Chapitre 3 : Espérance conditionnelle

Introduction :

La définition de l'espérance conditionnelle dans le cadre général, ce n'est pas si facile $\Rightarrow$ étude étape par étape

I Cadre discret

On considère un couple de variables aléatoires (X,Y), c'est à dire une application mesurable

$$(X,Y) : \begin{cases} (\Omega, \mathcal{T}, \mathbb{P}) \longrightarrow \mathcal{X} \times \mathcal{Y} \\ \omega \longrightarrow (X(\omega), Y(\omega)) \end{cases}$$

Puisque cadre discret $\mathcal{X} = (x_i)_{i \in I}$ et $\mathcal{Y} = (y_j)_{j \in J}$ au plus dénombrables

La loi de ce couple (X,Y) est définie par la donnée des quantités p_{ij} définies par :

$$p_{ij} = \mathbb{P}(X = x_i, Y = y_j)$$

Ces quantités satisfont

- $\forall (i,j) \quad p_{ij} \in [0,1]$
- $\sum_{i \in I, j \in J} p_{ij} = 1$

définition loi marginale

Soit (X,Y) un couple de variables aléatoires.

La loi de X, appelée loi marginale, est définie par les quantités $p_{i.}$

$$p_{i.} = \mathbb{P}(X = x_i) = \sum_{j \in J} \mathbb{P}(X = x_i, Y = y_j) = \sum_{j \in J} p_{ij}$$

De façon similaire, la loi marginale de Y est définie par les quantités $p_{.j}$

$$p_{.j} = \sum_{i \in I} p_{ij}$$

proposition

Les variables marginales X et Y sont indépendantes si et seulement si

$$\forall (i,j) \in I \times J, \quad P_{ij} = P_{i.} \times P_{.j}$$

définition : probabilité conditionnelle

Soit $x_i \in X$. La loi conditionnelle de Y sachant $X = x_i$ est la loi discrète prenant les valeurs y_j avec la probabilité $P_{j|i} = P(Y = y_j \mid X = x_i)$ où

$$P_{j|i} = \frac{P_{ij}}{P_{i.}} \quad \text{(probabilité conditionnelle)}$$

Rq

Soit $i \in I$. On peut avoir $P_{ij} = 0$ à savoir que l'évènement $X = x_i$ et $Y = y_j$ ne se réalise pas. Par contre, on ne peut pas avoir $P_{i.} = 0$ car alors $x_i \notin X$

définition : espérance conditionnelle

Supposons Y intégrable.

La variable aléatoire qui prend les valeurs $(E[Y \mid X = x_i])_i$ avec les probabilités $P_{i.}$ est l'espérance conditionnelle de Y sachant X, notée $E[Y|X]$

Rq : $$E[Y \mid X = x_i] = \sum_{j \in J} y_j \times P_{j|i}$$

/!\ l'espérance conditionnelle est une variable aléatoire

Théorème :

Soit Y une variable aléatoire intégrable.

Alors $E[Y|X]$ est intégrable et $E[E[Y|X]] = E[Y]$

preuve :

$$E[E[Y|X]] = \sum_{i \in I} E[Y \mid X = x_i] \cdot P_{i.}$$

$$= \sum_{i \in I} \sum_{j \in J} y_j P_{j|i} P_{i.}$$

$$= \sum_{i \in I} \sum_{j \in J} y_j P_{ij}$$

$$= \sum_{j \in J} y_j \underbrace{\left(\sum_{i \in I} P_{ij} \right)}_{= P_{\cdot j}}$$

$$= E[Y]$$

Propriétés :

Soit Y une variable aléatoire intégrable.

Si X et Y sont intégrables, alors $E[Y|X] = E[Y]$

Théorème de transfert :

Soit h une fonction de $\mathbb{R}^2$ dans $\mathbb{R}$.

Sous réserve d'intégrabilité :

$$E[h(X,Y)] = \sum_{\substack{i \in I \\ j \in J}} h(x_i, y_j) P_{ij}$$

Si $h(x,y) = f(x) \times g(y)$ et X et Y indépendants, alors

$$E[h(X,Y)] = E[f(X)]\, E[g(Y)]$$

Plus généralement

$$E[h(X,Y)] = \sum_{\substack{i \in I \\ j \in J}} h(x_i, y_j) P_{ij}$$

$$= \sum_{i \in I} \left(\sum_{j \in J} h(x_i, y_j) P_{j|i} \right) P_{i\cdot}$$

$$= \sum_{i \in I} E[h(x_i, Y) \mid X = x_i]\, P(X = x_i)$$

$$= E\big[\, E[h(X,Y) \mid X] \,\big]$$

où $E[h(X,Y) \mid X]$ est la variable qui prend les valeurs $E[h(x_i, Y) | X = x_i]$ avec les probabilités $P_{i\cdot}$

