Année universitaire 2010-2011

الجمهورية الجزائرية الديموقراطية الشعبية

République Algérienne Démocratique et Populaire

وزارة التعليم العالي و البحث العلمي

Minstère de l'Enseignement Supérieur et de la Recherche Scientifique

CENTRE UNIVERSITAIRE DE MILA

INSTITUT DES SCIENCES ET DE LA TECHNOLOGIE

Mémoire de fin d'étude

Présenté pour l'obtention du diplôme de

Licence Académique

Domaine: **Mathématique ET informatique.**

Filière : **Mathématiques.**

Spécialité : **Mathématiques Fondamentale.**

Thème

Espérance conditionnelle

Présenté par :

Zid mouncef Ramy

Boudabane Mahmoud

Dirigé par :

Bououden Rabeh

Année universitaire 2010-2011

Table des matières

Introduction

.

L'espèrance conditionnelle est un concept important en probabilités, notamment utilisé dans des domaines tels que l'étude des martingales et l'intégration stochastique. Néaumaines, sa définition dans le cas général n'est pas simple. c'est pour quoi nous présentant l'idée par étapes et de façon intuitive, cas discret, cas absolument continu, interprétation géométrique dans L^2.

Chapitre 1

Notions préliminaires

1.1 Rappel de la théorie de la mesure

Soient $(\Omega, \mathcal{A})$ un espace mesurable, P et Q deux mesures positives sur $(\Omega, \mathcal{A})$.

Définition 1.1 *:*
- Q est dite concentrée sur $A \in \mathcal{A}$ si :

$$Q(A \cap B) = Q(B) \quad \forall B \in \mathcal{A}\,;$$

c.a.d tout événement disjoint de$A$ est Q négligeable.
- Si Q est concentrée sur A et P est concentrée sur B avec A et B disjoints alors Q et$P$ sont dites singulières, on note $(Q \perp P)$.
- Q est dites absolument continue par rapport à P si :

$$\forall A \in \mathcal{A} : (P(A) = 0) \Longrightarrow (Q(A) = 0)\,;$$

on note $(Q \ll P)$.
Si $(Q(\Omega) < \infty)$, une C.N.S. pour que $(Q \ll P)$ est que :

$$\forall \varepsilon > 0,\ \exists \delta > 0 / \left(P(A) < \delta\right) \Longrightarrow \left(Q(A) < \varepsilon\right).$$

Théorème 1.2 ***(Radon Nykodim Lebesgue) :***
Si Q est une mesure sur $(\Omega, \mathcal{A})$ et P une mesure σ finie sur $(\Omega, \mathcal{A})$ alors il existe un couple unique (Q_{AC}, Q_S) de mesures tel que : $(Q = Q_{AC} + Q_S)$ où $(Q_S \perp P)$ (décomposition de Lebesgue).
De plus, il existe une classe unique de fonctions f égales P p.s de $(\Omega, \mathcal{A}) \longrightarrow (\overline{\mathbb{R}}^+, \mathcal{B}(\overline{\mathbb{R}}^+))$

telle que :

$$Q_{AC}(A) = \int_A f dP \quad \forall A \in \mathcal{A} \quad \textit{(Radon Nykodim).}$$

Ces fonction sont P p.s finies ssi Q est σ finie et sont intégrables ssi Q est bornée.
f est appelée la dérivée de Radon Nykodim de Q par rapport à P .elle est notée $\frac{dQ}{dP}$.

1.2 Espérance (ou moyenne)

Définition 1.3 *:*
- Si $X = \sum\limits_{i=1}^{n} x_i 1_{A_i}$ (v.a.r étagée),on appelle espérance mathématique de X par rapport à la mesure de probabilité P le réel, notée $E(X)$, définit par :

$$E(X) = \int X dP = \sum_{i=1}^{n} x_i P(A_i) \quad \textit{(c'est l'intégrale de } X\textit{).}$$

- Si X est une $(v.a.r \geq 0)$, $X = \lim\limits_{n \longrightarrow +\infty} X_n$ où (X_n) est une suite de v.a.r étagées,on a alors :

$$\begin{aligned} E(X) &= \int X dP \\ &= \lim_{n \longrightarrow +\infty} X_n dP \\ &= \lim_{n \longrightarrow +\infty} E(X_n). \end{aligned}$$

[On montre que $E(X)$ est indépendante du choix de la suite (X_n)] .
$E(X)$ peut être infinie, si elle est finie on dit que X est intégrable.
- Si X est une v.a.r quelconque, on a :

$$\left(X = X^{+} - X^{-}\right).$$

X est dite intégrable si $\left(\int \mid X \mid dP < +\infty\right)$,on pose dans ce cas :

$$E(X) = \int X^{+} dP - \int X^{-} dP.$$

Remarque 1.4 *:*
On définit aussi $E(X)$ pour X seulement quasi intégrable (c.f.Neveu).

Propriété d'une espérance (reprises de celle de L'intégrale) :
* Si X et Y sont des v.a.r intégrables alors $(aX + bY)$ l'est aussi $(a, b \in \mathbb{R}$, avec $aX + bY$ définit).

* Si X et Y sont des v.a.r telles que $(X \geq Y)$ alors $(E(X) \geq E(Y))$ (positivité).
* Si (X_n) est une suite croissante de v.a.r positives tendant p.s vers X alors :

$$E(X) = \lim_{n \longrightarrow \infty} E(X_n) \text{ (théorème de convergence monotone noté (T.C.M.).}$$

* Si (X_n) est une suite de v.a.r convergeant p.s vers X et s'il existe une v.a.r z telle que $(\forall n, \mid X_n \mid \leq z)$ p.s alors X est intégrable et $E(X) = \lim_{n \longrightarrow \infty} E(X_n)$.(théorème de convergence dominé noté T.C.D).
* Si (X_n) est une suite de v.a.r positives alors :

$$\left(\int_{\Omega} \liminf \ X_n dP \right) \leq \left(\liminf \int_{\Omega} X_n dP \right) \quad \text{(lemme de Fatou).}$$

* Si (X_n) est une suite de v.a.r ≥ 0 tendant p.s vers X et s'il existe une constante M telleque :

$$\int_{\Omega} X_n dP \leq M \quad \forall n \in \mathbb{N}.$$

Alors :

$$\int_{\Omega} X \ dP \leq M \text{ (autre forme du Lemme de Fatou).}$$

* Si X est une v.a.r positive, $A \in \mathcal{A}$ avec $\left(\int_A X \ dP < \infty\right)$ alors $(X < +\infty)$ p.s sur A.
* Si (X_n) est une suite croissante de v.a.r positives telle que :

$$\sup_n \int X_n dP < +\infty,$$

alors :

$$\lim_n X_n < \infty \text{ p.s.}$$

* Soit X une v.a de $(\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ et g v.a.r. $(\Omega', \mathcal{A}') \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$
i) ($g \geq 0$ où g P_X intégrable) $\Longrightarrow \left(\int_{\Omega'} g \ dP_X = \int_{\Omega} g \circ X \ dP\right)$;
ii) Si X est une v.a.r ≥ 0 où intégrable alors :

$$E(X) = \int_{\Omega} X \ dP = \int_{\mathbb{R}} x \ dP_X(x) = \int_{\mathbb{R}} x \ dF_X(x) = \int x \ f(x) dx.$$

Ou F_X est la fonction de répartition de X et f est la densité de X (si elle existe).

