Rappels de théorie des probabilités

CentraleSupélec

Hervé Frezza-Buet
Herve.Frezza-Buet@centralesupelec.fr

mis à jour le 2 septembre 2018

Table des matières

1 Tribus et mesures

Soit X un ensemble. Soit $T \subset \mathcal{P}(X)$ une famille de sous-ensembles de X. La famille T forme une tribu ssi :
- $T \neq \emptyset$
- $B \in T \Rightarrow \overline{B} \in T$
- $(\forall 1 \leq i \leq n,\ B_i \in T) \Rightarrow \displaystyle\bigcup_{1 \leq i \leq n} B_i \in T$

Il s'ensuit que $X \in T$ et $\emptyset \in T$.

Une fois qu'on a muni un ensemble d'une tribu, on peut le mesurer un associant un réel positif à chacun des éléments de la tribu. La mesure $\mu \in [0, +\infty]^T$ doit vérifier :

— $\mu(\emptyset) = 0$.
— La mesure d'une union *dénombrable* d'ensembles *disjoints deux à deux* est la somme des mesures de ces ensembles.

On dit que (X, T) est un ensemble mesurable. Soient (X, T) et (Y, T') deux ensembles mesurables, la fonction $f \in Y^X$ est dite (T, T')-mesurable si $B \in T' \Rightarrow f^{-1}(B) \in T$. Si on a seulement une tribu sur X, on peut facilement construire via f une tribu sur Y qui fasse de f une fonction mesurable (tribu image). Idem dans l'autre sens, quand on a une tribu sur Y (tribu image réciproque).

Le triplet (X, T, μ) est un ensemble mesuré.

2 Probabilité

En probabilités, l'ensemble que l'on considère est est appelé *univers*, il est noté Ω. C'est l'ensemble des résultats possibles des expériences. On lui associe une tribu $\mathcal{O}$, qui représente *les événements*. Il devient mesurable. On note les mesures $\mathbb{P}$ (au lieu de μ). Elles deviennent des probabilités du moment où on impose $\mathbb{P}(\Omega) = 1$ (donc $\mathbb{P} \in [0, 1]^{\mathcal{O}}$).

On dira que $(\Omega, \mathcal{O}, \mathbb{P})$ est un espace probabilisé : on pondère chaque événement par une probabilité, avec les bonnes propriétés d'addition en cas d'union et d'intersection d'événements.

Les probabilités conditionnelles et la formule de Bayes viennent alors :

$$\mathbb{P}(A \mid B) \stackrel{\text{def}}{=} \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \Rightarrow \mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

La fonction $\mathbb{P}(\bullet \mid B) \in [0, 1]^{\mathcal{O}}$ est aussi une mesure de probabilité, i.e $(\Omega, \mathcal{O}, \mathbb{P}(\bullet \mid B))$ est un ensemble probabilisé quel que soit $B \in \mathcal{O} \setminus \{\emptyset\}$.

3 Variables aléatoires

3.1 Définition

Une variable aléatoire X est une fonction mesurable. Soit $(E, \mathcal{E})$ un ensemble de *valeurs* associées aux événements, la variable aléatoire X à valeurs dans E est une fonction $X \in E^{\Omega}$ qui est $(\mathcal{O}, \mathcal{E})$-mesurable. On considère un ensemble Ω probabilisé, i.e $(\Omega, \mathcal{O}, \mathbb{P})$.

La ruse consiste à définir les probabilités que X prenne certaines valeurs (dans E) via $\mathbb{P}$, qui pourtant propabilise Ω (voir equation (1) et figure 1).

$$\forall B \in \mathcal{E},\ \mathbb{P}(X \in B) \stackrel{\text{def}}{=} \mathbb{P}\left(X^{-1}(B)\right) \tag{1}$$

