Espérance Conditionnelle

1. Rappels.

- Dans toute ce paragraphe, $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé, i.e.
  - ⋆ Ω est un ensemble non vide.
  - ⋆ $\mathcal{F} \subset \mathcal{P}(\Omega)$ est une tribu (ou σ-algèbre) sur Ω :
    1. $\emptyset \in \mathcal{F}$;
    2. Si $A \in \mathcal{F}$, alors $A^c \in \mathcal{F}$;
    3. Si $(A_n)_{n\in\mathbf{N}} \subset \mathcal{F}$, alors $\bigcup_{n\in\mathbf{N}} A_n \in \mathcal{F}$.
  - ⋆ $\mathbb{P}$ est (une mesure de) probabilité sur $(\Omega, \mathcal{F})$: $\mathbb{P} : \mathcal{F} \longrightarrow \overline{\mathbf{R}}_+ := \mathbf{R}_+ \cup \{+\infty\}$ t.q.
    1. $\mathbb{P}(\emptyset) = 0$;
    2. Pour $(A_n)_{n\in\mathbf{N}} \subset \mathcal{F}$ t.q. $A_n \cap A_k = \emptyset$ si $k \neq n$,
    $$\mathbb{P}\left(\bigcup_{n\in\mathbf{N}} A_n\right) = \sum_{n\geq 0} \mathbb{P}(A_n) \ ;$$
    3. $\mathbb{P}(\Omega) = 1$.

Remarque(s). Une mesure de probabilité est à valeurs dans $[0, 1]$.

Définition (Tribu engendrée)**.** Soit $\mathcal{C} \subset \mathcal{P}(\Omega)$. On appelle *tribu engendrée par* $\mathcal{C}$, notée $\sigma(\mathcal{C})$, la plus petite tribu sur Ω, au sens de l'inclusion, contenant $\mathcal{C}$.

- L'existence de $\sigma(\mathcal{C})$ résulte du fait qu'une intersection quelconque de tribus sur Ω est une tribu sur Ω.
- Si $A \subset \Omega$, $\sigma(\{A\}) = \{\emptyset, \Omega, A, A^c\}$.

Définition (Tribu borélienne)**.** On appelle *tribu borélienne de* $\mathbf{R}^n$, notée $\mathcal{B}(\mathbf{R}^n)$, la tribu engendrée par les ouverts (pour la topologie usuelle) de $\mathbf{R}^n$.

- On a $\mathcal{B}(\mathbf{R}) = \sigma\left(\{]-\infty, a] : a \in \mathbf{R}\}\right) = \sigma\left(\{[a, b] : a \in \mathbf{R}, b \in \mathbf{R}\}\right)$

Définition. Soit $X : \Omega \longrightarrow \mathbf{R}^n$ une application. On dit que X est une *variable aléatoire* si

$$\forall B \in \mathcal{B}(\mathbf{R}^n), \quad X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \overset{\text{not.}}{=} \{X \in B\} \in \mathcal{F}.$$

Remarque(s). Comme $\mathcal{B}(\mathbf{R}) = \sigma\left(\{]-\infty, a] : a \in \mathbf{R}\}\right)$, X est une v.a. réelle ssi, pour tout réel a, $\{X \leq a\} \in \mathcal{F}$.

Définition. Soient X une variable aléatoire et $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. On dit que X est $\mathcal{G}$-mesurable si

$$\forall B \in \mathcal{B}(\mathbf{R}^n), \quad \{X \in B\} \in \mathcal{G}.$$

Définition (Tribu engendrée par une v.a.)**.** Soit X une variable aléatoire. On appelle *tribu engendrée par* X, notée $\sigma(X)$, la plus petite tribu pour l'inclusion, qui rend X mesurable.

Lemme (Factorisation)**.** *Soit X une variable aléatoire à valeurs dans $\mathbf{R}^n$. Une variable aléatoire réelle Y est $\sigma(X)$-mesurable si et seulement si il existe une fonction $h : \mathbf{R}^n \longrightarrow \mathbf{R}$ borélienne telle que $Y = h(X)$.*

Remarque(s). Plus généralement, si $X_1, \ldots, X_n$ sont des v.a., $\sigma(X_1, \ldots, X_n)$ est la plus petite tribu pour l'inclusion qui rend $X_1, \ldots, X_n$ mesurables et Y est $\sigma(X_1, \ldots, X_n)$-mesurable ssi $Y = h(X_1, \ldots, X_n)$ avec h borélienne.

