# Conditionnement / Espérance conditionnelle

Jérémie Bigot

Cours de probabilités MA105
ISAE/SUPAERO 1A

**Année 2013 - 2014**

# Importance de la notion de probabilité conditionnelle

**Quand a-t-besoin de la la notion de probabilité conditionnelle ?**

- chaque fois que, pendant le déroulement d’une expérience aléatoire, une information partielle est fournie à l’expérimentateur

- un événement en conditionne un autre, si la réalisation de ce dernier dépend de la réalisation du premier.

**Les notions d’indépendance et de conditionnement sont donc étroitement liées.**

# Exemple de conditionnement discret

Envisageons les trois cas suivants :

1. On lance un dé à six faces : quelle est la probabilité d'obtenir le chiffre 3 ?
   **Réponse :** $1/6$

2. On lance un dé à six faces : quelle est la probabilité d'obtenir le chiffre 3 **sachant que le résultat est impair** ?
   **Réponse :** $1/3$

3. On lance un dé à six faces : quelle est la probabilité d'obtenir le chiffre 3 **sachant que le résultat est pair** ?
   **Réponse :** $0$

# Exemple de conditionnement continu

Modèle de déplacement d'un mobile (ex : véhicule) décrit pour un vecteur $x_t \in \mathbb{R}^2$ ou $\mathbb{R}^3$ au cours du temps $t \in \mathbb{R}_+$ :

- le vecteur $x_t$ est la position du mobile à l'instant $t$ dans le cas idéal d'un modèle théorique
- la position réelle $X_t$ est modélisée en introduisant un terme perturbateur aléatoire

$$X_t = x_t + V_t, \text{ avec } V_t \text{ v.a. centrée,}$$

- suivi du véhicule à partir de mesures $Y_t$ qui donnent une **approximation** de sa position réelle

$$Y_t = X_t + W_t, \text{ avec } W_t \text{ v.a. centrée.}$$

**Question :** connaissant les mesures $Y_{t_1}, \dots, Y_{t_n}$, comment prédire/estimer la position réelle du mobile au temps $t_{n+1}$ ?

**Réponse :** utilisation de l'espérance conditionnelle (filtre de Kalman)

# Objectifs du conditionnement

- Apprendre à utiliser de nouvelles informations pour augmenter la précision d'un modèle aléatoire (exemple : détection des spams)
- La définition du modèle conditionnel dans le cas discret est naturelle et remonte au XVIII siècle (Thomas Bayes, mathématicien britannique, 1702-1761). La définition générale est plus abstraite.
- Le formalisme des espaces de Hilbert et la notion projection orthogonale peuvent être utilisés pour définir les notions d'espérance conditionnelle et de loi conditionnelle.
- Le conditionnement ouvre la voie à **la statistique et l'inférence bayésienne** - Notion de modèle **a priori** et modèle **a posteriori**.

Approches fréquentistes $\neq$ Approches bayésiennes

# Définition de la probabilité conditionnelle

## Définition

*Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. Soit $B \in \mathcal{A}$ tel que $P(B) > 0$. On appelle* **probabilité de $A$ sachant $B$** *le nombre $P^B(A)$ (noté également $P(A|B)$) défini par*

$$P^B(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

**Lancé de dé** : on note $A$ “chiffre 3” et $B$ “chiffre impair”. On a que

$$A \cap B = A\ ;\ P(A \cap B) = \frac{\mathrm{card}(A \cap B)}{\mathrm{card}\Omega} = \frac{1}{6}\ ;\ P(B) = \frac{\mathrm{card}B}{\mathrm{card}\Omega} = \frac{3}{6} = \frac{1}{2}.$$

D’où

$$P^B(A) = P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

# Définition de la probabilité conditionnelle

## Définition

*Soit* $(\Omega, \mathcal{A}, P)$ *un espace probabilisé. Soit* $B \in \mathcal{A}$ *tel que* $P(B) > 0$*. On appelle* **probabilité de** $A$ **sachant** $B$ *le nombre* $P^B(A)$ *(noté également* $P(A|B)$*) défini par*

$$P^B(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

**Lancé de dé** : on note $A$ “chiffre 3” et $B$ “chiffre pair”. On a que

$$A \cap B = \emptyset \; ; \; P(A \cap B) = \frac{\mathrm{card}(A \cap B)}{\mathrm{card}\Omega} = 0 \; ; \; P(B) = \frac{\mathrm{card}B}{\mathrm{card}\Omega} = \frac{3}{6} = \frac{1}{2}.$$

