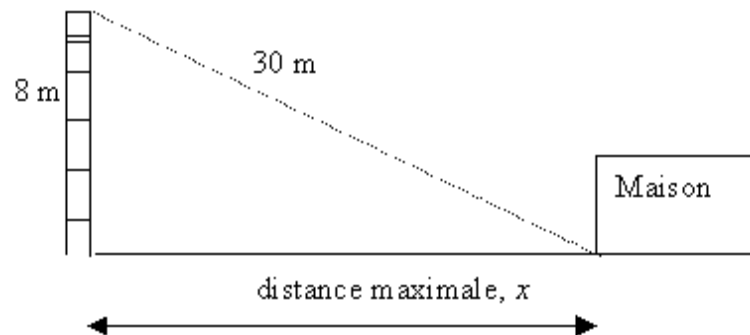


Préambule Processus aléatoires en recherche opérationnelle

1- Modélisation mathématique

L'objectif majeur des mathématiques appliquées est la mise en œuvre de *modèles mathématiques* destinés à décrire, représenter ou expliquer divers *phénomènes* rencontrés dans plusieurs domaines du savoir humain tels que la physique, la biologie, l'économie, la sociologie, etc. On entend par modèle mathématique une représentation simplifiée d'un phénomène sous étude, représentation soumise à des *hypothèses* simplificatrices au vu d'un objectif préalablement fixé. Par exemple, en physique l'équation d'un mouvement uniforme est un modèle mathématique du mouvement d'un mobile. L'équation est basée sur certaines hypothèses telles que la constance de la vitesse, l'absence de forces de frottement, etc. En chimie, la relation liant la pression, le volume et la température d'un système est un modèle qui n'est valable que sous certaines conditions de régularité. En biologie, la fonction exponentielle est un modèle de croissance de bactéries dans un milieu, en supposant entre autres que le taux de croissance est constant et qu'il n'y a pas de mortalité. Par ailleurs, l'élaboration d'un modèle mathématique est basée sur des outils mathématiques et un système scientifique rigoureux partant de l'observation du phénomène, de son analyse jusqu'à la validation du modèle correspondant. Une fois mis en œuvre, le modèle peut être exploité en vue de résoudre une problématique posée au préalable. Donnons-nous quelques exemples.

Exemple 1 Supposons qu'une personne désire bâtir une maison sur son propre terrain récemment acheté et déjà viabilisé (voir figure 1). Elle hésite sur l'endroit exact où elle va construire la maison parce qu'elle est contrainte par un endroit assez proche d'un poteau électrique de 8 mètres déjà installé à l'extrémité gauche du terrain et qui est destiné à alimenter la maison en électricité une fois prête. La société d'électricité Sonelgaz exige que le câble électrique menant l'extrémité du poteau à la base de la maison doive avoir une longueur au maximum égale à 30 mètres, longueur à partir de laquelle il y aura des pertes de charges. Autrement, la personne doit payer tous les frais relatifs à l'installation d'un autre poteau, ce qui lui revient très cher. C'est pourquoi elle cherche la distance maximale à partir de la quelle, elle ne peut construire sa maison.



De prime à bord, avant de chercher la valeur de cette distance maximale, la première chose qui vient à l'esprit est que l'on peut schématiser le problème sous forme d'un triangle rectangle

de deux cotés connus et dont le troisième est à déterminer. Ce schéma simplifié qu'est le triangle rectangle représente le modèle mathématique associé au problème de départ. Dans ce modèle on ne s'intéresse pas à la couleur du câble, à la matière du poteau ni à la hauteur de la maison. On a considéré seulement les données essentielles, soit un triangle rectangle dont deux cotés sont de longueur connue, pour la résolution du problème. On suppose par ailleurs que le fil est souple inélastique et bien tendu, ce qui constitue une autre hypothèse simplificatrice. Une fois le modèle établi, on pourra l'exploiter pour résoudre le problème qui consiste en la recherche de la distance maximale notée dorénavant x . Or, pour un triangle rectangle, il est bien connu que le théorème de Pythagore donne une relation liant les trois cotés du triangle. Dans ce cas, cette relation est donnée par l'équation suivante

$$8^2 + x^2 = 30^2,$$

équation du second degré admettant deux solutions, $x = \sqrt{836}$ et $x = -\sqrt{836}$ dont on ne retient que la première puisqu'une distance est évidemment positive. La deuxième solution est une solution du modèle et non du problème et sera donc rejetée.

Exemple 2 *Problème de gestion de production*

Une raffinerie produit chaque jour des essences, super (S) et normal (N), à partir de deux types de pétrole brut A et B. Les compositions des mélanges, les disponibilités en matières premières, les demandes en essences ainsi que les profits unitaires associés sont résumés dans le tableau suivant :

Type	Super (S)	Normal (N)	Disponibilité en pétrole brut
Pétrole Brut A	3	2	12
Pétrole Brut B	2	4	8
Demandes en essences	3	2	
Profits unitaires des essences		4	3

On demande de déterminer les quantités d'essence normale et d'essence super à produire chaque jour pour que le profit total de production soit le maximum possible.

Ceci est un problème classique de gestion de production. Notons x et y les quantités respectives de Super et de Normal à produire quotidiennement. Ces quantités sont des valeurs réelles non négatives, i.e. $x \geq 0$ et $y \geq 0$. Alors en supposant que les profits sont proportionnels à la quantité produite (c'en est une hypothèse simplificatrice), l'objectif sera de rendre maximum la fonction

$$z(x, y) = 4x + 3y,$$

en choisissant des valeurs adéquates pour x et y . Ces valeurs ne peuvent être quelconques car elles doivent respecter des contraintes de disponibilité et des contraintes de satisfaction de la demande. Concernant la disponibilité, la quantité de pétrole A utilisée pour produire x quantités de Super et y quantités de Normal, soit $3x + 2y$, ne doit pas excéder la disponibilité journalière du pétrole A qui est de 12. Ceci est traduit par l'inégalité suivante

$$3x + 2y \leq 12.$$

De même pour le pétrole B, la contrainte de sa disponibilité se traduit par

$$x + 2y \leq 4.$$

Concernant les contraintes de satisfaction de la demande, les quantités d'essence Super et de Normal doivent être égales au moins à 3 et 2 respectivement, i.e.

$$x \geq 3 \text{ et } y \geq 2.$$

Ainsi, le modèle représentatif à la problématique de la détermination des quantités optimales peut se mettre sous la forme d'un programme mathématique, linéaire en l'occurrence (PL), comme suit

$$(PL) \begin{cases} 4x + 3y = z(\max) \\ 3x + 2y \leq 12 \\ x + 2y \leq 4 \\ x \geq 3 \\ y \geq 2 \end{cases}$$

Le modèle ainsi mis en œuvre peut être utilisé pour déterminer les valeurs de x et y et par là même les quantités de Normal et de Super demandées. Il est entendu que les solutions trouvées du PL sont solutions du modèle qu'il faut adapter au problème sous-jacent.

Exemple 3 *Pendule simple*

Une masse m de dimension négligeable est suspendue à l'extrémité d'un fil souple inélastique dont l'autre extrémité étant fixe. On écarte la masse d'un angle θ par rapport à la position d'équilibre et on la lâche. On s'intéresse à *prédire* la valeur du temps T que met la masse pour retourner à sa position initiale pour la première fois. On détermine d'abord les facteurs influant sur T . Les physiciens stipulent qu'en supposant négligeable les force de frottement, T dépend de la longueur du fil l , de l'intensité de la pesanteur g et de l'angle d'écart θ selon une fonction f bien établie, où

$$T = f(l, g, \theta).$$

Pour des valeurs très faibles de θ le modèle associé s'exprime par

$$T = 2\pi\sqrt{l/g}, \quad (1.1.1)$$

qui est le modèle retenu pour le problème dit du pendule simple. La dernière relation peut être exploitée pour prédire la valeur de T lorsqu'on connaît les valeurs de l et de g .

Bien que les trois derniers exemples soient de nature très différente, ils ont le point commun de se formuler selon un cycle dont les étapes sont les mêmes. Ce cycle allant de l'analyse du problème à l'exploitation du modèle en passant par l'élaboration de celui-ci (voir figure 2) est dit cycle de modélisation mathématique.

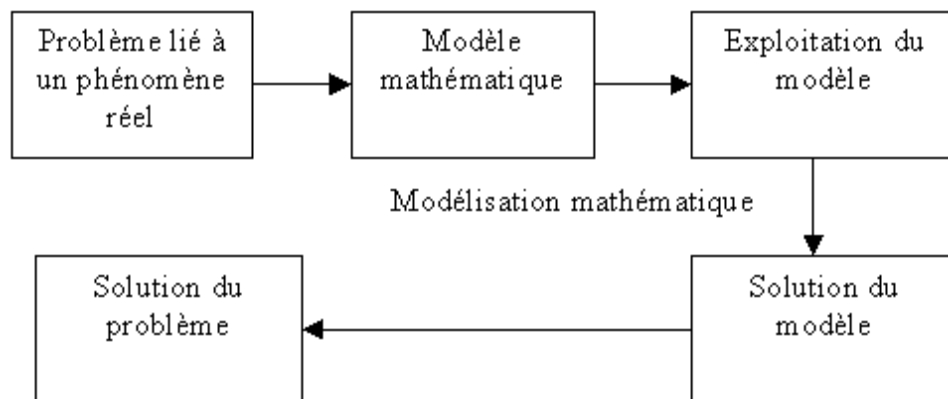


Figure 2 Méthodologie de modélisation mathématique

2- Phénomènes aléatoires en recherche opérationnelle

Les mathématiques appliquées peuvent être divisées en plusieurs branches en fonction du type du problème auquel elles sont destinées à résoudre. Aux problèmes de la physique correspondent les mathématiques appliquées à la physique, à ceux de nature économique correspond l'économétrie. Pour les problèmes rencontrés en biologie il y a la biométrie, etc. En outre, pour les problèmes de gestion, d'organisation et d'optimisation dans des structures complexes les mathématiques appliquées en offrent la *recherche opérationnelle*. Cette branche regroupe un ensemble de techniques et de méthodes mathématiques devant conduire à de meilleures politiques d'action et de décision face aux problèmes d'optimisation notamment dans les entreprises.

Or, les phénomènes étudiés par la recherche opérationnelle peuvent être regroupés en deux catégories : déterministes et stochastiques. Un phénomène est dit déterministe si, connaissant son état à une situation donnée, on peut déterminer complètement son état à toute situation possible. Le modèle correspondant est également dit déterministe. En outre, le mouvement de la terre peut être considéré comme déterministe car connaissant son état (ici position et vitesse) à un instant donné, on peut déterminer ses position et vitesse à tout instant ultérieur. L'équation régissant le mouvement de la terre est dite modèle déterministe. Par opposition, un phénomène est dit aléatoire (ou stochastique) si la connaissance de son état à une situation donnée ne détermine pas entièrement son état à une autre situation. Les modèles correspondants sont ainsi nommés aléatoires (ou *stochastiques*).

Dans l'exemple 2, le modèle proposé est déterministe car pour les mêmes conditions du problème le modèle fournira la même solution. Or, on a supposé que les paramètres du modèle tels que les prix, les disponibilités, et les demandes sont parfaitement connus. Pourtant, les demandes en essences ainsi que leurs prix ne peuvent être connus avec certitude, ainsi un modèle considérant ces quantités comme *aléatoires* sera plus réaliste.

Chapitre 1 Bagage mathématique

I- Théorie de la mesure et intégration

1) Théorie de la mesure

a- Notion de mesure La théorie de la mesure fournit les bases mathématiques pour mesurer des ensembles arbitraires, quelles que soient leurs formes ou leurs interprétations. Historiquement, elle a été développée dans le but de pouvoir assigner une grandeur à des ensembles physiquement interprétables, telles que la longueur d'un segment, l'air d'une surface dans un espace euclidien, la masse d'un corps physique et la charge électrique portée par une masse. Actuellement, la notion de mesure s'applique à des ensembles de structures plus complexes et de significations tout à fait abstraites. En fait, une mesure est une application qui associe à un ensemble un nombre positif exprimant sa "grandeur" selon un certain sens. Cette application est soumise à certaines règles édictées par l'intuition et la commodité. Cette définition est bien entendu incomplète puisque il reste encore à préciser avec soin l'ensemble de départ et l'ensemble d'arrivée correspondants à une telle application ainsi que la façon dont elle assigne les nombres à des ensembles.

Par analogie aux mesures physiques (longueur, volume, masse,...), une mesure abstraite que l'on désignera par $\mu(\cdot)$ est astreinte à satisfaire les conditions :

i) de positivité : $\mu(A) \geq 0$ pour tout ensemble A ,

ii) de nullité de l'ensemble vide : $\mu(\emptyset) = 0$, et

iii) d'additivité :

$$\mu(A \cup B) = \mu(A) + \mu(B), \quad (1.1.1)$$

pour tout A et B disjoints. La relation (1.1.1) s'étend évidemment au cas d'une réunion finie d'ensembles A_1, A_2, \dots, A_n deux à deux disjoints (en vertu de l'associativité de l'union). En effet,

$$\mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i). \quad (1.1.2)$$

Toutefois, il arrive que l'on veuille mesurer l'union dénombrable des ensembles A_1, A_2, \dots lorsqu'on connaît la mesure de chacun d'eux. Par exemple, le segment $[a, b[$ ($a < b$) de la droite réelle peut être considéré comme la réunion dénombrable des événements $A_n = [a, b - 1/n]$, $n \in \mathbb{N}^*$, i.e. $[a, b[= \bigcup_{i=1}^{\infty} [a, b - 1/i]$. Pourtant, la relation (1.1.2) n'entraîne pas nécessairement que la propriété d'additivité soit valable pour le cas de réunion infinie dénombrable. Donc, par commodité, on impose à une mesure de vérifier une telle condition qu'on qualifie de σ -additivité :

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i), \quad (1.1.3)$$

pour toute suite d'ensemble A_1, A_2, \dots deux à deux disjoints.

Par ailleurs, lorsqu'on écrit par exemple $\mu(A)$ ou $\mu\left(\bigcup_{i=1}^{\infty} A_i\right)$, les arguments A et $\bigcup_{i=1}^{\infty} A_i$ de la fonction ensembliste $\mu(\cdot)$ sont des éléments d'un ensemble de départ de $\mu(\cdot)$ qu'il faut spécifier avec soin. On convient que tous les ensembles que l'on veut mesurer sont des parties

d'un univers \mathcal{X} . L'ensemble de départ de $\mu(\cdot)$ sera une classe \mathcal{F} de parties de \mathcal{X} compatible avec les propriétés (1.1.1)-(1.1.3). En effet, si on a mesuré un ensemble A c'est que l'on a pu mesurer également son complémentaire A^c par rapport à \mathcal{X} et donc on a mesuré \mathcal{X} lui-même. Ainsi la classe \mathcal{F} , ensemble de départ de la mesure $\mu(\cdot)$, est astreinte à contenir \mathcal{X} , à contenir un ensemble A toutes les fois qu'elle contient son complémentaire A^c et à contenir la réunion dénombrable $\bigcup_{i \in I} A_i$ dès qu'elle contient A_i pour tout $i \in I$. Une telle classe sera dite algèbre ou σ -Algèbre (tribu) selon que I soit fini ou infini dénombrable.

Définition 1.1.1 (Tribu)

Soit \mathcal{X} un univers et \mathcal{F} une classe de parties de \mathcal{X} . \mathcal{F} est dite algèbre des parties de \mathcal{X} si elle vérifie

- i) $\mathcal{X} \in \mathcal{F}$,
- ii) si $A \in \mathcal{F}$ alors $A^c \in \mathcal{F}$,
- iii) si A_1, A_2, \dots, A_n sont des éléments de \mathcal{F} deux à deux disjoints alors $\bigcup_{i=1}^n A_i \in \mathcal{F}$.

De plus, si au lieu de la condition iii), \mathcal{F} vérifie plutôt la condition suivante,

- iv) si A_1, A_2, \dots est une suite d'éléments de \mathcal{F} deux à deux disjoints alors $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$,

alors \mathcal{F} est dite σ -Algèbre ou tribu des parties de \mathcal{X} .

Remarque 1.1.1

i) Il est clair que la condition iv) entraîne iii) et que la réciproque est fausse. Une tribu est donc une algèbre. La condition iv) est souvent formulée par : \mathcal{F} est stable par réunion dénombrable.

ii) Les ensembles membres de \mathcal{F} sont dit \mathcal{F} -mesurables ou tout court mesurables. \square

Remarque 1.1.2

On distingue deux σ -algèbres particulières : la σ -algèbre *triviale* $\{\emptyset, \mathcal{X}\}$, la plus petite dans le sens de l'inclusion et la σ -algèbre la plus large qu'est l'ensemble des parties de \mathcal{X} , $\mathcal{P}(\mathcal{X})$. Entre ces deux extrêmes, il existe plusieurs tribus intermédiaires. Si A est un ensemble non vide alors la plus petite tribu à laquelle appartient A est l'ensemble $\{\emptyset, A, A^c, \mathcal{X}\}$. Aussi, pour toute famille ξ de parties de \mathcal{X} , existe-t-il une tribu $\mathcal{B}(\xi)$, la plus petite contenant ξ . Pour $\mathcal{X} = \mathbb{R}$ la plus petite tribu, $\mathcal{B}(\mathbb{R})$, contenant tous les intervalles ouverts de \mathbb{R} est dite tribu de Borel. Elle est également la plus petite tribu générée par des intervalles de la forme $] - \infty, a]$, $a \in \mathbb{R}$. Chaque intervalle appartenant à une telle tribu est dit Borélien de \mathbb{R} . Cette notion s'étend aisément à \mathbb{R}^d , $d \in \mathbb{N}^*$. Il est clair que la tribu de Borel contient des ensembles de la forme $]a, b[, [a, b], [a, b[,]a, b],] - \infty, a[,]a, +\infty[$ et $[a, +\infty[$ ($a < b$). \square

Le couple $(\mathcal{X}, \mathcal{F})$, étant prêt à être mesuré par une mesure (bien entendu), mettons $\mu(\cdot)$, est ainsi dit *espace mesurable*. Sur un espace mesurable $(\mathcal{X}, \mathcal{F})$ il est donc possible de définir une mesure ayant comme ensemble de départ \mathcal{F} .

Définition 1.1.2 (Mesure)

Une fonction ensembliste $\mu(\cdot)$ définie sur \mathcal{F} à valeurs réelles positives (dans \mathbb{R}_+) est dite mesure si elle vérifie les conditions suivantes

- i) $\forall A \in \mathcal{F}$, $\mu(A) \in \mathbb{R}_+ \cup \{\infty\} := [0, \infty]$,
- ii) $\mu(\emptyset) = 0$
- ii) si A_1, A_2, \dots, A_n sont des éléments de \mathcal{F} deux à deux disjoints alors (1.1.3) est vérifiée.

Le triplet $(\mathcal{X}, \mathcal{F}, \mu)$ est dit espace mesuré ou espace de mesure.

La mesure $\mu(\cdot)$ est dite finie ou infinie selon que $\mu(\mathcal{X}) < \infty$ ou $\mu(\mathcal{X}) = \infty$. S'il existe une

suite d'ensembles A_1, A_2, \dots , telle que $\mathcal{X} = \bigcup_{i=1}^{\infty} A_i$ et $\mu(A_i) < \infty$ pour tout $i \in \mathbb{N}^*$ alors μ est dite σ -finie. Une mesure finie est donc σ -finie alors qu'une mesure σ -finie peut être finie ou infinie. Il est clair par ailleurs que si $\mu(\cdot)$ est une mesure alors, en vertu de (1.1.3) la condition (1.1.2) est systématiquement vérifiée.

Exemple 1.1.1 (Mesure de comptage)

Soit \mathcal{X} un ensemble arbitraire non vide, \mathcal{F} une tribu des parties de \mathcal{X} et $\mu(\cdot)$ une mesure définie sur $(\mathcal{X}, \mathcal{F})$ par

$$\begin{aligned} \mu(\cdot) : \mathcal{F} &\longrightarrow [0, \infty] \\ A &\longmapsto \mu(A) = \begin{cases} \text{card}(A) & \text{si } A \text{ est fini} \\ \infty & \text{si } A \text{ est infini} \end{cases} \end{aligned}$$

où $\text{card}(A)$ désigne le nombre d'éléments de A . La mesure $\mu(\cdot)$ est dite mesure de comptage. Elle est finie ou infinie selon que \mathcal{X} soit fini ou infini. $\mu(\cdot)$ est σ -finie si et seulement si \mathcal{X} est dénombrable. \square

Exemple 1.1.2 (Mesure de Lebesgue)

i) Soit $\mathcal{X} = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$, la plus petite tribu engendrée par les ensembles $[a, b]$, ($a < b \in \mathbb{R}$) et $\mu(\cdot)$ une mesure définie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ par

$$\begin{aligned} \mu(\cdot) : \mathcal{B}(\mathbb{R}) &\longrightarrow [0, \infty] \\ A = [a, b] &\longmapsto \mu([a, b]) = b - a \end{aligned}$$

avec $a < b$. La mesure $\mu(\cdot)$ est dite *mesure de Lebesgue*; intuitivement elle mesure la longueur d'un intervalle.

ii) On peut étendre la mesure de Lebesgue sur un espace euclidien de dimension finie, soit \mathbb{R}^d ($d \in \mathbb{N}^*$).

Dans ce cas la tribu de Borel associée à \mathbb{R}^d peut être formulée comme la plus petite tribu engendrée par les rectangles ouverts

$$A = \{(x_1, x_2, \dots, x_d) : a_i < x_i < b_i, i = 1, \dots, d\},$$

où $-\infty < a_i < b_i < +\infty$. La mesure de Lebesgue est alors définie sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ par

$$\begin{aligned} \mu(\cdot) : \mathcal{B}(\mathbb{R}^d) &\longrightarrow [0, \infty] \\ A &\longmapsto \mu(A) = (b_1 - a_1) \dots (b_d - a_d), \end{aligned}$$

représentant le volume de A .

iii) On démontre que tout ensemble fini ou dénombrable est de mesure (de Lebesgue) nulle : on dit qu'un tel ensemble est μ -négligeable. En particulier, \mathbb{N} est négligeable puisque :

$\mathbb{N} = \bigcup_{i=0}^{\infty} \{i\} = \bigcup_{i=0}^{\infty} [i, i]$ et alors $\mu(\mathbb{N}) = \mu\left(\bigcup_{i=0}^{\infty} [i, i]\right) = \sum_{i=0}^{\infty} \mu([i, i]) = 0$. De même, on démontre aisément que \mathbb{Q} , quoique partout dense dans \mathbb{R} , est μ -négligeable : $\mu(\mathbb{Q}) = 0$. \square

Propriétés d'une mesure

Soient $\mu(\cdot)$ une mesure et une \mathcal{F} tribu.

i) *Monotonie* : pour tout $A, B \in \mathcal{F}$ tels que $A \subset B$ alors

$$\mu(A) \leq \mu(B).$$

- ii) Continuité vers le haut : Si A_1, A_2, \dots est une suite croissante d'ensembles ($A_n \subset A_{n+1}$, $\forall n$) de \mathcal{F} , convergeant vers $A \in \mathcal{F}$, alors $\mu(A_n) \rightarrow \mu(A)$ lorsque $n \rightarrow \infty$.
- iii) Continuité vers le bas : Si A_1, A_2, \dots est une suite décroissante d'ensembles ($A_{n+1} \subset A_n$, $\forall n$) de \mathcal{F} , convergeant vers $A \in \mathcal{F}$ et si $\mu(A) < \infty$ alors $\mu(A_n) \rightarrow \mu(A)$ lorsque $n \rightarrow \infty$.
- iv) Sous-additivité : Si A_1, A_2, \dots est une suite d'ensembles de \mathcal{F} alors

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu(A_i).$$

Remarque 1.1.3 (La propriété Presque Partout)

i) Il arrive qu'une propriété soit vérifiée pour tous les éléments de l'univers \mathcal{X} sauf pour les éléments d'un ensemble N de mesure nulle ($\mu(N) = 0$). Alors on dit que la propriété est vérifiée presque partout (ou tout court p.p.) par rapport à la mesure considérée.

Par exemple si f est une fonction sur \mathbb{R} par :

$$f(x) = \begin{cases} 1 & \text{si } x = 3 \\ 0 & \text{sinon} \end{cases},$$

on dira alors que f est nulle presque partout par rapport à la mesure de Lebesgue car elle est non nulle sauf sur l'ensemble $\{3\}$ qui est, par la mesure de Lebesgue, de mesure nulle. \square
 Pour un espace de mesure $(\mathcal{X}, \mathcal{F}, \mu)$, il se peut qu'une partie d'un ensemble A , A étant de mesure nulle, ne soit pas dans \mathcal{F} et n'est donc pas mesurable, ce qui peut induire une difficulté d'un point de vue théorique. Par exemple lorsqu'une propriété $\mathbb{P}(x)$ est vraie pour tout x dans \mathcal{X} sauf dans un ensemble N_0 , partie d'un ensemble N de mesure nulle ($\mu(N) = 0$) mais qui ne soit pas mesurable, on est bien tenté de penser qu'elle est vraie presque partout (c'est à dire vraie partout sauf dans un ensemble de mesure nulle) mais on n'en a pas le droit puisque N_0 n'est pas mesurable. Un espace de mesure qui est tel qu'en retrouve pas ce type de problème est dit espace complet.

Définition 1.1.3 (Espace de mesure complet)

Un espace de mesure $(\mathcal{X}, \mathcal{F}, \mu)$ est dit complet si tout sous-ensemble d'un ensemble de \mathcal{F} , de mesure nulle, est aussi un membre de \mathcal{F} et qui est donc, par monotonie de la mesure, de mesure nulle.

Toutefois, lorsqu'un espace de mesure n'est pas complet, il est possible d'étendre l'espace de mesure en un espace contenant tous les ensembles, parties d'ensembles de mesure nulle. C'est le processus de complétion d'espace. Le nouveau espace complet ainsi construit est tel que les tribus correspondantes ne diffèrent que par des ensembles, parties d'ensembles de mesure nulle.

Produit de mesures

Soient $(\mathcal{X}, \mathcal{F}, \mu)$ et $(\mathcal{Y}, \mathcal{L}, \nu)$ deux espaces de mesure. On définit, $\mathcal{F} \otimes \mathcal{L}$, la plus petite tribu contenant les produits cartésiens de la forme $A \times B$ avec $A \in \mathcal{F}$ et $B \in \mathcal{L}$.

Alors, il existe une mesure unique λ sur l'espace mesurable $(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{L})$ telle que pour tout ensemble $C \in \mathcal{F} \otimes \mathcal{L}$ de la forme $C = A \times B$ on a

$$\lambda(A \times B) = \mu(A) \nu(B).$$

La mesure λ est dite mesure produit de μ et ν et est notée $\mu \times \nu$.

Ayant défini un espace de mesure et les modes d'assigner des mesures à des ensembles, il est fréquent de vouloir mesurer des ensembles se rattachant à des fonctions (image d'un

ensemble par une fonction, image réciproque d'un ensemble par une fonction) définies sur ces ensembles. Ces fonctions doivent satisfaire des conditions dites de mesurabilité.

b- Fonctions mesurables Soient $(\mathcal{X}, \mathcal{F})$ et $(\mathcal{X}', \mathcal{F}')$ deux espaces mesurables.

Définition 1.1.4

Une fonction $f, \mathcal{X} \rightarrow \mathcal{X}'$ définie sur $(\mathcal{X}, \mathcal{F})$ à valeurs dans $(\mathcal{X}', \mathcal{F}')$ est dite mesurable si pour tout $A' \in \mathcal{F}'$, $f^{-1}(A') := \{x \in \mathcal{X} : f(x) \in A'\} \in \mathcal{F}$. Autrement dit, l'image réciproque par f de tout membre de \mathcal{F}' , est un membre de \mathcal{F} .

Une fonction mesurable de $(\mathcal{X}, \mathcal{F})$ dans $(\mathcal{X}', \mathcal{F}')$ est aussi dite \mathcal{F} -mesurable ou encore \mathcal{F}/\mathcal{F}' -mesurable.

Soient $(\mathcal{X}, \mathcal{F})$ et $(\mathcal{X}', \mathcal{F}')$ deux espaces mesurables. Le théorème suivant rassemble quelques propriétés des fonctions mesurables.

Théorème 1.1.1

- i) Toute fonction continue de $(\mathcal{X}, \mathcal{F})$ dans $(\mathcal{X}', \mathcal{F}')$ est mesurable.
- ii) La somme, le produit, la division (lorsqu'elle est définie) et la composée de deux fonctions f et g mesurables est une fonction mesurable.
- iii) Si f_1, f_2, \dots des fonctions réelles \mathcal{F} -mesurables, alors : les fonctions suivantes : $\sup_n f_n$, $\inf_n f_n$, $\limsup_n f_n$, $\liminf_n f_n$ et $\lim_n f_n$ (si elle existe) sont \mathcal{F} -mesurables.

Exemple 1.1.4

- i) La fonction indicatrice d'un ensemble $A \in \mathcal{F}$, $I_A(\cdot)$ définie de \mathcal{X} dans $\{0, 1\}$ par

$$I_A(x) = \begin{cases} 1 & \text{si } x \in A, \\ 0 & \text{sinon} \end{cases}$$

est \mathcal{F} -mesurable.

- ii) La fonction étagée ξ définie sur $(\mathcal{X}, \mathcal{F})$, prenant un nombre fini de valeurs a_1, a_2, \dots, a_m , et pouvant s'écrire en termes de fonctions indicatrices comme suit

$$\xi(x) = \sum_{i=1}^m a_i I_{A_i}(x),$$

où $A_i = \{x : \xi(x) = a_i\}$, est \mathcal{F} -mesurable. On note que l'ensemble A_i est quelconque et peut avoir n'importe quelle structure. Dans le cas simple où A_i est un intervalle de \mathbb{R} , la fonction étagée $\xi(\cdot)$ est dite *en escalier*.

- iii) Soient ξ_1, ξ_2, \dots une suite de fonctions étagées définies sur $(\mathcal{X}, \mathcal{F})$, positives et croissantes et soit f une fonction définie sur $(\mathcal{X}, \mathcal{F})$ par

$$f(x) = \lim_{n \rightarrow \infty} \xi_n(x). \quad (1.1.4)$$

Il est clair que cette limite existe (mais elle peut être infinie). Si la limite $\lim_{n \rightarrow \infty} \xi_n(x)$ existe pour tout $x \in \mathcal{X}$ alors f est \mathcal{F} -mesurable.

Une définition équivalente de la mesurabilité d'une fonction positive est la suivante : Une fonction positive f est dite \mathcal{F} -mesurable si et seulement si il existe une suite de fonctions étagées, positives et croissantes, ξ_1, ξ_2, \dots pour lesquelles (1.1.4) est vérifiée.

iv) Pour une fonction f arbitraire, définie sur $(\mathcal{X}, \mathcal{F})$, on définit la partie positive f^+ et la partie négative f^- de f comme suit

$$\begin{aligned} f^+(x) &= \max(f(x), 0), \\ f^-(x) &= -\min(f(x), 0), \end{aligned}$$

de sorte que f^+ et f^- sont toutes les deux positives et

$$f(x) = f^+(x) - f^-(x).$$

Donc f est \mathcal{F} -mesurable si ses parties positive et négative le sont également. \square

Remarque 1.1.4 (Transformation de mesures)

Soient $(\mathcal{X}, \mathcal{F})$ et $(\mathcal{X}', \mathcal{F}')$ deux espaces mesurables et f une fonction \mathcal{F} -mesurable de \mathcal{X} dans \mathcal{X}' . Etant donné une mesure $\mu(\cdot)$ définie sur \mathcal{F} , on définit sur \mathcal{F}' la fonction ensembliste $\mu f^{-1}(\cdot)$ par

$$\begin{aligned} \mu f^{-1}(\cdot) &: \mathcal{F}' \longrightarrow [0, \infty] \\ A' &\longmapsto \mu f^{-1}(A') = \mu(f^{-1}(A')). \end{aligned}$$

Alors il est facile de montrer que $\mu f^{-1}(\cdot)$ est une mesure sur $(\mathcal{X}', \mathcal{F}')$. Si $\mu(\cdot)$ est finie alors il en est de même pour $\mu f^{-1}(\cdot)$. \square

2- Intégrale

La notion d'intégrale d'une fonction f sur un ensemble I a connu un développement important depuis la version originelle de Reimann. Celui-ci définit l'intégrale d'une fonction f sur un intervalle $[a, b]$ comme étant la limite d'une somme, relative à une subdivision donnée de $[a, b]$, dont chaque terme est le produit de la longueur d'un sous-intervalle de la subdivision par la valeur de la fonction f sur un point quelconque du sous-intervalle. La fonction f est alors dite intégrable au sens de Reimann si la limite de cette somme (dite de Riemann) lorsque le nombre de sous-intervalles de la subdivision tend vers l'infini existe indépendamment de la subdivision choisie. Cependant, pour certaines fonctions pathologiques qu'on retrouve en théorie des probabilités, en physique ou d'autres domaines scientifiques, cette limite (et donc l'intégrale de Reimann) peut ne pas exister. L'intégrale au sens de Reimann a été généralisée par Stieltjes de sorte que pour la somme considérée, ce n'est plus la longueur du pas qui est multipliée par la valeur de la fonction f mais plutôt une autre fonction g du pas, vérifiant certaines conditions, qui en est multipliée. Lorsque la limite d'une telle somme (dite somme de Riemann-Stieltjes) existe quelle que soit la subdivision choisie, on dira que la fonction f est intégrable au sens de Riemann-Stieltjes par rapport à g . L'intégrale de Riemann-Stieltjes existe pour des fonctions non Reimann intégrables. Pourtant, il existe certaines fonctions pour lesquelles l'intégrale de Stieltjes n'est pas défini. Lebesgue développa une théorie plus générale de l'intégrale même pour des fonctions sur des ensembles assez irréguliers et qui coïncide avec l'intégrale de Reimann dans les cas où ce dernier existe. Il suffit d'un espace mesurable sur lequel est définie une fonction mesurable, l'intégrale de Lebesgue peut être défini et est dit intégrale par rapport à une mesure. C'est ce qu'on étudiera dans cette section.

