



CHAPITRE 1 :

COMPLÉMENTS DE PROBABILITÉS

Table des matières

Chapitre 1 : Compléments de probabilités.....	1
1. Vecteurs aléatoires de dimension 2	2
1.1. Fonction de répartition, mesure de probabilité et lois marginales	2
1.2. Indépendance et corrélation.....	3
2. Vecteurs aléatoires de dimension supérieure à 2.....	4
2.1. Fonction de répartition et mesure de probabilité	4
2.2. Matrice de covariance.....	5
2.3. Fonction caractéristique et indépendance	5
2.4. Vecteurs gaussiens	6
2.5. Transformation de vecteurs aléatoires continus	7
3. Distributions et moments conditionnels.....	7
3.1. Loi conditionnelle	7
3.2. Espérance conditionnelle	9
3.3. Moments d'ordre p conditionnels	10
3.4. Variance conditionnelle.....	10
4. Espaces vectoriels L^1 et L^2	10
4.1. L^2 espace de Hilbert.....	11
4.2. Interprétations géométriques.....	11
4.3. Espérance conditionnelle et projection - Régression.....	12
Exercices	14

Références

- Bogaert P. (2006). *Probabilités pour scientifiques et ingénieurs*. Ed. De Boeck
- Saporta G. (1990). *Probabilités, analyse des données et statistiques*. Ed. Technip
- Lefèbvre M. (2005). *Processus Stochastiques appliqués*. Ed. Hermann

1. VECTEURS ALEATOIRES DE DIMENSION 2

Soit (Ω, \mathcal{A}, P) un espace probabilisé. Soient X et Y deux variables aléatoires :

$$X : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$$

$$Y : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$$

On étudie alors le couple (X, Y) à valeurs dans \mathbb{R}^2 muni de sa tribu borélienne.

1.1. Fonction de répartition, mesure de probabilité et lois marginales

La *fonction de répartition* du couple (X, Y) donnée par

$$F(x, y) = P[(X \leq x) \cap (Y \leq y)],$$

définit alors la loi jointe du couple.

On appelle *lois marginales* les lois de probabilités de X et Y pris séparément.

Cas de variables aléatoires continues

Si le couple (X, Y) admet une *fonction de densité conjointe* f alors on a

$$f(x, y) = \frac{\partial^2 F}{\partial x \partial y}(x, y),$$

et on retrouve

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(s, t) dt ds.$$

Les fonctions de répartition marginales se déduisent immédiatement,

$$F_X(x) = \int_{-\infty}^x \int_{-\infty}^{+\infty} f(s, t) dt ds$$

$$F_Y(y) = \int_{-\infty}^{+\infty} \int_{-\infty}^y f(s, t) dt ds$$

Les fonctions de densité marginales s'obtiennent en dérivant les fonctions de répartition marginales

$$\text{Loi marginale de } X : f_X(x) = \int_{-\infty}^{+\infty} f(x, t) dt \text{ (dérivée par rapport à } x)$$

$$\text{Loi marginale de } Y : f_Y(y) = \int_{-\infty}^{+\infty} f(s, y) ds \text{ (dérivée par rapport à } y)$$

De la même façon que pour une variable aléatoire, on calcule la probabilité d'un événement A de la tribu engendrée par le couple (X, Y) , par exemple $A = \{(X, Y), X+Y=0\}$, par l'intégrale

$$P(A) = \int \int_A f(x, y) dy dx = \int \int_{\{x+y=0\}} f(x, y) dy dx.$$

L'espérance s'obtient de la même manière

$$E[h(X,Y)] = \iint_{\mathbb{R}^2} h(x,y)f(x,y)dydx .$$

Exemple 1

Soit le vecteur aléatoire (X,Y) de fonction de densité conjointe suivante

$$f(x,y) = \begin{cases} 2ye^{-x} & \text{si } x \geq 0 \text{ et } 0 < y < 1 \\ 0 & \text{sinon} \end{cases} .$$

