CHAPITRE 1

# Notion de conditionnement

**Résumé.** Les notions de conditionnement (discret ou continu) par rapport à un événement, une variable aléatoire, plusieurs variables aléatoires, *etc.* sont généralisées par la notion unique d'*espérance conditionnelle par rapport à une tribu.* Ainsi, pour toute variable intégrable $X$ et toute sous-tribu $\mathcal{A}$ de la tribu définissant l'espace de probabilité, il existe une unique variable $\mathcal{A}$-mesurable $\mathbb{E}[X|\mathcal{A}]$ qui est l'espérance conditionnelle de $X$ sachant la tribu $\mathcal{A}$. D'un point de vue pratique, pour calculer l'espérance conditionnelle d'une variable aléatoire $X$ sachant une autre variable $Y$, on passe par le calcul de la *loi conditionnelle* de $X$ sachant $Y$, qui donne ensuite par intégration toutes les espérances conditionnelles $\mathbb{E}[f(X)|Y]$ pour $f$ mesurable bornée. Cette loi conditionnelle est donnée soit par un rapport de probabilités dans le cas discret, soit par un rapport de densités dans le cas continu.

On note en général $(\Omega, \mathcal{B}, \mathbb{P})$ un espace de probabilité ; ainsi, $\Omega$ est un ensemble, $\mathcal{B}$ est une tribu sur cet ensemble, et $\mathbb{P} : \mathcal{B} \to [0,1]$ est une mesure de probabilité. On utilise des lettres calligraphiques $\mathcal{F}, \mathcal{G}, \ldots$ pour des sous-tribus de $\mathcal{B}$ ; et des lettres capitales $X, Y, \ldots$ pour des variables aléatoires sur $(\Omega, \mathcal{B}, \mathbb{P})$, c'est-à-dire des fonctions mesurables $X : (\Omega, \mathcal{B}) \mapsto (\mathfrak{X}, \mathcal{B}_{\mathfrak{X}})$ à valeurs dans un autre espace mesurable (le plus souvent, $\mathbb{R}$ muni de la tribu des boréliens, ou $\mathbb{Z}$ muni de la tribu de l'ensemble des parties). La loi d'une variable $X$ est l'image de $\mathbb{P}$ par $X$, c'est-à-dire la mesure de probabilité sur $(\mathfrak{X}, \mathcal{B}_{\mathfrak{X}})$ donnée par $\mathbb{P}_X[B] = \mathbb{P}[X^{-1}(B)]$.

## 1. Espérance conditionnelle

De la même façon que la notion la plus générale d'indépendance est l'indépendance de tribus, la bonne notion de conditionnement est la *conditionnement par rapport à une sous-tribu* $\mathcal{A} \subset \mathcal{B}$. Dans ce qui suit, on fixe donc une telle sous-tribu. On peut s'imaginer par exemple que

$$\mathcal{A} = \sigma(X_1, \ldots, X_r)$$

est l'ensemble des événements associés à des observations $X_i$, *i.e.*, la plus petite sous-tribu de $\mathcal{B}$ qui contient toutes les parties $(X_i)^{-1}(B)$ avec $B$ mesurable dans l'espace d'arrivée de $X_i$, $1 \leq i \leq r$. On souhaite alors donner un sens à

“ *l'espérance d'une variable aléatoire $X$ sachant ces observations $X_1, \ldots, X_r$* ”, qui s'écrirait comme fonction mesurable $f(X_1, \dots, X_r)$ de ces observations $X_1, \dots, X_r$. Dans le cas d'une sous-tribu générale $\mathcal{A}$, l'espérance conditionnelle d'une variable $X$ sachant (toutes les réalisations d'événements dans) $\mathcal{A}$ se doit de même d'être une fonction $\mathcal{A}$-mesurable.

Théorème 1.1 (Espérance conditionnelle). *Soit $X \in \mathrm{L}^1(\Omega, \mathcal{B}, \mathbb{P})$, et $\mathcal{A} \subset \mathcal{B}$ une sous-tribu. Il existe une unique variable aléatoire dans $\mathrm{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ (donc, $\mathcal{A}$-mesurable), notée $\mathbb{E}[X|\mathcal{A}]$, telle que pour toute fonction bornée et A-mesurable $Y$,*

$$\mathbb{E}[XY] = \mathbb{E}[\mathbb{E}[X|\mathcal{A}]\, Y]. \qquad (\star)$$

*L'équation $(\star)$ est la propriété caractéristique de l'espérance conditionnelle, le membre de droite de cette identité étant l'espérance d'une fonction $\mathcal{A}$-mesurable.*