1- Intégrale de Reimann Soit f une fonction numérique positive à valeurs réelles, définie sur un intervalle $[a, b]$ ($a < b$). On considère une subdivision $\{x_0, \dots, x_n\}$ de $[a, b]$ telle que $x_0 = a < x_1 < \dots < x_n = b$, à partir de laquelle on construit la somme suivante S_n dite de Reimann

$$S_n = \sum_{i=1}^n (x_i - x_{i-1}) f(c_i),$$

où $c_i \in [x_{i-1}, x_i]$. Si lorsque $n \rightarrow \infty$, de sorte que le plus grand des pas $|x_i - x_{i-1}|$ tende vers zéro, S_n converge vers une limite finie quelle que soit la subdivision $\{x_0, \dots, x_n\}$ et quel que soit l'ensemble $\{c_1, \dots, c_n\}$ alors on dit que la fonction f est intégrable au sens de Reimann (ou Reimann intégrable) sur $[a, b]$. La limite de S_n lorsque $n \rightarrow \infty$ est dite intégrale de Reimann de f sur $[a, b]$ et est notée

$$\lim_{n \rightarrow \infty} S_n = \int_a^b f(x) dx.$$

Si f admet une primitive F sur $[a, b]$ alors

$$\int_a^b f(x) dx = F(b) - F(a).$$

L'intégrale de Reimann d'une fonction f positive peut s'interpréter comme étant la limite de la somme des aires des rectangles associés à une subdivision) dont les bases sont les longueurs des sous-intervalles successifs de la subdivision et dont les hauteurs correspondent aux valeurs de la fonction f à des points correspondants aux sous-intervalles. Ainsi, l'intégrale de f sur $[a, b]$ n'est autre que l'air de la surface délimitée par la courbe de f et les droites $x = a$, $x = b$ et $y = 0$ (voir Figure 1.1.1).

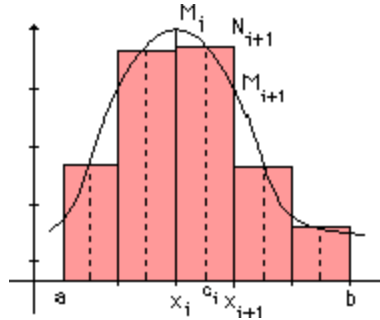


Figure 1.1.1

Si la fonction f n'est pas forcément positive alors son intégrale de Reimann représente l'aire signée (algébrique).

Il existe plusieurs définitions équivalentes de l'intégrale de Reimann dont parmi figure l'intégrale de Darboux. Au lieu de la somme S_n , Darboux a considéré pour $m_i = \inf_{x \in [x_{i-1}, x_i]} f(x)$ et

$M_i = \sup_{x \in [x_{i-1}, x_i]} f(x)$ les sommes suivantes

$$\begin{aligned}\overline{S}_n &= \sum_{i=1}^n (x_i - x_{i-1}) M_i, \\ \underline{S}_n &= \sum_{i=1}^n (x_i - x_{i-1}) m_i,\end{aligned}$$

et les bornes sup et inf

$$\begin{aligned}\overline{S} &= \inf_{\underline{x} \in \mathcal{P}} \{\overline{S}_n\}, \\ \underline{S} &= \sup_{\underline{x} \in \mathcal{P}} \{\underline{S}_n\},\end{aligned}$$

où \mathcal{P} est l'ensemble de toutes les subdivisions possibles de $[a, b]$. Si $\overline{S} = \underline{S}$ alors on dit que la fonction f est Darboux intégrable sur $[a, b]$ et son intégrale, notée $\int_a^b f(x) dx$, est égale à

$$\int_a^b f(x) dx = \overline{S} = \underline{S}.$$

Une propriété importante est la suivante.

Théorème 1.1.2

Une fonction f est Darboux intégrable sur $[a, b]$ si et seulement si elle y est Reimann intégrable et les valeurs des deux intégrales se confondent.

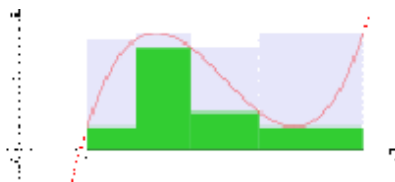


Figure 1.1.2

Figure 1.1.2 \underline{S} est représenté par l'aire en vert et \overline{S} par l'aire en lavande

L'intégrale de Riemann vérifie évidemment les propriétés :

i) de linéarité : $\int_a^b (\lambda f + \mu g)(x) dx = \lambda \int_a^b f(x) dx + \mu \int_a^b g(x) dx$ pour toutes fonctions f et g Riemann intégrables.

ii) de Challe : $\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx$ pour toute fonction f Riemann intégrable.

L'intégrale de Riemann peut être généralisé :

- i) au cas d'une fonction de plusieurs variables sur un domaine d'un espace Euclidien;
- ii) au cas où sur l'une des bornes de l'intégration la fonction n'est pas définie ou lorsque les bornes sont infinies,

$$\int_{-\infty}^{+\infty} f(x) dx = \lim_{a \rightarrow -\infty} \lim_{b \rightarrow +\infty} \int_a^b f(x) dx,$$

ce qui est appelée intégrale *impropre*.

2- Intégrale de Riemann-Stieltjes L'intégrale de Reimann-Steiltjes est une généralisation naturelle de l'intégrale de Reimann (due à T. Steiltjes) voyant son origine et ses applications dans des problèmes concrets de sciences physiques (distribution de masses) ou de calcul des probabilités.

Exemple 1.1.5 (introductif)

On considère sur un axe $[Ox]$ une masse m répartie sur le segment $[AB]$ et une masse ponctuelle m_c placée en C entre O et A (voir Figure 1.1.3).



Figure 1.1.3

Il s'agit de calculer l'attraction exercée par m sur m_c . Alors on applique à m et m_c la loi de Newton : la force d'attraction d'une particule infinitésimale dm de $[AB]$ attire m_c selon une force élémentaire, dF , telle que

$$dF = \frac{km_c dm}{(x - c)^2},$$

où k désigne la constante d'attraction universelle, c est l'abscisse de m_c et x l'abscisse de m sur $[AB]$. En parcourant tout l'intervalle $[a, b]$, la force totale F exercée sur m_c par la masse répartie sur le segment $[a, b]$ est donnée par :

$$\begin{aligned} F &= \int_a^b dF \\ &= km_c \int_a^b \frac{1}{(x - c)^2} dm. \end{aligned}$$

Or $dm = dm(x)$ dépend de l'abscisse x et la distribution de masse peut être répartie de façon très diverse et l'existence d'une fonction de densité n'est pas assurée. Si une densité de masse $\delta(x)$ existe, alors $dm = \delta(x)dx$ et le calcul de F se ramène à une intégrale de Riemann "ordinaire". Sinon, il faut envisager une intégrale selon la distribution de masse $g(\cdot)$, considérée comme une mesure sur l'intervalle $[AB]$. Cette mesure devra être additive : chaque portion $\Delta x_i = x_{i+1} - x_i$ de $[AB]$ sera mesurable et on sera en droit d'écrire $g(\Delta x_i) = g(x_{i+1}) - g(x_i)$. Dans ces conditions F apparaît comme la limite, pour n infini et $|x_{i+1} - x_i|$ tendant vers 0, des sommes du type :

$$\sum_{i=1}^n (g(x_i) - g(x_{i-1})) f(c_i).$$

□

Intégrale de Reimann-Steiltjes

Soient f et g deux fonctions numériques à valeurs réelles définies sur un intervalle $[a, b]$ ($a < b$). Pour une subdivision $\{x_0, \dots, x_n\}$ de $[a, b]$ telle que $x_0 = a < x_1 < \dots < x_n = b$, on considère la somme S_n dite de Reimann-Stieltjes

$$S_n = \sum_{i=1}^n (g(x_i) - g(x_{i-1})) f(c_i),$$

où $c_i \in [x_{i-1}, x_i]$. Si lorsque $n \rightarrow \infty$, de sorte que le plus grand des pas $|x_i - x_{i-1}|$ tende vers zéro, S_n converge vers une limite finie quelle que soit la subdivision $\{x_0, \dots, x_n\}$ et quelle que soit $\{c_1, \dots, c_n\}$ alors on dit que la fonction f est intégrable au sens de Reimann-Steiltjes (ou Reimann-Steiltjes intégrable) sur $[a, b]$. La limite de S_n lorsque $n \rightarrow \infty$ est dite intégrale de Reimann-Steiltjes de f sur $[a, b]$ par rapport à g et est notée

$$\lim_{n \rightarrow \infty} S_n = \int_a^b f(x) dg(x).$$

Les fonctions f et g sont dites *integrand* et *integrateur* respectivement. Usuellement, la fonction g est croissante mais ce n'est pas une condition nécessaire pour l'existence de l'intégrale. Par contre une condition nécessaire de l'existence d'une telle intégrale est que les fonctions f et g n'ont pas de point de discontinuité communs. Si g est différentiable et sa dérivée est continue sur tout l'intervalle alors l'intégrale de Reimann-Steiltjes $\int_a^b f(x) dg(x)$ se réduit à l'intégrale de Reimann

$$\int_a^b f(x) g'(x) dx.$$

Pourtant, cette dernière peut être différente de l'intégrale de Reimann-Steiltjes si la fonction g est même dérivable mais que sa dérivée n'est pas bornée.

Remarque 1.1.5

- i) L'intégrale de Riemann-Steiltjes se réduit à l'intégrale de Reimann pour $g(x) = x$ sur $[a, b]$.
- ii) L'intégrale de Reimann-Steiltjes admet une forme d'intégration par parties :

$$\int_a^b f(x) dg(x) = f(b)g(b) - f(a)g(a) - \int_a^b g(x) df(x),$$

où l'intégrale de droite existe pourvu que l'intégrale de gauche existe. \square

Existence de l'intégrale de Reimann-Steiltjes

Le plus simple théorème d'existence de l'intégrale de Reimann-Steiltjes stipule que si f est continue sur $[a, b]$ et si g est à *variations bornées* alors l'intégrale $\int_a^b f(x) dg(x)$ existe.

On rappellera d'abord quelques notions concernant la propriété de variation bornée.

Définition 1.1.5 La variation totale d'une fonction réelle g sur $[a, b]$ est donnée par

$$V_{[a,b]}(g) = \sup_{\underline{x} \in \mathcal{P}} \sum_{i=1}^n |g(x_i) - g(x_{i-1})|,$$

où \mathcal{P} est l'ensemble des partitions possibles de $[a, b]$.

Si g est différentiable sur $[a, b]$ alors il est clair que la variation totale de g sur $[a, b]$ est donnée par

$$V_{[a,b]}(g) = \int_a^b |g'(x)| dx.$$

Définition 1.1.6 Une fonction réelle g est dite à *variations bornées* sur $[a, b]$ si sa variation totale $V_{[a,b]}(g)$ est finie, i.e. si

$$\sup_{\underline{x} \in \mathcal{P}} \sum_{i=1}^n |g(x_i) - g(x_{i-1})| < \infty.$$

On note qu'une fonction g est à variation bornée si et seulement si elle peut s'écrire comme la différence de deux fonctions monotones.

Exemple 1.1.6

- i) La fonction g définie sur $[0, 2/\pi]$ par $g(x) = \begin{cases} \sin\left(\frac{1}{x}\right) & \text{si } x \neq 0, \\ 0 & \text{si } x = 0 \end{cases}$ n'est pas à variations bornée (voir Figure 1.1.4 (a)).
- ii) La fonction g définie sur $[0, 2/\pi]$ par $g(x) = \begin{cases} x \sin\left(\frac{1}{x}\right) & \text{si } x \neq 0, \\ 0 & \text{si } x = 0 \end{cases}$ n'est pas à variations bornée (voir Figure 1.1.4 (b)).
- iii) La fonction g définie sur $[0, 2/\pi]$ par $g(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right) & \text{si } x \neq 0, \\ 0 & \text{si } x = 0 \end{cases}$ est à variations bornée (voir Figure 1.1.4 (c)).

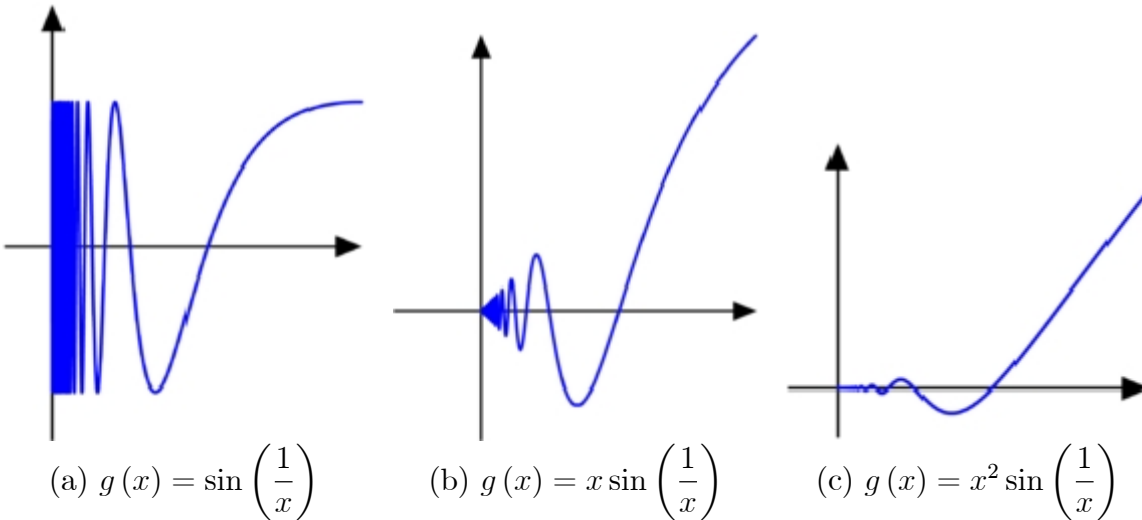


Figure 1.1.4

3- Intégrale par rapport à une mesure (Intégrale de Lebesgue) L'intégrale de Lebesgue est une construction mathématique généralisant l'intégrale de Reimann à une plus vaste classe de fonctions. Elle constitue aussi une extension à des domaines plus irréguliers sur lesquels ces fonctions peuvent être définies. Le terme "*intégration de Lebesgue*" peut se référer soit à une théorie générale de l'intégration par rapport à une mesure générale, comme introduit par Lebesgue, soit à un cas spécifique d'intégration d'une fonction définie sur un sous-ensemble de la droite réelle par rapport à la mesure de Lebesgue.

Soit $(\mathcal{X}, \mathcal{F}, \mu)$ un espace de mesure par rapport auquel on veut définir l'intégrale au sens de Lebesgue d'une fonction f , mesurable, définie sur \mathcal{X} à valeurs réelles. Alors, l'intégrale de Lebesgue est défini en trois étapes :

i) Si f est une fonction étagée, i.e., prenant un nombre fini de valeurs a_1, a_2, \dots, a_m , et pouvant s'écrire en termes de fonctions indicatrices comme suit

$$f(x) = \sum_{i=1}^m a_i I_{A_i}(x),$$

où $A_i = \{x : \xi(x) = a_i\}$, alors l'intégrale de Lebesgue de f est défini par

$$\int f d\mu = \sum_{i=1}^m a_i \mu(A_i), \quad (1.1.5)$$

avec la convention que $a\mu(A) = 0$ lorsque $a = 0$ et $\mu(A) = \infty$.

ii) Si f est une fonction positive, on sait qu'elle est la limite d'une suite de fonctions étagées croissantes et positives, comme donné par (1.1.4), i.e. $f(x) = \lim_{n \rightarrow \infty} f_n(x)$. Alors,

$$\int f d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu.$$

Les définitions i) et ii) n'excluent pas le cas où l'intégrale est infinie. Une fonction positive f est dite intégrable si son intégrale est finie, i.e. $\int f d\mu < \infty$.

iii) Si f est une fonction mesurable arbitraire alors elle est dite intégrable si ses parties positive et négatives sont toutes deux intégrables. Son intégrale est alors définie par

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu. \quad (1.1.6)$$

Remarque 1.1.6

i) L'intégrale précédente est aussi notée $\int f(x) d\mu(x)$, $\int_{\mathcal{X}} f(x) d\mu(x)$, $\int f(x) \mu(dx)$ ou $\int_{\mathcal{X}} f(x) \mu(dx)$.

ii) La mesure par rapport à laquelle est défini l'intégrale de Lebesgue n'est pas systématiquement la mesure de Lebesgue et peut être n'importe quelle mesure aussi abstraite soit-elle.

iii) Puisque $|f| = f^+ + f^-$ alors f est intégrable au sens de Lebesgue si et seulement si $\int |f| d\mu < \infty$. \square

Jusque là, on a défini l'intégrale de Lebesgue par rapport à une mesure sur tout le référentiel \mathcal{X} . Il est toutefois possible d'étendre la définition de l'intégrale de Lebesgue sur un ensemble $A \subset \mathcal{X}$ comme suit :

$$\int_A f d\mu = \int 1_A f d\mu. \quad (1.1.7)$$

Il s'ensuit à partir de (1.1.5) et (1.1.7) que

$$\int_A d\mu = \mu(A).$$

\square

Exemple 1.1.7

Soient $(\mathcal{X}, \mathcal{F}, \mu)$ un espace de mesure, tel que $\mathcal{X} = \mathbb{N}$, \mathcal{F} une tribu des parties de \mathbb{N} et μ la mesure de comptage, et f une fonction mesurable définie sur \mathcal{X} . Alors de (1.1.5) et (1.1.6) on a

$$\int f d\mu = \sum_{k \in \mathbb{N}} f(k).$$

\square

Exemple 1.1.8

Soient $(\mathcal{X}, \mathcal{F}, \mu)$ un espace de mesure, tel que $\mathcal{X} = \mathbb{N}$, \mathcal{F} une tribu des parties de \mathbb{N} et μ la mesure de Lebesgue, et f une fonction mesurable définie sur \mathcal{X} . Alors de (1.1.5) et (1.1.6) on a

$$\int f d\mu = 0.$$

□

Remarque 1.1.7

Soient $(\mathcal{X}, \mathcal{F}, \mu)$ un espace de mesure avec μ la mesure de Lebesgue et f une fonction mesurable définie sur \mathcal{X} . Alors l'intégrale $\int f d\mu$ existe toute les fois que l'intégrale au sens de Riemann de f sur \mathcal{X} , $\int_{\mathcal{X}} f(x) dx$, existe et on a

$$\int f d\mu = \int_{\mathcal{X}} f(x) dx.$$

Cependant, l'intégrale de Lebesgue $\int f d\mu$ est défini même pour des fonctions pour lesquelles l'intégrale de Riemann n'existe pas. Un exemple simple est celui de la fonction dite de Dirichlet, $1_{\mathbb{Q}}$, définie sur $[a, b] \subset \mathbb{R}$ par

$$1_{\mathbb{Q}}(x) = \begin{cases} 1 & \text{si } x \in \mathbb{Q}. \\ 0 & \text{sinon} \end{cases}$$

Il est clair que l'intégrale $\int_{[a,b]} 1_{\mathbb{Q}} d\mu$, μ étant la mesure de Lebesgue, existe puisque d'après (1.1.5) on a

$$\int_{[a,b]} 1_{\mathbb{Q}} d\mu = \mu(\mathbb{Q} \cap [a, b]) = 0,$$

montrant que l'intégrale de Lebesgue sur $[a, b]$ de la fonction $1_{\mathbb{Q}}$ est nulle.

Par contre si, dans les sommes de Riemann $S_n = \sum_{i=1}^n (x_i - x_{i-1}) f(c_i)$, on choisit c_i rationnel (ce qui est toujours possible puisque \mathbb{Q} est dense dans \mathbb{R}), l'intégrale de Riemann sur $[a, b]$ serait alors $S_n = \sum_{i=1}^n (x_i - x_{i-1}) \times 1 = b - a$ puisque les $1_{\mathbb{Q}}(c_i)$ valent 1. Et si, on choisit les c_i irrationnels, la somme de Riemann $S_n = \sum_{i=1}^n (x_i - x_{i-1}) \times 0$ est nulle et l'intégrale de Riemann serait nulle. Ce qui prouve que $1_{\mathbb{Q}}$ n'est pas intégrable au sens de Riemann. □

Remarque 1.1.8

L'intégrale de Lebesgue est invariant pour des fonctions qui se diffèrent sur un ensemble de mesure nulle (par rapport à la mesure convenue). En effet :

- Si f et g sont deux fonctions positives (pouvant prendre la valeur $+\infty$) telles que $f = g$ presque partout (p.p.) alors

$$\int f d\mu = \int g d\mu.$$

- Si f et g sont des fonctions telles que $f = g$ presque partout alors f est Lebesgue intégrable si et seulement si g est Lebesgue intégrable et les intégrales de f et g se confondent.

- Si $\int_A f d\mu = 0$ alors soit $f = 0$ sur A soit $\mu(A) = 0$. Cela peut se résumer comme suit : $f = 0$ p.p. sur A .

- Inversement, si f est positive p.p. sur A alors $\int_A f d\mu = 0$ entraîne que $f = 0$ p.p. sur A et si f est strictement positive p.p. alors $\int_A f d\mu = 0$ entraîne que $\mu(A) = 0$. □

Propriétés de L'intégrale de Lebesgue

L'intégrale de Lebesgue possède les propriétés suivantes :

i) *Linéarité* : Si f et g sont deux fonctions Lebesgue-intégrables et α et β sont deux nombres réels alors $\alpha f + \beta g$ est Lebesgue-intégrable et

$$\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu.$$

ii) *Monotonie* : Si f est mesurable et g Lebesgue-intégrable avec $0 \leq f \leq g$, alors

$$\int f d\mu \leq \int g d\mu.$$

iii) *Théorème de la convergence monotone* : Soit $\{f_n\}_{n \in \mathbb{N}}$ une suite de fonctions mesurables et positives, définies sur un ensemble E , telles que pour tout $n \in \mathbb{N}$ et tout $x \in E$

$$f_n(x) \leq f_{n+1}(x),$$

alors

$$\lim_{n \rightarrow \infty} \int f_n d\mu \leq \int \sup_{n \in \mathbb{N}} f_n d\mu,$$

où les valeurs de chacune des deux intégrales peuvent être infinies.

iv) *Théorème de la convergence dominée* : Si $\{f_n\}_{n \in \mathbb{N}}$ une suite de fonctions réelles (ou même complexes), mesurables et convergeant simplement vers une fonction f et s'il existe une fonction g Lebesgue-intégrable, telle que $|f_n| \leq g$ pour tout $n \in \mathbb{N}$, alors f est Lebesgue-intégrable et

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu.$$

v) *Lemme de Fatou* : Si $\{f_n\}_{n \in \mathbb{N}}$ une suite de fonctions positives et mesurables, alors

$$\int \left(\liminf_{n \rightarrow \infty} f_n \right) d\mu \leq \liminf_{n \rightarrow \infty} \int f_n d\mu,$$

où les valeurs de chacune des deux intégrales peuvent être infinies.

vi) *Théorème de Fubini* : Le théorème de Fubini stipule que pour une intégrale répétée d'une fonction positive l'ordre d'intégration n'est pas important. Soient $(\mathcal{X}, \mathcal{F}, \mu)$ et $(\mathcal{Y}, \mathcal{L}, \nu)$ deux espaces de mesure et f une fonction $\mathcal{F} \times \mathcal{L}$ -mesurable définie sur $\mathcal{X} \otimes \mathcal{Y}$. Alors

$$\begin{aligned} \int_{\mathcal{X}} \left(\int_{\mathcal{Y}} f(x, y) d\nu(y) \right) d\mu(x) &= \int_{\mathcal{Y}} \left(\int_{\mathcal{X}} f(x, y) d\mu(x) \right) d\nu(y) \\ &= \int_{\mathcal{X} \times \mathcal{Y}} f d(\mu \times \nu). \end{aligned}$$

Continuité absolue de mesure

Un thème central dans la théorie de la mesure est l'absolue continuité d'une mesure par rapport à une autre mesure. Avant d'étudier ce concept rappelons d'abord la notion de continuité absolue pour les fonctions réelles.

a- *Continuité absolue d'une fonction réelle* :

Définition 1.1.7 (*Continuité absolue d'une fonction*)

Une fonction réelle f , définie sur un intervalle I , est dite *absolument continue* sur I si pour tout $\varepsilon > 0$, aussi petit soit-il, il existe un nombre positif δ suffisamment petit de sorte que pour toute suite de sous-intervalles $\{[x_k, y_k]\}_{k=1, \dots, n}$ de I , deux à deux disjoints et vérifiant

$$\sum_{k=1}^n |y_k - x_k| < \delta,$$

alors

$$\sum_{k=1}^n |f(y_k) - f(x_k)| < \varepsilon.$$

Propriétés des fonctions absolument continues

- La somme, la différence et le produit de deux fonctions absolument continues sur I sont également absolument continues.
- Toute fonction absolument continue est uniformément continue et est donc continue.
- Toute fonction continue Lipschitzienne est absolument continue.
- Si $f : I \rightarrow \mathbb{R}$ est absolument continue, alors elle est à variations bornées sur I .
- Si $f : I \rightarrow \mathbb{R}$ est absolument continue, alors elle est dérivable p.p. sur I .

b- *Continuité absolue d'une mesure* : La notion de continuité absolue peut être adaptée à la notion de mesure.

Définition 1.1.8 (*Continuité absolue d'une mesure*)

Soient μ et ν deux mesures définies sur le même espace mesurable $(\mathcal{X}, \mathcal{F})$. On dit que μ est absolument continue par rapport à ν si pour tout $A \in \mathcal{F}$

$$\nu(A) = 0 \Rightarrow \mu(A) = 0.$$

Un résultat très important concernant la continuité absolue de mesure est connu sous le nom du théorème de Radon-Nikodym.

Théorème 1.1.3 (Radon-Nikodym)

Si μ est absolument continue par rapport à ν et ν est σ -finie alors μ admet une densité (dite aussi la dérivée de Radon-Nikodym) par rapport à ν , i.e. il existe une fonction ν -mesurable f (notée aussi $\frac{d\mu}{d\nu}$) à valeurs dans $[0, +\infty]$ telle que pour tout $A \in \mathcal{F}$

$$\mu(A) = \int_A f d\nu.$$

c- *Relation entre les deux notions de continuité absolue* : Le résultat suivant donne une connexion entre la continuité absolue d'une mesure (définie sur la tribu de Borel) par rapport à la mesure particulière de Lebesgue et la continuité absolue d'une fonction particulière jouant le rôle d'une distribution.

Théorème 1.1.4

Une mesure μ définie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est absolument continue par rapport à la mesure de Lebesgue si et seulement si la fonction $F : \mathbb{R} \rightarrow [0, +\infty[$ définie par

$$F(x) = \mu([-\infty, x]),$$

est absolument continue.

II- Théorie des probabilités

1- Introduction

La théorie des probabilités peut être vue comme la représentation mathématique des phénomènes aléatoires. Un phénomène est dit aléatoire s'il est impossible de prévoir sa réalisation à l'avance. Parfois, on parle d'*expérience* plutôt que de phénomène, expérience qu'on provoque à volonté et qu'on définit sous certaines conditions au vu d'un objectif fixé. Une expérience est alors dite aléatoire si pour les mêmes conditions elle peut fournir différents résultats possibles. La distinction entre phénomène et expérience provient du fait que par opposition aux expériences, les phénomènes se produisent sans que l'on participe à leur réalisation. Par exemple, le jet d'une pièce de monnaie et l'observation du résultat apparue sur la face supérieure est une expérience alors que la consommation d'électricité à Alger pendant le mois de Janvier est un phénomène. Cependant en admettant que l'observation d'un phénomène est une expérience, on peut réserver dans la suite le mot phénomène pour l'une et l'autre notion.

La théorie des probabilités permet de donner les règles de base pour le calcul des probabilités, domaine qui s'intéresse au calcul de la probabilité d'événements complexes sur la base de la connaissance de la probabilité de certains événements moins complexes. Elle a été développée dans les années 1930 par Kolmogorov qui, s'inspirant de la théorie de la mesure, a défini la probabilité en tant que mesure particulière sur un espace mesurable dont le référentiel est de mesure unité.

2- Espace de probabilité

Pour décrire un phénomène aléatoire, la théorie de probabilités offre trois éléments fondamentaux :

a) Un ensemble Ω , dit espace fondamental, collectionnant les différentes réalisations possibles du phénomène.

b) Une famille \mathcal{F} de parties de Ω telle que

i) $\Omega \in \mathcal{F}$,

ii) si $A \in \mathcal{F}$ alors $A^c \in \mathcal{F}$,

iii) si $\{A_i, i \in I\}$ est une famille dénombrable d'ensembles $A_i \in \mathcal{F}$ alors $\bigcup_{i \in I} A_i \in \mathcal{F}$.

c) Une mesure $P(\cdot)$ définie sur \mathcal{F} à valeurs réelles et vérifiant

Axiome i) $P(\Omega) = 1$,

Axiome ii) $P(A) \geq 0$ pour tout $A \in \mathcal{F}$,

Axiome iii) $P\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} P(A_i)$ pour toute famille d'événements deux à deux dis-

joints ($A_i \cap A_j = \emptyset$ pour $i \neq j$).

On appelle le couple (Ω, \mathcal{F}) espace *probabilisable* et le triplet (Ω, \mathcal{F}, P) espace probabilisé ou espace de probabilité.

Remarques

i) La famille \mathcal{F} est dite algèbre ou σ -algèbre (tribu) selon que I soit fini ou infini. Les parties de Ω appartenant à \mathcal{F} sont appelées *événements probabilisables (mesurables)* dont on distingue en particulier l'événement *certain* Ω , l'événement *impossible* \emptyset et l'événement *élémentaire* $\{\omega\}$, pour $\omega \in \Omega$. L'espace (Ω, \mathcal{F}) est dit probabilisable dans le sens où \mathcal{F} désignera les parties de Ω qui se prêtent à être probabilisées, c'est à dire qu'on peut leur

assigner une mesure de probabilité vérifiant les clauses i)-iii) de c). Ainsi, dans la définition de \mathcal{F} , la clause i) garantie que \mathcal{F} est non vide. La condition ii) est imposée car si l'on veut mesurer un évènement, il est naturel de vouloir mesurer son contraire. De même, la condition iii) est motivée par le fait que si l'on considère des évènements $A_i, i \in I$, il est légitime de s'intéresser à mesurer l'évènement "au moins l'un des évènements A_i se réalise". Avec les conditions i)-iii) on montre aisément que toutes les opérations ensemblistes usuelles (intersection, différence symétriques,...) appliquées à des évènements de \mathcal{F} forment encore des évènements probabilisables puisqu'ils appartiennent à \mathcal{F} .

ii) On distingue deux σ -algèbres particulières : la σ -algèbre *triviale* $\{\emptyset, \Omega\}$ la plus petite dans le sens de l'inclusion et la σ -algèbre la plus large qu'est l'ensemble des parties de Ω , $\mathcal{P}(\Omega)$. Entre ces deux extrêmes, il existe plusieurs tribus intermédiaires. Si A est un ensemble non vide alors la plus petite tribu à laquelle appartient A est l'ensemble $\{\emptyset, A, A^c, \Omega\}$. Aussi, pour toute famille ξ de parties de Ω il existe une tribu $\mathcal{B}(\xi)$, la plus petite contenant ξ . Pour $\Omega = \mathbb{R}$ la plus petite tribu, $\mathcal{B}(\mathbb{R})$, contenant tous les intervalles de \mathbb{R} est dite tribu de Borel. Elle peut s'écrire comme la plus petite tribu générée par des intervalles de la forme $] - \infty, a[$, $a \in \mathbb{R}$. Chaque intervalle appartenant à une telle tribu est dit Borélien de \mathbb{R} . Cette notion s'étend aisément à \mathbb{R}^d , $d \in \mathbb{N}^*$.

iii) Lorsque Ω est fini ou infini dénombrable, soit $\Omega = \{\omega_i, i \in I = \mathbb{N}\}$, on peut prendre $\mathcal{F} = \mathcal{P}(\Omega)$. Pour une suite de nombres non négatifs $\{p_i, i \in I\}$ telle que $\sum_{i \in \mathbb{N}} p_i = 1$, l'application

$P(\cdot)$ sur \mathcal{F} définie par

$$P(A) = \sum_{\omega_i \in A} p_i,$$

est une mesure de probabilité.

