3 Couples de variables aléatoires, conditionnement

L'étude de la loi de probabilité d'une variable aléatoire Y connaissant la valeur prise par une autre variable aléatoire X est fondamentale pour les problèmes d'approximation et de prévision. Il faut pour cela connaître en premier lieu la distribution de probabilité du couple (X, Y) qui est une application de $(\Omega, \mathscr{C}, P)$ dans $\mathbb{R}^2$ muni de sa tribu borélienne si il s'agit d'un couple de variables aléatoires réelles.

Il n'est cependant pas nécessaire que X et Y soient à valeurs dans $\mathbb{R}$.

3.1 ÉTUDE D'UN COUPLE DE VARIABLES DISCRÈTES

On étudiera ici la distribution d'un couple de variables aléatoires à valeurs dans des ensembles finis ou dénombrables ; par exemple la distribution simultanée de la somme et du produit des points amenés par deux dés.

3.1.1 Lois associées à un couple (X, Y)

Supposons que X et Y prennent des valeurs x_i et y_j en nombre fini ou dénombrable.

3.1.1.1 *Loi jointe*

La loi du couple (X, Y) P_{XY} est alors entièrement définie par l'ensemble des nombres :

$$P_{XY}(x_i\,;\,y_j) = P(X = x_i \cap Y = y_j)$$

dans le cas fini cette loi de probabilité conjointe peut se mettre sous la forme d'une table.

	y_1	y_j	y_q	
x_1				
x_i		p_{ij}		$p_{i.}$
x_p				
		$p_{.j}$		

On note $p_{ij} = P(X = x_i \cap Y = y_j)$ et bien sûr $\sum_i \sum_j p_{ij} = 1$.

3.1.1.2 *Lois marginales*

On appelle lois marginales les lois de probabilité de X et de Y pris séparément. On a d'après le théorème des probabilités totales :

– ***Loi marginale de X*** $\quad P(X = x_i) = \sum_{j=1}^{q} p_{ij} = p_{i.}$

– ***Loi marginale de Y*** $\quad P(Y = y_j) = \sum_{i=1}^{p} p_{ij} = p_{.j}$

3.1.1.3 Lois conditionnelles

Les évènements $\{X = x_i\}$ et $\{Y = y_j\}$ étant de probabilités non nulles on définit alors deux familles de lois conditionnelles selon que l'on connaît la « valeur » de X ou de Y. Rappelons qu'ici X et Y ne sont pas forcément des variables aléatoires réelles mais peuvent être des variables qualitatives. D'après le chapitre 1 on a :

– ***Lois conditionnelles de X si $Y = y_j$:***

$$P(X = x_i / Y = y_j) = \frac{p_{ij}}{p_{.j}} = \frac{P(X = x_j \cap Y = y_j)}{P(Y = y_j)}$$

– ***Lois conditionnelles de Y si $X = x_i$:***

$$P(Y = y_j / X = x_i) = \frac{p_{ij}}{p_{i.}} = \frac{P(X = x_i \cap Y = y_j)}{P(X = x_i)}$$

Le théorème des probabilités totales (deuxième forme) permet d'écrire :

$$\begin{aligned} P(X = x_i \cap Y = y_j) &= \sum_{j=1}^{q} P(X = x_i / Y = y_j) P(Y = y_j) \\ &= \sum_{i=1}^{p} P(Y = y_j / X = x_i) P(X = x_i) \end{aligned}$$

Remarques :

- Pour deux événements B_1 et B_2 relatifs à Y et X on a :

$$\begin{aligned} P((Y \in B_2) \cap (X \in B_1)) &= \sum_{x \in B_1} P(Y \in B_2 / X = x) P(X = x) \\ &= \int_{B_1} P(Y \in B_2 / X = x) \, \mathrm{d}P_X(x) \end{aligned}$$

formule qui servira pour étendre la notion de probabilité conditionnelle lorsque $X = x$ est de mesure nulle.

- Il arrive fréquemment dans les applications que l'on utilise la démarche inverse : on connaît la loi conditionnelle de Y à X fixé et celle de X et on en déduit alors la loi du couple.

Les formules de Bayes permettent d'exprimer une loi conditionnelle en fonction de l'autre :

$$P(X = x_i / Y = y_j) = \frac{P(Y = y_j / X = x_i) P(X = x_i)}{\sum_{i=1}^{p} P(Y = y_j / X = x_i) P(X = x_i)}$$

et :

$$P(Y = y_j / X = x_i) = \frac{P(X = x_i / Y = y_j) P(Y = y_j)}{\sum_{j=1}^{q} P(X = x_i / Y = y_j) P(Y = y_j)}$$

L'indépendance entre X et Y s'écrit :

$$p_{ij} = p_{i.}p_{.j} \quad \forall i \text{ et } j$$

ce qui revient à dire que les q lois conditionnelles de X à Y fixé (en faisant varier Y) sont identiques ; il en est de même pour les p lois conditionnelles de Y à X fixé.