D’où

$$P^B(A) = P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0}{\frac{1}{2}} = 0.$$

# Définition de la probabilité conditionnelle

## Proposition

*L'application $P^B : A \mapsto P^B(A)$ est une probabilité sur $(\Omega, \mathcal{A})$.*

**Preuve :** i) $P^B(\Omega) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$ et $P^B(A) = \frac{P(A \cap B)}{P(B)} \in [0,1]$ car $A \cap B \subset B$.

ii) Soit suite d'événements $A_n$, incompatibles 2 à 2 (i.e. $A_n \cap A_m = \emptyset$). Les événements $A_n \cap B$ sont aussi incompatibles 2 à 2 et donc :

$$\begin{aligned} P^B\left(\bigcup_n A_n\right) &= \frac{P\left(\left(\bigcup_n A_n\right) \cap B\right)}{P(B)} = \frac{P\left(\bigcup_n (A_n \cap B)\right)}{P(B)} \\ &= \frac{\sum_n P(A_n \cap B)}{P(B)} = \sum_n P^B(A_n) \end{aligned}$$

# Formule des probabilités totales

## Théorème

*Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. Soit $(B_i)_{i\in I}$ un système complet d'événements de probabilités non nulles i.e.*

(i) $B_i \in \mathcal{A}$ *pour tout* $i \in I$,

(ii) $\bigcup_{i\in I} B_i = \Omega$,

(iii) *si* $i \neq j$, *alors* $B_i \cap B_j = \emptyset$,

(iv) $P(B_i) \neq 0$ *pour tout* $i \in I$ *(ensemble dénombrable).*

*Alors, pour tout* $A \in \mathcal{A}$,

$$P(A) = \sum_{i\in I} P^{B_i}(A)P(B_i) = \sum_{i\in I} P(A|B_i)P(B_i).$$

# Formule des probabilités totales

**Cas particulier très utile :**

$$\begin{aligned} P(A) &= P^B(A)P(B) + P^{\overline{B}}(A)P(\overline{B}) \\ &= P(A|B)P(B) + P(A|\overline{B})P(\overline{B}) \end{aligned}$$

**Exemple :** Le taux de réussite d'un examen donné est de 60 % pour les candidats issus de l'établissement A et de 80 % pour ceux issus de B. D'autre part, 55 % des candidats proviennent de A et 45% de B.

Quel est le taux d'échec à cet examen ?

# Formule des probabilités totales

**Exemple :** Le taux de réussite d'un examen donné est de 60 % pour les candidats issus de l'établissement A et de 80 % pour ceux issus de B. D'autre part, 55 % des candidats proviennent de A et 45% de B.

**Solution :** soit les événements $A$ : "être issu de $A$", $B$ : "être issu de $B$", $R$ : "réussir" et $E$ : "être en échec".

Par la formule des probabilités totales on a que le taux de réussite est :

$$\mathbb{P}(R) = \mathbb{P}(R|A)\mathbb{P}(A) + \mathbb{P}(R|B)\mathbb{P}(B) = \frac{60 \times 55}{100^2} + \frac{80 \times 45}{100^2} = 69\%,$$

et donc

$$\mathbb{P}(E) = 1 - \mathbb{P}(R) = 31\% : \text{ taux d'échec.}$$

# Formule de Bayes

Soit $A$ et $B$ deux évènements. On a que

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ et } = P(B|A) = \frac{P(B \cap A)}{P(A)}$$

Donc, on peut aussi écrire que (**formule de Bayes**) :

$$\begin{aligned} P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \\ &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\overline{A})P(\overline{A})}, \end{aligned}$$

