

I - Rappels sur les variables aléatoires① - Espace fini ou dénombrablea) Quelques résultats utiles sur les séries.

On rappelle ici quelques résultats sur les séries, souvent utilisés dans l'étude des variables aléatoires sur un espace dénombrable. On introduit $(u_n)_{n \geq 1}$ une suite numérique, et $S_n = u_1 + \dots + u_n$ la somme partielle à l'ordre n .

- S1 - La série $\sum_n u_n$ est dite convergente si S_n CV vers une limite finie S .
- S2 - Si la série $\sum_n u_n$ CV, la suite $(u_n)_{n \geq 1}$ tend vers 0. Δ : réciproque fautive.
- S3 - La série $\sum_n u_n$ est dite absolument convergente si la série $\sum_n |u_n|$ CV.
- S4 - Si $u_n \geq 0$ pour tout n , alors la suite S_n est \uparrow , donc elle tend \forall s vers une limite S (parfois infinie) - On écrit $S = \sum_n u_n$, bien que la série CV $\Leftrightarrow S < +\infty$.
- S5 - Si $u_n \geq 0$ pour tout n , la somme $\sum_n u_n$ ne change pas en changeant l'ordre de sommation.
- S6 - Lorsque les u_n sont des réels, et que la série est absolument convergente, on peut modifier de façon arbitraire l'ordre des termes sans changer la propriété d'absolue convergence, ni même la somme de la série.
- S7 - Si $u_n \geq 0$, on peut "sommer par paquets".
- S8 - Si la série $\sum_n u_n$ est absolument convergente, S7 est encore vraie.
- S9 - Fubini: Soit $(a_{mn})_{m,n \in \mathbb{N}}$ une série double telle que la série de terme général $\sum_m |a_{mn}|$ converge. Alors les séries $\sum_n \sum_m a_{mn}$ et $\sum_m \sum_n a_{mn}$ sont CV et de même somme.

b) Contexte :

On suppose donnée une probabilité P sur Ω fini ou dénombrable.

P est munie de la tribu $\mathcal{P}(\Omega)$, et caractérisée par les $p_w = P(\{w\})$.

Alors toute application définie sur Ω est une variable aléatoire et l'ensemble F des valeurs de X (i.e. l'ensemble des $X(w)$ lorsque w parcourt Ω) est fini ou dénombrable. La loi de X est une probabilité sur F , notée P_X , caractérisée par :

$$p_i^X = P_X(\{x_i\}) = P(\{w, X(w) = x_i\}) = P(X = x_i) = \sum_{w, X(w) = x_i} p_w, \quad \forall x_i \in F.$$

Proposition : La loi d'une v.a. X à valeurs dans un espace dénombrable F est caractérisée par $\{(x_i, p_i^X), x_i \in F, \text{ avec } p_i^X = P(X = x_i)\}$.

c) Espérance d'une v.a. discrète

Imaginons qu'on puisse répéter de nb fois une expérience aléatoire, en notant X_1, X_2, \dots, X_n les résultats. Pour résumer le comportement de X , on peut utiliser la moyenne empirique. En regroupant suivant les résultats de l'expérience, on obtient

$$M_n = \sum_{w \in \Omega} f_n(\{w\}) X(w), \quad \text{où } f_n(\{w\}) \text{ est la fréquence de réalisation de } \{w\}.$$

On retrouve la notion d'espérance en passant à la limite.

$$\lim_{n \rightarrow +\infty} M_n = \lim_{n \rightarrow +\infty} \sum_{w \in \Omega} f_n(\{w\}) X(w) = \sum_{w \in \Omega} \lim_{n \rightarrow +\infty} (f_n(\{w\})) X(w)$$

C'est en fait la loi des grands nombres, que nous étudierons plus loin, qui est un des résultats fondamentaux de la théorie des probabilités.

Définition : L'espérance d'une v.a. discrète sur l'espace fini ou dénombrable Ω est le nombre $E[X] = \sum_{w \in \Omega} p_w X(w)$, pourvu que $\sum_{w \in \Omega} p_w |X(w)| < +\infty$.

Théorème: Considérons une v.a. X satisfaisant $\sum_{w \in \Omega} p_w |X(w)| = E[|X|] < \infty$. (2)

On a alors:
$$E[X] = \sum_{w \in \Omega} p_w X(w) = \sum_{x_i \in F} x_i P(X=x_i) = \sum_{x_i \in F} x_i p_i^X$$

→ L'espérance de X ne dépend que de la loi de X .

→ Propriétés: On appelle v.a. intégrable une v.a. X qui admet une espérance.

L'ensemble de toutes les v.a. intégrables est noté L^1 (dépend de Ω et P).

• L^1 est un espace vectoriel, et l'espérance est linéaire sur L^1 : $\forall X, Y \in L^1, \forall (a, b) \in \mathbb{R}$
 $E[aX + bY] = aE[X] + bE[Y]$.

• $X \in L^1 \Leftrightarrow |X| \in L^1$, et dans ce cas $|E[X]| \leq E[|X|]$

• Si $X \geq 0$ (i.e. $\forall w, X(w) \geq 0$) et $X \in L^1$, alors $E[X] \geq 0$.

• Si $X, Y \in L^1$ telles que $X \leq Y$ ($\forall w, X(w) \leq Y(w)$), alors $E[X] \leq E[Y]$.

• L^1 contient toutes les v.a. bornées (X est bornée si $\exists b \in \mathbb{R} \text{ k.p. } |X(w)| \leq b \ \forall w$).

→ Lien entre probabilité d'un événement et espérance d'une v.a.: on suppose A un événement élémentaire et on définit X comme:
$$\begin{cases} X(w) = 1 & \text{si } w \in A \\ X(w) = 0 & \text{si } w \notin A \end{cases}$$

On note $X = 1_A$, et on remarque que:

$$E[X] = E[1_A] = P(A)$$

d) Variance

On considère ici l'ensemble L^2 des v.a. X telles que X^2 soit intégrable. Dans ce cas, on dira que X est de carré intégrable.

Proposition: L^2 est un sous-espace vectoriel de L^1 , et si $X \in L^2$ on a

$$|E[X]| \leq E[|X|] \leq \sqrt{E[X^2]}$$

Interprétation: En considérant l'espérance mathématique, on se ramène à des quantités déterministes - les inégalités sont ensuite logiques et bien connues dans ce cadre.

Définition: Si $X \in L^2$, sa variance est définie par

$$\text{Var}(X) = E[(X - E[X])^2] = \sum_{x_i \in F} (x_i - E[X])^2 p_i^X.$$

→ la variance est toujours ≥ 0 ! On a aussi $\text{Var}(X) = E[X^2] - (E[X])^2$.

e) Moments d'une variable aléatoire et théorème de transfert:

On se ramène ici à un cas plus général.

Considérons une fonction $f: F \rightarrow \mathbb{R}$. Ainsi, $Y = f(X)$ est une v.a. réelle, qui prend un nombre fini ou dénombrable de valeurs (réelles!).

On peut aussi considérer f comme une v.a. sur l'espace probabilisé (F, \mathbb{P}_X) .

Théorème de transfert: Sous les hypothèses précédentes, la v.a. $Y = f(X)$ définie sur (Ω, \mathbb{P}) est intégrable \iff la v.a. f sur (F, \mathbb{P}_X) l'est également.

Dans ce cas, les espérances de ces 2 variables sont égales, et on a:

$$E[f(X)] = \sum_{\omega \in \Omega} f(X(\omega)) p_\omega = \sum_{x_i \in F} f(x_i) \mathbb{P}(X = x_i).$$

→ Idee: dans $Y = f(X)$, f est vue comme une fonction (et donc l'espace reste celui de X , c-à-d (Ω, \mathbb{P})). Ensuite on voit f comme une v.a., ce qui change l'espace probabilisé qui devient (F, \mathbb{P}_X) .