3.1.2 Covariance et corrélation linéaire

La covariance a été introduite au chapitre 2 pour des variables numériques.

$$\text{cov}(X\,;Y) = E((X - E(X))(Y - E(Y)) = E(XY) - E(X)E(Y)$$

On a : $$\text{cov}(X\,;X) = V(X) \quad \text{et} \quad \text{cov}(Y\,;Y) = V(Y)$$

On montrera plus loin que :

$$(\text{cov}(X\,;Y))^2 \leq V(X)V(Y)$$

ce qui permet de définir le coefficient de corrélation linéaire ρ, qui est donc toujours compris entre -1 et $+1$:

$$\rho = \frac{\text{cov}(X\,;Y)}{\sigma_X\sigma_Y}$$

Pour deux variables indépendantes $\rho = 0$. Cependant, la réciproque est en général inexacte et un coefficient de corrélation linéaire nul n'entraîne pas que les variables sont indépendantes. Deux exceptions notables où non-corrélation et indépendance sont équivalents : les couples $(X\,;Y)$ gaussiens (voir chapitre 4), et les couples de variables de Bernoulli (facile à montrer).

Les valeurs limites -1 et $+1$ sont atteintes si et seulement si il existe une relation linéaire entre Y et X.

3.1.3 Moments conditionnels

Supposons Y réelle mais pas nécessairement X qui peut être une variable qualitative. On peut alors définir, sous réserve de l'existence de ces expressions pour le cas dénombrable, l'espérance et la variance de Y à X fixé.

3.1.3.1 L'espérance conditionnelle

DÉFINITION

On appelle espérance de Y sachant que $X = x$ et on note $E(Y/X = x)$ la quantité définie par :

$$E(Y/X = x) = \sum_y yP(Y = y/X = x)$$

C'est donc l'espérance de Y prise par rapport à sa loi conditionnelle.

On note que $E(Y/X = x)$ est une fonction de x : $E(Y/X = x) = \varphi(x)$.

Cette fonction φ s'appelle ***fonction de régression***[1] ***de Y en X***. Son graphe est le lieu des moyennes conditionnelles de Y sachant X.

On voit donc que $E(Y/X = x)$ dépend des valeurs prises par X. On peut alors définir la variable aléatoire espérance conditionnelle, qui prend pour valeurs $E(Y/X = x)$ avec les probabilités $P(X = x)$:

DÉFINITION

On appelle variable aléatoire espérance conditionnelle de Y sachant X et on note $E(Y/X)$ la variable définie par :

$$E(Y/X) = \varphi(X)$$

Cette variable présente un certain nombre de propriétés remarquables.

Tout d'abord la linéarité comme conséquence de sa définition en tant qu'espérance :

$$E(Y_1 + Y_2/X) = E(Y_1/X) + E(Y_2/X)$$

mais surtout on a en prenant l'espérance de cette variable le :

THÉORÈME DE L'ESPÉRANCE TOTALE

$$\boxed{E[E(Y/X)] = E(Y)}$$

Démonstration

$$E[E(Y/X)] = \sum_x E(Y/X = x)P(X = x) = \sum_x \left(\sum_y yP(Y = y/X = x) \right) P(X = x)$$

$$= \sum_y y \sum_x P(Y = y/X = x)P(X = x) = \sum_y yP(Y = y) = E(Y)$$

Ce théorème est un outil très puissant pour calculer l'espérance mathématique d'une loi compliquée mais dont les lois conditionnelles sont simples : on voit même que l'on n'a pas besoin de connaître explicitement la loi de Y (voir plus loin).

Si $\psi(X)$ est une autre variable fonction de X on a $E[Y\psi(X)/X] = \psi(X)E[Y/X]$; la démonstration sans difficulté est omise. Concrètement cette formule signifie qu'à X fixé $\psi(X)$ est une constante et sort donc de l'espérance.

1 Ce terme de régression provient des travaux du statisticien Galton qui étudiait la taille des enfants Y en fonction de la taille de leur père X. Il avait constaté expérimentalement que la taille moyenne des fils dont le père avait une taille x supérieure à la moyenne $E(X)$ était elle-même supérieure à $E(Y)$ mais dans une moindre mesure $\dfrac{E(Y/X = x) - E(Y)}{x - E(X)}$ était inférieur à 1 ; il y avait donc régression au sens ordinaire du mot.

