Chapitre 3 : Espérance conditionnelle

1) Introduction :

La notion de probabilité conditionnelle a été étudiée au 1er semestre. Rappelons que si B est un événement de probabilité $\mathbb{P}(B) > 0$, la probabilité conditionnelle sachant B est la mesure de probabilité $\mathbb{P}_B$ définie pour tout événement A par :

$$\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

(on note aussi : $\mathbb{P}_B(A) = \mathbb{P}(A/B)$).

Si X est une variable ayant un moment d'ordre 1, l'espérance de X sachant B (i.e. sachant que l'événement B s'est réalisé) est le nombre

$$\mathbb{E}(X/B) = \int_\Omega X(\omega)\, d\mathbb{P}_B(\omega),$$

espérance de X pour la probabilité $\mathbb{P}_B$.

Considérons maintenant la question suivante : Soit B un événement quelconque. Dans une expérience aléatoire soit B soit $\overline{B}$ seront réalisés. Qu'en est-il pour l'espérance de X ?

sachant B, c'est $\mathbb{E}(X/B)$

sachant $\overline{B}$, c'est $\mathbb{E}(X/\overline{B})$

On est donc amené à considérer que deux valeurs sont possibles :

$\mathbb{E}(X/B)$ avec probabilité $\mathbb{P}(B)$

$\mathbb{E}(X/\overline{B})$ " " $\mathbb{P}(\overline{B})$

On a donc une variable aléatoire qui peut prendre 2 valeurs $\mathbb{E}(X/B)$ et $\mathbb{E}(X/\overline{B})$ avec les probabilités respectives $\mathbb{P}(B)$ et $\mathbb{P}(\overline{B})$.

Quel est le lien entre cette v.a. et la vraie valeur de $\mathbb{E}(X)$?

Supposons que $X = \mathbb{1}_A$ soit une v.a. indicatrice (d'un événement A). On a

$$\mathbb{E}(X/B) = \mathbb{E}(\mathbb{1}_A/B) = \int_\Omega \mathbb{1}_A\, d\mathbb{P}_B = \mathbb{P}_B(A)$$

et

$$\mathbb{E}(X/\overline{B}) = \mathbb{E}(\mathbb{1}_A/\overline{B}) = \int_\Omega \mathbb{1}_A\, d\mathbb{P}_{\overline{B}} = \mathbb{P}_{\overline{B}}(A)$$

Mais alors

$$\mathbb{E}(X/B)\mathbb{P}(B) + \mathbb{E}(X/\overline{B})\mathbb{P}(\overline{B}) =$$
$$\mathbb{P}_B(A)\mathbb{P}(B) + \mathbb{P}_{\overline{B}}(A)\,\mathbb{P}(\overline{B})$$
$$= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B})$$
$$= \mathbb{P}(A) = \mathbb{E}(X).$$

Ainsi on a :

$$\boxed{\mathbb{E}(X) = \mathbb{E}(X/B)\mathbb{P}(B) + \mathbb{E}(X/\overline{B})\mathbb{P}(\overline{B})}$$

② La formule de l'espérance totale

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé modélisant une expérience aléatoire. Soient $B_1, B_2, \ldots, B_n, \ldots$ un <u>système complet d'événements</u> (fini ou infini dénombrable, dont on suppose que les B_i sont <u>de probabilité non nulle</u>) :

Rappelons que $\forall i \neq j \quad B_i \cap B_j = \emptyset$ et $\Omega = \bigcup_n B_n$.

> <u>Rappel</u> : pour tout événement $A \in \mathcal{F}$, on a
> $$\mathbb{P}(A) = \sum_n \mathbb{P}(A|B_n)\mathbb{P}(B_n)$$
> (formule de la probabilité totale)

On va généraliser cette formule à l'espérance mathématique :

Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire (pas forcément discrète). On suppose que X a un moment d'ordre 1 c'est à dire $\int_\Omega |X(\omega)| d\mathbb{P}(\omega) < +\infty$. Alors :

X a une espérance $E(X) = \int_\Omega X(\omega) d\mathbb{P}(\omega)$.

