II : VARIABLES ALEATOIRES

II.1- Applications mesurables

1.Eléments aléatoires

Définition 1

On appelle élément aléatoire une application $X:(\Omega_1,\tau_1,P_1)\to(\Omega_2,\tau_2)$ d'un espace probabilisé vers un espace probabilisable telle que $\forall E\in\tau_2$ on a $X^{-1}(\tau_2)\in\tau_1$, autrement dit une *application mesurable* relativement à une paire d'espaces probabilisables $(\Omega_1,\tau_1,P),(\Omega_2,\tau_2)$, le premier étant muni d'une mesure de probabilité.

Propriété 1

L'espace (Ω_2,τ_2) peut alors également être muni d'une mesure de probabilité P_2, dite *mesure image* (ou encore *loi image*) *de* P_1 *par X* et qui est définie par :

$\forall E\in\tau_2\ P_2(E)=P_1(X^{-1}(E))$

Propriété 2

La famille d'événements $\{X^{-1}(E),E\in\tau_2\}\subset\tau_1$ est une tribu sur Ω_1, notée τ_X. Elle est appelée tribu engendrée par X sur Ω.

2. Variables aléatoires

Définition 2

Une VA (variable aléatoire) N dimensionnelle X sur un espace probabilisé (Ω,τ,P) est une application :

$X:\omega\in\Omega\to X(\omega)\in R^N$

mesurable relativement à (Ω,τ) et (R^N,B^N).

Pour vérifier si X est mesurable, il est possible d'utiliser les critères donnés ci-dessous (le second étant le plus simple).

Propriété 3 (critères de mesurabilité) :

- X est mesurable ssi pour tout pavé N-dimensionnel $E\in R^N$ on a $X^{-1}(E)=\{\omega:X(\omega)\in E\}\in\tau$.
- $X:\omega\in\Omega\to X(\omega)\in R^N$ est mesurable ssi

$\forall(x_1,...,x_N)\in R^N, X^{-1}(]-\infty,x_1[\times...\times]-\infty,x_N[)\in\tau$

Variables marginales

A une application X donnée sont associées les N applications $X_i:\Omega\to R$ qui correspondent, chacune , à ce qu'on appelle une VA *monodimensionnelle* :

$\omega\to X(\omega)=(X_1(\omega),...,X_N(\omega))$

Les X_i sont évidemment toutes déterminées lorsque X est donnée. Plus généralement, pour tout entier p tel que $1\le p<N$ et à p fixé pour tout ensemble S de p indices distincts $1\le i_1,...,i_p\le N$, $1\le p<N$, il est possible d'introduire une VA p-dimensionnelle :

$$\omega\to(X_{i_1}(\omega),...,X_{i_p}(\omega))\in R^p$$

Définition 3

- On appelle VA indicatrice d'un événement E dans τ l'application $\omega\to X(\omega)=1$ si $\omega\in E$ et $=0$ sinon. Elle est notée 1_E.

II : VARIABLES ALEATOIRES

- Une généralisation est d'introduire une partition finie $\Omega = \sum_{i=1..N} E_i, E_i \in \tau$ et N réels distincts $x_i, i = 1..N$ pour construire la VA notée $Y = \sum_{i=1..N} x_i 1_{E_i}$ et définie par $Y(\omega) = \sum_{i=1..N} x_i 1_{E_i}(\omega), \omega \in \Omega$. Une telle VA sera appelée VA simple.

Propriété 4

Toute VA $X \geq 0$ sur (Ω, τ, P) (cad qui ne prend sur Ω que des valeurs ≥ 0) peut s'écrire comme limite simple sur Ω d'une suite croissante de VA simples $Y_k, k \geq 1$ positives ou nulles.

II.2- Loi de probabilité d'une V.A.

2.1) Définition 4

La loi de probabilité d'une V.A $X : (\Omega, \tau, P) \to (R^N, B^N)$ est la loi image, notée P_X, de P par X.

VA presque sûrement égales

- Deux V.A. X_1 et X_2 sur le même espace probabilisé et à valeurs toute les deux dans (R^N, B^N) seront dites presque sûrement égales ou encore *égales presque partout* si $P(X_1 = X_2) = P(\{\omega / X_1(\omega) = X_2(\omega)\} = 1$.
- L'égalité presque sure entre VA est une relation d'équivalence.
- Si X_1 et X_2 sont équivalentes alors $P_{X_1} = P_{X_2}$ sur B^N mais la réciproque est fausse : 2 V.A. peuvent avoir la même loi et avoir leurs réalisations distinctes avec une probabilité strictement positive (et même avec probabilité égale à 1). L'exemple qui suit illustre cette possibilité :
  - $\Omega = \sum_{i=1..N} E_i, E_i \in \tau$, $P(E_i) = \dfrac{1}{N}, i = 1,..,N$,
  - $x_1,..,x_N$ N réels distincts, $y_1,..,y_N = x_{i_1},..,x_{i_N}$ où $(i_1,..,i_N) = \sigma((1,..,N))$ est une permutation de la suite $(1,..,N)$,
  - $X(\omega) = \sum_{i=1,..,N} x_i 1_{E_i}(\omega), \omega \in \Omega$ et $Y(\omega) = \sum_{i=1,..,N} y_i 1_{E_i}(\omega), \omega \in \Omega$
  - les VA X et Y ont la même loi :
  $$\forall E \in B : P_X(E) = \frac{1}{N} card((x_1,..,x_N) \cap E) = \frac{1}{N} card((y_1,..,y_N) \cap E) = P_Y(E)$$
  - si par ex $\sigma(.)$ correspond à une permutation circulaire et donc que $\forall i = 1,..,N : y_i \neq x_i$ on a $P(X = Y) = 0$!..