3.1.3.2 La variance conditionnelle

DÉFINITION

On appelle variance de Y sachant que $X = x$ et on note $V(Y/X = x)$ la quantité :

$$V(Y/X = x) = E[(Y - E(Y/X = x))^2/X = x]$$

Il s'agit donc de l'espérance conditionnelle du carré de l'écart à l'espérance conditionnelle.

Comme pour l'espérance, et puisque $V(Y/X = x) = \psi(x)$, on définit ensuite la variable aléatoire variance conditionnelle :

$$V(Y/X) = \psi(X) = E[(Y - E(Y/X))^2/X]$$

On a alors le résultat fondamental suivant :

THÉORÈME DE LA VARIANCE TOTALE

$$\boxed{V(Y) = E[V(Y/X)] + V[E(Y/X)]}$$

en donnant à l'espérance sa signification usuelle de moyenne on voit que la variance de Y est la somme de deux termes : la moyenne des diverses variances conditionnelles et la variance des diverses moyennes conditionnelles.

Démonstration

$$V(Y) = E[(Y - E(Y))^2] = E[(Y - E(Y/X) + E(Y/X) - E(Y))^2]$$

développons le carré en groupant $Y - E(Y/X)$ et $E(Y/X) - E(Y)$ il vient :

$$V(Y) = E[(Y - E(Y/X)^2] + 2E[(Y - E(Y/X))(E(Y/X) - E(Y))] + E[(E(Y/X) - E(Y))^2]$$

Le dernier terme est égal à $V[E(Y/X)]$ par définition de la variance puisque $E(Y)$ est l'espérance de $E(Y/X)$.

Le premier terme n'est autre que $E[V(Y/X)]$: en effet en appliquant le théorème de l'espérance totale :

$$E[(Y - E(Y/X))^2] = E[E[(Y - E(Y/X))^2/X]]$$

et on reconnaît l'expression de $V(Y/X)$. Notons que $V(Y/X)$ n'est pas égale à $(Y - E(Y/X))^2$ ce sont simplement deux variables ayant même espérance.

On vérifie que le double produit est nul en conditionnant à nouveau : l'espérance conditionnelle à X fixé de $(Y - E(Y/X))(E(Y/X) - E(Y))$ vaut alors :

$$[E(Y/X) - E(Y)][E(Y - E(Y/X))/X]$$

puisque $E(Y/X) - E(Y)$ est une constante à X fixé (voir la dernière propriété de l'espérance conditionnelle énoncée au sous-paragraphe précédent). Quant à :

$$E[(Y - E(Y/X))/X]$$

ce terme est nul, il suffit de développer. L'espérance conditionnelle du double produit est nul, il en est de même de son espérance.

(on trouvera plus loin une démonstration géométrique plus rapide et plus élégante) ■

3.1.3.3 Exemple d'utilisation de l'espérance et de la variance conditionnelle

Un examen se déroule sous forme d'un questionnaire à choix multiple (QCM) où on pose 20 questions ; chaque question comporte quatre réponses possibles, dont une et une seule est la bonne ; une réponse juste compte 1 point, sinon zéro.

On suppose que le programme de l'examen comporte 100 questions dont on tirera aléatoirement les 20 de l'examen.

Si l'on considère un candidat ayant appris une proportion p du programme, on étudie la distribution de sa note N.

Solution : Parmi les 20 questions, un certain nombre X va figurer dans la partie des 100 p questions révisées et fournir automatiquement X points. Les 20 questions étant tirées sans remise parmi les 100, la loi de X est une hypergéométrique $\mathcal{H}(100\ ;\ 20\ ;\ p)$.

Un certain nombre de réponses pourront être devinées par le jeu du hasard parmi les $20 - X$ questions non révisées, soit Y ce nombre. A chaque question non révisée est associée une variable de Bernoulli de paramètre $1/4$. Si $X = x$ est fixé, la loi de Y est alors une loi binomiale $\mathcal{B}(20 - x\ ;\ 1/4)$.

On a donc $N = X + Y$ avec $Y/X\ \mathcal{B}(20 - X\ ;\ 1/4)$. X et Y ne sont pas indépendantes puisque la distribution conditionnelle de $Y/X = x$ dépend de x.

