III ESPERANCE MATHEMATIQUE

I.Définition et calcul de l'espérance mathématique d'une VA

- La <u>définition</u> la plus générale de l'espérance d'un VA $X:\Omega \to R^+$ (*donc à valeurs positives ou nulles*) est obtenue en introduisant une suite de partitions Π_n de R^+ :

$R^+ = [0,x_1[\cup[x_1,x_2[\cup ... \cup[x_{2^n-1},x_{2^n}[\cup[x_{2^n},\infty[$ où $x_k = \dfrac{kn}{2^n}, k = 0,1,..,2^n$ et $x_{k+1} = \infty$

L'espérance de X est alors définie comme la limite de la somme des valeurs x_k pondérées par les probabilités des intervalles $[x_k, x_{k+1}[$ auxquels ils appartiennent

$E(X) = \lim_{n\to\infty} \sum_{i=0}^{2^n} x_i^n P_X([x_i^n, x_{i+1}^n[)$ et on note $E(X) = \int_{R^+} x dP_X(x)$

(Remarquer que $\forall n, \sum_{i=0}^{2^n} P_X([x_i^n, x_{i+1}^n[) = P(X \in R^+) = 1$)

- Pour une VA $X:\Omega \to R$ *pouvant prendre des valeurs négatives aussi bien que positives* on introduit la décomposition $X = \max(X,0) - \max(-X,0) = X^+ - X^-$ et on définit $E(X)$ par $E(X) =_{def} E(X^+) - E(X^-) = \int_{R^+} x dP_X(x) + \int_{R^-} x dP_X(x)$ si $E(X^+)$ et $E(X^-)$ ne sont pas simultanément infinis..
- <u>De cette définition on peut déduire, cas particulier par cas particulier des formules de calcul</u>[1] .

Si la fonction de répartition F_X présente des sauts (discontinuités) aux points $\alpha_i, i \in I$ (I dénombrable) d'amplitude $F_X(\alpha_i^+) - F_X(\alpha_i) = P(X = \alpha_i) = q_i, i \in I$ et qu'elle est dérivable ailleurs au sens ordinaire avec des valeurs de dérivée non nulles on a :

$$\boxed{E(X) = \sum_{i \in I} \alpha_i q_i + \int_{R - \{\alpha_i, i \in I\}} F_X^{'}(x)\, x\, dx \qquad (1)}$$

(où la somme continue se calcule à l'extérieur des points α_i de discontinuité)

Si la VA est de **loi discrète**, on a $\sum_{i \in I} q_i = 1$ et $F_X^{'}(x) = 0 \forall x \in R - \{\alpha_i, i \in I\}$ si bien que l'espérance devient :

$$\boxed{E(X) = \sum_{i \in I} \alpha_i q_i = \sum_{i \in I} \alpha_i P(X = \alpha_i) \qquad (2)}$$

Si la VA X admet une densité de probabilité p_X (cad si elle est de **loi continue**) on a $\forall x P(X = x) = 0$ (il n'y a pas de saut dans F_X) et $F_X^{'}(x) = p_X$. La somme discrète dans (1) devient alors nulle et l'espérance s'écrit :

[1] Il n'est pas nécessaire de connaître parfaitement la définition générale de l'espérance donnée si dessus pour appliquer ces formules et calculer des valeurs moyennes

$$E(X) = \int_R p_X(x)\, x\, dx \qquad (3)$$

Vocabulaire et notation : on dit couramment **valeur moyenne** pour espérance mathématique et on note $m_X =_{def} E(X)$.
Interprétation : si on réalise *n* fois la même expérience aléatoire pour obtenir *n* réalisations $X(\omega_i) = x_i, i = 1..,n$ et que l'on considère la moyenne arithmétique de ces résultats, $\frac{1}{n}\sum_1^n x_i$, cette dernière pour *n* très grand tendra vers une limite égale à $E(X)$ (on le montre théoriquement sous certaines hypothèses et on peut le 'constater' expérimentalement).