Si de plus $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ est fini et les évènements élémentaires $\{\omega_i\}$ sont équiprobables dans le sens où $P(\omega_i) = P(\omega_j)$ pour tout $i, j \in \{1, \dots, n\}$, alors pour tout évènement $A \in \mathcal{F} = \mathcal{P}(\Omega)$, on a

$$P(A) = \frac{\text{card}(A)}{\text{card}(\Omega)},$$

où $\text{card}(A)$ désigne le nombre d'issues réalisant A .

iv) Lorsque Ω est non dénombrable (par exemple $\Omega = \mathbb{R}$) l'ensemble des parties de \mathbb{R} , $\mathcal{P}(\mathbb{R})$, bien que formant une tribu, est trop grand pour pouvoir en définir une mesure de probabilité $P(\cdot)$ vérifiant les axiomes i)-iii). En effet, certaines parties de Ω sont dans ce cas la réunion non dénombrable d'une famille de parties de Ω auquel on ne peut appliquer l'axiome iii), à moins de considérer une autre version de l'axiome iii) prévoyant le cas d'une réunion non dénombrable, mais cela ajoutera une complexité inutile pour la théorie, puisque dans la pratique de tels cas ne se produisent guère. Selon une telle construction axiomatique, un évènement est alors une partie de Ω qui est membre de \mathcal{F} mais pas toutes les parties de Ω sont des évènements. Intuitivement un évènement est un fait rassemblant tous les résultats possibles le réalisant.

v) Lorsqu'un évènement A est de probabilité nulle, il est tout à fait légitime de penser que tout sous-ensemble de A est aussi de probabilité nulle. Cependant il arrive qu'une partie de A (A étant de probabilité nulle) ne soit pas dans \mathcal{F} et n'est donc pas probabilisable pourtant elle est primordiale d'un point de vue pratique. Cette difficulté est rencontrée particulièrement pour un évènement élémentaire qui n'appartient pas à \mathcal{F} et qui est partie d'un ensemble de probabilité nulle : une tribu \mathcal{F} ne contenant pas des évènements élémentaires n'a pas d'utilité

pratique. Cette difficulté peut être remédiée par le processus de completion d'espace. *Un espace de probabilité (Ω, \mathcal{F}, P) est dit complet si tout sous-ensemble d'un évènement de \mathcal{F} , de probabilité nulle, est aussi un évènement de \mathcal{F} et est donc, par monotonie de la probabilité, de probabilité nulle.* Lorsqu'un espace de probabilité n'est pas complet, il est toujours possible de lui associer un espace complet tel que les tribus correspondantes ne diffèrent que par des ensembles, parties d'ensembles de probabilité nulle.

vi) D'un point de vue purement mathématique la probabilité est un nombre attaché à un évènement. Pourtant intuitivement elle possède plusieurs interprétations. La probabilité d'un évènement représente parfois la limite de la fréquence relative d'apparition de cet évènement lorsque l'expérience est répétée indéfiniment. C'est l'approche fréquence-limite. Cependant, dans certains cas, le sens fréquence-limite est impraticable. Par exemple, si on s'intéresse à la probabilité qu'un accusé soit coupable, il est clair qu'on ne peut considérer une fréquence associée. En revanche une telle probabilité représente plutôt un degré de conviction qu'on a quant à la culpabilité de l'accusé. Une telle interprétation est dite subjective ou personnelle puisque le degré de conviction diffère d'une personne à une autre.

vii) Dans certains cas l'ensemble Ω peut être identifié à une région bornée du plan Euclidien d'où on tire un point au hasard. Si A est une surface dans Ω , alors la probabilité de tirer un point dans A est égale à

$$P(A) = \frac{\text{Air de } A}{\text{Air de } \Omega}.$$

Une telle probabilité est dite probabilité géométrique.

3- Probabilité conditionnelle

Lorsqu'on observe un phénomène aléatoire et qu'on s'interroge sur la probabilité d'apparition d'un évènement lié A , il arrive parfois qu'on obtienne des informations supplémentaires quant à la réalisation d'un autre évènement B . Ces informations alors peuvent changer la probabilité de A en une probabilité dite conditionnelle de A sachant B est réalisé.

Soit (Ω, \mathcal{F}, P) est un espace de probabilité et $B \in \mathcal{F}$ un évènement de probabilité non nulle. Alors on peut définir une autre mesure de probabilité $P(. / B)$ définie sur l'espace probabilisable (Ω, \mathcal{F}) par

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

Le triplet $(\Omega, \mathcal{F}, P(. / B))$ est dit espace de probabilité conditionnel sachant B . Evidemment lorsque $P(B) = 0$ la mesure de probabilité conditionnelle $P(. / B)$ n'est pas définie. Intuitivement, peut-on s'intéresser à la réalisation d'un évènement A sachant qu'un évènement B presque impossible (qui ne se réalise presque jamais) s'est réalisé ?

De la définition de la probabilité conditionnelle, on a

$$P(A \cap B) = P(B) P(A/B)$$

Paradoxalement cette relation est vérifiée même si $P(B) = 0$.

a- Formule des probabilités totales Il arrive parfois que la probabilité d'un évènement A , $P(A)$, assez difficile à obtenir, s'exprime en termes de probabilités conditionnelles plus

simples à calculer. Soit $\{B_n, n \in \mathbb{N}^*\}$ une partition de Ω (on l'appelle parfois système complet d'évènements). Alors,

$$P(A) = \sum_{n \in \mathbb{N}^*} P(B_n) P(A/B_n).$$

La dernière formule est dite formule des probabilités totales.

b- Formule de Bayes Une relation très utilisée en probabilité aussi bien qu'en statistique et qui est liée à la formule des probabilités totales est la formule de Bayes. Elle permet de calculer les probabilités $P(B_j/A)$ lorsqu'on connaît $P(B_n)$ et $P(A/B_n)$, $n \in \mathbb{N}^*$.

$$\begin{aligned} P(B_j/A) &= \frac{P(B_j \cap A)}{P(A)} \\ &= \frac{P(B_j) P(A/B_j)}{\sum_{n \in \mathbb{N}^*} P(B_n) P(A/B_n)}. \end{aligned}$$

c- Probabilités conditionnelles itérées - Formule des probabilité composées On suppose que sur l'espace probabilisable (Ω, \mathcal{F}) est définie la mesure de probabilité $P(\cdot)$. On peut construire la mesure $P(\cdot/A_1)$ en conditionnant par rapport à un évènement A_1 de probabilité non nulle. Si A_2 est un autre évènement de probabilité non nulle alors on peut encore définir une mesure de probabilité conditionnelle dérivée de $P(\cdot/A_1)$ en conditionnant par rapport à A_2 . Une telle mesure de probabilité, notée $P(\cdot/A_1/A_2)$, est naturellement définie par

$$P(B/A_1/A_2) = \frac{P(A_2 \cap B/A_1)}{P(A_2/A_1)},$$

pourvu que $P(A_2/A_1) > 0$. En remplaçant le numérateur et le dénominateur de la dernière relation par leur définitions on trouve

$$\begin{aligned} P(B/A_1/A_2) &= \frac{\frac{P(A_1 \cap A_2 \cap B)}{P(A_1)}}{\frac{P(A_2 \cap A_1)}{P(A_1)}} \\ &= \frac{P(A_1 \cap A_2 \cap B)}{P(A_2 \cap A_1)} \end{aligned}$$

qui n'est rien d'autre que la probabilité conditionnelle, $P(B/A_1 \cap A_2)$, de B sachant $A_1 \cap A_2$, ce qu'on note dorénavant $P(B/A_1 A_2)$.

Il est possible d'en déduire une formule dite "des probabilités composées" comme suit.

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) P(A_2/A_1) \dots P(A_n/A_{n-1} \dots A_2 A_1).$$

Une autre variante en est la suivante

$$P(A_1 \cap A_2 \cap \dots \cap A_n/C) = P(A_1/C) P(A_2/A_1 C) \dots P(A_n/A_{n-1} \dots A_2 A_1 C).$$

4- Indépendance- Produit d'espaces

a- Indépendance Deux évènements A et B sont dit (stochastiquement) *indépendants* si la réalisation de l'un n'apporte aucune modification à la probabilité de l'autre. Autrement dit $P(A/B)$ est confondue avec $P(A)$ pourvu que $P(B) > 0$ ou $P(B/A) = P(B)$ lorsque $P(B) > 0$. Une définition plus générale englobant ces deux cas est la suivante.

Définition 1.2.1 (Indépendance de deux évènements)

Deux évènements A et B sont dits indépendants si $P(A \cap B) = P(A)P(B)$.

La notion d'indépendance peut être généralisée au cas d'une famille finie d'évènements. Mais là deux cas sont à distinguer : L'indépendance deux-à-deux et l'indépendance mutuelle.

Définition 1.2.2 (Indépendance deux-à-deux)

Les évènements A_1, A_2, \dots, A_n sont dits deux-à-deux indépendants si pour tout $i, j \in \{1, \dots, n\}$ tels que $i \neq j$, A_i et A_j sont indépendants.

Définition 1.2.3 (Indépendance mutuelle d'un ensemble fini d'évènements)

Les évènements A_1, A_2, \dots, A_n sont dits mutuellement indépendants si pour et tout r ($2 \leq r \leq n$) distincts indices $i_1, i_2, \dots, i_r \in \{1, \dots, n\}$ on a

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_r}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_r}).$$

Ainsi, la notion d'indépendance mutuelle est plus forte que l'indépendance deux à deux. En effet, l'exemple suivant montre qu'on peut exhiber un ensemble d'évènements deux-à-deux indépendants mais qui ne sont pas mutuellement indépendants.

Exemple 1.2.1

Soit $\Omega = \{a, b, c, d\}$ dont les évènements élémentaires sont équiprobables et on considère les évènements $A_1 = \{a, b\}$, $A_2 = \{a, c\}$ et $A_3 = \{a, d\}$. Il est clair que ces évènements sont tous de probabilité $\frac{1}{2}$ et qu'ils sont deux-à-deux indépendants puisque $P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = \frac{1}{4}$. Pourtant $P(A_1 \cap A_2 \cap A_3) = P(\{a\}) = \frac{1}{4}$ et est différente de $P(A_1)P(A_2)P(A_3) = \frac{1}{8}$.

Ce résultat n'est pas surprenant puisqu'on peut rencontrer dans la nature des cas analogues. En chimie par exemple, certaines expériences montrent qu'en mélangeant deux-à-deux, trois produits chimiques A_1 , A_2 et A_3 aucune réaction chimique ne s'enclenche par contre si on les mélange simultanément une réaction peut surgir.

L'exemple suivant montre que pour vérifier l'indépendance mutuelle d'une famille d'évènements, il faut vérifier la formule de la Définition 1.2.3 pour toutes les intersections possibles. En particulier, il se pourrait que l'on ait $P(A \cap B \cap C) = P(A)P(B)P(C)$ sans que les évènements A , B et C ne soient indépendants.

Exemple 1.2.2

Soit Ω un espace d'épreuves partitionné en cinq évènements A_1, A_2, A_3, A_4 et A_5 , tels que

$$P(A_1) = P(A_2) = P(A_3) = \frac{15}{64}, P(A_4) = \frac{1}{64}, P(A_5) = \frac{18}{64}$$

On pose $B = A_1 \cup A_4$, $C = A_2 \cup A_4$ et $D = A_3 \cup A_4$. Alors on a $P(B) = P(C) = P(D) = \frac{1}{4}$ et $P(B \cap C \cap D) = P(A_4) = \frac{1}{64} = P(B)P(C)P(D)$. Pourtant $P(B \cap C) = P(A_4) = \frac{1}{64} \neq P(B)P(C) = \frac{1}{16}$.

Dorénavant lorsqu'on parle d'indépendance d'évènements, par défaut il s'agit d'indépendance mutuelle sauf mention contraire.

La notion d'indépendance mutuelle peut être encore généralisée à une famille infinie d'évènements.

Définition 1.2.4 (Indépendance mutuelle d'une famille infinie d'évènements)

Une famille infinie d'évènements $\{A_i, i \in I\}$ est dite mutuellement indépendante si toute sous-famille finie d'elle est mutuellement indépendante au sens de la Définition 1.2.3.

Définition 1.2.5 (Indépendance mutuelle d'une collection de familles d'évènements)

Une collection de familles d'évènements $\{\mathcal{A}_j, j \in J\}$ est dite mutuellement indépendante si toute famille d'évènements obtenus en sélectionnant un évènement A de chaque famille \mathcal{A}_j , $j \in J$ est mutuellement indépendante au sens de la Définition 1.2.4.

b- Expériences répétée - Produit d'espaces Il arrive souvent que l'on s'intéresse à une expérience aléatoire ξ qui est réalisée en répétant un nombre fini (ou infini) d'expériences élémentaires ξ_i , $i \in I$ où les issues sont mutuellement indépendants. Dans ce cas, l'espace de probabilité associé à l'expérience composée ξ peut être écrit en termes d'espaces de probabilités dont chacun est associé à l'une des expériences élémentaires ξ_i .

Cas de produit de deux espaces

Pour simplifier l'exposé on considère d'abord le cas d'une expérience ξ consistant en la composée de deux expériences ξ_1 et ξ_2 .

On considère deux expériences distinctes et indépendantes ξ_1 et ξ_2 dont les espaces de probabilités correspondants sont respectivement $(\Omega_1, \mathcal{F}_1, P_1)$ et $(\Omega_2, \mathcal{F}_2, P_2)$. L'objectif est de combiner ces deux expériences en une seule, soit ξ , puis décrire celle-ci au moyen d'un espace de probabilité (Ω, \mathcal{F}, P) .

Un tel espace sera décrit comme suit. Pour espace fondamental Ω associé à ξ , on prendra le produit cartésien de Ω_1 et Ω_2 , i.e. $\Omega = \Omega_1 \times \Omega_2$.

D'autre part, il est naturel d'imposer que les ensembles du type $A_1 \times \Omega_2$ et $\Omega_1 \times A_2$, pour tout $A_1 \in \mathcal{F}_1$ et $A_2 \in \mathcal{F}_2$ soient dans la tribu \mathcal{F} . De tels ensembles sont dits cylindriques. Ceci conduit à la définition suivante du produit de deux tribus.

Définition 1.2.6 (Produit de deux tribus)

Le produit $\mathcal{F}_1 \otimes \mathcal{F}_2$ de deux tribus \mathcal{F}_1 et \mathcal{F}_2 est la plus petite tribu des parties de $\Omega_1 \times \Omega_2$ contenant les ensembles cylindriques du type $A_1 \times \Omega_2$ et $\Omega_1 \times A_2$.

Il est clair que la tribu $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$ contient les rectangles $A_1 \times A_2$ pour $A_1 \in \mathcal{F}_1$ et $A_2 \in \mathcal{F}_2$. La tribu \mathcal{F} étant bien définie, on désire maintenant définir sur elle une mesure de probabilité $P(\cdot)$. Une telle mesure doit convenir avec $P_1(\cdot)$ sur des cylindres de la forme $A_1 \times \Omega_2$, $A_1 \in \mathcal{F}_1$ et avec $P_2(\cdot)$ sur des cylindres de la forme $\Omega_1 \times A_2$, $A_2 \in \mathcal{F}_2$. On exige donc que les familles d'évènements $\{A_1 \times \Omega_2, A_1 \in \mathcal{F}_1\}$ et $\{\Omega_1 \times A_2, A_2 \in \mathcal{F}_2\}$ soient indépendantes. Ceci entraîne que

$$\begin{aligned} P(A_1 \times A_2) &= P((A_1 \times \Omega_2) \cap (\Omega_1 \times A_2)) \\ &= P_1(A_1)P_2(A_2). \end{aligned}$$

La mesure de probabilité est définie pour tous les rectangles. En faisant appel au théorème d'extension on peut montrer que $P(\cdot)$ est uniquement déterminée pour tout $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$.

Théorème 1.2.1 (Théorème d'extension)

Il existe une seule mesure de probabilité $P(\cdot)$ définie sur le produit de tribus $\mathcal{F}_1 \otimes \mathcal{F}_2$ par

$$P(A_1 \times A_2) = P_1(A_1)P_2(A_2), \quad \text{pour } A_1 \in \mathcal{F}_1 \text{ et } A_2 \in \mathcal{F}_2.$$

Définition 1.2.7 (Produit de deux mesures)

La mesure P est dite produit des mesures P_1 et P_2 . Elle est notée $P_1 \times P_2$.

Définition 1.2.8 (Produit de deux espaces)

L'espace de probabilité (Ω, \mathcal{F}, P) défini par $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$ et $P = P_1 \times P_2$ est dit produit des espaces $(\Omega_1, \mathcal{F}_1, P_1)$ et $(\Omega_2, \mathcal{F}_2, P_2)$ dans cet ordre.

La notion de produit d'espaces peut être facilement étendue au cas d'un nombre fini ou infini dénombrable d'espaces de probabilités.

Cas de produit d'un nombre fini d'espaces

On considère n ($n \in \mathbb{N}^*$) expériences aléatoires ξ_i décrites par les espaces de probabilités $(\Omega_i, \mathcal{F}_i, P_i)$, $i \in \{1, \dots, n\}$. On désire combiner toutes ces expériences en une seule expérience ξ dont l'espace de probabilité est (Ω, \mathcal{F}, P) . Alors l'espace fondamental Ω sera donné par $\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_n$, l'ensemble de tous les n -uplets $(\omega_1, \omega_2, \dots, \omega_n)$ avec $\omega_i \in \Omega_i$. De même, la tribu \mathcal{F} est donné par le produit $\mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \dots \otimes \mathcal{F}_n$, la plus petite tribu contenant les ensembles de la forme

$$\Omega_1 \times \dots \times \Omega_{j-1} \times A_j \times \Omega_{j+1} \times \dots \times \Omega_n, \quad A_j \in \mathcal{F}_j, \quad j \in \mathbb{N}^*,$$

qui correspond à l'évènement "... à l'expérience j l'évènement A_j se réalise...". Ainsi la tribu \mathcal{F} est astreinte à contenir les évènements de la forme $A_1 \times A_2 \times \dots \times A_n$.

Comme pour le cas $n=2$, on définit la mesure de probabilité P sur les ensembles cylindriques de la forme $\Omega_1 \times \dots \times \Omega_{j-1} \times A_j \times \Omega_{j+1} \times \dots \times \Omega_n$ comme suit

$$P(\Omega_1 \times \dots \times \Omega_{j-1} \times A_j \times \Omega_{j+1} \times \dots \times \Omega_n) = P_j(A_j).$$

Puisque les rectangles du type $A_1 \times A_2 \times \dots \times A_n$ peuvent s'écrire comme

$$(A_1 \times \Omega_2 \times \dots \times \Omega_n) \cap (\Omega_1 \times A_2 \times \dots \times \Omega_n) \cap \dots \cap (\Omega_1 \times \Omega_2 \times \dots \times A_n),$$

alors il s'ensuit, grâce à l'indépendance des expériences, que

$$P(A_1 \times A_2 \times \dots \times A_n) = P_1(A_1) P_2(A_2) \dots P_n(A_n).$$

Ceci conduit à la définition de produits d'espaces comme suit

Définition 1.2.9 (Produit d'un nombre fini d'espaces)

L'espace de probabilité (Ω, \mathcal{F}, P) défini par $\Omega = \Omega_1 \times \dots \times \Omega_n$, $\mathcal{F} = \mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_n$ et $P = P_1 \times \dots \times P_n$ est dit produit des espaces $(\Omega_1, \mathcal{F}_1, P_1), \dots, (\Omega_n, \mathcal{F}_n, P_n)$ dans cet ordre.

Cas de produit d'une suite d'espaces

Le développement précédent peut être généralisé au cas d'une suite d'expériences aléatoires ξ_i , $i \in \mathbb{N}^*$) expériences aléatoires.

On considère une suite d'expériences ξ_i , $i \in \mathbb{N}^*$ indépendantes décrites par les espaces de probabilités $(\Omega_i, \mathcal{F}_i, P_i)$, $i \in \mathbb{N}^*$. De la même manière que précédemment on désire représenter ces expériences par une seule expérience dont l'espace de probabilité est (Ω, \mathcal{F}, P) . la construction en est similaire et est résumée par la définition suivante.

Définition 1.2.10 (Produit d'une suite d'espaces de probabilité)

Le produit des espaces de probabilité $(\Omega_i, \mathcal{F}_i, P_i)$, $i \in \mathbb{N}^*$ est un espace de probabilité (Ω, \mathcal{F}, P) tel que

i) $\Omega = \prod_{i=1}^{\infty} \Omega_i = \Omega_1 \times \Omega_2 \times \dots$ est l'ensemble des ∞ -uplets $(\omega_1, \omega_2, \dots)$ avec $\omega_i \in \Omega_i$, $i \in \mathbb{N}^*$.

ii) $\mathcal{F} = \bigotimes_{i=1}^{\infty} \mathcal{F}_i = \mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \dots$ est la plus petite tribu de parties de Ω contenant les ensembles de la forme $\Omega_1 \times \dots \times \Omega_{j-1} \times A_j \times \Omega_{j+1} \times \dots$, $A_j \in \mathcal{F}_j$, $j \in \mathbb{N}^*$.

iii) $P = \bigotimes_{i=1}^{\infty} P_i = P_1 \times P_2 \times \dots$ est l'unique mesure sur \mathcal{F} vérifiant pour tout $n \in \mathbb{N}^*$ et pour tout choix de n entiers $0 < i_1 < i_2 < \dots < i_n \in \mathbb{N}^*$

$$P(\Omega_1 \times \dots \times \Omega_{i_1-1} \times A_{i_1} \times \dots \times A_{i_n} \times \Omega_{i_n+1} \times \dots) = P_{i_1}(A_{i_1}) \dots P_{i_n}(A_{i_n}).$$

Le théorème d'extension garanti l'existence et l'unicité d'une telle mesure.

Dans le cas où les expériences indépendantes sont décrites par le même espace de probabilité on dit qu'on a affaire à une expérience répétée. Ce cas en est plus simple en raison de la symétrie additionnelle.

III. Variables aléatoires

1- Définitions de base

Lorsqu'on envisage d'observer un phénomène aléatoire et de lui associer un espace de probabilité (Ω, \mathcal{F}, P) , il arrive que l'on s'intéresse plutôt à une fonction (numérique) du résultat attendu qu'au résultat lui-même. Mais ce faisant, l'objectif reste de pouvoir assigner des probabilités à des événements concernant cette fonction du résultat. La fonction étudiée doit donc satisfaire certaines conditions de "probabilisabilité" (i.e. de mesurabilité) : Si $X(\cdot)$ est une telle fonction définie sur Ω à valeurs réelles (dans $\mathcal{X} \subseteq \mathbb{R}$), des événements concernant $X(\cdot)$, notamment $\{\omega \in \Omega : X(\omega) = a\}, \{\omega \in \Omega : X(\omega) > a\}, (a \in \mathbb{R})$, $\{\omega \in \Omega : a < X(\omega) < b\}, \{\omega \in \Omega : a \leq X(\omega) \leq b\}, a, b \in \mathbb{R}, a < b...$, doivent être dans \mathcal{F} pour qu'on puisse leur assigner des probabilités par $P(\cdot)$, car $P(A)$ n'est pas définie pour des A n'appartenant pas à \mathcal{F} . Cependant la quasi-totalité des événements concernant $X(\cdot)$ que l'on puisse formuler peuvent s'exprimer par des opérations ensemblistes élémentaires des événements $\{\omega \in \Omega : X(\omega) \leq a\}, a \in \mathbb{R}$ et donc il suffit d'exiger que ceux-ci soient membres de \mathcal{F} pour tout $a \in \mathbb{R}$. Une fonction vérifiant une telle condition est dite définie sur (Ω, \mathcal{F}, P) .

Définition 1.3.1 (Variable aléatoire)

Une fonction $X(\cdot)$ définie sur Ω à valeurs dans $\mathcal{X} \subseteq \mathbb{R}$ est dite variable aléatoire (v.a.) sur (Ω, \mathcal{F}, P) si pour tout $a \in \mathbb{R}, \{\omega \in \Omega : X(\omega) \leq a\} \in \mathcal{F}$.

En d'autres termes, l'image réciproque de $] -\infty, a[$ par X est un événement de \mathcal{F} pour tout $a \in \mathbb{R}$. De façon plus générale, l'image réciproque de tout borélien $B \in \mathcal{B}(\mathbb{R})$ est dans \mathcal{F} .

Remarque 1.3.1

Si $\mathcal{F} = \mathcal{P}(\Omega)$ ($\mathcal{P}(\Omega)$ étant l'ensemble des parties de Ω) alors il est clair que toute application de Ω à valeurs réelles vérifie la condition de mesurabilité et définit par conséquent une v.a. sur (Ω, \mathcal{F}, P) .

Définition 1.3.2 (Equivalence de v.a.)

Deux variables aléatoires X et Y sont dites équivalentes si elles se diffèrent seulement sur un ensemble de probabilité nulle, i.e. si $P(\omega \in \Omega, X(\omega) \neq Y(\omega)) = 0$.

Définition 1.3.3 (Tribu engendrée par une v.a.)

La tribu \mathcal{F}_X engendrée par une v.a. X définie sur un espace de probabilité (Ω, \mathcal{F}, P) est la plus petite tribu contenant les événements de la forme $\{\omega : X(\omega) \leq a\}, a \in \mathbb{R}$.

Proposition 1.3.1 (Fonction d'une v.a.)

Soient X une v.a. définie sur (Ω, \mathcal{F}, P) à valeurs dans \mathbb{R} et φ une fonction de \mathbb{R} dans \mathbb{R} . Alors la fonction composée $\varphi(X)$ définie sur (Ω, \mathcal{F}, P) à valeurs dans \mathbb{R} en est une v.a. pourvu que φ soit mesurable.

En effet, pour $a \in \mathbb{R}, \varphi(X)^{-1}(]-\infty, a]) = \{\omega : \varphi(X(\omega)) \leq a\} \in \mathcal{F}$ en vertu de la mesurabilité de φ .

2- Distribution de probabilité

Définition 1.3.4 (Distribution de probabilité)

Soit X une v.a. réelle définie sur un espace de probabilité (Ω, \mathcal{F}, P) . On appelle distribution de X la fonction ensembliste $P_X(\cdot)$ définie sur $\mathcal{B}(\mathbb{R})$ à valeurs dans $[0, 1]$ par

$$\begin{aligned} P_X(\cdot) : \mathcal{B}(\mathbb{R}) &\longrightarrow [0, 1] \\ B &\longmapsto P_X(B) = P(\omega : X(\omega) \in B) = P(X^{-1}(B)). \end{aligned}$$

De la remarque 1.1.4, il s'ensuit que $P_X(\cdot)$ est une mesure qui est de plus une mesure de probabilité puisque $P_X(\mathbb{R}) = 1$.

Puisque la tribu de Borel est aussi la plus petite tribu engendrée par les intervalles de la forme $] - \infty, a]$, donc la donnée de $P_X(\cdot)$ est équivalente à la donnée d'une fonction $F_X(\cdot)$ plutôt réelle, dite fonction de répartition de X , définie par

$$\begin{aligned} F_X(\cdot) : \mathbb{R} &\longrightarrow [0, 1] \\ x &\longmapsto F_X(x) = P(\omega : X(\omega) \leq x) = P(X^{-1}(] - \infty, a])). \end{aligned}$$

La fonction de répartition $F_X(\cdot)$ est aussi dite distribution de probabilité de X .

Théorème 1.3.1 (Propriétés de la fonction de répartition)

Soit $F_X(\cdot)$ la fonction de répartition d'une v.a. X . Alors,

i) F_X est une fonction croissante.

ii) F_X est continue à droite.

iii) $\lim_{x \rightarrow -\infty} F_X(x) = 0$.

iv) $\lim_{x \rightarrow +\infty} F_X(x) = 1$.

Remarque 1.3.2

Une fonction de répartition F_X d'une v.a. X peut être vue comme une mesure de probabilité sur $\mathcal{B}(\mathbb{R})$, la plus petite tribu engendrée par les intervalles $] - \infty, a]$, $a \in \mathbb{R}$, comme suit

$$\begin{aligned} F_X(\cdot) : \mathcal{B}(\mathbb{R}) &\longrightarrow [0, 1] \\] - \infty, a] &\longmapsto F_X(] - \infty, a]) = F_X(a). \end{aligned}$$

3- Variables aléatoires discrètes, continues et absolument continues

Définition 1.3.5 (v.a. discrète)

Une v.a. X définie sur (Ω, \mathcal{F}, P) est dite discrète si elle est équivalente à une v.a. prenant des valeurs dans \mathbb{Z} seulement. i.e., s'il existe un ensemble dénombrable $I \subset \mathbb{Z}$ tel que $P(\omega : X(\omega) = x) > 0$ pour tout $x \in I$.

Pour une v.a. discrète à valeurs dans un ensemble dénombrable I , la structure de probabilité est entièrement déterminée par les probabilités

$$P(\omega : X(\omega) = x) := P(X = x), \quad x \in I.$$

Définition 1.3.6 (Fonction de masse d'une v.a. discrète)

Pour une v.a. discrète X , définie sur (Ω, \mathcal{F}, P) à valeurs dans I , on définit la fonction de masse $p_X(\cdot)$ par

$$\begin{aligned} p_X(\cdot) : I &\longrightarrow [0, 1] \\ x &\longmapsto p_X(x) = P(\omega : X(\omega) = x) = P(X^{-1}(\{x\})). \end{aligned}$$

Proposition 1.3.2 (Relation entre F_X et p_X)

Pour une v.a. discrète X définie sur (Ω, \mathcal{F}, P) à valeurs dans I , on a

$$F_X(x) = \sum_{\{k \in I, k \leq x\}} p_X(k).$$

Définition 1.3.7 (v.a. continue)

Une v.a. X définie sur (Ω, \mathcal{F}, P) à valeurs dans \mathbb{R} est dite continue si $P(X = x) = 0$ pour tout $x \in \mathbb{R}$.

Définition 1.3.8 (v.a. absolument continue)

Une v.a. X définie sur (Ω, \mathcal{F}, P) à valeurs dans \mathbb{R} est dite absolument continue si sa fonction de répartition F_X est absolument continue au sens de la Définition 1.1.7.

Théorème 1.3.2 (Relation entre continuité et continuité absolue d'une v.a.)

i) Si X est absolument continue alors il existe une fonction $f_X(\cdot)$ positive de \mathbb{R} dans \mathbb{R} , dite densité de probabilité de X , telle que

$$F_X(x) = \int_{-\infty}^x f_X(t) dt.$$

f_X est aussi dite densité de Radon-Nikodym de la mesure F_X par rapport à la mesure de Lebesgue.

ii) Si X est absolument continue alors elle est continue, la réciproque est fausse.

Remarque 1.3.3

i) Si X admet une densité de probabilité alors X est absolument continue et est donc continue. Par contre il se peut que X soit continue sans qu'elle n'admette une densité de probabilité.

ii) La densité de probabilité $f_X(\cdot)$ de X ne représente évidemment pas une probabilité et peut prendre des valeurs très grandes. Mais elle est liée à la probabilité au moyen de la formule suivante

$$\begin{aligned} f_X(x) &= \lim_{\Delta x \rightarrow 0} \frac{F_X(x + \Delta x) - F_X(x)}{\Delta x} \\ &= \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X \leq x + \Delta x)}{\Delta x} \end{aligned}$$

à partir de laquelle on en déduit pour tout nombre suffisamment petit dx que

$$P(x < X < x + dx) \simeq f_X(x)dx.$$

Remarque 1.3.4

Si $f_X(\cdot)$ est la densité de probabilité de la v.a. X , alors :

i) $f_X(x) \geq 0, \forall x \in \mathbb{R}$.

ii) $\int_{-\infty}^{+\infty} f_X(x) dx = 1$.

iii) $f_X(x) = \frac{dF_X(x)}{dx}$.

