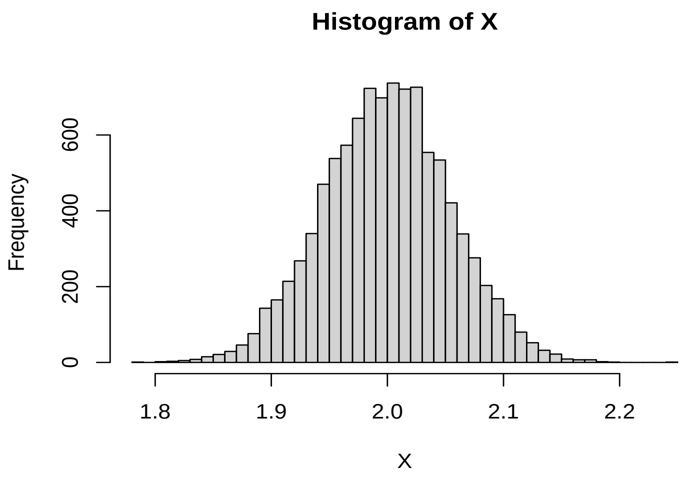
Université Paris I, Panthéon - Sorbonne

LICENCE M.I.A.S.H.S. TROISIÈME ANNÉE

Cours de Pré-rentrée: Théorie de la mesure et Probabilités

JEAN-MARC BARDET (UNIVERSITÉ PARIS 1, SAMM)



U.F.R. 27 et SAMM (Statistique, Analyse et Modélisation Multidisciplinaire)
Université Panthéon-Sorbonne, 90 rue de Tolbiac, 75013 Paris.

Plan du cours

Introduction

1. Rappels sur les ensembles
2. Evénements, tribus et mesure de probabilité
3. Probabilité conditionnelle et indépendance
4. Variables aléatoires

References

[1] Barbe et Ledoux, *Probabilités*, Belin.

[2] Dauxois, J. et Hassenforder, C. (2004). Toutes les probabilités et Statistiques. Cours et Exercices corrigés. Ellipses.

[3] Garet, O. et Kuntzmann, A., *De l'intégration aux probabilités*, Ellipses.

[4] Saporta, G. Probabilités, analyse des données et statistique (2nd édition), éditions Technip.

Introduction

Il demeure des choses inconnues à partir des connaissances antérieures en probabilités. Par exemple:

- Que se passe-t-il pour des probabilités d'événements moins classiques (par exemple l'ensemble des décimaux) ?
- Comment traiter une variable aléatoire qui est continue et discrète à la fois (par exemple le nombre de minutes passées devant la TV), notamment comment calculer son espérance?

1 Rappels sur les ensembles

Notation. Si Ω est un ensemble (fini ou infini), $\mathcal{P}(\Omega)$ est l'ensemble de tous les sous-ensembles (parties) de Ω.

Rappel. Union, intersection, complémentaire: $\overline{A} = \Omega \setminus A$ pour $A \subset \Omega$).
Si $I \subset \mathbb{N}$, et si les $(A_i)_{i \in I}$ sont des ensembles appartenant à Ω, alors:

$$\overline{\bigcup_{i \in I} A_i} = \bigcap_{i \in I} \overline{A_i}.$$

Rappel. Soit E un ensemble. E est dit dénombrable s'il existe une bijection entre E et $\mathbb{N}$ ou un sous-ensemble de $\mathbb{N}$.
Autre manière de formuler cette définition: E est dénombrable si $E = \{(u_i)_{i \in \mathbb{N}}\}$ où $(u_i)_{i \in \mathbb{N}}$ est une suite d'éléments.

Exemple.

Montrer que $\mathbb{Z}$ et $\mathbb{N}^2$ sont dénombrables. En déduire que $\mathbb{Q}$ est dénombrable.

Théorème. $\mathbb{R}$ *n'est pas dénombrable.*

Proof. Argument de la diagonale de Cantor. □

2 Evénements, tribus et mesure de probabilités

2.1 Evénements et tribus

Définition. Soit une famille $\mathcal{F}$ de parties de Ω (donc $\mathcal{F} \subset \mathcal{P}(\Omega)$). On dit que $\mathcal{F}$ est une algèbre si:

- $\Omega \subset \mathcal{F}$;
- lorsque $A \in \mathcal{F}$ alors $(\Omega \setminus A) \in \mathcal{F}$;
- pour tout $n \in \mathbb{N}^*$, lorsque $(A_1, \cdots, A_n) \in \mathcal{F}^n$ alors $A_1 \cup \cdots \cup A_n \in \mathcal{F}$.

Définition. Soit une famille $\mathcal{A}$ de parties de Ω (donc $\mathcal{A} \subset \mathcal{P}(\Omega)$). On dit que $\mathcal{A}$ est une tribu (ou σ-algèbre) sur Ω si :

- $\Omega \subset \mathcal{F}$;
- lorsque $A \in \mathcal{F}$ alors $(\Omega \setminus A) \in \mathcal{F}$;
- pour $I \subset \mathbb{N}$, lorsque $(A_i)_{i \in I} \in \mathcal{F}^I$ alors $\bigcup_{i \in I} A_i \in \mathcal{A}$.

