

C-13A

COURS de PROBABILITES

Chapitres I,II,III

G. MATHERON

Juin-Juil-Août 1965

COURS de PROBABILITES

Table des MatièresPages

<u>CHAPITRE I - AXIOMATIQUE</u>	1
I.- NECESSITE D'UNE AXIOMATIQUE	1
II.- EVENEMENTS ET ENSEMBLE D'EVENEMENTS ELEMENTAIRES	3
a) Somme (ou réunion, ou disjonction)	4
b) Produit (ou intersection, ou conjonction)	4
c) Passage au complémentaire (négation)	4
d) Inclusion (ou implication)	5
e) Evènement certain, évènement impossible	5
f) Evènements incompatibles	6
III.- DEFINITION DE LA PROBABILITE PAR LES EVENEMENTS ELEMENTAIRES EQUIPROBABLES	6
IV.- LES AXIOMES DU CALCUL DES PROBABILITES	8
Axiomes des σ -algèbres	9
Génération d'une σ -algèbre	10
Axiome des Probabilités	14
Théorème 1	17
Théorème 2	19
Théorème 3	21
V.- PROBABILITE CONDITIONNELLE - EVENEMENTS INDEPENDANTS	23
Théorème 4	24
Formule de Bayes	24
Théorème 5	27
<u>CHAPITRE II.- VARIABLES ALEATOIRES ET FONCTION DE REPARTITION</u>	33
I.- NOTION DE VARIABLE ALEATOIRE	33
Théorème 1	37
II.- FONCTIONS DE REPARTITION A UNE SEULE VARIABLE	38
Propriété 1	39
Propriété 2	40
Propriété 3	40
Propriété 4	41
III.- FONCTIONS DE REPARTITION DE PLUSIEURS VARIABLES	43
IV.- CHANGEMENTS DE VARIABLES	47
Théorème 1	48
Théorème 2	48
Théorème 3	48
Théorème 4	49

Table des Matières (suite)

V.-	INTEGRALES DE LEBESGUE ET DE STIELTJES	51
	Intégrale de LEBESGUE	51
	Propriété 1, 2, 3	52
	Propriété 4, 5	53
	Intégrale de STIELTJES	53
	Propriété 6	54
	Propriété 7	55
	Propriété 8	55
VI.-	ESPERANCE MATHEMATIQUE ET MOMENTS	57
	Propriétés de l'espérance mathématique	60
VII.-	CONVERGENCE EN LOI	67
	Premier théorème de Helly	69
	Deuxième théorème de Helly	72
 <u>CHAPITRE III.- LA FONCTION CARACTERISTIQUE</u>		76
I.-	DEFINITION ET PROPRIETES.	76
	Propriété 1	76
	Propriété 2	78
	Propriété 3	78
	Propriété 4	79
	Propriété 5	79
II.-	FONCTION CARACTERISTIQUE D'UNE A PLUSIEURS VARIABLES	81
	Propriété 1	82
	Propriété 2	82
	Propriété 3	82
	Propriété 4	82
	Propriété 5	83
	Propriété 6	83
III.-	PROPRIETE FONDAMENTALE DE LA FONCTION CARACTERISTIQUE	84
	Théorème fondamental	84
	Théorème d'unicité	85
	Premier théorème limite	87
	Théorème Réciproque 1.	87
IV.-	FONCTIONS DE TYPE POSITIF	91
	Définition	91
	Propriété 1	91
	Théorème de Bochner	92
	Propriété 2	93
	Propriété 3	94

Table des Matières (suite)

Propriété 4	94
V.- FONCTIONS GENERATRICES	94
Propriété 1	95
Propriété 2	95
Propriété 3	95
Propriété 4	96
VII.- LOIS COMPOSEES	98
<u>CHAPITRE IV.- CONVERGENCES ET LOIS DES GRANDS NOMBRES</u>	102
I.- CONVERGENCE EN PROBABILITE ET LOI ORDINAIRE DES GRANDS NOMBRES	102
Théorème 1	102
Théorème 2	105
II.- CONVERGENCE PRESQUE SURE, ET LOI FORTE DES GRANDS NOMBRES	106
Critère 1	107
Critère 2	108
Critère 3	108
Loi Forte des grands nombres	111
Théorème de Borel	112
Premier théorème de Kolmogorov	113
Deuxième théorème de Kolmogorov	113
III.- CONVERGENCE EN MOYENNE QUADRATIQUE	115
Critère de Cauchy	115

COURS de PROBABILITES

Chapitre I

AXIOMATIQUE

G. MATHERON

Septembre 1964

CALCUL DES PROBABILITES

CHAPITRE I

AXIOMATIQUE

I.- NECESSITE D'UNE AXIOMATIQUE

Comme toute théorie mathématique, le calcul des probabilités procède d'une axiomatique. Il ne cherchera pas à définir ce qu'est la probabilité "en soi", pas plus que la géométrie ne définit ce qu'est le point, ou la droite "en soi", pas plus que dans la théorie des espaces vectoriels on ne définit ce qu'est un vecteur "en soi". La notion d'espace vectoriel n'est pas liée aux particularités individuelles ou aux propriétés expérimentales des vecteurs, mais uniquement à un ensemble de relations, ou axiomes, que ceux-ci doivent vérifier par définition. Les axiomes une fois posés, la théorie se construit d'une manière purement déductive, sans recours à l'intuition ou au contrôle expérimental.

Ce congé formel donné au concret est, en fait, plus apparent que réel. En principe n'importe quel système d'axiomes, pourvu qu'il soit non-contradictoire, peut servir de base à une théorie mathématique ! mais cette théorie ne sera pas forcément intéressante. Les "bons" systèmes d'axiomes ne se sont pas constitués au hasard. Ils sont le fruit d'un effort de réflexion, poursuivi pendant des générations humaines, dont le but constant était de réduire à des principes de plus en plus simples et de plus en plus généraux les données de l'expérience sensible. Il n'est donc pas étonnant que l'on rencontre dans la nature des objets ou des phénomènes qui veulent bien vérifier (à l'approximation de nos mesures) les "bons" axiomes en question, et auxquels par suite toutes les déductions de la théorie sont applicables (avec la même approximation).

Il n'en reste pas moins qu'en bonne méthodologie scientifique on doit soigneusement distinguer la construction de la théorie, à partir de ses bases axiomatiques, et son application possible à tel ou tel domaine concret. Le fait que les architectes, les topographes, etc ... utilisent les théorèmes de la géométrie euclidienne ne constitue pas une démonstration de ces théorèmes, mais indique seulement que les axiomes

euclidiens représentent une approximation remarquablement exacte des propriétés de l'espace concret dans lequel nous vivons.

Le calcul des probabilités a rencontré des succès spectaculaires dans presque toutes les branches de la science, depuis la physique jusqu'à l'économie ou la sociologie. Ce sont l'ampleur et la généralité même de ces succès qui ont contraint les probabilistes à fonder leur discipline sur des bases axiomatiques rigoureuses. Tant qu'on se limitait à l'étude de problèmes qui, comme les jeux de hasard, ne font intervenir qu'un nombre fini d'évènements et peuvent se traiter par l'analyse combinatoire, il était possible de se contenter d'une notion toute intuitive de la probabilité. Dès que l'idée d'infini s'introduit (et elle s'introduit nécessairement dès que l'on considère un ensemble infini d'évènements possibles, ou que l'on envisage la répétition indéfinie d'une même expérience) cette notion intuitive se révèle insuffisante. Donnons deux exemples très simples : le premier, un peu artificiel, est le célèbre paradoxe de Bertrand. Le deuxième, peut être plus profond, est la loi dite des grands nombres.

Paradoxe de Bertrand

On choisit au hasard une corde sur un cercle et on cherche la probabilité pour que la longueur de cette corde soit supérieure au côté du triangle équilatéral inscrit.

1ère solution - Par raison de symétrie, on peut se donner la direction de la corde. Le milieu de la corde décrit alors un diamètre, et la corde convient lorsque ce milieu est compris entre le premier et le troisième quart du diamètre. D'où $p = \frac{1}{2}$.

2ème solution - On peut aussi fixer une extrémité de la corde, et choisir l'autre au hasard sur le cercle. On trouve immédiatement $p = \frac{1}{3}$.

3ème solution - La corde est parfaitement définie par son point médian, et répond à la question lorsque le point médian est pris intérieur au cercle concentrique de rayon deux fois moindre, d'où cette fois $p = \frac{1}{4}$.

Il est très clair que le paradoxe provient de ce que l'on n'a pas défini de la même manière, dans les trois solutions, ce que l'on n'entendait pas ! "choisir une corde au hasard".

Loi des grands nombres

Supposons qu'une expérience puisse comprendre dans son résultat la réalisation ou la non-réalisation d'un évènement E (par exemple, en lançant une pièce de monnaie on peut obtenir pile, ou non). Si l'on répète n fois l'expérience, soit u_n le nombre de réalisations de E. Lorsque n augmente, on constate que la fréquence empirique $\frac{1}{n} u_n$ tend à se rapprocher d'une certaine limite p , en ce sens que les fluctuations de $\frac{1}{n} u_n$ relativement à p semblent, en moyenne, diminuer d'amplitude.

On est tenté d'interpréter ce résultat expérimental en disant: "la fréquence tend vers p lorsque le nombre d'épreuves augmente indéfiniment", et de définir la probabilité comme la limite p de la fréquence. En fait, cette interprétation va très au delà du fait expérimental observé. En effet, dire que $\frac{1}{n} u_n$ tend vers p signifie exactement: "quel que soit ε , on peut trouver N tel que, pour $n \gg N$, on ait $|\frac{1}{n} u_n - p| \leq \varepsilon$ ". Pour justifier expérimentalement une telle proposition, il faudrait effectuer réellement une infinité d'épreuves, et constater que pour toutes les épreuves, en nombre infini, au delà du rang N , l'inégalité est bien vérifiée.

Dans la construction axiomatique, au contraire, la probabilité p sera introduite a priori, et l'on démontrera de manière déductive que l'évènement " $\frac{1}{n} u_n$ tend vers p " possède une probabilité égale à l'unité (loi forte des grands nombres). Le fait que, dans l'expérience réelle, les fluctuations de $\frac{u_n}{n}$ relativement à p aillent en s'amortissant sera considéré comme une indication que les axiomes probabilistes permettent une description correcte de cette expérience.

Avant d'énoncer les axiomes du calcul des probabilités, il convient d'examiner la notion d'évènements et la définition classique de la probabilité par la notion d'évènements élémentaires également probables.