4- Caractéristiques d'une variable aléatoire

a- Espérance mathématique

Soit X une v.a. définie sur (Ω, \mathcal{F}, P) à valeurs dans \mathbb{R} .

Définition 1.3.9 (Espérance mathématique)

L'espérance mathématique de X , notée $E(X)$, est l'intégrale (de Lebesgue) de X (sur Ω) par rapport à la mesure $P(\cdot)$. Autrement dit

$$E(X) = \int_{\Omega} X(\omega) dP(\omega).$$

Remarque 1.3.4

i) L'espérance mathématique de X représente la moyenne des valeurs possibles de X pondérées par leurs probabilités de réalisation. Elle représente en fait une valeur centrale des valeurs possibles d'une v.a.

ii) L'espérance $E(X)$ peut être également définie en terme de :

- la distribution de probabilité $P_X(\cdot)$ de X ,

$$E(X) = \int_{X(\Omega)} x dP_X(x),$$

où $X(\Omega)$ est l'ensemble des valeurs possibles prises par X ,

- la fonction de répartition F_X de X ,

$$E(X) = \int_{\mathbb{R}} x dF_X(x),$$

- de la fonction de masse $p_X(\cdot)$ si X est discrète

$$\begin{aligned} E(X) &= \sum_{x \in X(\Omega)} x p_X(x) \\ &= \sum_{x \in X(\Omega)} x P(X = x) \end{aligned}$$

- de la densité de probabilité $f_X(\cdot)$ de X , si f est absolument continue

$$\int_{\mathbb{R}} x f_X(x) dx.$$

L'espérance mathématique d'une v.a. peut être finie, infinie ou peut ne pas exister.

b- Espérance d'une fonction d'une v.a.

Soit X une v.a. définie sur (Ω, \mathcal{F}, P) et $\varphi(\cdot)$ une fonction mesurable (probabilisable) de $X(\Omega)$ vers \mathbb{R} . Alors l'espérance mathématique de la v.a. $\varphi(X)$ est définie par

$$\begin{aligned} E(\varphi(X)) &= \int_{\Omega} \varphi(X(\omega)) dP(\omega) \\ &= \int_{\mathbb{R}} \varphi(x) dF_X(x). \end{aligned}$$

c- Moments

Définition 1.3.10 (Moment d'ordre r d'une v.a.)

Le moment d'ordre r , m_r , ($r \in \mathbb{R}$) d'une v.a. X est l'espérance mathématique de la fonction de X , $\varphi(X) = X^r$, i.e.

$$m_r = E(X^r).$$

Ainsi le moment d'ordre un n'est autre que l'espérance mathématique de X .

d- Moments factoriels

Définition 1.3.11 (Moment factoriel d'ordre r d'une v.a.)

Le moment factoriel d'ordre r , mf_r , ($r \in \mathbb{N}$) d'une v.a. X est défini par

$$mf_r = E(X(X-1)\dots(X-r+1)).$$

e- Moments centrés

Définition 1.3.12 (Moment centré d'ordre r d'une v.a.)

Le moment centré d'ordre r , mc_r , d'une v.a. X est le moment d'ordre r de la v.a. $X - E(X)$.

f- Variance

Définition 1.3.13 (Variance d'une v.a.)

Le moment centré d'ordre 2 d'une v.a. X est dit variance de X et est noté $var(X)$.

$$var(X) = E \{ (X - E(X))^2 \}.$$

Remarque 1.3.5

i) La variance d'une v.a. représente une mesure du degré de dispersion des valeurs possibles de cette v.a. autour de son espérance mathématique. C'en est précisément une moyenne des carrés des déviations des valeurs possibles de leur espérance mathématique.

ii) Lorsqu'une v.a. X représente une grandeur physique mesurée au moyen d'une certaine unité de mesure, la variance de X sera un nombre exprimé en terme du carré de l'unité de mesure. Pour revenir à l'unité de mesure on préfère parfois travailler avec la racine carrée de la variance qu'on appelle écart-type de X et qu'on note σ_X .

5- Fonctions génératrices - Fonctions caractéristiques

a- Fonction génératrice des moments

Soit X une v.a. définie sur (Ω, \mathcal{F}, P) à valeurs dans \mathbb{R} .

Définition 1.3.14 (Fonction génératrice des moments d'une v.a.)

On appelle fonction génératrice des moments de X la fonction $M_X(\cdot)$ définie de \mathbb{R} dans \mathbb{R} par

$$\begin{aligned} M_X(\cdot) : \mathbb{R} &\longrightarrow \mathbb{R} \\ t &\longmapsto M_X(t) = E(e^{tX}). \end{aligned}$$

Le domaine de définition de $M_X(\cdot)$ est alors donné par

$$D_{M_X} = \{t \in \mathbb{R} : E(e^{tX}) < \infty\}.$$

b- Fonction génératrice des moments factoriels

Définition 1.3.15 (Fonction génératrice des moments factoriels d'une v.a.)

On appelle fonction génératrice des moments factoriels de X la fonction $G_X(\cdot)$ définie de \mathbb{R}^+ dans \mathbb{R} par

$$\begin{aligned} G_X(\cdot) : \mathbb{R}^+ &\longrightarrow \mathbb{R} \\ t &\longmapsto G_X(t) = E(t^X). \end{aligned}$$

c- Propriétés

i) Les fonctions $M_X(\cdot)$ et $G_X(\cdot)$ sont équivalentes dans le sens où l'existence de l'une est équivalente à l'existence de l'autre et à une forme près elle permettent d'en tirer les mêmes conclusions puisque $\forall t \in \mathbb{R} : M_X(t) = G_X(e^t)$.

ii) Les moments d'ordres entiers r de X peuvent être générés à partir de M_X au moyen de la relation $E(X^r) = \left. \frac{d^r M_X(t)}{dt^r} \right|_{t=0}$.

iii) Les moments factoriels d'ordres entiers r de X peuvent être obtenus au moyen de : $E(X(X-1)\dots(X-r+1)) = \left. \frac{d^r G_X(t)}{dt^r} \right|_{t=1}$.

c- Fonction caractéristique

Définition 1.3.16 (Fonction caractéristique d'une v.a.)

On appelle fonction caractéristique d'une v.a. X la fonction $\varphi_X(\cdot)$ définie de \mathbb{R} dans \mathbb{C} par

$$\begin{aligned}\varphi_X(\cdot) : \mathbb{R} &\longrightarrow \mathbb{C} \\ t &\longmapsto \varphi_X(t) = E(e^{itX}).\end{aligned}$$

6- Variables aléatoires infinies

Jusque là nous avons supposé que les valeurs prises par une v.a. X (définie sur (Ω, \mathcal{F}, P)) sont finies ou du moins presque sûrement (i.e. l'ensemble des ω pour lesquels $X(\omega)$ est infinie est de probabilité nulle). Cependant, il arrive souvent que l'on s'intéresse à des applications de Ω dans $\overline{\mathbb{R}} = [-\infty, +\infty]$, \mathcal{F} -probabilisables dans le sens où $\{\omega \in \Omega : X(\omega) = +\infty\}$ et $\{\omega \in \Omega : X(\omega) = -\infty\}$ appartiennent à \mathcal{F} , et telles que $P(\{\omega \in \Omega : X(\omega) = \infty\}) > 0$. Ceci est possible puisque \mathcal{F} étant une tribu, on a

$$\begin{aligned}\{\omega \in \Omega : X(\omega) = +\infty\} &= \bigcap_{n=1}^{\infty} \{\omega \in \Omega : X(\omega) > n\} \in \mathcal{F}, \\ \{\omega \in \Omega : X(\omega) = -\infty\} &= \bigcap_{n=1}^{\infty} \{\omega \in \Omega : X(\omega) < -n\} \in \mathcal{F}.\end{aligned}$$

La plus part des propriétés des v.a. finies restent valables pour les variables infinies (i.e. ayant une probabilité non nulle de prendre des valeurs infinies). Ce type de v.a. peut représenter des situations concrètes : par exemple, on convient que le temps de retour à un état d'une chaîne de Markov est infini si la chaîne ne revisite jamais cet état. De même, s'il y a une probabilité positive qu'un arbre généalogique s'étend indéfiniment alors le temps d'extinction du nom de famille correspondant peut être considéré infini.

7- Vecteurs aléatoires

Lors de l'observation d'un phénomène aléatoire, il est des cas où l'on s'intéresse à plusieurs fonctions du résultat possible. Ainsi on considère plusieurs fonctions définies sur un même espace de probabilité. Ces fonctions peuvent être formulées sous le cadre de ce qu'on appelle vecteur aléatoire (v.a.).

Définition 1.3.17 (Vecteur aléatoire)

Une application $\underline{X} = (X_1, X_2, \dots, X_m)'$ de Ω dans \mathbb{R}^m est dite vecteur aléatoire si pour tout $j = 1, \dots, m$, et pour tout $\underline{x} = (x_1, x_2, \dots, x_m)' \in \mathbb{R}^m$ on a $\{\omega \in \Omega : X(\omega) \leq x_j\} \in \mathcal{F}$.

Autrement dit, un vecteur aléatoire est un vecteur de v.a.

Définition 1.3.18 (Distribution de probabilité d'un vecteur aléatoire)

Soit \underline{X} un v.a. défini sur un espace de probabilité (Ω, \mathcal{F}, P) . On appelle distribution de \underline{X} la fonction ensembliste $P_{\underline{X}}(\cdot)$ définie sur $B(\mathbb{R}^m)$ à valeurs dans $[0, 1]$ par

$$\begin{aligned}P_{\underline{X}}(\cdot) : \mathcal{B}(\mathbb{R}^m) &\longrightarrow [0, 1] \\ \underline{B} &\longmapsto P_{\underline{X}}(\underline{B}) = P(\omega : \underline{X}(\omega) \in \underline{B}) = P(\underline{X}^{-1}(\underline{B})).\end{aligned}$$

Puisque la tribu de Borel $B(\mathbb{R}^m)$ est aussi la plus petite tribu engendrée par les intervalles de la forme $] -\infty, x_1] \times] -\infty, x_2] \times \dots \times] -\infty, x_m]$, donc la donnée de $P_{\underline{X}}(\cdot)$ est équivalente à la donnée d'une fonction $F_{\underline{X}}(\cdot)$ plutôt réelle, dite fonction de répartition de \underline{X} , définie par

$$\begin{aligned}F_{\underline{X}}(\cdot) : \mathbb{R}^m &\longrightarrow [0, 1] \\ \underline{x} = (x_1, x_2, \dots, x_m)' &\longmapsto F_{\underline{X}}(\underline{x}) = P\left(\omega : \bigcap_{i=1}^m X_i(\omega) \leq x_i\right).\end{aligned}$$

La fonction de répartition $F_{\underline{X}}(\cdot)$ est aussi dite distribution de probabilité de \underline{X} ou distribution conjointe de X_1, X_2, \dots, X_m .

Remarque 1.3.6 (Distribution marginale)

On peut retrouver la distribution de probabilité d'une v.a. X_1 lorsqu'on connaît la distribution conjointe de X_1, X_2, \dots, X_m . En effet,

$$\begin{aligned} F_{X_1}(x_1) &= P(X_1 \leq x_1) = P(X_1 \leq x_1, X_2 < \infty, \dots, X_m < \infty) \\ &= F_{\underline{X}}(x_1, \infty, \dots, \infty) \end{aligned}$$

La distribution $F_{X_1}(x_1)$ est alors dite distribution marginale.

Remarque 1.3.7

De la même manière que pour les v.a. scalaires, on distingue les vecteurs aléatoires discrets, continus et absolument continus.

Définition 1.3.19 (v.a. discret)

Un v.a. \underline{X} défini sur (Ω, \mathcal{F}, P) est dit discret si toutes ses composantes sont discrètes.

Pour un v.a. discret à valeurs dans un ensemble dénombrable $I \subset \mathbb{Z}^m$, la structure de probabilité est entièrement déterminée par les probabilités

$$P(\omega : \underline{X}(\omega) = \underline{x}) := P(\omega : \bigcap_{i=1}^m X_i(\omega) = x_i), \quad \underline{x} \in I.$$

La fonction $p_{\underline{X}}(\cdot)$ définie par

$$\begin{aligned} p_{\underline{X}}(\cdot) : I &\longrightarrow [0, 1] \\ \underline{x} &\longmapsto p_{\underline{X}}(\underline{x}) = P\left(\omega : \bigcap_{i=1}^m X_i(\omega) = x_i\right) \end{aligned}$$

est dite fonction de masse de \underline{X} . De plus on a

$$F_{\underline{X}}(\underline{x}) = \sum_{\{\underline{k} \leq \underline{x}\}} p_{\underline{X}}(\underline{k}),$$

pour $\underline{k} = (k_1, k_2, \dots, k_m)'$.

Définition 1.3.20 (v.a. continu)

Un v.a. \underline{X} défini sur (Ω, \mathcal{F}, P) à valeurs dans \mathbb{R}^m est dit continu si $P(\underline{X} = \underline{x}) = 0$ pour tout $\underline{x} \in \mathbb{R}^m$.

Autrement dit si toutes ses composantes sont continues.

Définition 1.3.21 (v.a. absolument continu)

Un v.a. \underline{X} défini sur (Ω, \mathcal{F}, P) à valeurs dans \mathbb{R}^m est dit absolument continu si toutes ses composantes sont absolument continues.

Théorème 1.3.3 (Relation entre continuité et continuité absolue d'un v.a.)

i) Si \underline{X} est absolument continu alors il existe une fonction $f_{\underline{X}}(\cdot)$ positive de \mathbb{R}^m dans \mathbb{R}^+ , dite densité de probabilité de \underline{X} , telle que

$$\begin{aligned} F_{\underline{X}}(\underline{x}) &= \int_{\underline{t} \leq \underline{x}} f_{\underline{X}}(\underline{t}) d\underline{t} \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_m} f_{\underline{X}}(t_1, t_2, \dots, t_m) dt_1 dt_2 \dots dt_m. \end{aligned}$$

$f_{\underline{X}}$ est aussi dite densité de Radon-Nikodym de la mesure $F_{\underline{X}}$ par rapport à la mesure de Lebesgue.

Remarque 1.3.8

Il est clair que d'après la formule précédente on a:

$$i) \int_{\mathbb{R}^m} f_{\underline{X}}(\underline{t}) d\underline{t} = 1,$$

$$ii) f_{\underline{X}}(\underline{x}) = \frac{\partial^m F_{\underline{X}}(\underline{x})}{\partial x_1 \partial x_2 \dots \partial x_m}.$$

Dans plusieurs cas de figure, on s'intéresse à une ou plusieurs fonctions de vecteurs aléatoires. En effet, la somme $X_1 + \dots + X_n$ de n v.a., le maximum $\max\{X_1, \dots, X_n\}$ et les puissances $X_1^r X_2^r \dots X_n^r$ ($r \in \mathbb{R}$) en sont des exemples importants. De même que pour les variables aléatoires, on peut construire des fonctions de vecteurs aléatoires qui restent des vecteurs aléatoires.

Définition 1.3.22 (Fonction d'un v.a.)

Soient φ une fonction de \mathbb{R}^m dans \mathbb{R}^k et $\underline{X} = (X_1, X_2, \dots, X_m)'$ un vecteur aléatoire défini sur (Ω, \mathcal{F}, P) à valeurs dans \mathbb{R}^m . Alors la fonction composée $\varphi(\underline{X})$ définie sur (Ω, \mathcal{F}, P) à valeurs dans \mathbb{R}^k en est un v.a. pourvu que φ soit mesurable.

Caractéristiques d'un vecteur aléatoire

La plus part des caractéristiques des v.a. sont conservées dans le cas vectoriel:

i) L'espérance mathématique d'un vecteur aléatoire \underline{X} est le vecteur des espérances de chacune des composantes, i.e.

$$\begin{aligned} E(\underline{X}) &= E(X_1, X_2, \dots, X_m)' \\ &: = (E(X_1), E(X_2), \dots, E(X_m))'. \end{aligned}$$

ii) Dans le cas d'un vecteur de deux dimensions, on définit la covariance de deux variables X et Y comme suit

$$\begin{aligned} cov(X, Y) &= E\{(X - E(X))(Y - E(Y))\} \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

Ainsi la covariance de X et Y mesure une variabilité simultanée myenne de X et Y . C'en est précisément la moyenne des croisements des déviations de chacune des variables de son espérance mathématique. Evidemment, la covariance d'une variable et elle même n'est autre que sa variance et la covariance d'une variable X et une constante c est égale à 0.

iii) On définit également la corrélation entre X et Y comme suit

$$corr(X, Y) = \frac{cov(X, Y)}{\sqrt{var(X) var(Y)}}.$$

On vérifie aisément que $|corr(X, Y)| \leq 1$ avec égalité s'il y a relation linéaire entre X et Y . Si $corr(X, Y) = 0$ alors X et Y sont dites non corrélées.

iii) Pour un v.a. \underline{X} on définit la matrice de variance-covariance comme suit,

$$cov(\underline{X}) = E\{(\underline{X} - E(\underline{X}))(\underline{X} - E(\underline{X}))'\}.$$

IV. Variables aléatoires conditionnelles

1- Distribution conditionnelle- Espérance conditionnelle par rapport à un évènement

a- Cas de v.a. discrètes Soient X et Y deux v.a. discrètes définies sur (Ω, \mathcal{F}, P) à valeurs dans \mathbb{R} et soit $y \in Y(\Omega)$ (ainsi $P(Y = y) \neq 0$).

Définition 1.4.1 (Distribution conditionnelle)

On appelle *distribution conditionnelle de X sachant $Y = y$* , la fonction $F_{X/Y=y}(\cdot)$ définie par

$$\begin{aligned} F_{X/Y=y}(\cdot) : \mathbb{R} &\longrightarrow [0, 1] \\ x &\longmapsto F_{X/Y=y}(x) = P(X \leq x / Y = y) \\ &= \frac{P(X \leq x, Y = y)}{P(Y = y)}. \end{aligned}$$

Même si $P(Y = y) = 0$, on peut tout de même définir arbitrairement $F_{X/Y=y}(\cdot)$.

Définition 1.4.2 (Fonction de masse conditionnelle)

On appelle *fonction de masse conditionnelle de X sachant $Y = y$* , la fonction $p_{X/Y=y}(\cdot)$ définie par

$$\begin{aligned} p_{X/Y=y}(\cdot) : X(\Omega) &\longrightarrow [0, 1] \\ x &\longmapsto p_{X/Y=y}(x) = P(X = x / Y = y) \\ &= \frac{P(X = x, Y = y)}{P(Y = y)}. \end{aligned}$$

Proposition 1.4.1 (Relation entre $F_{X/Y=y}(\cdot)$ et $p_{X/Y=y}(\cdot)$)

Pour tout $x \in \mathbb{R}$ on a

$$F_{X/Y=y}(x) = \sum_{k \leq x} p_{X/Y=y}(k).$$

Remarque 1.4.1

i) La fonction $F_{X/Y=y}(\cdot)$ définit pour tout y fixé, une distribution de probabilité en x sur l'espace de probabilité conditionnel $(\Omega, \mathcal{F}, P(\cdot / Y = y))$ sachant $Y = y$.

ii) $F_{X/Y=y}(x)$ est une fonction de y pour tout x fixé.

iii) On peut ainsi définir ses caractéristiques telles que: espérance mathématique conditionnelle, variance conditionnelle,...

Définition 1.4.2 (Espérance conditionnelle de v.a. discrètes par rapport à un évènement)

L'espérance conditionnelle de X sachant $Y = y$ est définie comme étant l'espérance mathématique de la distribution conditionnelle $F_{X/Y=y}(x)$, c'est à dire:

$$\begin{aligned} E(X / Y = y) &= \sum_{X(\Omega)} x p_{X/Y=y}(x) \\ &= \int_{\mathbb{R}} x dF_{X/Y=y}(x). \\ &= \int_{\Omega} X(\omega) dP_{X/Y=y}(\omega). \end{aligned}$$

Définition 1.4.3 (Espérance conditionnelle d'une fonction de v.a. discrètes par rapport à un évènement)

$\varphi(X)$ étant une fonction mesurable, l'espérance conditionnelle de $\varphi(X)$ sachant $Y = y$ est définie par

$$\begin{aligned} E(\varphi(X)/Y = y) &= \sum_{X(\Omega)} \varphi(x) p_{X/Y=y}(x) \\ &= \int_{\mathbb{R}} \varphi(x) dF_{X/Y=y}(x) \\ &= \int_{\Omega} \varphi(X)(\omega) dP_{X/Y=y}(\omega). \end{aligned}$$

Propriétés

L'espérance conditionnelle vérifie les mêmes propriétés que l'espérance ordinaire. En effet,

i) $E(g(X)/Y = y)$ est une fonction de y pour tout $g(\cdot)$ telle que $E(|g(X)|) < \infty$.

ii) Si $g(\cdot)$ est une fonction positive ou nulle alors $E(g(X)/Y = y) \geq 0$.

iii) $E(c/Y = y) = c$, pour tout $c \in \mathbb{R}$.

iv) $E(cg(X)/Y = y) = cE(g(X)/Y = y)$.

v) $E(g(X) + h(Z)/Y = y) = E(g(X)/Y = y) + E(h(Z)/Y = y)$.

Définition 1.4.4 (Variance conditionnelle de v.a. discrètes par rapport à un évènement)

La variance conditionnelle de X sachant $Y = y$ est définie par

$$\begin{aligned} Var(X/Y = y) &= E((X - E(X/Y = y))^2 / Y = y) \\ &= E(X^2/Y = y) - (E(X/Y = y))^2. \end{aligned}$$

b- Cas de v.a. absolument continues Dans le cas où Y est continue ou même absolument continue, l'évènement $Y = y$ sur lequel on conditionne est de probabilité nulle et par conséquent les définitions précédentes ne sont plus applicables. Cependant, d'une part d'un point de vue intuitif, il est légitime de considérer la probabilité d'un évènement sachant la réalisation d'un autre évènement presque impossible (mais non impossible). D'autres part, dans ce cas continu $F_{X/Y=y}(x) = P(X \leq x/Y = y) = \frac{P(X \leq x, Y = y)}{P(Y = y)} = \frac{0}{0}$, et donc il est possible d'enlever la forme d'indétermination en considérant la définition suivante

$$\begin{aligned} F_{X/Y=y}(x) &: = P(X \leq x/Y = y) \\ &= \lim_{h \rightarrow 0} P(X \leq x/y \leq Y < y + h). \end{aligned}$$

On montre que généralement la limite précédente existe et que par conséquent la probabilité conditionnelle sachant un évènement presque impossible peut avoir un sens mathématiquement. Cependant on aurait pu considérer que $F_{X/Y=y}(x) := P(X = x/Y = y)$ est la limite de l'expression suivante

$$\begin{aligned} F_{X/Y=y}(x) &: = P(X \leq x/Y = y) \\ &= \lim_{h \rightarrow 0} P(X \leq x/y - h \leq Y < y + h). \end{aligned}$$

mais cette limite si elle existe peut être différente de $\lim_{h \rightarrow 0} P(X \leq x/y \leq Y < y + h)$. Ainsi, la définition précédente n'est pas tout à fait consistante et pour régler ce problème les probabilistes, en s'inspirant de la définition de la fonction de masse conditionnelle (Définition

1.4.2), ont défini par analogie la notion de densité conditionnelle de X sachant $Y = y$, pourvu que $f_Y(y)$ soit positive.

Définition 1.4.5 (Densité conditionnelle)

Soient X et Y deux v.a. absolument continues définies sur (Ω, \mathcal{F}, P) à valeurs dans \mathbb{R} et soit $y \in \mathbb{R}$ tel que $f_Y(y) > 0$. La densité conditionnelle de X sachant $Y = y$, notée $f_{X/Y=y}$ est définie par

$$\begin{aligned} f_{X/Y=y}(\cdot) : \mathbb{R} &\longrightarrow \mathbb{R}^+ \\ x &\longmapsto f_{X/Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \end{aligned}$$

où $f_{X,Y}(x, y)$ est la densité conjointe de X et Y .

Remarque 1.4.2

i) La fonction $f_{X/Y=y}(\cdot)$ définit, pour tout y fixé, une densité de probabilité en x sur l'espace de probabilité conditionnel $(\Omega, \mathcal{F}, P(\cdot/Y = y))$ sachant $Y = y$.

ii) $f_{X/Y=y}(x)$ est une fonction de y pour tout x fixé.

Une fois définie la densité conditionnelle $f_{X/Y=y}(x)$, on peut ainsi définir sans ambiguïté la distribution conditionnelle $F_{X/Y=y}(\cdot)$ de X sachant $Y = y$.

Définition 1.4.6 (Distribution conditionnelle par rapport à une v.a. absolument continue)

La distribution conditionnelle de X sachant $Y = y$ est donnée par

$$\begin{aligned} F_{X/Y=y}(\cdot) : \mathbb{R} &\longrightarrow [0, 1] \\ x &\longmapsto F_{X/Y=y}(x) = \int_{-\infty}^x f_{X/Y=y}(t) dt \\ &= \int_{-\infty}^x \frac{f_{X,Y}(t, y)}{f_Y(y)} dt. \end{aligned}$$

Remarque 1.4.3 La distribution conditionnelle dans le cas continu possède les mêmes propriétés que celle dans le cas discret. On peut de même définir ses caractéristiques telles que: espérance mathématique conditionnelle, variance conditionnelle,...

Définition 1.4.7 (Espérance conditionnelle par rapport à une v.a. absolument continue)

$\varphi(X)$ étant une fonction mesurable, l'espérance conditionnelle de $\varphi(X)$ sachant $Y = y$ est donnée par

$$\begin{aligned} E(\varphi(X)/Y = y) &= \int \varphi(x) f_{X/Y=y}(x) dx \\ &= \int_{\mathbb{R}} \varphi(x) dF_{X/Y=y}(x) \\ &= \int_{\Omega} \varphi(X)(\omega) dP_{X/Y=y}(\omega). \end{aligned}$$

L'espérance conditionnelle vérifie les mêmes propriétés que l'espérance ordinaire.

Propriétés

Définition 1.4.7 (Variance conditionnelle par rapport à une v.a. absolument continue)

La variance conditionnelle de $\varphi(X)$ sachant $Y = y$ est donnée par

$$Var(\varphi(X)/Y = y) = E((\varphi(X) - E(\varphi(X)/Y = y))^2 / Y = y)$$

2- Espérance conditionnelle par rapport à une variable aléatoire

Soient X et Y deux v.a. définies sur (Ω, \mathcal{F}, P) à valeurs dans \mathbb{R} . On a vu que $E(X/Y = y)$ est une fonction de y , $y \in Y(\Omega)$. Lorsque y varie dans $Y(\Omega)$, on peut donc définir sur (Ω, \mathcal{F}, P) une variable aléatoire, notée $E(X/Y)$, dont les valeurs sont $E(X/Y = y)$, $y \in Y(\Omega)$. Une telle variable est dite v.a. conditionnelle.

Définition 1.4.8 (Espérance conditionnelle par rapport à une v.a.)

On définit l'application $E(X/Y)$ sur (Ω, \mathcal{F}, P) comme suit

$$\begin{aligned} E(X/Y) : \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto E(X/Y)(\omega) = E(X/Y = y), \end{aligned}$$

où y est l'image de ω par $Y(\cdot)$, i.e.

$$y = Y(\omega).$$

Une telle application définit bien une v.a. sur (Ω, \mathcal{F}, P) . En effet, pour $a \in \mathbb{R}$,

$$\begin{aligned} \{\omega \in \Omega : E(X/Y)(\omega) \leq a\} &= \{\omega \in \Omega : E(X/Y = y) \leq a, \text{ avec } y = Y(\omega)\} \\ &= \{\omega \in \Omega : Y(\omega) \leq c(a)\} \\ &\in \mathcal{F}, \end{aligned}$$

où $c(a)$ est telle que

$$E(X/Y = y) = h(y) = h(Y(\omega)) \leq a \text{ ssi } Y(\omega) \leq c(a).$$

Propriétés

La variable conditionnelle $E(X/Y)$ vérifie les propriétés suivantes :

i) **Positivité** : Si $g(\cdot)$ est une fonction positive ou nulle alors $E(g(X)/Y) \geq 0$.

ii) **Constance** : $E(c/Y) = c$, pour tout $c \in \mathbb{R}$,

iii) **Linéarité** : $E(\alpha g(X) + \beta h(Y)/Z) = \alpha E(g(X)/Z) + \beta E(h(Y)/Z)$,

iv) **Projection (stabilité)** : $E(X/X) = X$,

Preuve : Pour montrer l'égalité entre les deux applications $E(X/X)$ et X , il convient de montrer que pour tout $\omega \in \Omega$, $X(\omega) = E(X/X)(\omega)$. Soit $\omega \in \Omega$ tel que $X(\omega) = x$. Alors on a à montrer que $x = E(X/X = x)$. Or (on ne perd pas de généralité en se restreignant au cas discret),

$$\begin{aligned} E(X/X = x) &= \sum_{k \in X(\Omega)} k P(X = k/X = x) \\ &= x P(X = x/X = x) \\ &= x, \end{aligned}$$

puisque $P(X = k/X = x) = 1$ si $k = x$ et 0 sinon.

v) **Projection (stabilité)** : $E(h(X)/X) = h(X)$.

Preuve : Similaire à la preuve de iv).

vi) **Projection (stabilité)** : $E(h(X)g(Y)/X) = h(X)E(g(Y)/X)$.

Preuve : Similaire à la preuve de iv).

vii) **Projection (stabilité)** : $E(E(X/Y)/Y) = E(X/Y)$.

Preuve : Le résultat s'ensuit du fait que $E(X/Y)$ est une fonction de Y (soit, $h(Y)$) et du résultat v).

viii) **Information redondante** : $E(X/Y, h(Y)) = E(X/Y)$.

Preuve : Similaire à la preuve de v).

ix) **Information minimale** : $E(E(X/Y, Z)/Y) = E(X/Y) = E(E(X/Y)/Y, Z)$.

x) **Loi des espérances totales** : $E(X) = E(E(X/Y))$.

xi) **Loi des variances totales** : $Var(X) = E(Var(X/Y)) + Var(E(X/Y))$.

xii) **Définition** : $E(g(X)h(Y)) = \int E((g(X)/Y = y))h(y)dF_Y(y) = E(E(g(X)/Y)h(Y))$.

3- Espérance conditionnelle par rapport à une tribu

Avant d'exposer le concept d'espérance conditionnelle par rapport à une tribu, concept le plus général concernant l'espérance conditionnelle, rappelons d'abord la notion de tribu engendrée par une variable aléatoire.

Soit (Ω, \mathcal{F}, P) un espace de probabilité sur lequel est définie une v.a. X réelle. On appelle tribu engendrée par X , tribu qu'on note \mathcal{F}_X , la plus petite tribu par rapport à laquelle X est mesurable. C'est aussi la tribu engendrée par les événements $\{\omega : X(\omega) = x\}$, $x \in X(\Omega)$. Intuitivement parlant, c'est la plus petite tribu qu'on peut former par les événements concernant X uniquement.

Exemple 1.4.1

On lance une pièce de monnaie deux fois de suite et soit (Ω, \mathcal{F}, P) l'espace de probabilité correspondant avec $\Omega = \{(P, P), (P, F), (F, P), (F, F)\}$, $\mathcal{F} =$

$$\mathcal{P}(\Omega) = \left\{ \begin{array}{l} \Omega, \emptyset, \{(P, P)\}, \{(P, F)\}, \{(F, P)\}, \{(F, F)\}, \{(P, P), (P, F)\}, \\ \{(P, P), (F, P)\}, \{(P, P), (F, F)\}, \{(P, F), (F, P)\}, \{(P, F), (F, F)\}, \\ \{(F, P), (F, F)\}, \{(P, P), (P, F), (F, P)\}, \{(P, P), (P, F), (F, F)\}, \\ \{(P, F), (F, P), (F, F)\}, \{(P, P), (F, P), (F, F)\} \end{array} \right\}$$

et P une mesure de probabilité définie sur (Ω, \mathcal{F}) .