Exemple.

- Cas du Pile ou Face.
- Cas où Ω est infini: $\Omega = \mathbb{N}$ ou $\mathbb{R}$ par exemple.

Propriété. *Avec les notations précédentes :*

1. $\emptyset \in \mathcal{A}$;
2. *si A et B sont dans la tribu $\mathcal{A}$, alors $A \cap B$ est dans $\mathcal{A}$;*
3. *si $\mathcal{A}_1$ et $\mathcal{A}_2$ sont deux tribus sur Ω, alors $\mathcal{A}_1 \cap \mathcal{A}_2$ est une tribu sur Ω . Plus généralement, pour $I \subset \mathbb{N}$, si $(\mathcal{A}_i)_{i\in I}$ ensemble de tribus sur Ω, alors $\bigcap_{i\in I} \mathcal{A}_i$ est une tribu sur Ω;*
4. *si $\mathcal{A}_1$ et $\mathcal{A}_2$ sont deux tribus sur Ω, alors $\mathcal{A}_1 \cup \mathcal{A}_2$ n'est pas forcément une tribu sur Ω.*

Définition. Si $\mathcal{E}$ est une famille de parties de Ω (donc $\mathcal{E} \subset \mathcal{P}(\Omega)$), alors on appelle tribu engendrée par $\mathcal{E}$, notée $\sigma(\mathcal{E})$, la tribu engendrée par l'intersection de toutes les tribus contenant $\mathcal{E}$ (on peut faire la même chose avec des algèbres).

Remarque.

La tribu engendrée est la "plus petite" tribu (au sens de l'inclusion) contenant la famille $\mathcal{E}$.

Rappel.
- Un ensemble ouvert U dans un espace métrique X est telle que pour tout $x \in U$, il existe $r > 0$ tel que $B(x, r) \subset U$.
- On dit qu'un ensemble dans un espace métrique X est fermé si son complémentaire dans X est ouvert.

Définition. Soit Ω un espace métrique. On appelle tribu borélienne sur Ω, notée, $\mathcal{B}(\Omega)$, la tribu engendrée par les ouverts de Ω. Un ensemble de $\mathcal{B}(\Omega)$ est appelé borélien.

Exemple.

- Boréliens sur $\mathbb{R}$, sur $]0, 1[$.
- Boréliens sur $\mathbb{R}^2$.

Définition. Soit Ω un ensemble et soit $\mathcal{A}$ une tribu sur Ω. On dit que $(\Omega, \mathcal{A})$ est un espace mesurable.

Corollaire. *Quand on s'intéressera aux probabilités, on dira que $(\Omega, \mathcal{A})$ est un espace probabilisable.*

Propriété. *Si $(\Omega_i, \mathcal{A}_i)_i$ sont n espaces mesurables, alors un ensemble élémentaire de $\Omega = \Omega_1 \times \cdots \times \Omega_n$ est une réunion finie d'ensembles $A_1 \times \cdots \times A_n$ où chaque $A_i \in \mathcal{A}_i$. L'ensemble des ensembles élémentaires est une algèbre et on note $\mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_n$ (on dit $\mathcal{A}_1$ tensoriel $\mathcal{A}_2$ ... tensoriel $\mathcal{A}_n$) la tribu sur Ω engendrée par ces ensembles élémentaires.*

Exemple.

Pavés de $\mathbb{R}^d$.

Définition. On appelle espace mesurable produit des $(\Omega_i, \mathcal{A}_i)_i$ l'espace mesurable $\left(\prod_{i=1}^{n} \Omega_i, \bigotimes_{i=1}^{n} \mathcal{A}_i\right)$.

Exemple.

Pile / Face 2 fois.

2.2 Mesure et mesure de probabilités

Définition. Soit $(\Omega, \mathcal{A})$ un espace mesurable. L'application $\mu : \mathcal{A} \to [0, +\infty]$ est une mesure si:

- $\mu(\emptyset) = 0$.
- Pour tout $I \subset \mathbb{N}$ et pour $(A_i)_{i\in I}$ famille disjointe de $\mathcal{A}$ (telle que $A_i \cup A_j = \emptyset$ pour $i \neq j$), alors $\mu\left(\bigcup_{i\in I} A_i\right) = \sum_{i\in I} \mu(A_i)$ (propriété dite de σ-additivité).

Définition. Avec les notations précédentes :

- Si $\mu(\Omega) < +\infty$, on dit que μ est une mesure finie.
- Si $\mu(\Omega) = 1$, on dit que μ est une mesure de probabilité.

Définition. Soit $(\Omega, \mathcal{A})$ un espace mesurable et soit $x_0 \in \Omega$. On appelle masse de Dirac en x_0 la mesure de probabilité δ_{x_0} telle que pour tout $A \in \mathcal{A}$, $\delta_{x_0}(A) = 1$ si $x_0 \in A$ et $= 0$ sinon.

Exemple.