Définition (Indépendance)**.** Deux variables aléatoires X et Y sont indépendantes si les tribus $\sigma(X)$ et $\sigma(Y)$ le sont c'est à dire si

$$\begin{aligned}&\forall B \in \mathcal{B}(\mathbf{R}^n),\ \forall C \in \mathcal{B}(\mathbf{R}^m),\\ &\qquad\qquad \mathbb{P}\left(\{X \in B\} \cap \{Y \in C\}\right) = \mathbb{P}\left(\{X \in B\}\right) \mathbb{P}\left(\{Y \in C\}\right).\end{aligned}$$

Remarque(s). On peut également donner une définition « fonctionnelle » de l'indépendance : X et Y sont indépendantes si et seulement si, pour toutes fonctions f et g boréliennes et bornées,

$$\mathbb{E}\left[f(X)\, g(Y)\right] = \mathbb{E}\left[f(X)\right]\, \mathbb{E}\left[g(Y)\right].$$

Le passage de la définition ensembliste à la définition fonctionnelle se fait via la formule :

$$\forall B \in \mathcal{B}(\mathbf{R}^n), \qquad \mathbb{P}\left(\{X \in B\}\right) = \mathbb{E}\left[\mathbf{1}_B(X)\right].$$

2. Espérance conditionnelle.

- Soient X et Y des variables aléatoires et $\mathcal{G}$ une sous-tribu de $\mathcal{F}$.
  - ⋆ Connaissant l'information contenue dans $\mathcal{G}$, que peut-on dire de plus sur la v.a. X ?
  - ⋆ Quelle est la meilleure approximation de X connaissant $\sigma(Y)$?
  - ⋆ La notion d'espérance conditionnelle répond à ces questions

2.1. Exemple introductif.

- Soient X et Y deux v.a.r. ; X de carré intégrable et Y discrète à valeurs $y_1 < y_2 < y_3 < \ldots$
- Pour tout borélien B et tout $j \geq 1$, la probabilité conditionnelle

$$\mathbb{P}\left(\{X \in B\} | \{Y = y_j\}\right) = \frac{\mathbb{P}\left(\{X \in B\} \cap \{Y = y_j\}\right)}{\mathbb{P}\left(\{Y = y_j\}\right)}$$

représente la fréquence de réalisation de $\{X \in B\}$ parmi tous les événements où $Y = y_j$.