(3)

Définition : Soit $p \in \mathbb{N}^*$. La v.a. X admet un moment d'ordre p si la variable aléatoire $X^p \in L^1$, et ce moment vaut alors (en vertu du théorème de transfert) :

$$E[X^p] = \sum_{x_i \in F} x_i^p P(X=x_i)$$

g) Fonction génératrice.

Cette notion n'a de sens que dans le cas discret !

On considère une v.a. à valeurs dans \mathbb{N} . Sa loi est donc une proba. sur \mathbb{N} . Cette loi est caractérisée par une suite de nombres $\in [0,1]$, de somme 1. En réalité, cette loi peut aussi être caractérisée par la fonction génératrice.

Soit la v.a. X à valeurs dans \mathbb{N} , dont la loi est caractérisée par les nombres $p_n = p_n^X = P(X=n)$.

Définition : La fonction génératrice G_X de X est la fonction définie sur $[0,1]$ par : $\forall s \in [0,1], G_X(s) = E[s^X] = \sum_{n=0}^{+\infty} p_n s^n$.

→ G_X ne dépend que de la loi de X .

Proposition : La fonction génératrice est continue sur $[0,1]$ et indéfiniment dérivable sur $[0,1[$. Elle caractérise la loi de X .

Preuve : G_X est la somme d'une série entière qui converge absolument au point 1 puisque $\sum_n p_n = 1 \Rightarrow G_X$ est continue et indéfiniment dérivable sur $[0,1[$. Comme la $n^{\text{ième}}$ dérivée en 0 est $G_X^{(n)}(0) = p_n n!$, G_X caractérise les p_n , donc la loi de X .

Proposition : Soit X une v.a. à valeurs dans \mathbb{N} , de fonction génératrice G_X .

X intégrable $\Leftrightarrow G_X$ dérivable à gauche en $s=1$.

Dans ce cas, $E[X] = G'_X(1)$.

Preuve: On rappelle que en considérant des fonctions $s \mapsto u_n(s)$ \uparrow et positives sur $[0,1]$, on peut échanger la limite en s au point 1 et la

Somme :
$$\lim_{s \uparrow 1} \sum_{n \geq 0} u_n(s) = \sum_{n \geq 0} \lim_{s \uparrow 1} u_n(s).$$

Or, si $s < 1$,
$$\frac{G_X(s) - G_X(1)}{s-1} = \frac{1}{s-1} \left(\sum_n p_n s^n - \sum_n p_n \right) = \frac{1}{s-1} \sum_n p_n (s^n - 1)$$
$$= \sum_{n \geq 0} p_n \frac{s^n - 1}{s-1} = \sum_{n \geq 0} p_n (1 + s + \dots + s^{n-1}).$$

En posant $u_n(s) = p_n (1 + s + \dots + s^{n-1})$, on a $u_n(s)$ \uparrow et positives, avec

$$\lim_{s \uparrow 1} u_n(s) = p_n \times n.$$

Finalement donc,
$$\lim_{s \uparrow 1} \frac{G_X(s) - G_X(1)}{s-1} = \lim_{s \uparrow 1} \sum_{n \geq 0} u_n(s) = \sum_{n \geq 0} \lim_{s \uparrow 1} u_n(s) = \sum_{n \geq 0} n p_n = E[X].$$

\rightarrow Plus généralement, on a:

Proposition : La v.a. $X(X-1)\dots(X-p)$ est intégrable (donc X admet un moment d'ordre $(p+1)$) $\Leftrightarrow G_X$ est $(p+1)$ fois dérivable à gauche en $s=1$, et on a:

$$E[X(X-1)\dots(X-p)] = G_X^{(p+1)}(1).$$

Rp: En particulier, $E[X(X-1)] = G''_X(1)$, d'où

$$\text{Var}(X) = G''_X(1) - (G'_X(1))^2 + G'_X(1)$$

→ Intérêt de G_X : pour calculer les moments d'une v.a., il est parfois beaucoup plus rapide et simple d'utiliser la fonction génératrice! (4)

g) Lois conditionnelles.

Nous avons déjà évoqué les notions de probabilités conditionnelles et d'événements aléatoires indépendants. L'objectif ici est de faire le pont avec les variables aléatoires. On considère donc X et Y deux v.a. définies sur le m.s. fini ou dénombrable, muni de P . X est à valeurs dans F , Y dans G . (finis ou dénombrables).

→ La connaissance des deux lois P_X et P_Y ne donne aucune information sur les liens qui peuvent unir les comportements aléatoires de X et Y . Pour cela, il faut s'intéresser à la loi du couple $Z = (X, Y)$ à valeurs dans le produit cartésien $F \times G$. On note $P_Z = (P_k^Z, \exists k \in F \times G)$ cette loi.

→ Les lois P_X et P_Y de X et Y s'appellent les lois marginales du couple (X, Y) .

→ Exemple: on lance 2 dés. X est la v.a. du résultat du 1^{er} dé, et Y est celle de la somme des 2 dés. Clairement, la connaissance de X influe sur les valeurs possibles de Y et donc sur sa loi.

Définition : Soit $x_i \in F$ tel que $P(X=x_i) > 0$. On appelle loi conditionnelle de Y sachant $X=x_i$ la probabilité sur G définie par

$$P_j^{Y|X=x_i} = P(Y=y_j | X=x_i), \quad \forall y_j \in G.$$

→ Propriétés d'équivalence: entre:

- $P(X=x_i) = \sum_{y_j \in F} P(Z=(x_i, y_j))$ [On intègre sur la loi de Y pour retrouver la marginale de X]
- $P_j^{Y|X=x_i} = \frac{P(Z=(x_i, y_j))}{P(X=x_i)}$ si $P(X=x_i) > 0$.
- $P(Z=(x_i, y_j)) = \begin{cases} 0 & \text{si } P(X=x_i) = 0 \\ P(X=x_i) P_j^{Y|X=x_i} & \text{si } P(X=x_i) > 0. \end{cases}$

h) Espérance conditionnelle.

On a vu que pour i fixé tel que $p_i^x > 0$, la loi conditionnelle de $Y|X=\pi_i$, donnée par $\{(y_j, p_j^{Y|X=\pi_i}), y_j \in G\}$ définit une probabilité. Dès qu'ils existent, on peut donc accéder à ses moments.

Définition : Soit Y une v.a. intégrable. L'espérance conditionnelle de Y sachant $\{X=\pi_i\}$ est l'espérance de la loi conditionnelle de Y sachant $\{X=\pi_i\}$:

$$E[Y|X=\pi_i] = \sum_j y_j p_j^{Y|X=\pi_i} = \sum_j y_j P(Y=y_j | X=\pi_i)$$

→ C'est une fonction de π_i : on peut donc voir l'espérance conditionnelle comme une variable aléatoire : $E[Y|X] = \Psi(X)$ avec $\Psi(x) = \begin{cases} E[Y|X=x] & \text{si } P(X=x) > 0 \\ 0 & \text{sinon.} \end{cases}$

→ On a la propriété fondamentale suivante :

Théorème : Si Y est intégrable, alors $\Psi(X) = E[Y|X]$ est intégrable et

$$E[\Psi(X)] = E[E[Y|X]] = E[Y].$$

→ Ainsi, on a aussi $E[Y] = \sum_i E[Y|X=\pi_i] P(X=\pi_i)$.

→ Comme l'espérance conditionnelle dérive de l'espérance appliquée à la loi conditionnelle, elle hérite naturellement des propriétés de l'espérance :

- $E[aY + bZ | X] = a E[Y|X] + b E[Z|X]$.

- $E[Y|X] \geq 0$ si $Y \geq 0$.