- Cas où $\Omega = \{0, 1\}$.
- Cas où $\Omega = \{(x_i)_{1\leq i\leq n}\}$.
- On se place dans $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$. Montrer que l'application $\mu : \mathcal{P}(\mathbb{N}) \mapsto [0, +\infty]$ telle que pour $A \in \mathcal{P}(\mathbb{N})$, $\mu(A) = \sum_{k\in A} 2^{-k}$ est une mesure sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$.
- Cas où $\Omega = [0, 1]$.
- Cas où $\Omega = \mathbb{R}$.

Définition. Si $(\Omega, \mathcal{A})$ est un espace mesurable (resp. probabilisable) alors $(\Omega, \mathcal{A}, \mu)$ est un espace mesuré (resp. probabilisé quand μ est une probabilité).

Remarque.

Sur $(\Omega, \mathcal{A})$, on peut définir une infinité de mesures.

Propriété. *Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et $(A_i)_{i\in\mathbb{N}}$, une famille de $\mathcal{A}$.*

1. *Si $A_1 \subset A_2$, alors $\mu(A_1) \leq \mu(A_2)$.*
2. *Si $\mu(A_1) < +\infty$ et $\mu(A_2) < +\infty$, alors $\mu(A_1 \cup A_2) + \mu(A_1 \cap A_2) = \mu(A_1) + \mu(A_2)$.*
3. *Pour tout $I \subset \mathbb{N}$, on a $\mu\left(\bigcup_{i\in I} A_i\right) \leq \sum_{i\in I} \mu(A_i)$.*
4. *Si $A_i \subset A_{i+1}$ pour tout $i \in \mathbb{N}$ (suite croissante en sens de l'inclusion), alors $(\mu(A_n))_{n\in\mathbb{N}}$ est une suite croissante convergente telle que*
   $\mu\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \lim_{i\to+\infty} \mu(A_i)$ *(même si cette limite est $+\infty$).*

5. *Si $A_{i+1} \subset A_i$ pour tout $i \in \mathbb{N}$ (suite décroissante en sens de l'inclusion) et $\mu(A_0) < +\infty$, alors $(\mu(A_n))_{n\in\mathbb{N}}$ est une suite décroissante convergente telle que $\mu\left(\bigcap_{i\in\mathbb{N}} A_i\right) = \lim_{i\to+\infty} \mu(A_i)$.*

Exemple.

1. Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. On définit $\nu(A) = \mu(A \cap B)$ où $B \in \mathcal{A}$. ν mesure?
2. Si μ_1 et μ_2 mesures sur $(\Omega, \mathcal{A})$, $\alpha \in]0, \infty[$, $\mu_1 + \mu_2$, $\alpha\mu$ et $\mu_1\,\mu_2$ sont-elles des mesures? Et dans le cas de mesures de probabilités?

Définition. Définition de la mesure de Lebesgue sur $\mathbb{R}$, $\mathbb{R}^n$,...

Définition. Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré.

1. Pour $A \in \mathcal{A}$, on dit que A est μ-négligeable si $\mu(A) = 0$.
2. Soit une propriété $\mathcal{P}$ dépendant des éléments ω de Ω. On dit que $\mathcal{P}$ est vraie μ-presque partout (μ-presque sûrement sur un espace probabilisé) si l'ensemble des ω pour laquelle elle n'est pas vérifiée est μ-négligeable.

Exemple.

- Mesure de Lebesgue sur un espace dénombrable (typiquement $\mathbb{Q}$).
- La propriété " la suite de fonction $f_n(x) = x^n$ converge vers la fonction $f(x) = 0$" est vraie λ-presque partout sur $[0, 1]$.

Exercice.

On considère l'espace probabilisé $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ et on définit la fonction numérique F telle que pour $x \in \mathbb{R}$, $F(x) = \mu(]-\infty, x])$.

1. Montrer que F est croissante, et que $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to+\infty} F(x) = 1$.
2. Montrer que F admet des limites à gauche et à droite en tout point de $\mathbb{R}$, puis que F est continue à droite sur $\mathbb{R}$.
3. Montrer que $F(b) - F(a) = \mu(]a, b])$ et $F(a) - F(a^-) = \mu(a)$ pour $-\infty < a < b$ ($F(a^-)$ est la limite à gauche de F en a).
4. Soit D l'ensemble des points de $\mathbb{R}$ où F est discontinue, et soit $D_n = \{t \in \mathbb{R}, \mu(t) \geq \frac{1}{n}\}$. Montrer que $D = \bigcup_{n=1}^{\infty} D_n$ et en déduire que D est λ-négligeable où λ est la mesure de Lebesgue sur $\mathbb{R}$. D est-il μ-négligeable?

3 Probabilité conditionnelle et indépendance

3.1 Formule des probabilités totales

Définition. Soit Ω un ensemble et $J \subset \mathbb{N}$. On dit que $(E_i)_{i\in J}$ famille de sous-ensembles de Ω forme une partition de Ω si:

- Les E_i sont incompatibles deux à deux soit $E_i \cap E_j = \emptyset$ pour $i \neq j$.
- L'ensemble des E_i couvre Ω soit $\bigcup_{i\in J} E_i = \Omega$.