2.2) Fonction de répartition

Définition 5

La fonction de répartition d'une V.A. $X : \Omega \to R^N$ est l'application[1] :

$$F_X : R^N \to [0,1]$$
$$x \to F_X(x) = P(X < x) = P_X(]-\infty, x_1[\times]-\infty, x_2[\times \ldots \times]-\infty, x_N[)$$

[1] Une autre définition est également utilisée dans la littérature : $F_X(x) = P(X \leq x), x \in R^N$

où $x_1,...,x_N$ sont les N coordonnées de x . $F_X(x)$ et $P(X < x)$ pourront respectivement être notées $F_{X_1,...,X_N}(x_1,...,x_N)$ et $P(X_1 < x_1,...,X_N < x_N)$ pour bien distinguer les N coordonnées aléatoires.
La fonction de répartition d'une V.A. ne dépend de cette dernière que par l'intermédiaire de sa loi de probabilité. Deux V.A. distinctes pouvant avoir la même loi elles peuvent donc avoir la même fonction de répartition. Comme précisé dans les 2 prochains théorèmes *la relation entre fonction de répartition et mesure de probabilité est biunivoque.*

Théorème 1
La donnée de F_X est équivalente à la donnée de P_X , c.a.d. :
a) P_X détermine F_X (ce qui est évident d'après la définition),
b) F_X détermine P_X , (ce qui est beaucoup moins évident) :
si 2 mesures P_1 et P_2 sur (R^N, B^N) coïncident sur tous les boréliens de la forme $\{X < x\}, x \in R^N$, ce qui revient à dire que leurs fonctions de répartition respectives F_1 et F_2 sont égales sur R^N , alors elles coïncident également sur B^N *c.a.d. sur l'ensemble de tous les boréliens*.

Propriété 5 (propriétés des fonctions de répartition)
- $\forall i = 1,...,N$ les applications partielles $x_i \in R \to F_X(x_1,...,x_N)$ sont *non décroissantes* et *continues à gauche*[2],
- si $\exists i$ t.que $x_i = -\infty$ alors $F_X(x_1,...,x_N) = 0$
- si $\forall i = 1,...,N$ $x_i = +\infty$ alors $F_X(x_1,...,x_N) = 1$
- si $N = 1$ $F_X(+\infty) = 1$ et $F_X(-\infty) = 0$
- Avec $\Delta^x_{a,b} f(x, y, z,..) \overset{def}{=} f(b, y, z,..) - f(a, y, z,..)$, $\forall$ le sous-ensemble de K indices distincts $\{i_1,..,i_K\} \subset \{1,..,N\}$ et $\forall$ les K intervalles $[a_{i_1} b_{i_1}[\subset R$:

$$(\prod_{l=1}^{K} \Delta^{x_l}_{a_{i_l}, b_{i_l}}) F_X(x_1,..,x_N) \geq 0$$

Pour $N=2$ par exemple cette condition s'écrit :

$$\forall a, b, a', b' : \Delta^y_{a',b'} \Delta^x_{a,b} F_X(x, y) = F_X(b, b') - F_X(a, b') - F_X(a', b) + F_X(a, a') \geq 0 .$$

Théorème 2
Pour toute fonction $F : R^N \to [0,1]$ possédant les propriétés ci dessus *il existe* une mesure de probabilité sur (R^N, B^N) admettant F comme fonction de répartition, cette mesure étant *unique* du fait du théorème précédent. On appellera cette propriété *propriété de prolongement.*

Remarque (V.A. à valeur presque sûrement finie) : la définition classique d'une VA implique que les événements du type $\{X_i = \pm\infty\}$ correspondent à l'événement impossible. La définition peut cependant être étendue quand cela est nécessaire. Considérons par exemple la modélisation d'un jeu consistant à lancer de manière répétitive un dé choisi une fois pour toutes au hasard entre un dé A et un dé B. Introduisons alors la VA X égale au nombre d'essais nécessaires pour obtenir pour la première fois la suite de résultats consécutifs (1,1,4,4,5,5) et supposons que le dé A est normal alors que le dé B est truqué, avec 2 faces

[2] Continues à droite avec la définition $F_X(x) = P(X \leq x)$

portant le numéro 2 et aucune l'as. Dans ce cas non seulement l'événement $\{X = \infty\}$ n'est pas impossible mais de plus sa probabilité sera non nulle et égale à 1/2 (on peut le montrer).

II.3-Classification et décomposition des lois sur (R^N, B^N)

Les lois sur R^N, B^N peuvent être classées en 3 catégories :

- **lois discrètes**,
- **lois absolument continues,**
- **lois mixtes** (cad toutes celles qui ne sont ni discrète ni continue)

Lois discrètes
Une V.A. X est dite de loi discrète si il existe un sous-ensemble $V = \{x_i, i \in I, I \text{ dénombrable}\} \subset R^N$ tel que $P_X(V) = 1$.
Dans ce cas $\forall$ E borélien $P_X(E) = \sum_{x_i \in E} P_X(\{x_i\}) = \sum_{x_i \in E} P(X = x_i)$

La loi (et donc sa fonction de répartition) est complètement spécifiée par ce qu'on appelle sa distribution de probabilité, définie ci-après.

Définition 6
On appelle **distribution de probabilité** (de la VA de loi discrète X ou plus précisément de sa loi discrète P_X) sur l'ensemble de valeurs $x_i, i \in I$ l'ensemble des $p_i, i \in I$: les p_i sont 'distribuées' sur les x_i.