- $E[1|X] = 1$

- $E[Y g(X) | X] = g(X) E[Y|X]$: i.e., sachant que X est connu, on considère $g(X)$ comme une constante par l'on peut donc sortir de l'espérance.

i) Variables aléatoires indépendantes:

On rappelle que la connaissance de la loi jointe de $Z = (X, Y)$ permet de calculer les lois marginales P_X et P_Y , mais que la connaissance de P_X et P_Y ne suffit pas en général à trouver la loi de Z : il manque la structure de dépendance entre X et Y !

→ Exemple: on considère $Z = (X, Y)$ à valeurs dans $\{(-1, -1), (-1, 1), (1, -1), (1, 1)\}$,

telle que la loi de Z est donnée par

$$\text{On a } P(X=1) = P(Z=(1,1)) + P(Z=(1,-1)) = \frac{1}{2} \\ = P(Y=1)$$

$X \backslash Y$	-1	1
-1	$P/2$	$(1-P)/2$
1	$(1-P)/2$	$P/2$

⇒ Les lois de X et Y ne dépendent pas de p , alors que celle de Z en dépend!

→ Quand l'information que l'on a sur X ne change rien à la loi de Y , on parle d'indépendance.

Définition: Les variables X et Y sont indépendantes si pour toutes parties ACF et BCG, elles vérifient $P(X \in A, Y \in B) = P(X \in A) P(Y \in B)$.

Il en découle par exemple que $P_j^{Y|X=x_i} = P_j^Y$ pour

⇒ la loi conditionnelle est égale à la loi marginale.

(pour $\forall y_j \in G$
pour $\forall x_i \in F$ t.q. $P_i^X > 0$.)

→ Nous avons également le résultat suivant:

Proposition: Supposons les v.a. X et Y indépendantes. Soient f et g deux fonctions réelles définies sur F et G , telles que $f(X) \in L^1$ et $g(Y) \in L^1$. Alors le produit $f(X)g(Y)$ est aussi intégrable et vérifie:

$$E[f(X)g(Y)] = E[f(X)]E[g(Y)].$$

j) Somme de variables aléatoires II :

Proposition : Supposons que X et Y soient 2 variables aléatoires à valeurs dans \mathbb{Z} .
Notons $Z = (X, Y)$. Alors

$$P(X+Y=i) = \sum_{j \in \mathbb{Z}} P(Z=(j, i-j)) = \sum_{j \in \mathbb{Z}} P(Z=(i-j, j)).$$

En particulier, si $X \perp Y$,

$$P(X+Y=i) = \sum_{j \in \mathbb{Z}} P(X=j) P(Y=i-j) = \sum_{j \in \mathbb{Z}} P(X=i-j) P(Y=j)$$

→ C'est un calcul de convolution discrète.

→ On peut également utiliser la fonction génératrice pour la loi de la somme:

Proposition : Supposons $X \perp Y$, à valeurs dans $F=G=\mathbb{N}$, avec $U=X+Y$.

En notant G_X, G_Y et G_U les fonctions génératrices de X, Y et U , on a :

$$\forall s \in [0,1], \quad G_U(s) = G_X(s) G_Y(s).$$

→ Ce dernier résultat permet d'identifier très rapidement la loi d'une somme de v.a. \perp dans certains cas!

② - Support infini : cas des variables aléatoires réelles.

Nous considérons maintenant des v.a. pouvant être à valeurs dans \mathbb{R} tout entier. Dans ce cadre, vous avez vu en cours l'an dernier qu'il fallait considérer comme l'nbu la l'nbu borélienne $\mathcal{B}(\mathbb{R})$ de \mathbb{R} .

Définition : Soit Ω l'espace d'états muni de la l'nbu \mathcal{A} des événements. Une application $X: \Omega \rightarrow \mathbb{R}$ est une v.a. réelle si $[X^{-1}(B)] = \{ \omega \in \Omega \mid X(\omega) \in B \} \in \mathcal{A}$ pour tout $B \in \mathcal{B}(\mathbb{R})$.
notation abusive

⑥ Cette définition fait le lien entre espace probabilisé et variable aléatoire dans le cas où l'espace est infini (réel ici).

Proposition : Si X_1, \dots, X_n sont des v.a. réelles et si f est une fonction continue de \mathbb{R}^n dans \mathbb{R} , alors $Y = f(X_1, \dots, X_n)$ est une variable aléatoire réelle.

→ Remarque : si $n=1$, on peut généraliser cette proposition avec des fonctions mesurables.

Définition : Une fonction $f: \mathbb{R} \rightarrow \mathbb{R}$ est mesurable si l'image réciproque par f d'un ensemble de la tribu borélienne de \mathbb{R} est un ensemble de la tribu borélienne.

a) Loi d'une variable aléatoire réelle:

Soit X une v.a. réelle - La loi \mathbb{P}_X de X est la probabilité définie sur \mathbb{R} muni de la tribu borélienne $\mathcal{B}(\mathbb{R})$, qui vérifie $\forall B \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B).$$

→ La loi d'une v.a. peut être caractérisée par sa fonction de répartition:

Définition : On appelle fonction de répartition de X la fonction

$$F_X(x) = \mathbb{P}_X([-\infty; x]) = \mathbb{P}(X \leq x), \quad \forall x \in \mathbb{R}.$$

Une fonction de répartition (FDR) n'est pas forcément continue - Par exemple, si \mathbb{P}_X est la mesure de Dirac en 0 (càd $X=0$ p.s.), la FDR de X est la mesure de Heaviside en 0 : $H(x) = 0$ si $x < 0$ et $H(x) = 1$ si $x \geq 0$. (il y a un saut).

Proposition : La FDR F_x caractérise la probabilité \mathbb{P}_x , et elle vérifie les trois conditions suivantes :

i) F_x est croissante

ii) F_x est continue à droite

iii) $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$

→ Par sa définition, le fait que $F = F_x$ caractérise \mathbb{P}_x est "évident" en effet,

$$\mathbb{P}_x([x, y]) = F(y) - F(x) \text{ pour tout } x < y.$$

$$\text{Si } B = \bigcup_{i=1}^n [x_i, y_i] \text{ avec } x_i < y_i < x_{i+1}, \text{ on a } \mathbb{P}_x(B) = \sum_{i=1}^n \mathbb{P}_x([x_i, y_i])$$

$$= \sum_{i=1}^n (F(y_i) - F(x_i))$$

car les intervalles sont disjoints.

→ Comme F est \uparrow , elle admet une limite à gauche en chaque point, notée $F(x^-)$.

On dit que F est une fonction càdlàg (l'ensemble des fct càdlàg est appelé Espace de Skorokhod). En remarquant que $]-\infty, y[= \lim_{n \rightarrow \infty}]-\infty, y_n[$ si $y_n \rightarrow y$, on obtient que pour $x < y$:

$$\bullet \mathbb{P}_x([x, y]) = \mathbb{P}(x < X \leq y) = F(y) - F(x) = \mathbb{P}(X \leq y) - \mathbb{P}(X \leq x)$$

$$\bullet \mathbb{P}_x([x, y[) = \mathbb{P}(x < X < y) = F(y^-) - F(x) = \mathbb{P}(X < y) - \mathbb{P}(X \leq x)$$

$$\bullet \mathbb{P}_x([x, y]) = \mathbb{P}(x \leq X \leq y) = F(y) - F(x^-) = \mathbb{P}(X \leq y) - \mathbb{P}(X < x)$$

$$\bullet \mathbb{P}_x([x, y[) = \mathbb{P}(x \leq X < y) = F(y^-) - F(x^-) = \mathbb{P}(X < y) - \mathbb{P}(X < x)$$

$\Rightarrow \mathbb{P}_x(\{x\}) = F(x) - F(x^-)$ est le saut de la fonction au point x .

\Rightarrow On a aussi le résultat :

Proposition : F est continue en $x \Leftrightarrow \mathbb{P}_x(\{x\}) = \mathbb{P}(X=x) = 0$.