Les lois marginales sont données par

$$f_x(x) = \begin{cases} e^{-x} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases} \quad \text{et} \quad f_y(y) = \begin{cases} 2y & \text{si } 0 < y < 1 \\ 0 & \text{sinon} \end{cases}$$

La fonction de répartition conjointe est

$$F(s,t) = \begin{cases} (1-e^{-s})t^2 & \text{si } s \geq 0 \text{ et } 0 < t < 1 \\ 1-e^{-s} & \text{si } s \geq 0 \text{ et } t \geq 1 \\ 0 & \text{sinon} \end{cases} .$$

On peut aussi calculer la probabilité $P(X+Y > 1) = 2e^{-1}$.

Cas de variables aléatoires discrètes

Supposons que les variables aléatoires X et Y prennent des valeurs dénombrables respectivement $x_i, i \in I$ et $y_j, j \in J$. On définit la *fonction de masse conjointe* du couple (X,Y) par

$$\forall i \in I, \forall j \in J, p(x_i, y_j) = P[(X=x_i) \cap (Y=y_j)].$$

Les fonctions de masse marginales sont données par

$$\text{Loi marginale de } X : p_x(x_i) = \sum_{j \in J} p(x_i, y_j) \quad \forall i \in I.$$

$$\text{Loi marginale de } Y : p_y(y_j) = \sum_{i \in I} p(x_i, y_j) \quad \forall j \in J.$$

De la même façon, on peut calculer la probabilité d'un événement ainsi que l'espérance.

1.2 Indépendance et corrélation

Rappelons que deux variables aléatoires X et Y sont indépendantes si pour tout couple de boréliens B_i et B_j , on a

$$P[(X \in B_i) \cap (Y \in B_j)] = P(X \in B_i) \times P(Y \in B_j).$$

On en déduit que X et Y sont indépendantes si et seulement si la fonction de répartition du couple est égale au produit des fonctions de répartition marginales,

$$F(x,y) = F_x(x) \times F_y(y).$$

Cas continu

Si X et Y admettent pour fonctions de densité f et g respectivement, alors X et Y sont indépendantes si et seulement si h la fonction de densité du couple (X,Y) est égale au produit des fonctions de densité,

$$h(x,y)=f(x)\times g(y).$$

Exemple 1 (suite)

Dans l'exemple précédent, la densité conjointe est égale au produit des densités marginales donc les variables X et Y sont indépendantes.

Propriétés

Si X et Y sont deux variables aléatoires indépendantes alors :

$$\begin{aligned}E(XY) &= E(X)\times E(Y) \\ \text{var}(X+Y) &= \text{var}(X)+\text{var}(Y) \\ \text{cov}(X,Y) &= 0\end{aligned}$$

(attention, en général, les réciproques sont fausses)

Définition

On appelle *coefficient de corrélation linéaire* de X et Y le rapport

$$C = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}.$$

On dit que X et Y sont *complètement corrélées* si $|C|=1$ et sont *non corrélées* si $C=0$.

Remarque

Le coefficient de corrélation linéaire traduit l'existence ou non d'une relation linéaire entre X et Y (cf. §5).

Propriété

Si X et Y sont indépendantes alors elles sont non corrélées.

2. VECTEURS ALEATOIRES DE DIMENSION SUPERIEURE A 2

Soit (Ω, \mathcal{A}, P) un espace probabilisé. Soit $(X_i)_{i \in \{1, \dots, n\}}$ une suite de variables aléatoires :

$$X_i : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}, \forall i \in \{1, \dots, n\}$$

On étudie alors (X_1, \dots, X_n) le vecteur aléatoire à valeurs dans \mathbb{R}^n muni de sa tribu borélienne.

2.1. Fonction de répartition et mesure de probabilité

La fonction de répartition du vecteur (X_1, \dots, X_n) donnée par

$$F(x_1, \dots, x_n) = P[(X_1 \leq x_1) \cap \dots \cap (X_n \leq x_n)],$$

définit alors la loi jointe du vecteur.