Propriété 7
Une loi de probabilité discrète est complètement spécifiée par sa distribution de probabilité
Exercice : le montrer

Remarque : Ne pas confondre avec le terme anglais *distribution function* utilisé par les anglo-saxons pour désigner une fonction de répartition.
La fonction de répartition F_X est dans ce cas une *fonction étagée* (en forme d'escalier N-dimensionnel). Elle prend ses valeurs dans un ensemble dénombrable de réels éventuellement fini entre 0 et 1. Les points de discontinuité correspondent exactement à l'ensemble $\{x_i, i \in I\}$, F_X étant constante en dehors de ces points.

Lois absolument continues
Une V.A. est dite de loi absolument continue si il existe une fonction $p_X : R^N \to R^+$ appelée **densité de probabilité** et qui est telle que :
$\forall E \in B^N, P_X(E) = P(X \in E) = \int_{x \in E \in R^N} p_X(x)dx$ [3]
Dans ce cas :

- La fonction de répartition F_X est continue, dérivable, et s'obtient par intégration,

[3] En toute rigueur p_X doit être mesurable et l'intégrale prise au sens de Lebesgue mais dans les cas courants p_X est continue sauf sur un ensemble négligeable de points et la notion ordinaire d'intégrale de Rieman généralisée suffit.

II : VARIABLES ALEATOIRES

$$F_X(x_1,...,x_N) = \int\limits_{u_1<x_1} ... \int\limits_{u_N<x_N} p_X(u_1,...,u_N)du_1...du_N$$

- la densité s'obtient par dérivation de la fonction de répartition,

$$\frac{\partial^N}{\partial x_1..\partial x_N} F_X(x_1,...,x_N) = p_X(x_1,..,x_N),(x_1,..,x_N) \in R^N$$

Puisque la densité de probabilité d'une VA X de loi absolument continue spécifie la fonction de répartition, elle spécifie également sa la loi de probabilité.

Remarque : on utilise p (minuscule) pour noter une densité de probabilité alors que P (majuscule) est réservée pour désigner une mesure de probabilité.

Interprétation 'physique' :
Tout comme en intégrant une densité de masse sur un volume on obtient la masse de ce volume ou en intégrant une densité de charge électrostatique sur un domaine on obtient la charge totale supportée par ce domaine on peut dire qu'en intégrant une densité de probabilité sur un sous-ensemble de R^N on obtient la probabilité de ce sous-ensemble c'est à dire la probabilité pour qu'une réalisation $(X_1(\omega),..,X_N(\omega))$ de $(X_1,..,X_N)$ lui appartienne .

Propriétés 8

- $p_X \geq 0$ sur R^N [4]
- $\int\limits_{R^N} p_X(x) = 1$

Lois mixtes

Une V.A. est dite de loi mixte si elle n'est ni discrète ni absolument continue. On peut en fait démontrer le résultat suivant :

Théorème 3 (décomposition d'une loi de probabilité)

Toute mesure de probabilité P_X sur (R^N, B^N) peut se décomposer comme suit :
$\forall E$ borélien: $P_X(E) = (\alpha.P_D + \beta.P_{AC} + \gamma.P_{SG})(E)$ où

- $(\alpha.P_D + \beta.P_{AC} + \gamma.P_{SG})(E) \overset{def}{=} \alpha.P_D(E) + \beta.P_{AC}(E) + \gamma.P_{SG}(E)$,
- α,β,γ sont 3 réels ≥ 0 de somme 1, P_D et P_{AC} sont respectivement une mesure de probabilité discrète et une mesure de probabilité absolument continue sur B^N,
- P_{SG} est une troisième mesure de probabilité sur B^N dite *singulière* qui n'est ni discrète ni absolument continue .

$\rightarrow$ quand $\alpha = 1$ P_X est discrète,
$\rightarrow$ quand $\beta = 1$ P_X est absolument continue,
$\rightarrow$ une loi P_X qui n'est ni discrète ni absolument continue est donc une loi telle que $\alpha \neq 1$ et $\beta \neq 1$.

[4] Du moins en tout point où p_X est continue ; des valeurs négatives peuvent être admises sur un ensemble de 'volume nul' dans R^N dit ensemble négligeable.

II : VARIABLES ALEATOIRES

Dans le cadre des sciences de l'ingénieur si $N = 1$ on considère que $\gamma = 0$, cad que la partie singulière n'existe pas. Ceci n'est plus vrai pour $N > 1$ car alors la partie singulière peut intervenir dans des situations concrètes.

Exemple : Considérons un triangle de sommets A, B, C de coordonnées respectives (0,0), (0,1) et (1,1) dans R^2 . Un point M est choisi sur le triangle suivant la procédure aléatoire suivante :

- on lance un dé non pipé et on note le numéro N sortant,
- si N est 1 ou 2 on choisit au hasard l'un des 3 sommets,
- sinon, si N est 3 ou 4 on choisit un point au hasard sur le périmètre,
- sinon on choisit un point au hasard sur l'ouvert correspondant à l'intérieur du triangle.

Choisir 'au hasard' signifie avec équirépartition des chances.
On considère alors les 2 VA correspondant aux 2 coordonnées aléatoires X et Y dans R^2 pour le point ainsi choisi. La loi de la VA 2-dimensionnelle $(X,Y):\Omega \to R^2$ comporte une partie discrète, une partie absolument continue et une partie singulière avec $\alpha = \beta = \gamma = 1/3$.