Théorème : Si F est une fonction réelle sur \mathbb{R} vérifiant i), ii) et iii) de la proposition précédente, c'est la FDR d'une (unique) probabilité μ sur \mathbb{R} muni de sa tribu borélienne. On ne peut pas, en général, définir μ sur la tribu $\mathcal{P}(\mathbb{R})$ de toutes les parties de \mathbb{R} . (7)

→ Mesure de Lebesgue: prenons la fonction F telle que $F(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x \geq 1 \end{cases}$.
 F satisfait les hypothèses du théorème ci-dessus \Rightarrow il existe une unique probabilité.

Corollaire : Si $\mathcal{B}([0,1])$ désigne la tribu borélienne sur $[0,1]$, il existe une unique probabilité λ sur $([0,1], \mathcal{B}([0,1]))$ telle que la fonction de répartition soit égale à l'identité : $\lambda([x,y]) = y-x, \forall x < y, x,y \in [0,1]$.

On remarque qu'il s'agit de la longueur de l'intervalle.

On remarque aussi que $\mu(\mathbb{R}) = \lambda([0,1]) = 1$.

Cette (fonction) FDR $F = \text{Id}$ étant continue, elle ne charge pas les points : $\lambda(\{x\}) = 0 \quad \forall x \in [0,1]$.
 λ est appelé mesure de Lebesgue.

b) Variables aléatoires admettant une densité

Dans ce qui suit, nous parlerons de fonctions intégrables : on parle d'intégrabilité au sens de Riemann ou de Lebesgue (ce qui est plus général) - cela suppose que l'intégrale généralisée $\int_{-\infty}^{+\infty} |f(x)| dx$ est finie, et donc que $\int_{-\infty}^{+\infty} f(x) dx$ existe.

Nous verrons plus loin qu'une intégrale de Riemann est fondamentalement différente d'une intégrale de Lebesgue, au moins dans l'abord du calcul.

Définition : Une fonction réelle f définie sur \mathbb{R} est une densité de probabilité,

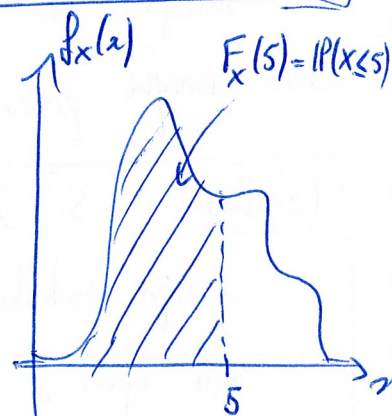
- si :
- elle est positive; $\forall x \in \mathbb{R}, f(x) \geq 0$
 - elle est intégrable: $\int_{-\infty}^{+\infty} f(x) dx < +\infty$
 - $\int_{-\infty}^{+\infty} f(x) dx = 1$.

Si f respecte ces conditions, alors $F(x) = \int_{-\infty}^x f(y) dy$ vérifie i), ii) et iii), et est donc la FDR d'une probabilité.

Définition : On dit que X a une loi de densité f si

P_X admet la densité f , et donc si $\forall x \in \mathbb{R}$,

$$P(X \leq x) = \int_{-\infty}^x f(y) dy.$$



Proposition : Soit X de loi de densité f . Alors :

- Sa FDR F est continue, d'où $P(X=x) = 0 \forall x \in \mathbb{R}$.

- La fonction F est dérivable en tout point x où f est continue, $F'(x) = f(x)$.

Inversement, si F est dérivable, ou seulement continue partout et dérivable par morceaux, alors X admet la densité $F'(x) = f(x)$.

→ Il existe des variables aléatoires qui n'ont pas de densité :

• les v.a. discrètes !

• les "mixtes" : on se donne une fonction $f \geq 0$, intégrable et d'intégrale strictement > 0 , et d'autre part une partie finie ou dénombrable I de \mathbb{R} (non vide) et des $q_i > 0$ indexés par $i \in I$, tels que $\int_{-\infty}^{+\infty} f(x) dx + \sum_{i \in I} q_i = 1$.

Alors $F(x) = \int_{-\infty}^x f(y) dy + \sum_{\substack{i \in I: \\ i \leq x}} q_i$ est une FDR, mais P_X associée n'admet pas de densité ! (mais n'est pas du tout non flou !).

c) Espérance des v.a. réelles: construction

On peut généraliser la notion d'espérance vue précédemment à toutes les v.a. réelles "suffisamment petites" (par exemple les v.a. bornées). Dans l'idée, on approche X par une suite de v.a. prenant un nombre fini de valeurs.

Cette suite de v.a. est obtenue par un découpage de l'espace d'arrivée (intégral de Lebesgue), bien connu pour une v.a. réelle (c'est \mathbb{R} !), à défaut de l'espace de départ (intégration de Riemann), Ω , très mal connu.

Définition : Une v.a. étagée est toute v.a. Y qui ne prend qu'un nombre fini de valeurs, notées ici a_1, \dots, a_p . Y admet donc une espérance donnée par

$$E[Y] = \sum_{i=1}^p a_i P(Y = a_i).$$

→ Pour définir l'espérance d'une v.a. réelle X , on s'intéresse d'abord à l'espérance des v.a. positives. Si X est une v.a. > 0 , on peut considérer une suite $(X_n)_n$ de v.a. > 0 étagées croissant vers X . Par exemple,

$$X_n(\omega) = \begin{cases} \frac{k}{2^n} & \text{si } \frac{k}{2^n} \leq \overset{>0}{X(\omega)} < \frac{k+1}{2^n} \text{ et } 0 \leq k \leq n2^n - 1. \\ n & \text{sinon} \end{cases}$$

→ on découpe l'espace d'arrivée ! (en effet $X(\omega)$ est dans l'espace d'arrivée).

Comme $X_n \leq X_{n+1}$, on a $E[X_n] \leq E[X_{n+1}]$.

La suite est croissante, donc $E[X] = \lim_{n \rightarrow \infty} E[X_n]$.

Cette limite existe toujours, elle est positive mais peut être infinie. Elle ne dépend pas du choix de (X_n) (à partir du moment où ce sont des v.a. > 0 , étagées, et \uparrow vers X).

Généralisons à X de signe quelconque - l'astuce est d'écrire :

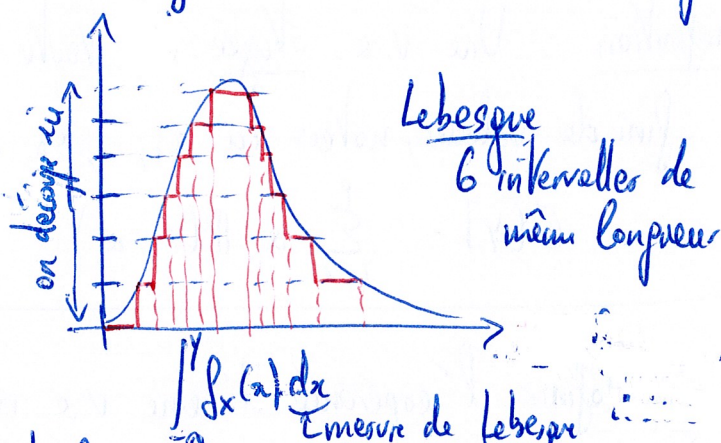
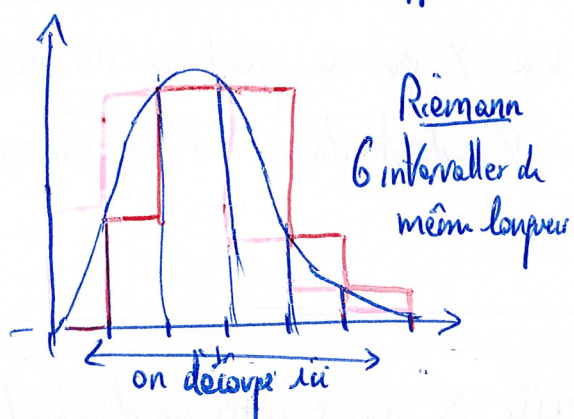
$$X = X^+ - X^- \quad \text{avec} \quad \begin{cases} X^+ = \sup(X, 0) & : \text{partie positive de } X \\ X^- = \sup(-X, 0) & : \text{partie négative de } X \end{cases}$$

On a $|X| = X^+ + X^-$, avec X^+ et X^- deux v.a. positives.