II.Espérance d'une VA fonction d'autres VA(formule de transfert)

Soit une VA *Y* définie à partir de *N* VA $X_1,..,X_N$ et d'une fonction $f : R^N \to R \; : Y = f(X_1,..,X_N)$). La formule de transfert permet de calculer $E(Y)$ sans exhiber préalablement sa loi P_Y. Elle s'écrit dans son expression la plus générale $E(Y) = \int_{R^N} f(x_1,..,x_N)\, dP_{X_1,..X_N}(x_1,..,x_N)$. Les formules de calcul à a utiliser en pratique dépendent de la nature de la loi conjointe des X_i.

- Si la loi conjointe $P_{X_1,..X_N}$ admet une densité $p_{X_1,..X_N}$ (loi de type continu) alors on aura :

$$E(Y) = \int_{R^N} f(x_1,..,x_N)\, p_{X_1,..X_N}(x_1,..,x_N) dx_1..dx_N$$

- Si la loi conjointe est discrète, cad si il existe un ensemble dénombrable de points de R^N $\alpha_i = (x_1^i,..,x_N^i), i \in I$ tel que : $\forall i\, P(X_1 = x_1^i,..,X_N = x_N^i) = q_i$ avec $\sum_{i\in I} q_i = 1$ alors $E(Y)$ se calcule par :

$$E(Y) = \sum_{i\in I} q_i f(x_1^i,..,x_N^i) = \sum_{i\in I} q_i f(\alpha_i)$$

- Le cas plus général d'une loi qui n'est ni de type continu ni de type discret n'est simple à écrire que pour $N = 1$ auquel cas on a :

$$E(Y) = \sum_{i\in I} f(\alpha_i) q_i + \int_{R-\{\alpha_i, i\in I\}} F_X'(x) f(x)\, dx$$

(avec les mêmes notations que pour (1))

Pour $N > 1$ des termes complémentaires du type intégrale curviligne ou intégrale de surface peuvent intervenir (on ne donne pas ici de formule générale correspondante).

III.Propriétés de l'espérance mathématique <u>utiles dans les calculs courants</u> (autres que la formule de transfert).

- *Positivité* : si $P(X \geq 0) = 1$ alors $E(X) \geq 0$
- *Espérance d'une constante K* : si $P(X = K) = 1, K = cte$ alors $E(X) = K$
- *Linéarité* : si, pour *N* VA $X_1,..,X_N$, $Y = \sum_1^N \lambda_k X_k$ alors $E(Y) = \sum_1^N \lambda_k E(X_k)$

- *Indépendance et factorisation* : soient *N* VA $X_1,..,X_N$ **indépendantes** dans l'ensemble et soient *N* VA $Y_1 = f_1(X_1),..,Y_N = f_N(X_N)$ construites à partir de *N* fonctions $f_k : R \to R, k = 1,..,N$. L'espérance de la VA $Y = \prod_{k=1}^{N} Y_k$ est alors $E(Y) = \prod_{k=1}^{N} E(Y_k)$

  - *Remarque* : les *N* VA $X_1,..,X_N$ étant indépendantes, les *N* VA $Y_1 = f_1(X_1),..,Y_{N1} = f_N(X_N)$ le sont aussi.
  - *Corollaire* : $X_1,..,X_N$ indépendantes $\Rightarrow E(\prod_{k=1}^{N} X_k)) = \prod_{k=1}^{N} E(X_k)$

  Cette propriété reste vraie si $X_1,..,X_N$ sont *N* VA à valeurs respectivement dans $R^{d_1},..,R^{d_N}$ et les f_k de la forme $R^{d_k} \to R^{d_k'}$

IV.Moments d'une VA, variance d'une VA

1.Définition : on appelle moment d'ordre *N* d'une VA X l'espérance $E(X^N)$ (si elle existe) .

2.Définition : A une VA *X* de valeur moyenne m_X on associe la VA notée X_c , appelée '*X* centrée' que l'on définit par $X_c = X - m_X$. *On dira également qu'une VA X est centrée si sa valeur moyenne est nulle, auquel cas* $X = X_c$.