Cas de variables aléatoires continues

Si le vecteur (X_1, \dots, X_n) admet une fonction de densité f alors on a

$$f(x_1, \dots, x_n) = \frac{\partial^n F}{\partial x_1 \dots \partial x_n}(x_1, \dots, x_n),$$

Cas de variables aléatoires discrètes

Supposons que les variables aléatoires X_i prennent des valeurs dénombrables. On définit la fonction de masse du vecteur (X_1, \dots, X_n) par

$$p(x_1^{(k)}, \dots, x_n^{(k)}) = P[(X_1 = x_1^{(k)}) \cap \dots \cap (X_n = x_n^{(k)})].$$

2.2. Matrice de covariance

Soit le vecteur $X = (X_1, \dots, X_n)$. Notons $\mu_i = E(X_i)$ l'espérance des variables, et σ_i^2 leur variance, $i=1, \dots, n$. L'espérance du vecteur X est donnée par le vecteur colonne,

$$E(X) = \mu = {}^t[\mu_1, \dots, \mu_n].$$

On définit alors la *matrice de covariance*

$$\begin{pmatrix} \sigma_1^2 & \text{cov}(X_1, X_2) & \dots & \text{cov}(X_1, X_n) \\ & \sigma_2^2 & & \\ \dots & & \dots & \\ & & & \sigma_n^2 \end{pmatrix}$$

2.3. Fonction caractéristique et indépendance

Soit (Ω, \mathcal{A}, P) un espace probabilisé. Soit $X = (X_1, \dots, X_n)$ un vecteur de variables aléatoires.

Définition

On définit la *fonction caractéristique* de X par

$$\forall \omega = (\omega_1, \dots, \omega_n) \in \mathbb{R}^n, \Phi_X(\omega) = E[e^{i \sum_{k=1}^n \omega_k X_k}].$$

Propriété

Les variables aléatoires X_1, \dots, X_n sont indépendantes si et seulement si la fonction caractéristique de X est égale au produit des fonctions caractéristiques des X_k ,

$$\forall \omega = (\omega_1, \dots, \omega_n) \in \mathbb{R}^n, \Phi_X(\omega) = \prod_{k=1}^n \Phi_{X_k}(\omega).$$

Propriété (Théorème de Cramer-Wold)

La loi de X est entièrement déterminée par celles de toutes les combinaisons linéaires de ses composantes.

2.4. Vecteurs gaussiens

Le vecteur X est dit un vecteur aléatoire *gaussien* ou *multinormal* à n dimensions si toutes combinaisons linéaires des X_k ,

$$Z = \sum_{i=1}^n a_i X_i,$$

est une variable aléatoire gaussienne (de loi normale).

La loi multinormale ou gaussienne à n dimensions d'espérance $\mu = (\mu_1, \dots, \mu_n)$ et de matrice de covariance $\Gamma \in M_n$ est notée $N_n(\mu, \Gamma)$.

Remarques

- Le théorème de Cramer-Wold permet d'établir que la loi de X est entièrement déterminée.
- On note aussi que la normalité de chaque composante du vecteur ne suffit pas pour avoir un vecteur gaussien.

Fonction caractéristique d'un vecteur gaussien

Afin de simplifier les notations, considérons un vecteur aléatoire $X = (X_1, \dots, X_n)$ gaussien centré, $\mathcal{N}_n(0, \Gamma)$. Alors la fonction caractéristique est donnée par

$$\Phi_X(\omega) = \exp\left(-\frac{1}{2} \omega^T \Gamma \omega\right).$$

Propriété

Les variables aléatoires X_k composantes du vecteur X sont indépendantes si et seulement si la matrice de covariance Γ est diagonale, c'est-à-dire si les X_k sont non corrélés.