Définition : La v.a. X est intégrable si les valeurs $E[X^+]$ et $E[X^-]$ sont finies.

Dans ce cas, $E[X] = E[X^+] - E[X^-]$, ou $E[X] = \int X(\omega) P(d\omega)$.

→ On peut illustrer la différence entre l'intégration de Riemann et de Lebesgue :



⇒ Dans le cas Lebesgue, ce qui définit la longueur des intervalles sur lesquels on va intégrer est la mesure de probabilité de X .

→ Les propriétés de l'espérance énoncées dans le cadre d'un espace d'états fini ou dénombrable demeurent dans le cas des v.a. réelles en ayant construit l'espérance ainsi.

→ On peut introduire les espaces L^1 et L^2 comme pour le cas fini ou dénombrable.

On remarque que si X et Y sont dans L^2 , la v.a. XY est dans L^1 . En effet, car $|XY| \leq \frac{1}{2}(X^2 + Y^2)$ - On peut alors définir :

Définition : Si X et $Y \in L^2$, la v.a. $(X - E[X])(Y - E[Y])$ est intégrable. On appelle covariance de X et Y l'espérance de cette v.a., où :

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

→ Si X est une variable aléatoire d'espérance $E[X]$, de carré intégrable, et d'écart-type $\sigma_X > 0$, alors la v.a. $\frac{X - E[X]}{\sigma_X}$ est d'espérance nulle et d'écart-type 1 : elle est centrée réduite. (9)

Proposition : Soient X_1, \dots, X_n des variables aléatoires de carré intégrable, alors il en est de même de la somme, et

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j).$$

→ Espérance pour une v.a. à densité : pour une v.a. $X : \Omega \rightarrow \mathbb{R}$, de loi P_X , on a vu que $h(X)$ est une v.a. sur Ω si h est mesurable de $\mathbb{R} \rightarrow \mathbb{R}$. La fonction h est intégrable si $h \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$.

On a :

Proposition : Sous les hypothèses ci-dessus, h est intégrable $\Leftrightarrow h(X) \in L^1(\Omega, \mathcal{F}, P)$.

Dans ce cas, les espérances, respectivement de h par rapport à P_X et de $h(X)$ par rapport à P , sont égales : $E[h(X)] = \int_{\Omega} h(X(\omega)) P(d\omega) = \int_{\mathbb{R}} h(x) P_X(dx)$.

Proposition : Soit X une v.a. réelle de loi à densité f , et g une fonction mesurable de \mathbb{R} dans \mathbb{R} . Alors :

$$g(X) \in L^1 \Leftrightarrow \int_{-\infty}^{+\infty} |g(x)| f(x) dx < \infty.$$

Dans ce cas, $E[g(X)] = \int_{-\infty}^{+\infty} g(x) f(x) dx$.

II - Inégalités célèbres.

① - Inégalité de Bienaymé-Chebyshev.

Théorème: Soit $p \geq 1$ et $a > 0$. On a les inégalités:

- inégalité de Markov: soit $X \in L^p$, alors $P(|X| > a) \leq \frac{E[|X|^p]}{a^p}$
- inégalité de Bienaymé-Chebyshev: soit $X \in L^2$, $P(|X - E[X]| \geq a) \leq \frac{\text{Var}(X)}{a^2}$.

Preuve: On a $|X|^p \geq a^p \mathbb{1}_{[a, \infty[}(|X|)$ naturellement, donc

$$E[|X|^p] \geq a^p E[\mathbb{1}_{[a, \infty[}(|X|)] = a^p P(|X| \geq a). \quad (\text{Markov}).$$

On voit aisément que la 2^e inégalité est un cas particulier de la première.

→ Ces inégalités sont des inégalités de concentration: elles donnent une information sur le comportement d'une v.a. autour d'un point (comme l'espérance).

→ L'inégalité de Bienaymé-Chebyshev est très utile: elle permet de mesurer la probabilité de grands écarts à la moyenne: par exemple en prenant $a = 10 \sigma_x$, la proba. qu'une var. dévie de $E[X]$ de plus de 10 fois son écart-type est inférieure à 0,01 (donc c'est improbable).

→ Toutefois, l'inégalité de Bienaymé-Chebyshev est peu précise! Par exemple, en notant $m = E[X]$ et $\sigma^2 = \text{Var}(X)$, $\forall a > 0$, $P(|X - m| \geq a) \leq \frac{\sigma^2}{a^2}$.

Donc $P(|X - m| \geq a\sigma) \leq \frac{1}{a^2} \Rightarrow P(|X - m| < a\sigma) = P(X \in]m - a\sigma; m + a\sigma[) \geq 1 - \frac{1}{a^2}$.

A.N: pour $a = 2$, on obtient $P(X \in]m - 2\sigma; m + 2\sigma[) \geq 1 - \frac{1}{4} \approx 0,75$.

Cette approximation est universelle, mais très grossière. Par exemple, pour une loi normale, on sait que cette proba avoisine 0,95.

② - Inégalité de Cauchy-Schwarz:

(10)

Proposition: Soient X et Y deux v.a. de carré intégrable, alors $XY \in L^1$.

On a: $|E[XY]| \leq E[|XY|] \leq \sqrt{E[X^2]E[Y^2]}$

Le cas d'égalité correspond au cas où les v.a. sont presque sûrement proportionnelles.

Preuve: Comme $|XY| \leq \frac{1}{2}(X^2 + Y^2)$, on a $XY \in L^1$ dès que $X, Y \in L^2$. En effet le membre de droite est intégrable car de carré intégrable, donc le membre de gauche était inférieur il l'est aussi. De plus, $\forall x \in \mathbb{R}$, par linéarité et positivité de l'espérance: $x^2 E[X^2] + 2x E[XY] + E[Y^2] = E[(xX + Y)^2] \geq 0$. Mais ce n'est possible que si ce trinôme en x a au plus une seule racine réelle. \Rightarrow Son discriminant doit donc être négatif ou nul, ce qui donne l'inégalité. De plus, le discriminant est nul \Leftrightarrow il y a une racine double x_0 , et donc $Y(\omega) = -x_0 X(\omega)$ pour presque tout ω .

\rightarrow On peut déduire de cette inégalité que le coefficient de corrélation, $\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$ vérifie $-1 \leq \rho(X, Y) \leq 1$.

③ - Inégalité de Jensen.

Théorème: Soit X une v.a. réelle intégrable, et f une fonction mesurable telle que $f(X)$ soit intégrable. On suppose de plus que f est convexe. Alors $E[f(X)] \geq f(E[X])$.

Preuve: Puisque f est convexe, on a: $\forall a \in \mathbb{R}, \exists \lambda_a \in \mathbb{R}$ t.p. $\forall x, f(x) \geq f(a) + \lambda_a(x - a)$. On applique ce résultat avec $x = X(\omega)$ et $a = E[X]$, puis on prend l'espérance.

→ On peut aisément appliquer cette inégalité pour obtenir certains résultats pratiques:

• Si X est une v.a. r.g. $E[X] = 0$, alors $\forall \lambda \in \mathbb{R}, E[e^{\lambda X}] \geq 1$.

• Soit $G \sim \mathcal{N}(0,1)$. Alors $\forall x, \sigma, K \in \mathbb{R}$, on a:

$$E\left[\left(xe^{\sigma G - \frac{\sigma^2}{2}} - K\right)^+\right] \geq (x-K)^+, \text{ où } (x-K)^+ = \sup(0, x-K).$$

Cette inégalité est bien utile en finance. (ici, la fonction $x \mapsto (x-K)^+$ est convexe, et $E[e^{\sigma G - \frac{\sigma^2}{2}}] = 1$).