II. EVENEMENTS ET ENSEMBLE D'EVENEMENTS ELEMENTAIRES

Nous considérons une épreuve, ou une expérience (accomplie dans des conditions précisées à l'avance et fixées une fois pour toutes) susceptible d'entraîner la réalisation de l'un ou l'autre de n résultats e_1, e_2, \dots, e_n ou évènements élémentaires. Par exemple, si l'on jette un dé, les six évènements élémentaires sont e_1, \dots, e_6 , où e_i est l'évènement: "le dé marqué i ". De même si l'on jette deux dés, (discernables l'un de l'autre) il y a ici 36 évènements élémentaires $e_{11}, e_{12}, \dots, e_{66}$,

e_{ij} désignant l'évènement "le premier dé marque i et le deuxième marque j ".

Ainsi l'ensemble E des évènements élémentaires $e_1, e_2 \dots e_n$ constitue l'ensemble des résultats possibles de l'épreuve, et chaque e_i constitue, par définition, une description exhaustive d'un résultat. On appellera évènement composé, ou simplement évènement tout sous-ensemble de E , c'est-à-dire tout ensemble $\{e_{i_1}, e_{i_2}, \dots, e_{i_k}\}$ constitué par la réunion de k évènements élémentaires quelconques. On peut effectuer sur les évènements les trois opérations fondamentales, qui sont celles de la théorie des ensembles aussi bien que de la logique !

a - Somme - (ou réunion, ou disjonction).

On appelle somme de deux évènements A et B l'ensemble des évènements élémentaires e appartenant soit à A , soit à B , soit aux deux. La notation ensembliste est $A \cup B$, la notation logique est "A ou B". En calcul des probabilités, on écrit en général simplement $A + B$, aucune confusion n'étant à craindre avec l'addition usuelle. Dire que l'évènement $A + B$ s'est réalisé signifie indifféremment :

- l'épreuve a amené un évènement élémentaire e qui appartient à A , ou à B , ou aux deux (point de vue ensembliste),
- l'épreuve a amené la réalisation de l'évènement A ou de l'évènement B , ou des deux (points de vue logique).

b - Produit (ou intersection, ou conjonction).

Le produit de deux évènements A et B est l'ensemble des évènements élémentaires qui appartiennent à la fois à A et B . La notation ensembliste est $A \cap B$, la notation logique "A et B". En calcul des probabilités on écrit simplement $A B$. Dire que l'évènement AB s'est réalisé signifie aussi bien :

- l'épreuve a amené un évènement élémentaire e qui appartient à la fois à A et B (point de vue ensembliste).
- l'épreuve a amené la réalisation des deux évènements A et B à la fois (point de vue logique).

c - Passage au complémentaire (négation).

Le contraire de l'évènement A est l'ensemble des évènements élémentaires qui n'appartiennent pas à A . On le note \bar{A} . Dire que \bar{A} s'est réalisé signifie :

... / ...

- l'épreuve a amené la réalisation d'un évènement élémentaire qui n'appartient pas à A.

- l'épreuve n'a pas amené la réalisation de A.

Propriétés

- Somme et produit sont associatifs, commutatifs et distributifs l'un par rapport à l'autre.

- le complémentaire d'une somme d'évènements A_i est identique au produit des complémentaires \bar{A}_i , et inversement :

$$\overline{\sum A_i} = \prod \bar{A}_i \qquad \overline{\prod A_i} = \sum \bar{A}_i$$

Ces théorèmes élémentaires de la théorie des ensembles restent vrais même s'il y a une infinité (dénombrable ou non) d'évènements A_i . Naturellement l'égalité de deux évènements $A = B$ doit s'entendre dans le sens "tout évènement élémentaire qui appartient à A appartient aussi à B et réciproquement".

Outre les trois opérations, on définit une relation d'ordre entre les évènements.

d - Inclusion (ou implication).

On dira que A est contenu dans B, ou que A entraîne B, ou que A est un cas particulier de B, et on notera $A \subset B$ si tout évènement élémentaire appartenant à A appartient aussi à B. Lorsqu'il en est ainsi, une épreuve ne peut pas amener la réalisation de A sans amener ipso facto celle de B. Ou encore, deux circonstances seulement sont possibles : ou bien \bar{A} , ou bien AB se réalise. Ou enfin : on a l'égalité $A = AB$.

Nous utiliserons encore les définitions suivantes :

e - Evènement certain, évènement impossible.

Par définition, l'ensemble E de tous les évènements élémentaires est l'évènement certain : quel que soit l'évènement élémentaire θ amené par l'épreuve, celui-ci appartient à E, et par suite la réalisation de E est certaine. Le complémentaire de E est l'évènement impossible. C'est l'ensemble vide de la théorie des ensembles

(noté \emptyset), c'est-à-dire l'évènement qui ne contient aucun évènement élémentaire. Quel que soit le résultat de l'épreuve, celle-ci amènera un évènement élémentaire et par suite \emptyset ne se réalisera pas.

f - Evènements incompatibles

Deux évènements A et B sont dits incompatibles si $AB = \emptyset$, c'est-à-dire s'il n'existe aucun évènement élémentaire appartenant à la fois à A et B. Une épreuve quelconque ne peut alors amener que l'un ou l'autre des trois résultats : $\bar{A}\bar{B}$, $A\bar{B}$, $\bar{A}B$.

Un évènement A et son complémentaire \bar{A} sont incompatibles par définition. Plus précisément : pour que deux évènements soient complémentaires, il faut et il suffit que leur somme soit l'évènement certain ($A + B = E$) et leur produit l'évènement impossible ($AB = \emptyset$).

III.- DEFINITION DE LA PROBABILITE PAR LES EVENEMENTS ELEMENTAIRES EQUIPROBABLES.

Dans un certain nombre de cas (et notamment dans l'étude des jeux de hasard d'où s'est dégagée au 17ème siècle la notion même de probabilité) la structure de l'épreuve possède des propriétés de symétrie qui imposent la notion d'égale probabilité des évènements élémentaires (ex : pile ou face, le dé à jouer, l'ensemble des données possibles au jeu de bridge, etc...). Si les évènements élémentaires sont au nombre de n, on peut convenir que chacun d'eux a la probabilité $\frac{1}{n}$. Plus généralement, à tout évènement A contenant m évènements élémentaires, on peut convenir de faire correspondre le nombre

$$(1) \quad P(A) = \frac{m}{n}$$

égal au rapport du nombre des évènements élémentaires favorables (contenus dans A) au nombre n total, et donner à P(A) le nom de probabilité de l'évènement A. Le calcul d'une probabilité se ramène alors au dénombrement des cas favorables et relève des méthodes de l'analyse combinatoire.

Indiquons tout de suite les limites d'une telle définition :

- toutes les épreuves ne posséderont pas nécessairement les propriétés de symétrie qui rendent évidentes l'équiprobabilité des évènements élémentaires (le dé peut être pipé, etc ...). En fait, en dehors du domaine restreint des jeux de hasard, cette équiprobabilité intuitive sera plutôt l'exception que la règle.

- la définition n'a plus de sens lorsqu'il y a une infinité d'évènements élémentaires (même équiprobables), et, cependant, l'introduction de l'infini est une nécessité impérative.

Néanmoins, cette définition élémentaire de la probabilité $P(A)$ permet de démontrer un certain nombre de propriétés fondamentales : ces mêmes propriétés, convenablement généralisées et érigées en axiomes, (et non plus démontrées) serviront de base à toute la construction probabiliste.

Propriété 1 - $P(A)$ est une fonction non négative de l'ensemble A . Cela résulte de la définition même ($m \geq 0$ et $n > 0$).

Propriété 2 - L'évènement certain a une probabilité égale à l'unité $P(E) = 1$: en effet, pour E , on a $m = n$.

Propriété 3 - Si les évènements $A_1, A_2 \dots A_k$ sont deux à deux incompatibles, (c'est-à-dire $A_i \cap A_j = \emptyset$ pour $i \neq j$) leurs probabilités sont additives :

$$(2) \quad P(A_1 + A_2 + \dots + A_k) = P(A_1) + P(A_2) + \dots + P(A_k)$$

Ici encore, la démonstration est immédiate : Comme les évènements sont incompatibles, le nombre d'évènements élémentaires contenus dans la somme est manifestement égal à la somme des évènements élémentaires contenus dans chacun des A_i .

De ces trois propriétés se déduisent plusieurs autres qui seront énoncées au paragraphe suivant.

La notion de probabilité conditionnelle s'introduit aussi de façon toute naturelle : la probabilité conditionnelle de l'évènement A , dans l'hypothèse où l'évènement B s'est réalisée, notée $P(A|B)$ sera égale, par définition, au rapport

$$(3) \quad P(A|B) = \frac{m(A \cap B)}{n(B)}$$

du nombre $m(A \cap B)$ d'évènements élémentaires contenus dans l'intersection de A et B au nombre $n(B)$ d'évènements élémentaires contenus dans B . En comparant (1) à (3), on

voit que l'on passe de la probabilité $P(A)$, dite a priori, à la probabilité conditionnelle $P(A|B)$ en restreignant à B l'ensemble E des événements élémentaires. Naturellement, ceci n'a de sens que si B n'est pas l'évènement impossible, c'est-à-dire si $n(B) \neq 0$.

Comme conséquence immédiate des définitions (1) et (3) on a la propriété fondamentale :

$$(4) \quad P(A|B) = \frac{P(AB)}{P(B)}$$

valable pourvu que $P(B) \neq 0$. Sous forme symétrique, cette relation s'écrira encore :

$$(5) \quad P(AB) = P(A|B) P(B) = P(B|A) P(A)$$

et sera valable cette fois même si $P(B)$ ou $P(A)$ est nulle.

IV. - LES AXIOMES DU CALCUL DES PROBABILITES

Les axiomes du calcul des probabilités ne sont autres que les trois propriétés ci-dessus convenablement généralisées au cas où il y a une infinité d'évènements élémentaires.

Nous supposons donné un ensemble E d'éléments e appelés ici encore évènements élémentaires, ensemble qui peut être fini ou non, dénombrable ou non. Du point de vue pratique, E pourra encore être interprété comme l'ensemble des résultats possibles d'une épreuve définie : mais, du point de vue axiomatique, cette interprétation est superflue, et E est simplement un ensemble abstrait quelconque.