Fonction de densité d'un vecteur gaussien

Soit un vecteur aléatoire gaussien de dimension n , $X = (X_1, \dots, X_n)$ gaussien, $\mathcal{N}_n(\mu, \Gamma)$. Si la matrice Γ est régulière (Γ^{-1} existe), alors X admet pour fonction de densité,

$$f_X(x) = \frac{1}{(\sqrt{2\pi})^n (\det \Gamma)^{1/2}} \exp\left[-\frac{1}{2} (x - \mu)^T \Gamma^{-1} (x - \mu)\right].$$

Cas particulier

Si X est un vecteur aléatoire gaussien dont les composantes sont des variables aléatoires normales centrées réduites indépendantes, alors

$$f_X(x) = \frac{1}{(\sqrt{2\pi})^n} \exp\left[-\frac{1}{2} \sum_{k=1}^n x_k^2\right].$$

Convergence (Théorème de la limite centrale multidimensionnel)

Soient X_1, \dots, X_p une suite de vecteurs aléatoires de taille n , indépendants, de même loi, d'espérance le vecteur μ et de matrice de covariance Γ , alors

$$\frac{1}{\sqrt{p}} \sum_{i=1}^p (X_i - \mu) \xrightarrow[p \rightarrow \infty]{\mathcal{L}} N_n(0, \Gamma).$$

2.5. Transformation de vecteurs aléatoires continus

Soient (Ω, \mathcal{A}, P) un espace probabilisé et (X_1, \dots, X_n) un vecteur aléatoire de fonction de densité f . Soit ϕ une application bijective de \mathbb{R}^n dans \mathbb{R}^n ,

$$\phi : x = (x_1, \dots, x_n) \mapsto \phi(x_1, \dots, x_n) = (\phi_1(x), \dots, \phi_n(x)).$$

La fonction de densité du vecteur aléatoire transformé $Y = \phi(X)$, est donnée par

$$g(y) = \frac{f[\phi^{-1}(y)]}{|\det J|},$$

où J est la matrice jacobienne de ϕ .

Exemple 2

Soient X_1 et X_2 deux variables aléatoires indépendantes de loi normale centrée réduite. Alors la loi conjointe de (X_1, X_2) est

$$f(x_1, x_2) = \frac{1}{2\pi} e^{-\frac{1}{2}(x_1^2 + x_2^2)}$$

On note $Y_1 = X_1 + X_2$ et $Y_2 = X_1 + 2X_2$. Alors l'application ϕ est définie par

$$\phi(x_1, x_2) = (y_1, y_2) = (x_1 + x_2, x_1 + 2x_2)$$

d'où

$$\phi^{-1}(y_1, y_2) = (x_1, x_2) = (2y_1 - y_2, y_2 - y_1) \text{ et } J = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

Finalement, la densité conjointe du vecteur (Y_1, Y_2) est

$$g(y_1, y_2) = \frac{1}{2\pi} e^{-\frac{1}{2}(5y_1^2 + 2y_2^2 - 6y_1y_2)}.$$

3. DISTRIBUTIONS ET MOMENTS CONDITIONNELS

3.1. Loi conditionnelle

Cas discret

Soient X et Y deux variables aléatoires sur (Ω, \mathcal{A}, P) . Supposons que les variables aléatoires X et Y prennent des valeurs dénombrables respectivement $x_i, i \in I$ et $y_j, j \in J$.

On note F la fonction de répartition du couple (X, Y) et F_Y celle de Y . On considère l'événement $\{Y = y_j\}$. La *fonction de répartition conditionnelle* de X sachant $Y = y_j$ est donnée par

$$F_{X|Y=y_j}(x) = F(x|Y=y_j) = \frac{P[X \leq x, Y = y_j]}{P(Y = y_j)}.$$

On note p la fonction de masse conjointe du couple (X, Y) et p_Y celle de Y . On définit la *fonction de masse conditionnelle* de X sachant $Y = y_j$ par

$$p_{X|Y=y_j}(x_i) \equiv p(x_i|y_j) = \frac{p(x_i, y_j)}{p_Y(y_j)}, \quad \forall i \in I.$$

Remarque : On a bien $\sum_{i \in I} p(x_i | y_j) = 1$.