III - Vecteurs aléatoires

On peut généraliser les concepts déjà vus au cas où la v.a. peut prendre ses valeurs dans \mathbb{R}^n . La tribu borélienne sur \mathbb{R}^n est la tribu engendrée par les ensembles de la forme $\prod_{i=1}^n]-\infty, x_i]$ pour $x_i \in \mathbb{R}$. On munit \mathbb{R}^n de cette tribu.

→ Une fonction mesurable $f: \mathbb{R}^n \rightarrow \mathbb{R}$ est dite intégrable si

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} |f(x_1, \dots, x_n)| dx_1 \dots dx_n < +\infty.$$

Théorème: Soit f une fonction mesurable sur \mathbb{R}^2 .

1) Théorème de Tonelli: Si f est positive alors $\int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y) dy \right) dx = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y) dx \right) dy$

2) Théorème de Fubini: Si f est de signe quelconque mais vérifie:

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(x,y)| dy \right) dx < +\infty \quad \text{ou} \quad \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(x,y)| dx \right) dy < +\infty, \text{ alors on a le}$$

même résultat et on peut en plus définir $\int_{\mathbb{R}} \int_{\mathbb{R}} f(x,y) dx dy$ comme valeur commune.

Théorème : On considère une suite de fonctions $(f_k)_k$ telle que $\sum_k \int_{\mathbb{R}} |f_k(x)| dx < +\infty$,

$$\text{Alors } \sum_k \int_{\mathbb{R}} f_k(x) dx = \int_{\mathbb{R}} \left(\sum_k f_k(x) \right) dx.$$

Ces résultats vont se révéler utiles dans un cadre multidimensionnel. On peut aussi généraliser la mesure de Lebesgue (qui devient une aire en dim. 2, un volume en dim. 3, ...) et la notion d'équiprobabilité vue sur $[a, b]$.

→ En effet, la probabilité uniforme sur un ensemble borélien V de \mathbb{R}^n de mesure de Lebesgue $\lambda(V) \in]0, +\infty[$ est définie pour tout borélien $A \subset V$ par

$$P_V(A) = \lambda(A) / \lambda(V).$$

① - Introduction

Définition : Un vecteur aléatoire X à valeurs dans \mathbb{R}^n est formé de n variables aléatoires réelles qui sont les composantes de X : $X = (X_1, \dots, X_n)$.

→ La loi du vecteur aléatoire est caractérisée par sa FDR : on a

$$F: \mathbb{R}^n \rightarrow [0, 1]$$

$$(x_1, \dots, x_n) \mapsto F(x_1, \dots, x_n) = P\left(X \in \prod_{i=1}^n]-\infty, x_i]\right).$$

→ On dit que X admet la densité f si :

- $f \geq 0$ sur \mathbb{R}^n

- f est intégrable

- $\int_{\mathbb{R}^n} f(x) dx = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, \dots, x_n) dx_1 \dots dx_n = 1.$

- $P(X_1 \leq x_1, \dots, X_n \leq x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(y_1, \dots, y_n) dy_1 \dots dy_n.$

→ On peut aussi généraliser le calcul de l'espérance:

Proposition: Soit X une v.e. à valeurs dans \mathbb{R}^n , admettant la densité f ,
et g une fonction mesurable de \mathbb{R}^n dans \mathbb{R} .

Alors $g(X)$ est intégrable ^{Lebesgue} $\Leftrightarrow \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} |g(x_1, \dots, x_n)| f(x_1, \dots, x_n) dx_1 \dots dx_n < +\infty$.

Dans ce cas, nous avons

$$E[g(X)] = \int_{\mathbb{R}} \dots \int_{\mathbb{R}} g(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \dots dx_n.$$

② - Moments d'un vecteur aléatoire.

Définition: Si les composantes X_i du vecteur aléatoire $X = (X_1, \dots, X_n)$ sont intégrables,
on peut définir le vecteur espérance $E[X] = (E[X_1], \dots, E[X_n])$.

• Si ces composantes sont de carré intégrable, on définit la matrice de covariance de X , notée $C_X = (c_{ij})_{i,j \in \{1, \dots, n\}}$ de taille $n \times n$, par

$$c_{ij} = \text{Cov}(X_i, X_j).$$

Proposition: La matrice de covariance est symétrique non-négative.

Preuve: Symétrie: évident car $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$. $\forall i, j \in \{1, \dots, n\}$.

Non-négative signifie que $\sum_{i=1}^n \sum_{j=1}^n a_i a_j c_{ij} \geq 0$. On peut montrer simplement
que $\sum \sum a_i a_j c_{ij} = \text{Var}\left(\sum_{i=1}^n a_i X_i\right)$, or une variance est ≥ 0 . (car $\text{Var}(X_i) = \text{Cov}(X_i, X_i)$)

→ Transformation linéaire :

(12)

Proposition : Soit X un vecteur aléatoire n -dimensionnel, de matrice de covariance C - Soit A une matrice de taille $m \times n$, et Y le vecteur m -dimensionnel $Y = AX$.

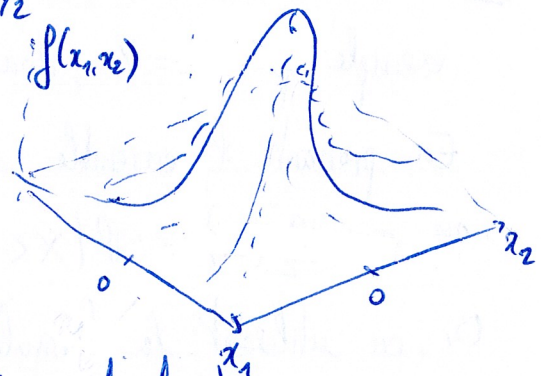
La matrice de covariance de Y vaut $AC^T A$, ou A^T est la transposée de A .

→ Exemple d'un vecteur gaussien n -dimensionnel : soit $m \in \mathbb{R}^n$ et C une matrice symétrique positive. Le vecteur $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ est un vecteur aléatoire gaussien d'espérance m et de matrice de covariance C si sa densité s'écrit :

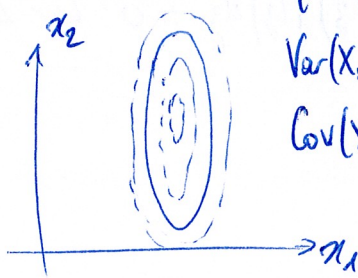
$$f(x) = f(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} \sqrt{\det(C)} e^{-\frac{1}{2} \langle x-m, \frac{1}{C} \langle x-m \rangle}$$

où $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$.

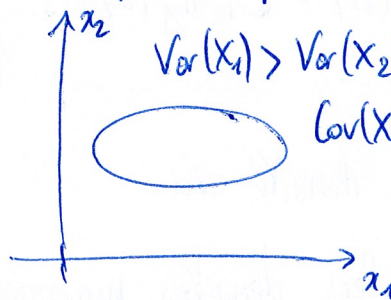
⇒ Illustration avec $n=2$;
et $m = (0, 0)$



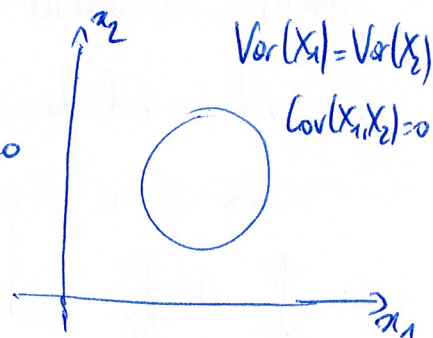
• On peut aussi représenter cette densité avec une vue "par-dessus" ; ce qui donnera des ellipsoïdes par niveau de densité.