On a toujours $E(X_c) = E(X - m_X) = E(X) - m_X = m_X - m_X = 0$

3.Définition : on appelle moment *centré* d'ordre *N* d'une VA X la quantité $E(X_c^N)$ (si elle existe) .

4.Propriété (inégalité de Markov) : $\forall \varepsilon > 0 \forall n \geq 0 : P(|X| \geq \varepsilon) \leq \dfrac{E(|X|^n)}{\varepsilon^n}$

5.Définition : la variance d'une VA *X* est son moment centré d'ordre 2, $VAR(X) = E(X_c^2)$

6.Propriétés de la variance :

- $E(X^2) = VAR(X) + m_X^2$
- $VAR(\alpha X + \beta) = \alpha^2 VAR(X) \; \forall \alpha, \beta$ réels
- Si $X_1,..,X_N$ indépendantes et $Y = X_1 + .. + X_N$ alors $VAR(Y) = VAR(X_1) + .. + VAR(X_N)$
- De manière plus générale $VAR(\sum_1^N \lambda_i X_i) = \sum\sum_{1 \leq i,j \leq N} \lambda_i \lambda_j E(X_{ic} X_{jc})$

  qui devient $VAR(\sum_1^N \lambda_i X_i) = \sum_{i=1}^{N} \lambda_i^2 E(X_{ic}^2)$ **si** $i \neq j \Rightarrow E(X_{ic} X_{jc}) = 0$ (condition qui sera réalisée en particulier si les *N* VA sont indépendantes 2 à 2).
- Inégalité de Bienaymé Tchebychef (faire $n = 1$, remplacer X par X_c dans Markov) :

  $\forall \varepsilon > 0 : P(|X| \geq \varepsilon) \leq \dfrac{VAR(X)}{\varepsilon^2}$

V.Fonction caractéristique et calculs de moments

V.1.Variables aléatoires à valeurs complexes

1.Définition :une VA sur (Ω,τ,P) à valeur dans C (corps des complexes) est une application $X:\omega\in\Omega\to X(\omega)=U(\omega)+iV(\omega)$ où (U,V) est une paire de VA sur (Ω,τ,P), chacune à valeurs dans R.

Remarque : la définition se généralise sans problème au cas de VA *N*dimensionnelles à valeur dans C^N.

2.Loi de probabilité.

La loi de *Z* correspond à la loi conjointe du couple (U,V). En notant $z=u+iv$ on écriera :

$F_Z(z)=F_{U,V}(u,v),(u,v)\in R^2$

$p_Z(z)=p_{U,V}(u,v),(u,v)\in R^2$ si (U,V) est de loi conjointe continue

Ceci se généralise pour une VA à valeurs dans C^N par

$F_{Z_1,..,Z_n}(z_1,..,z_N)=F_{U_1,..,U_N,V_1,..,V_N}(u_1,..u_N,v_1,..,v_N),(u_1,..u_N,v_1,..,v_N)\in R^{2N}$

$p_{Z_1,..,Z_n}(z_1,..,z_N)=p_{U_1,..,U_N,V_1,..,V_N}(u_1,..u_N,v_1,..,v_N),(u_1,..u_N,v_1,..,v_N)\in R^{2N}$

3.Définitions *de la moyenne et de la variance* :

$E(X)=_{def}(E(U)+iE(V))\in C$, '*X* centrée' : $X_c=U_c+iV_c$

$VAR(X)=_{def}E(|X_c|^2)=E(U_c^2)+E(V_c^2)=VAR(U)+VAR(V)$

V.2.Fonction caractéristique et moments

1.Définition :

- La fonction caractéristique d'une VA *X* à valeurs dans *R* est l'application $\varphi_X:u\in R\to\varphi_X(u)=E(e^{iuX})=E(\cos uX)+iE(\sin uX)\in C$
- La fonction caractéristique d'une VA *N*-dimensionnelle $(X_1,..,X_N)$ à valeurs dans R^N est l'application

$$\varphi_{X_1,..,X_N}:(u_1,..,u_N)\in R^N\to\varphi_{X_1,..,X_N}(u_1,..,u_N)=E(\exp(\sum_{k=1..,N}iu_kX_k)\in C$$