Proposition. (Formule des probabilités totales) *Soit $\mathbb{P}$ est une probabilité sur $(\Omega, \mathcal{A})$ et soit $(E_j)_{j\in J}$ des évenements de $\mathcal{A}$ constituant une partition de Ω. Alors, pour tout $A \in \mathcal{A}$,*

$$\mathbb{P}(A) = \sum_{j\in J} \mathbb{P}(A \cap E_j).$$

Rappel. Cas particulier de l'équiprobabilité

Définition.
- Cardinal d'un ensemble fini.
- Equiprobabilité (ou probabilité uniforme) sur un ensemble fini.

Propriété. *Si Ω est fini et si $\mathbb{P}$ est une probabilité uniforme sur $(\Omega, \mathcal{A})$, alors pour tout $E \in \mathcal{A}$,* $\mathbb{P}(E) = \dfrac{card(E)}{card(\Omega)}$.

Remarque: Pour calculer le cardinal d'un ensemble, on peut utiliser les résultats combinatoires suivants: on considère un ensemble de n éléments et on tire k éléments:

- S'il y a remise, et que l'ordre compte, un tirage est un k-uplet, et le nombre total de k-uplets est: n^k.
- S'il n'y a pas remise, et que l'ordre compte, un tirage est un arrangement, et le nombre total d'arrangements est: $A_n^k = n(n-1) \times \cdots \times (n-k+1) = n!/(n-k)!$.
- S'il n'y a pas remise, et que l'ordre ne compte pas, un tirage est une combinaison, et le nombre total de combinaisons est: $C_n^k = A_n^k/k! = n!/(k!(n-k)!)$.

3.2 Probabilité conditionnelle et indépendance

Définition. Soit $\mathbb{P}$ est une probabilité sur $(\Omega, \mathcal{A})$ et soit $(E, F) \in \mathcal{A}^2$. On appelle probabilité conditionnelle de B sachant A, la probabilité que B soit réalisé sachant que A est réalisé et $\mathbb{P}(B \mid A) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$ si $\mathbb{P}(A) \neq 0$.

Remarque : Calculer une probabilité sachant A revient à travailler avec une nouvelle probabilité sur l'ensemble fondamental A et la tribu qui lui est associée.

Définition. A et B, événements de $(\Omega, \mathcal{A})$ sont indépendants pour $\mathbb{P}$ si $\mathbb{P}(B \mid A) = \mathbb{P}(B)$.

Conséquence : Deux événements A et B de $(\Omega, \mathcal{A})$ sont indépendants pour $\mathbb{P}$ si et seulement si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Remarque : Ne pas confondre indépendance et incompatibilité! De la dernière conséquence, on montre facilement que l'on peut être incompatible sans être indépendant (puisqu'on a toujours $\mathbb{P}(A \cap \overline{A}) = 0$, différent de $\mathbb{P}(A)\mathbb{P}(\overline{A})$ sauf si l'un des deux est nul) et l'on peut être indépendant sans être incompatible (penser à deux lancers successifs d'une pièce équilibrée et aux évenèments A: le premier lancer est Pile et B: le second lancer est Pile: on a $A \cap B = \{(PP)\}$).

Proposition. (Formule de Bayes) *Soit $\mathbb{P}$ est une probabilité sur $(\Omega, \mathcal{A})$ et soit $(E_j)_{j\in J}$ des évenements de $\mathcal{A}$ constituant une partition de Ω. Soit A un événement de $\mathcal{A}$). On suppose que l'on connaît $\mathbb{P}(E_j)$ et $\mathbb{P}(A \mid E_j)$ pour tout $j \in J$. Alors, pour $k \in J$:*

$$\mathbb{P}(E_k \mid A) = \frac{\mathbb{P}(A \mid E_k)\mathbb{P}(E_k)}{\sum_{j\in J}\mathbb{P}(A \mid E_j)\mathbb{P}(E_j)}.$$

Définition. Soit $(A_i)_{i\in I}$ une famille d'événements de $(\Omega, \mathcal{A})$. On dit que la famille $(A_i)_{i\in I}$ est une famille d'événements (mutuellement) indépendants si et seulement si pour tout $k \in \mathbb{N}$, pour tout $j_1, \cdots, j_k \in I^k$ distincts, $\mathbb{P}\big(A_{j_1} \cap \cdots \cap A_{j_k}\big) = \mathbb{P}(A_{j_1}) \times \cdots \times \mathbb{P}(A_{j_k})$.

Remarque : Etre mutuellement indépendant est plus fort que d'être indépendant deux à deux. Prendre à ce sujet l'exemple précédent de deux lancers successifs d'une pièce équilibrée et considérer les évenèments A: le premier lancer est Pile et B: le second lancer est Pile et C: les deux lancers donnent la même chose. On aura bien A et B indépendants, A et C également, de même que B et C, et pourtant A, B et C ne sont pas mutuellement indépendants...

4 Variables aléatoires

4.1 Variables aléatoires

Définition. On dit que X est une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité si: $X : \Omega \to \mathbb{R}$ est une application mesurable de $(\Omega, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Pourquoi demander à X d'être mesurable? Parce que l'on veut pouvoir définir une fonction de répartition pour X, soit $F_X(x) = \mathbb{P}(\text{"}X \leq x\text{"}) = \mathbb{P}\big(\{\omega \in \Omega,\ X(\omega) \leq x\}\big) = X^{-1}(]-\infty, x])$: il faut donc que l'ensemble $X^{-1}(]-\infty, x])$ soit un évenement de $\mathcal{A}$ pour tout $x \in \mathbb{R}$.