écrite (souvent !) sous la forme

$$P(A|B) \propto P(B|A)P(A)$$

# Lois conditionnelles pour un couple discret

Soit $X$ et $Y$ deux v.a.r. discrètes, définies sur $(\Omega, \mathcal{A}, P)$, avec

$$X(\Omega) = \{x_i \; ; \; i \in I\} \text{ et } Y(\Omega) = \{y_j \; ; \; j \in J\}, \; I, J \text{ dénombrables .}$$

**Loi de probabilité du couple** $(X, Y)$ est déterminée par

$$p_{ij} = P([X = x_i] \cap [Y = y_j]) \text{ pour } i \in I \text{ et } j \in J.$$

**Lois marginales** de $X$ et $Y$

$$P(X = x_i) = p_{i.} = \sum_j p_{ij} \quad \text{et} \quad P(Y = y_j) = p_{.j} = \sum_i p_{ij}.$$

## Définition

*Si $p_{i.} \neq 0$, la loi conditionnelle de $Y$ sachant $[X = x_i]$ est définie par :*

$$P^{[X=x_i]}(Y = y_j) = P(Y = y_j | X = x_i) = \frac{P([X = x_i] \cap [Y = y_j])}{P(X = x_i)} = \frac{p_{ij}}{p_{i.}}, \; j \in J.$$

*Si $p_{i.} = 0$, alors, par convention, $P^{[X=x_i]}(Y = y_j) = 0$.*

# Espérance conditionnelle dans le cas discret

## Définition

*Supposons que* $\mathbb{E}(|Y|) < +\infty$.

*L'espérance d'une v.a. dont la loi est la loi conditionnelle de* $Y$ *à l'événement* $[X = x_i]$ *est appelée* **espérance conditionnelle de** $Y$ **à l'événement** $[X = x_i]$.

*Elle est notée* $\mathbb{E}^{[X=x_i]}(Y)$ *ou* $\mathbb{E}(Y|X = x_i)$. *On a donc*

$$\mathbb{E}^{[X=x_i]}(Y) = \sum_j y_j P^{[X=x_i]}(Y = y_j),$$

*notée également*

$$\mathbb{E}(Y|X = x_i) = \sum_j y_j P(Y = y_j|X = x_i).$$

# Espérance conditionnelle dans le cas discret

**Exemple** : soient $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ indépendantes (loi de Poisson de paramètres $\lambda > 0$ et $\mu > 0$).

Les v.a. $X$ et $Y$ sont discrètes et à valeurs dans $\mathbb{N}$ de loi

$$P([X = i]) = e^{-\lambda} \frac{\lambda^i}{i!},\ i \in \mathbb{N},$$

et

$$P([Y = j]) = e^{-\mu} \frac{\mu^j}{j!},\ j \in \mathbb{N}.$$

**Proposition**

*La v.a. $X + Y$ suit une de loi de Poisson de paramètre $\lambda + \mu$.*

# Espérance conditionnelle dans le cas discret

**Exemple** : soient $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ indépendantes.

Etant donné que $X+Y \sim \mathcal{P}(\lambda+\mu)$, on a que pour $0 \leq i \leq n$,

$$\begin{aligned}
P^{[X+Y=n]}([X=i]) &= \frac{P([X=i]\cap[X+Y=n])}{P([X+Y=n])} = \frac{P([X=i]\cap[Y=n-i])}{P([X+Y=n])} \\
&= \frac{P([X=i])P([Y=n-i])}{P([X+Y=n])} = \frac{e^{-\lambda}\frac{\lambda^i}{i!}\, e^{-\mu}\frac{\mu^{n-i}}{(n-i)!}}{e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^n}{n!}} \\
&= C_n^i \left(\frac{\lambda}{\lambda+\mu}\right)^i \left(1-\frac{\lambda}{\lambda+\mu}\right)^{n-i} = P_Z(\{i\})
\end{aligned}$$

où $Z$ est de loi binomiale $\mathcal{B}\left(n, \frac{\lambda}{\lambda+\mu}\right)$. Ainsi, la loi conditionnelle de $X$ à $[X+Y=n]$ est la loi binomiale $\mathcal{B}\left(n, \frac{\lambda}{\lambda+\mu}\right)$, d'espérance $\frac{n\lambda}{\lambda+\mu}$ et donc