Cas continu

Supposons que les variables aléatoires X et Y sont continues. On ne peut pas conditionner sur l'évènement $\{Y=y\}$ directement car $P(Y=y)=0$. On doit prendre la limite lorsque ε décroît vers 0 des fonctions définies en conditionnant sur l'évènement $\{y < Y < y + \varepsilon\}$. On obtient alors les définitions suivantes.

On note f la fonction de densité conjointe du couple (X,Y) et f_Y celle de Y. On la *fonction de répartition conditionnelle* de X sachant $Y=y$, où $f_Y(y) > 0$, est donnée par

$$F_{X|Y=y}(x) = F(x|Y=y) = \frac{\int_{-\infty}^x f(u,y) du}{f_Y(y)}.$$

On définit la *fonction de densité conditionnelle* de X sachant $Y=y$ où $f_Y(y) > 0$ par

$$f_{X|Y=y}(x) \equiv f(x|y) = \frac{f(x,y)}{f_Y(y)}.$$

Remarque : On a bien $\int_{-\infty}^{+\infty} f(x|y) dx = 1$.

Exemple 1 (suite)

Dans le cas du 1^{er} exemple, les variables étant indépendantes, les fonctions de densité conditionnelles sont égales aux fonctions de densités marginales.

Exemple 3

Soit la couple de variables aléatoires (X, Y) de fonction de densité conjointe

$$f(x,y) = \begin{cases} \ln x / x & \text{si } 1 \leq x \leq e \text{ et } 0 < y < x \\ 0 & \text{sinon} \end{cases}.$$

On calcule les densités marginales,

$$f_X(x) = \begin{cases} \ln x & \text{si } 1 \leq x \leq e \\ 0 & \text{sinon} \end{cases} \quad \text{et} \quad f_Y(y) = \begin{cases} 1/2 & \text{si } 0 < y < 1 \\ (1 - \ln^2 y) / 2 & \text{si } 1 \leq y < e \\ 0 & \text{sinon} \end{cases}.$$

La densité conditionnelle par rapport à $\{X=x\}$ n'existe que pour $x \in [1, e]$ et on a

$$f_{Y|X}(y) = \begin{cases} 1/x & \text{si } 0 < y < x \\ 0 & \text{sinon} \end{cases}.$$

On remarque que la variable aléatoire $Y | \{X=x\}$ suit une loi uniforme sur l'intervalle $[0, x]$.

La densité conditionnelle par rapport à $\{Y=y\}$ n'existe que pour $y \in]0, e[$ et on a :

$$\text{si } 0 < y < 1, f_{X|Y}(x) = \begin{cases} 2 \frac{\ln x}{x} & \text{si } 1 \leq x \leq e \\ 0 & \text{sinon} \end{cases} \quad \text{et} \quad \text{si } 1 \leq y < e, f_{X|Y}(x) = \begin{cases} \frac{2}{1 - \ln^2 y} \frac{\ln x}{x} & \text{si } x > y \\ 0 & \text{sinon} \end{cases}.$$

3.2. Espérance conditionnelle

Soient X et Y deux variables aléatoires sur (Ω, \mathcal{A}, P) et h une application telle $h(X)$ soit une variable aléatoire.

Cas discret

Supposons que les variables aléatoires X et Y prennent des valeurs dénombrables respectivement $x_i, i \in I$ et $y_j, j \in J$. On considère l'événement $\{Y=y_j\}$ et on définit l'espérance conditionnelle de $h(X)$ par

$$E_{X|Y=y_j}[h(X)] \equiv E[h(X)|Y=y_j] = \sum_{i \in I} h(x_i) p(x_i | y_j).$$

Cas continu

Supposons X et Y continues et on considère l'événement $\{Y=y\}$. On définit l'espérance conditionnelle de $h(X)$ par

$$E_{X|Y=y}[h(X)] \equiv E[h(X)|y] = \int_{-\infty}^{+\infty} h(x) f(x|y) dx.$$

Exemple 3 (suite)

Soit $x \in [1, e]$, $E[Y|X=x] = \int_{\mathbb{R}} y f(y|x) dy = \int_0^x y/x dy = x/2$.