II Cadre continu (au sens à densité)

Soit (X,Y) un couple continu de variables aléatoires. La densité jointe f satisfait

- $\forall x,y \quad f(x,y) \geq 0$
- $\displaystyle\int_{\mathbb{R}^2} f(x,y)\,dx\,dy = 1$

et $P_{X,Y}(B) = \displaystyle\int_B f(x,y)\,dx\,dy$ où B partie de $\mathbb{R}^2$

<u>définition</u> : loi marginale

Si (X,Y) est un couple continu, alors X et Y sont aussi des variables continues de loi de densité respective f_X et f_Y définies par

$$f_X(x) = f(x) = \int_{\mathbb{R}} f(x,y)\,dy \quad \text{et} \quad f_Y(y) = f(y) = \int_{\mathbb{R}} f(x,y)\,dx$$

<u>proposition</u> :

Les variables X et Y sont indépendantes si et seulement si

$$\forall x,y \quad , \quad f(x,y) = f_X(x) \times f_Y(y)$$

<u>définition</u> : loi conditionnelle

La densité conditionnelle de Y sachant $X = x$ est

$$f(y|x) = \begin{cases} \dfrac{f(x,y)}{f_X(x)} & \text{si } f_X(x) \neq 0 \\ 0 & \text{si } f_X(x) = 0 \end{cases}$$

<u>définition</u> : Espérance conditionnelle

La variable aléatoire qui prend la valeur $E[Y|X=x]$ avec la densité $f_X(x)$ est l'espérance conditionnelle de Y sachant X, notée $E[Y|X]$

<u>Rq</u> $E[Y|X=x] = \displaystyle\int_{\mathbb{R}} y \times f(y|x)\,dy$: fonction de x

$\Rightarrow \quad E[Y|X] = \varphi(X)$ variable aléatoire

Théorème :

Soit Y une variable intégrable. Alors $\mathbb{E}[Y|X]$ est intégrable et

$$\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$$

preuve :

$$\mathbb{E}[\mathbb{E}[Y|X]] = \int_{\mathbb{R}} \mathbb{E}[Y|X=x] \, f_X(x) \, dx$$

$$= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} y \cdot f(y|x) \, dy \right) f_X(x) \, dx$$

$$= \int_{\mathbb{R}} y \left(\int_{\mathbb{R}} f(x,y) \, dx \right) dy$$

$$= \int_{\mathbb{R}} y \, f_Y(y) \, dy$$

$$= \mathbb{E}[Y]$$

définition :

Soit h une fonction de $\mathbb{R}^2 \to \mathbb{R}$.

Sous réserve d'intégrabilité, l'espérance de $h(X,Y)$ sachant $X=x$ est

$$\mathbb{E}[h(X,Y) \mid X=x] = \int_{\mathbb{R}} h(x,y) \, f(y|x) \, dy = \mathbb{E}[h(x,Y) \mid X=x]$$

Donc l'espérance conditionnelle de $h(X,Y)$ sachant X, notée $\mathbb{E}[h(X,Y)|X]$ est la variable aléatoire qui prend la valeur $\mathbb{E}[h(x,Y) \mid X=x]$ avec la densité $f_X(x)$.

propriétés :

Sous réserve d'intégrabilité

- $\mathbb{E}[\mathbb{E}[h(X,Y) \mid X]] = \mathbb{E}[h(X,Y)]$
- si X et Y indépendants, $\mathbb{E}[g(Y)|X] = \mathbb{E}[g(Y)]$

  en particulier, $\mathbb{E}[Y|X] = \mathbb{E}[Y]$ dans ce cas
- $\mathbb{E}[g(X) \mid X] = g(X)$

  en particulier $\mathbb{E}[X|X] = X$
- $\mathbb{E}[\alpha g(X) + \beta h(Y) \mid X] = \alpha g(X) + \beta \, \mathbb{E}[g(Y)|X]$
- $\mathbb{E}[g(X) h(Y) \mid X] = g(X) \, \mathbb{E}[h(Y)|X]$

III) Applications

1) Probabilités conditionnelles

Soit A un évènement qui s'exprime en fonction de X et Y

exemple $A = \{X < Y\}$

$$P(A) = E[\mathbb{1}_A] = \int_{\mathbb{R}^2} \mathbb{1}_{\{x<y\}} f(x,y)\,dx\,dy$$

<u>définition</u> probabilité conditionnelle

$$P(A \mid X = x) = E[\mathbb{1}_A \mid X = x] = \int \mathbb{1}_A(x,y)\, f(y \mid x)\,dy$$

$P(A \mid X)$ probabilité conditionnelle de A sachant Y est la variable aléatoire prenant les valeurs $P(A \mid X = x)$ avec densité $f_X(x)$.