$$\mathbb{E}^{[X+Y=n]}(X) = \mathbb{E}(X|X+Y=n) = \frac{n\mu}{\lambda+\mu}.$$

# Espérance conditionnelle dans le cas discret

## Définition

*Soit $X$ une v.a. réelle discrète sur $(\Omega, \mathcal{A}, P)$ telle que, pour tout $x \in X(\Omega)$, $P([X = x]) \neq 0$ et soit $Y$ une v.a. réelle discrète sur $(\Omega, \mathcal{A}, P)$ admettant une espérance i.e. $\mathbb{E}(|Y|) < +\infty$.*

*On appelle* **espérance conditionnelle de $Y$ sachant $X$**,

*la* **variable aléatoire** *discrète*

$$\mathbb{E}^X(Y) = h(X),$$

*où $h : X(\Omega) \to \mathbb{R}$ est la fonction définie par*

$$h(x) = \mathbb{E}^{[X=x]}(Y),$$

*pour tout $x \in X(\Omega)$.*

**Remarque :** on la note également $\mathbb{E}(Y|X) = \mathbb{E}^X(Y)$.

# Espérance conditionnelle dans le cas discret

**Exemple** : soient $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ indépendantes.

On a que

$$h(n) = \mathbb{E}^{[X+Y=n]}(Y) = \mathbb{E}(X|X+Y=n) = \frac{n\mu}{\lambda+\mu},\ n \in \mathbb{N},$$

et donc

$$\mathbb{E}^{X+Y}(X) = \mathbb{E}(Y|X+Y) = \frac{(X+Y)\lambda}{\lambda+\mu}$$

**Remarque importante :** $\mathbb{E}^{X+Y}(X)$ est une variable aléatoire !

# Espérance conditionnelle dans le cas discret

**Théorème (Théorème de l'espérance totale)**

*Soit $X$ et $Y$ deux v.a.r. discrètes définies sur le même espace telles que $\mathbb{E}(|Y|) < +\infty$. Alors la v.a.r. discrète $\mathbb{E}^X(Y)$ admet une espérance et*

$$\mathbb{E}(\mathbb{E}^X(Y)) = \mathbb{E}(Y),$$

*relation notée également*

$$\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X)).$$

# Espérance conditionnelle dans le cas discret

**Preuve** (Théorème de l'espérance totale) : on a que

$$h(x_i) = \mathbb{E}(Y|X = x_i) = \sum_j y_j P(Y = y_j|X = x_i).$$

L'espérance de la v.a. $h(X) = \mathbb{E}(Y|X)$ peut donc s'écrire

$$\begin{aligned}
\mathbb{E}(\mathbb{E}(Y|X)) &= \mathbb{E}(h(X)) = \sum_i h(x_i) P(X = x_i) \\
&= \sum_i \left( \sum_j y_j P(Y = y_j|X = x_i) \right) P(X = x_i) \\
&= \sum_j y_j \left( \sum_i P(Y = y_j|X = x_i) P(X = x_i) \right) \\
&= \sum_j y_j P(Y = y_j) = \mathbb{E}(Y).
\end{aligned}$$

# Lois conditionnelles dans le cas continu

Soit $(X, Y) \in \mathbb{R}^2$ un couple de variables aléatoires de densité $f_{X,Y}$ admettant pour densités marginales $f_X$ et $f_Y$.

## Définition

*Pour tout $x \in \mathbb{R}$ tel que $f_X(x) \neq 0$, la* **densité conditionnelle** *de $Y$ sachant $[X = x]$ est la fonction $f_Y^{X=x} : \mathbb{R} \to \mathbb{R}_+$ définie par*

$$f_Y^{X=x} = \frac{f_{X,Y}(x, y)}{f_X(x)}.$$

**Parallèle avec les couples discrets** : on peut faire l'analogue avec la loi de $Y$ sachant $[X = x_i]$ :

$$P^{[X=x_i]}(Y = y_j) = \frac{P([X = x_i] \cap [Y = y_j])}{P(X = x_i)} = \frac{p_{ij}}{p_{i.}},\ j \in J.$$