Soient X , X_1 et X_2 les v.a. définies sur (Ω, \mathcal{F}, P) par

$$\begin{aligned} X : \Omega &\longrightarrow X(\Omega) = \{0, 1, 2\} \\ \omega &\longmapsto X(\omega) = \text{nombre de piles dans } \omega = (\omega_1, \omega_2). \end{aligned}$$

$$\begin{aligned} X_1 : \Omega &\longrightarrow X_1(\Omega) = \{0, 1\} \\ \omega &\longmapsto X_1((\omega_1, \omega_2)) = \text{nombre de piles dans } \omega_1. \end{aligned}$$

$$\begin{aligned} X_2 : \Omega &\longrightarrow X_2(\Omega) = \{0, 1\} \\ \omega &\longmapsto X_2((\omega_1, \omega_2)) = \text{nombre de piles dans } \omega_2. \end{aligned}$$

i) La tribu \mathcal{F}_{X_1} engendrée par X_1 est la plus petite tribu générée par

$$\begin{aligned} \{\{X_1 = x\}, x \in X_1(\Omega)\} &= \{\{X_1 = 0\}, \{X_1 = 1\}\} \\ &= \{\{(F, F), (F, P)\}, \{(P, P), (P, F)\}\}. \end{aligned}$$

Donc

$$\mathcal{F}_{X_1} = \{\Omega, \emptyset, \{(F, F), (F, P)\}, \{(P, P), (P, F)\}\}.$$

ii) De même, la tribu \mathcal{F}_{X_2} engendrée par X_2 est la plus petite tribu générée par

$$\begin{aligned} \{\{X_2 = x\}, x \in X_2(\Omega)\} &= \{\{X_2 = 0\}, \{X_2 = 1\}\} \\ &= \{\{(F, F), (P, F)\}, \{(P, P), (F, P)\}\}. \end{aligned}$$

Donc

$$\mathcal{F}_{X_2} = \{\Omega, \emptyset, \{(F, F), (P, F)\}, \{(P, P), (F, P)\}\}.$$

iii) Enfin, a tribu \mathcal{F}_X engendrée par X est la plus petite tribu générée par

$$\begin{aligned} \{\{X = x\}, x \in X(\Omega)\} &= \{\{X = 0\}, \{X = 1\}, \{X = 2\}\} \\ &= \{\{(F, F)\}, \{(P, F), (F, P)\}, \{(P, P)\}\}. \end{aligned}$$

Donc

$$\mathcal{F}_X = \left\{ \begin{array}{l} \Omega, \emptyset, \{(F, F)\}, \{(P, P)\}, \{(P, F), (F, P)\}, \{(P, P), (P, F), (F, P)\}, \\ \{(P, F), (F, P), (F, F)\}, \{(P, P), (F, F)\} \end{array} \right\}.$$

iv) La collection d'ensembles

$$\mathcal{B} = \{\Omega, \emptyset, \{(F, F)\}, \{(P, P), (P, F), (F, P)\}\}$$

est une tribu qui n'est pas engendrée par X puisque, par exemple, $X^{-1}([2, +\infty[) = \{(P, P)\} \notin \mathcal{B}$. i.e. X n'est pas mesurable par rapport à \mathcal{B} . De même $\mathcal{P}(\Omega)$ n'est pas la tribu engendrée par X car elle n'est pas la plus petite tribu générées par les évènements de X .

Remarque 1.4.4

Pour $c \in \mathbb{R}$, la tribu \mathcal{F}_c engendrée par la v.a. dégénérée c (dite aussi constante, triviale) définie sur (Ω, \mathcal{F}, P) par

$$\begin{aligned} c : \Omega &\longrightarrow \{c\} \\ \omega &\longmapsto X(\omega) = c \end{aligned}$$

est la plus petite tribu engendrée par $\{\{\omega : c(\omega) = c\}\} = \{\Omega\}$. Donc $\mathcal{F}_c = \{\Omega, \emptyset\}$ la tribu triviale.

Dans ce qui suit, on introduira l'intuition derrière la notion d'espérance conditionnelle par rapport à une tribu.

On a vu que $E(Y/A)$: espérance mathématique de Y (qu'on suppose discrète pour simplifier) sachant A s'est réalisé (A est donc de probabilité non nulle) est définie par

$$\begin{aligned} E(Y/A) &= \sum_{y \in Y(\Omega)} y P(Y = y/A) \\ &= \sum_{y \in Y(\Omega)} y \frac{P(Y = y, A)}{P(A)} \\ &= \frac{1}{P(A)} E(Y \mathbf{1}_A) \end{aligned}$$

où $Y \mathbf{1}_A$ est la v.a. définie par

$$\begin{aligned} Y \mathbf{1}_A : \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto Y \mathbf{1}_A(\omega) = \begin{cases} Y(\omega) & \text{si } \omega \in A \\ 0 & \text{sinon} \end{cases} \end{aligned}$$

Il est clair, par ailleurs, que les évènements A qu'on sait réalisés ou non forment une tribu. Par exemple, on sait que Ω se réalise toujours et que \emptyset ne se réalise jamais. Il est clair également que si on sait que A s'est ou non réalisé alors on sait aussi si son contraire A^c s'est

ou non réalisé, etc. Ces informations peuvent être déterminées par \mathcal{F}_A la tribu engendrée par A contenant l'information concernant la réalisation ou la non-réalisation de A . Cette tribu ne peut contenir plus d'informations concernant la réalisation de A que A lui-même et ainsi on peut substituer le nombre $E(Y/A)$ par $E(Y/\mathcal{F}_A)$.

De même, on peut considérer $E(Y/\mathcal{F}_X)$ comme étant une v.a. égale à $E(Y/X)$: espérance conditionnelle d'une v.a. mettons Y sachant X . Cette égalité est évidente puisque \mathcal{F}_X contient les mêmes informations contenue dans $\{\{X = x\}, x \in X(\Omega)\}$. $E(Y/\mathcal{F}_X)$ définit donc une v.a. mesurable par rapport à \mathcal{F}_X . Cependant, alors que toute v.a. X génère une tribu \mathcal{F}_X il n'est pas vrai que toute tribu détermine une v.a. par laquelle elle est la tribu engendrée. Ainsi, le concept d'espérance conditionnelle par rapport à une tribu est une stricte extension du concept d'espérance conditionnelle par rapport à une v.a. On donne maintenant la définition rigoureuse de l'espérance conditionnelle par rapport à une tribu.

Définition 1.4.9 (Espérance conditionnelle par rapport à une tribu)

Soient X une v.a. réelle définie sur un espace de probabilité (Ω, \mathcal{F}, P) , telle que $E|X| < \infty$ et \mathcal{B} une sous-tribu de \mathcal{F} . L'espérance conditionnelle de X par rapport à la tribu \mathcal{B} est une v.a. $E(X/\mathcal{B})$ vérifiant :

- i) $E(X/\mathcal{B})$ est \mathcal{B} -mesurable.
- ii) $E(X\mathbf{1}_A) = E(E(X/\mathcal{B})\mathbf{1}_A)$ pour tout $A \in \mathcal{B}$.

Remarque 1.4.5

- i) On peut montrer que $E(X/\mathcal{B})$ vérifiant i) et ii) existe pourvu que $E|X| < \infty$.
- ii) La définition de $E(X/\mathcal{B})$ est ambiguë dans le sens où $E(X/\mathcal{B})$ n'est pas unique, i.e. il peut y exister plus d'une espérance conditionnelle $E(X/\mathcal{B})$ vérifiant i) et ii). On parle alors de différentes versions d'espérance conditionnelle. Cependant, tout deux espérances conditionnelles vérifiant i) et ii) sont égales presque sûrement, i.e. si $E^{(1)}(X/\mathcal{B})$ et $E^{(2)}(X/\mathcal{B})$ sont deux espérances conditionnelles vérifiant i) et ii) alors

$$P(E^{(1)}(X/\mathcal{B}) \neq E^{(2)}(X/\mathcal{B})) = 0.$$

iii) La condition ii) peut être remplacée par :

ii') $E(XZ) = E(E(X/\mathcal{B})Z)$ pour toute v.a. Z bornée et \mathcal{B} -mesurable. En effet, il est clair que ii') entraîne ii) puisque $\mathbf{1}_A$ est une v.a. \mathcal{B} -mesurable pour $A \in \mathcal{B}$. La réciproque peut être établie en approximant une v.a. Z par une fonction étagée de la forme

$$\begin{aligned} Z_n : \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto Z_n(\omega) = \sum_{i=1}^n \alpha_{in} \mathbf{1}_{A_i}(\omega) \end{aligned}$$

où $\{A_1, \dots, A_n\}$ est une partition finie de Ω avec $A_1 \in \mathcal{B}$. Ainsi, ii') est satisfaite pour Z_n et en passant à la limite lorsque n tend vers l'infini, ii) est vérifiée pour Z (puisque toute v.a. Z bornée et \mathcal{B} -mesurable est la limite d'une suite de fonctions étagées).

Cas particuliers

i) Si $\mathcal{B} = \mathcal{F}_Y$ est la tribu engendrée par Y alors

$$E(X/\mathcal{F}_Y) = E(X/Y).$$

ii) Si $\mathcal{B} = \mathcal{F}$ la tribu de l'espace, alors

$$E(X/\mathcal{F}) = X.$$

iii) Si $\mathcal{B} = \mathcal{F}_c = \{\Omega, \emptyset\}$ la tribu triviale alors

$$E(X/\{\Omega, \emptyset\}) = E(X).$$

iv) Si $\Omega = \{\omega_1, \omega_2, \dots\}$ est dénombrable, $\mathcal{F} = \mathcal{P}(\Omega)$, P est une mesure de probabilité et \mathcal{B} est une sous-tribu de \mathcal{F} alors pour tout X définie sur (Ω, \mathcal{F}, P) , $E(X/\mathcal{B})$ est définie par

$$\begin{aligned} E(X/\mathcal{B}) : \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto E(X/\mathcal{B})(\omega) = E(X/A_i) \text{ lorsque } \omega \in A_i \end{aligned}$$

où A_1, A_2, \dots est une partition de Ω et où \mathcal{B} est la tribu engendrée par A_1, A_2, \dots . Ainsi, $E(X/\mathcal{B})$ est constante sur A_i .

Exemple 1.4.1 (suite)

i) Calculons $E(X/\mathcal{F}_{X_1})$. Puisque Ω est dénombrable et \mathcal{F}_{X_1} est la tribu engendrée par $\{(F, F), (F, P)\}, \{(P, P), (P, F)\}$ qui forme une partition de Ω donc

$$\begin{aligned} E(X/\mathcal{F}_{X_1}) : \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto E(X/\mathcal{F}_{X_1})(\omega) = E(X/A_i) \text{ lorsque } \omega \in A_i \end{aligned}$$

on a

$$\begin{aligned} (P, P) &\in \{(P, P), (P, F)\} \implies (P, P) \longmapsto E(X/\mathcal{F}_{X_1})((P, P)) = E(X/\{(P, P), (P, F)\}) \\ (P, F) &\in \{(P, P), (P, F)\} \implies (P, F) \longmapsto E(X/\mathcal{F}_{X_1})((P, F)) = E(X/\{(P, P), (P, F)\}) \\ (F, P) &\in \{(F, F), (F, P)\} \implies (F, P) \longmapsto E(X/\mathcal{F}_{X_1})((F, P)) = E(X/\{(F, F), (F, P)\}) \\ (F, F) &\in \{(F, F), (F, P)\} \implies (F, F) \longmapsto E(X/\mathcal{F}_{X_1})((F, F)) = E(X/\{(F, F), (F, P)\}) \end{aligned}$$

ii) Vérifier que $E(X/\mathcal{F}_{X_1}) = E(X/X_1)$.

iii) Déterminer $E(X/\mathcal{F}_{X_2})$.

vi) Déterminer $E(X/\mathcal{B})$ où \mathcal{B} est la tribu engendrée par $\{(P, P)\}, \{(P, F)\}, \{(F, P)\}, \{(F, F)\}$.

Propriétés

L'espérance conditionnelle par rapport à une tribu $E(X/\mathcal{B})$ vérifie les propriétés suivantes :

i) **Positivité** : Si $g(\cdot)$ est une fonction positive ou nulle alors $E(g(X)/\mathcal{B}) \geq 0$.

ii) **Constance** : $E(c/\mathcal{B}) = c$, pour tout $c \in \mathbb{R}$,

iii) **Linéarité** : $E(\alpha g(X) + \beta h(Y)/\mathcal{B}) = \alpha E(g(X)/\mathcal{B}) + \beta E(h(Y)/\mathcal{B})$,

iv) **Projection (stabilité)** : $E(X/\mathcal{B}) = X$ pour X \mathcal{B} -mesurable.

v) **Projection (stabilité)** : $E(h(X)/\mathcal{B}) = h(X)$ pour X \mathcal{B} -mesurable.

vi) **Projection (stabilité)** : $E(h(X)g(Y)/\mathcal{B}) = h(X)E(g(Y)/\mathcal{B})$ pour X \mathcal{B} -mesurable.

vii) **Projection (stabilité)** : $E(E(X/\mathcal{B})/\mathcal{B}) = E(X/\mathcal{B})$.

viii) **Information minimale** : $\mathcal{B} \subset \mathcal{H}$ alors $E(E(X/\mathcal{B})/\mathcal{H}) = E(X/\mathcal{B}) = E(E(X/\mathcal{H})/\mathcal{B})$.

xi) **Loi des espérances totales** : $E(X) = E(E(X/\mathcal{B}))$.

4- Distribution conditionnelle par rapport à une tribu

Soit (Ω, \mathcal{F}, P) un espace de probabilité sur lequel est définie une v.a. X réelle et soit B un borélien de $\mathcal{B}(\mathbb{R})$. On veut définir $P(X \in B/\mathcal{B})$ où \mathcal{B} est une sous-tribu de \mathcal{F} .

Définition 1.4.10 (Distribution conditionnelle par rapport à une tribu)

i) La distribution conditionnelle par rapport à la tribu \mathcal{B} est la mesure $P_X(\cdot/\mathcal{B})$ définie par

$$\begin{aligned} P_X(\cdot/\mathcal{B}) : \mathcal{B}(\mathbb{R}) &\longrightarrow [0, 1] \\ B &\longmapsto P_X(B/\mathcal{B}) = E(\mathbf{1}_B/\mathcal{B}). \end{aligned}$$

ii) La fonction de répartition conditionnelle $F_X(.|\mathcal{B})$ de X par rapport à la tribu \mathcal{B} est définie par

$$F_X(.|\mathcal{B}) : \mathbb{R} \longrightarrow [0, 1] \\ x \longmapsto F_X(x|\mathcal{B}) = P(X \leq x|\mathcal{B}) = E(\mathbf{1}_{[X \leq x]}|\mathcal{B}).$$

iii) Si la distribution $F_X(.|\mathcal{B})$ est absolument continue par rapport à la mesure de Lebesgue alors il existe une fonction $f_X(.|\mathcal{B})$, dite densité de Radon-Nikodym définie par

$$f_X(x|\mathcal{B}) = \frac{\partial F_X(x|\mathcal{B})}{\partial x}.$$

iv) Si X est discrète alors on définit la fonction de masse conditionnelle par rapport à comme \mathcal{B} suit :

$$p_X(.|\mathcal{B}) : \mathbb{R} \longrightarrow [0, 1] \\ x \longmapsto p_X(x|\mathcal{B}) = P(X = x|\mathcal{B}) = E(\mathbf{1}_{[X=x]}|\mathcal{B}).$$

vi) La variance conditionnelle de X par rapport à \mathcal{B} est définie par

$$\begin{aligned} \text{var}(X|\mathcal{B}) &= E((X - E(X|\mathcal{B}))^2|\mathcal{B}) \\ &= E(X^2|\mathcal{B}) - (E(X|\mathcal{B}))^2. \end{aligned}$$

Remarque 1.4.6

Contrairement à $F_{X/Y=y}(x)$ qui représente une valeur, $F_X(x|\mathcal{B})$ est une fonction \mathcal{B} -mesurable.

5- Espérance conditionnelle comme meilleure approximation

Le concept d'espérance conditionnelle est crucial dans les théories de l'estimation et de la prédiction statistiques. Etant données deux v.a. X et Y définies sur un même espace de probabilité (Ω, \mathcal{F}, P) , une fonction $h(Y)$ (étant une variable aléatoire) définie par $h(Y) = E(X|Y)$ est dite *fonction de regression* de X sur Y . Cette régression est vue comme *estimateur* des valeurs possible de X lorsque l'on dispose d'une information tirée en observant une valeur de Y . Quelle est la qualité de cet estimateur ?

En se référant au critère de l'erreur quadratique moyenne, l'estimateur $E(X|Y)$ représente, en fait, la meilleure approximation de X dans le sens où $E((X - Z)^2)$ est minimum pour $Z = E(X|Y)$, dans l'ensemble des v.a. $Z = h(X)$ fonction de X . Ce résultat est énoncé dans une forme plus générale dans le théorème suivant. Juste avant, on note que pour que le critère de l'erreur quadratique moyenne soit bien défini, les v.a. à considérer doivent être de carré intégrable, i.e. $E(Z^2) < \infty$. Ainsi on se restreint sur l'ensemble $L^2(\Omega, \mathcal{F}, P)$ de v.a. de carré intégrable définies sur (Ω, \mathcal{F}, P) .

Théorème 1.4.1

Soit X une v.a. de $L^2(\Omega, \mathcal{F}, P)$ et \mathcal{B} une sous-tribu de \mathcal{F} . Alors,

$$E(X - E(X|\mathcal{B}))^2 = \min_{Z \in L^2(\Omega, \mathcal{B}, P)} (E(X - Z)^2).$$

Preuve

Pour toute v.a. $Z \in L^2(\Omega, \mathcal{B}, P)$, on a

$$\begin{aligned} E(X - Z)^2 &= E((X - E(X|\mathcal{B})) - (Z - E(X|\mathcal{B})))^2 \\ &= E(X - E(X|\mathcal{B}))^2 + E((Z - E(X|\mathcal{B}))^2) - 2E[(X - E(X|\mathcal{B}))(Z - E(X|\mathcal{B}))]. \end{aligned}$$

Or, en vertu de la \mathcal{B} -mesurabilité de Z on a

$$\begin{aligned}
E[(X - E(X/\mathcal{B}))(Z - E(X/\mathcal{B}))] &= E[E[(X - E(X/\mathcal{B}))(Z - E(X/\mathcal{B}))/\mathcal{B}]] \\
&= E[(Z - E(X/\mathcal{B}))E[(X - E(X/\mathcal{B}))/\mathcal{B}]] \\
&= E[(Z - E(X/\mathcal{B}))[E(X/\mathcal{B}) - E(E(X/\mathcal{B})/\mathcal{B})]] \\
&= E[(Z - E(X/\mathcal{B}))[E(X/\mathcal{B}) - E(X/\mathcal{B})]] \\
&= 0.
\end{aligned}$$

Donc

$$\begin{aligned}
E(X - Z)^2 &= E(X - E(X/\mathcal{B}))^2 + E((Z - E(X/\mathcal{B}))^2) \\
&\geq E((Z - E(X/\mathcal{B}))^2).
\end{aligned}$$

■

Ainsi, l'espérance conditionnelle peut être considérée comme opérateur de projection de l'espace $L^2(\Omega, \mathcal{F}, P)$ dans l'espace $L^2(\Omega, \mathcal{B}, P)$.

6- Variables aléatoires indépendantes, indépendantes en moyenne et non corrélation

Le concept d'indépendance entre deux événements peut être étendu au cas de variables aléatoires.

Définition 1.4.11 (Variables aléatoires indépendantes)

i) Des v.a. X_1, X_2, \dots, X_n définies sur (Ω, \mathcal{F}, P) à valeurs réelles sont dites indépendantes si pour tout boréliens B_1, B_2, \dots, B_n , les événements $\{X_1 \in B_1\}, \{X_2 \in B_2\}, \dots, \{X_n \in B_n\}$ sont mutuellement indépendants.

ii) Une suite $(X_n, n \in \mathbb{N})$ de v.a. définies sur (Ω, \mathcal{F}, P) à valeurs réelles est dite indépendante si pour tout $n \geq 2$ les v.a. X_1, X_2, \dots, X_n sont indépendantes.

Il est clair que si X_1, X_2, \dots, X_n sont indépendantes alors

$$P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) = P(X_1 \leq x_1) P(X_2 \leq x_2) \dots P(X_n \leq x_n)$$

pour tout $x_1, x_2, \dots, x_n \in \mathbb{R}$. Il suffit de prendre $B_i =]-\infty, x_i]$ ($i = 1, \dots, n$). La réciproque est cependant plus utile et plus profonde.

Théorème 1.4.2

Si pour tout $x_1, x_2, \dots, x_n \in \mathbb{R}$, on a

$$P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) = P(X_1 \leq x_1) P(X_2 \leq x_2) \dots P(X_n \leq x_n)$$

alors les v.a. X_1, X_2, \dots, X_n sont indépendantes.

Conséquences

i) Des v.a. X_1, X_2, \dots, X_n sont indépendantes si et seulement si les tribus correspondantes $\mathcal{F}_{X_1}, \mathcal{F}_{X_2}, \dots, \mathcal{F}_{X_n}$ sont indépendantes dans le sens de la Définition 1.2.5.

ii) Si X_1, X_2, \dots, X_n sont indépendantes alors pour toutes fonctions mesurables f_1, f_2, \dots, f_n , les v.a. $f_1(X_1), f_2(X_2), \dots, f_n(X_n)$ sont également indépendantes.

iii) Si X_1, X_2, \dots, X_n sont indépendantes alors

$$E[f_1(X_1) f_2(X_2) \dots f_n(X_n)] = E(f_1(X_1)) E(f_2(X_2)) \dots E(f_n(X_n))$$

pourvu que les espérances précédentes existent.

iv) Cependant, la reciproque de iii) n'est pas vraie : On pourrait avoir deux v.a. X et Y telles que

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

mais qui ne soient pas independantes. En revanche, l'independance exige la condition forte suivante :

X et Y sont independantes si et seulement si pour tout couple de fonctions mesurables $(f(\cdot), g(\cdot))$ on a

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

iv) Si X_1, X_2, \dots, X_n sont independantes alors $E(X_{i_1}/X_{i_2}, \dots, X_{i_n}) = E(X_{i_1})$.

V- Convergence stochastique et théorèmes de limites

On a vu plus haut, lors de l'interprétation de la probabilité en fréquentiste et subjective que l'interprétation fréquentiste se base sur le fait observé par Bernoulli et qui stipule que la fréquence relative d'apparition d'un évènement A converge "dans un certain sens à préciser plus bas", lorsque le nombre d'expériences augmente indéfiniment, vers un nombre $P(A)$ qu'on a baptisé *probabilité de A*. Ce résultat fait émerger, au milieu du désordre, un ordre, une loi qui n'est valide que pour de grands nombres d'expériences, aussi grands que l'on veut. Voilà pourquoi cette loi est dite *loi des grands nombres*. En terme plus précis si n_A désigne le nombre de réalisations de l'évènement A sur le nombre n de répétitions de l'expérience aléatoire alors la loi des grands nombres stipule que la fréquence relative d'apparition de A , qui n'est autre que $\frac{n_A}{n}$, converge "dans un certain sens" vers $P(A)$ lorsque $n \rightarrow \infty$. Or si on associe à la suite des répétitions de l'expérience la suite de variables aléatoires X_1, X_2, \dots , où $X_i = 1$ si l'évènement A se réalise et 0 sinon, alors la fréquence relative de A n'est autre que la moyenne empirique $\frac{1}{n}(X_1 + \dots + X_n)$ et $P(A) = E(X_1)$ (puisque X_i suit une loi de Bernoulli de paramètre $P(A)$). Ainsi, la loi des grands nombres énoncée plus haut peut aussi s'exprimer sous forme plus générale : $\frac{1}{n}(X_1 + \dots + X_n)$ converge "dans un certain sens" vers $E(X_1)$, signifiant que la suite des moyennes empiriques $(\bar{X}_n, n \in \mathbb{N})$ converge vers une limite qui n'est autre que la moyenne théorique, $E(X_1)$. Cette loi représente l'un des resultats centraux de la théorie des probabilités et de la statistique. Une autre loi toute aussi importante et centrale, d'ailleurs souvent nommée *théorème central limite*, stipule que la somme "pendérée" de n v.a. vérifiant certaines conditions converge "dans un certain sens" vers une loi normale. Cette loi représente la justification et le fondement du fait bien reconnu et hypothétique que la loi normale peut représenter la loi de bruits, d'innovations, de chocs, d'erreurs, des erreurs de modèles..., puisque ces erreurs sont en fait le cumul, la conjugaison et l'interaction d'une infinité de chocs indépendants. Cependant, jusque là, nous avons parlé de convergence dans un certain sens sans avoir à préciser aucun sens. Puisqu'il s'agit là de suite de variables aléatoires, donc de fonctions réelles, il existe plusieurs modes possibles de convergence d'une suite de v.a. vers une autre v.a. dont on distingue : convergence presque sûre, convergence en probabilité, convergence en moyenne quadratique et convergence en distribution. Dans cette section on exposera ces différents modes de convergence et les deux théorèmes limites les plus important de la théorie des probabilités : loi des grands nombres et théorème central limite. En fait, il existe plusieurs versions de la loi des grands nombre et

du théorème central limite en fonction du mode de convergence (par exemple, loi faible des grands nombres, loi forte des grands nombres...) et de la structure de dépendance de la suite de v.a. (loi des grands nombres pour v.a. i.i.d., théorème central limite pour v.a. i.i.d.,...). Avant d'exposer ces concepts, on a besoin de rappeler la notion de suite de v.a. et quelques inégalités connues et dont on aura besoin par la suite.

1- Suite de variables aléatoires

Pour plusieurs problèmes, spécialement lorsqu'il s'agit d'étudier l'évolution d'un phénomène dans le temps ou dans l'espace, on a besoin de considérer plutôt le concept de *suites de variables aléatoires*. Cependant des problèmes théoriques surgissent nécessitant des concepts de limites, concept dont on n'avait pas besoin lors de l'étude des vecteurs aléatoires ou plus généralement d'ensembles finis de v.a. L'étude des suites de v.a. est un sous-produit de la théorie des processus aléatoires, théorie dont on reportera l'étude au chapitre suivant.

Définition 1.5.1 (Suite de v.a.)

Une application $\mathbb{X}(\cdot) = (X_0(\cdot), X_1(\cdot), \dots)$ définie sur (Ω, \mathcal{F}, P) à valeurs dans $\mathbb{R}^{\mathbb{N}} = \{x = (x_0, x_1, \dots), x_i \in \mathbb{R}, i \in \mathbb{N}\}$, c'est à dire qui pour tout résultat de l'expérience ω associe une suite de nombres $(X_0(\omega), X_1(\omega), \dots)$, est dite suite de variables aléatoires sur (Ω, \mathcal{F}, P)

si pour tout $i \in \mathbb{N}$ et pour tout nombre $a_i \in \mathbb{R}$, l'ensemble $\left\{ \omega \in \Omega : \bigcap_{i \in \mathbb{N}} X_i(\omega) \leq a_i \right\} \in \mathcal{F}$.

Autrement dit, pour tout $i \in \mathbb{N}$, l'image réciproque de tout Borélien B_i de $\mathcal{B}(\mathbb{R})$ par X_i est un évènement de \mathcal{F} .

Une conséquence de la Définition 1.5.1 et du fait que \mathcal{F} soit une tribu est que pour tout ensemble fini de nombres $\{a_{i_1}, a_{i_2}, \dots, a_{i_n}\}$, l'ensemble

$$\{\omega \in \Omega : X_{i_1}(\omega) \leq a_{i_1}, X_{i_2}(\omega) \leq a_{i_2}, \dots, X_{i_n}(\omega) \leq a_{i_n}\}$$

est dans \mathcal{F} pour tout $n \in \mathbb{N}^*$ et pour tout $i_1, i_2, \dots, i_n \in \mathbb{N}$.

Comme pour les vecteurs aléatoires on peut définir la distribution de probabilité d'une suite de v.a.

Définition 1.5.2 (Distribution infini-dimensionnelle d'une suite de v.a.)

Soit $\mathbb{X} = (X_0(\cdot), X_1(\cdot), \dots)$ une suite de v.a. définie sur un espace de probabilité (Ω, \mathcal{F}, P) . On appelle distribution (infini-dimensionnelle) de \mathbb{X} la fonction ensembliste $P_{\mathbb{X}}(\cdot)$ définie sur $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$ à valeurs dans $[0, 1]$ par

$$P_{\mathbb{X}}(\cdot) : \mathcal{B}(\mathbb{R}^{\mathbb{N}}) \longrightarrow [0, 1]$$

$$\mathbb{B} \longmapsto P_{\mathbb{X}}(\mathbb{B}) = P(\omega : \mathbb{X}(\omega) \in \mathbb{B}) = P\left(\omega : \bigcap_{i=0}^{\infty} X_i(\omega) \in B_i\right) = P\mathbb{X}^{-1}(\mathbb{B})$$

où $\mathbb{B} = B_1 \times B_2 \times \dots$ avec $B_i \in \mathcal{B}(\mathbb{R})$. Puisque la tribu de Borel $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$ est aussi la plus petite tribu engendrée par les intervalles de la forme $] - \infty, x_0] \times] - \infty, x_1] \times \dots$, donc la donnée de $P_{\mathbb{X}}(\cdot)$ est équivalente à la donnée d'une fonction $F_{\mathbb{X}}(\cdot)$ plutôt réelle, dite fonction de répartition infini-dimensionnelle de la suite \mathbb{X} , et qui est définie par

$$F_{\mathbb{X}}(\cdot) : \mathbb{R}^{\mathbb{N}} \longrightarrow [0, 1]$$

$$\mathbf{x} = (x_0, x_1, \dots)' \longmapsto F_{\mathbb{X}}(\mathbf{x}) = P\left(\omega : \bigcap_{i=1}^{\infty} X_i(\omega) \leq x_i\right).$$

La fonction de répartition $F_{\mathbb{X}}(\cdot)$ est aussi dite distribution (infini-dimensionnelle) de probabilité de \mathbb{X} ou distribution conjointe de X_0, X_1, \dots

La distribution infini-dimensionnelle est difficile à manipuler puisqu'elles nécessitent des outils analytiques puissants faisant appel aux limites. Le théorème suivant qui est une version du théorème d'extension permet de proposer un outil pour simplifier l'analyse de la distribution infini-dimensionnelle : Il s'agit de ce qu'on appelle *distributions fini-dimensionnelles* d'une suite de v.a. On en expose d'abord la définition.

Définition 1.5.3 (Distribution fini-dimensionnelle d'une suite de v.a.)

Soit $\mathbb{X} = (X_0(\cdot), X_1(\cdot), \dots)$ une suite de v.a. définie sur un espace de probabilité (Ω, \mathcal{F}, P) . Pour $n \in \mathbb{N}^*$ et pour tout $i_1, i_2, \dots, i_n \in \mathbb{N}$, on appelle *distribution fini-dimensionnelle* de \mathbb{X} la fonction ensembliste $P_{X_{i_1}, X_{i_2}, \dots, X_{i_n}}(\cdot)$ définie sur $\mathcal{B}(\mathbb{R}^n)$ à valeurs dans $[0, 1]$ par

$$P_{X_{i_1}, X_{i_2}, \dots, X_{i_n}}(\cdot) : \mathcal{B}(\mathbb{R}^n) \longrightarrow [0, 1]$$

$$\tilde{B} \longmapsto P_{X_{i_1}, X_{i_2}, \dots, X_{i_n}}(\tilde{B}) = P \left(\omega : \bigcap_{j=1}^n X_{i_j}(\omega) \in B_{i_j} \right).$$

Ainsi la distribution fini-dimensionnelle d'une suite \mathbb{X} est la distribution de toute sous-suite finie $(X_{i_1}, X_{i_2}, \dots, X_{i_n})$ de \mathbb{X} .

Théorème 1.5.1 (Théorème d'extension)

Soit $\mathbb{X} = (X_0(\cdot), X_1(\cdot), \dots)$ une suite de v.a. définie sur un espace de probabilité (Ω, \mathcal{F}, P) à valeurs dans $\mathbb{R}^{\mathbb{N}}$. Alors les probabilités infini-dimensionnelles

$$P(\omega : \mathbb{X}(\omega) \in \mathbb{B}) = P \left(\omega : \bigcap_{i=0}^{\infty} X_i(\omega) \in B_i \right),$$

(\mathbb{B} étant un Borélien de $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$) sont uniquement déterminées par les probabilités fini-dimensionnelles $P \left(\omega : \bigcap_{j=1}^n X_{i_j}(\omega) \in B_{i_j} \right)$ pour tout $n \in \mathbb{N}^*$ et pour tout $i_1, i_2, \dots, i_n \in \mathbb{N}$.

2- Quelques inégalités

Dans ce qui suit on présentera certaines inégalités connues et utiles dans le développement asymptotique en théorie des probabilités et qui le plus souvent font intervenir l'espérance mathématique d'une v.a. Toutes les v.a. qui apparaissent sont supposées définies sur un espace de probabilités (Ω, \mathcal{F}, P) .