A la différence de ce qui se passe lorsque E est fini, la dénomination d'évènement ne s'appliquera pas à n'importe quel sous ensemble de E , mais seulement à une classe particulière de sous ensembles privilégiés, classe que nous désignerons ici par \mathcal{A} : les propositions "A appartient à \mathcal{A} " et "A est un évènement" sont, par définition, équivalentes. La nécessité de la restriction de la notion d'évènement à une classe \mathcal{A} privilégiée résulte de la théorie générale de la mesure, et ne peut pas

être justifiée ici : en fait, cette classe \mathcal{A} sera en général si large qu'elle contiendra non seulement tous les sous-ensembles A pratiquement utiles à considérer, mais même tous ceux qu'il est possible à l'imagination humaine de construire. La classe \mathcal{A} des sous-ensembles de E appelés "événements" ne peut pas être choisie tout-à-fait arbitrairement. Il est nécessaire, en effet, que le résultat d'opérations logiques quelconques effectuées sur des événements soit encore un événement : Si A et B sont des événements, il doit en être de même de \bar{A} , $A + B$ ("A ou B") et AB ("A et B"). Enfin - en vue de permettre l'introduction de l'infini - cette propriété doit subsister lorsque l'on effectue une suite illimitée (c'est-à-dire une infinité dénombrable) d'opérations logiques. De plus, l'événement certain E doit lui-même être un événement (et par suite aussi sa négation $\bar{E} = \emptyset$ qui est l'événement impossible).

Ces propriétés définissent une structure ensembliste appelée " σ - algèbre". Nous dirons : l'ensemble \mathcal{A} des événements de E doit constituer une σ - algèbre sur E , entendant par là que \mathcal{A} doit vérifier les trois axiomes suivants :

Axiomes des σ - algèbres

- 1 - Si A appartient à \mathcal{A} alors \bar{A} appartient à \mathcal{A}
- 2 - Si les ensembles d'une suite $A_1, A_2, \dots, A_n, \dots$ appartiennent à \mathcal{A} , leur somme $\sum_{n=1}^{\infty} A_n$ et leur produit $\prod_{n=1}^{\infty} A_n$ appartiennent à \mathcal{A}
- 3 - L'ensemble plein (ou événement certain) E appartient à \mathcal{A}

En langage bref : \mathcal{A} est une σ - algèbre sur E si les complémentaires, les réunions dénombrables, les intersections dénombrables d'ensembles de \mathcal{A} et E lui-même appartiennent à \mathcal{A}

Remarque : les deux propriétés énoncées dans l'axiome 2 ne sont pas indépendantes. Elles se déduisent l'une de l'autre, compte tenu de l'axiome 1. En effet, supposons que toutes les sommes $\sum_{n=1}^{\infty} A_n$ d'ensembles de \mathcal{A} appartiennent à \mathcal{A} . Alors un produit $P = \prod_{n=1}^{\infty} B_n$ d'ensembles B_n de \mathcal{A} pour complémentaire $\bar{P} = \sum_{n=1}^{\infty} \bar{B}_n$ la somme des \bar{B}_n , qui appartiennent à \mathcal{A} d'après l'axiome 1. Par suite \bar{P} appartient à \mathcal{A} . D'après le même axiome 1, le complémentaire de \bar{P} , qui est P lui-même, appartient à \mathcal{A} . La démonstra-

tion est terminée.

tion est la même si l'on suppose que ce sont les produits qui appartiennent à \mathcal{A}

Génération d'une σ -algèbre.

Il arrive souvent que les conditions concrètes d'un problème nous imposent de considérer comme des événements certains sous-ensembles privilégiés de E constituant une famille \mathcal{F} , laquelle n'est pas une σ -algèbre. Mais alors, le principe posé ci-dessus nous impose aussi de considérer comme des événements E lui-même (s'il n'appartient pas déjà à \mathcal{F}) et tous les sous-ensembles de E que l'on peut construire en appliquant à des ensembles appartenant à \mathcal{F} (ou égaux à E) une suite infinie d'opérations logiques quelconques (somme, produit ou négation). On est ainsi conduit à construire la plus petite σ -algèbre possible contenant les ensembles de \mathcal{F} . Son existence résulte des considérations suivantes :

Soit \mathcal{F} une famille quelconque de sous-ensembles de E . Il existe au moins une σ -algèbre contenant \mathcal{F} : à savoir, l'ensemble de tous les sous-ensembles de E (vérification immédiate des 3 axiomes des σ -algèbres). Par ailleurs, si \mathcal{A}_1 et \mathcal{A}_2 sont deux σ -algèbres contenant \mathcal{F} , leur intersection (qui est la famille des sous-ensembles A de E appartenant à la fois à \mathcal{A}_1 et à \mathcal{A}_2) est encore une σ -algèbre contenant \mathcal{F} (vérification immédiate des axiomes), et cette propriété est encore vraie de l'intersection d'un ensemble infini de σ -algèbres contenant \mathcal{F} . En particulier, l'intersection de toutes les σ -algèbres contenant \mathcal{F} est elle-même une σ -algèbre contenant \mathcal{F} : plus précisément, c'est la plus petite σ -algèbre contenant \mathcal{F} , et on l'appelle σ -algèbre engendrée par \mathcal{F} (on la note quelquefois $\sigma(\mathcal{F})$).

Exemple - si E est l'ensemble \mathbb{R} des nombres réels et \mathcal{F} l'ensemble des intervalles, la σ -algèbre engendrée par \mathcal{F} constitue la famille \mathcal{B} des ensembles boréliens qui joue un grand rôle en probabilité. De même, si E est l'espace euclidien \mathbb{R}^n à n dimension, et \mathcal{F} l'ensemble des pavés, les ensembles de $\sigma(\mathcal{F})$ sont les ensembles boréliens de \mathbb{R}^n .

Définition - Espace probabilisable : Par définition, nous appellerons espace probabilisable le couple (E, \mathcal{A}) constitué par un ensemble E et par une σ -algèbre \mathcal{A} composée de sous-ensembles de E .

Par définition, également, les éléments de l'ensemble E seront appelés

évènements élémentaires, et les sous-ensembles A de E appartenant à la σ -algèbre \mathcal{A} seront appelés évènements.

Dans cette définition abstraite, plus rien, en apparence, n'évoque les notions intuitives d'évènements et de probabilité. On verra cependant que leurs caractéristiques essentielles sont conservées. En particulier, les trois axiomes des σ -algèbres nous garantissent que le résultat d'une suite (finie ou dénombrable) d'opérations logiques effectuées sur des évènements sera toujours lui-même un évènement. De même, la définition suivante (produit d'espaces probabilisables) est nécessaire si l'on veut formuler sur des bases axiomatiques correctes la notion d'alternative répétée et, plus généralement, celle de répétition indéfinie d'une même épreuve: une telle formulation est nécessaire, par exemple, à la démonstration de la loi des grands nombres.

Produit d'espaces probabilisables.

Soient (E_1, \mathcal{A}_1) et (E_2, \mathcal{A}_2) deux espaces probabilisables. Le produit de (E_1, \mathcal{A}_1) et (E_2, \mathcal{A}_2) est l'espace probabilisable (E, \mathcal{A}) défini de la manière suivante :

1 - L'ensemble E est le produit (cartésien) $E_1 \times E_2$, c'est-à-dire l'ensemble des couples (e_1, e_2) où e_1 appartient à E_1 et e_2 à E_2 : un évènement élémentaire de E est défini comme l'ensemble (e_1, e_2) de deux évènements élémentaires de E_1 et E_2 . En termes intuitifs, e représente le résultat d'une épreuve composée de deux épreuves élémentaires ayant comme résultats e_1 et e_2 respectivement.

2 - La σ -algèbre \mathcal{A} est la σ -algèbre engendrée par les sous-ensembles de E de la forme $A_1 \times A_2$, où A_1 appartient à \mathcal{A}_1 et A_2 à \mathcal{A}_2 . Le produit $A_1 \times A_2$ des évènements A_1 et A_2 représente l'ensemble des évènements élémentaires e de la forme (e_1, e_2) , où e_1 appartient à A_1 et e_2 à A_2 . En termes intuitifs, dire que l'épreuve composée amène l'évènement $A_1 \times A_2$ signifie que les deux épreuves composantes ont amené, respectivement, les évènements A_1 et A_2 . On notera qu'outre les évènements - produits $A_1 \times A_2$ (qui ne constituent pas une σ -algèbre) \mathcal{A} contient d'autres évènements de structure plus complexe. Cela se comprend intuitivement en considérant le passage de l'espace euclidien R^1 à une dimension à l'espace produit $R^2 = R^1 \times R^1$ qui est l'espace euclidien à 2 dimensions.

La notion de produit se généralise au cas d'un nombre fini ou d'une infinité dénombrable d'espaces probabilisables. Plus précisément, étant donné une suite infinie d'espaces probabilisables (E_n, \mathcal{A}_n) on appellera produit l'espace probabilisable (E, \mathcal{A}) où :

1 - E est l'ensemble produit $E_1 \times E_2 \times \dots$, c'est-à-dire l'ensemble des suites $e = (e_1, e_2, \dots)$ d'évènements élémentaires de E_1, E_2, \dots

2 - \mathcal{A} est la σ -algèbre engendrée par les sous-ensembles de E de la forme :

$$(6) \quad A = A_1 \times A_2 \times \dots \times A_k \times E_{k+1} \times E_{k+2} \times \dots$$

que nous appellerons évènements simples.

k est un entier quelconque, et A_i appartient à \mathcal{A}_i . L'évènement simple A est composé des évènements élémentaires e de la forme (e_1, e_2, \dots) , où e_1, e_2, \dots, e_k appartiennent respectivement à A_1, A_2, \dots, A_k , tandis que e_{k+1} et les suivants sont des évènements élémentaires quelconques de E_{k+1}, E_{k+2}, \dots

Pour décrire la répétition indéfinie d'une même épreuve, on devra considérer le produit infini d'espaces probabilisables (E_i, \mathcal{A}_i) , où chaque ensemble E_i est identique à un même ensemble E , et chaque σ -algèbre \mathcal{A}_i identique à une même σ -algèbre \mathcal{A} . Le produit est ici l'espace probabilisable $(E^{(\omega)}, \mathcal{A}^{(\omega)})$ où :

1 - $E^{(\omega)}$ est l'ensemble des suites $e = (e_1, e_2, \dots, e_n, \dots)$ d'évènements élémentaires de E : en termes intuitifs, l'ensemble des résultats possibles d'une suite infinie d'épreuves identiques à l'épreuve associée à (E, \mathcal{A}) .

2 - $\mathcal{A}^{(\omega)}$ représente la σ -algèbre engendrée par les sous-ensembles de $E^{(\omega)}$ de la forme (6), ou évènements simples, A_1, A_2, \dots, A_k appartenant à \mathcal{A} tandis que E_{k+1}, E_{k+2}, \dots coïncident avec E . La réalisation de l'évènement simple A a la signification suivante : au cours d'une suite infinie d'épreuves, les k premières épreuves (en nombre fini) ont amené respectivement les évènements A_1, A_2, \dots, A_k , aucune spécification n'étant faite sur les résultats des épreuves suivantes (en nombre infini).