Dans la suite, pour une variable aléatoire X définie sur $(\Omega, \mathcal{A}, \mathbb{P})$, on n'utilisera presque jamais la forme explicite de la fonction $\omega \to X(\omega)$ qui restera inconnue. En revanche, on préfèrera travailler avec la **loi** de X.
Qu'appelle-t-on loi de X? Il y a en réalité plusieurs moyens de la définir:

1. On définit la loi par $F_X(x) = \mathbb{P}(X \leq x)$ pour $x \in \mathbb{R}$, la **fonction de répartition** de X;

2. On définit la loi par la **mesure de probabilité** $\mathbb{P}_X$ **induite par** X: pour $B \in \mathcal{B}(\mathbb{R})$, $\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \mathbb{P}\big(X^{-1}(B)\big)$. On notera que $\mathbb{P}_X$ est une mesure de probabilité définie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ou bien sur $(X(\Omega), \mathcal{T})$, avec $\mathcal{T}$ une tribu sur $X(\Omega) \subset \mathbb{R}$.

3. Dans 2 cas particuliers, on préfèrera travailler avec:

   - Si $X(\Omega) = \big\{x_i\big\}_{i\in I}$ avec $I \subset \mathbb{N}$ et $x_i \in \mathbb{R}$ for all $i \in I$, on parlera de variable aléatoire **discrète** et la loi sera donnée par les $\mathbb{P}(X = x_i) = \mathbb{P}_X\big(\{x_i\}\big)$ pour $i \in I$. On remarque alors que pour tout $B \in \mathcal{B}(\mathbb{R})$, $\mathbb{P}_X(B) = \sum_{i\in I}\mathbb{P}(X = x_i)\,\delta_{\{x_i\}}(B)$, avec δ masse (mesure) de Dirac.
   - Si $X(\Omega)$ est une union finie ou dénombrable d'intervalles de I, et si $\mathbb{P}_X$ est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}$, alors on peut définir la **densité de probabilité** f_X de X par rapport à la mesure de Lebesgue sur $\mathbb{R}$, et l'on a:

   $$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt \quad \text{pour tout } x \in \mathbb{R}.$$

   La fonction F_X est une fonction **absolument continue** sur $\mathbb{R}$, elle est même dérivable λ-presque partout sur $\mathbb{R}$ et $F'_X(x) = f_X(x)$ pour presque tout x dans $\mathbb{R}$. On dira souvent pour simplifier que X est une variable aléatoire **"absolument continue"** et même parfois que X est une variable aléatoire **"continue"**.

On va dans la propriété suivante utiliser pleinement le fait qu'une variable aléatoire est une fonction mesurable:

Propriété. *Soit X et Y deux v.a. sur $(\Omega, \mathcal{A}, \mathbb{P})$ et $g : \mathbb{R}^2 \to \mathbb{R}$ une fonction borélienne. Alors $Z = g(X, Y)$ est une v.a. sur $(\Omega, \mathcal{A}, \mathbb{P})$.*

Proof. Voir cours Intégration et Probabilités: il s'agit de montrer que la composition d'une fonction mesurable par une autre fonction mesurable est une fonction mesurable. Et également que le vecteur composé de fonctions mesurables est une fonction mesurable. □

Par itération du procédé, la propriété est aussi vraie pour une fonction g à n variables. Mais aussi bien-sûr pour une fonction à 1 variable.
De la même manière, toujours du fait qu'une variable aléatoire est une fonction mesurable:

Propriété. *Soit X et Y deux v.a. sur $(\Omega, \mathcal{A}, \mathbb{P})$ telles que $\mathbb{P}_X = \mathbb{P}_Y$, ce que l'on peut aussi noter $X \overset{\mathcal{L}}{\sim} Y$. Alors pour et $g : \mathbb{R} \to \mathbb{R}$ une fonction borélienne, $g(X) \overset{\mathcal{L}}{\sim} g(Y)$.*

Proof. Pour tout $B \in \mathcal{B}(\mathbb{R})$, $\mathbb{P}_{g(X)}(B) = \mathbb{P}(g(X) \in B) = \mathbb{P}(X \in g^{-1}(B)) = \mathbb{P}_X(g^{-1}(B))$. Mais on a aussi $\mathbb{P}_{g(Y)}(B) = \mathbb{P}_Y(g^{-1}(B)) = \mathbb{P}_X(g^{-1}(B))$ puisque $\mathbb{P}_X = \mathbb{P}_Y$. D'où le résultat. □

Remarque: Si $X \overset{\mathcal{L}}{\sim} \mathcal{U}\big([0,1]\big)$ et $Y = 1 - X$, alors on a aussi $Y \overset{\mathcal{L}}{\sim} \mathcal{U}\big([0,1]\big)$, donc $\mathbb{P}_X = \mathbb{P}_Y$ (on le montre en utilisant par exemple la fonction de répartition). Mais pourtant on n'a pas du tout $X = Y$...