<u>proposition</u>

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

2) La régression

<u>Approximation par une constante</u>

Soit Y variable aléatoire de carré intégrable.

On veut approcher Y par une constante en considérant l'erreur quadratique.

<u>proposition</u>

Parmi les réels a, la quantité $E[(Y-a)^2]$ est minimale en $a = E[Y]$ et alors

$$E[(Y-a)^2] = V[Y]$$

<u>preuve</u>

$$E[(Y-a)^2] = E\left[\left((Y - E[Y]) + (E[Y] - a)\right)^2\right]$$

$$= E[(Y - E[Y])^2] + (E[Y] - a)^2 + 2\underbrace{E[(Y - E[Y])(E[Y] - a)]}_{=0}$$

Approximation par une droite

Soit le couple (X,Y) dont on suppose connaître la loi jointe

On suppose X et Y de carré intégrable.

Supposons que l'on observe X, mais pas Y. On voudrait deviner Y. Impossible car Y sachant $X=x$ reste aléatoire.

On peut cependant chercher à faire la plus petite erreur en moyenne.

Idée la plus simple : approcher Y par une fonction affine de X

On veut minimiser

$$\begin{aligned} E\left[(Y-(aX+b))^2\right] &= a^2 E[X^2] + 2ab\,E[X] + b^2 + E[Y^2] \\ &\quad - 2a\,E[XY] - 2b\,E[Y] \\ &= \phi(a,b) \end{aligned}$$

ϕ atteint son maximum en

$$a = \frac{\operatorname{cov}(X,Y)}{\sigma^2(X)}$$

$$b = E[Y] - \frac{\operatorname{cov}(X,Y)}{\sigma^2(X)} E[X]$$

Introduisons $\rho = \dfrac{\operatorname{cov}(X,Y)}{\sigma(X)\,\sigma(Y)}$

alors $\min\limits_{a,b} E\left[(Y-(aX+b))^2\right] = \sigma^2(Y)(1-\rho^2)$

Rq : cela suppose $\sigma(X) \neq 0$

Rq : erreur d'autant plus faible que ρ^2 proche de 1 ! et $\sigma(Y)$ proche de 0 (Y peu dispersée autour de sa moyenne !)

Rq : En statistique, on ne connaît pas la loi jointe mais que des points

$\rightarrow$ droite de régression

Approximation par une fonction

Maintenant, on ne se restreint pas à une f^ct affine !

définition : Courbe de régression

La courbe $x \longrightarrow \mathbb{E}[Y | X = x]$ est appelée courbe de régression de Y en X

Théorème :

Soit Y de carré intégrable.

Parmi toutes les fonctions $u : \mathbb{R} \to \mathbb{R}$, l'erreur d'approximation $\mathbb{E}[(Y - u(X))^2]$ est minimale lorsque $u(X) = \mathbb{E}[Y | X]$.

preuve :

Posons $m(X) = \mathbb{E}[Y | X]$

$$\mathbb{E}[(Y - u(X))^2] = \mathbb{E}[(Y - m(X))^2] + \mathbb{E}[(m(X) - u(X))^2]$$
$$+ 2\,\mathbb{E}[(Y - m(X))(m(X) - u(X))]$$

Or on sait que

$$\mathbb{E}[(Y - m(X))(m(X) - u(X))] = \mathbb{E}[\mathbb{E}[(Y - m(X))(m(X) - u(X)) | X]]$$
$$= \mathbb{E}[(m(X) - u(X))\underbrace{\mathbb{E}[(Y - m(X)) | X]}_{= 0}]$$

d'où

$$\mathbb{E}[(Y - u(X))^2] = \mathbb{E}[(Y - m(X))^2] + \mathbb{E}[(m(X) - u(X))^2]$$

définition :

La quantité $\sigma^2 = \min_u \mathbb{E}[(Y - u(X))^2] = \mathbb{E}[(Y - \mathbb{E}[Y | X])^2]$

est l'erreur quadratique moyenne, ou variance résiduelle.

Rq : On peut montrer que $\mathbb{E}[Y | X]$ est la projection $\perp$ de Y sur $\mathcal{L}^2(X)$