Des exemples classiques de lois de probabilité en dimension 1 :

- Exemples de lois discrètes :
  - Aléa de **Bernoulli** de paramètre p: $X(\omega) \in \{0,1\}$, $P(X=1) = 1 - P(X=0) = p \in]0,1[$
  - Loi **Binomiale** de paramètres N et p:
    $X(\omega) \in N\,, P(X=k) = C_N^k p^k (1-p)^{N-k}, k \geq 0, p \in]0,1[$
  - Loi de **Poisson** de paramètre λ **:** $X(\omega) \in N$, $P(X=k) = e^{-\lambda}\dfrac{\lambda^k}{k!} k \geq 0, \lambda > 0$
- Exemples de lois absolument continues :
  - Loi **uniforme** (**équirépartie**) sur $[a,b] \subset R$: $p_X(x) = 1/(b-a)$ si $x \in [a,b], = 0$ sinon
  - Loi **normale** (**gaussienne**) de paramètres m et σ :
    $$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp - \frac{1}{2\sigma^2}(x-m)^2,\ \sigma \in R^{*+}, m \in R$$
  - Loi **exponentielle** de paramètre *a :* $p_X(x) = a\exp - axU(x), a \in R^{*+}$
  - Loi de **Laplace** de paramètre *a :* $p_X(x) = \frac{1}{2}a\exp - a|x|, x \in R, a \in R^{*+}$

II.4-Lois marginales

X étant une VA (variable aléatoire) N dimensionnelle sur un espace probabilisé (Ω, τ, P), c'est à dire une application :

$X : \omega \in \Omega \to X(\omega) \in R^N$

Pour tout entier p tel que $1 < p < N$ et à p fixé pour toute sous-suite $S = (i_1,..,i_p)$ de $\{1,..,N\}$ constituée de p indices distincts $1 \leq (i_1,...,i_p) < N$, il est possible d'introduire une VA p-dimensionnelle :

$$\omega \to \pi_S(X(\omega)) \overset{def}{=} (X_{i_1}(\omega),...,X_{i_p}(\omega)) \in R^p$$

où $\pi_S : R^N \to R^p$ est l'application 'projection' qui retient les p coordonnées dans S de $X(\omega)$. Il est possible de vérifier que l'application π_S est mesurable. L'application composée $\pi_S \circ X : \Omega \to R^p$ (notée également $\pi_S(X)$) est donc également mesurable et définit effectivement une VA[5]

Définition 8
On appelle lois marginales de la loi P_X les lois de probabilité de toutes les VA du type $\pi_S(X)$ obtenues pour toutes les suites d'indices $S = (i_1,...,i_p), 1 \le p < N$ introduits ci-dessus. Il y a au plus $\sum_{k=1}^{N-1} C_1^k$ lois marginales distinctes si on impose $i_1 < ... < i_p$.

Dans le but d'alléger les notations décomposons le groupe de N VA $X_1,..,X_N$ en 2 sous-groupes $Y_1,..,Y_p, 1 \le p \le N-1$ et $Z_1,..,Z_q$, avec $p + q = N$, si bien que $\{X_1,..,X_N\} = \{Y_1,..,Y_p\} \cup \{Z_1,..,Z_q\}$. Ces 2 sous-groupes correspondent respectivement à une VA p-dimensionnelle pour le premier et à une VA q-dimensionnelle pour le second. Nous allons voir comment obtenir leurs lois respectives, qui sont des lois marginales comme défini ci-dessus, à partir de la loi conjointe de $X_1,..,X_N$. En fait il suffit pour cela d'obtenir les fonctions de répartition correspondantes.

Fonctions de répartition marginales
Considérons par exemple la loi marginale de $Y_1,..,Y_p$ et plus spécifiquement sa fonction de répartition. On a :

$$F_{Y_1,..,Y_p}(y_1,..,y_p) = \Pr(Y_1 < y_1,..,Y_p < y_p) = \Pr(Y_1 < y_1,..,Y_p < y_p, Z_1 < +\infty,..,Z_q < +\infty))$$

où la 2ème égalité s'explique par le fait qu'on a la relation entre événements :

$$\{ Y_1 < y_1,..,Y_p < y_p, Z_1 < +\infty,..,Z_q < +\infty \} = \{Y_1 < y_1,..,Y_p < y_p\}$$ [6]

La fonction de répartition p dimensionnelle de Y s'exprime donc en fonction de la fonction de répartition N dimensionnelle de X :

$$F_{Y_1,..,Y_p}(y_1,..,y_p) = F_{X_1,..,X_N}(y_1,..,y_p,+\infty,..,+\infty), (y_1,..,y_p) \in R^p$$

Densités de probabilité marginales.
Dans le cas où il existe une fonction densité de probabilité $p_{X_1,..,X_N}$ il est possible d'en déduire une fonction densité pour le sous-ensemble des p VA $Y_1,..,Y_p$. En effet :

$$\forall (y_1,..,y_p) \in R^p, P(Y_1 < y_1,..,Y_p < y_p) = \int_{-\infty}^{y_1} .. \int_{-\infty}^{y_p} [\int_{-\infty}^{\infty} .. \int_{-\infty}^{\infty} p_{X_1,..,X_N}(x_1,..,x_N) dx_{p+1}..dx_N]\, dx_1..dx_p$$

$$= \int_{-\infty}^{y_1} .. \int_{-\infty}^{y_p} p_{Y_1,..,Y_p}(x_1,..,x_p) dx_1..dx_p$$

La fonction de p variables

---

[5] Voir paragraphe sur loi image de $Y = f(X)$ avec X variable aléatoire et f mesurable.