- On remarque que
$$\mathbb{P}\left(\{X \in B\}|\{Y = y_j\}\right) = \frac{\mathbb{E}\left[\mathbf{1}_B(X)\mathbf{1}_{\{Y=y_j\}}\right]}{\mathbb{P}\left(\{Y = y_j\}\right)},$$
et on définit, pour toute fonction f borélienne (bornée ou positive)
$$\mathbb{E}\left[f(X) \mid \{Y = y_j\}\right] = \frac{\mathbb{E}\left[f(X)\mathbf{1}_{\{Y=y_j\}}\right]}{\mathbb{P}\left(\{Y = y_j\}\right)}.$$
  - ⋆ En fait, la probabilité conditionnelle sachant $\{Y = y_j\}$, $A \longmapsto \mathbb{P}(A|\{Y = y_j\})$, est la probabilité de densité $\mathbf{1}_{\{Y=y_j\}}/\mathbb{P}\left(\{Y = y_j\}\right)$ par rapport à $\mathbb{P}$.
- On définit une variable aléatoire Z en posant
$$Z(\omega) = \sum_{j\geq 1} \mathbb{E}\left[X \mid \{Y = y_j\}\right] \mathbf{1}_{Y(\omega)=y_j}, \quad \text{i.e. } Z = \mathbb{E}\left[X \mid \{Y = y_j\}\right] \text{ si } Y = y_j.$$
  - ⋆ Z est $\sigma(Y)$-mesurable puisque $Z = h(Y)$ avec $h(y) = \sum_{j\geq 1} \mathbb{E}\left[X \mid \{Y = y_j\}\right] \mathbf{1}_{y=y_j}$
  - ⋆ Par ailleurs, si $G = g(Y)$ est $\sigma(Y)$-mesurable et de carré intégrable, on a
$$\begin{aligned}
\mathbb{E}\left[GZ\right] &= \mathbb{E}\left[g(Y)\sum_{j\geq 1} \mathbb{E}\left[X \mid \{Y = y_j\}\right] \mathbf{1}_{Y=y_j}\right] = \sum_{j\geq 1} \mathbb{E}\left[g(Y)\mathbb{E}\left[X \mid \{Y = y_j\}\right] \mathbf{1}_{Y=y_j}\right], \\
&= \sum_{j\geq 1} \mathbb{E}\left[g(y_j)\mathbb{E}\left[X \mid \{Y = y_j\}\right] \mathbf{1}_{Y=y_j}\right] = \sum_{j\geq 1} g(y_j)\mathbb{E}\left[X \mid \{Y = y_j\}\right]\mathbb{P}(\{Y = y_j\}), \\
&= \sum_{j\geq 1} g(y_j)\mathbb{E}\left[X\mathbf{1}_{Y=y_j}\right] = \sum_{j\geq 1} \mathbb{E}\left[g(y_j)X\mathbf{1}_{Y=y_j}\right] = \sum_{j\geq 1} \mathbb{E}\left[g(Y)X\mathbf{1}_{Y=y_j}\right], \\
&= \mathbb{E}\left[Xg(Y)\sum_{j\geq 1} \mathbf{1}_{Y=y_j}\right] = \mathbb{E}\left[Xg(Y)\right], \quad \text{puisque } \sum_{j\geq 1} \mathbf{1}_{Y=y_j} = 1, \\
&= \mathbb{E}\left[GX\right].
\end{aligned}$$
- La variable aléatoire Z vérifie les deux propriétés suivantes :
  - ⋆ Z est $\sigma(Y)$-mesurable ;
  - ⋆ pour toute v.a. G, $\sigma(Y)$-mesurable et de carré intégrable,
$$\mathbb{E}\left[G(X - Z)\right] = \langle G, X - Z\rangle = 0$$
  où $\langle\cdot,\cdot\rangle$ désigne le produit scalaire de L^2.
- Par conséquent, Z est la projection orthogonale de X sur $\mathrm{L}^2(\sigma(Y))$, le sous-espace des v.a. $\sigma(Y)$-mesurables de carré intégrable
- Z est vérifie donc
$$\mathbb{E}\left[|X - Z|^2\right] = \min\left\{\mathbb{E}\left[|X - G|^2\right] : G \in \mathrm{L}^2(\sigma(Y))\right\}.$$
- Z est la meilleure approximation de X, au sens des moindres carrés, connaissant l'information contenue dans $\sigma(Y)$.

2.2. Définition.

- Soit $\mathcal{G}$ une sous-tribu de $\mathcal{F}$
- Notons $P_{\mathcal{G}}$ la projection orthogonale sur $\mathrm{L}^2(\mathcal{G})$, le sous-espace vectoriel des « variables aléatoires » $\mathcal{G}$-mesurables et de carré intégrable.
- Si $X \in \mathrm{L}^2(\mathcal{F})$, $Z = P_{\mathcal{G}}(X)$ vérifie
  - ⋆ Z est $\mathcal{G}$-mesurable ;
  - ⋆ Z est de carré intégrable ;
  - ⋆ Pour toute v.a. G, $\mathcal{G}$-mesurable et de carré intégrable, $\mathbb{E}[XG] = \mathbb{E}[ZG]$
- On a également,

$$\mathbb{E}\left[|X-Z|^2\right] = \min\left\{\mathbb{E}\left[|X-G|^2\right] : G \in \mathrm{L}^2(\mathcal{G})\right\}.$$

- Il s'agit d'une généralisation du calcul introductif !
- On peut remarquer que $\mathbb{E}[P_{\mathcal{G}}(X)] = \mathbb{E}[X]$ et que $P_{\mathcal{G}}(X) \geq 0$ p.s. si $X \geq 0$ p.s.
- Cette construction peut être étendue, d'une certaine manière, aux v.a. intégrables

Définition (Espérance conditionnelle)**.** Soient X une v.a.r. intégrable et $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. On appelle *(version de l')espérance conditionnelle* de X sachant $\mathcal{G}$ toute variable aléatoire Z vérifiant :

1. Z est $\mathcal{G}$-mesurable ;
2. Z est intégrable i.e. $\mathbb{E}[|Z|] < +\infty$;
3. Pour toute v.a. G, $\mathcal{G}$-mesurable et bornée, $\mathbb{E}[XG] = \mathbb{E}[ZG]$.