Soit $y \in]0, 1[$, $E[X|Y=y] = \int_{\mathbb{R}} x f(x|y) dx = \int_1^e 2 \ln x dx = 2$.

Propriétés

Soient X et Y deux variables aléatoires de nature quelconque.

- L'espérance conditionnelle de X sachant Y est une variable aléatoire fonction de Y .
- L'espérance conditionnelle est linéaire.
- Soit g une fonction telle que $g(Y)$ soit une variable aléatoire, alors $E[g(Y)X|Y] = g(Y)E[X|Y]$ presque sûrement.

Théorème de l'espérance totale

$$E(E[X|Y]) = E[X]$$

Inégalité de Jensen

Soit g une fonction convexe sur \mathbb{R} telle que $E[|g(X)|] < \infty$ alors

$$g(E[X|Y]) \leq E[g(X)|Y].$$

En particulier pour $g : x \mapsto |x|$.

Remarque

Soient X et Y deux variables aléatoires. On dit que $X=Y$ presque sûrement (noté p.s.) si $X(\omega)=Y(\omega)$ pour tout $\omega \in \Omega$ sauf pour un sous-ensemble de Ω de mesure nulle (en des points isolés).

3.3. Moments d'ordre p conditionnels

Le moment conditionnel d'ordre p se déduit des expressions précédentes en considérant la fonction $h : x \rightarrow x^p$.

Cas discret

$$E_{X|Y=y_j}[X^p] \equiv E[X^p | y_j] = \sum_{i \in I} x_i^p p(x_i | y_j).$$

Cas continu

$$E_{X|Y=y}[X^p] \equiv E[X^p | y] = \int_{-\infty}^{+\infty} x^p f(x | y) dx.$$

3.4. Variance conditionnelle

Soient X et Y deux variables aléatoires sur (Ω, \mathcal{A}, P) . Notons μ_X l'espérance de X. De même que la variance, la *variance conditionnelle* est l'espérance conditionnelle de la variable centrée au carré.

Cas discret

Supposons que les variables aléatoires X et Y prennent des valeurs dénombrables respectivement $x_i, i \in I$ et $y_j, j \in J$. On considère l'événement $\{Y=y_j\}$ et on définit la variance conditionnelle par

$$\text{var}_{X|Y=y_j}(X) \equiv \text{var}(X | y_j) = E[(X - E(X | Y))^2 | y_j] = \sum_{i \in I} (x_i - E(X | Y=y_j))^2 p(x_i | y_j).$$

Cas continu

Supposons X et Y continues et on considère l'événement $\{Y=y\}$. On définit la variance conditionnelle par

$$\text{var}_{X|Y=y}(X) \equiv \text{var}(X | y) = E[(X - E(X | Y))^2 | y] = \int_{-\infty}^{+\infty} (x - E(X | Y=y))^2 f(x | y) dx.$$

Propriété de la variance totale

$$\text{var}(X) = E[\text{var}(X | Y)] + \text{var}[E(X | Y)]$$

Remarque

Les définitions de ce paragraphe peuvent se généraliser au cas où X et Y sont des vecteurs aléatoires.

4. ESPACES VECTORIELS L^1 ET L^2

Soit (Ω, \mathcal{A}, P) un espace probabilisé.