L'unicité du théorème s'entend à modification sur un ensemble de mesure nulle près, donc au sens presque sûr. D'autre part, en utilisant la propriété caractéristique, on montre sans difficulté que l'espérance conditionnelle partage de nombreuses propriétés avec l'espérance simple, en particulier : la linéarité, la positivité, les inégalités liées à la convexité, les théorèmes de convergence, *etc.* (notons que l'espérance simple est le cas particulier de l'espérance conditionnelle par rapport à la tribu triviale $\mathcal{A} = \{\emptyset, \Omega\}$). D'autres conséquences importantes et immédiates de la propriété caractéristique sont :

(1) pour toute variable $X$, $\mathbb{E}[X|\mathcal{A}]$ est intégrable d'espérance $\mathbb{E}[\mathbb{E}[X|\mathcal{A}]] = \mathbb{E}[X]$.

(2) si $X$ est $\mathcal{A}$-mesurable, alors $\mathbb{E}[X|\mathcal{A}] = X$.

(3) inversement, si $X$ est indépendante de $\mathcal{A}$, alors $\mathbb{E}[X|\mathcal{A}] = \mathbb{E}[X]$.

(4) conditionnements successifs : si $\mathcal{A}_1 \subset \mathcal{A}_2$ sont deux sous-tribus de $\mathcal{B}$, alors $\mathbb{E}[\mathbb{E}[X|\mathcal{A}_2]|\mathcal{A}_1] = \mathbb{E}[X|\mathcal{A}_1]$ pour toute variable intégrable $X$.

Exemple. Si $\mathcal{A} = \sigma(Z)$ est la tribu engendrée par une variable aléatoire $Z$, alors les fonctions mesurables pour $\mathcal{A}$ sont exactement les fonctions mesurables $f(Z)$ de $Z$, et notant dans ce cas $\mathbb{E}[X|\sigma(Z)] = \mathbb{E}[X|Z]$, on a, pour toute fonction mesurable bornée $f$,

$$\mathbb{E}[X f(Z)] = \mathbb{E}[\mathbb{E}[X|Z]\, f(Z)],$$

le membre de droite de l'identité étant l'espérance d'une fonction de $Z$ (il existe $g$ mesurable telle que $\mathbb{E}[X|Z] = g(Z)$).

Exemple. Supposons que $\mathcal{A} = \sigma(A) = \{\emptyset, A, A^{\mathrm{c}}, \Omega\}$ est la tribu engendrée par un événement de probabilité comprise strictement entre 0 et 1, et comparons l'espérance conditionnelle $\mathbb{E}[X|\mathcal{A}]$ à la notion plus simple

$$\mathbb{E}[X|A] = \frac{\mathbb{E}[X\, \mathbb{1}_A]}{\mathbb{E}[\mathbb{1}_A]} = \frac{\mathbb{E}[X\, \mathbb{1}_A]}{\mathbb{P}[A]}.$$

Les fonctions $\mathcal{A}$-mesurables sont exactement les combinaisons linéaires de $\mathbb{1}_A$ et de $\mathbb{1}_{A^{\mathrm{c}}}$, donc $\mathbb{E}[X|\mathcal{A}]$ doit être une combinaison de ces fonctions. On peut alors vérifier que

$$\mathbb{E}[X|A]\, \mathbb{1}_A + \mathbb{E}[X|A^{\mathrm{c}}]\, \mathbb{1}_{A^{\mathrm{c}}}$$

vérifie la propriéte caractéristique $(\star)$, donc est l'espérance conditionnelle de $X$ sachant la tribu $\mathcal{A}$.

Il existe une formule qui généralise simultanément les propriétés (2) et (3) énoncées précédemment. Soit $X$ une variable $\mathcal{A}$-mesurable, et $Y$ une variable indépendante de $\mathcal{A}$. Lors du calcul de l'espérance conditionnelle par rapport à $\mathcal{A}$ d'une fonction de $X$ et de $Y$, tout ce qui dépend de $X$ est laissé tel quel, et tout ce qui dépend de $Y$ est intégré contre la loi de $Y$. Ainsi, si $f(X,Y)$ est une fonction mesurable bornée avec $X$ variable $\mathcal{A}$-mesurable et $Y$ variable $\mathcal{A}$-indépendante, alors

$$\mathbb{E}[f(X,Y)|\mathcal{A}] = \int_{\mathfrak{Y}} f(X,y)\,\mathbb{P}_Y(dy).$$