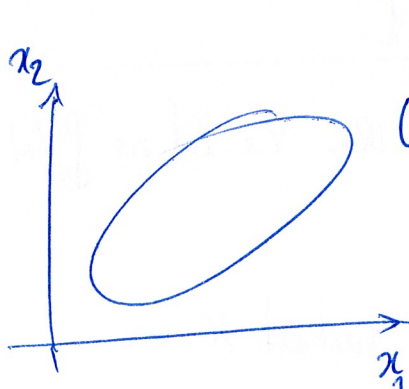
$\text{Var}(X_2) > \text{Var}(X_1)$
 $\text{Cov}(X_1, X_2) = 0$



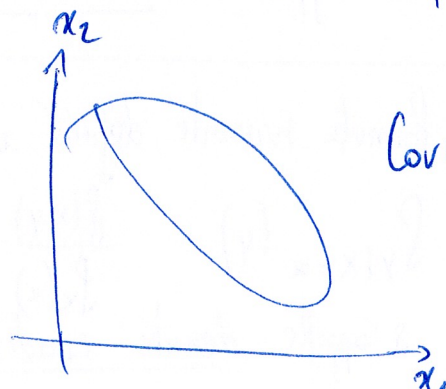
$\text{Var}(X_1) > \text{Var}(X_2)$
 $\text{Cov}(X_1, X_2) = 0$



$\text{Var}(X_1) = \text{Var}(X_2)$
 $\text{Cov}(X_1, X_2) = 0$



$\text{Cov}(X_1, X_2) > 0$



$\text{Cov}(X_1, X_2) < 0$

③ - Densités marginales et conditionnelles.

Pour simplifier, on prendra $n=2$.

On considère une v.a. Z à valeurs dans \mathbb{R}^2 , de loi à densité. On note X et Y ses composantes, d'où $Z = (X, Y)$. On étend maintenant le cas discuté.

Proposition : Supposons que Z admette une densité f . Alors X et Y admettent les densités f_X et f_Y sur \mathbb{R} , données par :

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy \quad \text{et} \quad f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx.$$

⚠ : X et Y peuvent avoir des densités sans que Z n'en ait : supposons par exemple que $X=Y$, donc $Z = (X, Y) = (X, X)$.

En prenant l'ensemble $\Delta = \{(x, x) : x \in \mathbb{R}\}$ (la diagonale/bissectrice dans \mathbb{R}^2), on a $P_Z(\Delta) = P(X < +\infty, Y < +\infty) = P(X < +\infty, Y < +\infty) = 1$.

Or, en utilisant la formule de calcul de l'espérance si la loi de Z était à densité, on aurait $P_Z(\Delta) = E[1_\Delta(Z)] = \int_{\mathbb{R}} \int_{\mathbb{R}} 1_\Delta(z) f(z) dz = 0$ car Δ a un volume nul dans \mathbb{R}^2 .

$\Rightarrow Z$ n'a pas de densité ii.

$\rightarrow f_X$ et f_Y sont appelées densités marginales de f .

Proposition : La formule suivante définit une densité sur \mathbb{R} , $\forall x$ tel que $f_X(x) > 0$:

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

Cette fonction s'appelle densité conditionnelle de Y sachant $X=x$.

Preuve : immédiat car $f_{Y|X=x}$ est une fonction positive, d'intégrale 1.

④ - Espérance conditionnelle.

(13)

Puisque $f_{Y|X=x}$ est une densité, on peut définir l'espérance qui lui est associée.

Définition: Soit Y une v.a. intégrable.

- L'espérance conditionnelle de Y sachant $X=x$ vaut $E[Y|X=x] = \int_{\mathbb{R}} y f_{Y|X=x}(y) dy$.
- On peut noter comme dans le cas discret:

$$E[Y|X] = \Psi(X) \text{ avec } \Psi(x) = E[Y|X=x].$$

C'est donc une variable aléatoire.

→ On retrouve l'ensemble des propriétés énoncées dans le cas discret, par exemple:
(si Y est intégrable)

- $E[Y] = E[E[Y|X]] = \int_{\mathbb{R}} E[Y|X=x] f_X(x) dx$

- Si $Y \geq 0$ alors $E[Y|X] \geq 0$.

- $E[1|X] = 1$.

- Si g est mesurable positive ou bornée, $E[Y g(X)|X] = g(X) E[Y|X]$.

- Pour toute fonction h mesurable positive ou bornée sur \mathbb{R}^2 ,

$$E[h(X,Y)] = \int_{\mathbb{R}} \int_{\mathbb{R}} f(x,y) h(x,y) dx dy$$

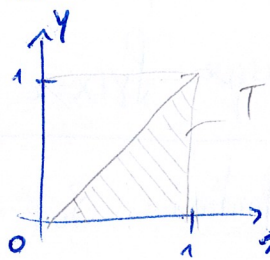
$$= \int_{\mathbb{R}} \int_{\mathbb{R}} h(x,y) f_X(x) f_{Y|X=x}(y) dx dy$$

$$= \int_{\mathbb{R}} \int_{\mathbb{R}} h(x,y) f_Y(y) f_{X|Y=y}(x) dx dy$$

→ Exemple: Soit X et Y de densité jointe $f_{(X,Y)}(x,y) = \frac{1}{x} 1_T(x,y)$,
où T est le triangle $T = \{0 < y < x < 1\}$.

La densité de X est donnée par

$$f_X(x) = \int f_{(X,Y)}(x,y) dy = \int_0^1 \frac{1}{x} 1_T(x,y) dy = \frac{1}{x} \int_0^x dy$$



donc $f_X(x) = \frac{1}{x} [y]_0^x = 1_{]0,1[}(x)$. Ainsi, pour $x \in]0,1[$,

$$f_{Y|X=x}(y) = \frac{f_{(X,Y)}(x,y)}{f_X(x)} = \frac{\frac{1}{x} 1_T(x,y)}{1_{]0,1[}(x)} = \frac{1}{x} 1_{]0,x[}(y).$$

$\Rightarrow X \sim \text{Unif}(]0,1[)$ et $Y|X=x \sim \text{Unif}(]0,x[)$ ($0 < x < 1$).

$\Rightarrow E[Y|X=x] = \frac{x}{2}$ et donc $E[Y|X] = \frac{X}{2}$.

(4) - Variables aléatoires indépendantes: cas continu.

On considère un couple (X,Y) de vecteurs aléatoires (X à valeurs dans \mathbb{R}^m ,
 Y à valeurs dans \mathbb{R}^n).

Définition: Les vecteurs aléatoires X et Y sont indépendants si pour tous boréliens A et B dans les espaces correspondants,
$$P(X \in A, Y \in B) = P(X \in A) P(Y \in B).$$

En considérant deux fonctions mesurables g et h définies sur \mathbb{R}^m et \mathbb{R}^n , on a:

Proposition: Avec les notations précédentes, et si X et Y sont \perp , les v.a. $g(X)$ et $h(Y)$ sont aussi \perp . Si de plus, $g(X)$ et $h(Y)$ sont dans L^1 , alors le produit $g(X)h(Y)$ est aussi intégrable, et nous avons:

$$E[g(X)h(Y)] = E[g(X)] E[h(Y)]$$

Il en découle que :

14

Proposition : Si les variables X et Y sont indépendantes et dans L^2 , alors $\text{Cov}(X, Y) = 0$ et $\rho(X, Y) = 0$ par suite.

A : Ce n'est pas parce que $\rho(X, Y) = 0$ que X et Y sont indépendantes.
Prendre par exemple $X \sim \mathcal{N}(0, 1)$ et $Y = X^2$.

Proposition : Soient X et Y deux v.a. réelles admettant les densités f_X et f_Y .
Pour qu'elles soient \perp , il faut et il suffit que $Z = (X, Y)$ admette la densité suivante sur \mathbb{R}^2 : $f(x, y) = f_X(x) f_Y(y)$.

⑤. Suite de vecteurs aléatoires \perp .