3.2 Loi de probabilité

3.2.1 Définition

La fonction $\mathbb{P}_X \in [0, 1]^{\mathcal{E}}$ définie par

$$\mathbb{P}_X(B) \stackrel{\text{def}}{=} \mathbb{P}(X \in B)$$

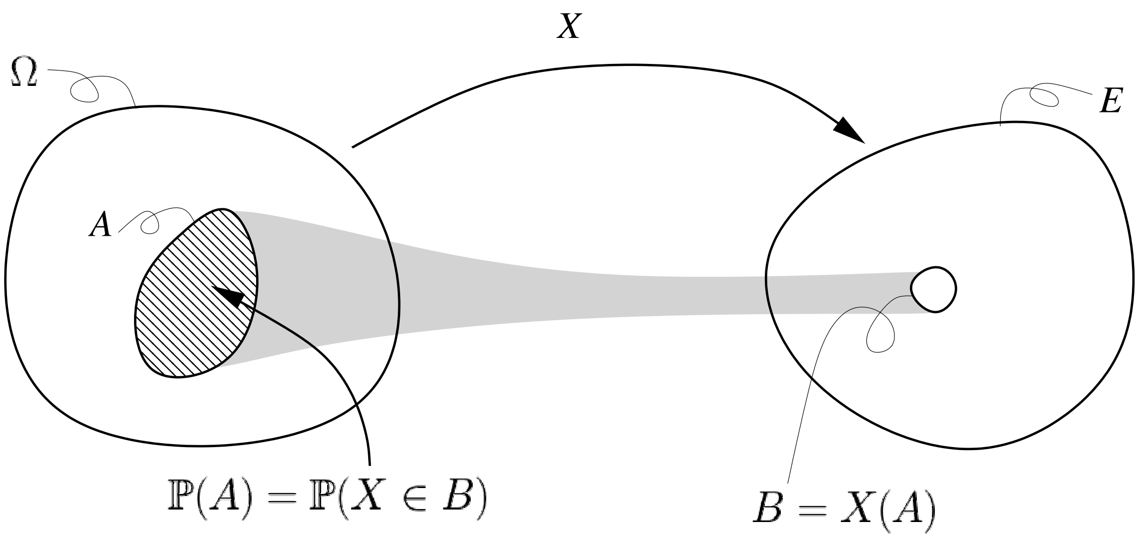

FIGURE 1 – La variable aléatoire X. Les événements s'interprètent graphiquement comme des sous-régions des ensembles, la probabilité peut être vue comme l'aire de la région (l'aire totale valant 1)... pour les régions de Ω. On a $\mathbb{P}(X \in B) = \mathbb{P}\left(X^{-1}(B)\right)$.

est ce qu'on appelle la *loi de probabilité* de X. Si on définit cette loi, sans la déduire de $\mathbb{P}(X \in B)$, on ignore $(\Omega, \mathcal{O}, \mathbb{P})$, ou du moins on s'épargne de l'expliciter. C'est souvent ce que l'on fait quand on parle d'une variable aléatoire à valeurs dans E. La fonction $\mathbb{P}_X$ est une mesure sur $(E, \mathcal{E})$.

La densité de probabilité est une fonction $\mathrm{p}_X \in (\mathbb{R}^+)^E$ qui, *quand elle existe*, vérifie

$$\forall B,\ \mathbb{P}(X \in B) = \int_E \mathbb{1}_B(x)\, \mathrm{p}_X(x)\, \mathrm{d}x$$

3.2.2 Espérance

L'espérance d'une variable aléatoire X est définie par

$$\mathbb{E}[X] \stackrel{\text{def}}{=} \int_\Omega X(\omega)\, \mathrm{d}\mathbb{P}(\omega) = \int_E x \mathrm{d}\mathbb{P}_X(x) \tag{2}$$

Ces intégrales sont des intégrales de Lebesgue [1], qui se définissent à partir des mesures justement. Mesurer les ensembles (Ω ici est mesuré par $\mathbb{P}$) perment de définir proprement leur « taille » par intégration de Lebesgue.

La définition de l'espérance par l'expression de gauche dans l'équation (2), ou la loi de probabilité $\mathbb{P}_X$ uniquement intervient, est justement le cas où l'on se passe d'expliciter $(\Omega, \mathcal{O}, \mathbb{P})$. Tout est caché dans $\mathbb{P}_X$.