<u>Alors</u> X a aussi un moment d'ordre 1 relativement aux mesures de probabilité $\mathbb{P}_{B_n}$ ⊛ (i.e. $\int_\Omega |X(\omega)| d\mathbb{P}_{B_n}(\omega) < +\infty$) et donc une espérance $\mathbb{E}(X|B_n) = \int_\Omega X(\omega) d\mathbb{P}_{B_n}(\omega)$ appelé espérance de X sachant B_n.

<u>Remarque</u> le fait ⊛ est admis pour l'instant (ceci fera l'objet d'<u>un exercice</u>).

> <u>Théorème</u> (formule de l'espérance totale)
>
> Si X admet un moment d'ordre 1, alors
> $$\mathbb{E}(X) = \sum_n \mathbb{E}(X|B_n)\mathbb{P}(B_n)$$
> (somme finie ou série convergente suivant que le système complet est fini ou dénombrable)

dém^n :

étape 1 : Si $X = \mathbb{1}_A$ est une v.a. indicatrice,

$\mathbb{E}(X) = \mathbb{P}(A)$, $\mathbb{E}(X/B_n) = \mathbb{P}(A/B_n)$ et la formule du théorème n'est autre que la formule de la probabilité totale.

étape 2 : Si $X = \sum_{i=1}^{N} \alpha_i \mathbb{1}_{A_i}$ est une v.a. simple

$\mathbb{E}(X) = \sum_{i=1}^{N} \alpha_i \mathbb{P}(A_i)$ (linéarité de l'espérance)

$\mathbb{E}(X/B_n) = \sum_{i=1}^{N} \alpha_i \mathbb{P}(A_i/B_n)$ (idem) ⊛

pour tout i $\mathbb{P}(A_i) = \sum_n \mathbb{P}(A_i/B_n)\mathbb{P}(B_n)$

en multipliant par α_i les deux membres de cette égalité puis en sommant sur tous les indices i, on obtient

$$\sum_{i=1}^{N} \alpha_i \mathbb{P}(A_i) = \sum_{i=1}^{N} \sum_n \alpha_i \mathbb{P}(A_i/B_n)\mathbb{P}(B_n)$$
$$= \sum_n \left(\sum_{i=1}^{N} \alpha_i \mathbb{P}(A_i/B_n) \right) \mathbb{P}(B_n)$$

d'où la formule du théorème compte tenu de ⊛.

étape 3 : Si X est une v.a. positive, on sait

qu'il existe une suite croissante $(X_k)_{k\geq 1}$ de v.a. simples et positives telle que $X_k \nearrow X$ si $k \to +\infty$.

Plus précisément :

$\forall \omega \in \Omega$, $X_1(\omega) \leq X_2(\omega) \leq \cdots \leq X_k(\omega) \leq X_{k+1}(\omega) \leq \cdots$

et $\lim_{k \to +\infty} X_k(\omega) = X(\omega)$

(voir le cours de probabilités 1).

Cette propriété d'approximation de toute v.a. positive par une suite croissante de v.a. simples est un résultat théorique très important qui permet souvent de démontrer qu'une propriété vraie pour les v.a. simples se prolonge aux v.a. positives comme on va le voir dans le résultat qui nous occupe :

La formule du théorème est vraie pour chaque v.a. X_k :

$$\mathbb{E}(X_k) \overset{⊛}{=} \sum_n \mathbb{E}(X_k/B_n)\mathbb{P}(B_n) \quad (k \in \mathbb{N}^*)$$

Comme $X_k \nearrow X$, le thm de convergence monotone (cours d'intégration) montre que :

$$\int_\Omega X_k(\omega)\,d\mathbb{P}(\omega) \underset{(k\to+\infty)}{\nearrow} \int_\Omega X(\omega)\,d\mathbb{P}(\omega)$$

c'est à dire

$$\mathbb{E}(X_k) \nearrow \mathbb{E}(X) \text{ si } k\to+\infty$$

De même pour tout n fixé,

$$\mathbb{E}(X_k|B_n) \nearrow \mathbb{E}(X|B_n)\ (k\to+\infty)$$

Mais

$$\lim_{k\to+\infty} \sum_n \mathbb{E}(X_k|B_n)\mathbb{P}(B_n)$$

$$= \sum_n \lim_{k\to+\infty} \mathbb{E}(X_k|B_n)\mathbb{P}(B_n)$$

(encore le thm de convergence monotone pour la sommation qui est une intégration pour la mesure de comptage !).