On se donne une famille finie X_1, \dots, X_n de vecteurs aléatoires à valeurs dans $\mathbb{R}^{p(i)}$, $1 \leq i \leq n$.

Définition : Les v.a. X_1, \dots, X_n sont \perp si pour tous boréliens A_1, \dots, A_n , on a :

$$\underbrace{P(X_1 \in A_1, \dots, X_n \in A_n)}_{\text{vecteur}} = \prod_{i=1}^n \underbrace{P(X_i \in A_i)}_{\text{vecteur}}.$$

→ Considérons maintenant une suite infinie : $(X_n)_{n \in \mathbb{N}^*}$.

Définition : La suite $(X_n)_{n \in \mathbb{N}^*}$ de v.a. est \perp si $\forall n$, la famille finie X_1, \dots, X_n est \perp .

→ Il s'ensuit que : l'indépendance de la suite $(X_n)_n$ entraîne celle de

- toute sous-suite $(X_{k_n})_{k_n}$
- toute " " de vecteurs issus de $(X_n)_n$.
- toute suite de la forme $(f_1(X_1), \dots, f_n(X_n))_n$ où les f_i sont continues (ou mesurables).

IV - Vecteurs gaussiens

Définition : Un vecteur aléatoire $X = (X_1, X_2, \dots, X_n)$ à valeurs dans \mathbb{R}^n est appelé un vecteur gaussien si toute combinaison linéaire $\sum_{j=1}^n a_j X_j = \langle a, X \rangle$, pour $a = (a_1, \dots, a_n) \in \mathbb{R}^n$, suit une loi normale (avec la convention que la mesure de Dirac au point m est la loi normale $\mathcal{N}(m, 0)$).

→ Bien entendu, cela implique que chaque composante X_j suit elle-même une loi normale. En revanche, avoir un vecteur composé de lois marginales gaussiennes ne garantit pas que c'est un vecteur gaussien !

→ Exemples :

- Si les X_i sont i.i.d., gaussiennes, alors le vecteur X est gaussien.

- Si les X_i ne sont pas i.i.d., ce n'est pas garanti : prendre $X_1 \sim \mathcal{N}(0, 1)$ et

$$X_2 = \begin{cases} X_1 & \text{si } |X_1| \leq 1, \\ -X_1 & \text{sinon} \end{cases}$$

Alors $X_2 \sim \mathcal{N}(0, 1)$, mais $X = (X_1, X_2)$ n'est pas un vecteur gaussien. En effet, $0 < \mathbb{P}(X_1 + X_2 = 0) < 1$, alors

que cette probabilité est nulle si le vecteur était gaussien.

Corollaire : Si X est un vecteur gaussien, alors ses composantes sont i.i.d. \Leftrightarrow la matrice de covariance est diagonale.

⚠ : Ce résultat peut être faux si X n'est pas gaussien ... (cf exemple ci-dessus).

Proposition : Soit X un vecteur gaussien à valeurs dans \mathbb{R}^n , de moyenne m . Il existe des v.a. réelles i.i.d. Y_1, \dots, Y_n , de lois normales $\mathcal{N}(0, \lambda_j)$ avec $\lambda_j \geq 0$ (si $\lambda_j = 0$ on convient que $Y_j = 0$) et une matrice orthogonale A telles que

$$X = m + AY, \text{ où } Y = (Y_1, \dots, Y_n)$$

Preuve: Comme C (la matrice de covariance du vecteur gaussien) est une matrice symétrique positive, il existe une matrice orthogonale A et une matrice diagonale Λ dont les éléments diagonaux vérifient $\lambda_j \geq 0$, et telle que $C = A \Lambda A^T$. Soit $Y = A^T(X - m)$, alors Y est un vecteur gaussien de covariance $C' = A^T C A = \Lambda$ et de moyenne nulle. Les composantes Y_j de Y répondent à la question.

Proposition: Le vecteur gaussien X admet une densité sur \mathbb{R}^n si et seulement si sa matrice de covariance C est non-dégénérée (ou inversible, ou de valeurs propres toutes strictement > 0).

Dans ce cas, cette densité vaut
$$f_X(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det(C)}} e^{-\frac{1}{2} \langle x - m, C^{-1} (x - m) \rangle}$$

Définition: On appelle variable aléatoire de χ^2 (chi-deux) à d degrés de liberté toute variable aléatoire positive qui a même loi que la somme de d carrés de v.a. de loi $\mathcal{N}(0,1)$ indépendantes.

→ Cette loi est très utilisée en statistique car elle permet de construire des tests d'indépendance ou d'adaptation à une loi.

→ Si $X \sim \mathcal{N}(0,1)$, alors $X^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2})$ (loi Gamma)

$\left\{ \begin{array}{l} \text{Si } X \sim \Gamma(\alpha, \theta) \\ Y \sim \Gamma(\beta, \theta) \\ X \perp Y \end{array} \right\} \Rightarrow \text{alors } U+V \sim \Gamma(\alpha+\beta, \theta)$

⇒ Grâce à ces 2 propriétés, on retrouve pour la loi du χ_d^2 (qui est un cas particulier de la loi Gamma): si $Y \sim \chi_d^2$ alors $\left\{ \begin{array}{l} E[Y] = d \\ \text{Var}(Y) = 2d \end{array} \right.$

$\Gamma(\frac{d}{2}, \frac{1}{2})$

Proposition : Soient n v.a. $(X_1, \dots, X_n) \perp$, de loi $\mathcal{P}(m, \sigma^2)$ - (i.i.d)

Considérons: $\bar{X} = \frac{X_1 + \dots + X_n}{n}$ et $V = \frac{1}{n} \left((X_1 - \bar{X})^2 + \dots + (X_n - \bar{X})^2 \right)$.

Alors

1) $\bar{X} \sim \mathcal{P}(m, \sigma^2/n)$

2) $\frac{n}{\sigma^2} \left((X_1 - \bar{X})^2 + \dots + (X_n - \bar{X})^2 \right) \sim \chi_{n-1}^2$

3) \bar{X} et V sont indépendantes.

Preuve: 1) Trivial.

2) Nous simplifions en nous ramenant au cas centré réduit. On pose pour chaque $1 \leq i \leq n$, $\xi_i = \frac{X_i - m}{\sigma}$, donc $\xi_i \sim \mathcal{P}(0, 1)$.

On peut réécrire V avec l'expression: $V = \frac{\sigma^2}{n} \left(\sum_{i=1}^n \xi_i^2 - n_1^2 \right)$ où $n_1 = \frac{\xi_1 + \dots + \xi_n}{\sqrt{n}}$.

Prendre une matrice U de taille $n \times n$, orthogonale, dont la 1^{ère} ligne est formée de coordonnées toutes égales à $\frac{1}{\sqrt{n}}$ (autres termes choisis arbitrairement).

Désignons par n_j ($1 \leq j \leq n$) les coordonnées du vecteur aléatoire obtenu en appliquant U au vecteur aléatoire (colonne) ξ , soit

$$\begin{pmatrix} n_1 \\ \vdots \\ n_n \end{pmatrix} = U \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix} \quad \text{Alors on a bien } n_1 = \frac{\xi_1 + \dots + \xi_n}{\sqrt{n}}.$$

Puisque U est orthogonale et que ξ est un vecteur gaussien formé de v.a. \perp , il en est de même pour n . De plus, on a $\sum_{i=1}^n n_i^2 = \sum_{i=1}^n \xi_i^2$, ce qui entraîne que $V = \frac{\sigma^2}{n} \left(\sum_{i=1}^n n_i^2 - n_1^2 \right) = \frac{\sigma^2}{n} \sum_{i=2}^n n_i^2$.

$$\Rightarrow \frac{n}{\sigma^2} V \sim \chi_{n-1}^2, \text{ avec } V \perp n_1, \text{ et donc } V \perp \bar{X}.$$