On note dans ce cas $Z = \mathbb{E}[X \mid \mathcal{G}]$.

Théorème. *Soient X une v.a.r. intégrable et $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. Il existe une version de l'espérance conditionnelle de X sachant $\mathcal{G}$.*
Si Z et Z' sont deux versions de l'espérance conditionnelle de X sachant $\mathcal{G}$, alors $Z = Z'$ p.s. i.e. $\mathbb{P}(Z = Z') = 1$.

- Lorsque $\mathcal{G} = \sigma(Y)$ où Y est une v.a. dans $\mathbf{R}^m$, on note $\mathbb{E}[X \mid Y]$ au lieu de $\mathbb{E}[X \mid \sigma(Y)]$
- $\mathbb{E}[X \mid Y]$ est la meilleure approximation de X connaissant Y
- D'après le lemme de factorisation, $Z = \mathbb{E}[X \mid Y]$ si
  1. $Z = h(Y)$ avec $h : \mathbf{R}^m \longrightarrow \mathbf{R}$ borélienne ;
  2. $\mathbb{E}[|h(Y)|] < \infty$;
  3. Pour toute $g : \mathbf{R}^m \longrightarrow \mathbf{R}$ borélienne et bornée, $\mathbb{E}[Xg(Y)] = \mathbb{E}[h(Y)g(Y)]$.

Exemple(s). 1. Soient X et Y indépendantes ; X intégrable. On a $\mathbb{E}[X \mid Y] = \mathbb{E}[X]$. En effet, notant $Z = \mathbb{E}[X]$:

(a) $Z = h(Y)$ avec h fonction constante égale à $\mathbb{E}[X]$;

(b) Z intégrable ;

(c) Pour g borélienne bornée, puisque X et Y son indépendantes

$$\mathbb{E}[Xg(Y)] = \mathbb{E}[X]\ \mathbb{E}[g(Y)] = \mathbb{E}[\mathbb{E}[X]\, g(Y)] = \mathbb{E}[Zg(Y)]\,.$$

2. Soient X et Y deux v.a. ; X intégrable et $\sigma(Y)$-mesurable. Alors $\mathbb{E}[X \mid Y] = X$. En effet,

(a) X est $\sigma(Y)$-mesurable ;

(b) X intégrable ;

(c) Pour g borélienne bornée,

$$\mathbb{E}[Xg(Y)] = \mathbb{E}[Xg(Y)]\,.$$

3. Soit Y discrète prenant les valeurs distinctes $(y_j)_{j\geq 1}$. Si X est intégrable

$$\mathbb{E}[X \mid Y] = \sum_{j\geq 1} \mathbb{E}[X \mid Y = y_j]\ \mathbf{1}_{Y=y_j}, \qquad \text{où } \mathbb{E}[X \mid Y = y_j] = \frac{\mathbb{E}\left[X\mathbf{1}_{Y=y_j}\right]}{\mathbb{P}(Y = y_j)}.$$

C'est le calcul introductif.

Proposition (Propriétés élémentaires)**.** *Soit $\mathcal{G}$ une sous-tribu de $\mathcal{F}$.*

1. Pour toute constante réelle c, $\mathbb{E}[c \,|\, \mathcal{G}] = c$;

2. Si a et b sont deux réels, U et V deux v.a. intégrables

$$\mathbb{E}[aU + bV \,|\, \mathcal{G}] = a\,\mathbb{E}[U \,|\, \mathcal{G}] + b\,\mathbb{E}[V \,|\, \mathcal{G}] \quad ;$$