Le calcul de la distribution de N conduit en tout état de cause à une expression difficilement manipulable :

$$P(N = n) = \sum_{x=0}^{x=n} P(X = x)P(Y = n - x/X = x)$$

$$= \sum_{x=0}^{x=n} \frac{C_{100p}^{x} C_{100(1-p)}^{20-x}}{C_{100}^{20}} C_{20-x}^{n-x} \left(\frac{1}{4}\right)^{n-x} \left(\frac{3}{4}\right)^{20-n}$$

On peut cependant trouver aisément $E(N)$ et $V(N)$:

- Calcul de $E(N)$:

$$E(N) = E(X) + E(Y) = E(X) + E[E(Y/X)]$$

$$E(X) = 20p \text{ (loi hypergéométrique)}$$

$$E(Y/X) = (20 - X)\frac{1}{4} = 5 - \frac{X}{4}$$

$$E[E(Y/X)] = 5 - \frac{E(X)}{4} = 5 - 5p$$

soit :

$$E(N) = 15p + 5$$

- Calcul de $V(N)$:

$$V(N) = E[V(N/X)] + V[E(N/X)]$$

$$V(N/X = x) = V[x + Y/X = x] = V[Y/X = x] = (20 - x)\frac{1}{4}\cdot\frac{3}{4}$$

$$V(N/X) = (20 - X)\frac{3}{16}\quad E[V(N/X)] = 20(1 - p)\frac{3}{16} = \frac{15(1 - p)}{4}$$

$$E[N/X = x] = x + \frac{1}{4}(20 - x) = 5 + \frac{3x}{4}$$

$$E[N/X] = 5 + \frac{3X}{4}\quad V[E(N/X)] = \frac{9}{16}V(X)$$

$$= \frac{9}{16}20p(1 - p)\frac{100 - 20}{100 - 1}$$

$$V[E(N/X)] = \frac{100p(1 - p)}{11}$$

$$V(N) = \frac{15(1 - p)}{4} + \frac{100p(1 - p)}{11} = (1 - p)\left[\frac{15}{4} + \frac{100p}{11}\right]$$

La figure 3.1 donne les variations de $E(N)$ et de $V(N)$ en fonction de p.

Un taux de révision de 0.6 à 0.7 devrait donc assurer la réussite à l'examen avec une forte probabilité.

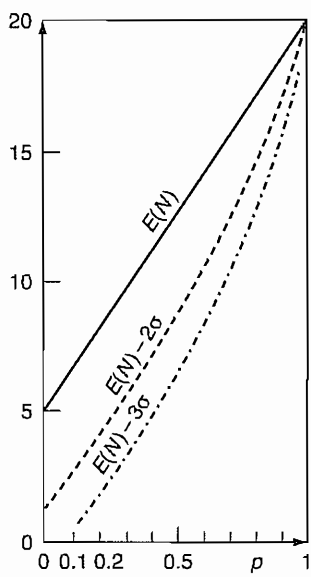

Figure 3.1

3.1.4 Extension au conditionnement d'une variable continue Y par une variable discrète X

Ce cas ne présente pas de difficulté. On définira d'abord la fonction de répartition conditionnelle :

$$P(Y < y/X = x) = \frac{P(Y < y \cap X = x)}{P(X = x)} = G(y/x)$$

puis si elle existe la densité conditionnelle $g(y/x)$ qui sera la dérivée de G en y.

La densité marginale de Y s'obtient par :

$$g(y) = \sum_x g(y/x)P(X = x)$$

Si $E(Y)$ existe, on prouve aisément que $E(Y/X = x)$ existe également et vaut :

$$E(Y/X = x) = \int_{\mathbb{R}} yg(y/x)\mathrm{d}y$$

Les formules de l'espérance totale et de la variance totale sont également valables. La formule de Bayes donne :

$$P(X = x/Y < y) = \frac{G(y/x)P(X = x)}{G(y)}$$

mais l'écriture formelle :

$$P(X = x/Y = y) = \frac{g(y/x)P(X = x)}{g(y)}$$

ne peut être pour l'instant justifiée car $P(Y = y) = 0$.

3.1.5 Somme d'un nombre aléatoire de variables *iid*

Le problème suivant est courant en assurance : au cours d'une période de temps donnée le nombre de sinistres survenus est une variable aléatoire N. Chaque sinistre a un coût aléatoire représenté par une variable X.

Le montant total des sinistres est alors :

$$S = X_1 + X_2 + \cdots X_N$$

Si les X_i sont indépendantes et de même loi, les théorèmes de l'espérance et de la variance totale, en conditionnant par N, permettent de montrer facilement que :

$$E(S) = E(N)E(X)$$
$$V(S) = E(N)V(X) + V(N)(E(X))^2$$

3.2 EXTENSION À DES VARIABLES QUELCONQUES

3.2.1 Lois conjointes et lois marginales d'un couple de variables aléatoires réelles

Si (X, Y) est à valeurs dans $\mathbb{R}^2$ rappelons que la fonction de répartition du couple $H(x, y)$ se définit par :

$$H(x, y) = P(X < x \cap Y < y)$$

Les fonctions de répartition marginales s'en déduisent immédiatement par :

$$F(x) = H(x\,;\infty) = P(X < x)$$
$$G(y) = H(\infty\,;y) = P(Y < y)$$

Si le couple (X, Y) admet une densité $h(x, y)$ on a :

$$h(x, y) = \frac{\partial^2 H}{\partial x \partial y}$$

les densités marginales s'obtiennent par :

$$f(x) = \int_{\mathbb{R}} h(x, y)\, \mathrm{d}y$$

$$g(x) = \int_{\mathbb{R}} h(x, y)\, \mathrm{d}x$$

Rappelons que si et seulement si les variables X et Y sont indépendantes on a :

$$H(x, y) = F(x)G(y) \quad \forall x\,;y$$
$$h(x, y) = f(x)g(y) \quad \forall x\,;y$$