Convenons d'appeler évènement simple un évènement de la forme (6) dont la définition n'intéresse que les résultats des k premières épreuves (k fini). La σ -algèbre \mathcal{A}^{ω} contient, évidemment, des évènements de structure beaucoup plus complexe. Mais, en général, les propriétés les plus intéressantes pourront être établies à l'aide des seuls évènements simples : concrètement, cela signifie que l'étude d'une suite infinie d'épreuves se ramènera à celle de suites d'un nombre fini (mais quelconques) de ces mêmes épreuves : on reconnaît là un procédé très général de traitement de la notion d'infinité en mathématique.

Exemple : alternative répétée.

Soit une épreuve susceptible d'amener deux résultats seulement (un "succès" et un "échec"). L'ensemble E est constitué de deux évènements élémentaires : le succès, que nous noterons S et l'échec que nous noterons \bar{S} . Comme σ -algèbre \mathcal{A} sur E , nous prendrons la famille constituée par les quatre sous-ensembles de E , à savoir :

- \emptyset évènement impossible
- S le succès
- \bar{S} l'échec
- $E = S + \bar{S}$ évènement certain

L'espace probabilisable (E, \mathcal{A}) peut représenter, par exemple, une partie jouée entre deux joueurs.

Les évènements élémentaires de E^{ω} sont les suites infinies $A_1, A_2, \dots, A_n, \dots$ où chaque A_i est S, \bar{S} . Les évènements simples de la forme (6) sont définis par la donnée du résultat des k premières parties (k quelconques), par exemple : $A_1 = S, A_2 = A_3 = \bar{S}, A_4 = S$ etc ... $A_k = \bar{S}$

Pour donner un exemple d'évènement composé, supposons que chacun des joueurs possède un capital initial (a francs pour le joueur n° 1, b francs pour son adversaire - a et b étant des entiers positifs) et qu'à l'issue de chaque partie le joueur n°1 reçoive un franc de son adversaire ou lui paye un franc selon que S ou \bar{S} s'est produit. Considérons la proposition "le joueur n°1 se ruine à la $n^{\text{ième}}$ partie" dont le sens exact est celui-ci : "le capital d'aucun des deux joueurs ne

s'est annulé à aucune des $n-1$ premières parties, et le capital du joueur n°1 s'est annulé à la $n^{\text{ième}}$ partie ". Cette proposition décrit un évènement constitué de la réunion d'un nombre fini d'évènements simples, choisis parmi les 2^n évènements simples de la forme :

$$A_1 A_2 \dots A_n \text{ E E E } \dots \quad (A_i = S \text{ ou } A_i = \bar{S})$$

On a d'ailleurs ici, obligatoirement, $A_n = \bar{S}$. Appelons R_n l'évènement "le joueur n°1 se ruine à la $n^{\text{ième}}$ partie " et R'_n l'évènement "le joueur n°2 se ruine à la $n^{\text{ième}}$ partie ". D'après les axiomes des σ -algèbres, la somme des R_n ou des R'_n est un évènement. Ainsi l'évènement

$$R = \sum_{n=1}^{\infty} R_n$$

est l'évènement "le joueur n°1 se ruine à une partie quelconque " ou "est perdant ", de même R' est l'évènement "le joueur N°2 est perdant " ou "le joueur N°1 est gagnant ".

La somme $R + R'$ est l'évènement "l'un des deux joueurs est gagnant " ou encore "le jeu s'arrête après un nombre fini de parties ". Le complémentaire de $R + R'$, qui est $\overline{R + R'} = \bar{R} \bar{R}'$ est l'évènement "aucun des deux joueurs ne gagne " ou encore "le jeu ne s'arrête jamais ". Ce dernier évènement n'a rien d'irréaliste, en particulier si l'un des joueurs, le n°2 par exemple, dispose d'un capital b illimité. Dans ce cas, R' est l'évènement impossible, et l'évènement $\bar{R} \bar{R}' = \bar{R}$ "le jeu ne s'arrête pas " se réduit à $\bar{R} = \bar{R}$ "le joueur n°1 ne perd pas ".

Définition : Probabilité, espace probabilisé.

Etant donné un espace probabilisable (E, \mathcal{A}) , c'est-à-dire un ensemble E et une σ -algèbre \mathcal{A} sur E , et une fonction d'ensemble $P(A)$ définie sur \mathcal{A} , c'est-à-dire une application de \mathcal{A} dans l'ensemble des nombres réels associant à tout évènement A de la σ -algèbre \mathcal{A} un nombre réel noté $P(A)$ - nous dirons que la fonction d'ensemble P est une probabilité si elle vérifie les trois axiomes suivants :

Axiomes des Probabilités :

- 1 - Pour tout évènement A de \mathcal{A} , on a $P(A) \geq 0$
- 2 - L'évènement certain a la probabilité unité : $P(E) = 1$.
- 3 - Si $A_1, A_2, \dots, A_n, \dots$ sont une suite d'évènements deux à deux incompatibles,

on a :

$$P(A_1 + A_2 + \dots + A_n + \dots) = \sum_{n=1}^{\infty} P(A_n)$$

On énonce parfois l'axiome 3 en disant que P est une fonction additive (ou complètement additive) d'ensemble. Dans le langage de la théorie de la mesure une probabilité se définit comme une mesure non négative, sommable et de somme unité. Dans un langage plus intuitif, la probabilité P peut être considérée comme décrivant la répartition d'une masse unité dans l'ensemble E des événements élémentaires, cette répartition étant telle que tout sous-ensemble A de E (qui soit un événement) porte, ou contienne la masse P(A).

L'ensemble (E, \mathcal{A}, P) d'un espace probabilisable (E, \mathcal{A}) et d'une probabilité P définie sur (E, \mathcal{A}) s'appelle espace probabilisé.

Remarques - Les trois axiomes des probabilités sont une généralisation évidente des trois propriétés démontrées au N°3 dans le cadre de la définition par les événements élémentaires équiprobables. Cependant :

- l'introduction de l'infini est rendu possible, du fait qu'il peut y avoir une infinité (dénombrable ou non) d'événements élémentaires, et que l'axiome 3 concerne une infinité (dénombrable) d'événements incompatibles.

- les événements élémentaires ne sont plus nécessairement équiprobables. Bien mieux, rien, dans les axiomes énoncés, ne s'oppose à ce que l'on définisse plusieurs probabilités distinctes sur un même espace probabilisable (E, \mathcal{A}) . Par exemple, dans l'exemple simple où E est constitué d'un succès S et d'un échec \bar{S} , la fonction P définie par :

$$\begin{aligned} P(\emptyset) &= 0 \\ P(S) &= p & 0 \leq p \leq 1 \\ P(\bar{S}) &= 1 - p \\ P(E) &= 1 \end{aligned}$$

est une probabilité (vérification immédiate des axiomes). Mais p peut être un nombre quelconque compris entre 0 et 1.

En ce sens, on peut dire que les 3 axiomes ne forment pas un système complet :

mais cette "incomplétude" est une nécessité. On rencontre effectivement de nombreux exemples où un même espace probabilisable est muni de probabilités différentes suivant les problèmes traités. Ainsi, dans l'exemple ci-dessus, $p = \frac{1}{2}$ pourra représenter une partie de pile ou face, et $p = \frac{1}{6}$ une partie de dé où le succès est défini par "amener le chiffre 6". Du point de vue abstrait, n'importe quelle fonction d'ensemble est une probabilité, pourvu qu'elle vérifie les axiomes. Dans les applications, naturellement, n'importe quelle probabilité ne conviendra pas à la description d'une épreuve ou d'un processus naturel donné. La probabilité convenable sera imposée par les propriétés objectives de l'épreuve ou du processus, et c'est en ce sens seulement que le calcul des probabilités permet, dans les applications, une description correcte des propriétés objectives des phénomènes naturels. A titre d'exercice, on pourra réfléchir au problème suivant : la proposition "il y a une chance sur dix pour que la planète Mars soit habitée" peut elle recevoir un sens, soit du point de vue de l'axiomatique pure, soit du point de vue de la méthodologie scientifique ?

Des trois axiomes des probabilités découlent les propriétés élémentaires suivantes :

Propriété 1- Quel que soit l'évènement A, on a

$$P(A) + P(\bar{A}) = 1$$

En effet, A et \bar{A} sont incompatibles, et l'on a (axiome 3) :

$$P(A + \bar{A}) = P(A) + P(\bar{A})$$

Mais $A + \bar{A} = E$ est l'évènement certain, et l'axiome 2 permet de conclure.

En particulier, l'évènement impossible $\emptyset = \bar{E}$ a une probabilité nulle, puisque $P(E) = 1$.

On notera que la réciproque n'est pas vraie ! un évènement de probabilité nulle n'est pas nécessairement l'évènement impossible. De même, un évènement de probabilité 1 n'est pas nécessairement l'évènement certain. Nous appellerons évènement presque impossible un évènement A tel que $P(A) = 0$ et évènement presque certain un évènement tel que $P(A) = 1$. La notion d'évènement presque impossible est

identique à celle d'ensemble de mesure nulle (pour la mesure P).

Propriété 2 - Quel que soit l'évènement A, on a

$$0 \leq P(A) \leq 1$$

C'est une conséquence immédiate de la propriété 1 et de l'axiome 1.

Propriété 3 - Si A est contenu dans B, on a

$$P(A) \leq P(B)$$

En effet, B est somme des deux évènements incompatibles A et $\bar{A} \cap B$.

Théorème 1 - Si A et B sont deux évènements quelconques, on a :

$$(7) \quad P(A + B) = P(A) + P(B) - P(AB)$$

En effet, les évènements A + B et B peuvent s'écrire comme somme d'évènements incompatibles :

$$\begin{aligned} A + B &= A + B \bar{A} \\ B &= AB + B \bar{A} \end{aligned}$$

L'axiome 3 entraîne alors :

$$\begin{aligned} P(A + B) &= P(A) + P(B \bar{A}) \\ P(B) &= P(AB) + P(B \bar{A}) \end{aligned}$$

Il suffit de soustraire membre à membre pour obtenir (7).