Chapitre 2

Probabilité et esperance conditionnelle

2.1 Probabilité conditionnelle

2.1.1 Par rapport à une v.a discrète

Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité et soient A et B deux événements de $\mathcal{A}$.
La réalisation de A peut influer sur celle de B (soit en augmentant la chance de B sachant A et en la diminuant).
On définit la probabilité de B sachant A et on note $P(A/B)$ le nombre :

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \quad \text{si} \quad P(A) > 0.$$

Ce nombre est appelé probabilité conditionnelle de B par rapport à A (ou sachant A).
Le conditionnement par rapport à une v.a discrète est alors immédiat :
On définit pour une v.a X prenant les valeurs $(X_n)_{n \in N}$ la probabilité conditionnelle de A sachant $(X = x_n)$ et on note $P(A/X = x_n)$ le nombre :

$$P(A/X = x_n) = \begin{cases} \dfrac{P(A \cap (X = x_n))}{P(X = x_n)} & \text{si} \quad P(X = x_n) \neq 0. \\ 0 & \text{sinon} \end{cases}$$

Remarque 2.1 *:*
Si $(P(X = x_n) \neq 0)$ *alors l'application :*

$$\begin{array}{rcl} \mathcal{A} & \rightarrow & [0,1] \\ A & \rightarrow & P(A/X = x_n) \end{array}$$

Définit une mesure de probabilité.
On peut donc définir l'éspérance d'une v.a. Y *inégrable par rapport à cette probabilité par :*

$$E(Y/X = x_n) = \int Y(w)\ P(dw/X = x_n).$$

On montre que :

$$E(Y/X = x_n) = \frac{\displaystyle\int_{X=x_n} Y(w)P(dw)}{P(X = x_n)}$$

(Commencer par Y *étagée, puis positive et enfin intégrable).*
(C'est ce qu'on appelle l'espérance conditionnelle de Y *sachant* $(X = x_n)$*).*

2.1.2 Par rapport à une v.a continue

Soit X une v.a.r définie sur $(\Omega, \mathcal{A}, P)$ à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
Pour $A \in \mathcal{A}$ et $B \in \mathcal{B}(\mathbb{R})$ on a :

$$P(A/X \in B) = \frac{P(A \cap (X \in B))}{P(X \in B)} \quad \text{si} \quad P(X \in B) \neq 0.$$

Si on cherche $P(A/X = x)$, la définition donnée plus haut ne peut marcher que pour x tel que $P(X = x) \neq 0$, or dans le cas d'une v.a continue $P(X = x) = 0 \quad \forall x \in \mathbb{R}$.
L'idée de poser :

$$P(A/X = x) = \lim_{h_1, h_2 \underset{\rightarrow}{>} 0} \frac{P\left[A \cap (X \in]x - h_1, x + h_2[)\right]}{P(X \in]x - h_1, x + h_2[)}$$

Se heurte au problème d'existence de cette limite dans un cadre général.
Ce pendant,si on définit les deux mesures suivantes :

$$\begin{array}{rcl} \hat{Q}(B) & = & P(A \cap (X \in B)) \\ \hat{P}(B) & = & P(X \in B) = P_x(B) \end{array}$$

On a :$\left(\hat{Q} \ll \hat{P}\right)$.
D'après le théorème de Radon Nykodim, il existe une classe de fonctions $\mathcal{C}$, $\mathcal{B}(\mathbb{R})$ mesurables et intégrables telles que :

$$\hat{Q}(B) = \int_B \mathcal{C}(x) P_X(dx) = \int_{X^{-1}(B)} (\mathcal{C} \circ X)(w)\ P(dw) \ \ \forall B \in \mathcal{B}(\mathbb{R}).$$

D'ou :

Définition 2.2 *:*
$P(A/X = x)$ est n'importe quelle fonction $\mathcal{B}(\mathbb{R})$ mesurable telle que :

$$\forall B \in \mathcal{B}(\mathbb{R}),\ P(A \cap (X \in B)) = \int_B P(A/X = x)\ P_X(dx).$$

Remarque 2.3 *:*
Deux versions différentes de $P(A/X = x)$ sont P_X p.s égales, en effet :
Si g est $\mathcal{B}(\mathbb{R})$ mesurable avec :

$$\forall B \in \mathcal{B}(\mathbb{R}),\ \ P(A \cap (X \in B)) = \int_B g(x) P_X(dx).$$

Alors on a :

$$\int_B (g(x) - P(A/X = x)) P_X(dx) = 0 \qquad \forall B \in \mathcal{B}(\mathbb{R}).$$

D'ou :

$$\left(\int_{B_1} (g(x) - P(A/X = x)) P_X(dx) = 0\right) \Longrightarrow (g(x) = P(A/X = x)\ P_X\ p.s\ sur\ B_1)\,,$$

$$avec : B_1 = \{x \in B / g(x) - P(A/X = x) \geq 0\}\,.$$

Et

$$\left(\int_{B_2} (g(x) - P(A/X = x)) P_X(dx) = 0\right) \Longrightarrow (g(x) = P(A/X = x)\ P_X\ p.s\ sur\ B_2)\,.$$

$$avec : B_2 = \{x \in B / g(x) - P(A/X = x) < 0\}\,.$$

Donc :

$$g(x) = P(A/X = x)\ P_X\ p.s.$$

L'intégration par rapport à la loi image P_X se rapportant à une intégration par rapport à P on obtient :

Définition 2.4 *:*
La probabilité de A sachant X est tout v.a définie sur Ω notée $P(A/X)$, $\sigma(x)$ mésurable et vérifiant :

$$P(A \cap (X \in B)) \ = \int_{X^{-1}(B)} P(A/X)dP \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

Ou ce qui est équivalent :

$$P(A \cap B) = \int_{B} P(A/X)dP \qquad \forall B \in \sigma(x).$$

Remarque 2.5 *:*
- Deux versions de $P(A/X)$ sont P p.s égales.
- Si une version de $P(A/X = x) = \mathcal{F}(x)$ alors une version de $P(A/X)$ est égale à $(\mathcal{F} \circ X)$;
- $P(A/X)$ étant $\sigma(X)$ mesurable, son écriture $(\mathcal{F} \circ X)$ est prévisible ;
- La probabilité conditionnelle par rapport à une v.a ne dépend que de la tribu engendrée par cette v.a et non des valeurs qu'elle prend. Aussi deux v.a qui engendrent la même tribu définissent-elles la même probabilité conditionnelle.