3.2.3 Variance

La variance est définie par :

$$\operatorname{var}(X) \stackrel{\text{def}}{=} \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \tag{3}$$

mais cette définition est mal posée à ce stade. $\mathbb{E}[X]$ est un scalaire, une constante. Notons-là e. Posons $f \in \mathbb{R}^\mathbb{R}$ la fonction $f(x) = (x-e)^2$. L'idée est d'utiliser f pour définir une nouvelle

1. Non détaillées ici. Elles sont de la forme $\int_X f(x)\, \mathrm{d}\mu(x)$, avec μ une mesure sur X.

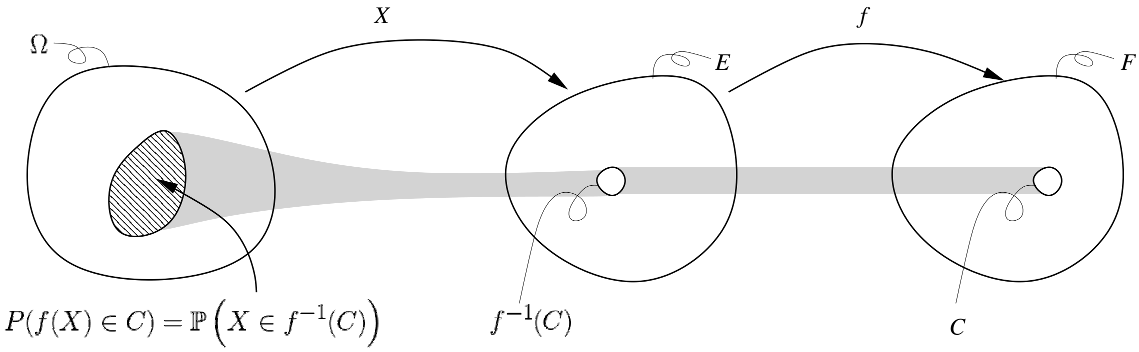

FIGURE 2 – Loi de probabilité de $f(X)$. Voir aussi la figure 1.

variable aléatoire notée $f(X)$ à partir de X. La variance $\mathrm{var}(X)$ définie par l'équation (3) est alors l'espérance de cette variable aléatoire que nous avons noté $f(X)$. La définition de l'application d'une fonction à une variable aléatoire est justement l'objet du paragraphe suivant.

3.3 Fonction d'une variable aléatoire

Soit $X \in E^{\Omega}$ une variable aléatoire, et $f \in F^{E}$ une fonction, on peut définir la variable aléatoire $f(X) = Y \in F^{\Omega}$ (il faut munir F d'une tribu... à creuser). La loi de probabilité de Y est donnée, pour $C \subset F$ par l'équation (4) (voir figure 2)

$$\mathbb{P}(Y \in C) \stackrel{\text{def}}{=} \mathbb{P}(f(X) \in C) \stackrel{\text{def}}{=} \mathbb{P}\left(X \in f^{-1}(C)\right) \tag{4}$$

3.4 Couple de variables aléatoires

3.4.1 Définition

On pourrait penser que l'on peut considérer deux variables aléatoires (pas nécessairement définies sur les mêmes univers Ω) et en faire un couple. En fait, il n'en est rien dans le cas général [2]. On va définir *une* variable aléatoire, à valeur dans $E = E_1 \times E_2$. À partir de cette variable-là, dite *variable jointe*, on séparera les deux variables du couple, par une opération de *marginalisation*.

Notons $(X, Y) \in (E_X \times E_Y)^{\Omega}$ *la* variable jointe, définie sur l'univers $(\Omega, \mathcal{O}, \mathbb{P})$. Usuellement, les $\omega \in \Omega$ sont des paires... mais rien ne l'exige. Notons qu'il faut définir une tribu sur $E_X \times E_Y$.