3. Pour toute v.a. intégrable $\mathbb{E}[\mathbb{E}[X \,|\, \mathcal{G}]] = \mathbb{E}[X]$;

4. Si X est positive, $\mathbb{E}[X \,|\, \mathcal{G}]$ est positive ; en particulier, si X est intégrable

$$|\mathbb{E}[X \,|\, \mathcal{G}]| \leq \mathbb{E}[|X| \,|\, \mathcal{G}]\,.$$

Démonstration. La preuve est laissée en exercice. □

Exemple(s). 1. Si X et Y sont indépendantes, alors

$$\begin{aligned}\mathbb{E}\left[(X+Y)^2 \,|\, Y\right] &= \mathbb{E}\left[X^2 \,|\, Y\right] + 2\,\mathbb{E}[XY \,|\, Y] + \mathbb{E}\left[Y^2 \,|\, Y\right] = \mathbb{E}\left[X^2\right] + 2Y\,\mathbb{E}[X \,|\, Y] + Y^2, \\ &= \mathbb{E}\left[X^2\right] + 2Y\,\mathbb{E}[X] + Y^2 = \mathbb{V}(X) + (Y - \mathbb{E}[X])^2\,.\end{aligned}$$

2. Soient X une v.a. de loi uniforme sur $[0,1]$ et Y une v.a. à valeurs dans $\mathbf{N}^*$ telle que

$$\forall k \in \mathbf{N}^*, \qquad \mathbb{P}(Y=k \mid X) = \mathbb{E}\left[\mathbf{1}_{Y=k} \mid X\right] = (1-X)\, X^{k-1}.$$

On obtient facilement la loi de Y : pour $k \in \mathbf{N}^*$,

$$\mathbb{P}(Y=k) = \mathbb{E}\left[\mathbb{P}\left(Y=k \mid X\right)\right] = \mathbb{E}\left[(1-X)X^{k-1}\right] = \int_0^1 (1-x)x^{k-1}\, dx = \frac{1}{k(k+1)}.$$

Par ailleurs, Y étant discrète à valeurs dans $\mathbf{N}^*$, on a

$$\mathbb{E}\left[X \mid Y\right] = \sum_{k\geq 1} \mathbb{E}\left[X \mid Y=k\right]\, \mathbf{1}_{Y=k} = \sum_{k\geq 1} \mathbb{E}\left[X\mathbf{1}_{Y=k}\right]\, \mathbb{P}(Y=k)^{-1}\, \mathbf{1}_{Y=k},$$

et, pour tout $k \in \mathbf{N}^*$,

$$\mathbb{E}\left[X\mathbf{1}_{Y=k}\right] = \mathbb{E}\left[\mathbb{E}\left[X\mathbf{1}_{Y=k} \mid X\right]\right] = \mathbb{E}\left[X\mathbb{E}\left[\mathbf{1}_{Y=k} \mid X\right]\right] = \mathbb{E}\left[(1-X)X^{k}\right] = \frac{1}{(k+1)(k+2)}.$$

Par conséquent,

$$\mathbb{E}\left[X \mid Y\right] = \sum_{k\geq 1} \frac{k(k+1)}{(k+1)(k+2)}\, \mathbf{1}_{Y=k} = \sum_{k\geq 1} \frac{k}{k+2}\, \mathbf{1}_{Y=k} = \frac{Y}{Y+2}.$$

Calculer $\mathbb{E}\left[X \mid Y\right]$.

2.3. Propriétés.

Proposition (Projections emboitées)**.** *Soient X une variable aléatoire intégrable, $\mathcal{H} \subset \mathcal{G}$ des sous-tribus de $\mathcal{F}$. Alors,*

$$\mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{G}\right] \mid \mathcal{H}\right] = \mathbb{E}\left[X \mid \mathcal{H}\right].$$

- Cette propriété généralise $\mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{G}\right]\right] = \mathbb{E}\left[X\right]$.