Lemme 1.5.1 (Inégalité de Markov)

Soit Y une v.a. réelle non négative définie sur (Ω, \mathcal{F}, P) telle que $E(Y) < \infty$. Alors,

$$\forall \eta > 0, \quad P(Y \geq \eta) \leq \frac{E(Y)}{\eta}.$$

Preuve

On suppose sans perte de généralité que Y est absolument continue par rapport à la mesure

de Lebesgue. On a pour tout $\eta > 0$,

$$\begin{aligned}
 E(Y) &= \int_0^{+\infty} y f_Y(y) dy \\
 &= \int_0^{\eta} y f_Y(y) dy + \int_{\eta}^{+\infty} y f_Y(y) dy \\
 &\geq \int_{\eta}^{+\infty} y f_Y(y) dy \\
 &\geq \int_{\eta}^{+\infty} \eta f_Y(y) dy = \eta P(Y \geq \eta).
 \end{aligned}$$

■

Lemme 1.5.2 (Inégalité de Tchebychev)

Soit X une v.a. réelle définie sur (Ω, \mathcal{F}, P) telle que $E(Y^2) < \infty$. Alors,

$$\forall \varepsilon > 0, \quad P(|X - E(X)| \geq \varepsilon) \leq \frac{\text{var}(Y)}{\varepsilon^2}.$$

Preuve

En posant $Y = |X - E(X)|^2$, alors Y est non négative et comme $E(X^2) < \infty$ donc $E(Y) < \infty$. Ainsi, les hypothèses d'applicabilité de l'inégalité de Markov sont satisfaites. En appliquant celle-ci à Y , on a

$$\forall \eta > 0, \quad P(Y \geq \eta) \leq \frac{E(Y)}{\eta},$$

ce qui équivaut à

$$\forall \eta > 0, \quad P(|X - E(X)|^2 \geq \eta) \leq \frac{E(|X - E(X)|^2)}{\eta},$$

ou encore à

$$\forall \eta > 0, \quad P(|X - E(X)| \geq \sqrt{\eta}) \leq \frac{\text{var}(X)}{\eta}.$$

En posant $\varepsilon = \sqrt{\eta}$ (donc $\eta = \varepsilon^2$), sachant que la fonction $x \mapsto \sqrt{x}$ est bijective sur $[0, +\infty[$, on sait que pour tout $\eta > 0$ correspond un et un unique $\varepsilon > 0$, vérifiant

$$\forall \varepsilon > 0, \quad P(|X - E(X)| \geq \varepsilon) \leq \frac{\text{var}(X)}{\varepsilon^2}.$$

■

Remarque 1.5.1

De façon générale, le théorème de Tchebychev peut se formuler comme suit.

$$\forall \varepsilon > 0, \quad P(X \geq \varepsilon) \leq \frac{E(g(X))}{g(\varepsilon)},$$

pour toute fonction g positive, mesurable, strictement croissante et telle que $E(g(X)) < \infty$. Il suffit d'appliquer l'inégalité de Markov à la v.a. $Y = g(X)$.

3- Convergence stochastique

On a vu plus haut qu'une suite de v.a. $\mathbb{X} = (X_0(\cdot), X_1(\cdot), \dots)$ (notée aussi $(X_n, n \in \mathbb{N})$) définie sur (Ω, \mathcal{F}, P) est une application probabilisable de Ω dans $\mathbb{R}^{\mathbb{N}}$ qui pour tout issu ω associe une suite de nombres réels $\mathbb{X}(\omega) = (X_0(\omega), X_1(\omega), \dots)$. Comme pour le cas de suites réelles, il est donc très important de connaître le comportement des suites $\mathbb{X}(\omega) = (X_n(\omega), n \in \mathbb{N})$ lorsque n croît indéfiniment. Cependant la question n'est pas aussi simple que pour les suites réelles, puisque dans le cas de suite de v.a. il existe plusieurs façons (ou modes) selon lesquelles une suite de v.a. \mathbb{X} peut s'approcher ou tendre vers une limite X . Un premier critère est celui de convergence d'une suite $\mathbb{X}(\omega) = (X_n(\omega), n \in \mathbb{N})$ pour un $\omega \in \Omega$ donné.

Définition 1.5.4 (Convergence en un issu)

Une suite de v.a. $(X_n, n \in \mathbb{N})$ définie sur (Ω, \mathcal{F}, P) est dite avoir une limite (ou convergente) en un point $\omega \in \Omega$ si la suite de nombres réels $\mathbb{X}(\omega) = (X_n(\omega), n \in \mathbb{N})$ converge vers une limite $X(\omega)$.

Cette définition n'est pas utile en soi puisqu'on ne s'intéresse pas au comportement d'une suite en un résultat spécifique ω de l'expérience. Cependant elle peut servir pour d'autres concepts de limite plus importants.

Il est utile de noter qu'une suite de v.a. étant une suite de fonctions définies sur Ω , on peut en définir le concept de convergence simple.

Définition 1.5.5 (Convergence simple d'une suite de v.a.)

Une suite de v.a. $(X_n, n \in \mathbb{N})$ définie sur (Ω, \mathcal{F}, P) est dite converger simplement vers une v.a. X définie sur le même espace (Ω, \mathcal{F}, P) , si pour tout $\omega \in \Omega$,

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega).$$

Autrement dit,

$$\forall \omega \in \Omega, \forall \varepsilon > 0, \exists N(\omega, \varepsilon) > 0 : \forall n \geq N(\omega, \varepsilon), |X_n(\omega) - X(\omega)| < \varepsilon.$$

Une telle définition est cependant trop forte puisqu'elle exige la convergence pour tout $\omega \in \Omega$; en pratique, il existe rarement de suites de v.a. convergeant simplement vers une v.a. C'est pourquoi, une telle définition est allégée en exigeant la convergence seulement sur un ensemble particulier inclu dans Ω .

Définition 1.5.5 (Ensemble de convergence)

Etant donnée une suite de v.a. $(X_n, n \in \mathbb{N})$ définie sur (Ω, \mathcal{F}, P) , l'ensemble de points $\omega \in \Omega$ pour lequel $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$ est dit ensemble de convergence de la suite $(X_n, n \in \mathbb{N})$.

Il est possible de montrer que l'ensemble de convergence d'une suite de v.a. est un évènement dans \mathcal{F} , donc probabilisable. Ainsi, on peut parler de la probabilité de convergence d'une suite $(X_n, n \in \mathbb{N})$ en faisant référence à la probabilité de son ensemble de convergence.

Une définition de convergence moins forte que la convergence simple mais qui reste quand même assez forte est le concept de convergence *presque sûre*.

Définition 1.5.5 (Convergence presque sûre)

On dit qu'une suite de v.a. $(X_n, n \in \mathbb{N})$ définie sur (Ω, \mathcal{F}, P) converge presque sûrement vers une v.a. X définie sur (Ω, \mathcal{F}, P) et on écrit $X_n \xrightarrow[p.s.]{n \rightarrow \infty} X$, si son ensemble de convergence est de probabilité 1. Autrement dit, si elle converge sauf dans un ensemble $E \in \mathcal{F}$ de probabilité nulle. On écrit alors

$$P\left(\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right) = 1.$$

Exemple 1.5.1

Soit $(X_n, n \in \mathbb{N})$ une suite v.a. indépendantes et identiquement distribuées (i.i.d.) définie sur (Ω, \mathcal{F}, P) et de fonction de répartition $F(\cdot)$. On suppose que $F(x) < 1$ pour tout x . On considère la suite des maxima $(X_{(n)}, n \in \mathbb{N}^*)$ définie par $X_{(n)} = \max_{1 \leq i \leq n} \{X_i\}$, $n \geq 1$. On va montrer que

$$X_{(n)} \xrightarrow[n \rightarrow \infty]{p.s.} \infty.$$

On note que la suite $(X_{(n)}(\omega), n \in \mathbb{N}^*)$ est croissante pour tout $\omega \in \Omega$, et en considérant ∞ comme élément de $\overline{\mathbb{R}} = [0, \infty]$, donc $\lim_{n \rightarrow \infty} X_{(n)}(\omega)$ existe pour tout $\omega \in \Omega$. Pour montrer que $X_{(n)} \xrightarrow[n \rightarrow \infty]{p.s.} \infty$, il suffit de montrer que

$$P(\omega : \exists M > 0 : \forall n_0 \in \mathbb{N}, n \geq n_0 \text{ et } X_{(n)}(\omega) \leq M) = 0.$$

ou encore

$$P\left(\omega : \exists M > 0 : \forall n_0 \in \mathbb{N}, \bigcap_{n \geq n_0} X_{(n)}(\omega) \leq M\right) = 0.$$

Si on pose $A = \bigcap_{n \geq 1} \{\omega : X_{(n)}(\omega) \leq M\}$ et $A_k = \bigcap_{1 \leq n \leq k} \{\omega : X_{(n)}(\omega) \leq M\}$, alors il est clair que

$$A \subset A_k \text{ pour tout } k \in \mathbb{N}^*,$$

et donc

$$P(A) \leq P(A_k) \text{ pour tout } k \in \mathbb{N}^*,$$

En particulier, en passant à la limite quand $k \rightarrow \infty$, on trouve

$$\begin{aligned} P(A) &\leq \lim_{k \rightarrow \infty} P(A_k) \\ &= \lim_{k \rightarrow \infty} P\left(\bigcap_{1 \leq n \leq k} \{\omega : X_{(n)}(\omega) \leq M\}\right) \\ &= \lim_{k \rightarrow \infty} P(\omega : X_{(n)}(\omega) \leq M)^k = \lim_{k \rightarrow \infty} F(M)^k = 0. \end{aligned}$$

■

On peut exhiber des exemples dans lesquels la limite $\lim_{n \rightarrow \infty} X_n(\omega)$ ne converge pas null part (pour tout ω) mais que $P(\omega : |X_n(\omega) - X(\omega)| \geq \varepsilon)$ tend à être petite pour tout $\varepsilon > 0$. Ceci conduit à une définition plus faible de convergence : la convergence en probabilité.

Définition 1.5.6 (Convergence en probabilité)

On dit qu'une suite de v.a. $(X_n, n \in \mathbb{N})$ définie sur (Ω, \mathcal{F}, P) converge en probabilité vers une v.a. X définie sur (Ω, \mathcal{F}, P) et on écrit $X_n \xrightarrow[n \rightarrow \infty]{p} X$, si

$$\forall \varepsilon > 0 : \lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0.$$

Exemple 1.5.2

Soit $(X_n, n \in \mathbb{N})$ une suite v.a. indépendantes et identiquement distribuées (i.i.d.) définie sur (Ω, \mathcal{F}, P) de moyenne μ et de variance σ^2 . Soit $(S_n, n \in \mathbb{N})$ la suite des sommes partielles associées à $(X_n, n \in \mathbb{N})$, i.e. $S_n = X_1 + \dots + X_n$. Alors on va montrer que

$$\frac{1}{n} S_n \xrightarrow[n \rightarrow \infty]{p} \mu.$$

Puisque $E(S_n) = n\mu$ et $\text{var}(S_n) = n\sigma^2$, en utilisant l'inégalité de Tchebychev on sait que

$$\forall \varepsilon > 0 : \lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} S_n - E \left(\frac{1}{n} S_n \right) \right| > \varepsilon \right) \leq \frac{\text{var}(S_n)}{\varepsilon^2}.$$

Donc

$$\begin{aligned} \forall \varepsilon > 0 : \lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} S_n - \mu \right| > \varepsilon \right) &\leq \frac{\text{var}(S_n)}{\varepsilon^2} \\ &\xrightarrow{n \rightarrow \infty} 0, \end{aligned}$$

montrant que $\frac{1}{n} S_n \xrightarrow[n \rightarrow \infty]{p} \mu$. ■

Le concept de convergence en probabilité fait surgir de fait d'autres concepts tels que la borne en probabilité.

Définition 1.5.7

On dit qu'une suite $(X_n, n \in \mathbb{N})$ définie sur (Ω, \mathcal{F}, P) à valeurs dans \mathbb{R} est bornée en probabilité si

$$\forall \varepsilon > 0 \exists \delta(\varepsilon) > 0 : \forall n \in \mathbb{N}, P(|X_n| > \delta(\varepsilon)) < \varepsilon.$$

La notion de convergence en probabilité peut être étendue sans peine au cas de suites de vecteurs aléatoires $(\underline{X}_n, n \in \mathbb{N})$ définis sur (Ω, \mathcal{F}, P) à valeurs dans \mathbb{R}^m ($m \in \mathbb{N}^*$).

On dit alors que la suite $(\underline{X}_n, n \in \mathbb{N})$ converge en probabilité vers un vecteur aléatoire \underline{X} (et on note $\underline{X}_n \xrightarrow{p} \underline{X}$) si

$$\forall \varepsilon > 0 : \lim_{n \rightarrow \infty} P(\|\underline{X}_n - \underline{X}\| > \varepsilon) = 0,$$

où $\|\cdot\|$ désigne la norme euclidienne, i.e. $\|V\| = \sqrt{\sum_{j=1}^m V_j^2}$ avec $V = (V_1, \dots, V_m)'$.

On démontre facilement le résultat suivant.

Théorème 1.5.2 (Stabilité de la convergence en probabilité)

Si $\underline{X}_n \xrightarrow{p} \underline{X}$ et si $g : \mathbb{R}^m \rightarrow \mathbb{R}^k$ est une fonction continue, alors $g(\underline{X}_n) \xrightarrow{p} g(\underline{X})$.

Plusieurs concepts de limite sont basés sur la notion de distance. Un cas particulier dans le cas de convergence stochastique est le suivant.

Définition 1.5.8 (Convergence en moyenne quadratique)

On dit qu'une suite de v.a. $(X_n, n \in \mathbb{N})$ définie sur (Ω, \mathcal{F}, P) converge en moyenne quadratique vers une v.a. X définie sur (Ω, \mathcal{F}, P) et on écrit $X_n \xrightarrow[n \rightarrow \infty]{m.q.} X$, si

$$\lim_{n \rightarrow \infty} E(|X_n - X|^2) = 0.$$

En général, lorsque pour $r \geq 1$, on a $\lim_{n \rightarrow \infty} E(|X_n - X|^r) = 0$, alors on dit que $(X_n, n \in \mathbb{N})$ converge en moyenne d'ordre r vers X et on écrit $X_n \xrightarrow[n \rightarrow \infty]{r} X$.

Un autre mode de convergence très répandu et qui n'implique pas directement les valeurs de X_n mais plutôt leurs distributions est connu sous le nom convergence en distribution.

Définition 1.5.9 (Convergence en distribution)

On dit qu'une suite de v.a. $(X_n, n \in \mathbb{N})$ définie sur (Ω, \mathcal{F}, P) converge en distribution (ou en loi) vers une v.a. X définie sur (Ω, \mathcal{F}, P) et on écrit $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$, si

$$\lim_{n \rightarrow \infty} F_n(x) = F(x),$$

pour tout x où F est continue, où $F_n(\cdot)$ et $F(\cdot)$ sont respectivement les fonctions de répartition de X_n et X .

La définition peut être généralisée au cas de suite de vecteurs aléatoires.

Le résultat suivant donne un pont entre convergence de vecteurs et convergence de v.a. scalaires.

Théorème 1.5.3 (Cramér-Wold devise)

Soit $(\underline{X}_n, n \in \mathbb{N})$ une suite de vecteurs aléatoires définis sur (Ω, \mathcal{F}, P) à valeurs dans \mathbb{R}^m . Alors

$$\underline{X}_n \xrightarrow{\mathcal{L}} \underline{X} \text{ ssi } \lambda' \underline{X}_n \xrightarrow{\mathcal{L}} \lambda' \underline{X}, \quad \forall \lambda = (\lambda_1, \dots, \lambda_m)' \in \mathbb{R}^m.$$

La stabilité de convergence par application de fonctions continues reste valable pour la convergence en loi.

Théorème 1.5.4 (Stabilité de la convergence en loi)

Si $\underline{X}_n \xrightarrow{\mathcal{L}} \underline{X}$ et si $g : \mathbb{R}^m \rightarrow \mathbb{R}^k$ est une fonction continue, alors $g(\underline{X}_n) \xrightarrow{\mathcal{L}} g(\underline{X})$.

Remarque 1.5.2

Pour établir une convergence en loi la manipulation de la convergence de suites de fonctions (de répartition) est souvent délicate. Cependant, le théorème suivant permet d'établir la convergence en loi à partir de la convergence de fonctions caractéristiques dont la manipulation est plus simple.

Théorème 1.5.5 (de Levy)

$X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$ si et seulement si $\varphi_{X_n}(t) := E(e^{itX_n}) \rightarrow \varphi_X(t) := E(e^{itX})$ pour tout $t \in \mathbb{R}$.

Relation entre modes de convergence

Les résultats suivants font apparaître la relation entre les différents modes étudiés.

Théorème 1.5.6 (Convergence p.s. implique convergence en probabilité)

$X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$ implique $X_n \xrightarrow[n \rightarrow \infty]{p} X$.

Preuve

En supposant que $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$, soit pour $\varepsilon > 0$ l'évènement $B_n(\varepsilon) = \bigcup_{k \geq n} \{|X_k - X| \geq \varepsilon\}$,

$n \geq 1$. La suite d'évènements $\{B_n(\varepsilon), n \geq 1\}$ est décroissante et donc tendant vers $B(\varepsilon) = \bigcap_{n \geq 1} B_n(\varepsilon)$. Il est clair que $B(\varepsilon)$ est un sous-ensemble du complément de l'ensemble de convergence de $(X_n, n \in \mathbb{N})$, pour tout $\varepsilon > 0$. Ainsi, $P(B(\varepsilon)) = 0$ et donc $\lim_{n \rightarrow \infty} P(B_n(\varepsilon)) = 0$.

Or, pour tout $\varepsilon > 0 : \{|X_k - X| \geq \varepsilon\} \subset B_n(\varepsilon)$ et donc

$$\begin{aligned} 0 &\leq P(|X_k - X| \geq \varepsilon) \leq P(B_n(\varepsilon)) \\ &\xrightarrow[n \rightarrow \infty]{} 0, \end{aligned}$$

montrant que $X_n \xrightarrow[n \rightarrow \infty]{p} X$. ■

Théorème 1.5.7 (Convergence en m.r. implique convergence en probabilité)

$X_n \xrightarrow[n \rightarrow \infty]{r} X$ implique $X_n \xrightarrow[n \rightarrow \infty]{p} X$.

Preuve

En appliquant l'inégalité de Markov à la v.a. $Y = |X_k - X|$, on obtient pour tout $\varepsilon > 0$ et pour tout $r > 0$,

$$P(|X_k - X| \geq \varepsilon) \leq \frac{1}{\varepsilon^r} E|X_k - X|^r.$$

Si $X_n \xrightarrow[n \rightarrow \infty]{r} X$ alors le membre droit de la dernière inégalité tend vers 0 quand $n \rightarrow \infty$, montrant le résultat. ■

Exercice 1.5.1

Montrer que si $X_n \xrightarrow{p} \mu$ (μ étant une constante réelle) et $\lim_{n \rightarrow \infty} \text{var}(X_n) = 0$ alors $X_n \xrightarrow{m,q} \mu$.

Théorème 1.5.8 (Convergence en probabilité implique convergence en loi)

$X_n \xrightarrow[n \rightarrow \infty]{p} X$ implique $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$.

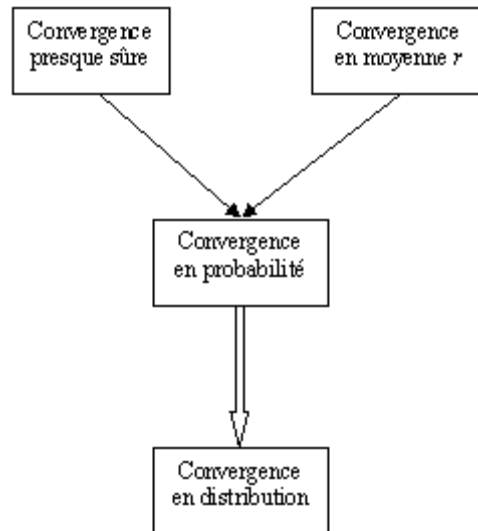
La réciproque n'est pas vraie à l'exception du cas où X est une v.a. certaine.

Exercice 1.5.2

Montrer que si $\underline{X}_n \xrightarrow{\mathcal{L}} a$ (a étant un vecteur constant) alors $\underline{X}_n \xrightarrow{P} a$.

Remarque 1.5.3

Il n'existe pas de relation entre la convergence presque sûre et la convergence en moyenne r , sauf si l'on considère d'autres hypothèses sur la suite en question.



4- Loi des grands nombres

Il existe autant de versions de lois des grands nombres qu'il y ait de modes de convergence et de structure de dépendance de suite de v.a. Dans le cas de v.a. *i.i.d.*, il existe deux lois des grands nombres : forte et faible.

Théorème 1.5.9 (Loi faible des grands nombres)

Soit $(X_n, n \in \mathbb{N})$ une suite de v.a. indépendantes et identiquement distribuées (*i.i.d.*) définies sur (Ω, \mathcal{F}, P) à valeurs dans \mathbb{R} . On suppose que $\mu := E(X_n) < \infty$. Alors, la suite de sommes partielles $(S_n, n \in \mathbb{N})$ définie par $S_n = X_1 + \dots + X_n$ converge en probabilité vers μ . Autrement dit,

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{p} \mu.$$

Preuve

Dans le cas de v.a. de variance finie, simple application de l'inégalité de Tchebychev, démontre le résultat (voir exemple 5.1.2). Dans le cas de v.a. de variance infinie, la démonstration est plus délicate et se base sur la preuve du théorème fort des grands nombres qu'on énonce sans démonstration.

Théorème 1.5.10 (Loi forte des grands nombres)

Soit $(X_n, n \in \mathbb{N})$ une suite de v.a. indépendantes et identiquement distribuées (i.i.d.) définies sur (Ω, \mathcal{F}, P) à valeurs dans \mathbb{R} . On suppose que $\mu := E(X_n) < \infty$. Alors, pour la suite des sommes partielles pondérées $(S_n, n \in \mathbb{N})$ définie par $S_n = \frac{X_1 + \dots + X_n}{n}$, converge presque sûrement vers μ . Autrement dit,

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow[n \rightarrow \infty]{p.s.} \mu.$$

5- Théorème central limite

Comme on a dit plus haut ce théorème stipule sous certaines conditions que la somme de v.a. se distribue asymptotiquement suivant une loi normale, justifiant ainsi l'utilisation de la loi normale comme lois d'erreurs de modèles statistiques. Il existe en fait plusieurs versions du théorème central limite en fonction de la structure de dépendance de la suite de v.a. (ici on ne considère que les suites indépendantes) mais aussi de la distribution de ses termes. On considère d'abord le cas le plus simple de suite de v.a. i.i.d.

5-1- Cas i.i.d. Théorème 1.5.11 (Théorème Central Limite pour i.i.d.)

Soit $(X_n, n \in \mathbb{N})$ une suite de v.a. indépendantes et identiquement distribuées (i.i.d.) définies sur (Ω, \mathcal{F}, P) à valeurs dans \mathbb{R} . On suppose que $\mu := E(X_n) < \infty$ et que $\sigma^2 := E(X_n^2) < \infty$. Alors,

$$\frac{X_1 + \dots + X_n - n\mu}{\sqrt{n}\sigma} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, 1).$$

Preuve

Une des preuves les plus simples du théorème se base sur le théorème de Levy (Théorème 1.5.5). On sait que la fonction caractéristique d'une v.a. Y centrée et réduite admet le développement limité suivant

$$\varphi_Y(t) = 1 - \frac{t^2}{2} + o(t^2)$$

au voisinage de 0. Or la fonction caractéristique de la v.a. $S_n = \frac{X_1 + \dots + X_n - n\mu}{\sqrt{n}\sigma} :=$

$\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma}$ est donnée, compte tenu de la propriété i.i.d., par

$$\begin{aligned} \varphi_{S_n}(t) &= \varphi_Y(t)^n \\ &= \left(1 - \frac{t^2}{2} + o(t^2)\right)^n \\ &\xrightarrow[n \rightarrow \infty]{} \exp\left(-\frac{t^2}{2}\right). \end{aligned}$$

Cette limite n'est autre que la fonction caractéristique de la loi normale centrée réduite. Application du théorème de Levy (Théorème 1.5.5) démontre complètement le théorème. Une extension du théorème précédent au cas où les v.a. ne sont pas nécessairement identiquement distribuées est donnée ci-après.

5-2- Cas de v.a. indépendantes Si les termes de la suite ne sont pas identiquement distribués mais toujours indépendants, les termes de la somme S_n peuvent ne pas jouer le même rôle et la convergence en loi si elle en a lieu peut ne pas être vers la loi normale. Pour éviter que la somme ne converge vers une loi normale des conditions ont été données (d'abord par Lyapunov, puis par Lindeberg, Levy et Feller) assurant la "petitesse uniforme" des termes centrés (par leurs moyenne respectives) et réduits (par la somme des variances). Formellement soit $(X_n, n \in \mathbb{N})$ une suite de v.a. indépendantes définies sur (Ω, \mathcal{F}, P) à valeurs dans \mathbb{R} . On suppose que $\mu_n := E(X_n) < \infty$ et que $\sigma_n^2 := E(X_n^2) < \infty$ et on pose $s_n^2 = \sum_{k=1}^n \sigma_k^2$. Une condition de petitesse uniforme peut être exprimée par le fait que les variances des termes sont petites par rapport à la somme des variances, i.e.

$$\lim_{n \rightarrow \infty} \max_{1 \leq k \leq n} \frac{\sigma_k}{s_n} = 0.$$

Malheureusement cette condition n'est pas suffisante à la réalisation du théorème central limite. Une première condition assurant la "petitesse uniforme" des termes sommés est la condition de Lyapunov

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^n E(|X_k - \mu_k|^{2+\delta}) = 0,$$

pour un certain $\delta > 0$ (dans certains ouvrages on énonce le théorème pour $\delta = 1$ mais ça semble restrictif). Sous cette condition Lyapunov démontre le résultat suivant :

Théorème 1.5.12 (Théorème central limite de Lyapunov)

S'il existe $\delta > 0$ tel que

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^n E(|X_k - \mu_k|^{2+\delta}) = 0,$$

alors

$$\frac{(X_1 - \mu_1) + \dots + (X_n - \mu_n)}{s_n} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, 1).$$

Avec les mêmes définitions et les mêmes notations que précédemment, Lindeberg a remplacé la condition de Lyapunov par une condition plus faible :

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{k=1}^n \int_{|x - \mu_k| > \varepsilon s_n} (x - \mu_k)^2 dF_k(x) = 0.$$

ou encore

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{k=1}^n E((X_k - \mu_k) 1_{|X_k - \mu_k| > \varepsilon s_n}) = 0.$$

où F_k est la fonction de répartition de X_k .

Sous cette condition, on a le théorème suivant.

Théorème 1.5.13 (Théorème central limite de Lindeberg)

Si

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{k=1}^n E \left((X_k - \mu_k) 1_{|X_k - \mu_k| > \varepsilon s_n} \right) = 0,$$

alors

$$\frac{(X_1 - \mu_1) + \dots + (X_n - \mu_n)}{s_n} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} N(0, 1).$$

On vérifie que la condition de Lindeberg garrantie la condition de petitesse uniforme

$$\lim_{n \rightarrow \infty} \max_{1 \leq k \leq n} \frac{\sigma_k}{s_n} = 0.$$

Problème (Méthode des moindres carrés pour un modèle de régression)

Soit X , Y et ε trois variables aléatoires définies sur un même espace de probabilité et liées par la relation linéaire suivante (dite modèle de population)

$$Y = \alpha_0 + \beta_0 X + \varepsilon, \quad (1)$$

où $\theta_0 = (\alpha_0, \beta_0)' \in \mathbb{R}^2$ est la vraie valeur du vecteur des paramètres du modèle (1) et ε est telle que :

$H1 : E(\varepsilon) = 0$ (ε est centrée),

$H2 : E(\varepsilon X) = 0$ (ε et X sont non corrélées),

$H3 : Var(\varepsilon) = \sigma^2$.

I) On suppose que la valeur de $\theta_0 = (\alpha_0, \beta_0)'$ est inconnue et qu'on désire en trouver une approximation $\theta = (\alpha, \beta)'$, de sorte que l'on ait la famille de modèles

$$Y = \alpha + \beta X + \varepsilon(\theta), \quad (2)$$

paramétrée par θ , où $\varepsilon(\theta)$ est dite la fonction erreur.

1) **a-** Quelle relation existe-t-il entre $\varepsilon(\theta)$ et ε . La fonction erreur $\varepsilon(\theta)$ vérifie-t-elle les hypothèses $H1$ - $H3$?

b- Sous quelle condition sur α et β a-t-on : $E(\varepsilon(\theta)) = 0$.

2) Si on fixe le critère de l'erreur quadratique moyenne

$$V(\theta) = E[(Y - (\alpha + \beta X))^2],$$

comme critère de choix de la meilleure approximation θ ,

- montrer que

$$V(\theta) = E(\varepsilon(\theta)^2),$$

est minimum pour $\theta = \theta_0$.

- En déduire la valeur minimum de $V(\theta)$.

- Déterminer la meilleure approximation (au sens de l'erreur quadratique moyenne) de Y sur la base de X .

3- Interpréter géométriquement les relations (1) et (2) sous les hypothèses $H1$ - $H3$.

II) On considère toujours que $\theta_0 = (\alpha_0, \beta_0)'$ est inconnu et on suppose que l'on désire "estimer" la valeur de θ_0 sur la base d'une observation, au vu du critère de l'erreur quadratique moyenne. Soit (X_i) , (Y_i) et (ε_i) , $i = 1, n$ trois suites de n variables aléatoires indépendantes de même loi (*i.i.d.*) que X , Y et ε respectivement, telles que

$$Y_i = \alpha_0 + \beta_0 X_i + \varepsilon_i, \quad i = 1, \dots, n. \quad (3)$$

La relation (3) est dite modèle de l'échantillon pour lequel on peut considérer la famille de modèles

$$Y_i = \alpha + \beta X_i + \varepsilon_i(\theta), \quad i = 1, \dots, n. \quad (4)$$

1- Expliquer pourquoi le critère de l'erreur quadratique moyenne est "superflu" pour estimer, sur la base d'un échantillon (Y_i, X_i) , $i = 1, \dots, n$, la valeur inconnue de $\theta_0 = (\alpha_0, \beta_0)'$.

2- Soit

$$V_n(\theta) = \frac{1}{n} \sum_{i=1}^n (Y_i - (\alpha + \beta X_i))^2,$$

la fonction critère de l'erreur quadratique moyenne empirique (de l'échantillon).

- Montrer que

$$V_n(\theta) \xrightarrow[n \rightarrow \infty]{p} V(\theta).$$

- Qu'en concluez-vous ?

3- Proposer une méthode pour l'estimation de α_0 et β_0 sur la base d'une observation (y_i, x_i) de (Y_i, X_i) , $i = 1, \dots, n$.

- Trouver le meilleur estimateur (des moindres carrés) $\hat{\theta}_n$ de θ_0 sur la base de (Y_i, X_i) , $i = 1, \dots, n$.

- Montrer que $\hat{\theta}_n$ converge presque sûrement vers θ_0 .

- Trouver la distribution asymptotique de $\sqrt{n}(\hat{\theta}_n - \theta_0)$ puis en déduire une distribution approximative de $\hat{\theta}_n$.