Conséquence 1 - Comme P(AB) est positive ou nulle (axiome 1), on déduit de la relation (7) l'inégalité suivante :

$$P(A + B) \leq P(A) + P(B)$$

qui se généralise, par une récurrence immédiate, à un nombre fini quelconque d'évènements.

$$P(A_1 + A_2 + \dots + A_n) \leq P(A_1) + P(A_2) + \dots + P(A_n)$$

Conséquence 2 - La formule (7) se généralise elle-même à un nombre fini quelconque d'évènements A_1, A_2, \dots, A_n . Posant :

$$S_1 = \sum_{i=1}^n P(A_i)$$

$$S_2 = \sum_{1 \leq i < j \leq n} P(A_i A_j)$$

$$S_3 = \sum_{1 \leq i < j < k \leq n} P(A_i A_j A_k)$$

$$S_n = P(A_1 A_2 \dots A_n)$$

on démontrera à titre d'exercice la formule :

$$(8) \quad P(A_1 + A_2 + \dots + A_n) = S_1 - S_2 + S_3 - \dots + (-1)^{n+1} S_n$$

(procéder par récurrence à partir de (7)).

Conséquence 3 - La formule (8) donne la probabilité pour que l'un au moins des n évènements $A_1, A_2 \dots A_n$ se produise. On peut aussi s'intéresser à la probabilité de l'évènement B défini comme suit " un et un seul des évènements $A_1, A_2 \dots A_n$ se produit ". On met B sous la forme d'une somme d'évènements incompatibles.

$$B = A_1 \bar{A}_2 \bar{A}_3 \dots \bar{A}_n + \bar{A}_1 A_2 \bar{A}_3 \dots \bar{A}_n + \dots + \bar{A}_1 \bar{A}_2 \dots \bar{A}_{n-1} A_n$$

L'axiome 3 montre qu'il suffit de calculer les probabilités des évènements du type $A_1 \bar{A}_2 \bar{A}_3 \dots \bar{A}_n$ et de les ajouter.

Pour calculer $P(A_1 \bar{A}_2 \bar{A}_3 \dots \bar{A}_n)$, on remarque que A_1 est somme de deux évènements incompatibles : " A_1 se produit seul " ou bien " A_1 se produit en même temps que l'un quelconque au moins des $A_2, A_3 \dots A_n$ ", soit

$$A_1 = A_1 \bar{A}_2 \bar{A}_3 \dots \bar{A}_n + A_1(A_2 + A_3 + \dots + A_n)$$

L'axiome 3 donne ainsi :

$$P(A_1 \bar{A}_2 \bar{A}_3 \dots \bar{A}_n) = P(A_1) - P(A_1 A_2 + A_1 A_3 + \dots + A_1 A_n)$$

Appliquons la formule (8)

$$P(A_1 \bar{A}_2 \dots \bar{A}_n) = P(A_1) - \sum_{2 \leq j \leq n} P(A_1 A_j) + \sum_{2 \leq j < k \leq n} P(A_1 A_j A_k) - \dots$$

Plus généralement, désignons par P_i la probabilité de $\bar{A}_1 \bar{A}_2 \dots \bar{A}_{i-1} A_i \bar{A}_{i+1} \dots \bar{A}_n$ (événement "A_i se produit seul"). On a :

$$P_i = P(A_i) - \sum_{j \neq i} P(A_i A_j) + \sum_{\substack{1 \leq j < k \leq n \\ j, k \neq i}} P(A_i A_j A_k) + \dots$$

D'où, enfin :

$$(9) \quad P(B) = \sum_{i=1}^n P_i = S_1 - 2 S_2 + 3 S_3 - \dots + (-1)^{n+1} S_n$$

Plus généralement, on démontrera, à titre d'exercice, que la probabilité pour que m exactement ($m \leq n$) des événements A_1, A_2, \dots, A_n se produise est donnée par :

$$P_m = S_m - C_{m+1}^m S_{m+1} + C_{m+2}^m S_{m+2} + \dots + (-1)^{n-m} C_n^m S_n$$

Théorème 2 (ou théorème de la continuité) - Si une suite d'évènements B_1, B_2, \dots, B_n est décroissante (c'est-à-dire si $B_1 \supset B_2 \supset \dots \supset B_n \supset \dots$) et si le produit $\prod_{n=1}^{\infty} B_n$ est un évènement presque impossible, alors $P(B_n)$ tend vers 0 lorsque n tend vers l'infini.

La démonstration de ce théorème fondamental est un excellent exemple de la manière dont l'idée d'infinité peut être traitée en calcul des probabilités.

Pour démontrer le théorème, écrivons B_n sous la forme d'une somme (infinie) d'évènements incompatibles :

$$(11) \quad B_n = \sum_{k=n}^{\infty} B_k \bar{B}_{k+1} + \prod_{k \gg n} B_k$$

La démonstration de cette égalité est immédiate : comme les B_n sont décroissants, un évènement élémentaire e de B_n peut soit appartenir à tous les B_k pour $k \gg n$, soit appartenir à un nombre fini seulement : dans cette dernière éventua-

lité, il existe $k \geq n$ tel que ω appartienne à B_k et n'appartienne pas à B_{k+1} . Ainsi le premier membre de l'équation (11) est contenu dans le deuxième. Mais, réciproquement, tout événement élémentaire appartenant au 2ème membre appartient soit à tous les B_k ($k \geq n$) soit à un nombre fini d'entre eux, donc à B_n . Ainsi le 2ème membre est contenu dans le premier, et (11) en résulte.

L'axiome 3 donne alors :

$$P(B_n) = \sum_{k=n}^{\infty} P(B_k \bar{B}_{k+1}) + P\left[\prod_{k \geq n} B_k\right]$$

Mais la probabilité du produit infini est nulle, car :

$$P\left[\prod_{k \geq n} B_k\right] \leq P\left(\prod_{k=1}^{\infty} B_k\right) = 0$$

L'inégalité résulte de la propriété 3, et la nullité de l'hypothèse que le produit des B_n est un événement presque impossible. Finalement, on a :

$$P(B_n) = \sum_{k=n}^{\infty} P(B_k \bar{B}_{k+1})$$

et, en particulier :

$$(12) \quad P(B_1) = \sum_{k=1}^{\infty} P(B_k \bar{B}_{k+1})$$

Alors $P(B_n)$ tend vers 0 lorsque n augmente indéfiniment, puisque $P(B_n)$ représente le reste de la série convergente (12).

Réciproque

Le théorème 2 admet une réciproque élémentaire : si une suite d'événements B_1, B_2, \dots est telle que $\lim_{n \rightarrow \infty} P(B_n) = 0$, alors le produit $\prod_{n=1}^{\infty} B_n$ est presque impossible.

En effet, quel que soit n on a $\prod_{n=1}^{\infty} B_n \subset B_n$, et par suite, d'après la propriété 3 :

$$P\left(\prod_{n=1}^{\infty} B_n\right) \leq P(B_n)$$

Comme $P(B_n)$ tend vers 0, la proposition est établie :

Théorème 3 - Soient (E_n, \mathcal{A}_n) une suite d'espaces probabilisables (E, \mathcal{A}) l'espace probabilisable produit, et P_n une suite de probabilités définies sur chacun des (E_n, \mathcal{A}_n) respectivement. Alors, il existe une probabilité P et une seule, définie sur le produit (E, \mathcal{A}) et telle que, pour tout événement simple de la forme

$$A = A_1 A_2 \dots A_k E_{k+1} E_{k+2} \dots$$

on ait

$$P(A) = P(A_1) P(A_2) \dots P(A_k)$$

Nous ne démontrerons pas ce théorème. Son intérêt provient de ce qu'il permet de définir la probabilité attachée à une suite infinie d'épreuves indépendantes, identiques ou non (pour la notion d'indépendance, voir le paragraphe suivant).

Exemple 1 : reprenons l'exemple simple de l'alternative répétée. L'espace E , composé des deux seuls événements élémentaires S et \bar{S} a été probabilisé plus haut par la donnée d'un nombre

$$p = P(S) \quad \text{avec } 0 < p < 1$$

L'espace produit E peut être probabilisé par application du théorème 3. On obtient ainsi la description de l'alternative indéfiniment répétée, avec indépendance des épreuves successives. L'événement simple

$$A = A_1 A_2 \dots A_n E_{n+1} E_{n+2} \dots$$

où k des A_i sont des succès S et $n-k$ des échecs \bar{S} possède alors, selon le théorème 3, la probabilité

$$(13) \quad P(A) = p^k (1-p)^{n-k}$$

Désignons par B_n l'événement simple $A_1 = A_2 = \dots = A_n = S$ (les n premières épreuves sont des succès). On a, d'après (13) :

$$P(B_n) = p^n$$

Soit $B = \prod_{n=1}^{\infty} B_n$ l'événement $S S S \dots$ (toutes les épreuves sont des succès) B est bien un événement d'après l'axiome 2 des σ -algèbres. Montrons que c'est un événement presque impossible.

En effet, $P(B_n) = p^n$ tend vers 0 lorsque n tend vers l'infini, et il suffit d'appliquer la réciproque du théorème 2. On a :

$$P(B) = 0$$

Ainsi l'évènement B_n "il n'y a que des succès" est un évènement presque impossible (de probabilité nulle). On notera que ce n'est pas l'évènement impossible.

De même, soit l'évènement C_k "au delà du rang k il n'y a que des succès". C_k est le produit des évènements simples de la forme $E_1 E_2 \dots E_k A_{k+1} \dots A_{k+r} \dots E_{k+r+1} \dots$ où $A_{k+1} = \dots = A_{k+r} = S$. On montre comme ci-dessus que C_k est presque impossible

$$P(C_k) = 0$$

Enfin, soit C l'évènement "il n'y a qu'un nombre fini d'échecs". C se décompose en évènements incompatibles :

$$C = C_0 + \bar{C}_0 C_1 + \bar{C}_1 C_2 + \dots = C_0 + \sum_{k=0}^{\infty} \bar{C}_k C_{k+1}$$

Comme $\bar{C}_k C_{k+1}$ est contenu dans C_{k+1} et $P(C_{k+1}) = 0$, on a aussi

$$P(\bar{C}_k C_{k+1}) = 0$$

et le troisième axiome donne :

$$P(C) = P(C_0) + \sum_{k=0}^{\infty} P(\bar{C}_k C_{k+1}) = 0$$

Ainsi, il est presque impossible qu'il n'y ait qu'un nombre fini d'échecs. De même, il est presque impossible qu'il n'y ait qu'un nombre fini de succès. Appelons D ce dernier évènement. De la conséquence 1 du théorème 1, on déduit :

$$P(C + D) \leq P(C) + P(D)$$

et, comme $P(C) = P(D) = 0$, on a $P(C+D) = 0$. La négation de $C + D$, qui est l'évènement $\bar{C} \bar{D}$: "il y a une infinité de succès et une infinité d'échecs" a donc une probabilité unité : c'est un évènement presque certain.