D'autre part, la notion d'espérance conditionnelle est définie plus simplement dans $\mathrm{L}^2(\Omega,\mathcal{B},\mathbb{P})$. En effet, pour toute variable $X \in \mathrm{L}^2(\Omega,\mathcal{B},\mathbb{P})$, l'espérance conditionnelle $\mathbb{E}[X|\mathcal{A}]$ est la projection orthogonale de $X$ sur $\mathrm{L}^2(\Omega,\mathcal{A},\mathbb{P})$, et pour toute variable $Y \in \mathrm{L}^2(\Omega,\mathcal{A},\mathbb{P})$,

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

Cette interprétation fournit *ipso facto* une preuve de l'existence de l'espérance conditionnelle pour des variables de carré intégrable, et la densité de $\mathrm{L}^2(\Omega,\mathcal{B},\mathbb{P})$ dans $\mathrm{L}^1(\Omega,\mathcal{B},\mathbb{P})$ mène ensuite au théorème 1.1. On peut également définir l'espérance conditionnelle dans le cadre des variables aléatoires positives (sans hypothèse d'intégrabilité), par exemple par application du théorème de convergence monotone.

## 2. Conditionnement discret

Dans le cas discret, il est fréquent de manipuler simplement des probabilités conditionnelles d'événements $\mathbb{P}[A|B] = \mathbb{E}[\mathbb{1}_A|B] = \frac{\mathbb{P}[A\cap B]}{\mathbb{P}[B]}$. Une propriété importante de ces quantités est la formule des probabilités totales :

$$\mathbb{P}[B] = \sum_{i=1}^{r} \mathbb{P}[B|A_i]\,\mathbb{P}[A_i]$$

si $\Omega = A_1 \sqcup A_2 \sqcup \cdots \sqcup A_r$ est un système complet d'événements ; et son corollaire, la formule de Bayes, qui permet d'inverser le rôle joué par des événements lors d'un conditionnement :

$$\mathbb{P}[A_i|B] = \frac{\mathbb{P}[A_i \cap B]}{\mathbb{P}[B]} = \frac{\mathbb{P}[B|A_i]\,\mathbb{P}[A_i]}{\sum_{j=1}^{r} \mathbb{P}[B|A_j]\,\mathbb{P}[A_j]}.$$

Voyons maintenant comment calculer en pratique des espérances conditionnelles dans le cas discret. Plus précisément, considérons deux variables $X$ et $Y$ à valeurs dans $\mathbb{Z}$, et voyons comment calculer $\mathbb{E}[X|Y]$. Pour les mêmes raisons que lors du conditionnement par rapport à un événement, on a :

$$\mathbb{E}[X|Y] = \sum_{\substack{k\in\mathbb{Z} \\ \mathbb{P}[Y=k]\neq 0}} \mathbb{E}[X|Y=k]\,\mathbb{1}_{Y=k},$$

qui est bien une fonction de $Y$, à savoir la fonction

$$k \mapsto \begin{cases} \frac{\mathbb{E}[X\,\mathbb{1}_{Y=k}]}{\mathbb{P}[Y=k]} & \text{si } \mathbb{P}[Y=k]>0; \\ 0 & \text{si } \mathbb{P}[Y=k]=0. \end{cases}$$

Exemple. Soient $X$ et $Y$ deux variables indépendantes suivant une loi Poisson de paramètre 1, et $Z = X + Y$, qui suit une loi de Poisson de paramètre 2. L'espérance conditionnelle de $X$ sachant $Z = k$ est :

$$\mathbb{E}[X|Z=k] = \frac{\mathbb{E}[X\,\mathbb{1}_{Z=k}]}{\mathbb{P}[Z=k]} = \mathrm{e}^2\,2^{-k}\,k!\,\sum_{j=0}^{k} j\,\mathbb{P}[X=j]\,\mathbb{P}[Y=k-j]$$
$$= 2^{-k}\,\sum_{j=0}^{k} j\,\frac{k!}{j!\,(k-j)!} = k\,2^{-k}\sum_{j=0}^{k-1}\frac{(k-1)!}{j!\,(k-j-1)!} = \frac{k}{2},$$

donc $\mathbb{E}[X|Z] = \frac{Z}{2}$.