Chapitre 2 Généralités sur les processus aléatoires

I- Introduction

1- De la variable aléatoire au processus aléatoire

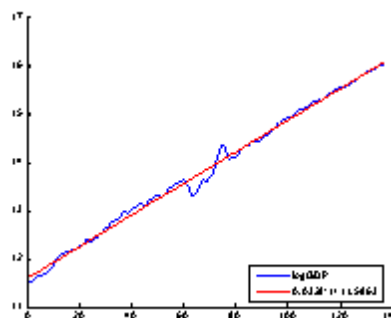
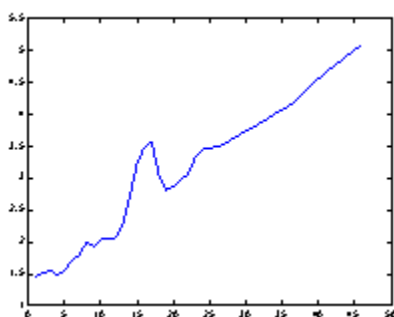
On a vu plus haut que la théorie des probabilités a pour but ultime la représentation mathématique des phénomènes aléatoires. En fait, on peut classer les phénomènes aléatoires en deux catégories : les phénomènes statiques et les phénomènes évolutifs (ou dynamiques). Dans le premier cas, il s'agit d'un fait qu'on admet fixe qui ne se rattache ou ne dépend d'aucun autre fait, dusse le même fait à un autre instant ou à une autre région. L'observation de la taille d'un individu choisi au hasard, de certaines caractéristiques d'une espèce biologique, ou de la défectuosité ou non d'une pièce sont des exemples de phénomènes ou expériences qui n'évoluent pas. Le modèle mathématique de phénomènes aléatoires statiques est la variable (ou vecteur) aléatoire. Lorsqu'on veut inférer des informations concernant la distribution de probabilité des valeurs possibles d'un phénomène aléatoire statique, on peut répéter des observations identiques et indépendantes, comme par exemple, choisir plusieurs individus, plusieurs espèces ou plusieurs pièces au hasard et mesurer leurs caractéristiques. Ce faisant, on obtient un ensemble fini de variables aléatoires indépendantes et identiquement distribuées. Cependant, si l'on désire appréhender certains phénomènes comme le prix du pétrole à un instant donné, la taille d'une population à une génération ou la consommation d'électricité à une région donnée, on peut encore obtenir un ensemble d'observations prises chaque jour, chaque génération, ou chaque région; mais à ce moment là les hypothèses d'indépendance et d'équidistributivité ne sont plus admissibles. Il est plus judicieux, ici, de considérer que les observations prises à des instants ou à des générations ou encore à des régions différentes font partie d'une même famille infinie représentant tout le phénomène à tout instant, à chaque génération ou à chaque région. Le modèle de variable aléatoire s'avère ainsi insuffisant pour décrire les phénomènes qui évoluent par rapport à un ensemble renfermant une infinité de membres, parfois non dénombrable. Ainsi, on est amené à considérer plutôt une famille de variables aléatoires, évoluant sur un domaine, et dont chacune est destinée à décrire un moment correspondant de l'évolution du phénomène. Un tel modèle est appelé processus aléatoire ou stochastique et est donc caractérisé par une *structure de dépendance* entre membres de la famille et une *distribution de probabilité des valeurs*, en raison de l'aléa qui le caractérise. Cette structure de dépendance stochastique est la source principale d'information grâce à laquelle on peut représenter le phénomène sous-jacent et exploiter cette représentation en vue de répondre à une problématique au préalable. La théorie des processus aléatoires a pour but de donner les règles de base pour le calcul des probabilités d'événements se rattachant plutôt à des processus aléatoires. C'en est en fait un complément au calcul de probabilités appliqué plutôt aux processus aléatoires qu'aux variables aléatoires. La statistique des processus, quant à elle, permet sur la base d'une observation d'une partie du phénomène d'induire des informations concernant la distribution de probabilité d'un ou de plusieurs membres du processus ayant généré cette observation. Ce cours met l'accent tant sur l'aspect probabiliste des processus aléatoires que sur l'étude statistique.

2- Quelques exemples de phénomènes évolutifs

On rencontre assez souvent une multitude de phénomènes évolutifs dans des domaines scientifiques assez variés : la physique, l'électronique, l'environnement, la biologie, l'économie, les finances, les sciences de gestion... La théorie des processus aléatoires permet d'en résoudre certaines problématiques attachées. La liste suivante est bien entendu non exhaustive.

a- Evolution de grandeurs économiques

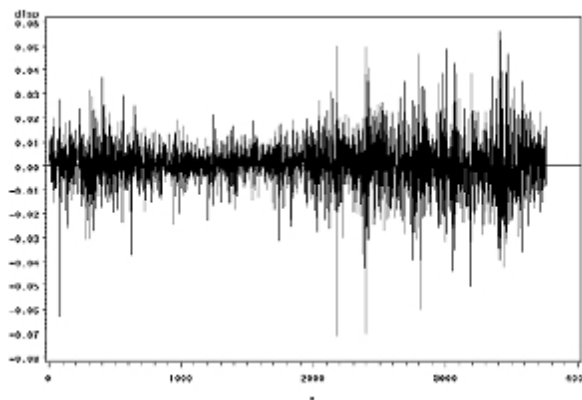
En économie on rencontre plusieurs grandeurs aléatoires qui évoluent dans le temps ou dans des régions de l'espace. Par exemple, évolution du volume (journaliers, mensuels, annuels...) de production d'une entreprise dans le temps, évolution d'indices macro-économiques (produit intérieur brut, indice de consommation, volume d'exportation d'un pays) dans le temps, évolution du chômage dans une contrée... L'étude de tels phénomènes a pour but de : prévoir à un horizon donné l'état du phénomène, identifier des régimes de croissance (expansion, récession), trouver des relations stochastiques entre grandeurs économiques, infirmer ou confirmer des théories économiques...



Evolution du PIB à l'USA 1940-1990 Evolution du PIB à l'USA de 1870 à 2006

b- Evolution de grandeurs financières et actuarielles

En actuariat, le prix d'un produit dans une bourse, le taux de change d'une monnaie par rapport à une autre, la ruine d'une compagnie d'assurance sont des exemples de phénomènes auxquels on s'intéresse afin de prévoir des valeurs futures, prévoir des risques dû à une variation du prix, calibrer des investissements, fixer des primes d'assurances appropriées afin d'éviter la ruine tout en étant compétitif...



Evolution du prix d'un actif financier

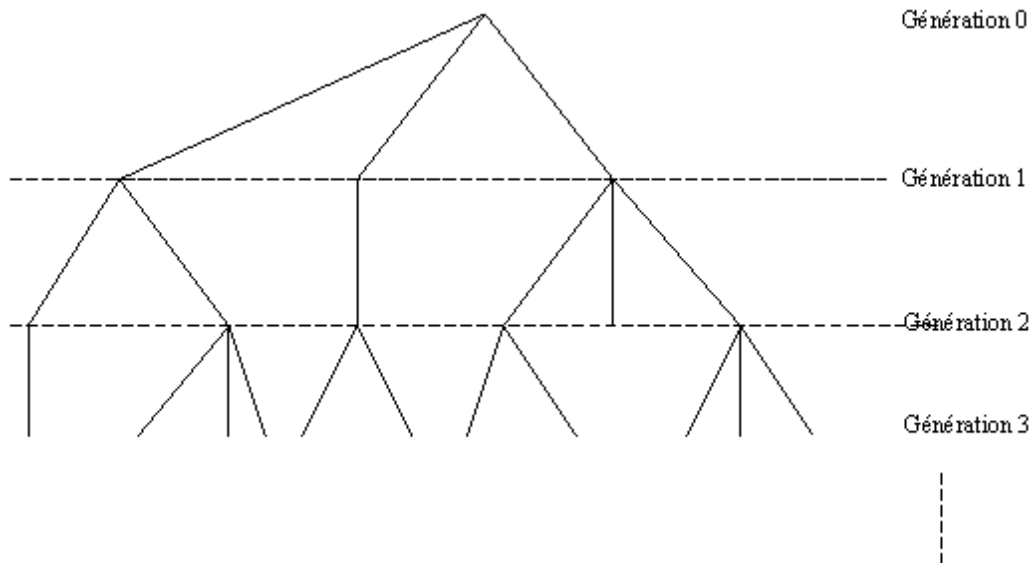
c- Systèmes et réseaux d'attente

Des clients de nature assez générale, (à savoir individus, véhicules, requêtes informatiques etc.) se présentent, selon une certaine discipline (premier venu premier servi : FIFO, premier venu dernier servi : LIFO, service aléatoire : RS...), à des serveurs (guichets, intersections, serveur de base de données,...) en des instants aléatoires afin d'y recevoir un service dont la durée est aléatoire. Le déroulement purement aléatoire des phénomènes d'arrivées et de services dans le système induit souvent des files d'attente dans lesquelles des individus attendent leurs tours pour en être servis. La connaissance de la distribution du nombre de clients dans le système à un instant donné, du nombre moyen de clients dans une file, du temps moyen de séjour dans le système sont d'importantes questions dans la théorie des files d'attente dont l'étude ayant pour but le redimensionnement et l'optimisation des performances du système.



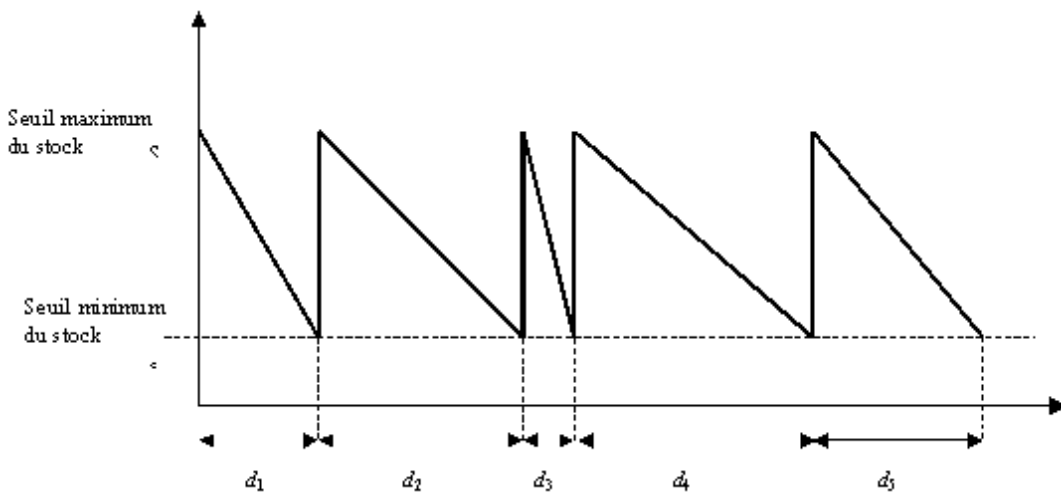
d- Phénomènes de branchement

L'étude de l'évolution des populations est un domaine de prédilection de la biologie, de la démographie, de la sociologie pour ne citer que ceux là. Les individus d'une population (de bactéries, d'animaux, d'êtres humains...) peuvent donner lieu, à une certaine génération, à des descendants en nombres aléatoires (ou non) et dont des caractéristiques (taille, volume, comportement...) peuvent en être similaires. Les descendants des individus d'une $n^{\text{ème}}$ génération constituent la $(n + 1)^{\text{ème}}$ génération. On peut s'intéresser à la taille moyenne de la population à une certaine génération, à la probabilité d'extinction ou d'explosion d'une espèce, au temps moyen d'extinction, à une loi d'évolution d'une caractéristique (taille, QI, aptitude au suicide,...) à travers les générations...



e- Phénomènes de stocks

Une quantité de matière première est stockée afin de satisfaire un certain besoin d'une entreprise de production. Deux valeurs critiques s et S étant fixées ($s < S$), et dès que le niveau de stock est au seuil minimum s , une commande est faite. Généralement l'arrivée de la commande se fait après une certaine durée (aléatoire ou non) et le remplissage est effectué immédiatement (ou après un délais). Dans ce genre de problèmes on s'intéresse aux durées de temps successives pendant lesquelles il n'y a pas eu de pénurie, le nombre moyen de remplissages futurs, le temps nécessaire pour un prochain remplissage... L'étude d'un tel phénomène permet d'appréhender les pénuries, d'organiser les commandes et de façon générale d'opter pour une meilleure politique de gestion des stocks.



f- Les mouvements aléatoires

La physique offre de nombreux phénomènes représentant des mouvements aléatoires de mobiles : mouvement brownien d'une molécule dans un milieu, mobilité d'un groupe de connectés à un réseau téléphonique... L'étude des mouvements permet de prévoir des positions

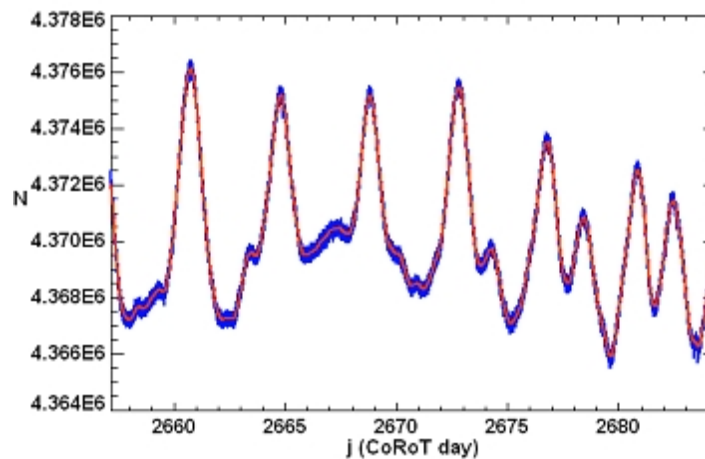
moyennes, des taux d'occupation d'une région,...



Simulation d'un mouvement Brownien sur une droite

g- Evolution de signaux

Un signal est souvent entaché de bruit indésirable que l'on veut éliminer : sons bruités, images floues, mesures imprécises... Un signal évolue (dans le temps ou dans un espace) et l'étude de son évolution sur une tranche permet de prévoir des valeurs possibles de son évolution future (prédiction), d'éliminer le bruit et d'extraire le signal principal...

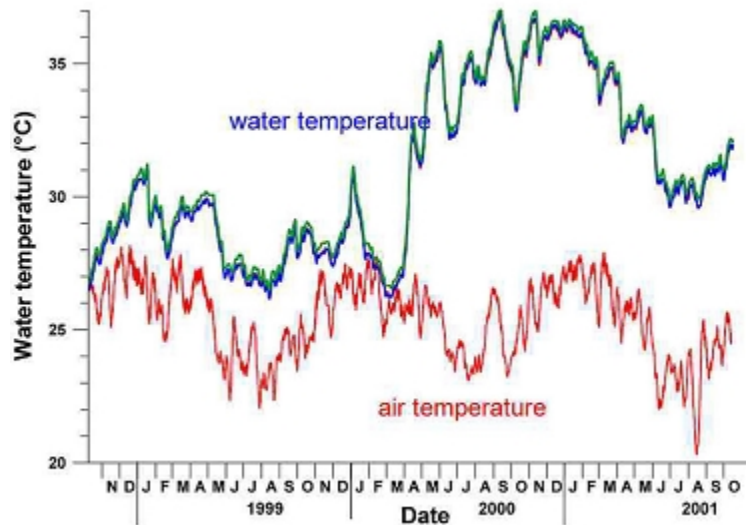


Exemple de filtrage. Série temporelle du flux de l'étoile HD1 75726 observée par le photomètre CoR

h- Evolution de grandeurs environnementales

L'évolution des précipitations dans une région, les crues de rivières, la température d'un milieu, l'intensité de secousse dans une région sont des exemples de phénomènes, environnementaux incluant : la météorologie, l'hydrologie, la géophysique... L'étude de l'évolution de ces phénomènes permet de prévoir leurs comportements futurs sur la base de la connais-

sance de leurs évolutions passées.



3- Caractéristiques des phénomènes évolutifs et problématiques associées

Les exemples précédents révèlent trois éléments importants caractérisant les phénomènes aléatoires évolutifs, à savoir :

- i) Un domaine d'évolution par rapport auquel se dynamise le phénomène. Ce domaine peut être dénombrable (domaine d'évolution d'une consommation journalière, domaines d'évolution des générations d'une population) ou non dénombrable (domaine d'évolution de la température d'un milieu, domaine d'évolution du nombre de clients dans le système à un instant...). Il peut représenter concrètement l'axe des temps ou une région de l'espace, et il peut même désigner une autre notion abstraite telle que l'ensemble de générations d'une population, position d'un pixel dans une image, etc. Généralement, il consiste en un ensemble de scalaires : \mathbb{N} , \mathbb{Z} , \mathbb{R} mais il peut être un ensemble quelconque pour autant qu'on puisse en définir une relation d'ordre total comme par exemple les treillis de \mathbb{Z}^2 .
- ii) Un ensemble des états possibles du phénomène à tout moment de l'évolution. Cet ensemble dépend implicitement du domaine d'évolution en ce sens qu'il représente la réunion d'ensembles renfermant chacun les états possibles à un moment donné de l'évolution. L'ensemble des états peut être dénombrable (nombre de clients arrivant à un guichet, taille d'une population à une génération donnée,...) ou non dénombrable (température d'un milieu, prix d'un actif financier, niveau de stock à un instant donné...).
- iii) Un caractère aléatoire des valeurs possibles du phénomène. Si l'on a effectivement observé un phénomène sur une tranche du domaine d'évolution, le phénomène aurait pu prendre d'autres réalisations sur tout le domaine d'évolution. Ces réalisations possibles peuvent donc être caractérisées par une loi de probabilité qui varie dans le domaine de l'évolution exprimant une structure de dépendance stochastique bien spécifique.

Le but de l'étude d'un phénomène aléatoire évolutif est dès lors de déceler une structure de dépendance (stochastique) appropriée relativement à son évolution, et à travers laquelle on peut résoudre des problèmes associés tels que :

- la prévision (extrapolation): prévoir une valeur future sur la base d'une information de l'évolution passée.

- le lissage (interpolation) : prédire, sur la base d'une information sur une tranche d'évolution du phénomène, une valeur possible à un moment dans le domaine de délimitation de la tranche observée.

- le filtrage (extraction du signal) : certaines grandeurs sont mesurées avec incertitude (l'écoute d'une chanson entachée de bruits de fond, mesure de la température d'un milieu lorsque le vent influence l'appareil de mesure,...) et sur la base d'une observation sur une tranche d'évolution on désire extraire une approximation de la vraie valeur de la grande sur une tranche d'évolution.

Ces trois objectifs peuvent être associés comme sous-produit à une problématique plus générale, comme par exemple élaborer un programme de gestion de production, de gestion de portefeuille, où l'on suppose que les paramètres sont sujets à l'aléa, ou gérer un système de files d'attente où plusieurs phénomènes aléatoires rentrent en interaction.

Dans la section suivante on présentera rigoureusement le concept de processus aléatoire.

II- Définitions de base

1- Espace de probabilité associé à un phénomène évolutif

Pour représenter mathématiquement un phénomène aléatoire évolutif donné, la première étape consiste à lui associer un espace de probabilités (Ω, \mathcal{F}, P) où Ω désigne l'ensemble des ses réalisations possibles, \mathcal{F} est une tribu de parties de Ω représentant les événements que l'on puisse formuler, a priori, à propos des réalisations possibles de ce phénomène et P est une mesure de probabilité sur \mathcal{F} , mesurant les chances de réalisation des événements de \mathcal{F} associés.

L'espace fondamental Ω est supposé dépendre implicitement d'un ensemble T représentant le domaine d'évolution du phénomène. A chaque moment $t \in T$ de l'évolution on peut associer un ensemble Ω_t représentant les résultats possibles du phénomène à l'instant t . Alors sur tout le domaine T , l'ensemble des réalisations possibles Ω sera le produit cartésien des ensembles Ω_t , $t \in T$, i.e. $\Omega = \prod_{t \in T} \Omega_t$ où \prod désigne le produit cartésien de tous les membres de T .

Exemple 2.2.1

i) Si on observe une expérience d'un jet indéfiniment répété d'une pièce de monnaie et l'on s'intéresse à la suite des résultats possibles successifs affichés sur la face supérieure, alors le domaine d'évolution peut être représenté par $T = \mathbb{N}$. Pour tout $t \in \mathbb{N}$, les résultats possibles de l'expérience au $t^{\text{ème}}$ lancé sont $\Omega_t = \{\text{pile}, \text{face}\}$. L'espace fondamental du phénomène global est alors

$$\begin{aligned} \Omega &= \prod_{t \in \mathbb{N}} \{\text{pile}, \text{face}\} \\ &= \{\text{pile}, \text{face}\}^{\mathbb{N}} \\ &= \{\omega = (\omega_0, \omega_1, \omega_2, \dots), \omega_i \in \{\text{pile}, \text{face}\}, i \in \mathbb{N}\}. \end{aligned}$$

ii) Pour le même exemple, si on s'intéresse au nombre de piles obtenus jusqu'à une répétition

$t \in T = \mathbb{N}$, alors, $\Omega_t = \{0, \dots, t\}$ et

$$\begin{aligned}\Omega &= \prod_{t \in \mathbb{N}} \{0, \dots, t\} \\ &= \{\omega = (\omega_0, \omega_1, \omega_2, \dots), \omega_i \in \{0, \dots, i\}, i \in \mathbb{N}\} \\ &= \mathbb{N}^{\mathbb{N}}.\end{aligned}$$

La tribu \mathcal{F} , quant elle, dépend également implicitement du domaine d'évolution T . A tout moment $t \in T$ de l'évolution on peut associer également une tribu élémentaire \mathcal{F}_t représentant les évènements liés au phénomène à l'instant t . Sur tout le domaine T , la tribu \mathcal{F} des évènements liés au phénomène est le produit

$$\mathcal{F} = \bigotimes_{t \in T} \mathcal{F}_t$$

qui est la plus petite tribu générée par les ensembles $\prod_{t \in T} A_t$ avec $A_t \in \mathcal{F}_t$ et pour éviter des situations de dégénérescence, cette tribu est telle qu'elle doit renfermer les évènements élémentaires associés à tout moment de l'évolution (e.g. $\{\omega_t\} \in \mathcal{F}$, pour tout t). De tels évènements sont dits projections des évènements élémentaires $\{\omega = (\omega_t, t \in T)\}$. Puisque Ω représente un produit d'espaces élémentaires Ω_t , les évènements élémentaires projections peuvent s'exprimer à travers des ensembles cylindriques que la tribu \mathcal{F} doit contenir. La représentation d'expériences finiment répétées vue à la section 1.2.4 en est un cas particulier. Par exemple, dans le cas par exemple où $T = \mathbb{N}$, les ensembles cylindriques sont de la forme

$$\prod_{k=0}^{j-1} \Omega_k \times A_j \times \prod_{k=j+1}^{\infty} \Omega_k, \quad A_j \in \mathcal{F}_j, \quad j \in \mathbb{N},$$

L'espace probabilisable (Ω, \mathcal{F}) est donc le produit d'espaces élémentaires $(\Omega_t, \mathcal{F}_t)$

$$(\Omega, \mathcal{F}) = \prod_{t \in T} (\Omega_t, \mathcal{F}_t).$$

La mesure de probabilité P définie sur \mathcal{F} dépend également de T . Pour tout t , soit P_t une mesure de probabilité sur l'espace probabilisable élémentaire $(\Omega_t, \mathcal{F}_t)$. Alors sur tout le domaine d'évolution T , la mesure de probabilité P est fonction des $(P_t, t \in T)$. Dans le cas où le phénomène est soumis à l'indépendance mutuelle, alors on a

$$P = \prod_{t \in T} P_t.$$

L'espace de probabilité (Ω, \mathcal{F}, P) est donc le produit d'espaces élémentaires $(\Omega_t, \mathcal{F}_t, P_t)$

$$(\Omega, \mathcal{F}, P) = \prod_{t \in T} (\Omega_t, \mathcal{F}_t, P_t).$$

2- Définitions d'un processus aléatoire

L'espace de probabilité (Ω, \mathcal{F}, P) étant complètement spécifié, on peut en associer une application qui permet de numériser l'ensemble Ω des réalisations possibles, lequel peut être quelconque. Par numériser on entend associer à chaque réalisation du phénomène, à tout moment $t \in T$ de l'évolution, un nombre ou un vecteur de nombres réels. Lorsque Ω est d'emblée numérique dans le sens où les composantes de $\omega \in \Omega$ sont des nombres ou des vecteurs de réels, on peut prendre pour application l'identité et l'espace (Ω, \mathcal{F}, P) est alors dit canonique. Une telle application numérisant Ω , dite processus aléatoire, et qui doit vérifier certaines condition de probabilisabilité (mesurabilité), peut être définie de plusieurs manières. On réservera ici trois définitions dont chacune se base sur un angle de vue différent. La première, vision verticale, regarde un processus du point de vue de ses membres, c'est à dire de variables aléatoires le constituant. La deuxième définition, d'une vision horizontale, définit un processus par rapport à ses réalisations. Enfin, la troisième selon une vision croisée, définit le processus du point de vu du croisement d'un membre du processus et d'une réalisation possible.

Définition 2.2.1 (Processus aléatoire, vision verticale)

Un processus aléatoire (p.a.) de domaine d'évolution T , défini sur un espace de probabilité (Ω, \mathcal{F}, P) et à valeurs dans un espace d'état E qui est muni d'une tribu \mathcal{E} , est une famille de variables aléatoires $(X_t, t \in T)$ chacune définie sur (Ω, \mathcal{F}, P) à valeur dans (E, \mathcal{E}) .

Autrement dit, $(X_t, t \in T)$ est une application qui associe pour tout t dans T une variable aléatoire X_t dans $\mathcal{D}((\Omega, \mathcal{F}), (E, \mathcal{E}))$, ensemble des variables aléatoires \mathcal{F}/\mathcal{E} -probabilisables, définies de Ω à valeurs dans E . Autrement dit,

$$(X_t, t \in T) : T \longrightarrow \mathcal{D}((\Omega, \mathcal{F}), (E, \mathcal{E})) \\ t \longmapsto X_t.$$

Remarque 2.2.1

- i) Cette définition, d'essence récursive, ne se préoccupe pas a priori de la condition de mesurabilité qui est déjà garantie par le fait d'avoir défini un processus en tant que famille de variables aléatoires, donc d'applications déjà mesurables.
- ii) Les variables aléatoires $X_t, t \in T$ peuvent être discrètes ou continues, réelles, complexes ou vectorielles, finies ou infinies.
- iii) Le domaine d'évolution T peut être fini, infini dénombrable, ou non dénombrable.
- iv) Lorsque $E \subset \mathbb{R}$, on peut associer à \mathbb{R} sa tribu borélienne.
- v) Dans la Définition 2.2.1, on peut permettre le cas où pour tout élément t de l'évolution, l'ensemble des états associé à la variable X_t correspondante dépend de t , soit E_t . Dans ce cas, la Définition 2.2.1 devient :

Un processus aléatoire $(X_t, t \in T)$ défini sur un espace de probabilité (Ω, \mathcal{F}, P) et à valeurs dans un espace d'état E muni d'une tribu \mathcal{E} est une application qui associe pour tout t dans T une variable aléatoire X_t dans $\mathcal{D}((\Omega, \mathcal{F}), (E_t, \mathcal{E}_t))$, ensemble des variables aléatoires $\mathcal{F}/\mathcal{E}_t$ -probabilisables, définies de Ω à valeurs dans un espace E_t muni de la tribu \mathcal{E}_t avec

$$E = \bigcup_{t \in T} E_t.$$

$$(X_t, t \in T) : T \longrightarrow \mathcal{D}((\Omega, \mathcal{F}), (E_t, \mathcal{E}_t)) \\ t \longmapsto X_t.$$

La définition suivante définit un processus en terme de ses réalisations. Soit $\mathcal{E}^{\otimes T}$ la tribu engendrée par les parties de $E^T = \prod_{t \in T} E$, et contenant les ensembles cylindriques (voir

Définition 1.2.10).

Définition 2.2.2 (Processus aléatoire, vision horizontale)

Un processus aléatoire (p.a.), $\mathbb{X} = (X_t, t \in T)$ de domaine d'évolution T , défini sur un espace de probabilité (Ω, \mathcal{F}, P) et à valeurs dans un espace d'état (E, \mathcal{E}) , est une application mesurable de (Ω, \mathcal{F}) dans $(E^T, \mathcal{E}^{\otimes T})$.

Autrement dit, c'est une application qui pour toute réalisation possible ω du phénomène, associe une famille de nombres $\mathbb{X}(\omega) = (X_t(\omega), t \in T)$, et est telle que l'image réciproque de tout Borélien \mathbb{B} de $\mathcal{E}^{\otimes T}$ de la forme $\mathbb{B} = \prod_{t \in T} B_t$ (où $B_t \in \mathcal{E}, t \in T$), est un membre de \mathcal{F} ;

i.e.

$$\begin{aligned} \mathbb{X} : \Omega &\longrightarrow E^T \\ \omega &\longmapsto \mathbb{X}(\omega) = (X_t(\omega), t \in T) \\ \text{où } \forall \mathbb{B} = \prod_{t \in T} B_t \in \mathcal{E}^{\otimes T}, &\left\{ \omega \in \Omega : \bigcap_{t \in T} X_t(\omega) \in B_t \right\} \in \mathcal{F}, \end{aligned}$$

ou encore

$$\forall \mathbb{B} \in \mathcal{E}^{\otimes T} : \mathbb{X}^{-1}(\mathbb{B}) \in \mathcal{F}.$$

■

Remarque 2.2.2

- i) Pour tout $\omega \in \Omega$, $\mathbb{X}(\omega) = (X_t(\omega), t \in T)$ est dit réalisation ou trajectoire du processus.
- ii) Dans le cas où $T = \mathbb{N}$ (ou de manière générale T est dénombrable) et $E \subset \mathbb{R}$ (E est un ensemble de scalaires), le processus $\mathbb{X} = (X_t, t \in T)$ se réduit à une suite de v.a. dans le sens de la Définition 1.5.1. Ainsi, pour tout $\omega \in \Omega$, la réalisation $\mathbb{X}(\omega) = (X_t(\omega), t \in T)$ définit une suite de nombres réels et c'est la raison pour laquelle, un tel processus est souvent appelé *suite aléatoire*.
- iii) Lorsque T est non dénombrable, le processus $\mathbb{X} = (X_t, t \in T)$ génère pour tout $\omega \in \Omega$ une famille de nombre $(X_t(\omega), t \in T)$ ou précisément une fonction $\mathbb{X}(\omega)$ définie sur T à valeurs dans E , telle que

$$\begin{aligned} \mathbb{X}(\omega) &: T \rightarrow E \\ t &\mapsto X_t(\omega). \end{aligned}$$

C'est pourquoi, un processus dans ce cas est souvent appelé *fonction aléatoire* et est souvent désigné par $(X(t), t \in T)$. ■

Une troisième définition mais qui n'est pas tout à fait équivalente aux deux premières dans le sens où elle exige des conditions supplémentaires, regarde un processus du point de vue et de ses membres et de ses réalisations.

Définition 2.2.3 (Processus aléatoire, vision croisée)

Un processus aléatoire (p.a.), $\mathbb{X} = (X_t, t \in T)$ de domaine d'évolution T , défini sur un espace de probabilité (Ω, \mathcal{F}, P) et à valeurs dans un espace d'état (E, \mathcal{E}) , est une application mesurable de $(\Omega \times T, \mathcal{F} \otimes \mathcal{T})$ dans (E, \mathcal{E}) , où \mathcal{T} est une tribu associée à T et $\mathcal{F} \otimes \mathcal{T}$ la plus

petite tribu engendrée par les parties de $\Omega \times T$ et contenant les ensembles cylindriques. Ceci se traduit symboliquement par

$$\begin{aligned}\mathbb{X} : (\Omega \times T, \mathcal{F} \otimes \mathcal{T}) &\longrightarrow (E, \mathcal{E}) \\ (\omega, t) &\longmapsto \mathbb{X}(\omega, t) = X_t(\omega) \\ \text{où } \forall A \in \mathcal{E}, \{(\omega, t) \in \Omega \times T : X_t(\omega) \in A\} &\in \mathcal{F} \otimes \mathcal{T}.\end{aligned}$$

Remarque 2.2.3

i) On prendra pour \mathcal{T} la tribu de Borel si T est une partie non dénombrable de \mathbb{R} et $\mathcal{P}(T)$ s'il est dénombrable.

ii) Cette troisième définition exige une condition de mesurabilité conjointe à l'ensemble des réalisations possibles et au domaine d'évolution, ce qui induit une difficulté supplémentaire d'un point de vue théorique. En effet, on peut définir un processus aléatoire via l'une des deux premières définitions (Définition 2.2.1 ou Définition 2.2.2) sans que la condition de mesurabilité

$$\forall A \in \mathcal{E}, \{(\omega, t) \in \Omega \times T : X_t(\omega) \in A\} \in \mathcal{F} \otimes \mathcal{T},$$

qui peut être restrictive ne soit vérifiée. Cette condition s'avère ainsi une propriété d'un processus aléatoire (si l'on se restreint à l'une des deux premières définitions) plutôt qu'un élément de sa définition. Un processus vérifiant une telle propriété est dit *mesurable*.

iii) Si un processus $\mathbb{X} = (X_t, t \in T)$ est mesurable, alors sa réalisation $\mathbb{X}(\omega) = (X_t(\omega), t \in T)$ pour ω fixé est une fonction mesurable de (T, \mathcal{T}) dans (E, \mathcal{E}) .

iv) Pour un certain $t \in T$ fixé, X_t est une v.a. sur (Ω, \mathcal{F}) à valeurs dans (E, \mathcal{E}) .

v) Les processus aléatoires peuvent être classés selon la dénombrabilité ou non des espaces d'états E et du domaine d'évolution T .