Définition. Si X est une v.a. sur $(\Omega, \mathcal{A}, \mathbb{P})$ et $0 < p < 1$, le quantile d'ordre p de X est : $q_X(p) = \inf\big\{y \in \mathbb{R},\ F_X(y) \geq p\big\}$.

Cas particulier: Si $p = 1/2$, alors $q_X(p)$ est la médiane (théorique) de X.

Propriété. *Si X est une v.a. continue, $q_X(p) = \widetilde{F}_X^{-1}(p)$ où $\widetilde{F}_X : x \in X(\Omega) \mapsto F_X(x)$.*

Exemple: Si X suit une distribution $\mathcal{E}(\lambda)$ (exponentielle),

$$F_X(x) = (1 - e^{-\lambda x})\mathbb{I}_{x \geq 0} \quad \text{and} \quad q_X(1/2) = \frac{\ln(2)}{\lambda} \neq \mathbb{E}[X] = \frac{1}{\lambda}.$$

4.2 Plus généralement: fonctions mesurables

Rappel. Soit $f : E \mapsto F$, où E et F sont 2 espaces métriques.

- Pour $I \subset F$, on appelle ensemble réciproque de I par f, l'ensemble $f^{-1}(I) = \{x \in E, f(x) \in I\}$.
- (f continue) $\Longleftrightarrow$ (pour tout ouvert U de F alors $f^{-1}(U)$ est un ouvert de E).

Définition. Soit $f : E \mapsto F$ et soit $\mathcal{I}$ une tribu sur F. On note $f^{-1}(\mathcal{I})$ l'ensemble de sous-ensembles de Ω tel que $f^{-1}(\mathcal{I}) = \{f^{-1}(I), I \in \mathcal{I}\}$.

Propriété. *Soit $(\Omega', \mathcal{A}')$ un espace mesurable et soit $f : \Omega \mapsto \Omega'$. Alors $f^{-1}(\mathcal{A})$ est une tribu sur Ω appelée tribu engendrée par f.*

Définition. Soit $(\Omega, \mathcal{A})$ et $(\Omega', \mathcal{A}')$ deux espaces mesurables. Une fonction $f : \Omega \mapsto \Omega'$ est dite mesurable pour les tribus $\mathcal{A}$ et $\mathcal{A}'$ si et seulement si $f^{-1}(\mathcal{A}') \subset \mathcal{A}$ (donc si et seulement si $\forall A' \in \mathcal{A}'$, alors $f^{-1}(A') \in \mathcal{A}$).

Exemple.

- Fonction indicatrice.
- Combinaison linéaire de fonctions indicatrices.

Remarque.

Dans le cas où $(\Omega, \mathcal{A})$ est un espace probabilisable, et si $f : \Omega \mapsto \mathbb{R}$, alors si f est une fonction mesurable sur $\mathcal{A}$ et $\mathcal{B}(\mathbb{R})$, alors f est une variable aléatoire.

Exemple.

Nombre de Piles dans un jeu de Pile/Face.

Remarque.

Dans le cas où $(\Omega, \mathcal{A})$ est un espace mesurable, et si $f : \Omega \mapsto (\Omega', \mathcal{B}(\Omega'))$, où Ω' est un espace métrique et $\mathcal{B}(\Omega')$ l'ensemble des boréliens de Ω', si f est une fonction mesurable sur $\mathcal{A}$ et $\mathcal{B}(\Omega')$, alors f est dite fonction borélienne.

Proposition. *Soit $(\Omega, \mathcal{A})$ et $(\Omega', \mathcal{A}')$ deux espaces mesurables et $f : \Omega \mapsto \Omega'$. Soit $\mathcal{F}$ une famille de sous-ensembles de Ω' telle que $\sigma(\mathcal{F}) = \mathcal{A}'$. Alors*

1. *$f^{-1}(\mathcal{F})$ engendre la tribu $f^{-1}(\mathcal{A})$.*
2. *(f mesurable) $\Longleftrightarrow$ ($f^{-1}(\mathcal{F}) \subset \mathcal{A}$)*

Conséquence.
- Si $(\Omega, \mathcal{A})$ et $(\Omega', \mathcal{A}')$ sont deux espaces mesurables boréliens, alors toute application continue de $\Omega \mapsto \Omega'$ est mesurable.
- Pour montrer qu'une fonction $f : \Omega \mapsto \mathbb{R}$ est mesurable, il suffit de montrer que la famille d'ensemble $(\{\omega \in \Omega, f(\omega) \leq a\})_{a \in \mathbb{R}} \in \mathcal{A}$.