Donc, soit un ensemble de couple de valeurs $B \in E_X \times E_Y$, la loi de (X, Y) est définie, comme nous l'avons vu, par

$$\mathbb{P}((X,Y) \in B) = \mathbb{P}\left((X,Y)^{-1}(B)\right)$$

Du coup, on peut considérer X et Y séparément. Leurs lois s'obtiennent par une opération de marginalisation, selon l'equation (5) (voir figure 3). On écrira $\forall B \subset E_X$ pour simplifier [3], sachant que l'on passe sous silence la définition de tribus adéquates (on a une tribu sur $E_X \times E_Y$, dont on dérive une tribu sur E_X (ou E_Y) uniquement, ainsi qu'une sous-tribu dans Ω).

$$\forall B \subset E_X,\ \mathbb{P}(X \in B) \stackrel{\text{def}}{=} \mathbb{P}((X,Y) \in B \times E_Y) \tag{5}$$

2. On peut le faire pour construire des variables indépendantes en fait, on définira l'indépendance plus loin. Mais d'un point de vue théorique, il n'y a pas deux variables aléatoires qui existent seules que l'on peut réunir en un couple.
3. au lieu d'écrire $\forall B \in \mathcal{E}_X$, où $\mathcal{E}_X$ est une tribu sur E_X.

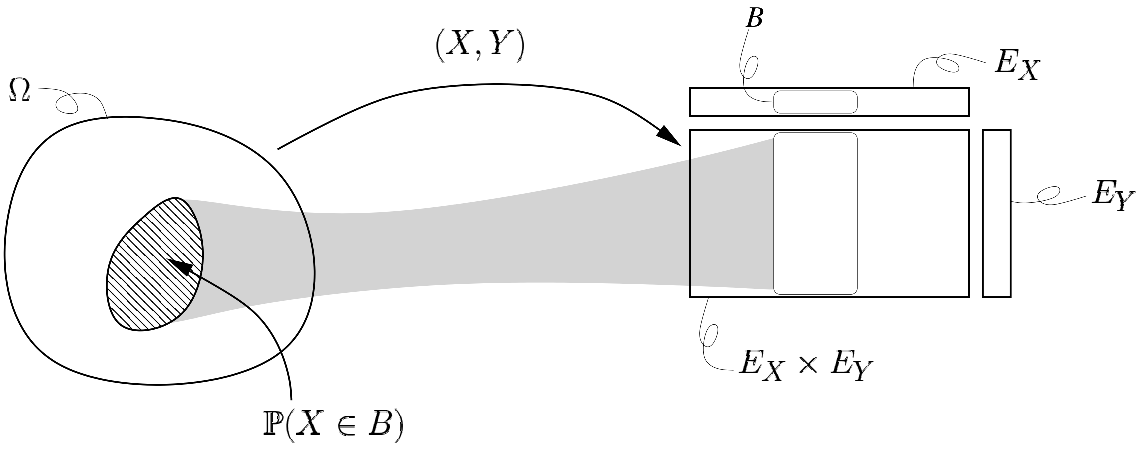

FIGURE 3 – Loi de probabilité de la variable aléatoire X obtenue par marginalisation de la variable aléatoire (X, Y). Voir aussi la figure 1.

On peut alors définir des lois conditionnelles :

$$\mathbb{P}\left(X \in A \mid Y \in B\right) \stackrel{\text{def}}{=} \frac{\mathbb{P}\left((X,Y) \in A \times B\right)}{\mathbb{P}\left(Y \in B\right)}$$

qui définit « la probabilité de X ait ses valeurs dans A sachant que Y a ses valeurs dans B ». Cette loi conditionnelle définit la variable aléatoire conditionnelle $X \mid Y \in B$ (à comprendre comme $X \mid (Y \in B))$:

$$\mathbb{P}_{X \mid Y \in B}\left(A\right) \stackrel{\text{def}}{=} \mathbb{P}\left(X \in A \mid Y \in B\right) \tag{6}$$

Le cas usuel est le cas où A et B sont des singletons $\{x\}$ et $\{y\}$, on écrit alors $X = x$ au lieu de $X \in \{x\}$. On a alors la formule plus classique (mais moins générale) :

$$\mathbb{P}_{X \mid Y=y}\left(x\right) \stackrel{\text{def}}{=} \frac{\mathbb{P}\left((X,Y) = (x,y)\right)}{\mathbb{P}\left(Y = y\right)}$$