le passage à la limite dans l'égalité (*) p.6 donne donc

$$\mathbb{E}(X) = \sum_n \mathbb{E}(X|B_n)\mathbb{P}(B_n)$$

étape 4 : Si X est de signe quelconque, on a $X = X^+ - X^-$, où $X^+ = \sup(X,0)$ et

$X^- = \sup(-X,0)$ sont des v.a. positives la formule du thm est vraie pour X^+ et X^- d'après l'étape 3 :

$$\mathbb{E}(X^+) = \sum_n \mathbb{E}(X^+|B_n)\mathbb{P}(B_n)$$

$$\mathbb{E}(X^-) = \sum_n \mathbb{E}(X^-|B_n)\mathbb{P}(B_n)$$

par différence des 2 égalités et compte tenu de la linéarité de l'espérance ($\mathbb{E}(X^+)-\mathbb{E}(X^-) = \mathbb{E}(X^+-X^-) = \mathbb{E}(X)$ et idem pour l'espérance sachant B_n) on obtient le résultat du théorème

CQFD.

Exemple : On lance n pièces identiques telles que $p = \mathbb{P}(\text{pile})$ où $0<p<1$. On conserve uniquement celles qui tombent sur pile (les autres sont éliminées). On relance les piles conservées. Soit X celles qui tombent sur pile. Sans calculer la loi de X, montrer que $\mathbb{E}(X) = np^2$

Solution : Soit X_1 = le nombre de pièces tombées sur pile au 1er lancer. On a

$$\mathbb{E}(X|X_1=k) = kp$$

car lorsqu'on lance k pièces le nombre de celles qui tombent sur pile est de loi $\mathcal{B}(k,p)$

dont l'espérance est kp. Ainsi

$$\mathbb{E}(X) = \sum_{k=0}^{m} \mathbb{E}(X|X_1=k)\,\mathbb{P}(X_1=k)$$

(on applique le thm. avec le système complet $B_k = [X_1=k]$, $k=0,1,\dots,m$). Mais $\mathbb{P}(X_1=k) = C_m^k\, p^k(1-p)^{m-k}$. D'où

$$\mathbb{E}(X) = \sum_{k=0}^{m} kp\, C_m^k\, p^k(1-p)^{m-k}$$

$$= p \sum_{k=0}^{m} k C_m^k\, p^k(1-p)^{m-k}$$

$= p\,mp = mp^2$ (car la somme représente l'espérance d'une loi $\mathcal{B}(m,p)$) CQFD.

Interprétation de la formule de l'espérance totale :

On reprend les hypothèses du théorème de la p.4 : Considérons la v.a. discrète Y qui prend les valeurs $x_n = \mathbb{E}(X|B_n)$ avec la probabilité $p_n = \mathbb{P}(B_n)$:

Y	x_1	x_2	$\dots$	x_n	$\dots$
	p_1	p_2		p_n	$\dots$

l'espérance de Y est justement $\mathbb{E}(X)$ d'après le thm.

Le nombre $\mathbb{E}(X)$ apparaît donc comme l'espérance d'une variable aléatoire discrète Y prenant les valeurs $\mathbb{E}(X|B_n)$ avec la probabilité $p_n = \mathbb{P}(B_n)$. Cette nouvelle variable aléatoire qu'on appelle l'espérance conditionnelle de X sachant la partition B_n va faire l'objet de l'étude de tout ce chapitre dans une version un peu plus générale.

③ Une approche élémentaire de l'espérance conditionnelle

On cherche une interprétation de la v.a. Y précédente. Supposons pour simplifier que le système complet $B_1, B_2, \dots, B_N$ est fini.

On peut définir la v.a. Y sur l'espace de départ $(\Omega, \mathcal{F}, \mathbb{P})$ en posant

$$Y = \sum_{n=1}^{N} \mathbb{E}(X|B_n)\,\mathbb{1}_{B_n}$$

lorsque $\omega \in B_n$, on a $Y(\omega) = \mathbb{E}(X|B_n)$. La v.a. Y est constante sur les ensembles B_n et les valeurs prises par Y sont donc bien les valeurs $x_n = \mathbb{E}(X|B_n)$ avec probabilité $p_n = \mathbb{P}(B_n)$.