Démonstration. Notons $Z = \mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{G}\right] \mid \mathcal{H}\right]$ et montrons que $Z = \mathbb{E}\left[X \mid \mathcal{H}\right]$. Il s'agit de montrer que :

1. Z est $\mathcal{H}$-mesurable ;
2. Z est intégrable ;
3. Pour toute v.a. H bornée et $\mathcal{H}$-mesurable,

$$\mathbb{E}\left[ZH\right] = \mathbb{E}\left[ZX\right].$$

Pour le premier point, comme $Z = \mathbb{E}\left[TRUC \mid \mathcal{H}\right]$, Z est $\mathcal{H}$–mesurable. Par construction de l'espérance conditionnelle, X étant intégrable, $\mathbb{E}\left[X \mid \mathcal{G}\right]$ est intégrable et $Z = \mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{G}\right] \mid \mathcal{H}\right]$ aussi. Soit H bornée et $\mathcal{H}$-mesurable. Comme $\mathcal{H} \subset \mathcal{G}$, H est aussi $\mathcal{G}$-mesurable et donc

$$\mathbb{E}\left[ZH\right] = \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{G}\right] \mid \mathcal{H}\right] H\right] = \mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{G}\right] H\right] = \mathbb{E}\left[XH\right].$$

$\square$

Proposition. *Soient X et Y deux v.a. et $\mathcal{G}$ une sous-tribu. On suppose que X et XY sont intégrable et que Y est $\mathcal{G}$–mesurable. Alors*

$$\mathbb{E}\left[XY \,|\, \mathcal{G}\right] = Y\,\mathbb{E}\left[X \,|\, \mathcal{G}\right].$$

- Il s'agit d'une propriété d'usage très fréquent.

Démonstration. Faisons la preuve dans le cas où Y est bornée. Notons $Z = Y\,\mathbb{E}\left[X \,|\, \mathcal{G}\right]$ et montrons que $Z = \mathbb{E}\left[XY \mid \mathcal{G}\right]$. Les deux premiers points de la définition sont évidents. Pour le 3$^{\text{e}}$, soit G une va $\mathcal{G}$-mesurable et bornée. Comme YG est $\mathcal{G}$-mesurable et bornée, on a, par définition de $\mathbb{E}\left[X \,|\, \mathcal{G}\right]$,

$$\mathbb{E}\left[GYX\right] = \mathbb{E}\left[GY\mathbb{E}\left[X \,|\, \mathcal{G}\right]\right] = \mathbb{E}\left[GZ\right].$$

$\square$

Proposition. *Soient X et Y deux variables aléatoires indépendantes à valeurs dans $\mathbf{R}^n$ et $\mathbf{R}^m$ et $h : \mathbf{R}^n \times \mathbf{R}^m \longrightarrow \mathbf{R}$ une fonction borélienne telle que $\mathbb{E}\left[|h(X,Y)|\right] < \infty$. Alors*

$$\mathbb{E}\left[h(X,Y) \,|\, Y\right] = H(Y), \quad \textit{avec} \quad H(y) = \mathbb{E}\left[h(X,y)\right].$$

- La preuve sera vue en TD dans le cas où (X,Y) possède une densité.

Exemple(s). Si X et Y sont des v.a. indépendantes, X de loi de Poisson de paramètre $\lambda > 0$, Y de loi exponentielle de paramètre 2λ, on a, Y^X étant positive,

$$\mathbb{E}\left[Y^X \,|\, Y\right] = H(Y) \quad \text{avec} \quad H(a) = \mathbb{E}\left[a^X\right] = e^{\lambda(a-1)}\ ; \qquad \mathbb{E}\left[Y^X \,|\, Y\right] = e^{\lambda(Y-1)}.$$

Par conséquent,

$$\mathbb{E}\left[Y^X\right] = \mathbb{E}\left[\mathbb{E}\left[Y^X \,|\, Y\right]\right] = e^{-\lambda}\,\mathbb{E}\left[e^{\lambda Y}\right] = e^{-\lambda}\int_0^{+\infty} e^{\lambda y}\,(2\lambda)e^{-2\lambda y}\,dy = 2e^{-\lambda}.$$

Proposition. *Soient X une v.a. intégrable, $\mathcal{G}$ et $\mathcal{H}$ deux sous-tribus de $\mathcal{F}$. On suppose que $\mathcal{H}$ est indépendante de la sous-tribu $\sigma\left(\sigma(X), \mathcal{G}\right)$. Alors*

$$\mathbb{E}\left[X \,|\, \sigma(\mathcal{G},\mathcal{H})\right] = \mathbb{E}\left[X \,|\, \mathcal{G}\right].$$