2.Propriété (relations avec les moments)

Pour X VA à valeurs dans R :

- Si le moment $E(X^n)$ est défini alors on a $E(X^n)=\frac{1}{i^n}\frac{\partial^n}{\partial u^n}\varphi_X(0)$ et la fonction caractéristique admet le développement de Taylor à l'ordre *n* autour de l'origine :

$$\varphi_X(u)=\sum_{k=0..n}E(X^n)\frac{i^nu^n}{n!}+\varepsilon(u^{n+1})u^{n+1}$$

- Si le moment $E(X^n)$ existe pour tout *n* on a le développement infini

$$\varphi_X(u)=\sum_{k=0..\infty}E(X^n)\frac{i^nu^n}{n!}$$

On retiendra que :

les moments d'une VA peuvent donc être calculés en dérivant la fonction caractéristique où en la développant en série de Taylor autour de l'origine.

3.Fonction caractéristique et transformée de Fourier (TF)
Si la VA X est de loi continue on a

$$\varphi_X : u \in R \to \varphi_X(u) = E(e^{iuX}) = \int_R e^{iux} p_X(x) dx$$

ce qui montre, en notant $\hat{p}_X$ la TF de p_X, que $\varphi_X(u) = \hat{p}_X(-\frac{u}{2\pi}), u \in R$

et donc, qu'au changement de variable près, la fonction caractéristique est la transformée de Fourier de la densité de probabilité. La transformée de Fourier étant une bijection (la transformation de Fourier inverse permet de retrouver la fonction d'origine[2]) ceci montre qu'il est possible de retrouver la densité de probabilité à partir de la fonction caractéristique et qu'il y a donc correspondance biunivoque entre une loi de probabilité continue et la fonction caractéristique . On montre que ceci reste vrai pour des lois quelconques, la fonction caractéristique s'avérant ainsi être toujours une spécification exacte de la loi de probabilité correspondante.

VI.Coefficient de corrélation entre 2 VA réelles

1.Meilleure approximation affine d'une VA à partir d'une autre VA.

Soit 2 variables aléatoires X et Y . Supposons que l'on observe $X(\omega) = x$. Peut on alors calculer une approximation de la réalisation $Y(\omega) = y$ au moyen d'une fonction $y = f(x)$. Plus précisément existe-il une fonction $f : R \to R$ telle que, pour toute autre fonction $g : R \to R$ on ait $E((Y - f(X))^2) \leq E((Y - g(X))^2)$, ce qui revient à rechercher :

$$f = \arg\min_f E((Y - g(X))^2)$$

L'espérance $E((Y - g(X))^2)$ est appelée *erreur quadratique moyenne* (*EQM*) entre la variable 'cible' et son approximation $g(X)$. Elle ne peut être que positive ou nulle. Pour être nulle il y a nécessité que $P(Y = g(X)) = 1$ (on peut le montrer en utilisant l'inégalité de B.T.). Cette erreur permet d'évaluer l'erreur d'approximation sur l'ensemble des cas rencontrés $(X(\omega), Y(\omega))$ en tenant compte de leurs fréquences relatives d'apparition.

On peut contraindre le problème en imposant à f d'appartenir à une certaine classe Ψ de fonctions :

$$f = \arg\min_{f \in \Psi} E((Y - g(X))^2)$$

Cherchons la solution du problème dans le cas Ψ où est la classe des fonctions affines. Il faut alors trouver 2 constantes réelles a et b telles que $(a,b) = \arg\min_{(A,B) \in R^2} E((Y - AX - B))^2)$. On a :

$$E((Y - AX - B)^2) = E((Y_C + m_Y - AX_C - Am_X - B)^2) =$$
$$E((Y_C - AX_C)^2) + (m_Y - Am_X - B)^2 - 2E(Y_C - AX_C)(m_Y - Am_X - B) =$$
$$E((Y_C - AX_C)^2) + (m_Y - Am_X - B)^2$$

