par exemple la loi de Cauchy. Posons  $f(x) = \frac{1}{\pi} \frac{1}{1+x^2} \mathbf{1}_{\mathbb{R}}(x)$ . On a alors  $\int_{\mathbb{R}} f(x) dx = 1$ . Mais, en utilisant le critère de Riemann, il vient directement que l'espérance n'est pas définie dans ce cas.

Définissons à présent de manière "générale" l'espérance, et remarquons en quoi cette définition englobe le cas discret et le cas continu.

**Définition 2.14.** Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace probabilisé et soit  $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  une variable aléatoire. On définit alors l'espérance de  $X$  par :

$$\mathbb{E}(X) = \int_{\Omega} X d\mathbb{P}. \quad (24)$$

*Remarque 2.15* (Le cas discret). Si  $\mathbb{P}_X$  est une mesure discrète sur  $\mathbb{N}$ , elle peut s'écrire :

$$\mathbb{P}_X = \sum_{n \in \mathbb{N}} \alpha_n \delta_n, \quad (25)$$

avec  $\mathbb{P}_X(\{n\}) = \mathbb{P}(X = n) = \alpha_n$ .

On a  $\int_{\Omega} X d\mathbb{P} = \int_{\Omega} X(\omega) d\mathbb{P}(\omega)$ . En utilisant le théorème 1.9 (théorème de transfert), on obtient que

$$\int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \int_{\mathbb{R}} x d\mathbb{P}_X(x). \quad (26)$$

On a alors :

$$\begin{aligned} \int_{\Omega} X(\omega) d\mathbb{P}(\omega) &= \int_{\mathbb{R}} x d\mathbb{P}_X(x) = \int_{\mathbb{R}} x \left( \sum_{n \in \mathbb{N}} \alpha_n d\delta_n(x) \right) \\ &= \sum_{n \in \mathbb{N}} \alpha_n \int_{\mathbb{R}} x d\delta_n(x) = \sum_{n \in \mathbb{N}} \alpha_n n \\ &= \sum_{n \in \mathbb{N}} n \mathbb{P}_X(\{n\}) \end{aligned}$$

Ce qui redonne la formule énoncée précédemment.

*Remarque 2.16 (Le cas continu).* Si  $\mathbb{P}_X$  est une mesure de densité  $f$  par rapport à la mesure de Lebesgue sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , elle peut s'écrire :

$$d\mathbb{P}_X(x) = f(x)d\lambda(x). \quad (27)$$

On obtient alors :

$$\begin{aligned} \mathbb{E}(X) &= \int_{\Omega} X d\mathbb{P} = \int_{\mathbb{R}} x d\mathbb{P}_X(x) \text{ (théorème de transport)} \\ &= \int_{\mathbb{R}} f(x)d\lambda(x) \text{ (théorème de Radon-Nikodym)} \end{aligned}$$

On peut aussi citer le résultat suivant :

**Proposition 2.17.** *Pour toute fonction  $\phi$  telle que  $\phi(X) \in \mathcal{L}^1$ , on a :*

$$\mathbb{E}(\phi(X)) = \int_{\Omega} \phi(X) d\mathbb{P} = \int_{\mathbb{R}} \phi(x) d\mathbb{P}_X(x) = \int_{\mathbb{R}} \phi(x) f(x) d\lambda(x).$$

Citons pour finir quelques propriétés de l'espérance.

**Proposition 2.18.** *Soient  $X$  et  $Y$  deux variables aléatoires à valeurs réelles ou bien toutes deux positives ou bien toutes deux intégrables. Soit  $a \in \mathbb{R}$ .*

1. *La variable aléatoire  $X$  est intégrable si et seulement si  $\mathbb{E}(|X|) < +\infty$ . De plus, on a :*

$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|). \quad (28)$$

2. *On a  $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$  et  $\mathbb{E}(aX) = a\mathbb{E}(X)$  (linéarité de l'espérance). En particulier,  $X + Y$  est intégrable si  $X$  et  $Y$  sont intégrables.*

3. *Si  $\mathbb{P}(X \leq Y) = 1$ , alors,  $\mathbb{E}(X) \leq \mathbb{E}(Y)$ .*

4. *Si  $\mathbb{P}(X = Y) = 1$ , alors,  $\mathbb{E}(X) = \mathbb{E}(Y)$ .*

5. *Si  $X$  est positive presque sûrement, alors,*

$$\mathbb{E}(X) = 0 \Leftrightarrow \mathbb{P}(X = 0) = 1. \quad (29)$$

**Définition 2.19.** Soit  $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  une variable aléatoire à valeur réelle. Lorsque la variable aléatoire  $X$  vérifie  $\mathbb{E}(X^2) < +\infty$ , alors, la variance de  $X$  est le nombre réel positif, noté  $\mathbb{V}(X)$  défini par

$$\mathbb{V}(X) = \mathbb{E}((X - \mathbb{E}(X))^2). \quad (30)$$

*Rappel 2.20.* Soit  $(X, \mathcal{B}, \mu)$  un espace mesuré tel que  $\mu(X) < +\infty$ . Alors, si  $p > q$ , on a  $\mathcal{L}^p(X, \mu) \subseteq \mathcal{L}^q(X, \mu)$ .