Domaine d'évolution T	Espace d'état E	Dénombrable	Non dénombrable
Dénombrable		Processus à temps discret et à espace d'état discret	Processus à temps discret et à espace d'état continu
Non dénombrable		Processus à temps continu et à espace d'état discret	Processus à temps continu et à espace d'état continu

Exemple 2.2.1 (suite)

i) Reprenons l'exemple 2.2.1 du jet indéfiniment répété d'une pièce de monnaie avec $\Omega = \{\text{pile, face}\}^{\mathbb{N}} = \{\omega = (\omega_0, \omega_1, \omega_2, \dots), \omega_i \in \{\text{pile, face}\}, i \in \mathbb{N}\}$, \mathcal{F} est une tribu des parties de Ω et P une mesure de probabilité définie sur \mathcal{F} . Considérons pour tout $t \in \mathbb{N}$ la variable aléatoire X_t définie de Ω à valeurs dans $\{0, 1\}$ par

$$\begin{aligned}X_t : (\Omega, \mathcal{F}) &\longrightarrow (\{0, 1\}, \mathcal{P}(\{0, 1\})) \\ \omega = (\omega_0, \omega_1, \omega_2, \dots) &\mapsto X_t(\omega) = \begin{cases} 1 & \text{si } \omega_t = \text{pile} \\ 0 & \text{si } \omega_t = \text{face}. \end{cases}\end{aligned}$$

La famille de v.a. $(X_t, t \in \mathbb{N})$ définit un processus aléatoire défini sur (Ω, \mathcal{F}, P) à valeurs dans $E = \{0, 1\}$.

Nous avons utilisé la 1^{ère} vision pour définir le processus $(X_t, t \in \mathbb{N})$. On aurait pu le définir en utilisant la vision horizontale.

En effet, sur Ω on définit l'application \mathbb{X} à valeurs dans $\{0, 1\}^{\mathbb{N}} = \{x = (x_0, x_1, x_2, \dots), x_i \in \{0, 1\}, i \in \mathbb{N}\}$ comme suit :

$$\begin{aligned} \mathbb{X} &= (X_t, t \in \mathbb{N}) : (\Omega, \mathcal{F}) \longrightarrow \left(\{0, 1\}^{\mathbb{N}}, \mathcal{P}(\{0, 1\}^{\otimes \mathbb{N}}) \right) \\ \omega &= (\omega_0, \omega_1, \omega_2, \dots) \mapsto \mathbb{X}(\omega) = x = (x_0, x_1, x_2, \dots) \\ \text{avec } x_i &= \begin{cases} 1 & \text{si } \omega_i = \text{pile} \\ 0 & \text{si } \omega_i = \text{face} \end{cases}, i \in \mathbb{N}, \end{aligned}$$

où $\mathcal{P}(\{0, 1\}^{\otimes \mathbb{N}})$ est la plus petite tribu engendrée par des parties de $\{0, 1\}^{\mathbb{N}}$ et contenant les ensembles cylindriques de la forme $\prod_{k=0}^{j-1} \{0, 1\} \times A_j \times \prod_{k=j+1}^{\infty} \{0, 1\}$, $A_j \in \mathcal{P}(\{0, 1\})$, $j \in \mathbb{N}$.

L'application \mathbb{X} est un processus aléatoire numérisant Ω .

Exercice Définir le processus $(X_t, t \in \mathbb{N})$ selon la vision croisée. ■

3- Distribution de probabilité d'un processus aléatoire

a- Distributions fini-dimensionnelles Pour un processus aléatoire $\mathbb{X} = (X_t, t \in T)$ défini sur un espace de probabilité (Ω, \mathcal{F}, P) , à valeurs dans un espace d'état (E, \mathcal{E}) et de domaine d'évolution T dénombrable, on a vu dans la section 1.5.1 (cf, suites de variables aléatoires) que sa structure de probabilité est caractérisée par la distribution infini-dimensionnelle $P_{\mathbb{X}}(\cdot)$ définie sur $\mathcal{E}^{\otimes T}$ à valeurs dans $[0, 1]$ comme suit

$$\begin{aligned} P_{\mathbb{X}}(\cdot) : \mathcal{E}^{\otimes T} &\longrightarrow [0, 1] \\ \mathbb{B} = \prod_{t \in T} B_t &\longmapsto P_{\mathbb{X}}(\mathbb{B}) = P(\omega : \mathbb{X}(\omega) \in \mathbb{B}) \\ &= P\left(\omega : \bigcap_{t \in T} X_t(\omega) \in B_t\right) \end{aligned}$$

avec $B_t \in \mathcal{E}$. On a vu également que cette distribution infini-dimensionnelle est difficile à manipuler et qu'il existe un outil simple, la *distribution fini-dimensionnelle*, permettant de simplifier son analyse. L'introduction de cette distribution est justifié par le théorème d'extension (cf, Théorème 1.5.1) qui stipule que la distribution infini-dimensionnelle est uniquement déterminée par les probabilités fini-dimensionnelles. La structure probabiliste d'un processus à temps discret (de domaine d'évolution dénombrable) est donc entièrement déterminée par les distributions fini-dimensionnelles.

Lorsque le domaine d'évolution T du processus n'est pas dénombrable, on est à première vue tenté de définir la distribution infini-dimensionnelle analogiquement au cas dénombrable, comme suit

$$\begin{aligned} P_{\mathbb{X}}(\cdot) : \mathcal{E}^{\otimes T} &\longrightarrow [0, 1] \\ \mathbb{B} = \prod_{t \in T} B_t &\longmapsto P_{\mathbb{X}}(\mathbb{B}) = P(\omega : \mathbb{X}(\omega) \in \mathbb{B}) \\ &= P\left(\omega : \bigcap_{t \in T} X_t(\omega) \in B_t\right) = P\mathbb{X}^{-1}(\mathbb{B}) \end{aligned}$$

Cependant, en raison de la définition même de la probabilité (voir axiome iii) section 1.2.2), on ne sait calculer la probabilité des événements du type $\left\{ \omega : \bigcap_{t \in T} X_t(\omega) \in B_t \right\}$ définis en ter-

mes d'intersection non dénombrable et par conséquent la distribution infinie-dimensionnelle ne peut être définie pour des processus à temps continu. Ainsi, le seul moyen de caractériser la distribution de probabilité des processus à temps continu demeure les distributions fini-dimensionnelles dont la définition est similaire que celle des processus à temps discret. La définition suivante est valable pour tout type de processus.

Définition 2.3.1 (Distributions fini-dimensionnelles)

Pour $n \in \mathbb{N}^*$ et pour tout $t_1, t_2, \dots, t_n \in T$, la distribution fini-dimensionnelle de $\mathbb{X} = (X_t, t \in T)$ est la distribution de toute sous-suite finie $(X_{i_1}, X_{i_2}, \dots, X_{i_n})$ de \mathbb{X} . Autrement dit, c'est la fonction ensembliste $P_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(\cdot)$ définie sur $\mathcal{E}^{\otimes n}$ à valeurs dans $[0, 1]$ par

$$P_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(\cdot) : \mathcal{E}^{\otimes n} \longrightarrow [0, 1]$$

$$\tilde{B} = \prod_{j=1}^n B_{t_j} \longmapsto P_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(\tilde{B}) = P\left(\omega : \bigcap_{j=1}^n X_{t_j}(\omega) \in B_{t_j}\right).$$

où $B_{t_j} \in \mathcal{E}$.

Remarque 2.3.1

i) Contrairement au cas à temps discret, la structure probabiliste du processus n'est pas entièrement déterminée via les distributions fini-dimensionnelles, à l'exception de certains processus dit "**séparables**".

ii) Afin d'éviter la manipulation de boréliens, on peut de manière équivalente caractériser la structure probabiliste du processus au moyen de la fonction de répartition fini-dimensionnelle.

Définition 2.3.2 (Fonction de répartition fini-dimensionnelle)

Pour $n \in \mathbb{N}^*$ et pour tout $t_1, t_2, \dots, t_n \in T$, la fonction de répartition fini-dimensionnelle de $\mathbb{X} = (X_t, t \in T)$ est la fonction de répartition de toute sous-suite finie $(X_{i_1}, X_{i_2}, \dots, X_{i_n})$ de X . Autrement dit, c'est la fonction $F_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(\cdot)$ définie sur E^n à valeurs dans $[0, 1]$ par

$$F_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(\cdot) : E^n \longrightarrow [0, 1]$$

$$(x_1, x_2, \dots, x_n) \longmapsto F_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(x_1, x_2, \dots, x_n) = P\left(\omega : \bigcap_{j=1}^n X_{t_j}(\omega) \leq x_j\right).$$

A partir de la fonction de répartition on peut définir :

i) La fonction de masse fini-dimensionnelle $p_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(\cdot)$ lorsque l'espace d'état E est dénombrable, comme suit

$$p_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(\cdot) : E^n \longrightarrow [0, 1]$$

$$(x_1, x_2, \dots, x_n) \longmapsto p_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(x_1, x_2, \dots, x_n) = P\left(\omega : \bigcap_{j=1}^n X_{t_j}(\omega) = x_j\right).$$

ii) La fonction densité de masse fini-dimensionnelle $f_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(\cdot)$ lorsque les variables aléatoires $X_t, t \in T$ sont absolument continues (a.c.).

$$f_{t_1, t_2, \dots, t_n}(\cdot) : E^n \longrightarrow [0, 1]$$

$$(x_1, x_2, \dots, x_n) \longmapsto f_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(x_1, x_2, \dots, x_n) = \frac{\partial F_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \dots \partial x_n}.$$

Remarque 2.2.2

On a évidemment

$$F_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(x_1, x_2, \dots, x_n) = \begin{cases} \sum_{k_1 \leq x_1} \dots \sum_{k_n \leq x_n} p_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(k_1, k_2, \dots, k_n) & \text{si } E \text{ est discret} \\ \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} p_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(k_1, k_2, \dots, k_n) dk_1 dk_2 \dots dk_n & \text{si } X_t \text{ est a.c.} \end{cases}$$

b- Théorème d'existence de Kolmogorov Pour tout processus aléatoire $\mathbb{X} = (X_t, t \in T)$ défini sur un espace de probabilité (Ω, \mathcal{F}, P) , à valeurs dans un espace d'état (E, \mathcal{E}) , il est clair que les fonctions de répartition fini-dimensionnelles $\{F_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}, n \in \mathbb{N}^*, t_1, \dots, t_n \in T\}$ vérifient les trois propriétés suivantes :

- i) $F_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(\cdot, \dots, \cdot)$ est une fonction de répartition pour tout t_1, t_2, \dots, t_n fixés.
- ii) $F_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(x_1, x_2, \dots, x_n) = F_{X_{t_{i_1}}, X_{t_{i_2}}, \dots, X_{t_{i_n}}}(x_{i_1}, x_{i_2}, \dots, x_{i_n})$ pour toute permutation i_1, i_2, \dots, i_n de $1, 2, \dots, n$ et tout $x_i \in E$.
- iii) $F_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(x_1, x_2, \dots, x_{n-1}, \infty) = F_{X_{t_1}, X_{t_2}, \dots, X_{t_{n-1}}}(x_1, x_2, \dots, x_{n-1})$.

Cependant, le problème inverse n'est pas tout à fait évident. En effet, étant donnée une famille de fonctions $\{F_{t_1, t_2, \dots, t_n}, n \in \mathbb{N}^*, t_1, \dots, t_n \in T\}$ existe-t-il un processus aléatoire ayant comme distributions fini-dimensionnelle cette même famille de fonction. Ce problème a été résolu par Kolmogorov en 1933.

Théorème 2.3.1

Etant donné une famille de fonctions $\{F_{t_1, t_2, \dots, t_n}, n \in \mathbb{N}^*, t_1, \dots, t_n \in T\}$ définie sur \mathbb{R}^n à valeurs dans $[0, 1]$, une condition nécessaire et suffisante pour qu'il existe un processus aléatoire $\mathbb{X} = (X_t, t \in T)$ défini sur un certain (Ω, \mathcal{F}, P) à valeurs dans (E, \mathcal{E}) et admettant $\{F_{t_1, t_2, \dots, t_n}, n \in \mathbb{N}^*, t_1, \dots, t_n \in T\}$ comme fonctions de répartition fini-dimensionnelles est que les trois conditions suivantes soient vérifiées :

- i) $F_{t_1, t_2, \dots, t_n}(\cdot, \dots, \cdot)$ est une fonction de répartition pour tout t_1, t_2, \dots, t_n fixés.
- ii) $F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) = F_{t_{i_1}, t_{i_2}, \dots, t_{i_n}}(x_{i_1}, x_{i_2}, \dots, x_{i_n})$ pour toute permutation i_1, i_2, \dots, i_n de $1, 2, \dots, n$ et tout $x_i \in \mathbb{R}$.
- iii) $F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_{n-1}, \infty) = F_{t_1, t_2, \dots, t_{n-1}}(x_1, x_2, \dots, x_{n-1})$.

Le processus $(X_t, t \in T)$ peut être construit de la manière suivante :

- Pour domaine d'évolution on prendra T ,
- Pour espace d'état on prendra $E = \mathbb{R}$ et $\mathcal{E} = \mathcal{B}(\mathbb{R})$.
- Pour espace de probabilité on prend (Ω, \mathcal{F}, P) avec Ω l'ensemble de toutes les fonctions $\omega = \omega(\cdot)$ définies de T à valeurs dans \mathbb{R}

$$\omega : \begin{matrix} T \rightarrow E \\ t \mapsto \omega(t) \end{matrix},$$

\mathcal{F} est une tribu des parties de Ω engendrés par les ensembles cylindriques de la forme

$$\begin{aligned} & \{\omega \in \Omega : \omega(t_1) \leq x_1, \omega(t_2) \leq x_2, \dots, \omega(t_n) \leq x_n\} \\ & : = C_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n), x_i \in \mathbb{R} \end{aligned}$$

et P une mesure de probabilité définie sur \mathcal{F} par

$$\begin{aligned} P(C_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n)) & : = F_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(x_1, x_2, \dots, x_n) \\ & = F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_{n-1}, x_n). \end{aligned}$$

- Enfin, $(X_t, t \in T)$ est défini comme étant une application mesurable de (Ω, \mathcal{F}) dans $(E^T, \mathcal{E}^{\otimes T})$

$$\begin{aligned}\mathbb{X} : \Omega &\longrightarrow E^T \\ \omega &\longmapsto \mathbb{X}(\omega) = (X_t(\omega), t \in T) \\ \text{avec } X_t(\omega) &= \omega(t).\end{aligned}$$

4- Caractéristiques de la distribution d'un processus aléatoire

Comme pour les v.a., la distribution (infini-dimensionnelle) d'un processus aléatoire $(X_t, t \in T)$ est aussi caractérisée par certaines familles particulières définies sur T à valeurs dans E , à savoir : la fonction moyenne, la fonction variance, la fonction d'autocovariance... Dans la suite on considère un processus aléatoire $(X_t, t \in T)$ défini sur un espace de probabilité (Ω, \mathcal{F}, P) à valeurs dans un espace d'état $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et de domaine d'évolution T .

a- Fonction moyenne La fonction moyenne $\mu(\cdot)$ est une fonction de T dans \mathbb{R} qui pour tout $t \in T$ associe l'espérance mathématique du membre X_t . Autrement dit, c'est la fonction définie par

$$\begin{aligned}\mu(\cdot) : T &\longrightarrow \mathbb{R} \\ t &\longmapsto \mu(t) = E(X_t),\end{aligned}$$

et ayant comme domaine de définition l'ensemble $D_\mu = \{t \in T : E(X_t) < \infty\}$.

b- Fonction variance La fonction variance $\sigma^2(\cdot)$ est une fonction de T dans \mathbb{R} qui pour tout $t \in T$ associe la variance du membre X_t . Autrement dit, c'est la fonction définie par

$$\begin{aligned}\sigma^2(\cdot) : T &\longrightarrow \mathbb{R} \\ t &\longmapsto \sigma^2(t) = \text{var}(X_t),\end{aligned}$$

et ayant comme domaine de définition l'ensemble $D_{\sigma^2} = \{t \in T : E(X_t^2) < \infty\}$.

c- Fonction d'autocovariance La fonction d'autocovariance $\gamma(\cdot, \cdot)$ est une fonction de $T \times T$ dans \mathbb{R} qui pour tout couple $(t, s) \in T \times T$ associe la covariance entre les membres X_t et X_s . Autrement dit, c'est la fonction définie par

$$\begin{aligned}\gamma(\cdot, \cdot) : T \times T &\longrightarrow \mathbb{R} \\ (t, s) &\longmapsto \gamma(t, s) = \text{cov}(X_t, X_s),\end{aligned}$$

et ayant comme domaine de définition l'ensemble $D_\gamma = \{(t, s) \in T \times T : E(X_t X_s) < \infty\}$.

d- Fonction d'autocorrélation On définit également la fonction d'autocorrélation $\rho(\cdot, \cdot)$ comme une fonction de $T \times T$ dans $[-1, 1]$ qui pour tout couple $(t, s) \in T \times T$ associe la corrélation entre les membres X_t et X_s . Autrement dit, c'est la fonction définie par

$$\begin{aligned}\rho(\cdot, \cdot) : T \times T &\longrightarrow \mathbb{R} \\ (t, s) &\longmapsto \rho(t, s) = \text{corr}(X_t, X_s),\end{aligned}$$

et ayant comme domaine de définition l'ensemble $D_\rho = D_\gamma$.

e- Fonction génératrice - On appelle fonction génératrice des moments de $\mathbb{X} = (X_t, t \in T)$ la fonction $M_{\mathbb{X}}(.,.)$ définie de \mathbb{R} dans \mathbb{R} par

$$M_{\mathbb{X}}(.,.) : T \times \mathbb{R} \longrightarrow \mathbb{R} \\ (t, z) \longmapsto M_{\mathbb{X}}(t, z) = E(e^{zX_t}).$$

Le domaine de définition de $M_{\mathbb{X}}(.,.)$ est alors donné par

$$D_{M_{\mathbb{X}}} = \{(t, z) \in T \times \mathbb{R} : E(e^{zX_t}) < \infty\}.$$

- On appelle fonction génératrice des moments factoriels de $\mathbb{X} = (X_t, t \in T)$ la fonction $G_{\mathbb{X}}(.,.)$ définie de $T \times \mathbb{R}^+$ dans \mathbb{R} par

$$G_{\mathbb{X}}(.) : T \times \mathbb{R}^+ \longrightarrow \mathbb{R} \\ (t, z) \longmapsto G_{\mathbb{X}}(t, z) = E(z^{X_t}).$$

f- Fonction caractéristique On appelle fonction caractéristique de $\mathbb{X} = (X_t, t \in T)$ la fonction $\varphi_{\mathbb{X}}(.,.)$ définie de $T \times \mathbb{R}$ dans \mathbb{C} par

$$\varphi_{\mathbb{X}}(.,.) : T \times \mathbb{R} \longrightarrow \mathbb{C} \\ (t, z) \longmapsto \varphi_{\mathbb{X}}(t, z) = E(e^{izX_t}).$$

IV- Structures de dépendance usuelles des processus aléatoires

On a vu plus haut que l'une des caractéristiques importantes des phénomènes évolutifs est la structure de dépendance stochastique qui caractérise l'évolution d'un membre à un autre. De même, les processus aléatoires qui sont destinés à représenter les phénomènes évolutifs sont aussi caractérisés par des structures spécifiques. Puisqu'un processus aléatoire peut être considéré comme suite ou fonction aléatoire (cf, Définition 2.2.2) selon que son domaine d'évolution soit dénombrable ou non dénombrable, des propriétés usuelles concernant les suites et les fonctions numériques peuvent être adaptées au cas de processus aléatoires selon le principe suivant :

(P) Principe d'adaptation du déterminisme au stochastisme : *une propriété exprimant une relation entre les valeurs des membres d'une suite (ou d'une fonction) déterministe est adaptée pour une suite aléatoire (processus aléatoire) en exprimant la même relation mais plutôt entre les probabilités des valeurs des membres i.e. les probabilités que les valeurs des termes du processus soient dans un certain ensemble.*

Dans ce qui suit nous en étudions quelques-unes des plus connues.

1- Indépendance stochastique

La structure de dépendance la plus simple voire triviale que pourrait avoir un processus aléatoire est la propriété d'absence de dépendance, dite aussi *indépendance stochastique*, indépendance qu'on a étudié dans le cadre de variables aléatoires liées aux phénomènes statiques. Ce concept a été lui aussi adapté, selon le principe d'adaptation **(P)**, à partir de la propriété d'indépendance déterministe entre deux grandeurs déterministes. En effet, deux grandeurs (variables) déterministes sont dites *déterministiquement indépendantes* si la connaissance de la valeur de l'une ne détermine pas ou pour ainsi dire n'influence pas la valeur

de l'autre. Par exemple, pour le modèle de pendule étudié plus haut (dans le préambule), les variables période T et masse m sont indépendantes pour des valeurs négligeables de l'angle θ , puisque la connaissance de la masse m du corps ne détermine pas par la relation (1.1.1) la période, même connaissant toutes les autres variables intervenantes. En revanche, dans le cas de variables aleatoires, on a vu plus haut que deux v.a. sont dites *stochastiquement indépendantes* si la connaissance de la réalisation de l'une ne détermine pas ou n'influence pas plutôt la probabilité de réalisation de l'autre. On a vu aussi que cette propriété s'étend au cas d'un nombre fini de v.a. si l'indépendance stochastique est vérifiée pour chaque couple de variables. Cette propriété s'étend également dans le cadre de processus aleatoires, en ce sens qu'un processus est dit indépendant si ses membres sont stochastiquement indépendants les uns les autres. Un tel processus est en soi d'une utilité superflue et représente rarement des phénomènes évolutifs réels, puisqu'il n'exprime aucune dépendance stochastique. Cependant il est l'un des processus les plus utilisés dans la composition d'autres processus avec structure de dépendance appropriées.

2- Markovianite

3- Stationnarité

La propriété de stationnarité (stochastique), qui caractérise plutôt une certaine régularité stochastique dans l'évolution, joue un rôle crucial dans la théorie des processus aléatoires. Dans plusieurs problèmes du monde réel, on rencontre des phénomènes aléatoires qui évoluent dans un régime "d'équilibre stochastique" dans le sens où les caractéristiques fréquentistes du phénomène ne changent pas dans le domaine d'évolution. De tels phénomènes peuvent être représentés par lesdits *processus stationnaires* dont la définition a été empruntée et adaptée (selon le principe d'adaptation du déterminisme au stochastiquisme) à partir du concept de stationnarité des suites et des fonctions numériques.

Pour rappel, une suite numérique $(U_t, t \in \mathbb{N})$ est dite stationnaire si la valeur des termes est invariante dans le temps, i.e. si

$$U_{t+1} = U_t \quad \forall t \in \mathbb{N}, \quad (2.4.1a)$$

ou de manière équivalente si

$$U_{t+h} = U_t \quad \forall t, h \in \mathbb{N}. \quad (2.4.1b)$$

Pour adapter cette propriété à une suite aléatoire $(X_t, t \in \mathbb{N})$, c'est à dire un processus aléatoire défini sur un certain espace de probabilité (Ω, \mathcal{F}, P) à valeurs dans disons $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, il est clair qu'il est peu convenable de considérer une définition analogue à (2.4.1) du style

$$X_{t+1} = X_t \quad \forall t \in \mathbb{N}, \quad (2.4.2a)$$

ou encore

$$X_{t+1}(\omega) = X_t(\omega) \quad \forall t \in \mathbb{N}, \forall \omega \in \Omega, \quad (2.4.2b)$$

puisque de tels processus seraient inadéquats à représenter des situations réelles où règne l'aléa. Au lieu d'opérer directement sur les valeurs via (2.4.2), il paraît plus judicieux, en vertu du principe d'adaptation du déterminisme au stochastiquisme, d'adapter (2.4.1) aux probabilités des événements attachés aux valeurs plutôt qu'aux valeurs elles-mêmes. Ainsi,

on dira dans un premier temps qu'un processus processus aléatoire est stationnaire si la probabilité qu'un terme X_t se trouve dans une région quelconque B est invariante dans le domaine d'évolution, i.e. si

$$P(X_t \in B) = P(X_{t+1} \in B) \quad \forall t \in \mathbb{N}, \forall B \in \mathcal{B}(\mathbb{R}), \quad (2.4.3a)$$

ou de manière équivalente si

$$P(X_t \in B) = P(X_{t+h} \in B) \quad \forall t, h \in \mathbb{N}, \forall B \in \mathcal{B}(\mathbb{R}). \quad (2.4.3b)$$

Une telle définition peut être rapportée aux fonctions de répartitions comme suit :

$$P(X_t \leq x) = P(X_{t+h} \leq x) \quad \forall t, h \in \mathbb{N}, \forall x \in \mathbb{R}. \quad (2.4.3c)$$

Bien que la relation (2.4.3) semble adaptée au cas stochastique, elle ne traduit en fait qu'une stationnarité marginale comme dans la figure 2.4.1

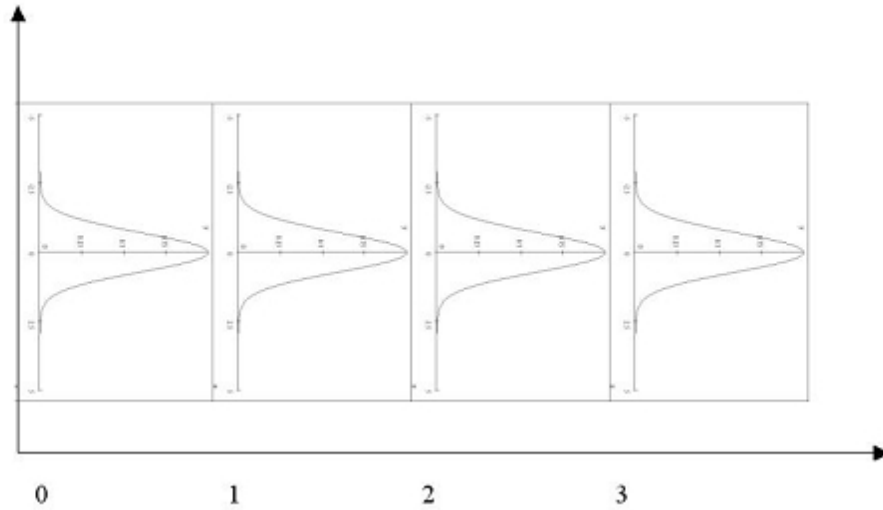


Figure 2.4.1 Stationnarité marginale

et il n'est pas clair si c'en est le cas pour les lois conjointes de deux, trois membres et plus... On peut ainsi généraliser la relation (2.4.3c) au cas de loi conjointe à deux termes comme suit

$$P(X_{t_1} \leq x_1, X_{t_2} \leq x_2) = P(X_{t_1+h} \leq x_1, X_{t_2+h} \leq x_2) \quad \forall t_1, t_2, h \in \mathbb{N}, \forall x_1, x_2 \in \mathbb{R},$$

au cas de lois conjointes à trois termes

$$P(X_{t_1} \leq x_1, X_{t_2} \leq x_2, X_{t_3} \leq x_3) = P(X_{t_1+h} \leq x_1, X_{t_2+h} \leq x_2, X_{t_3+h} \leq x_3) \quad \forall n_i, h \in \mathbb{N}, \forall x_i \in \mathbb{R}, i = 1, 2,$$

et de manière générale au cas de n termes ($n \in \mathbb{N}^*$)

$$P(X_{t_1} \leq x_1, X_{t_2} \leq x_2, \dots, X_{t_n} \leq x_n) = P(X_{t_1+h} \leq x_1, X_{t_2+h} \leq x_2, \dots, X_{t_n+h} \leq x_n) \quad \forall n_i, h \in \mathbb{N}, \forall x_i \in \mathbb{R}, i$$

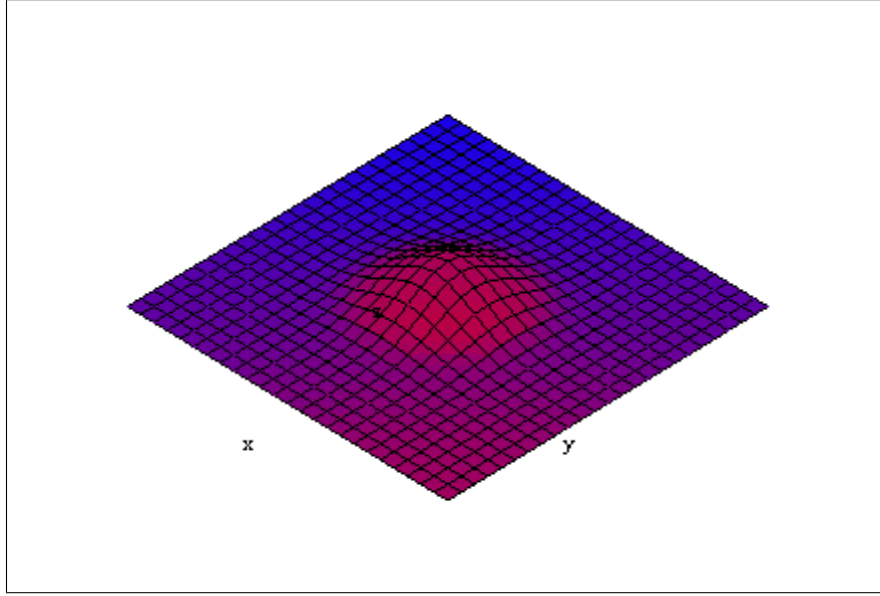
traduisant ainsi la stationnarité en termes des distributions fini-dimensionnelles et donc (dans ce cas à temps discret) en terme de la loi de probabilité du processus entier. C'est donc la définition de stationnarité qu'on retient de façon générale. Elle a été introduite par Kintchine (1931) pour le cas d'un domaine d'évolution T quelconque et est dite stationnarité stricte.

Définition 2.4.1 (Stationnarité stricte)

Un processus aléatoire $(X_t, t \in T)$ défini sur un espace de probabilité (Ω, \mathcal{F}, P) à valeurs dans un espace d'état $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et de domaine d'évolution T est dit strictement stationnaire si

$$P(X_{t_1} \leq x_1, X_{t_2} \leq x_2, \dots, X_{t_n} \leq x_n) = P(X_{t_1+h} \leq x_1, X_{t_2+h} \leq x_2, \dots, X_{t_n+h} \leq x_n) \quad \forall n, h \in T, \forall x_i \in \mathbb{R},$$

tel que $t_i + h \in T, i = 1, \dots, n$. Autrement dit, si les distributions fini-dimensionnelles sont invariantes par translation admissible dans le domaine d'évolution.



Il est important de noter que tous les moments d'un processus strictement stationnaire, lorsqu'ils existent, sont invariants dans le temps. Le processus est ainsi dit stationnaire en tous les moments ou à l'ordre ∞ . Ainsi, la définition de stationnarité stricte semble contraignante puisque lorsque existent tous les moments, elle exige l'invariance de tous ces moments par rapport au temps. De plus, elle repose sur la connaissance des lois fini-dimensionnelles du processus qui ne peuvent être connues en pratique, sauf dans des cas très spéciaux. Cependant, plusieurs propriétés probabilistes essentielles des processus aléatoires peuvent être obtenues juste à partir des deux premiers moments (lorsqu'ils existent) et pour les moments restants, la distinction est souvent négligeable. La stationnarité de ces deux premiers moments peut donc être suffisante pour expliquer, du moins avec bonne précision, la stationnarité dans la distribution du processus. En particulier, pour les processus dits gaussiens qui sont très répandus en pratique, la stationnarité de ces deux premiers moments est équivalente à la stationnarité stricte. C'est pourquoi, on a souvent besoin d'un concept de stationnarité moins fort mais qui peut être rencontré en pratique.

Définition 2.4.2 (Stationnarité au second ordre)

Un processus aléatoire $(X_t, t \in T)$ défini sur un espace de probabilité (Ω, \mathcal{F}, P) à valeurs dans un espace d'état $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et de domaine d'évolution T est dit stationnaire au second ordre (faiblement stationnaire, covariance stationnaire) si

- i) La fonction variance $\sigma^2(t) := \text{var}(X_t) < \infty$ existe ($\sigma^2(t) < \infty$) pour tout t dans le domaine d'évolution T (et donc nécessairement la fonction moyenne $\mu(t) = E(X_t)$ existe : $E(X_t^2) < \infty$).
- ii) La fonction moyenne $\mu(t)$ est constante sur le domaine d'évolution et la fonction d'autocovariance $\gamma(t, t+h) = \text{cov}(X_t, X_{t+h})$ dépend seulement de h , ce qui entraîne que la fonction variance $\sigma^2(t)$ est constante dans le domaine d'évolution T .