Propriété.
- *Soit f mesurable de $(\Omega, \mathcal{A})$ dans $(\Omega', \mathcal{A}')$ et g mesurable de $(\Omega', \mathcal{A}')$ dans $(\Omega'', \mathcal{A}'')$. Alors $g_0 f$ est mesurable dans $\mathcal{A}$ et $\mathcal{A}'$.*
- *Soit f_1 mesurable de $(\Omega, \mathcal{A})$ dans $(\Omega_1, \mathcal{A}_1)$ et f_2 mesurable de $(\Omega, \mathcal{A})$ dans $(\Omega_2, \mathcal{A}_2)$. Alors $h : \Omega \mapsto \Omega_1 \times \Omega_2$ telle que $h(\omega) = (f_1(\omega), f_2(\omega))$ est mesurable dans $\mathcal{A}$ et $\mathcal{A}_1 \otimes \mathcal{A}_2$.*
- *Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions mesurables de $(\Omega, \mathcal{A})$ dans $(\Omega', \mathcal{B}(\Omega'))$, où Ω' est un espace métrique, telle qu'il existe une fonction f limite simple de (f_n) (donc $\forall \omega \in \Omega$, $\lim_{n \to \infty} f_n(\omega) = f(\omega)$). Alors f est mesurable dans $\mathcal{A}$ et $\mathcal{B}(\Omega')$.*

Définition. Soit f mesurable de $(\Omega, \mathcal{A}, \mu)$ dans $(\Omega', \mathcal{A}')$ et soit $\mu_f : \mathcal{A}' \mapsto [0, +\infty]$ telle que pour tout $A' \in \mathcal{A}'$, on ait $\mu_f(A') = \mu(f^{-1}(A'))$. Alors μ_f est une mesure sur $(\Omega', \mathcal{A}')$ appelée mesure image de μ par f.

Cas particulier.

Si μ est une mesure de probabilité et si X est une variable aléatoire alors μ_X est la mesure (loi) de probabilité de la variable aléatoire X.

Cas des fonctions réelles mesurables

Propriété. *Soit f et g deux fonctions réelles mesurables (de $(\Omega, \mathcal{A}, \mu)$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$). Alors $\alpha.f$, $f+g$, $\min(f,g)$ et $\max(f,g)$ sont des fonctions réelles mesurables.*

Propriété. *Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions réelles mesurables. Alors $\inf(f_n)$ et $\sup(f_n)$ sont des fonctions réelles mesurables.*

Définition. Soit $f : \Omega \to \mathbb{R}$. Alors f est dite étagée s'il existe une famille d'ensembles disjoints $(A_i)_{1\leq i\leq n}$ de Ω et une famille de réels $(\alpha_i)_{1\leq i\leq n}$ telles que pour tout $\omega \in \Omega$, on ait $f(\omega) = \sum_{i=1}^{n} \alpha_i \mathbb{I}_{A_i}(\omega)$.

Remarque.

Si les A_i sont tous dans $\mathcal{A}$ tribu sur Ω, alors f est $\mathcal{A}$-mesurable.

Théorème. *Toute fonction réelle mesurable à valeurs dans $[0, +\infty]$ est limite simple d'une suite croissante de fonctions étagées.*

Conséquence. Soit f une fonction réelle mesurable. Alors f est limite simple de fonctions étagées.

Mesures induites et densités

Théorème (Théorème du Transport)**.** *Soit f une fonction mesurable de $(\Omega, \mathcal{A}, \mu)$ dans $(\Omega', \mathcal{A}')$ telle que μ_f soit la mesure induite par f (donc $\mu_f(A') = \mu(f^{-1}(A'))$ pour $A' \in \mathcal{A}'$) et soit ϕ une fonction mesurable de $(\Omega', \mathcal{A}')$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors, si $\phi_0 f \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$,*

$$\int_{\Omega'} \phi \, d\mu_f = \int_{\Omega} \phi_0 f \, d\mu.$$

Définition. Soit μ et ν deux mesures sur $(\Omega, \mathcal{A})$. On dit que μ domine ν (ou ν est dominée par μ) et que ν est absolument continue par rapport à μ lorsque pour tout $A \in \mathcal{A}$, $\mu(A) = 0 \Longrightarrow \nu(A) = 0$.

Propriété. *Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et f une fonction définie sur $(\Omega, \mathcal{A})$ mesurable et positive. On suppose que pour $A \in \mathcal{A}$, $\nu(A) = \int_A f \, d\mu$. Alors, ν est une mesure sur $(\Omega, \mathcal{A})$, dominée par μ. De plus, pour toute fonction g définie sur $(\Omega, \mathcal{A})$ mesurable et positive,*

$$\int g d\nu = \int g.f d\mu.$$

Enfin, g est ν intégrable si et seulement si $g.f$ est μ intégrable.

Définition. On dit que μ mesure sur $(\Omega, \mathcal{A})$ est σ-finie lorsqu'il existe une famille $(A_i)_{i\in I}$, avec I dénombrable, d'ensembles de $\mathcal{A}$ telle que $\bigcup A_i = \Omega$ et $\mu(A_i) < +\infty$ pour tout $i \in I$.