3.2.2 Conditionnement

Le problème essentiel est de donner un sens aux expressions du type $P(Y \in B/X = x)$ et $E(Y/X = x)$ lorsque $X = x$ est un évènement de probabilité nulle ce qui est toujours le cas lorsque X est une variable admettant une densité.

3.2.2.1 *Présentation naïve*

Lorsque X est une variable continue on peut songer à définir la fonction de répartition conditionnelle de Y sachant que $X = x$ comme la limite pour ε tendant vers 0 de :

$$\frac{P(Y < y \cap (x < X < x + \varepsilon))}{P(x < X < x + \varepsilon)} = \frac{H(x + \varepsilon\,;y) - H(x\,;y)}{F(x + \varepsilon) - F(x)}$$

Lorsque X possède une densité $f(x)$ on « voit » que la limite de cette expression est $\dfrac{\partial H(x\,;y)}{\partial x}\Big/ f(x)$ et que si (X, Y) a une densité $h(x, y)$ la densité conditionnelle de Y à $X = x$ fixé vaut alors :

$$\frac{h(x\,;y)}{f(x)} = g(y/x)$$

On conçoit cependant aisément qu'une telle approche est peu rigoureuse et ne recouvre en plus qu'une partie du problème : dans certaines applications il faut pouvoir conditionner par rapport à une variable quelconque pas nécessairement à valeur dans $\mathbb{R}$ ni dans un ensemble fini. Pour définir une espérance conditionnelle il faut seulement que Y soit réelle et que $E(Y)$ existe.

3.2.2.2 Aperçus théoriques

Vu sa complexité nous ne donnerons que les résultats les plus importants sans rentrer dans les détails des démonstrations qui figurent dans les ouvrages de « Théorie des probabilités » (Neveu (1964) ou Métivier (1972) par exemple).

- Première présentation

X étant une variable aléatoire quelconque de $(\Omega, \mathscr{C}, P)$ dans un ensemble mesurable $(E, \mathscr{E})$ on définira la probabilité conditionnelle d'un événement A par rapport à X grâce au théorème suivant :

THÉORÈME

Soit $A \in \mathscr{C}$, alors $\forall B \in \mathscr{E}$ il existe une classe d'équivalence unique de fonctions de $(E, \mathscr{E})$ dans $[0\,;1]$ notée $P(A/X = x)$ telle que :

$$P(A \cap \{X \in B\}) = \int_B P(A/X = x)\, \mathrm{d}P_X(x)$$

La fonction $P(A/X = x)$ n'est pas unique car une modification de celle-ci sur un ensemble de probabilité P_X nulle ne change pas le résultat de l'intégrale.

Peut-on choisir un représentant de cette classe pour tout A qui définisse une loi de probabilité conditionnelle sur Ω ? Ce n'est pas sûr si X est quelconque et $P(./X = x)$ n'est pas nécessairement une mesure de probabilité : ici se trouve la difficulté majeure de la théorie. Si un tel choix est possible on dit que c'est une « version régulière » de la probabilité conditionnelle par rapport à X, notée $P(./X = x)$.

On peut alors définir l'espérance conditionnelle d'une variable Y intégrable par :

$$E(Y/X = x) = \int_\Omega Y(\omega)\, \mathrm{d}P(\omega/X = x)$$

- Deuxième présentation

Les ouvrages récents de théorie des probabilités préfèrent partir de la définition de l'espérance conditionnelle grâce au théorème suivant qui étend la formule de l'espérance totale en intégrant sur un événement quelconque de E au lieu d'intégrer sur E tout entier.

THÉORÈME

Soit Y une variable aléatoire réelle de $(\Omega, \mathscr{C}, P)$ dans $(\mathbb{R}, \mathscr{B})$ telle que $E(Y)$ soit fini, et X une variable quelconque de $(\Omega, \mathscr{C}, P)$ dans $(E, \mathscr{E})$ de loi de probabilité P_X.