3.4.2 Indépendance

Un couple de variables (X, Y) est constitué de deux variables *indépendantes* si on a :

$$\forall A \in \mathcal{E}_X,\ \forall B \in \mathcal{E}_Y,\ \mathbb{P}\left((X,Y) \in A \times B\right) = \mathbb{P}\left(X \in A\right) \times \mathbb{P}\left(Y \in B\right)$$

On en déduit facilement, à l'aide de l'équation (6), que pour deux variables (X, Y) indépendantes, on a :

$$\forall A \in \mathcal{E}_X,\ \forall B \in \mathcal{E}_Y,\ \left\{\begin{array}{lll} \mathbb{P}\left(Y \in B\right) \neq 0 & \Rightarrow & \mathbb{P}\left(X \in A \mid Y \in B\right) = \mathbb{P}\left(X \in A\right) \\ \mathbb{P}\left(X \in A\right) \neq 0 & \Rightarrow & \mathbb{P}\left(Y \in B \mid X \in A\right) = \mathbb{P}\left(Y \in B\right) \end{array}\right.$$

Dit autrement, connaître la valeur d'une des variables ne change pas les probabilité des valeurs possibles pour l'autre. La figure 4 montre graphiquement comment sont les distributions quand deux variables sont indépendantes. Un exemples de variable dépendentes est illustré figure 5.

3.4.3 Corrélation

Soit (X, Y) une paire de variables aléatoires (et donc X et Y les deux variables marginalisées). On définit leur covariance par :

$$\text{cov}\left(X, Y\right) \stackrel{\text{def}}{=} \mathbb{E}\left[(X - \mathbb{E}\left[X\right])(Y - \mathbb{E}\left[Y\right])\right]$$

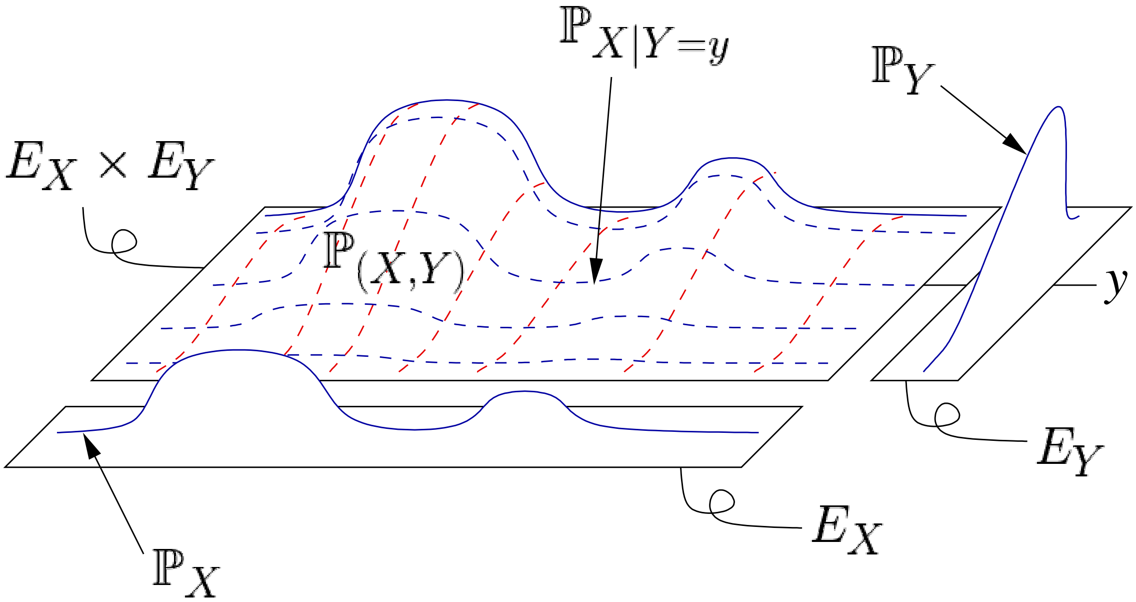

FIGURE 4 – La figure montre les distributions $\mathbb{P}_X$, $\mathbb{P}_Y$ et $\mathbb{P}_{(X,Y)}$. Les lignes pointillées bleues sont les distibutions conditionnelles $\mathbb{P}_{X \mid Y=y}$. Comme les variables sont indépendantes, ces lignes ont, à une échelle près, la même forme que $\mathbb{P}_X$. Le raisonnement est symétrique si l'on considère $\mathbb{P}_{Y \mid X=x}$ (lignes pointillées rouges).