2.2 Espérance conditionnelle

2.2.1 D'une v.a par rapport à une v.a

En procédant par analogie avec la probabilité conditionnelle on va définir l'espérance conditionnelle.
Soient :

$$\begin{aligned} X \ v.a \ &: \ (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}') \\ Y \ v.a \ &: \ (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})) \end{aligned}$$

Si $B \in \mathcal{A}'$ avec $P(X \in B) \neq 0$, il parait naturel de vouloir définir l'ésperance conditionnelle de Y sachant $(X \in B)$ notée $E(Y/X \in B)$ comme suit :

$$E(Y/X \in B) = \int Y(w)P(dw/X \in B),$$

où :

$$P(A/X \in B) = \frac{P(A \cap (X \in B))}{P(X \in B)}$$

Plus généralement définissons :

$$\begin{aligned}\hat{Q}(B) &= \int Y(w)\ P(dw \cap (X \in B)) \\ &= \int_{X \in B} Y(w)\ P(dw)\end{aligned}$$

$$P_x(B) = P(X \in B).$$

$\hat{Q}$ est une mesure absolument continue par rapport à P_x. D'où :

Définition 2.6 *:*
Si Y est une v.a.r intégrable sur $(\Omega, \mathcal{A}, P)$, $E(Y/X = x)$ est la classe de fonctions numériques $\mathcal{A}'$ mesurable et vérifiant :

$$\forall B \in \mathcal{A}'; \int_B E(Y/X = x) P_x(dx) = \int_{X^{-1}(B)} Y dP$$

$E(Y/X = x)$ est l'éspérance conditionnelle de Y sachant X .
Par définition de la mesure image on obtient :

Définition 2.7 *:*
Si Y est une v.a.r intégrable sur $(\Omega, \mathcal{A}, P)$, $E(Y/X)$ est toute fonction numérique $\sigma(x)$ mesurable et vérifiant :

$$\forall A \in \sigma(X) \int_A E(Y/X) dP = \int_A Y dP.$$

Remarque 2.8 *:*
- Deux versions de $E(Y/X = x)$ (resp.$E(Y/X\ )$) sont P_X p.s.égales (resp. P p.s égales)
- $E(Y/X)$ est aussi notée $E^X(Y)$.
- Si Y est $\sigma(X)$ mesurable alors $E(Y/X) = Y$ P p.s .
- $E(1_A\ /X = x) = P(A/X = x)$ P_X p.s et $E(1_A/X) = P(A/X)$ P p.s.

Proposition 2.9 *:*
X et Y sont deux v.a indépendantes ssi :

$$\forall A \in \mathcal{B}(\mathbb{R}),\ P(Y \in A/X) = k(A) \quad P \text{ p.s.(cste)}$$

Et on a alors :

$$k(A) = P(Y \in A) \quad P \text{ p.s.}$$

Preuve. :
Si X et Y sont indépendantes alors $\forall B \in \mathcal{A}^{'}$:

$$\begin{aligned}\int_{X^{-1}(B)} P(Y &\in A)dP = P(Y \in A)P(X^{-1}(B)) \\ &= P((Y \in A) \cap (X \in B)) \\ &= \int_{X\in B} 1_{(Y\in A)} \\ &= \int_{X\in B} P(Y \in A/X)dP.\end{aligned}$$

D'ou :

$$P(Y \in A/X) = P(Y \in A) \quad P \ p.s.$$

- **Réciproquement :**
Si $P(Y \in A/X) = k(A)$ on a $\forall B \in \mathcal{A}^{'}$:

$$P((Y \in A) \cap (X \in B)) = \int_{(X\in B)} P(Y \in A/X)dP = \int_{(X\in B)} k(A)dP = k(A)P(X \in B).$$

Si $\left(B = \Omega^{'}\right)$ on obtient :

$$\begin{aligned} P(Y &\in A) = k(A), \\ \text{d'où} \quad P((Y &\in A) \cap (X \in B)) = P(Y \in A)P(X \in B).\end{aligned}$$

Donc X et Y sont des v.a indépendantes. ■

Proposition 2.10 *:*
Si X et Y sont indépendantes alors :

$$E(Y/X) = E(Y) \quad P \ p.s.$$

Preuve. :
Soit $B \in \mathcal{A}^{'}$:

$$\begin{aligned}\int_{X\in B} YdP &= \int 1_{(X\in B)} YdP \\ &= E(1_{(X\in B)}Y) \\ &= P(X \in B)E(Y) \\ &= \int_{X\in B} E(Y)dP.\end{aligned}$$

D'ou :

$$E(Y/X) = E(Y) \quad P\ p.s.$$

■

Remarque 2.11 *:*
La réciproque de cette proposition est fausse en général.
En effet :
Soit $(Y = ZX)$ *avec* Z *et* X *des v.a.r. indépendantes et* Z *prenant les valeurs* 1 *et* -1 *avec probabilités* $\frac{1}{2}$*.*
Dans ce cas :

$$EY = EZ\ EX = 0.$$

Soit $B \in \mathcal{B}(\mathbb{R})$:

$$\begin{aligned}
\int_{X\in B} 0dP &= 0 \\
&= EZ\ E(X\ 1_{X\in B}) \\
&= E(ZX\ 1_{X\in B}) \\
&= \int_{X\in B} ZXdP \\
&= \int_{X\in B} YdP.
\end{aligned}$$

Donc :

$$E(Y/X) = 0 = EY\ \ P\ p.s.$$

Et pour tant X *et* Y *ne sont pas indépendantes.*

2.2.2 D'une variable par rapport à une tribu

Nous savons que $E(Y/X)$ ne dépend que de $\sigma(X)$, aussi peut on noter $E(Y/X)$ par $E(Y/\sigma(X))$; d'ou :