# Espérance conditionnelle dans le cas continu

## Définition

*Supposons que* $\mathbb{E}(|Y|) < +\infty$.**L'espérance conditionnelle** *de* $Y$ *à "l'événement* $[X = x]$*" est le réel :*

$$\mathbb{E}^{X=x}(Y) = \int_{\mathbb{R}} y f_Y^{X=x}(y)\, dy.$$

**L'espérance conditionnelle** *de* $Y$ *sachant* $X$ *est la* **variable aléatoire** *réelle*

$$\mathbb{E}^X(Y) = \mathbb{E}(Y|X) = h(X)$$

*avec* $h : \mathbb{R} \to \mathbb{R}$ *la fonction définie par* $h(x) = \mathbb{E}^{X=x}(Y),\ x \in \mathbb{R}$*.*

**Théorème de l'espérance totale** (comme dans le cas discret).
Si $\mathbb{E}(Y)$ existe :

$$\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}^X(Y)) = \mathbb{E}(\mathbb{E}(Y|X)).$$

# Espérance conditionnelle dans le cas continu

**Exemple :** soit un couple aléatoire $(X, Y)$ de densité

$$f_{X,Y}(x,y) = \frac{1}{x} e^{-x-y/x} \mathbb{1}_{\{[0,+\infty[\times[0,+\infty[\}}(x,y).$$

**On veut calculer $\mathbb{E}(Y)$ ?**

Calculons tout d'abord la densité conditionnelle

$$f_Y^{X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_{X,Y}(x,y)}{\int_{\mathbb{R}} f_{X,Y}(x,y)dy} = \frac{1}{x} e^{-y/x},$$

qui est une loi exponentielle de paramètre $\frac{1}{x}$. On a donc que

$$\mathbb{E}^{X=x}(Y) = \int_{\mathbb{R}} y f_Y^{X=x}(y)\, dy = \int_0^{+\infty} y \frac{1}{x} e^{-y/x} dy = x,$$

et donc $\mathbb{E}(Y|X) = X$.

# Espérance conditionnelle dans le cas continu

**Exemple :** soit un couple aléatoire $(X, Y)$ de densité

$$f_{X,Y}(x,y) = \frac{1}{x} e^{-x-y/x} \mathbb{1}_{\{[0,+\infty[\times[0,+\infty[\}}(x,y).$$

**On veut calculer $\mathbb{E}(Y)$ ?**

Etant donné que

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)dy = e^{-x} \mathbb{1}_{\{[0,+\infty[\}}(x)$$

(loi exponentielle de paramètre 1) et $\mathbb{E}(Y|X) = X$, par le Théorème de l'espérance totale, on obtient que

$$\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(X) = \int_{\mathbb{R}} x f_X(x)dx = \int_0^{+\infty} e^{-x}dx = 1.$$

# Extensions de la notion de conditionnement

- Conditionnement dans le cas d'un couple mixte $Y$ v.a. discrète par rapport à $X$ v.a. continue / $Y$ v.a. continue par rapport à $X$ v.a. discrète
- Extension au cas multidimensionnel $(X, Y) \in \mathbb{R}^n \times \mathbb{R}^m$

# Calcul de probabilités par conditionnement

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. Soit $A \in \mathcal{A}$ un évènement quelconque et $X$ une v.a. réelle. Posons

$$Y(\omega) = \mathbb{1}_A(\omega), \ \forall \omega \in \Omega.$$

On peut remarquer que $P(A) = \mathbb{E}(\mathbb{1}_A) = \mathbb{E}(Y)$. Par le théorème de l'espérance totale, on a que

$$P(A) = \mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(\mathbb{E}(\mathbb{1}_A|X)) = \mathbb{E}(P(A|X)),$$

où par convention on pose que

$$P(A|X) = \mathbb{E}(\mathbb{1}_A|X).$$

# Calcul de probabilités par conditionnement

**Cas discret :** si $X$ est une v.a.r. discrètes avec $X(\Omega) = \{x_i \ ; \ i \in I\}$ on obtient alors que

$$P(A) = \mathbb{E}(P(A|X)) = \sum_i P(A|X = x_i)P(X = x_i),$$

et on retrouve donc la formule des probabilités totales !