Exemple 2 - Risque de ruine

L'espace probabilisé est le même que dans le premier exemple. A chaque épreuve, le premier joueur reçoit de son adversaire, ou lui verse 1 franc selon qu'il y a succès ou échec. Les capitaux initiaux sont a et b (a, b entiers positifs). Nous avons défini plus haut l'évènement : "le premier joueur se ruine à la n^{ième} partie" comme étant l'évènement "les capitaux des deux joueurs ne s'annulent à l'issue d'aucune des n-1 premières parties, et le capital du 1er joueur s'annule à l'issue de la n^{ième} partie". Soit A_n cet évènement et P_n(a,b) = P(A_n) sa probabilité. Soit de même B_n l'évènement : "le 2ème joueur se ruine à la n^{ième} partie" et Q_n(a,b) sa probabilité.

Les évènements

$$\left\{ \begin{array}{l} A = \sum_{n=1}^{\infty} A_n \\ B = \sum_{n=1}^{\infty} B_n \end{array} \right.$$

qui signifient "le premier joueur perd" et "le 2ème joueur perd", sont écrits sous forme de sommes d'évènements incompatibles, et on a (axiome 3) :

$$(14) \quad \left\{ \begin{array}{l} P(a,b) = P(A) = \sum_{n=1}^{\infty} P_n(a,b) \\ Q(a,b) = P(B) = \sum_{n=1}^{\infty} Q_n(a,b) \end{array} \right.$$

Les probabilités de ruine P(a,b) et Q(a,b) des deux joueurs seront calculées à la fin du paragraphe suivant.

V.- PROBABILITE CONDITIONNELLE - EVENEMENTS INDEPENDANTS

Définition

Si (E, A, P) est un espace probabilisé, la probabilité conditionnelle d'un évènement A relativement à un évènement B (on dit aussi : dans l'hypothèse où l'évènement B s'est réalisé) est définie comme le rapport

$$(15) \quad P(A|B) = \frac{P(AB)}{P(B)} \quad (\text{avec } P(B) \neq 0)$$

des probabilités (on dit aussi : probabilité a priori, pour les distinguer des probabilités conditionnelles) des événements AB et B . Cette définition généralise exactement la relation (4) obtenue dans le cas des événements élémentaires équiprobables. Son interprétation est la même : (15) représente la restriction au sous-ensemble B de E de l'espace probabilisé (E, \mathcal{A}, P) . En effet, sur l'ensemble $B = EB$, la famille \mathcal{A}_B des sous-ensembles de B de la forme AB , A appartenant à \mathcal{A} , constitue une σ -algèbre, (vérification immédiate des axiomes) et $P(AB)$ est une fonction additive d'ensemble sur (B, \mathcal{A}_B) qui vérifie tous les axiomes des probabilités sauf l'axiome 2. Il suffit de renommer $P(AB)$ en la divisant par $P(B)$ pour que cet axiome soit à son tour vérifié. Ainsi $P(A|B)$ est bien une probabilité pour (B, \mathcal{A}_B) .

On notera bien que la probabilité conditionnelle $P(A|B)$ n'est définie que si l'évènement B n'est pas presque impossible ($P(B) \neq 0$). La relation (15) peut s'écrire aussi sous forme symétrique :

$$(16) \quad P(A \cap B) = P(B) P(A|B) = P(A) P(B|A)$$

où il n'est plus nécessaire de supposer $P(B)$ ou $P(A)$ différents de 0.

Énonçons ce résultat :

Théorème 4

La probabilité du produit de deux événements est égale à la probabilité a priori de l'un des deux multipliée par la probabilité conditionnelle de l'autre dans l'hypothèse où le premier s'est produit.

Formule de Bayes

Soit un événement B et n événements incompatibles A_1, A_2, \dots, A_n . Supposons que B soit contenu dans la somme $A_1 + \dots + A_n$. Dans ces conditions, la réalisation de B s'accompagne nécessairement de la réalisation d'un et d'un seul des A_i , et B se décompose en somme d'évènements incompatibles :

$$B = B \cap A_1 + B \cap A_2 + \dots + B \cap A_n$$

soit, d'après la formule (16) et l'axiome 3 :

$$(17) \quad P(B) = \sum_{i=1}^n P(A_i) P(B|A_i)$$

Proposons-nous de déterminer la probabilité conditionnelle de l'évènement A_1 , sachant que B s'est réalisé. On a :

$$P(A_1 | B) = \frac{P(A_1 B)}{P(B)} = \frac{P(A_1) P(B | A_1)}{P(B)}$$

Soit encore, compte tenu de (17)

$$(18) \quad P(A_1 | B) = \frac{P(A_1) P(B | A_1)}{\sum_{k=1}^n P(A_k) P(B | A_k)}$$

C'est la formule de Bayes, appelée aussi formule de la probabilité des causes (ou des hypothèses). On peut l'interpréter comme suit :

Soient $A_1 \dots A_n$ des hypothèses de travail (incompatibles) dont les probabilités a priori sont les $P(A_i)$ et qui sont les seules possibles pour expliquer la réalisation d'un évènement. Soit $P(B | A_i)$ la probabilité pour que l'évènement B se réalise dans l'hypothèse où A_i est vraie. Alors la formule de Bayes donne la probabilité a posteriori pour que l'hypothèse A_i soit la bonne, sachant que B s'est réalisé.

Remarque

Dans les applications pratiques, la formule de Bayes doit être maniée avec précaution. Il faut s'assurer que les probabilités conditionnelles $P(B | A_i)$, et surtout les probabilités a priori $P(A_i)$ ont une signification concrète réelle. Exemple de raisonnement sophistic : On sait, (évènement B) que le soleil s'est levé 1.800.000 matins consécutifs sans interruption, depuis que l'histoire de l'humanité repose sur des témoignages écrits. Adoptons deux hypothèses de travail ! A_1 : " le soleil se lèvera certainement " soit $P(B | A_1) = 1$, et A_2 : " il y a une chance sur mille pour que le soleil ne se lève pas ". soit $P(B | A_2) = \frac{999}{1000} \approx e^{-1800} \approx 10^{-781}$ et admettons que ces deux hypothèses soient également probables : $P(A_1) = P(A_2) = \frac{1}{2}$. La formule (18) donne :

$$P(A_1 | B) = \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{2} e^{-1800}} = \frac{1}{1 + e^{-1800}} = 1 - 10^{-781}$$

Une telle probabilité équivaut à une "quasi presque certitude". Par contre, si l'on prend comme hypothèse A_2 : " il y a une chance sur un million pour que le soleil ne se lève pas demain " ; on a $P(B | A_2) \approx e^{-1,8} \approx 0.165$ et, dans ce cas

$P(A_1 | B) \neq 0,75$. L'hypothèse reste plausible. Mais si, avec la même hypothèse A_2 , on prend $P(A_2) = \frac{999}{1000}$, alors il vient :

$$P(A_1 | B) = \frac{1/1000}{1/1000 + 0,165 \times \frac{999}{1000}} \neq \frac{6}{1000}$$

C'est alors l'hypothèse A_2 qui devient la plus plausible.

Définition : Evènements indépendants. Un évènement A est dit indépendant d'un évènement B si l'on a

$$(19) \quad P(A | B) = P(A)$$

c'est-à-dire si la probabilité conditionnelle de A relativement à B est égale à la probabilité a priori P(A). Si la relation (19) n'est pas vérifiée, les évènements A et B sont dépendants.

La signification concrète de la notion d'indépendance est très claire : dire que la probabilité de A n'est pas modifiée si l'on sait que B s'est réalisé signifie qu'aucune information sur l'évènement B ne peut nous renseigner quant aux possibilités de réalisation de A.

Conséquence 1 - Si A est indépendant de B, alors B est indépendant de A. En effet, on a

$$P(B | A) = \frac{P(AB)}{P(A)} = \frac{P(B) P(A | B)}{P(A)} = \frac{P(B) P(A)}{P(A)} = P(B)$$

l'indépendance est une notion réciproque.

Conséquence 2 - Si A et B sont indépendants, il en est de même de \bar{A} et B, A et \bar{B} et de \bar{A} et \bar{B} .

Montrons par exemple que \bar{A} et B sont indépendants. De

$$P(A | B) + P(\bar{A} | B) = 1$$

et $P(A | B) = P(A)$

on tire immédiatement :

$$P(\bar{A} | B) = 1 - P(A) = P(\bar{A})$$

Théorème 5 -

Si deux évènements A et B sont indépendants, on a

$$(20) \quad P(AB) = P(A) P(B)$$

et réciproquement, si (20) est vérifié A et B sont indépendants. Ce théorème, qui se déduit immédiatement des formules (16) et (19), constitue la propriété principale des évènements indépendants.

Généralisation. Pour généraliser la relation (20), nous dirons que n évènements $A_1, A_2 \dots A_n$ sont indépendants (ou mutuellement indépendants) si quel que soit l'entier $k < n$ et quel que soit le choix des indices $i_1, i_2 \dots i_k$ on a :

$$(21) \quad P(A_{i_1} A_{i_2} \dots A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \dots P(A_{i_k})$$

Pour qu'il en soit ainsi il faut et il suffit que chacun des A_i soit indépendant de chacun des produits $A_{j_1} A_{j_2} \dots A_{j_k}$ ($k < n$ quelconque, $j_1, j_2 \dots j_k$ différents de i). La démonstration est immédiate.

Remarque

Pour que n évènements $A_1 \dots A_n$ soient mutuellement indépendants, il ne suffit pas qu'ils soient indépendants 2 à 2, c'est-à-dire que chaque A_i soit indépendant de chaque A_j pour $i \neq j$.

Par exemple : soit un tétraèdre régulier dont les 4 faces portent : la première A un point rouge, la 2ème B un point bleu, la troisième C un point jaune, la 4ème O trois points, l'un rouge, l'autre bleu, le troisième jaune. Lançons le tétraèdre, et admettons qu'il y a une chance sur 4 pour qu'il repose sur l'une quelconque de ses faces.

Il y a, a priori, une chance sur deux pour que la face ainsi déterminée contienne un point rouge ($P(A) + P(O) = \frac{1}{2}$). Il y a une chance sur deux pour que cette face contienne un point rouge sachant qu'elle contient un point bleu ($\frac{P(O)}{P(A) + P(O)} = \frac{1}{2}$). Ainsi les trois évènements "rouge", "bleu", "jaune" sont deux à deux indépendants. Mais ils ne sont pas mutuellement indépendants, car :

$$P(\text{rouge et bleu et jaune}) = P(O) = \frac{1}{4} \neq P(\text{rouge})P(\text{bleu})P(\text{jaune}) = \frac{1}{8}$$

Dans la plupart des applications, cependant, l'indépendance mutuelle est ou bien postulée a priori (sur la base de considérations théoriques), ou bien considérée comme hautement plausible pourvu que l'indépendance deux à deux ait été vérifiée expérimentalement.