Cette dernière quantité est minimale pour $B = m_Y - Am_X$ et pour A qui minimise

[2] En toute rigueur à quelques détails 'négligeables' près (notion mathématique de fonction presque partout égales)

$E(Y_C^2) - 2AE(X_C Y_C) + A^2 E(X_C^2)$ qui est un trinôme du second degré en A. Ce trinôme admet un seul minimum (en supposant $E(X_C^2) \neq 0$) en $A = \dfrac{E(X_C Y_C)}{E(X_C^2)}$. On a donc :

$$\left(\frac{E(X_C Y_C)}{E(X_C^2)}, m_Y - \frac{E(X_C Y_C)}{E(X_C^2)} m_X\right) = \underset{(A,B)\in R^2}{\arg\min} E((Y - AX - B))^2)$$

et si on développe les calculs, pour ces valeurs optimales des coefficients A et B on trouve que la valeur minimale de $E((Y - AX - B)^2)$ est égale à :

$$\min_{A,B} E((Y - AX - B)^2) = E(Y_c^2)(1 - \rho_{X,Y}^2) \text{ où } \rho_{X,Y} =_{def} \frac{E(X_c Y_c)}{\sqrt{VAR(X)VAR(Y)}} = \frac{E(X_c Y_c)}{\sigma_X \sigma_Y}$$

Exercice : vérifier la première égalité ci-dessus

2.Définition du coefficient de corrélation entre 2 VA

On appelle coefficient de corrélation $\rho_{X,Y}$ entre les VA X et Y la quantité $\dfrac{E(X_c Y_c)}{\sigma_X \sigma_Y}$

(rappelons que $E(X_c Y_c) = E(X\ Y) - m_X m_Y$)

Calcul de $\rho_{X,Y}$ *:*

Il suffit de calculer $m_X, m_Y, \sigma_X, \sigma_Y$ et $E(XY)$

à partir d'une densité conjointe $p_{X,Y}$ *on calculera :*

$$E(XY) = \iint_{R^2} p_{X,Y}(x, y) xy dx dy$$

dans le cas d'une loi discrète à 2 dimensions on calculera:

$$E(XY) = \sum_{x_i, y_j} x_i y_j P(X = x_i, Y = y_j)$$

3.Propriétés *du coefficient de corrélation*

- Si X et Y sont indépendantes alors $\rho_{X,Y} = 0$ (**attention : réciproque fausse**)
- $|\rho_{X,Y}| \leq 1$
- $\rho_{X,Y} = +1 \Rightarrow X_c = \lambda Y_c$ pour un certain $\lambda > 0$
- $\rho_{X,Y} = -1 \Rightarrow X_c = \lambda Y_c$ pour un certain $\lambda < 0$

4.Approche par le produit scalaire entre VA

Introduisons l'ensemble de toutes les VA d'ordre 2 (correspondant à une même expérience aléatoire (Ω, τ, P), cad celui de toutes les VA Z telles que $E(Z^2)$ est bien définie (certaines lois de probabilité n'admettent pas de moment d'ordre 2 comme la loi de Cauchy par exemple qui n'en admet aucun). Pour 2VA quelconques Z_1, Z_2 de cet ensemble on montre qu'il est toujours possible de calculer l'espérance du produit $Z_1 Z_2$. Du fait des propriétés de l'espérance mathématique cette opération a toutes les propriétés d'un produit scalaire :

- *symétrie* : $E(Z_1 Z_2) = E(Z_2 Z_1)$
- *linéarité* : $E((aZ_1 + bZ_2)Z_3) = aE(Z_1 Z_3) + bE(Z_2 Z_3)$
- *positivité* : $E(Z^2) \geq 0$, *caractère défini* : $(E(Z^2) = 0) \Rightarrow P(Z = 0) = 1$