# Calcul de probabilités par conditionnement

**Cas continu :** si $X$ est une v.a.r. continue de densité $f_X$ on a alors que

$$P(A) = \mathbb{E}(P(A|X)) = \int_{\mathbb{R}} P(A|X = x) f_X(x) dx,$$

où l'on pose que $P(A|X = x) = \lim_{h\to 0} P(A \mid x \leq X < x + h)$.

**Exemple :** soient $X$ et $Y$ deux v.a. indépendantes de densité $f_X$ et $f_Y$. On peut alors calculer la densité de $X + Y$ par conditionnement :

$$\begin{aligned} F_{X+Y}(t) &= P(X + Y < t) = \int_{\mathbb{R}} P(X + Y < t|X = x) f_X(x) dx, \\ &= \int_{\mathbb{R}} P(x + Y < t|X = x) f_X(x) dx \\ &= \int_{\mathbb{R}} P(Y < t - x) f_X(x) dx \text{ (par indépendance de } X \text{ et } Y) \\ &= \int_{\mathbb{R}} F_Y(t - x) f_X(x) dx, \end{aligned}$$

et donc $f_{X+Y}(t) = F'_{X+Y}(t) = \int_{\mathbb{R}} f_Y(t - x) f_X(x) dx$.

# Lien avec la projection orthogonale ?

Soit $X$ et $Y$ deux v.a.r. discrètes, définies sur $(\Omega, \mathcal{A}, P)$, avec

$$X(\Omega) = \{x_i \; ; \; i \in I\} \text{ et } Y(\Omega) = \{y_j \; ; \; j \in J\}, \; I, J \text{ dénombrables .}$$

On rappelle que $\mathbb{E}(Y|X = x_i) = \sum_j y_j P(Y = y_j|X = x_i)$.

Soit $h : \mathbb{R} \to \mathbb{R}$ une fonction déterministe. On a alors que

$$\begin{aligned}
\mathbb{E}\left(\mathbb{E}(Y|X)h(X)\right) &= \sum_i \mathbb{E}(Y|X = x_i)h(x_i)P(X = x_i) \\
&= \sum_i \left(\sum_j y_j P(Y = y_j|X = x_i)\right) h(x_i)P(X = x_i) \\
&= \sum_i \sum_j y_j h(x_i) P([Y = y_j] \cap [X = x_i]) = \mathbb{E}\left(Yh(X)\right)
\end{aligned}$$

# Lien avec la projection orthogonale ?

On a donc que pour toute fonction $h : \mathbb{R} \to \mathbb{R}$

$$\mathbb{E}\left( \, (Y - \mathbb{E}(Y|X)) \, h(X) \, \right) = 0.$$

## Proposition

*Soit $(V, \langle \cdot, \cdot \rangle)$ un espace de Hilbert et $V_0 \subset V$ un sous-espace vectoriel fermé de $V$. Soit $v \in V$. La projection orthogonale de $v$ sur $V_0$ est le vecteur de $v_0 \in V_0$ tel que*

$$\|v - v_0\|^2 \leq \|v - u\|^2, \text{ pour tout } u \in V_0.$$

*Il est aussi caractérisé par la propriété*

$$\langle v - v_0, u \rangle = 0, \text{ pour tout } u \in V_0.$$

# Lien avec la projection orthogonale ?

On a donc que pour toute fonction $h : \mathbb{R} \to \mathbb{R}$

$$\mathbb{E}\left(\underbrace{(Y - \mathbb{E}(Y|X))}_{v-v_0}\ \underbrace{h(X)}_{u}\right) = 0 \text{ avec } v = Y,\ v_0 = \mathbb{E}(Y|X).$$

## Proposition

*Soit $(V, \langle\cdot,\cdot\rangle, \|\cdot\|)$ un espace de Hilbert et $V_0 \subset V$ un sous-espace vectoriel fermé de $V$. Soit $v \in V$. La projection orthogonale de $v$ sur $V_0$ est le vecteur de $v_0 \in V_0$ tel que*

$$\|v - v_0\|^2 \leq \|v - u\|^2, \text{ pour tout } u \in V_0.$$

*Il est aussi caractérisé par la propriété*

$$\langle v - v_0, u\rangle = 0, \text{ pour tout } u \in V_0.$$

# Structure hilbertienne

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé, et $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ une v.a.r. (discrète ou continue) i.e. une application mesurable.