Théorème (Théorème de Radon-Nikodym)**.** *On suppose que μ et ν sont deux mesures σ-finies sur $(\Omega, \mathcal{A})$ telles que μ domine ν. Alors il existe une fonction f définie sur $(\Omega, \mathcal{A})$ mesurable et positive, appelée densité de ν par rapport à μ, telle que pour tout $A \in \mathcal{A}$, $\nu(A) = \int_A f \, d\mu$.*

4.3 Espérance de variables aléatoires

Définition. Soit X une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Alors si $X \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ (donc si $\int_\Omega |X(\omega)|\, d\mathbb{P}(\omega) < \infty$), on définit l'espérance de X par le réel:

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

Plus généralement, si $\phi : \mathbb{R} \mapsto \mathbb{R}$ est borélienne et si $\phi(X) \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$, on définit l'espérance de $\phi(X)$ par

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

Propriété. *Si X est une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$, si $\phi : \mathbb{R} \mapsto \mathbb{R}$ est borélienne telle que $\phi(X) \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$, alors :*

$$\mathbb{E}\left[\phi(X)\right] = \int_{\mathbb{R}} \phi(x)\, d\mathbb{P}_X(x).$$

Proof. Théorème du transport... □

Conséquence.

- Si $\mathbb{P}_X$ est absolument continue par rapport à la mesure de Lebesgue (donc X est une v.a. dite absolument continue), de densité f_X, alors $\mathbb{E}[\phi(X)] = \displaystyle\int_{\mathbb{R}} \phi(x)\, f_X(x)\, dx$.
- Si $\mathbb{P}_X$ est absolument continue par rapport à une mesure de comptage sur $\{x_i\}_{i \in I}$ avec $I \subset \mathbb{N}$ (donc X est une v.a. dite discrète), de densité p_X avec $p_X(i) = \mathbb{P}(X = x_i)$, alors $\mathbb{E}[\phi(X)] = \displaystyle\sum_{i \in I} p_X(i)\, \phi(x_i)$.

Propriété.

1. *Soit X et Y des variables aléatoires telles que X et $Y \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$. Alors pour tout $(a, b) \in \mathbb{R}^2$, $aX + bY \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ et*

$$\mathbb{E}[a\, X + b\, Y] = a\, \mathbb{E}[X] + b\, \mathbb{E}[Y].$$

2. *Soit X une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$, et soit $B \in \mathcal{B}(\mathbb{R})$. Alors $\mathbb{E}[\mathbb{I}_B(X)] = \mathbb{P}(X \in B)$.*

Définition. Pour X et Y des variables aléatoires telles que X et $Y \in \mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$, on définit:

- la variance de X, $\mathrm{var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.
- la covariance de X et Y par

$$\mathrm{cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right] = \mathbb{E}[X\, Y] - \mathbb{E}[X]\, \mathbb{E}[Y].$$

Remarque: On a $\mathrm{var}(X) = \mathrm{cov}(X, X)$.

Propriété. *Sur l'espace vectoriel $E = \Big\{X \in \mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P}),\ \mathbb{E}[X] = 0\Big\}$, on définit $< X, Y >= cov(X, Y)$ pour $X, Y \in E$. Alors $< \cdot, \cdot >$ définit un produit scalaire.*

Conséquence. Pour X et Y deux v.a. sur $\mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$, alors

$$|\mathrm{cov}(X, Y)|^2 \leq \mathrm{var}(X)\, \mathrm{var}(Y).$$

Propriété. *Pour X et Y deux v.a. sur $\mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$, on peut définir le coefficient de corrélation entre X et Y par:*

$$cor(X, Y) = \frac{cov(X, Y)}{\sqrt{var(X)\, var(Y)}} \quad et \quad \big|cor(X, Y)\big| \leq 1.$$

Définition. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et $I \subset \mathbb{N}$.

- Soit $(A_i)_{i\in I}$ une famille d'événements de $\mathcal{A}$. On dit que les événements $(A_i)_{i\in I}$ sont indépendants si et seulement si pour tous les sous-ensembles finis $K \subset I$,

$$\mathbb{P}\left(\bigcap_{i\in K} A_i\right) = \prod_{i\in K} \mathbb{P}(A_i).$$

- Soit $(\mathcal{A}_i)_{i\in I}$ une famille de sous-tribus de $\mathcal{A}$ (donc pour tout $i \in I$, $\mathcal{A}_i \subset \mathcal{A}$). On dit que les tribus $(\mathcal{A}_i)_{i\in I}$ sont indépendantes si et seulement si pour tous les sous-ensembles finis $K \subset I$, et pour tous les événements $A_k \in \mathcal{A}_k$ avec $k \in K$, les A_k sont indépendants.

- Soit $(X_i)_{i\in I}$ une famille de variables aléatoires sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On dit que les v.a. $(X_i)_{i\in I}$ sont indépendantes si et seulement si les tribus engendrées $(X_i^{-1}(\mathcal{B}(\mathbb{R})))_{i\in I}$ sont indépendantes, ou encore pour tous les sous-ensembles finis $K \subset I$:

$$\mathbb{P}\big(\bigcap_{i\in K} X_i \in B_i\big) = \prod_{i\in K} \mathbb{P}(X_i \in B_i) \quad \text{pour tous } B_i \in \mathcal{B}(\mathbb{R}).$$

4.4 Lois à connaître

Définition. Les lois à connaître sont:

- Lois discrètes: loi uniforme, loi de Bernoulli, loi binomiale, loi de Poisson.

- Lois continues: loi uniforme, loi exponentielle, loi gamma, loi normale.