Définition 2.12 *:*
Si Y *est une v.a.r intégrable sur* $(\Omega, \mathcal{A}, P)$ *est* $\mathcal{D}$ *une sous tribu de* $\mathcal{A}$ *alors* $E(Y/\mathcal{D})$ *est toute v.a.r* $\mathcal{D}$ *mesurable verifiant :*

$$\forall D \in \mathcal{D} \int_D YdP = \int_D E(Y/\mathcal{D})dP.$$

Remarque 2.13 *:*
- Deux versions de $E(Y/\mathcal{D})$ *sont égales* P *p.s.*

- *$E(Y/\mathcal{D})$ n'est autre que la dérivée de Radon Nykodim de la mesure $\left(\hat{Q}(D) = \int_D Y dP\right)$ par rapport à la restriction de P sur $\mathcal{D}$*
- *Si $\mathcal{D} = \{\Phi, \Omega\}$ alors $E(Y/\mathcal{D}) = EY$ P p.s.*
- *$E(Y/\mathcal{D})$ se définit pour Y positive même non intégrable(voir le probléme à la fin de ce chapitre).*
- *$E(Y/\mathcal{D})$ est aussi noté $E^{\mathcal{D}}Y$.*

2.3 Propriétés de l'espérance conditionnelle

Il suffit de donner ces propriétés pour $E(Y/\mathcal{D})$ car $E(Y/X) = E(Y/\sigma(X))$:

a) Si $Y \geq 0$ alors $E^{\mathcal{D}}Y \geq 0$ P $p.s$;

b) L'application :

$$\begin{array}{rcl} L^1(\Omega, \mathcal{A}, P) & \longrightarrow & L^1(\Omega, \mathcal{D}, P) \\ Y & \longrightarrow & E^{\mathcal{D}}Y \end{array}$$

est linéaire ;

c) $E(1/\mathcal{D}) = 1$ P $p.s$;

d) Si $E(Y) \subset \mathcal{D}$ alors $E(Y/\mathcal{D}) = Y$ P $p.s.$

e) Si $\sigma(Y)$ et $\mathcal{D}$ sont indépendantes alors $E(Y/\mathcal{D}) = E(Y)$ P $p.s$; même démonstration que la proposition2.

f) $E(E(Y/\mathcal{D})) = E(Y)$ en effet :

$$\begin{aligned} E(Y) &= \int Y dP \\ &= \int E(Y/\mathcal{D}) dP \\ &= E(E(Y/\mathcal{D})) \end{aligned}$$

g) Si $\mathcal{D}_1 \subset \mathcal{D}_2$ alors :

$$\begin{aligned} E(E(Y/\mathcal{D}_2)/\mathcal{D}_1) &= E(E(Y/\mathcal{D}_1)/\mathcal{D}_2) \\ &= E(Y/\mathcal{D}_1). \end{aligned}$$

Preuve. :
$E(E(Y/\mathcal{D}_1)/\mathcal{D}_2) = E(Y/\mathcal{D}_1)$ car $E(Y/\mathcal{D}_1)$ étant $\mathcal{D}_1$ mesurable est $\mathcal{D}_2$ mesurable.
Reste à voir que $E(E(Y/\mathcal{D}_2)/\mathcal{D}_1) = E(Y/\mathcal{D}_1)$.

Soit $D \in \mathcal{D}_1$, $\int_D E(E(Y/\mathcal{D}_2)/\mathcal{D}_1)dP = \int_D E(Y/\mathcal{D}_2)dP = \int_D Y dP$ car $D \in \mathcal{D}_2(\mathcal{D}_{1\subset}\mathcal{D}_2)$
Donc :

$$\left.\begin{array}{c} \forall D \in \mathcal{D}_1,\ \int E(E(Y/\mathcal{D}_2)/\mathcal{D}_1)dP = \int Y dP \\ \text{et} \quad E(E(Y/\mathcal{D}_2)/\mathcal{D}_1) \text{ est } \mathcal{D}_1 \text{ mesurable} \end{array}\right\} \Longrightarrow E(E(Y/\mathcal{D}_2)/\mathcal{D}_1) = E(Y/\mathcal{D}_1).$$

■

h) théorème converergence monotone conditionnel (T.C.M.C.) :

Soit (Y_n) une suit croissante de v.a.r positives avec :

$$Y_n \nearrow Y \ P \ p.s \text{ et } (E \mid Y \mid < \infty).$$

Alors :

$$\lim_n E(Y_n/\mathcal{D}) = E(Y/\mathcal{D}).$$

Preuve. :
Posons $(Z_n = Y - Y_n)$, on a : $(Z_n \searrow 0)$ P p.s et $(EZ_n \searrow 0)$.
$(Z_n \geq Z_{n+1}) \Longrightarrow (E(Z_n/\mathcal{D}) \geq E(Z_{n+1}/\mathcal{D}) \geq 0)$.(D'aprés les propriétés a et b)
Donc :

$$E(Z_n/\mathcal{D}) \searrow U \quad (U \geq 0).$$

Par ailleurs,d'aprés le T.C.D. on a :

$$\lim_n E(Z_n/\mathcal{D}) = U \quad (U \geq 0)$$

$$\begin{aligned} E(U) &= \int_\Omega U \, dP \\ &= \int_\Omega \lim_n E(Z_n/\mathcal{D})dP \\ &= \lim_n \int_\Omega E(Z_n/\mathcal{D})dP \qquad \text{(T.C.M)} \\ &= \lim_n \int_\Omega Z_n dP \\ &= \int_\Omega \lim_n Z_n dP \qquad \text{(T.C.M)} \\ &= \int_\Omega 0 dP = 0 \end{aligned}$$

$(E(U) = 0)$ avec $(U \geq 0) \Longrightarrow (U = 0)$. ■

i) Théorème de convergence conditionnel (T.C.D.C) :

Si (Y_n) est une suite de v.a.r. Z convergeant P $p.s$ vers Y et s'il existe une v.a.r Z intégrable telle que :$\mid Y_n \mid \leq Z$ P $p.s$ $\forall n \in \mathbb{N}$ Alors :

$$E(Y/\mathcal{D}) = \lim_n E(Y_n/\mathcal{D}).$$

Preuve. :

Posons :

$$X_n = \sup_{k \geq n} \mid Y_k - Y \mid, \text{ on a} : X_n \to 0 \quad P \ p.s.$$

Et

$$\mid E(Y_n/\mathcal{D}) - E(Y/\mathcal{D}) \mid = \mid E(Y_n - Y/\mathcal{D}) \mid \leq \mid E(\mid Y_n - Y \mid /\mathcal{D}) \mid \leq E(X_n/\mathcal{D})$$

Il suffity donc de montrer que $(E(X_n/\mathcal{D}) \to 0)$ P $p.s.$

La suite (X_n) étant décroissante et positive, il en esty de même de $E(X_n/\mathcal{D})$, appelons X sa limite puisqu'elle existe, on a :

$$E(X) = \int \lim E(X_n/\mathcal{D}) \overset{T.C.D}{=} \lim \int E(X_n/\mathcal{D})dP$$

Car :

$$E(X_n/\mathcal{D}) \leq 2\ E(Z/\mathcal{D}) \text{ intégrable.}$$

Or :

$$\lim_n \int E(X_n/\mathcal{D})dP = \lim_n \int X_n dP \overset{T.C.D}{=} \int \lim X_n dP = 0.$$

Donc :

$$\text{et} \begin{array}{c} X \geq 0 \\ EX = 0 \end{array} \Bigg\} \Longrightarrow X = 0 \quad P\ p.s.\text{(c.q.f.d.)}$$

■

j) Si Y est $\mathcal{D}$ mesurable alors $(E(YZ/\mathcal{D}) = YE(Z/\mathcal{D}))$.