## Définition

*On note par $L^2(\Omega, \mathcal{A}, P)$ l'espace des v.a.r. $X$ de carré intégrable i.e telles que $\mathbb{E}(X^2) < +\infty$.*

## Théorème

*L'espace $L^2(\Omega, \mathcal{A}, P)$ est un espace de Hilbert lorsqu'il est muni du produit scalaire*

$$\langle X, Y \rangle = \mathbb{E}(XY),$$

*et de la norme quadratique associée $\|X\|_2 = \sqrt{\mathbb{E}(X^2)}$.*

# Structure hilbertienne

## Définition (Information disponible par la connaissance de $X$)

*Soit* $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ *une v.a.r. On note par*

$$\mathcal{B}_X = X^{-1}(\mathcal{B}_{\mathbb{R}}) = \{X^{-1}(B)\ ;\ B \in \mathcal{B}_{\mathbb{R}}\} \subset \mathcal{A}$$

*la tribu engendrée par* $X$*.*

## Théorème (Doob)

*Soit* $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ *et* $Y : \Omega \to \mathbb{R}$ *deux v.a.r. Alors* $Y$ *est mesurable par rapport à la tribu* $\mathcal{B}_X$ *ssi il existe une fonction* $h : (\mathbb{R}, \mathcal{B}_{\mathbb{R}}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ *mesurable telle que*

$$Y = h(X).$$

# Structure hilbertienne

## Définition

*Soit $L^2(\Omega, \mathcal{B}_X, P)$ l'espace des v.a.r. $Z : \Omega \to \mathbb{R}$ de carré intégrable et mesurables par rapport à $\mathcal{B}_X$.*

*Alors $L^2(\Omega, \mathcal{B}_X, P)$ est un sous-espace vectoriel fermé de $L^2(\Omega, \mathcal{A}, P)$.*

## Théorème (et Définition)

*Soit $X$ et $Y$ deux v.a. quelconques de $L^2(\Omega, \mathcal{A}, P)$.*

*Alors,*

$\mathbb{E}(Y|X)$ *est l'espérance conditionnelle de $Y$ sachant $X$*

$$\Updownarrow$$

$\mathbb{E}(Y|X)$ *est* **la projection orthogonale** *de $Y$ sur $L^2(\Omega, \mathcal{B}_X, P)$).*

# Quelques propriétés de l'espérance conditionnelle

- L'espérance conditionnelle de $Y$ sachant $X$ est la meilleure approximation quadratique que l'on puisse avoir de $Y$ par une v.a.r. de la forme $h(X)$ i.e.

$$\|Y-\mathbb{E}(Y|X)\|^2=\mathbb{E}\left(Y-\mathbb{E}(Y|X)\right)^2\leq\|Y-h(X)\|^2=\mathbb{E}((Y-h(X))^2)$$

pour toute fonction $h:(\mathbb{R},\mathcal{B}_{\mathbb{R}})\to(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ mesurable.

- Par la propriété de projection orthogonale

$$\begin{aligned}\langle h(X),Y\rangle &= \langle h(X),\mathbb{E}(Y|X)\rangle\\ \mathbb{E}(h(X)Y) &= \mathbb{E}\left(h(X)\mathbb{E}(Y|X)\right)\end{aligned}$$

pour toute fonction $h:(\mathbb{R},\mathcal{B}_{\mathbb{R}})\to(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ mesurable.