Il existe alors une classe d'équivalence unique de fonctions P_X intégrables de $(E, \mathscr{E})$ dans $(\mathbb{R}, \mathscr{B})$ notée $E(Y/X = x)$ telle que :

$$\forall B \in \mathscr{E} \quad \int_{X^{-1}(B)} Y(\omega)\, \mathrm{d}P(\omega) = \int_B E(Y/X = x)\, \mathrm{d}P_X(x)$$

Ceci définit alors de manière (presque sûrement) unique la variable aléatoire espérance conditionnelle $E(Y/X)$.

On en déduit alors la probabilité d'un événement A quelconque de Ω conditionnellement à X en prenant pour Y la variable indicatrice de A :

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

Comme $\mathbb{1}_A$ est intégrable la probabilité conditionnelle de A existe toujours. Le problème de l'existence d'une version régulière de la probabilité conditionnelle reste cependant entier, cette existence est nécessaire pour pouvoir calculer l'espérance conditionnelle par la formule :

$$E(Y/X = x) = \int_\Omega Y(\omega)\, \mathrm{d}P(\omega/X = x)$$

et pour pouvoir parler de distribution conditionnelle de Y sachant X.

La distribution conditionnelle de Y sachant $X = x$ est en effet définie comme la mesure image de $P(./X = x)$ par Y pour chaque x. Il faut donc que $P(./X = x)$ soit une mesure de probabilité sur Ω.

La preuve directe de l'existence de distributions conditionnelles dans les cas les plus usuels est donné par le théorème de Jirina : il suffit que E soit un espace métrique complet séparable (ou espace polonais), c'est-à-dire admettant un sous-ensemble partout dense, ce qui est le cas de $\mathbb{R}^p$.

3.2.2.3 *Ce qu'il faut retenir*

Il ressort des résultats précédents les propriétés utiles suivantes : si (X, Y) est un couple de variables aléatoires où Y est à valeurs dans $\mathbb{R}$ et X à valeurs dans un ensemble fini ou dénombrable, où à valeurs dans $\mathbb{R}$ ou $\mathbb{R}^p$:

- Il existe une mesure de probabilité conditionnelle $P(./X = x)$ sur Ω.
- Il existe une distribution conditionnelle de $Y/X = x$.

- Si $E(Y)$ existe, alors il existe une variable aléatoire espérance conditionnelle : $E(Y/X)$ qui prend les valeurs $E(Y/X = x)$ avec la loi de probabilité P_X :

$$E(Y/X = x) = \int_{\Omega} Y(\omega)\, \mathrm{d}P(\omega/X = x) = \int_{\mathbb{R}} y\, \mathrm{d}P(y/X = x)$$

et $E[E(Y/X)] = E(Y)$.
- Si $V(Y)$ existe on a $V(Y) = E(V(Y/X)) + V(E(X/Y))$.
- Si le couple (X, Y) est à valeur dans $\mathbb{R}^2$ et possède une densité $h(x, y)$ les densités conditionnelles existent et sont données par :

$$g(y/x) = \frac{h(x\,;y)}{f(x)} \qquad f(x/y) = \frac{h(x\,;y)}{g(y)}$$

et on a $E(Y/X = x) = \int_{\mathbb{R}} yg(y/x)\,\mathrm{d}y$ ainsi que les formules de Bayes pour les densités :

$$g(y/x) = \frac{f(x/y)g(y)}{\int_{\mathbb{R}} f(x/y)g(y)\,\mathrm{d}y} \qquad f(x/y) = \frac{g(y/x)f(x)}{\int_{\mathbb{R}} g(y/x)f(x)\,\mathrm{d}x}$$

- Lorsque l'une des variables est discrète et l'autre possède une densité il suffit de remplacer là où c'est nécessaire les intégrales par des sommes finies et les densités par des probabilités ponctuelles.

3.3 SYNTHÈSE GÉOMÉTRIQUE

Le cas où on n'étudie que des variables aléatoires réelles de moment d'ordre 2 fini est un des plus importants en pratique et est susceptible d'interprétations géométriques très éclairantes.

3.3.1 Espace de Hilbert des classes de variables aléatoires de carré intégrables

L'ensemble de toutes les variables aléatoires définies sur un même univers (en fait l'ensemble des classes de variables aléatoires presque partout égales) forme un espace de Hilbert L^2 si l'on le munit du produit scalaire :

$$\langle X, Y \rangle = E(XY) \qquad \text{et de la norme : } \|X\| = (E(X^2))^{1/2}$$

L'écart-type est donc la norme des variables centrées, et la covariance le produit scalaire des variables centrées.

Si l'on considère l'ensemble des variables aléatoires constantes, on obtient une droite D de L^2. Car si X est constante, aX l'est aussi.