En effet :

- Si

$$(Y = \alpha\ 1_{D_o}) \text{ où } (D_o \in \mathcal{D}) \text{ alors}$$

Soit :

$$D \in \mathcal{D}, \int_D YZdP = \int_{D \cap D_o} \alpha Z dP = \alpha \int_{D \cap D_o} E(Z/\mathcal{D})dP = \int_D Y\ E(Z/\mathcal{D})dP.$$

- Si Y est une v.a.r. étagée, la propriété de linéarité de l'espérance conditionnelle (propriété b) permet de conclure.

- Si Y est une $(v.a.r. \geq 0)$, il suffit d'appliquer la proprieté b et le $T.C.M.C$ pour avoir le résultat.
- Enfin si Y est qulconque, l'application de la propriété b permet de conclure. $(c.q.f.d.)$.

k) Inégalité de Jensen :

Soit Y une v.a.r. intégrable sur $(\Omega, \mathcal{A}, P)$ à valeurs dans un intervalle I de $\mathbb{R}$ et soit $\mathcal{C}$ une fonction $\mathcal{B}(I)$ mesurable convexe sur I avec $\mathcal{C}(Y)$ intégrable. nous avons :

$$E^X(\mathcal{C}(Y)) \geq \mathcal{C}(E^X Y) \quad P \ p.s.$$

Preuve. :

$$(\Omega, \mathcal{A}, P) \underset{\rightarrow}{Y} (I, \mathcal{B}(I)) \underset{\rightarrow}{\mathcal{C}} (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

Commençons par remarquer que si $(Y \in I)$ alors $\left(E^X Y \in I\right)$, en effet
Si $(Y > a)$ alors $\left(E^X Y - a = E^X(Y - a) > 0\right)$.
Par ailleurs $\mathcal{C}$ étant convexe on peut écrire :

$$\forall x, \zeta \in I \geq \mathcal{C}(x) - \mathcal{C}(\zeta) \geq \mathcal{C}'_d(\zeta)(x - \zeta)$$

(où $\mathcal{C}_d$ est la dérivé à droitede $\mathcal{C}$).

Posons :

$$x = Y(w), \ \mathcal{C} = E^X Y(w).$$

On obtient :

$$\mathcal{C}(Y(w)) - \mathcal{C}(E^X Y(w)) \geq \mathcal{C}'_d(E^X Y(w))(Y(w) - E^X Y(w)) \forall w \in \Omega$$

$$\Longrightarrow \mathcal{C}(Y) - \mathcal{C}'_d(E^X Y)(Y - E^X Y) \geq \mathcal{C}(E^X Y).$$

$$\Longrightarrow E^X(\mathcal{C}(Y) - \mathcal{C}'_d(E^X Y)(Y - E^X Y) \geq E^X(\mathcal{C}(E^X Y)) = \mathcal{C}(E^X Y)$$

(d'aprés les propriétés a,b et d).

Or :

$$E^X(\mathcal{C}'_d(E^X Y)(Y - E^X Y)) = \mathcal{C}'_d(E^X Y)(E^X(Y - E^X Y)) = 0 \text{ (propriété j)}.$$

D'ou :

$$E^X(\mathcal{C}(Y)) \geq \mathcal{C}(E^X Y). \ (c.q.f.d.).$$

■

Corollaire 2.14 *:*
Si $\mathcal{C}$ et Y vérifient les hypothéses de l'inégalité de Jensen alors :
- $E^{\mathcal{D}}(\mathcal{C}(Y)) \geq \mathcal{C}(E^{\mathcal{D}}(Y))$ $\forall \mathcal{D}$ *sous tribu de* $\mathcal{A}$.
- $E(\mathcal{C}(Y)) \geq \mathcal{C}(E(Y))$.

2.4 L'espérance conditionnelle en tant que projecteur :

Soient X et Y deux v.a.r de carré intégrable sur $(\Omega, \mathcal{A}, P)$ et $\mathcal{D}$ une sous tribu de $\mathcal{A}$. La covariance est un produit scalaire sur l'espace vectoriel des v.a centrées et de carrrée intégrables.

1-Probléme des moindres carrés :

On cherche une v.a. Z $\sigma(X)$ mesurable $(Z = \mathcal{C}(X))$ telle que Y et Z soient les plus proches possible dans L^2.

a) Si $(X = Cste)$ on cherche $Z = c$ $(Cste)$ qui minimise $E(Y - c)^2$ donc $(Z = EY)$.

b) Soit Z une v.a $\sigma(X)$ mesurable

Posons :

$$T = E(Y/\sigma(X)) - Z.$$

T est $\sigma(X)$ mesurable

$$E(Y - Z)^2 = E(Y - E(Y/\sigma(X)))^2 + ET^2.$$

Car :

$$\begin{aligned} E(T(Y - E(Y/\sigma(X)))) &= EE((T(Y - E(Y/\sigma(X))/\sigma(X)) \\ &= ET(E(Y - E(Y/\sigma(X)/\sigma(X))) \\ &= ET(E(Y/\sigma(X) - E(Y/\sigma(X)) = 0. \end{aligned}$$

On a donc :

$E(Y - Z)^2$ est minimum pour $(ET^2 = 0)$ c.a.d pour $(T^2 = 0 \quad P\ p.s)$.

En conclusion :

Parmi les v.a $\sigma(X)$ mesurables, $(Z = E(Y/X))$ est celle qui est la plus proche de Y au sens de L^2.

2-Cas général :

Soit $P_{\mathcal{D}}$ la restriction de P à $\mathcal{D}$, on est tenté de considérer $E^{\mathcal{D}}$ comme le projecteur orthogonal de $L^2(\Omega, \mathcal{A}, P) \longrightarrow L^2(\Omega, \mathcal{D}, P_{\mathcal{D}})$.