[6] En supposant que $\pm\infty \notin Z_i(\Omega), i = 1,..,q$ et donc que $\{Z_i < \infty\} = \Omega$ pour tout i.

$$p_{Y_1,..,Y_p}(x_1,..,x_p) = \int_{-\infty}^{\infty}..\int_{-\infty}^{\infty} p_{X_1,..,X_N}(x_1,..,x_N)dx_{p+1}..dx_N$$

correspond donc à une fonction densité de probabilité pour $Y_1,..,Y_p$.

Remarque : on a raisonné ci-dessus pour $Y_1,..,Y_p$ correspondant aux p premières coordonnées de $X_1,..,X_N$ mais en procédant de manière analogue on peut calculer la densité marginale pour $Y_1 = X_{i_1},..,Y_p = X_{i_p},(i_1,..,i_p) = S$. On trouve alors :

- $F_{Y_1,..,Y_p}(y_1,..,y_p) = F_{X_1,..,X_N}(x_1,..,x_N)$ où $x_k = \infty$ si $\notin S$ et $x_k = y_l$ si $k = i_l \in S$
- $p_{Y_1,..,Y_p}(x_{i_1},..,x_{i_p}) = p_{X_{i_1},..,X_{i_p}}(x_{i_1},..,x_{i_p}) = \int_{-\infty}^{\infty}..\int_{-\infty}^{\infty} p_{X_1,..,X_N}(x_1,..,x_N)dx_{k_1}..dx_{k_{N-p}}$

  où $X_{k_1},..,X_{k_{N-p}}$ sont les *N-p* composantes du vecteurs de départ qui restent (l'ordre étant conservé) lorsque les p composantes $Y_1 = X_{i_1},..,Y_p = X_{i_p}$ ont été supprimées .

II.5-Indépendance entre variables aléatoires

Indépendance entre événements (rappels)

- Soit un espace probabilisé (Ω,τ,P). Par définition les *N* événements $E_1,..,E_N$ pris dans τ (et qui correspondent donc à *N* sous-ensembles de Ω) sont dits indépendants dans l'ensemble pour la mesure de probabilité *P* si pour tout sous-ensemble de *K* indices distincts $i_1,..,i_K$ pris dans $\{1,..,N\}$, avec $2 \le K \le N$ on a $P(\bigcap_{k=1..K} E_k) = \prod_{k=1..K} P(E_k)$.
- On a donc en particulier
  $P(\bigcap_{k=1..N} E_k) = \prod_{k=1..N} P(E_k)$ et $\forall (i,j), 1 \le i,j \le N), i \ne j, P(E_i \cap E_j) = P(E_i)P(E_j)$
  <u>mais ces seules conditions ne sont pas en général équivalentes à l'indépendance dans l'ensemble (elles ne suffisent pas).</u>
- Des événements de (Ω,τ) indépendants dans l'ensemble pour une mesure *P* ne seront pas, en général, indépendants dans l'ensemble pour une autre mesure *Q* sur le même espace.
- Indépendance entre tribus
  N tribus $\tau_1,..,\tau_N$ sur (Ω,τ,P) sont indépendantes dans l'ensemble si $\forall E_1,..,E_N$ pris respectivement dans $\tau_1,..,\tau_N$, $E_1,..,E_N$ est une famille indépendante dans l'ensemble.

Définition 9

Considérons *N* VA $X_1,..,X_N$ sur un même espace probabilisé à valeurs respectivement dans $R^{d_1},..,R^{d_N}$ où $d_1,..,d_N$ sont *N* entiers chacun ≥ 1. Un cas particulier est $d_1 = .. = d_N = 1$ auquel cas les *N* VA sont, chacune, à valeurs dans *R*. On dira que $X_1,..,X_N$ sont indépendantes dans l'ensemble si les tribus $\tau_{X_1},..,\tau_{X_N}$ sont indépendantes dans l'ensemble, cad si <u>pour toute famille de</u> <u>boréliens</u> $E_1,..,E_N$ inclus dans $R^{d_1},..,R^{d_N}$ on a :

$$P(X_1 \in E_1,..,X_N \in E_N) = \prod_{i=1..N} P(X_i \in E_i)$$

Remarque : la condition d'indépendance inclus les cas où pour certains indices *i* on a $E_i = R^{d_i}$ (ce qui fait que $P(X_i \in E_i) = 1$ pour ces indices).

II : VARIABLES ALEATOIRES

Propriété 9 (critères d'indépendance)

Soit p VA $Z_1,..,Z_p$ de dimensions respectives $d_1,..,d_N$ définies sur le même espace probabilisé .

- $Z_1,..,Z_p$ sont indépendantes dans l'ensemble ssi $F_{Z_1,..,Z_p}(z_1,..,z_p) = \prod_{k=1}^{p} F_{Z_k}(z_k)$ sur $R^{d_1+..+d_p}$
- Si $Z_1,..,Z_p$ admettent une densité de probabilité conjointe, elles sont indépendantes dans l'ensemble ssi $p_{Z_1,..,Z_p}(z_1,..,z_p) = \prod_{k=1}^{p} p_{Z_k}(z_k)$ sur $R^{d_1+..+d^p}$

A titre d'exemple pour $Z_i : \Omega \to R, i = 1,..,4$ et $X_1 = (Z_1, Z_3)$ à valeurs dans R^2 et $X_2 = Z_2$, $X_3 = Z_4$:

X_1 , X_2 et X_3 sont indépendantes dans l'ensemble

- ssi $F_{Z_1,..,Z_4}(z_1,..,z_4) = F_{Z_1,Z_3}(z_1,z_3)F_{Z_2}(z_2)F_{Z_4}(z_4)$ sur R^4
- ssi $p_{Z_1,..,Z_4}(z_1,..,z_4) = p_{Z_1,Z_3}(z_1,z_3)p_{Z_2}(z_2)p_{Z_4}(z_4)$ sur R^4 pour une loi continue

Exercice : proposer d'autres exemples
Exercice : donner un critère d'indépendance dans le cas d'une loi conjointe discrète

On retiendra que la possibilité de factoriser la fonction de répartition conjointe, la densité de probabilité conjointe (dans le cas continu), la distribution de probabilité conjointe (dans le cas discret) sont des condition NS d'indépendance dans l'ensemble pour des VA de dimensions quelconques.