# Quelques propriétés de l'espérance conditionnelle

- Si $Y$ est indépendante de $X$, alors $\mathbb{E}(Y|X) = \mathbb{E}(Y)$
- Pour toute v.a.r. $Y$ intégrable et toute fonction $h : \mathbb{R} \to \mathbb{R}$ bornée
$$\mathbb{E}(h(X)Y|X) = h(X)\mathbb{E}(Y|X)$$
- Propriété de l'espérance totale : pour tout couple aléatoire $(X, Y)$
$$\mathbb{E}(\mathbb{E}^X(Y)) = \mathbb{E}(Y)$$
- Pour toute v.a.r. $X$ de $L^2(\Omega, \mathcal{A}, P)$, $\mathbb{E}(X)$ s'identifie à **la projection orthogonale de $X$ sur le sous-espace des v.a.r. constantes.**

# Variance conditionnelle

## Définition

*Étant données deux v.a.r. $X$ et $Y$, la* **variance conditionnelle** *de $Y$ sachant $X$ est la variable aléatoire :*

$$\mathbb{E}^X\left((Y-\mathbb{E}^X(Y))^2\right) =_{(noté)} \mathrm{var}^X(Y).$$

## Théorème (Propriété de la variance totale)

*Pour tout couple de variables aléatoires $(X,Y)$ :*

$$\begin{aligned}
\mathrm{var}(Y) &= \mathbb{E}(\mathrm{var}^X(Y)) + \mathrm{var}(\mathbb{E}^X(Y)) \\
\mathbb{E}\left(Y-\mathbb{E}(Y)\right)^2 &= \mathbb{E}\left(\mathbb{E}^X\left((Y-\mathbb{E}^X(Y))^2\right)\right) + \mathbb{E}\left(\mathbb{E}^X(Y)-\mathbb{E}(\mathbb{E}^X(Y))\right)^2 \\
\|Y-\mathbb{E}(Y)\|^2 &= \|Y-\mathbb{E}^X(Y)\|^2 + \|\mathbb{E}^X(Y)-\mathbb{E}(Y)\|^2
\end{aligned}$$

*Dans le cadre de la géométrie hilbertienne de $L^2(\Omega,\mathcal{A},P)$, la propriété de la variance totale équivaut donc au théorème de Pythagore.*

# Variance conditionnelle

Soit $(X_i)_{i\in\mathbb{N}}$ une suite de v.a. indépendantes entre elles et de même loi qu'une v.a. $X$.

**Exemple 1** - Espérance et variance de la somme d'un nombre $n$ **fixé** de v.a.

On a que

$$\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = n\,\mathbb{E}(X) \quad \text{et} \quad \operatorname{var}\left(\sum_{i=1}^{n} X_i\right) = n\,\operatorname{var}(X).$$

**Exemple 2** - Espérance et variance de la somme d'un nombre **aléatoire** de v.a.

Soit $N \in \mathbb{N}^*$ une v.a. discrète indépendante des $X_i$, et considérons la somme aléatoire

$$S_N = \sum_{i=1}^{N} X_i$$

**Question :** quelle est l'espérance et la variance de $S$ ?

# Variance conditionnelle

Soit $(X_i)_{i\in\mathbb{N}}$ une suite de v.a. indépendantes entre elles et de même loi qu'une v.a. $X$.

On a que

$$\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = n\,\mathbb{E}(X) \quad \text{et} \quad \mathrm{var}\left(\sum_{i=1}^{n} X_i\right) = n\,\mathrm{var}(X).$$

Soit $N \in \mathbb{N}^*$ une v.a. discrète indépendante des $X_i$, et considérons la somme aléatoire $S = \sum_{i=1}^{N} X_i$

On établit (cf. PC) que $\mathbb{E}(S|N) = N\mathbb{E}(X)$ et $\mathrm{var}(S|N) = N\mathrm{var}(X)$, ce qui implique

$$\mathbb{E}(S) = \mathbb{E}\left(\mathbb{E}(S|N)\right) = \mathbb{E}(N)\mathbb{E}(X)$$

et

$$\begin{aligned}\mathrm{var}(S) &= \mathbb{E}(\mathrm{var}^N(S)) + \mathrm{var}(\mathbb{E}^N(S)) \\ &= \mathbb{E}(N)\mathrm{var}(X) + \mathrm{var}(N)(\mathbb{E}(X))^2.\end{aligned}$$