L'espérance mathématique de X est alors la projection orthogonale de X sur cette droite (fig. 3.2) : en effet, on sait que le minimum de $E((X - a)^2)$ est atteint pour $a = E(X)$, ce qui définit la projection orthogonale de X sur D.

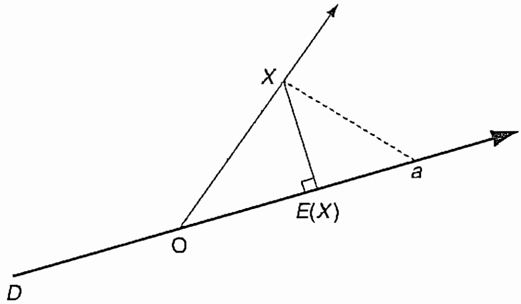

FIGURE 3.2

La formule de König-Huyghens :

$$E((X-a)^2) = V(X) + (E(X)-a)^2$$

s'interprète comme le théorème de Pythagore appliqué au triangle rectangle X, $E(X)$, a.

$E(X)$ est, en d'autres termes, la meilleure approximation de la variable X par une constante (au sens de la norme de L^2).

Comme $\text{cov}(X, Y) = \langle X - E(X) ; Y - E(Y) \rangle$ l'inégalité de Schwarz donne :

$$|\text{cov}(X, Y)| \leq \|X - E(X)\| \; \|Y - E(Y)\|$$

soit :

$$|\text{cov}(X, Y)| \leq \sigma_X \sigma_Y$$

Le cosinus de l'angle formé par $X - E(X)$ et $Y - E(Y)$ vaut donc $\dfrac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$. On retrouve le coefficient de corrélation linéaire ρ entre X et Y.

Dans cet espace, la non corrélation se traduit donc par l'orthogonalité

$\rho = \pm 1$ si $|\text{cov}(X, Y)| = \sigma_X \sigma_Y$ donc si $(X - E(X))$ et $(Y - E(Y))$ sont proportionnelles soit : $X - E(X) = a(Y - E(Y))$.

Le coefficient de corrélation linéaire est donc égal à ± 1 s'il y a une relation linéaire entre les deux variables X et Y.

La nullité de ce coefficient exclut la relation linéaire, mais n'exclut pas l'existence d'autres relations. Il est facile de fabriquer des contre-exemples de dépendance fonctionnelle avec un coefficient de corrélation linéaire nul : ainsi, X et X^2 ou $\sin X$ et $\cos X$ lorsque la loi de X est symétrique.

3.3.2 Espérance conditionnelle et projection

Soit L_X^2 le sous-espace de L^2 constitué des variables aléatoires fonctions seulement de X du type $\varphi(X)$: L_X^2 est convexe et contient la droite des constantes D.

C'est donc un sous-espace de Hilbert fermé.

Alors l'espérance conditionnelle de Y sachant X, $E(Y/X)$, s'interprète comme la projection orthogonale de Y sur L_X^2.

Soit en effet l'opérateur qui associe à toute variable aléatoire son espérance conditionnelle à X. C'est un opérateur linéaire ; pour montrer que c'est un projecteur orthogonal il suffit de vérifier qu'il est idempotent et auto-adjoint :

- il est idempotent : $E(E(Y/X)/X) = E(Y/X)$;
- et auto-adjoint : $\langle Z ; E(Y/X) \rangle = \langle E(Z/X) ; Y \rangle$.

En effet, les deux membres de cette relation sont égaux à $E[E(Z/X)E(Y/X)]$.

Le théorème de l'espérance totale $E(Y) = E(E(Y/X))$ est alors un cas particulier du théorème des trois perpendiculaires, comme l'illustre la figure 3.3.

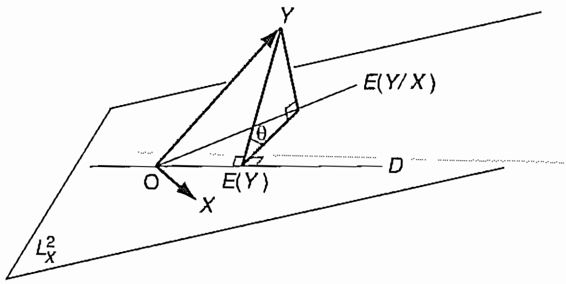

FIGURE 3.3

$E(Y/X)$ étant une projection orthogonale, ceci montre que le minimum de :

$$E[(Y - \varphi(X))^2]$$

est atteint pour $\varphi(X) = E(Y/X)$, résultat qui sera repris lors de l'étude de la régression. On peut dire que si $E(Y)$ est la meilleure approximation de Y par une constante, $E(Y/X)$ est la meilleure approximation de Y par une fonction de X.

Il est alors immédiat que le « résidu » $Y - E(Y/X)$ est non corrélé avec X par suite de l'orthogonalité.