Remarque(s). Attention, il ne suffit pas que $\mathcal{H}$ et X soient indépendantes pour appliquer ce résultat ! $\mathcal{G}$ peut amener un couplage. Par exemple, si X et Y sont indépendantes, notant $\mathcal{H} = \sigma(Y)$ et $\mathcal{G} = \sigma(X+Y)$

$$\mathbb{E}\left[X \,|\, \sigma(\mathcal{G},\mathcal{H})\right] = X, \quad \mathbb{E}\left[X \,|\, \mathcal{G}\right] = \frac{X+Y}{2}.$$

Proposition (Intégrabilité)**.** *Soient $(X_n)_{n\geq 1}$ et X des v.a. intégrables.*

Convergence monotone. *On suppose que, pour tout $n \geq 1$, $0 \leq X_n \leq X_{n+1}$ presque sûrement. Alors,*

$$\lim_{n\to\infty} \mathbb{E}\left[X_n \,|\, \mathcal{G}\right] = \sup_{n\geq 1} \mathbb{E}\left[X_n \,|\, \mathcal{G}\right] = \mathbb{E}\left[\lim_{n\to\infty} X_n \,|\, \mathcal{G}\right] = \mathbb{E}\left[\sup_{n\geq 1} X_n \,|\, \mathcal{G}\right].$$

Lemme de Fatou. *On suppose que, pour tout $n \geq 1$, $X_n \geq 0$ presque sûrement. Alors,*

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n \,|\, \mathcal{G}\right] \leq \liminf_{n\to\infty} \mathbb{E}\left[X_n \,|\, \mathcal{G}\right].$$

Convergence dominée. *On suppose que $(X_n)_{n\geq 1}$ converge vers X presque sûrement et que $\sup_{n\geq 1} |X_n|$ est intégrable. Alors,*

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X| \,|\, \mathcal{G}\right] = 0 \; ; \quad \textit{en particulier, } \lim_{n\to\infty} \mathbb{E}\left[X_n \,|\, \mathcal{G}\right] = \mathbb{E}\left[X \,|\, \mathcal{G}\right].$$

- Ces propriétés se déduisent facilement de leurs analogues pour l'espérance classique.

Proposition (Inégalité de Jensen). *Soient X une v.a. intégrable et $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. Soit $g : \mathbf{R} \longrightarrow \mathbf{R}$ une fonction convexe telle que $\mathbb{E}\left[|g(X)|\right] < \infty$. Alors,*

$$g\left(\mathbb{E}\left[X \,|\, \mathcal{G}\right]\right) \leq \mathbb{E}\left[g(X) \,|\, \mathcal{G}\right].$$

- Une fonction g est convexe si, pour tous x, y et $\lambda \in [0,1]$,

$$g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)y.$$

  - ⋆ On a alors $g(y) \geq g(x) + g'(x)(y-x)$
  - ⋆ Si g est dérivable, g est convexe ssi g' est croissante.

- On utilise souvent l'inégalité de Jensen avec les fonctions $x \longmapsto |x|^p$ convexe dès que $p \geq 1$ et $x \longmapsto e^{ax}$ convexe pour tout réel a.

3. Loi conditionnelle.

`Paragraphe non traité en 2019/2020`

Définition. On appelle *noyau de transition* toute fonction $K : \mathbf{R}^m \times \mathcal{B}(\mathbf{R}^n) \longrightarrow [0,1]$ telle que :

1. Pour tout $B \in \mathcal{B}(\mathbf{R}^n)$, $y \longmapsto K(y,B)$ est borélienne ;
2. Pour tout $y \in \mathbf{R}^m$, $B \longmapsto K(y,B)$ est une mesure de probabilité notée $K(y,dx)$.

- $K(y,dx)$ est une famille de probabilités sur $\mathbf{R}^n$ indexée par $y \in \mathbf{R}^m$.