**Proposition 2.21.** Soit  $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  une variable aléatoire à valeur réelle. Lorsque la variable aléatoire  $X$  vérifie  $\mathbb{E}(X^2) < +\infty$ , alors,

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \quad (31)$$

Pour conclure cette section, nous allons réécrire le théorème de changement de variables dans le cas qui nous intéressera par la suite. On va montrer comment déterminer la loi de probabilité de  $\phi(X_1, \dots, X_n)$  si  $\mathbb{P}_{(X_1, \dots, X_n)}$  est connue.

On prend  $\Psi$  une application borélienne positive.

$$\begin{aligned}\mathbb{E}(\Psi(\phi(X_1, \dots, X_n))) &= \int_{\Omega} \Psi(\phi(X_1, \dots, X_n)) d\mathbb{P} \\ &= \int_{\mathbb{R}^n} \Psi(\phi(x)) d\mathbb{P}_{(X_1, \dots, X_n)}(x) \quad (\text{théorème de transport})\end{aligned}$$

Nous allons à présent faire un exemple complet pour voir comment cela fonctionne.

**Exemple 2.22.** Considérons  $X$  et  $Y$  deux variables aléatoires réelles indépendantes identiquement distribuées absolument continues de densité  $f_X(x) = \lambda e^{-\lambda(x-1)} \mathbb{1}_{]1, +\infty[}(x)$  avec  $\lambda > 0$ . On pose  $U = \frac{X}{X+Y}$  et  $V = X+Y$ .

1. Déterminer  $\mathbb{P}_{(U,V)}$ .

On sait que  $X$  et  $Y$  sont indépendantes. Ainsi, pour tout  $(x, y) \in \mathbb{R}^2$ , on a  $f_{(X,Y)}(x, y) = f_X(x)f_Y(y)$ . Ainsi, on a :

$$f_{(X,Y)}(x, y) = \lambda^2 e^{-\lambda(x-1)} e^{-\lambda(y-1)} \mathbb{1}_{]1, +\infty[}(x) \mathbb{1}_{]1, +\infty[}(y).$$

Considérons  $\Psi$  une fonction borélienne positive définie sur  $\mathbb{R}^2$ . D'un côté, on a :

$$\begin{aligned}\mathbb{E}(\Psi(U, V)) &= \int_{\Omega} \Psi(U, V) d\mathbb{P} \\ &= \int_{\mathbb{R}^2} \Psi(u, v) d\mathbb{P}_{(U,V)}(u, v) \\ &= \int_{\mathbb{R}^2} \Psi(u, v) f_{(U,V)}(u, v) dudv\end{aligned}$$

De l'autre côté,

$$\begin{aligned}\mathbb{E}(\Psi(U, V)) &= \mathbb{E}\left(\Psi\left(\frac{X}{X+Y}, X+Y\right)\right) \\ &= \int_{\Omega} \Psi\left(\frac{X}{X+Y}, X+Y\right) d\mathbb{P} \\ &= \int_{\mathbb{R}^2} \Psi\left(\frac{x}{x+y}, x+y\right) d\mathbb{P}_{(X,Y)}(x, y) \\ &= \int_{\mathbb{R}^2} \Psi\left(\frac{x}{x+y}, x+y\right) f_{(X,Y)}(x, y) dx dy \\ &= \int_{\mathbb{R}^2} \Psi(u, v) A(u, v) dudv\end{aligned}$$

Par identification,  $A(u, v)$  correspond à la fonction de densité du couple  $(U, V)$ . Il faut donc déterminer l'ensemble sur lequel on va réaliser ce changement de variables et le jacobien de ce dernier.