Exemple : Risque de ruine (suite de l'exemple 2 du paragraphe précédent). Un évènement simple

$$A = A_1 A_2 \dots A_k E_{k+1} \dots$$

a comme probabilité, selon le théorème 3 :

$$P(A) = P(A_1) P(A_2) \dots P(A_k)$$

ce qui signifie que les k évènements $A_1 E_2 E_3 \dots, E_1 A_2 E_3 \dots, \dots, E_1 E_2 \dots E_{k-1} A_k E_{k+1} \dots$ sont mutuellement indépendants. L'évènement A_n (le joueur n°1 se ruine à la n^{ième} partie) est somme des deux évènements incompatibles "le joueur 1 gagne la l^{ère} partie et se ruine à la n^{ième}" dont la probabilité est $p P_{n-1}(a+1, b-1)$ et "le joueur 1 perd la l^{ère} partie et se ruine à la n^{ième}" de probabilité $q P_{n-1}(a-1, b+1)$ (avec $q = 1 - p$). D'où

$$P_n(a, b) = p P_{n-1}(a+1, b-1) + q P_{n-1}(a-1, b+1)$$

Sommons en n (ce qui est légitime, d'après (14)). Il vient

$$(22) \quad P(a, b) = p P(a+1, b-1) + q P(a-1, b+1)$$

Considérée comme équation aux différences finies en a , (22) admet des solutions de la forme λ^a avec

$$p \lambda^2 - \lambda + q = 0$$

Les racines de cette équation caractéristiques sont 1 et q/p , et par suite la solution générale de (22) est de la forme

$$(23) \quad P(a, b) = A + B \left(\frac{q}{p}\right)^a$$

Les constantes A et B se déterminent par les conditions aux limites évidentes :

$$\begin{cases} P(0, a+b) = 1 \\ P(a+b, 0) = 0 \end{cases}$$

D'où la probabilité de ruine du 1^{er} joueur :

$$P(a,b) = \frac{1 - \left(\frac{p}{q}\right)^b}{1 - \left(\frac{p}{q}\right)^{a+b}}$$

et de même celle du deuxième joueur

$$Q(a,b) = \frac{1 - \left(\frac{q}{p}\right)^a}{1 - \left(\frac{q}{p}\right)^{a+b}}$$

On vérifie immédiatement que $P(a,b) + Q(a,b) = 1$. Ainsi, l'évènement "l'un des deux joueurs gagne" est presque certain, et son complémentaire "la partie se prolonge indéfiniment" est presque impossible.

Dans le cas particulier $p = q = \frac{1}{2}$, l'équation caractéristique admet $\lambda = 1$ comme racine double, et la solution générale est de la forme

$$(24) \quad P(a,b) = A + B a$$

Les mêmes conditions aux limites que ci-dessus donnent :

$$P(a,b) = 1 - \frac{a}{a+b} = \frac{b}{a+b}$$

et, de même

$$Q(a,b) = \frac{a}{a+b}$$

Ici encore $P + Q = 1$, et il est presque impossible que la partie se prolonge indéfiniment.

Passage à la limite

Examinons ce que deviennent les probabilités de ruine P et Q lorsque, a restant constant, b tend vers l'infini. Deux cas se présentent :

$$\begin{array}{lll} \text{Pour } q \geq p & P \rightarrow 1 & \text{et } Q \rightarrow 0 \\ \text{Pour } q < p & P \rightarrow \left(\frac{q}{p}\right)^a & \text{et } Q \rightarrow 1 - \left(\frac{q}{p}\right)^a \end{array}$$

Dans le cas $q \geq p$, l'interprétation du résultat ne prête à aucun malentendu. Pour $q < p$, on peut seulement énoncer : "jouant contre un adversaire extrêmement riche (b très grand, mais non infini) le joueur n°1 a la probabilité $\left(\frac{q}{p}\right)^a$ de gagner, et il est presque certain que la partie s'arrêtera". Ce résultat, ob-

tenu par passage à la limite, ne peut pas s'appliquer au cas où l'adversaire dispose d'un capital illimité (b infini) : dans ce cas, en effet, on a :

$$Q(a, \infty) = \sum_{n=1}^{\infty} Q_n(a, \infty) = 0$$

car chacun des Q_n est nul (l'adversaire ne peut pas être ruiné en un nombre fini de parties), alors que le passage à la limite donne

$$\lim_{b \rightarrow \infty} Q(a, b) = 1 - \left(\frac{q}{p}\right)^a \neq 0 \quad (p > q)$$

Calcul de $P(a, \infty)$ - Si le capital de l'adversaire est illimité, les mêmes raisonnements que ci-dessus conduisent encore à l'équation aux différences (22) dont la solution générale est de la forme (23), ou (24) pour $p = q = \frac{1}{2}$. Mais il n'y a plus qu'une seule condition aux limites :

$$P(0, \infty) = 1$$

Si $q \geq p$, la condition supplémentaire $P(a, \infty) \leq 1$ montre que l'on a

$$P(a, \infty) = 1 \quad (q \geq p)$$

La ruine est presque certaine, et la partie s'arrête presque certainement. Par contre, pour $q < p$ on obtient seulement

$$(25) \quad P(a, \infty) = A + (1 - A) \left(\frac{q}{p}\right)^a$$

avec $0 < A \leq 1$, mais le raisonnement ne donne pas la valeur de la constante A . On peut penser que $A = 0$, car, a tendant vers l'infini, c'est-à-dire si le joueur n°1 devient lui aussi infiniment riche, son risque de ruine doit tendre vers 0. Mais nous avons vu les dangers que présentent ces passages à la limite. En fait, il est exact que $A = 0$, nous le montrerons dans un instant à l'occasion d'une étude plus approfondie. A étant nul, on a :

$$\begin{cases} P(a, \infty) = \left(\frac{q}{p}\right)^a \\ Q(a, \infty) = 0 \end{cases} \quad (q < p)$$

Ainsi, pour $q < p$, il y a une probabilité non nulle $1 - \left(\frac{q}{p}\right)^a$ pour que la partie se prolonge indéfiniment.

Etude spéciale du cas $b = \infty$, $a = 1$. Pour déterminer la constante A (indépendante de Q) de l'équation (25), nous allons calculer directement $P(1, \infty)$ par une méthode dans laquelle on reconnaîtra, par la suite, un exemple particulier d'emploi de la fonction caractéristique. Pour abrégé, posons

$$P(1, \infty) = P$$

$$P_n(1, \infty) = P_n$$

et proposons nous de calculer P_n , probabilité de ruine à la $n^{\text{ième}}$ partie. On a nécessairement $P_n \leq 1$, de sorte que la fonction (dite fonction génératrice)

$$(26) \quad \bar{\Phi}(s) = \sum_{n=0}^{\infty} P_n s^n$$

est absolument convergente pour $s < 1$, et même pour $s = 1$ (car $P = \sum_{n=0}^{\infty} P_n$).

Par ailleurs on a

$$\left\{ \begin{array}{l} P_0 = 0 \\ P_1 = q \end{array} \right.$$

Pour $n \geq 1$, un raisonnement probabiliste simple (que le lecteur pourra expliciter à titre d'exercice) conduit à la relation de récurrence :

$$P_{n+1} = p \sum_{k=1}^n P_k P_{n-k}$$

Multipliant par s^{n+1} et sommant de $n = 1$ à l'infini, on trouve :

$$\bar{\Phi}(s) - q s = s p \bar{\Phi}(s)^2$$

d'où

$$\bar{\Phi}(s) = \frac{1 \pm \sqrt{1 - 4 p q s^2}}{2 p s}$$

Mais $\bar{\Phi}(s)$ est continue en $s = 0$ (puisque la série (26) est absolument convergente) et $\bar{\Phi}(0) = P_0 = 0$. Donc seule convient la solution :

$$(27) \quad \bar{\Phi}(s) = \frac{1 - \sqrt{1 - 4 p q s^2}}{2 p s}$$

Il suffit de faire $s = 1$ pour obtenir

$$P = \underline{\Phi}(1) = \frac{1 - \sqrt{1 - 4pq}}{2p} = \frac{1 - |p - q|}{2p} = \begin{cases} 1 & \text{pour } q \geq p \\ \frac{q}{p} & \text{pour } q < p \end{cases}$$

On a ainsi montré que la constante A de l'équation (25) est bien nulle.

En fait, on peut aller plus loin. Du développement en série

$$(1 - 4pq s^2)^{\frac{1}{2}} = 1 - 2 \sum_{n=1}^{\infty} \frac{1}{2n-1} C_{2n-1}^n (pq)^n s^{2n}$$

on tire, en portant dans (27)

$$\underline{\Phi}(s) = \sum_{n=1}^{\infty} \frac{1}{2n-1} C_{2n-1}^n p^{n-1} q^n s^{2n-1}$$

D'où, en identifiant avec (26)

$$(28) \quad \begin{cases} P_{2n} = 0 \\ P_{2n+1} = \frac{1}{2n+1} C_{2n+1}^n p^n q^{n+1} = \frac{1}{n+1} C_{2n}^n p^n q^{n+1} \end{cases}$$

On a ainsi obtenu l'expression de la probabilité de ruine à la $n^{\text{ième}}$ partie (avec $a = 1$ et $b = \infty$)

Remarque : probabilité d'absorption

On peut schématiser un mouvement brownien en imaginant qu'une particule, en mouvement sur l'axe des x , accomplit, aux temps $t = 1, 2 \dots n \dots$ soit un déplacement de $+1$ (avec une probabilité p), soit un déplacement -1 (probabilité $q = 1-p$) C'est le problème dit de la promenade aléatoire. A la notion de ruine répond celle d'absorption (la particule étant placée en $t = 0$ au point d'abscisse a , deux barrières absorbantes sont placées en $x = 0$ et $x = a+b$, ou, pour b infini, une seule barrière absorbante est placée en $x = 0$). Ainsi, dans le problème à une seule barrière absorbante et pour $p > q$, il y a une probabilité non nulle $1 - \left(\frac{q}{p}\right)^a$ pour que la particule ne soit jamais absorbée.