On note $L^1(\Omega)$ l'espace vectoriel des variables aléatoires intégrables au sens de Lebesgue, c'est-à-dire si X est une variable aléatoire de fonction de densité f,

$$E[|X|] = \int |x| f(x) dx < +\infty.$$

On note $L^2(\Omega)$ l'espace vectoriel des variables aléatoires de carré sommable au sens de Lebesgue, c'est-à-dire si X est une variable aléatoire de fonction de densité f ,

$$E[X^2] = \int x^2 f(x) dx < +\infty.$$

On dit des variables aléatoires de L^2 qu'elles sont d'ordre 2 (le moment d'ordre 2 est fini).

4.1. L^2 espace de Hilbert

Soient (Ω, \mathcal{A}, P) un espace probabilisé et $(X, Y) \in L^2 \times L^2$ un couple de variables aléatoires d'ordre 2.

On définit un produit scalaire sur L^2 par la forme bilinéaire symétrique, définie positive non dégénérée suivante :

$$\langle X, Y \rangle = E[XY].$$

Il en résulte la norme,

$$\|X\|^2 = E[X^2].$$

Propriété

L^2 muni de la distance associée à la norme est un espace de Hilbert.

4.2. Interprétations géométriques

On remarque que :

- L'écart-type est la norme des variables centrées : $\sigma_x = \|X - E(X)\|$.
- La covariance est le produit scalaire des variables centrées :

$$\text{cov}(X, Y) = \langle X - E(X), Y - E(Y) \rangle.$$

Interprétation de l'espérance

Si on considère l'ensemble des variables aléatoires constantes, on obtient une droite D de L^2 . L'espérance mathématique est alors la projection orthogonale de X sur cette droite. 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 . En d'autre terme, $E(X)$ est la meilleure approximation de la variable X par une constante (au sens de la norme L^2).

Interprétation du coefficient de corrélation linéaire

Le coefficient de corrélation linéaire correspond au cosinus de l'angle formé par les vecteurs $X - E(X)$ et $Y - E(Y)$. Donc la non corrélation ($C=0$) signifie que les vecteurs sont orthogonaux et la corrélation ($C=\pm 1$) que les vecteurs sont proportionnels. On a alors $Y - E(Y) = k(X - E(X))$, c'est-à-dire une relation linéaire entre X et Y . La nullité du coefficient de corrélation exclut la relation linéaire entre X et Y mais n'exclut pas d'autres relations.

4.3. Espérance conditionnelle et projection - Régression

Soit L_x^2 le sous-espace de L^2 constitué des variables aléatoires fonctions seulement de X , c'est-à-dire du type $\varphi(X)$. Cet espace est un sous-espace de Hilbert fermé. L'espérance conditionnelle de Y sachant X peut donc s'interpréter, $E[Y|X]$, comme la projection orthogonale de Y sur L_x^2 .

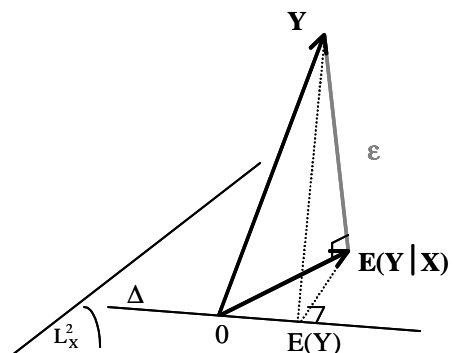
Soit en effet l'opérateur qui associe à toute variable aléatoire son espérance conditionnelle à X . Cet opérateur vérifie les propriétés d'un projecteur orthogonal :

- linéaire
- idempotent
- auto-adjoint

$E[Y|X]$ étant une projection orthogonale, ceci montre que le minimum de $E[(Y-\varphi(X))^2]$ est atteint pour $\varphi(X)=E[Y|X]$. On peut donc dire que, de même que $E(Y)$ est la meilleure approximation de Y par une constante, $E[Y|X]$ est la meilleure approximation de Y par une fonction de X . La fonction,

$$x \mapsto E[Y|x],$$

est appelée *fonction de régression* de Y en X .



Propriété

Les variables $Y-E[Y|X]$ et X sont non corrélés.