On a posé  $u = \frac{x}{x+y}$  et  $v = x + y$ . Ainsi, on a  $u = \frac{x}{v}$ , i.e.  $uv = x$ . Ainsi,  $y = v - x = v(1 - u)$ . On remarque tout d'abord que  $v \in ]2, +\infty[$  et  $u \in ]0, 1[$ . Le domaine d'intégration, noté  $\Delta$ , devient alors :

$$\Delta = \{(u, v) \in ]0, 1[ \times ]2, +\infty[ / 1 \leq uv, 1 \leq v(1 - u)\}.$$

Le jacobien du changement de variable, noté  $Jac_{(U,V)}(u, v)$ , est donné par :

$$Jac_{(U,V)}(u, v) = \det \begin{pmatrix} \frac{\partial dx}{\partial du} & \frac{\partial dx}{\partial dv} \\ \frac{\partial dy}{\partial du} & \frac{\partial dy}{\partial dv} \end{pmatrix},$$

i.e.

$$Jac_{(U,V)}(u, v) = \det \begin{pmatrix} v & u \\ -v & 1 - u \end{pmatrix},$$

On obtient alors :

$$f_{(U,V)}(u, v) = \lambda^2 v e^{-\lambda(v-2)} \mathbb{1}_{\Delta}(u, v).$$

2. Les variables aléatoires  $U$  et  $V$  sont-elles indépendantes ?

Non, les variables aléatoires  $U$  et  $V$  ne sont pas indépendantes, car  $f_{(U,V)}(u, v) \neq f_U(u)f_V(v)$ .

3. Déterminer  $\mathbb{P}_U$ .

On souhaite déterminer la densité marginale de la variable aléatoire  $U$ , notée  $f_U$ . Par définition, on a :

$$f_U(u) = \int_{\mathbb{R}} f_{(U,V)}(u, v) dv.$$

On a alors :

$$\begin{aligned} f_U(u) &= \int_{\mathbb{R}} f_{(U,V)}(u, v) dv \\ &= \int_{\mathbb{R}} \lambda^2 v e^{-\lambda(v-2)} \mathbb{1}_{\Delta}(u, v) dv \\ &= \left( \int_{\frac{1}{u}} \lambda^2 v e^{-\lambda(v-2)} dv \right) \mathbb{1}_{]0, 1/2]}(u) + \left( \int_{\frac{1}{1-u}} \lambda^2 v e^{-\lambda(v-2)} dv \right) \mathbb{1}_{]1/2, 1[}(u) \\ &= \lambda^2 e^{2\lambda} \left( \frac{-1}{\lambda} [v e^{-\lambda v}]_{1/u}^{+\infty} + \frac{1}{\lambda} \int_{1/u}^{+\infty} e^{-\lambda v} dv \right) \mathbb{1}_{]0, 1/2]}(u) \\ &\quad + \lambda^2 e^{2\lambda} \left( \frac{-1}{\lambda} [v e^{-\lambda v}]_{1/(1-u)}^{+\infty} + \frac{1}{\lambda} \int_{1/(1-u)}^{+\infty} e^{-\lambda v} dv \right) \mathbb{1}_{]1/2, 1[}(u) \\ &= \left( \lambda e^{2\lambda} \frac{1}{u} e^{-\frac{\lambda}{u}} + \lambda e^{2\lambda} \left[ \frac{e^{-\lambda v}}{-\lambda} \right]_{1/u}^{+\infty} \right) \mathbb{1}_{]0, 1/2]}(u) + \left( \lambda e^{2\lambda} \frac{1}{1-u} e^{-\frac{\lambda}{1-u}} + \lambda e^{2\lambda} \left[ \frac{e^{-\lambda v}}{-\lambda} \right]_{1/u}^{+\infty} \right) \mathbb{1}_{]1/2, 1[}(u) \\ &= \left( \frac{\lambda e^{2\lambda}}{u} e^{-\frac{\lambda}{u}} + e^{2\lambda} e^{-\frac{\lambda}{u}} \right) \mathbb{1}_{]0, 1/2]}(u) + \left( \frac{\lambda e^{2\lambda}}{1-u} e^{-\frac{\lambda}{1-u}} + e^{2\lambda} e^{-\frac{\lambda}{1-u}} \right) \mathbb{1}_{]1/2, 1[}(u) \end{aligned}$$

4. Déterminer  $\mathbb{E}(U)$  si elle existe.

On remarque rapidement que l'intégrale  $\int_0^1 u f_U(u) du$  est divergente. Ainsi, l'espérance de la variable aléatoire  $U$  n'est pas définie.

## Références

- [1] Maryse Béguin. *Théorie de la mesure et de l'intégration pour les probabilités*. Ellipses, 2013.
- [2] Gerald B. Folland. *Real Analysis - Modern Techniques and Their Applications*. John Wiley and Sons, Inc, 1999.
- [3] Janos Galambos. *Introductory Probability Theory*. Marcel Dekker, Inc, 1984.
- [4] Gilles Pagès Marc Briane. *Théorie de l'intégration*. Vuibert, 2004.