L'ennui est que $L^2(\Omega, \mathcal{D}, P_{\mathcal{D}})$ n'est pas exactement un sous espace de $L^2(\Omega, \mathcal{A}, P)$ car les relations d'équivalence mod P et mode $P_{\mathcal{D}}$ ne sont pas identiques.Une classe d'équivalence dans $L^2(\Omega, \mathcal{D}, P_{\mathcal{D}})$ n'est pas forcément une classe de $L^2(\Omega, \mathcal{A}, P)$,d'autres v.a. non $\mathcal{D}$ mesurables peuvent apparaitre.

Mais fort heureusement il existe une isométrie de $L^2(\Omega, \mathcal{D}, P_{\mathcal{D}})$ donnée par :

$$\begin{array}{ccc} X \in L^2(\Omega, \mathcal{D}, P_{\mathcal{D}}) & \longrightarrow & X \in L^2(\Omega, \mathcal{A}, P) \\ \uparrow & & \uparrow \\ \text{Classe de } X & & \text{classe de } X \\ \text{Dans } L^2(\Omega, \mathcal{D}, P_{\mathcal{D}}) & & \text{dans } L^2(\Omega, \mathcal{A}, P) \end{array}$$

En effet :

$$\| X \|_{\mathcal{D}} = \int_{\Omega} X^2 dP_{\mathcal{D}} = \int_{\Omega} X^2 dP = \| X \|_{\mathcal{A}}$$

L'image de $L^2(\Omega, \mathcal{D}, P_{\mathcal{D}})$ par cette isométrie est un sous espace de $L^2(\Omega, \mathcal{D}, P)$.

Aussi désormais identifiera-t-on $L^2(\Omega, \mathcal{D}, P_{\mathcal{D}})$ et son image par cette isométrie.Autrement dit $L^2(\Omega, \mathcal{D}, P_{\mathcal{D}})$ est considéré comme l'espace des éléments de $L^2(\Omega, \mathcal{D}, P)$ dont au moins un représentant est $\mathcal{D}$ mesurable.

Proposition 2.15 *:*

$E^{\mathcal{D}}$ est le projecteur orthogonal de $L^2(\Omega, \mathcal{A}, P)$, notons $P_r^{\mathcal{D}} Y$ la projection de Y sur $L^2(\Omega, \mathcal{D}, P_{\mathcal{D}})$.

Preuve. :

Soit $Y \in L^2(\Omega, \mathcal{A}, P)$, notons $P_r^{\mathcal{D}} Y$ la projection de Y sur $L^2(\Omega, \mathcal{D}, P_{\mathcal{D}})$.

On a :

$$\begin{aligned} \int_D P_r^{\mathcal{D}} Y dP &= \int 1_D P_r^{\mathcal{D}} Y \\ &= E(1_D P_r^{\mathcal{D}} Y) + E(1_D (Y - P_r^{\mathcal{D}} Y)). \end{aligned}$$

Car 1_D est $Y - P_r^{\mathcal{D}} Y$ sont orthogonaux.

Donc :

$$\int_D P_r^{\mathcal{D}} Y dP = E(1_D Y) = \int_D Y dP \qquad \forall D \in \mathcal{D}$$

D'ou :

$$P_r^{\mathcal{D}} Y = E^{\mathcal{D}} Y \qquad \text{(c.q.f.d).}$$

■

2.5 Version régulière de la probabilité conditionnelle :

Si $(Y = 1_A)$ on a :($E(Y/\mathcal{D})$ =$E(1_A/\mathcal{D})$) est la classe de v.a.r. $\mathcal{D}$ mesurable telle que :

$$\forall D \in \mathcal{D} \int_D E(1_A/\mathcal{D})dP = \int_D 1_A dP = P(A \cap D).$$

Si $\mathcal{D}$ =$\sigma(X)$ on retrouve $P(A/X) = E(1_A/X)$, d'où :

Définition 2.16 *:*
Si $\mathcal{D}$ est une sous tribu de $\mathcal{A}$ et $A \in \mathcal{A}$, la probabilité conditionnelle de A sachant $\mathcal{D}$ est toute v.a.
$\mathcal{D}$ mesurable notée $P(A/\mathcal{D})$ qui vérifie :

$$\forall D \in \mathcal{D} \int_D P(A/\mathcal{D})dP = P(A \cap D).$$

En fait :

$$P(A/\mathcal{D}) = E(1_A/\mathcal{D}).$$

On remarque que :
Pour A fixé, $P(A/\mathcal{D})$ est un classe de v.a $\mathcal{D}$ mesurables.
Si A décrit $\mathcal{A}$ et $\mathcal{D}$ est fixée, $P(./\mathcal{D})$ se comporte presque comme une probabilité,en effet :
i) $(P(A/\mathcal{D}) \geq 0)$ P p.s.
ii) $(P(\Omega, \mathcal{D}) = 1)$ P p.s.
iii) $P(\sum\limits_{i\in\mathbb{N}} A_i/\mathcal{D}) = \sum\limits_{i\in\mathbb{N}} P(A_i/\mathcal{D})$ P p.s.
Mais $A \longrightarrow P(A_i/\mathcal{D})$ n'est pas nécessairement une probabilité p.s car en changeant de suite (A_n).
l'evénement négligeable en dehors du quel iii) est vérifiée varie et il n'y a pas de raison pour qu'il existe un négligeable universel valable pour toutes les suites (A_n), c'est pourquoi on ne peut, en général, extraire pour une classe $P(A/\mathcal{D})$ un représentant $P^*(A/\mathcal{D})$ tel que l'application :

$$\begin{array}{ccc} \Omega \times \mathcal{A} & \longrightarrow & [0,1] \\ (w, A) & \longrightarrow & P^*(A/\mathcal{D})(w) \end{array}$$

Soit, pour W fixée, une probabilité sur $\mathcal{A}$.cela nous conduit à :

Définition 2.17 *:*
$P^*(A/\mathcal{D})$ est une version réguliére de la probabilité conditionnelle (notée v.r.p.c.) de A sachant $\mathcal{D}$ si :
a) $\forall A \in \mathcal{A}$ fixé, $P^*(A/\mathcal{D})$ est une version de $P(A/\mathcal{D})$;
b) $\forall w \in \Omega$ fixé, $P^*(./\mathcal{D})\ (w)$ est une probabilité sur $\mathcal{A}$.