Propriété 10

Si $X_1,..,X_N$ indépendantes dans l'ensemble alors $\forall$ les N fonctions $g_1,..,g_N$ (mesurables) de la forme $g_i : R^{d_i} \to R^{d_i^{'}}$ alors les N VA $g_i(X_i)$ à valeurs respectivement dans $R^{d_i^{'}}$ sont également indépendantes dans l'ensemble.

II.6-Introduction aux lois de probabilité conditionnelles

La notion de loi conditionnelle est à traiter avec précaution car elle oblige en général à considérer des probabilités conditionnelles avec des événements conditionnants de probabilité nulle.

1.Définition

Soit une paire (X,Y) de VA (définies évidemment sur un même espace probabilisé). On appelle loi de Probabilité de Y conditionnelle à $X = x$ l'application :

$$P_{Y/X=x} : E \in B \to \lim_{dx \to 0} \frac{P(Y \in E, X \in [x, x+dx[)}{P(X \in [x, x+dx[)}$$

Avec $B =]-\infty, y[$ est introduite la fonction de répartition conditionnelle

$$F_{Y/X=x} : y \to \lim_{dx \to 0} \frac{P(Y < y, X \in [x, x+dx[)}{P(X \in [x, x+dx[)}$$

lorsque la limite existe.

Commentaires :

II : VARIABLES ALEATOIRES

- Il est vérifiable que cette cette limite existe $\forall E$ ou $\forall y$ pour tout x sauf pour un ensemble de valeurs de x contenu dans un borélien A tel que $P(X \in A) = 0$ et qu'à $E \in B$ ou à y fixé l'application $x \to P_{Y/X=x}(E)$ est τ_X mesurable.

- On ne peut pas utiliser en général une définition + directe (cad sans passage à la limite) car on peut avoir $P(X = x) = 0$ (ce qui est vrai pour tout x si X est de loi continue) et sans l'utilisation d'un passage à la limite on aurait une forme indéterminée 0/0. Poser $P_{Y/X=x}(E) = 0$ pour lever l'indétermination ne mène à rien dans le cas où l'ensemble des x amenant 0/0 est de probabilité non nulle.
- La loi conditionnelle, à $x \in A$ fixé *est une loi de probabilité* sur *R*. Elle admet toujours une fonction de répartition. Elle peut être continue (et elle admet alors une densité de probabilité dite conditionnelle), discrète où mixte suivant les cas (et suivant la valeur de x).

2.Propriété :

- Si X est de loi continue :
  - $\forall E, F$ boréliens de R, $P(X \in E, Y \in F) = \int_{x \in E} P_{Y/X=x}(F) p_X(x) dx$
  - et pour $E = R : P(Y \in F) = \int_R P_{Y/X=x}(F) p_X(x) dx$
- Si X est de loi discrète à valeurs dans $X(\Omega) = \{\alpha_k, k \in I\}$:
  - $\forall E, F$ boréliens de R, $P(X \in E, Y \in F) = \sum_{k \in K} P_{Y/X=\alpha_k}(F) P(X = \alpha_k)$ où $E = \{\alpha_k, k \in K\}$
  - et pour $E = R$, $P(Y \in F) = \sum_{k \in I} P_{Y/X=\alpha_k}(F) P(X = \alpha_k)$
- Que la loi de X soit continue, discrète ou mixte la loi $P_{Y/X=x}$, à x fixé, peut être continue, discrète ou mixte. Si la loi de Y est discrète $P_{Y/X=x}$ est obligatoirement discrète. Si la loi de Y est continue il est possible que $P_{Y/X=x}$ soit discrète.

3.Cas d'une loi conjointe absolument continue

Dans ce cas les lois de Y si $X = x$ et de Y si $X = x$ admettent des densités conditionnelles :

$$y \to p_{Y/X=x}(y) = \frac{p_{X,Y}(x,y)}{p_X(x)} \text{ si } p_X(x) > 0 \text{ et} = 0 \text{ sinon}$$

$$x \to p_{X/Y=y}(y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} \text{ si } p_Y(y) > 0 \text{ et} = 0 \text{ sinon}$$

Exercice : montrer la formule de Bayes qui relie les 2 densités conditionnelles :

$$p_{Y/X=x}(y) = \frac{p_{X/Y=y}(x) p_Y(y)}{\int_{y \in R} p_{X/Y=y}(x) p_Y(y) dy}$$

4.Cas d'une loi conjointe discrète

Dans ce cas on a $P(X \in \{\alpha_k, k \in I \subset Z\}) = 1$ et $P(Y \in \{\beta_l, l \in J \subset Z\}) = 1$ pour 2 ensembles de valeurs (les α_k et les β_l). La loi conjointe est spécifiée par toutes les valeurs de probabilités $P(X = \alpha_k, Y = \beta_l), (k,l) \in I \times J$ (dont certaines peuvent être nulles[7]).