Théorème (Jirina). *Soient X et Y deux v.a. à valeurs dans $\mathbf{R}^n$ et $\mathbf{R}^m$. Il existe un noyau de transition $K(y,dx)$ tel que : pour toute fonction $f : \mathbf{R}^n \longrightarrow \mathbf{R}$ borélienne et bornée,*

$$\mathbb{E}\left[f(X) \,|\, Y\right] = \int_{\mathbf{R}^n} f(x)\, K(Y,dx) \quad \mathbb{P}-p.s.$$

- Un tel noyau de transition $K(y,dx)$ est appelée *(version de la) loi conditionnelle de X sachant Y.*

- On désigne $K(y,dx)$ par $\mathcal{L}(X \,|\, Y)$ ou $\mathcal{L}(X \,|\, Y = y)$.

Remarque(s). La loi du couple (X, Y) est obtenue à partir de $\mathcal{L}(X \mid Y)$ et $\mathcal{L}(Y)$. En effet, si f et g sont boréliennes et bornées

$$\begin{aligned}\mathbb{E}\left[f(X)g(Y)\right] &= \mathbb{E}\left[\mathbb{E}\left[g(Y)f(X) \mid Y\right]\right] = \mathbb{E}\left[g(Y)\,\mathbb{E}\left[f(X) \mid Y\right]\right], \\ &= \mathbb{E}\left[g(Y) \int_{\mathbf{R}^n} f(x)K(Y, dx)\right] = \int_{\mathbf{R}^m} g(y) \int_{\mathbf{R}^n} f(x)K(y, dx)\,\mathbb{P}_Y(dy), \\ &= \iint_{\mathbf{R}^n \times \mathbf{R}^m} g(y)f(x)\,K(y, dx)\,\mathbb{P}_Y(dy).\end{aligned}$$

Plus généralement, si $h : \mathbf{R}^n \times \mathbf{R}^m \longrightarrow \mathbf{R}$ est borélienne et bornée

$$\mathbb{E}\left[h(X, Y)\right] = \iint_{\mathbf{R}^n \times \mathbf{R}^m} h(x, y)\,K(y, dx)\,\mathbb{P}_Y(dy).$$

On retient $\mathbb{P}_{(X,Y)}(dx, dy) = K(y, dx)\,\mathbb{P}_Y(dy)$.

Exemple(s). 1. Supposons que X et Y sont indépendantes. On a, dans ce cas, pour toute f borélienne bornée

$$\mathbb{E}\left[f(X) \mid Y\right] = \mathbb{E}\left[f(X)\right] = \int_{\mathbf{R}^n} f(x)\,\mathbb{P}_X(dx).$$

Par conséquent, $K(y, dx) = \mathbb{P}_X(dx)$ soit $\mathcal{L}(X \mid Y = y) = \mathcal{L}(X)$.

2. Supposons que X est $\sigma(Y)$-mesurable. D'après le lemme de factorisation, $X = h(Y)$ avec $h : \mathbf{R}^m \longrightarrow \mathbf{R}^n$ borélienne. On a, pour toute f borélienne bornée,

$$\mathbb{E}\left[f(X) \mid Y\right] = f(X) = f(h(Y)) = \int_{\mathbf{R}^n} f(x)\,\delta_{h(Y)}(dx).$$

Il s'en suit $K(y, dx) = \delta_{h(y)}(dx)$.

3. On suppose que (X, Y) possède une densité $p(x, y)$. Pour f et g boréliennes bornées,

$$\begin{aligned}\mathbb{E}\left[f(X)g(Y)\right] &= \iint_{\mathbf{R}^n \times \mathbf{R}^m} f(x)g(y)\,p(x, y)\,dxdy = \int_{\mathbf{R}^m} g(y) \left(\int_{\mathbf{R}^n} f(x)\,p(x, y)\,dx\right) dy \\ &= \int_{\mathbf{R}^m} g(y) \left(\frac{\int_{\mathbf{R}^n} f(x)\,p(x, y)\,dx}{p_Y(y)}\right) p_Y(y)dy,\end{aligned}$$

où p_Y désigne la densité de Y i.e.

$$p_Y(y) = \int_{\mathbf{R}^n} p(x, y)\,dx.$$

Par conséquent,

$$\mathbb{E}\left[f(X) \mid Y\right] = \int_{\mathbf{R}^n} f(x)\frac{p(x, Y)}{p_Y(Y)}\,dx, \qquad K(y, dx) = \frac{p(x, y)}{p_Y(y)}\,dx.$$