Proposition 2.18 *:*
Si une v.r.p.c. de A sachant $\mathcal{D}$ existe et si $(E \mid Y \mid < \infty)$ alors :

$$E(Y/\mathcal{D}) = \int Y P^*(dy/\mathcal{D}) \qquad P\ p.s.$$

Preuve. :
- Si $(Y = 1_A)$ $A \in \mathcal{A}$ on a :

$$E(Y/\mathcal{D}) = P(A/\mathcal{D})$$

Et

$$\begin{aligned} \int_\Omega Y P^*(dy/\mathcal{D}) &= \int_A P^*(dy/\mathcal{D}) \\ &= P^*(A/\mathcal{D}) \\ &= P(A/\mathcal{D}) \qquad P\ p.s. \\ &\quad \text{Car } P^*(A/\mathcal{D}) \text{ est une version de } P(A/\mathcal{D}). \\ \Longrightarrow &\int Y P^*(dy/\mathcal{D}) = E(Y/\mathcal{D}) \qquad P\ p.s. \end{aligned}$$

- Si Y est étagée le résultat reste vrai par linéarité de l'espérance conditionnelle et de l'espérance.
- Si Y est positive le résultat se généralise par application du $T.C.M.C.$ et du $T.C.M.$
- Enfin si Y est intégrable, le résultat découle de la linéarité de l'espérance conditionnelle et de l'espérance $(c.q.f.d.)$.
En général la $v.r.p.c.$ n'existe pas mais le probléme se simplifie lorsqu'on passe aux espaces image.
Supposons que Y prenne ses valeurs dans $(E, \mathcal{B})$. ■

Définition 2.19 *:*
$\hat{P}(B/\mathcal{D})$ définie sur $\mathcal{B}\times\Omega$ est une version réguliére de la loi conditionnelle (notée v.r.l.c.) de Y sachant $\mathcal{D}$ si :
i) $\forall B \in \mathcal{B}$ fixé, $\hat{P}(B/\mathcal{D})$ est une version de $P(Y \in B/\mathcal{D})$;
ii) $\forall w$ fixé $(w \in \Omega)$, $\hat{P}(./\mathcal{D})(w)$ est une probabilité sur $\mathcal{B}$.

Proposition 2.20 *:*
Soit :

$$\begin{aligned} Y \ v.a. \ &: \ (\Omega, \mathcal{A}, P) \longrightarrow (E, \mathcal{B}) \\ \mathcal{C} \ v.a. \ &: \ (E, \mathcal{B}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})) \end{aligned}$$

Avec $(E \mid \mathcal{C}(Y) \mid < \infty)$.
Si $\hat{P}(B/\mathcal{D})$ *est une version réguliére de la loi conditionnelle de* Y *sachant* $\mathcal{D}$ *alors :*

$$E(\mathcal{C}(Y)/\mathcal{D}) = \int \mathcal{C}(Y) \ \hat{P}(dY/\mathcal{D}) \quad P \ p.s.$$

Preuve. :
Provient du fait que la *v.r.l.c.* se comporte comme un *v.r.p.c.* (*d.c.q.f.d.*). ■

Définition 2.21 *:*
Soit X *une v.a :*$(\Omega, \mathcal{A}, P) \longrightarrow (S, \mathcal{C})$
On définit une v.r.l.c. de Y *sachant* $(X = x)$*,notée* $\hat{P}(B/X = x)$ *par :*
i) $\forall x \in S$, $\hat{P}(./X = x)$ *est une probabilité sur* $\mathcal{B}$.
ii) $\forall B \in \mathcal{B}$, $\hat{P}(B/X = x)$ *est une version de* $P(Y \in B/X = x)$.

Proposition 2.22 *:*
Soit $\mathcal{C}(x, y)$ *une v.a sur* $(S, \mathcal{C}) \otimes (E, \mathcal{B})$ *telle que* $(E \mid \mathcal{C}(x, y) \mid < \infty)$.
Si $\hat{P}(B/X = x)$ *est une v.r.l.c. de* Y *sachant* $(X = x)$ *alors :*

$$E(\mathcal{C}(x, y)/X = x) = \int \mathcal{C}(x, y)\hat{P}(dY/X = x) \quad p.s.$$

Preuve. :
Si :

$$\mathcal{C}(x, y) = 1_C(x)1_D(y),$$

alors :

$$E(\mathcal{C}(x, y)/X = x) = 1_C(x)E(1_D(y)/X = x) \quad \text{(D'aprée la propriété j)}$$

$$\Longrightarrow E(\mathcal{C}(x, y)/X = x) = 1_C(x)P(Y \in \mathcal{D}/X = x) = 1_C(x) \int 1_D(y)\hat{P}(dY/X = x)$$

$$= \int \mathcal{C}(x, y)\hat{P}(dY/X = x).$$

La proposition reste vraie pour $\mathcal{C}$ intégrable par linéarité et passage à la limite monotone. (*c.q.f.d.*). ■

Corollaire 2.23 *:*
Si X et Y sont des v.a. indépendantes et sous les mêmes conditions que la proposition précédente on a :

$$E(\mathcal{C}(X,Y)/X=x)=E(\mathcal{C}(x,Y)) \quad p.s$$

Preuve. :
Il suffit de prendre $\hat{P}(B/X=x)=P(Y\in B)$ et d'appliquer la proposition précédente.$(c.q.f.d.)$.
Dans le cas où Y est une v.a réelle, nous allons montrer l'existence d'une $v.r.l.c.$ de Y sachant $\mathcal{D}$ pour cela nous avons recours aux versions réguliéres de la fonction de répartition conditionnelle dont voici la définition : ∎

Définition 2.24 *:*
$P(x/\mathcal{D})$ définie sur $\mathbb{R}\times\Omega$ est une version réguliére de la fonction de répartition conditionnelle (notée v.r.f.r.c.) de Y sachant $\mathcal{D}$ si :
i) $\forall x$ fixée, $F(x/\mathcal{D})$ est une version de $P(Y<x/\mathcal{D})$;
ii) $\forall w$ fixée, $F(./\mathcal{D})(w)$ est une fonction de répartition.

Proposition 2.25 *:*
Si Y est une v.a.r., il existe une v.r.f.r.c. de Y sachant $\mathcal{D}$.

Preuve. :
Choisissons une version de $P(Y<x/\mathcal{D})$.
Soit x un nombre rationnel.
$\mathcal{Q}=\{x_i\}_{i\in\mathbb{N}}$ l'ensemble des rationnels.
- Soit :

$$\begin{aligned} M &= \underset{r_j>r_i}{\cup}\{w/P(Y<r_j/\mathcal{D})<P(Y<r_i/\mathcal{D})\} \\ &= \underset{r_j>r_i}{\cup}\left\{w/E(1_{Y<r_j}-1_{Y<r_i}/\mathcal{D})\right\} \end{aligned}$$

(Ensemble où $P(Y<x/\mathcal{D})$ n'est pas croissante dans $\mathcal{Q}$).