Pour donner une meilleure explication de l'exemple précédent, il est utile d'introduire la notion de *loi conditionnelle*. Soient $X$ et $Y$ deux variables aléatoires sur un espace $(\Omega, \mathcal{B}, \mathbb{P})$, à valeurs dans des espaces $(\mathfrak{X}, \mathcal{B}_{\mathfrak{X}})$ et $(\mathfrak{Y}, \mathcal{B}_{\mathfrak{Y}})$, qu'on suppose polonais (métrisables complets et séparables). L'application

$$f \mapsto \mathbb{E}[f(X)|Y]$$

est une fonctionnelle linéaire positive de $f$ ; on s'attend donc à pouvoir la représenter par une mesure de probabilité, qui serait aléatoire et $Y$-mesurable.

Théorème 1.2 (Loi conditionnelle). *Il existe une unique application mesurable*

$$\mathfrak{Y} \to \mathscr{P}(\mathfrak{X}, \mathcal{B}_{\mathfrak{X}})$$
$$y \mapsto \nu(\cdot, y)$$

*appelée loi conditionnelle de $X$ sachant $Y$, telle que pour toute fonction $f : \mathfrak{X} \to \mathbb{R}$ mesurable bornée,*

$$\mathbb{E}[f(X)|Y] = \int_{\mathfrak{X}} f(x)\,\nu(dx, Y).$$

*On note souvent $\nu(dx, Y) = \mathbb{P}_{X|Y}(dx)$. Dans le cas discret, la loi conditionnelle est donnée par la fonction $\nu(x, y) = \mathbb{E}[\mathbb{1}_{X=x}|\mathbb{1}_{Y=y}]$.*

Exemple. Revenons à l'exemple précédent, et calculons la loi conditionnelle de $X$ sachant $Z$. Pour tous $j \leq k$,

$$\mathbb{E}[\mathbb{1}_{X=j}|\mathbb{1}_{Z=k}] = \frac{\mathbb{P}[X=j]\,\mathbb{P}[Y=k-j]}{\mathbb{P}[Z=k]} = 2^{-k}\,\frac{k!}{j!\,(k-j)!}.$$

Ainsi, sachant $Z = k$, $X$ suit une loi binomiale $\mathcal{B}(k, \frac{1}{2})$. Autrement dit,

$$\mathbb{P}_{X|Z} = \mathcal{B}\left(Z, \frac{1}{2}\right).$$

On retrouve bien en particulier $\mathbb{E}[X|Z] = \sum_x x\,\mathcal{B}(Z, \frac{1}{2})(x) = \frac{Z}{2}$.

## 3. Conditionnement continu

Dans le cas continu, on peut également calculer l'espérance conditionnelle, et plus généralement la loi conditionnelle d'une variable réelle $X$ sachant une autre variable réele $Y$, si l'on connaît la loi jointe du couple $(X,Y)$, donnée par une densité de probabilité $f(x,y)\,dx\,dy$ (les raisonnements qui suivent s'adaptent sans problème au cas de variables vectorielles). Soit $g(y) = \int_{\mathbb{R}} f(x,y)\,dx$ ; c'est la densité de la variable $Y$. Au vu du cas discret et par analogie, il est naturel de poser

$$\mathbb{E}[X|Y] = \begin{cases} \frac{\int_{\mathbb{R}} x\, f(x,Y)\,dx}{g(Y)} & \text{si } g(Y) > 0; \\ 0 & \text{si } g(Y) = 0. \end{cases}$$

Il s'agit bien d'une fonction mesurable de $Y$, et elle vérifie bien la propriété caractéristique $(\star)$, donc c'est effectivement $\mathbb{E}[X|Y]$. Plus généralement, la loi conditionnelle de $X$ sachant $Y$ est donnée par

$$\mathbb{P}_{X|Y}(dx) = \begin{cases} \frac{f(x,Y)\,dx}{g(Y)} & \text{si } g(Y) > 0; \\ \delta_0(dx) & \text{si } g(Y) = 0, \end{cases}$$

le deuxième cas étant simplement une convention pratique (la loi conditionnelle est définie à un ensemble de mesure nulle près).