Les lois de probabilité conditionnelles $P_{Y/X=x}$ et $P_{X/Y=x}$ sont alors également discrètes, la première étant spécifiée par les probabilités conditionnelles :

$$P(Y = \beta_l / X = \alpha_k) = \frac{P(X = \alpha_k, Y = \beta_l)}{P(X = \alpha_k)} = \frac{P(X = \alpha_k, Y = \beta_l)}{\sum_{l \in J} P(X = \alpha_k, Y = \beta_l)}, l \in J, k \in I$$

et la 2ème par la même formule en échangeant les rôles de *X* et de *Y*. Ces probabilités conditionnelles constituent une distribution de probabilité (dite conditionnelle) sur les $\beta_l, l \in J$

Exercice : écrire une formule pour le cas discret analogue à la formule de Bayes introduite ci-dessus dans le cas continu.

5.Généralisation : conditionnement en dimension > 2

Soit $(Y_1,..,Y_M, X_1,..,X_N)$ un ensemble de VA sur le même (Ω, τ, P), chacune à valeur respectivement dans $R^{d_1},..,R^{d_M}, R^{d_1'},..,R^{d_N'}$. Il est toujours possible d'introduire la loi conditionnelle de $(Y_1,..,Y_M)$ conditionnelle à $(X_1 = x_1,..,X_N = x_n)$ que l'on note $P_{Y_1,..,Y_M / X_1 = x_1,..,X_N = x_n}$.

Par exemple, dans le cas où les $M + N$ VA admettent une loi conjointe absolument continue la loi conditionnelle est également de type absolument continu sur $(R^{\sum_1^M d_i}, B^{\sum_1^M d_i})$ avec une densité de la forme :

$$p_{Y_1,..,Y_M / X_1 = x_1,..,X_N = x_n}(y_1,..,y_M) = \frac{p_{X_1,..,X_N,Y_1,..,Y_M,}(x_1,..,y_M)}{p_{X_1,..,X_N}(x_1,..,x_N)} \quad \text{si } p_{X_1,..,X_N}(x_1,..,x_N) \neq 0$$

$$p_{Y_1,..,Y_M / X_1 = x_1,..,X_N = x_n}(y_1,..,y_M) = 0 \text{ sinon}$$

pour $(x_1,..,x_N) \in R^{\sum_1^N d_i'}$.

II.7-Calculs de la loi image d'une VA fonction d'une autre VA

Le problème est, pour une VA $X : \Omega \to R^N$ sur (Ω, τ, P) et $f : \mathbb{R}^N \to \mathbb{R}^M$ mesurable, de trouver la loi de la VA $Y = f \circ X$, autrement dit de trouver la loi P_Y image de P_X par f (loi qui est donc aussi image de P par $f \circ X$).

1.Fonction de répartition image

Le principe du calcul est élémentaire (bien que le calcul lui-même puisse s'avérer compliqué en pratique). On a:

$$F_Y(y) = P(Y < y) = P_X(\{ \ x / f(x) < y \ \}) \text{ où } Y < y \Leftrightarrow \forall i : Y_i < y_i$$

2.Image d'une loi discrète

---

[7] En général l'ensemble des (α_k, β_l) t. que $P(X = \alpha_k, Y = \beta_l) > 0$ ne représente qu'un sous-ensemble de l'ensemble produit $\{\alpha_k, k \in I\} \times \{\beta_l, l \in J\}$

II : VARIABLES ALEATOIRES

Si X est de loi discrète, cad si $\exists E = \{x_i \in \mathbb{R}^N, i \in I \subset \mathbb{N}\} : P_X(E) = 1$, alors évidemment Y est <u>toujours de loi discrète</u> puisque l'image $f(E)$ de E par f est dénombrable et que par construction $P_Y(f(E)) = 1$. En considérant $f(E) = \{y_j \in \mathbb{R}^M, j \in J \subset \mathbb{N}\}$ la distribution de probabilité pour Y s'obtient par :

$$P(Y = y_j) = \sum_{x_i : f(x_i) = y_j} P(X = x_i),\ j \in J$$

3.Image d'une loi continue

L'image P_Y d'une loi absolument continue P_X de densité p_X peut être <u>absolument continue, discrète ou mixte</u>. Tout dépend du couple (p_X, f)**.** En particulier:

> si l'image $f(E)$ de E est dénombrable de la forme $f(E) = \{y_j \in \mathbb{R}^M, j \in J \subset \mathbb{N}\}$
>
> alors obligatoirement la loi de Y est discrète et s'écrit:
>
> $$P(Y = y_j) = \int_{x : f(x) = y_j} p_X(x)dx,\ j \in J$$

Transformation par un difféomorphisme par morceaux d'une densité de probabilité

Soit $X : \Omega \to \mathbb{R}^N$ de loi absolument continue avec une densité p_X sur $\mathbb{R}^N$ et soit $f : \mathbb{R}^N \to \mathbb{R}^N$. Supposons que:

- il existe une famille $O_i, i = 1...K$ d'ouverts disjoints dans $\mathbb{R}^N$ tels que $P_X(\bigcup_{i=1...K} O_i) = 1$ (cad que la probabilité pour que X ne tombe dans aucun de ces ouverts est nulle).
- la restriction de f à chacun des O_i étant notée f_i et l'image de O_i par f_i étant notée $O_i^{'}$, pour tout i les fonctions $f_i : O_i \to O_i^{'}$ correspondent à des difféomorphismes (cad des fonctions continuement différentiables et admettant, chacune, une inverse continuement différentiable) $f_i^{-1} : O_i^{'} \to O_i$.