Or si $(r_j>r_i)$ alors :

$$\begin{aligned} \left((1_{Y<r_j}-1_{Y<r_i})\geq 0\right) &\Longrightarrow \left(E(1_{Y<r_j}-1_{Y<r_i}/\mathcal{D})\geq 0 \quad p.s\right) \\ &\Longrightarrow P(w:E(1_{Y<r_j}-1_{Y<r_i}/\mathcal{D})<0)=0. \end{aligned}$$

Donc $P(M)=0$ (comme réunion dénombrable d'ensembles négligeables)
- Soit :

$$N=\underset{i}{\cup}\left\{w/\lim_{r_j\nearrow r_i}P(Y<r_j/\mathcal{D})\neq P(Y<r_i/\mathcal{D})\right\}$$

(Ensemble où $P(Y < x/\mathcal{D})$ n'est pas continue à gauche)

On a :

$$\lim_{y \nearrow r_i} P(Y < r_j/\mathcal{D}) = P(\lim(Y < r_j/\mathcal{D}) = P(Y < r_i) \quad p.s$$

Donc : $(P(N) = 0)$.
De plus :
Si $(r \nearrow +\infty)$ alors $(P(Y < r/\mathcal{D}) \longrightarrow 1 \;\; p.s)$.
Si $(r \searrow -\infty)$ alors $(P(Y < r/\mathcal{D}) \longrightarrow 0 \;\; p.s)$.
Si on note L l'ensemble pour lequel $\left(P(Y < r/\mathcal{D}) \underset{r \searrow -\infty}{\nrightarrow} 0\right)$ .et $\left(P(Y < r/\mathcal{D}) \underset{r \nearrow +\infty}{\nrightarrow} 1\right)$ alors $(P(L) = 0)$.
Donc $\forall w \notin M \cup N \cup L$ et pour x rationnel on a :$(x \longrightarrow P(Y < x/\mathcal{D}))$ est monotone,continue à gauche, tend vers 1 si $(x \longrightarrow +\infty)$ et tend vers 0 si $(x \longrightarrow -\infty)$.
Si G est une fonction de répartition quelconque, posons :

$$F_Y(x/\mathcal{D}) = \begin{cases} G(x) & \text{si } w \in N \cup L \cup M \\ \lim\limits_{r_i \nearrow x} P(Y < r_j & \text{si } w_j \notin N \cup L \cup M \end{cases}$$

Pour tout w fixé, $F_Y(x/\mathcal{D})$ est une fonction de répartition, et c'est bien une version de $P(Y < x/\mathcal{D})$ pour x fixé. ■

Corollaire 2.26 *:*
Si Y est une v.a. réelle, il existe une v.r.l.c. de Y sachant $\mathcal{D}$.

Preuve. :
D'aprés la proposition 7, il existe $F_Y(x/\mathcal{D})$ qui est une v.r.f.r.c.
Pour chaque w, $F_Y(x/\mathcal{D})$ définit une probabilité $P_{Y/\mathcal{D}}(B)$ qui coincide avec $\hat{P}(B/\mathcal{D})$ pour $B =]-\infty, x[\; \forall x \in \mathbb{R}$ donc partout. ■

Remarque 2.27 *:*
Le cas où Y prend ses valeurs dans $\mathbb{R}^n$ est analogue.
Plus généralement que le corollaire précèdent,il existe une v.r.l.c. de Y sachant $\mathcal{D}$ pour tout vecteur Y à valeurs dans $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

Proposition 2.28 *:*
Soit $X = (X_1, ..., X_P)$ et $Y = (X_{p+1}, X_{p+2}, ..., X_n)$ deux $\overrightarrow{v.a}$ à valeurs respectivement dans $\mathbb{R}^p$ et $\mathbb{R}^{n-p}$.
Si le vecteur $(X_1, ..., X_n)$ admet une densité $f(x_1, ..., x_n)$ par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$ alors Y admet une loi conditionnelle par rapport à X qui pour presque toute

valeur $x_1, x_2, ..., x_p$ de X a la densité :

$$f_Y^{x_1...x_p} = (x_{p+1}, ..., x_n) = \frac{f(x_1...x_n)}{f_X(x_1...x_p)}$$

où f_X est la densité marginale de X.

Preuve. :
Soit $B \in \mathcal{B}(\mathbb{R}^p)$ et $B_1 \in \mathcal{B}(\mathbb{R}^{n-p})$ on a :

$$\begin{aligned} P((X &\in B) \cap (Y \in B_1)) = P((x_1...x_n) \in B \times B_1) \\ &= \int_{B\times B_1} f(x_1...x_n) dx_1..dx_n \\ &= \int_B dx_1..dx_p \int_{B_1} f(x_1...x_p) dx_{p+1}..dx_n \\ &= \int_B f_x(x_1...x_p) dx_1...dx_p \int_{B_1} \frac{f(x_1...x_n)}{f_X(x_1...x_p)} dx_{p+1}..dx_n \end{aligned}$$

$$\text{car : } f_X(x_1...x_p) \neq 0 \quad P_X \ p.s.$$

D'où :

$$P((X \in B) \cap (Y \in B_1)) = \int_B dP_X(x_1...x_p) \int_{B_1} \frac{f(x_1...x_n)}{f_X(x_1...x_p)} dx_{p+1}..dx_n.$$

Or :

$$P((X \in B) \cap (Y \in B_1)) = \int P(Y \in B_1/X = (x_1...x_p)) dP_X(x_1...x_p)$$

D'où :

$$P(Y \in B_1/X = (x_1...x_p)) = \int_{B_1} \frac{f(x_1...x_n)}{f_X(x_1...x_p)} dx_{p+1}..dx_n.$$

qui est mesure de probabilité sur $(\mathbb{R}^{n-p}, \mathcal{B}(\mathbb{R}^{n-p}))$ de densité par rapport à la mesure de Lenesgue sur $\mathbb{R}^n$: $\frac{f(x_1...x_n)}{f_X(x_1...x_p)}$. ■

Exemple 2.29 *:*
Si X et Y sont deux v.a.r telles que (X, Y) a pour densité f alors :

$$E(Y/X \ = \ x) = \int y \frac{f(x,y)}{f(x,y)dy} dy \qquad \textit{(D'après les propositions 5 et 8).}$$

(Y est supposé intégrable).

Bibliographie

[1] J.Neuveu :Bases mathématiques du calcule des probabilités. Manou (1970).

[2] Khoan Vokhac :Théorie des probabilités. Ellipses (1984).

[3] J.Pierre Lecoutre :Statistique et probabilités. Dunod Paris (2009).

[4] F.Messaci,E.Belaloui :Notion fondamentales probabilités.Université mentouri constantine (2000/2001)