Exemple. Considérons une gaussienne standard $(X,Y)$ de dimension 2, et calculons la loi de $X$ sachant $R = \sqrt{X^2+Y^2}$. La densité de $R$ est donnée par

$$\mathbb{E}[f(R)] = \frac{1}{2\pi} \int_{\mathbb{R}^2} f\left(\sqrt{x^2+y^2}\right) \mathrm{e}^{-\frac{x^2+y^2}{2}}\,dx\,dy = \int_{r=0}^{\infty} f(r)\,\mathrm{e}^{-\frac{r^2}{2}}\,r\,dr,$$

elle est donc donnée par $g(r) = r\,\mathrm{e}^{-\frac{r^2}{2}}$ pour $r > 0$, et 0 sinon. D'autre part, le jacobien de $(x,y) \mapsto (x, \sqrt{x^2+y^2})$ est $\frac{\sqrt{r^2-x^2}}{r}$, et chaque couple $(x,r)$ avec $r \neq 0$ est atteint 2 fois, donc, la densité du couple $(X,R)$ est

$$f(x,r) = \frac{\mathrm{e}^{-\frac{r^2}{2}}}{\pi} \frac{r}{\sqrt{r^2-x^2}} \mathbb{1}_{0 \le |x| \le r}\,dx\,dr.$$

La loi conditionnelle de $X$ sachant $R$ est donc

$$\mathbb{P}_{X|R}(dx) = \begin{cases} \frac{\mathbb{1}_{|X| \le R}}{\pi\sqrt{R^2-x^2}}\,dx & \text{si } R > 0; \\ \delta_0(dx) & \text{si } R \le 0, \end{cases}$$

c'est-à-dire une version de la loi de l'arcsinus dépendant continuement de $R$. On en déduit par exemple la valeur de

$$\mathbb{E}\big[|X|\,\big|R\big] = \int_{-R}^{R} \frac{|x|}{\pi\sqrt{R^2-x^2}}\,dx = \frac{2R}{\pi} \int_0^{\pi/2} \sin\theta\,d\theta = \frac{2R}{\pi}.$$

Exemple. Soit $(N_t)_{t\in\mathbb{N}}$ un processus de Poisson d'intensité 1. Ceci veut dire qu'il existe des variables aléatoires $E_1, \ldots, E_k, \ldots$ indépendantes de loi exponentielle $\mathbb{1}_{x>0}\, \mathrm{e}^{-x}\, dx$ telles que si

$$\begin{aligned} S_1 &= E_1; \\ S_2 &= E_1 + E_2; \\ \vdots \quad & \quad \vdots \\ S_k &= E_1 + \cdots + E_k \end{aligned}$$

alors les variables $S_k$ déterminent les temps des sauts de $N_t$ : $N_t = k$ sur l'intervalle $[S_k, S_{k+1})$. Les lois marginales de $N_t$ sont alors des variables de Poisson : $\mathbb{P}[N_t = k] = \frac{t^k\, \mathrm{e}^{-t}}{k!}$.

Fixons un entier $k$, et déterminons la loi de $(S_1, \ldots, S_{k-1})$ sachant $S_k$. La loi jointe de $(S_1, \ldots, S_k)$ est donnée par la densité :

$$\mathbb{1}_{e_1>0,\ldots,e_k>0}\, \mathrm{e}^{-e_1+\cdots+e_k}\, de_1 \,\cdots\, de_k = \mathbb{1}_{0\leq s_1\leq\cdots\leq s_k}\, \mathrm{e}^{-s_k}\, ds_1 \,\cdots\, ds_k.$$

La loi de $S_k$ est donc donnée par intégration par rapport aux $k-1$ premières variables :

$$\frac{(s_k)^{k-1}}{(k-1)!}\, \mathrm{e}^{-s_k}\, ds_k$$

ce qui constitue une loi $\Gamma(k,1)$. Par conséquent, la loi conditionnelle de $(S_1, \ldots, S_{k-1})$ sachant $S_k$ est

$$\frac{\mathbb{1}_{0\leq s_1\leq\cdots\leq S_k}\, \mathrm{e}^{-S_k}}{\frac{(S_k)^{k-1}}{(k-1)!}\, \mathrm{e}^{-S_k}}\, ds_1 \,\ldots\, ds_{k-1} = (k-1)!\, \mathbb{1}_{0\leq s_1\leq\cdots\leq s_{k-1}\leq S_k}\, \frac{ds_1 \,\ldots\, ds_{k-1}}{(S_k)^{k-1}},$$

c'est-à-dire que c'est la loi uniforme sur le simplexe $\{0 \leq s_1 \leq \cdots \leq s_{k-1} \leq S_k\}$. C'est l'une des manifestations du phénomène d'absence de mémoire des processus de Poisson.

**Références.** On trouvera un traitement plus complet de la notion de probabilité et d'espérance conditionnelle dans

(1) P. Billingsley, *Probability and Measure*, 3rd edition, Wiley, 1995 ; §33-34.

(2) R. Durrett, *Probability : Theory and Examples*, 4th edition, Cambridge University Press, 2010 ; §5.1.

On s'est ici inspiré du traitement proposé par ce second ouvrage.