Soit d'autre part la VA $Y = f \circ X$ (on peut montrer que f est obligatoirement mesurable étant données les hypothèses). On va montrer que Y admet une densité de probabilité: $p_Y(y) = \sum_{i : y \in O_i^{'}} p_X(f_i^{-1}(y)) \left|[f_i^{-1}]'(y)\right|, y \in \bigcup_{i=1,..K} O_i^{'}$ où $[f_i^{-1}]'$ désigne la matrice des dérivées partielles d'éléments $\dfrac{\partial f_i^{-1}(y_k)}{\partial x_l}, (i,k) \in \{1,...,N\}^2$ et où $|A|$ désigne ici la valeur absolue du déterminant de la matrice A.

<u>Preuve:</u>

Pour un borélien quelconque $E \subset \mathbb{R}^N$:

$$P_Y(E) = P_Y(E \bigcap (\bigcup_{1...K} O_i^{'})) = P_X(f^{-1}(\bigcup_{1...K} (O_i^{'} \bigcap E))) = P_X(\bigcup_{1...K} f^{-1}(O_i^{'} \bigcap E))$$
$$= P_X(\bigcup_{1...K} f^{-1}(O_i^{'}) \bigcap f^{-1}(E)) = P_X(\bigcup_{1...K} O_i \bigcap f^{-1}(E)) = \sum_{1...K} P_X(f_i^{-1}(E))$$

Or:

$$P_X(f_i^{-1}(E)) = \int_{x \in f_i^{-1}(E)} p_X(x)dx = \int_{y: f_i^{-1}(y) \in f_i^{-1}(E \cap O_i^{'})} p_X(f_i^{-1}(y)) \left|[f_i^{-1}]'(y)\right| dy = \int_{y \in E \cap O_i^{'}} p_X(f_i^{-1}(y)) \left|[f_i^{-1}]'(y)\right| dy$$

où la 2ème égalité correspond à l'utilisation de la formule du changement de variable dans les intégrales multiples. En Conséquence:

$$\sum_{1...K} P_X(f_i^{-1}(E)) = \sum_{1...K} \int_{y \in E \cap O_i^{'}} p_X(f_i^{-1}(y)) \left|[f_i^{-1}]'(y)\right| dy = \int_{y \in E} \sum_{\substack{i=1...K \\ i: y \in O_i^{'}}} p_X(f_i^{-1}(y)) \left|[f_i^{-1}]'(y)\right| dy$$

et puisque

$$\forall E \in B^N : P_Y(E) = \int_{y \in E} [\sum_{\substack{i=1...K \\ i: y \in O_i^{'}}} p_X(f_i^{-1}(y)) \left|[f_i^{-1}]'(y)\right|] \; dy$$

on a par identification:

$$\boxed{p_Y(y) = \sum_{\substack{i=1...K \\ i: y \in O_i^{'}}} p_X(f_i^{-1}(y)) \left|[f_i^{-1}]'(y)\right| , y \in \mathbb{R}^N \quad (y \notin \bigcup_i O_i^{'} \Rightarrow p_Y(y) = 0)}$$

Cas particulier:

Si $\exists O \subset \mathbb{R}^N : P_X(O) = 1$ et si $f : O \to O^{'} = f(O)$ est continuement différentiable et admet une inverse f^{-1} sur $O^{'}$ continuement différentiable la formule devient :

$$p_Y(y) = p_X(f^{-1}(y)) \left|[f^{-1}]'(y)\right| \; , y \notin O^{'} \Rightarrow p_Y(y) = 0$$

<u>Remarque</u>

Dans le cas $X : \Omega \to \mathbb{R}$, $f : \mathbb{R} \to \mathbb{R}$ où f est un difféomorphisme le résultat est très facile à démontrer en passant par le calcul de la fonction de répartition de $Y = f(X)$:

$F_Y(y) = P(Y < y) = P(f(X) < y) = P(X < f^{-1}(y)) = F_X(f^{-1}(y))$ dans le cas où f est croissante

$F_Y(y) = P(Y < y) = P(f(X) < y) = P(X > f^{-1}(y)) = 1 - F_X(f^{-1}(y))$ dans le cas où f est décroissante et en se souvenant que, X étant de loi absoluement continue, $P(X = f^{-1}(y)) = 0$

Pour obtenir la densité p_Y il suffit de dériver par rapport à y:

- $p_{Y(y)} = p_X(f^{-1}(y)) \dfrac{\partial f^{-1}(y)}{\partial y}$ dans le cas d'une fonction croissante,
- $p_{Y(y)} = -p_X(f^{-1}(y)) \dfrac{\partial f^{-1}(y)}{\partial y}$ dans le cas d'une fonction décroissante,

ce qui s'écrit, en regroupant les 2 cas (puisque $\dfrac{\partial f^{-1}(y)}{\partial y} < 0$ dans le deuxième):

II : VARIABLES ALEATOIRES

$$\boxed{p_{Y(y)} = p_X(f^{-1}(y))\left|\frac{\partial f^{-1}(y)}{\partial y}\right|, y \in R}$$

Il est évidemment possible de particulariser au cas où $P_X(O) = 1$ pour un ouvert O de $\mathbb{R}$ et où

$f : O \to O^{'} = f(O)$ est un difféomorphisme (croissant ou décroissant) puisque comme d'habitude on peut toujours évacuer les ensembles de probabilité nulle sans que cela change les résultats.