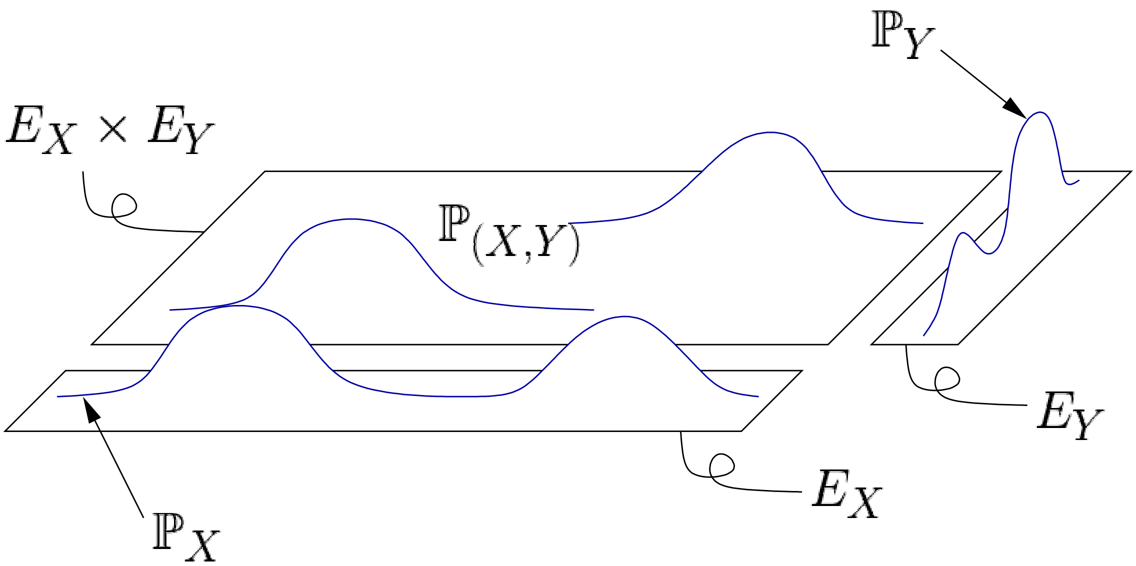

FIGURE 5 – La figure montre les distributions $\mathbb{P}_X$, $\mathbb{P}_Y$ et $\mathbb{P}_{(X,Y)}$. Comme les variables sont dépendantes, si l'on traçait les distributions conditionnelles comme sur la figure 4, on ne retrouverait plus l'effet d'échelle.

Si $\text{cov}(X,Y)=0$, on dit que X et Y sont décorrélées. L'indépendance implique la décorrélation, mais pas l'inverse.

Le coefficient de corrélation est :

$$\text{cov}(X,Y) \overset{\text{def}}{=} \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} \in [-1,1]$$

3.5 Familles de variables aléatoires

On peut aisément généraliser ce qu'on a vu sur les couples aux $n-uplets$. Mais on peut aussi aller plus loin. Un n-uplet de variables aléatoires $(X_1,\cdots,X_i,\cdots,X_n)$ peut être vu comme une famille $\{X_i\}_{i\in[1,2,\cdots n]}$, c'est-à-dire une famille indexée par l'ensemble fini $[1,2,\cdots n]$.

Ce qui n'est pas évident du tout [4], c'est que l'on peut définir une famille de variables aléatoires qui soit indexée par des ensembles continus. Par exemple, une famille de redimensionnements de valeurs définis à partir de la variable aléatoire X peut-être $\{\alpha X+\beta\}_{(\alpha,\beta)\in\mathbb{R}^2}$.

4. Essayez d'imaginez les tribus associées....