A ce produit scalaire peut être associé une norme : $\|V\|^2 = 1$

Une propriété de tout produit scalaire (V_1,V_2) est l'inégalité de Schwartz : $|(V_1,V_2)| \le \|V_1\|\|V_2\|$ (avec égalité ssi $\exists$ réel $\lambda \neq 0 : V_1 = \lambda V_2$)

Avec $V_1 = \frac{Z_{1c}}{\sigma_{X_1}}, V_2 = \frac{Z_{2c}}{\sigma_{X_2}}$, $\|V_1\| = \|V_2\| = 1$ et en appliquant l'inégalité on arrive à :

$$\left| E\left(\frac{Z_{1c}}{\sigma_{X_1}} \frac{Z_{2c}}{\sigma_{X_2}}\right) \right|^2 \le E\left(\left[\frac{Z_{1c}}{\sigma_{X_1}}\right]^2\right) E\left(\left[\frac{Z_{2c}}{\sigma_{X2}}\right]^2\right) = 1$$

ce qui correspond à $|\rho_{X,Y}| \le 1$ en tenant compte des définitions de la variance et du coefficient de corrélation .

5.Retour sur le problème d'approximation

L'erreur d'approximation dans le problème introduit plus haut valait

$$\varepsilon^2 = \min_{A,B} E((Y - AX - B)^2) = E(Y_c^2)(1 - \rho_{X,Y}^2)$$

On voit donc que cette erreur est comprise entre une valeur minimale nulle quand le coefficient de corrélation atteint une valeur maximale en valeur absolue égale à 1 (et on sait alors que cela correspond à l'existence d'une relation linéaire exacte entre les variables centrées, du moins avec probabilité 1) et une valeur maximale égale à *VAR(Y)* lorsque le coefficient est nul. Dans ce dernier cas la valeur optimale de *A* est nulle et on peut dire que si les variables sont décorrélées (cad $\rho_{X,Y} = 0$) alors la meilleure approximation affine de *Y* se ramène à la valeur constante $m_Y = E(Y)$: il ne sert à rien d'utiliser $X(\omega)$ pour évaluer $Y(\omega)$.

Conclusion :Il y a une correspondance entre la valeur plus ou moins élevée de $\rho_{X,Y}$ et la possibilité de prédire linéairement Y_c à partir de X_c .

VII Espérances conditionnelles.

1.Définition de l'espérance conditionnelle.
Soit un couple (X,Y) de VA, chacune à valeurs dans *R*. La définition la plus directe de l'espérance de *Y* si $X = x$ est :

$$E(Y/X=x)=\int_{y\in R} y dP_{Y/X=x}(y)dy$$

Autrement dit $E(Y/X=x)$ est la moyenne pour la loi conditionnelle $P_{Y/X=x}$. En toute rigueur cette loi n'est définie que P_X presque sûrement (cad pour un ensemble de valeurs de *x* contenant un borélien *A* tel que $P(X\in A)=1$). Pour chacune de ces valeurs de *x* la loi conditionnelle $P_{Y/X=x}$ peut être discrète, continue ou mixte. <u>La variable aléatoire conditionnante *X* peut être à valeurs dans *R* où dans</u> R^N.

2.Formules pratiques de calcul.
Le calcul de l'espérance conditionnelle s'effectue suivant les mêmes méthodes que pour une espérance ordinaire (non conditionnelle). Les formules qui suivent permettent de calculer l'espérance conditionnelle de $f(Y)$ conditionnellement à $X = x$. Elles correspondent à la formule de transfert dans le cas conditionnel. Pour obtenir l'espérance conditionnelle de Y conditionnellement à $X = x$ il suffit d'y remplacer $f(.)$ par l'application identité. Les V.A. *Y* et *X* peuvent être à valeurs respectivement dans R^M et R^N, $M\geq 1$ et $N\geq 1$. On considère ici f de la forme $f:R^N\to R$. Si $f(.)$ est l'application identité on considère $M=1$.