Le théorème de la variance totale s'interprète comme le théorème de Pythagore appliqué au triangle rectangle Y, $E(Y)$, $E(Y/X)$:

$$\begin{aligned} \|Y - E(Y)\|^2 = V(Y) &= \|E(Y/X) - E(Y)\|^2 + \|Y - E(Y/X)\|^2 \\ &= V(E(Y/X)) + E[(Y - E(Y/X))^2] \\ &= V(E(Y/X)) + E[E(Y - E(Y/X))^2] \\ &= V(E(Y/X)) + E(V(Y/X)) \end{aligned}$$

3.3.3 Rapport de corrélation de Y en X

Le coefficient de corrélation linéaire ρ est une mesure symétrique de dépendance, qui est maximale dans le cas de la liaison linéaire.

Le théorème de la variance totale permet de définir une autre mesure de liaison non symétrique cette fois : le rapport de corrélation $\eta_{Y/X}$ tel que :

$$\boxed{\eta^2_{Y/X} = \frac{V(E(Y/X))}{V(Y)}}$$

Ce rapport est le cosinus carré de l'angle formé par $Y - E(Y)$ et l'espace L^2_X.

On a donc : $$0 \leq \eta^2_{Y/X} \leq 1$$

PROPRIÉTÉ

Si $\eta^2_{Y/X} = 1$, $E(V(Y/X)) = 0$.

On en déduit donc que $V(Y/X) = 0$ presque sûrement, car c'est une variable positive. Ce qui veut dire qu'à X fixé la variance de Y est nulle, donc que Y ne prend qu'une seule valeur.

$$\eta^2_{Y/X} = 1 \Rightarrow Y = \varphi(X)$$

Le rapport de corrélation est maximal si Y est lié fonctionnellement à X.

PROPRIÉTÉ

Si $\eta^2_{Y/X} = 0$, $V(E(Y/X)) = 0$, $E(Y/X)$ est donc presque sûrement une constante.

On dit que Y est non corrélé avec X, il y a absence de dépendance en moyenne. C'est en particulier le cas si X et Y sont indépendantes mais la réciproque est inexacte. On montre en fait que l'indépendance entre Y et X est équivalente à l'orthogonalité des espaces L^2_X et L^2_Y engendrés par X et Y le long de la droite des constantes (fig. 3.4) :

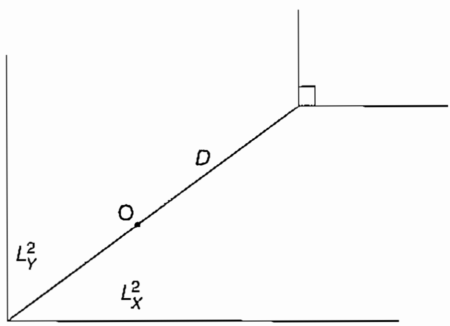

FIGURE 3.4 Indépendance de X et Y.

$\eta^2_{Y/X} = 0$ signifie seulement que $Y - E(Y)$ est orthogonal à L^2_X ;
η^2 est une mesure de liaison fonctionnelle alors que ρ est une mesure de liaison linéaire ;
$\eta^2_{Y/X}$ est toujours supérieur ou égal à ρ^2 car ρ^2 est le cosinus carré de l'angle formé par $Y - E(Y)$ avec le sous-espace de dimension 2 de L^2_X engendré par la droite des constantes D et la variable X.

Le cas ou $\eta^2_{Y/X} = \rho^2$ signifie donc que $E(Y/X)$ appartient à ce sous-espace de dimension 2, donc que :

$$E(Y/X) = \alpha + \beta X$$

c'est celui de la régression linéaire dont l'étude sera effectuée en détail au chapitre 16.

Si $E(Y/X) = \alpha + \beta X$, on ne peut trouver de transformation de X augmentant ρ.

En effet d'une part $\eta^2_{Y/X} = \sup_{\varphi} \rho^2(Y\,;\varphi(X))$, et d'autre part la linéarité de la régression implique $\eta^2_{Y/X} = \rho^2(Y\,;X)$.

Lorsque $(Y\,;X)$ est un couple gaussien on a simultanément $E(Y/X) = \alpha + \beta X$ *et* $E(X/Y) = \gamma + \delta Y$

On en déduit le théorème suivant :

THÉORÈME

Si $(Y\,;X)$ est un couple gaussien, on ne peut pas trouver de transformations $\varphi(X)$ et $\psi(Y)$ augmentant en valeur absolue le coefficient de corrélation :

$$\rho^2(\varphi(X)\,;\psi(Y)) \leq \rho^2$$

Les prévisions optimales (en moyenne quadratique) sont donc linéaires.