- Si $P_{Y/X=x}$ admet une densité $p_{Y/X=x}$ (loi de type continu) alors :

$$E(f(Y)/X=x)=\int_{R^M} f(y)\,p_{Y/X=x}(y)dy$$

- Si la loi conditionnelle est discrète, cad si il existe un ensemble dénombrable de points de R^M, $y_j, j\in I$ tel que $\sum_{j\in J} P_{Y/X=x}(\{y_j\})=\sum_{j\in J} P(Y=y_j/X=x)=1$ alors $E(f(Y)/X=x)$ se calcule par :

$$\sum_{j\in J} f(y_j)P_{Y/X=x}(\{y_j\})=\sum_{j\in J} f(y_j)P(Y=y_j/X=x)$$

- Si *Y* <u>est à valeurs dans *R*</u> et que la loi conditionnelle est mixte avec une fonction de répartition conditionnelle $F_{Y/X=x}$:

$$E(f(Y)/X=x)=\sum_{i\in I} f(\alpha_i)[F'_{Y/X=x}(\alpha_i^+)-F'_{Y/X=x}(\alpha_i)]+\int_{R-\{\alpha_i,i\in I\}} f(y)F'_{Y/X=x}(x)dy$$

  où les α_i sont les points de discontinuité de $F_{Y/X=x}$.

3.Propriétés de l'espérance conditionnelle.
- Positivité :

$P(Y\geq 0)=1\Rightarrow E(Y/X=x)\geq 0$ et ceci $P_X\,ps$ (cad presque sûrement dans la loi P_X)
- Linéarité :

$E(aY_1+bY_2\ /X=x)=E(aY_1/X=x)+E(_1bY_2\ /X=x)$,*a* et *b* ctes ($P_X\,ps$)

- Formule de déconditionnement .

Cette formule est fondamentale dans les applications. Elle utilise le fait que l'application $x \to h(x) = E(Y / X = x)$ est mesurable (on le montre) et que $h \circ X = h(X)$ correspond donc à une variable aléatoire dont on peut chercher à calculer l'espérance. Elle s'écrit :

$$\boxed{E(Y) = E(h(X)) = \int_{x \in R^N} h(X)\, dP_X(x) = \int_{x \in R^N} E(Y / X = x)\, dP_X(x)}$$

où $\int_{x \in R^N} (.)\, dP_X(x)$ se calcule en utilisant les formules appropriées suivant que la loi de X est absolument continue, discrète ou encore mixte.

Ceci montre au passage que si Y est d'espérance finie il en est de même pour $E(Y / X)$ mais la réciproque est en général fausse.

> *Remarque* : l'utilisation de la variable aléatoire auxiliaire X et la chaîne de calculs conditionnement + déconditionnement pour calculer *E(Y)* sont recommandés lorsque le calcul de $E(Y / X = x)$ s'avère facile et 'naturel', voire évident, dans le contexte de l'étude (généralement parce que la loi conditionnelle est elle même évidente) . Le passage par ces deux étapes de calcul peut s'avérer alors économique par rapport à un calcul plus direct de *E(Y)* dans la loi P_Y si cette dernière n'est pas connue a priori et qu'elle est difficile à calculer.

4 Approche géométrique de l'espérance conditionnelle : projections orthogonales.

On considère l'espace $L^2(\Omega, \tau, P)$ (qu'on notera plus brièvement L^2) des VA $X : \Omega \to R$ sur (Ω, τ, P) admettant un moment d'ordre 2. Plus précisément c'est l'ensemble des classes d'équivalences de VA qui est considéré, deux VA X et Y étant équivalentes par définition si $\|X - Y\|^2 = E((X - Y)^2) = 0$, ce qui sera vrai chaque fois que $P(X = Y) = 1$ (ce point de détail pouvant cependant être négligé en première lecture). On montre que cet espace est un espace de Hilbert pour le produit scalaire introduit en **VI.4**, cad. un espace vectoriel complet pour la norme $\|X\|^2 = E(X^2)$. Dans ces conditions on peut appliquer tous les théorèmes existants pour les espaces de Hilbert, en particulier le théorème des projections orthogonales:

Théorème (projections orthogonales):

Soit $(E, <.,.>)$ un espace de Hilbert muni de son produit scalaire et F un sous espace de Hilbert inclus dans E (cad un sous espace vectoriel complet pour la norme associée au produit scalaire). Pour X arbitraire dans E :

1) $\exists!\, X_P \in F : \forall Z \in F : \{Z \neq X_P \Rightarrow \|X - Z\| > \|X - X_P\|\}$

2) X_p est l'unique solution du système $< X - X_P, Z >= 0\ \forall Z \in F$

Le point 2) correspond à ce qu'on appelle la condition d'orthogonalité

Dans le cas où l'on dispose d'un système générateur G_E pour E (en particulier d'une base) le système d'équations dans 2) est équivalent à $< X - X_P, Z >= 0\ \forall Z \in G_E$, ce qui en général simplifie le problème. X_p s'appelle la projection orthogonale de X sur E.

Considérons (X, Y) une paire de VA de $E = L^2$ et introduisons l'ensemble F_X (ne pas confondre ici avec une fonction de répartition) des VA de L^2 égales (presque sûrement) à $f(X)$ pour une certaine fonction $f : R \to R$. Il est possible de vérifier que F_X est un sous espace fermé de E et donc de considérer la projection de Y sur F_X .

Définition
Pour Y dans L^2 on appelle espérance conditionnelle de Y si X la projection orthogonale, notée $E(Y/X)$, de Y sur F_X.

Dans cette définition il n'est pas exigé que $X \in L^2$ mais par contre il est obligatoire de supposer Y dans L^2. Sous réserve que cette dernière hypothèse est vraie et si on introduit $h : x \to h(x) = E(Y / X = x) = \int\limits_{y \in R} y dP_{Y/X=x}(y)dy$, ce qui correspond à la première définition de l'espérance donnée en **VII.1**, on a l'égalité (presque sûre pour prendre les précautions habituelles, toujours aussi négligeables en première lecture) :

$$E(Y/X) \overset{ps}{=} h(X)$$

La plupart des propriétés de l'espérance conditionnelle sont vérifiées pour les 2 définitions. Elles peuvent ainsi être démontrées par 2 approches différentes. A titre d'exemple considérons la propriété: $E(E(Y/X)) = E(Y)$.
Montrons d'abord cette propriété à partir de la première définition et dans le cas où la loi conjointe de (X,Y) admet une densité:

$$E(E(Y/X)) = E(h(X)) = \int\limits_R h(x) p_X(x)dx = \int\limits_R [\int\limits_R y p_{Y/X=x}(y)dy] p_X(x)dx$$

$$= \iint\limits_{R^2} y\, p_{Y/X=x}(y) p_X(x)\, dxdy = \iint\limits_{R^2} y\, p_{X,Y}(x,y)\, dxdy$$

$$= \int\limits_R y [\int\limits_R p_{X,Y}(x,y)\, dx] dy = \int\limits_R y\, p_Y(y)\, dy = E(Y)$$

Considérons à présent la preuve dans le cadre des projections orthogonales. La condition d'orthogonalité s'écrit:
$\forall f(X) \in L^2 : E((Y - E(Y/X)) f(X)) = 0$, soit encore:
$\forall f(X) \in L^2 : E((Y f(X)) = E((Y/X) f(X))$ et donc, en particularisant à $f : x \to f(x) = 1$ (fonction égale à 1 partout sur R) il vient $E(Y) = E(E(Y/X))$, cqfd.

<u>Exercice</u>: montrer que si $Z = g(X)$ on a $E(E(Y/X)/Z)) = E(E(Y/Z)/X)) = E(Y/Z)$ et interpréter en termes géométriques (2 projections) cette propriété.

<u>Exercice</u>: montrer que $E(X.Y/Y) = Y.E(X/Y)$ et que

$$E(g(X,Y)/Y = y) = E(g(X,y)/X = y)$$