CHAPITRE II

VARIABLES ALEATOIRES ET FONCTION DE REPARTITION

I. - NOTION DE VARIABLE ALEATOIRE

En termes intuitifs, une variable aléatoire X , attachée à une épreuve ou expérience définie, est une grandeur susceptible de prendre l'une ou l'autre de plusieurs valeurs numériques selon le résultat de l'épreuve ou de l'expérience. Par exemple, si l'expérience consiste à mesurer une certaine grandeur physique à l'aide d'un appareillage donné et selon une procédure définie, le résultat de l'expérience peut être défini comme l'ensemble de toutes les circonstances ayant accompagné la réalisation de la mesure, y compris les erreurs de lecture, les variations fortuites de température et pression, etc ... Chacun des facteurs de ce résultat exerce une certaine influence sur la mesure, et la valeur numérique X de cette mesure est fonction de tous ces facteurs. Il est clair qu'une même valeur numérique X de la mesure peut être fournie par plusieurs, ou même une infinité de constellations possibles de facteurs du résultat, tandis qu'à chaque constellation correspond une valeur unique, bien définie de X . Dans la terminologie probabiliste, ces "résultats" ou ces "constellations de facteurs du résultat" portent le nom d'évènements élémentaires. On voit qu'une variable aléatoire X est, en réalité, une fonction numérique $X(e)$ définie sur l'ensemble E des évènements élémentaires e .

Cette fonction $X(e)$ ne peut pas être absolument quelconque. En effet, il est nécessaire, x , a et b étant des nombres quelconques, que des propositions du type: "l'expérience a donné le résultat numérique $X = x$ " ou "l'expérience a donné le résultat $a \leq X < b$ " constituent des évènements. Mais nous avons vu que, parmi les sous-ensembles de l'espace E des évènements élémentaires e , seuls les sous-ensembles privilégiés appartenant à une σ -algèbre \mathcal{A} constituaient des évènements. Ainsi donc, la fonction $X(e)$ doit être telle que l'ensemble :

$\left\{ e : a \leq X(e) < b \right\}$ des évènements élémentaires e de E , pour lesquels les inégalités $a \leq X(e) < b$ sont vérifiées, appartienne à la σ -algèbre \mathcal{A} . L'image inverse par X de tout intervalle doit appartenir à \mathcal{A} . Par ailleurs, si la

proposition " X appartient à l'intervalle (a,b) " définit un évènement, il doit en être de même de toute proposition obtenue en appliquant à des propositions de ce type une suite illimitée d'opérations logiques. Nous avons appelé \mathcal{B} la σ -algèbre sur l'ensemble \mathbb{R} des nombres réels engendrés par les intervalles (a,b), ou famille des ensembles boréliens. Alors, si B est un ensemble borélien quelconque, la proposition " X appartient à B " doit définir un évènement. Autrement dit, la fonction numérique $X(e)$ doit être telle que l'image inverse de tout ensemble borélien appartienne à la σ -algèbre \mathcal{A} . Une telle fonction est dite mesurable :

Définition : Fonction mesurable. Une fonction numérique $X(e)$ définie sur un espace probabilisable (E, \mathcal{A}) est dite mesurable si l'image inverse $X^{-1}(B)$ de tout ensemble borélien B est un évènement, c'est-à-dire si :

$$\forall B \in \mathcal{B} \text{ on a : } X^{-1}(B) = \{ e : X(e) \in B \} \in \mathcal{A}$$

Nous sommes maintenant à même de donner la définition générale d'une variable aléatoire.

Définition : Variable aléatoire. Etant donné un espace probabilisable (E, \mathcal{A}) , on appelle variable aléatoire (sous entendu : sur (E, \mathcal{A})) une fonction numérique $X(e)$ mesurable définie sur (E, \mathcal{A}) .

Si (E, \mathcal{A}, P) est un espace probabilisé et $X(e)$ une variable aléatoire sur (E, \mathcal{A}) , on voit que l'évènement "X(e) appartient à B " où B est un ensemble borélien, possède une probabilité bien définie :

$$P[X(e) \in B] = P[X^{-1}(B)]$$

Posons :

$$(1) \quad P'(B) = P[X^{-1}(B)]$$

Nous définissons ainsi une fonction d'ensemble P' qui, à tout ensemble borélien B , fait correspondre un nombre $P'(B)$ compris entre 0 et 1. Cette fonction d'ensemble $P'(B)$ est elle une probabilité pour l'espace probabilisable $(\mathbb{R}, \mathcal{B})$ constitué par l'ensemble des nombres réels \mathbb{R} et la σ -algèbre de Borel \mathcal{B} ? De (1) résulte immédiatement $P'(B) \geq 0$, $P'(\mathbb{R}) = P(E) = 1$, de sorte que les axiomes 1 et 2 des probabilités sont vérifiés. Montrons qu'il en est de même de l'axiome 3.

Soit B_1, B_2, \dots des ensembles boréliens disjoints ($B_i B_j = \emptyset$ pour $i \neq j$). On a :

$$X^{-1}(B_1 + \dots + B_n + \dots) = \sum_{n=1}^{\infty} X^{-1}(B_n)$$

En effet, la proposition : "l'évènement élémentaire e appartient à $X^{-1}(\sum B_n)$ " équivaut par définition à : " $X(e)$ appartient à $\sum B_n$ " qui équivaut à : " $X(e)$ appartient à l'un au moins des B_n ", qui équivaut, à son tour, à : " e appartient à l'un au moins des $X^{-1}(B_n)$ ", c'est-à-dire : " e appartient à $\sum X^{-1}(B_n)$ ".

Dans ces conditions, on peut écrire :

$$\begin{aligned} P'(\sum B_n) &= P[X^{-1}(\sum B_n)] \\ &= P[\sum X^{-1}(B_n)] \end{aligned}$$

et, comme P est une probabilité et que les $X^{-1}(B_n)$ appartiennent à \mathcal{A} , on a :

$$P'(\sum B_n) = \sum P[X^{-1}(B_n)] = \sum P'(B_n)$$

L'axiome 3 est vérifié, et P' est bien une probabilité sur (R, \mathcal{B}) . Tant qu'on se limite à l'étude de la variable aléatoire $X(e)$, pour laquelle les seuls évènements observables sont du type " X appartient à l'ensemble borélien B ", il n'y a aucun intérêt à conserver l'espace probabilisé initial (E, \mathcal{A}, P) . Toutes les propriétés de la variable aléatoire X peuvent s'obtenir directement à partir de l'espace probabilisé (R, \mathcal{B}, P') , dans lequel l'étude de X est en général plus facile. Le passage de (E, \mathcal{A}, P) à (R, \mathcal{B}, P') est un changement d'espace probabilisé. Mais on prendra garde que si $Y(e)$ est une deuxième variable aléatoire différente de $X(e)$, il ne sera pas possible, en général, d'étudier ses propriétés dans l'espace (R, \mathcal{B}, P') adapté à la seule variable $X(e)$. Il ne sera cependant pas nécessaire, en général, de travailler directement dans (E, \mathcal{A}, P) . Pour le montrer, nous allons généraliser les notions introduites ci-dessus.

Généralisation

Etant donnés deux espaces probabilisables (E_1, \mathcal{A}_1) et (E_2, \mathcal{A}_2) , une application α de E_1 dans E_2 (c'est-à-dire une correspondance qui, à tout évènement élémentaire e_1 de E_1 fait correspondre un évènement élémentaire e_2 de E_2 noté $\alpha(e_1)$, est dite mesurable si l'image inverse de tout évènement de (E_2, \mathcal{A}_2) est un évènement de (E_1, \mathcal{A}_1) , autrement dit si :

$$\forall A_2 \in \mathcal{A}_2, \text{ on a : } \alpha^{-1}(A_2) = \{ e_1 : \alpha(e_1) \in A_2 \} \in \mathcal{A}_1$$

Une application mesurable α de (E_1, \mathcal{A}_1) dans (E_2, \mathcal{A}_2) sera aussi appelée variable aléatoire (sous entendue : définie sur E_1 et à valeur dans E_2).

Supposons maintenant l'espace (E_1, \mathcal{A}_1) probabilisé par la donnée d'une probabilité P , et soit α une variable aléatoire, application mesurable de (E_1, \mathcal{A}_1) dans (E_2, \mathcal{A}_2) . A tout évènement A_2 de \mathcal{A}_2 , nous pouvons faire correspondre le nombre

$$P'(A_2) = P[\alpha^{-1}(A_2)]$$

On démontre alors, exactement comme ci-dessus, que la fonction d'ensemble P' vérifie les 3 axiomes des probabilités, et constitue donc une probabilité pour (E_2, \mathcal{A}_2) . Dans ces conditions, toutes les propriétés observables de la variable aléatoire α (c'est-à-dire : toutes les propriétés liées aux probabilités d'évènements du type : " α appartient à A_2 " A_2 étant un évènement de \mathcal{A}_2) pourront s'étudier directement sur l'espace probabilisé transformé (E_2, \mathcal{A}_2, P')

Définition. : Variable aléatoire à n composantes. On peut, en particulier, prendre comme espace probabilisable l'espace (R^n, \mathcal{B}^n) , où R^n est l'espace euclidien à n dimensions et \mathcal{B}^n la σ -algèbre engendrée par les pavés de R^n , (ensembles de points $x = (x_1 \dots x_n)$ de R^n vérifiant des inégalités du type $a_i \leq x_i < b_i, i = 1, 2, \dots, n$) ou famille des ensembles boréliens de R^n .

Etant donné un espace probabilisable (E, \mathcal{A}) , on appellera variable aléatoire vectorielle, ou variable aléatoire à n composantes une application mesurable X de (E, \mathcal{A}) dans (R^n, \mathcal{B}^n) .

A tout évènement élémentaire e de E , X fait correspondre un point $X(e)$

de \mathbb{R}^n . Soient $X_1(e), \dots, X_n(e)$ les coordonnées de $X(e)$. On vérifie immédiatement que chacune des coordonnées $X_i(e)$ définit une application mesurable de (E, \mathcal{A}) dans $(\mathbb{R}, \mathcal{B})$, c'est-à-dire une variable aléatoire à une seule composante. Ainsi la variable à n composantes $X(e)$ représente l'ensemble des n variables à une composante $X_1(e) \dots X_n(e)$. Dans la suite, une variable aléatoire à une composante sera, en général, appelée simplement variable aléatoire.

Réciproquement, soient $X_1 \dots X_n$ n variables aléatoires (à une composante), c'est-à-dire n applications mesurables de (E, \mathcal{A}) dans $(\mathbb{R}, \mathcal{B})$. Leur donnée simultanée définit l'application $X(e) = [X_1(e), \dots, X_n(e)]$ de E dans \mathbb{R}^n . $X(e)$ est-elle une variable aléatoire vectorielle ? Pour le montrer, il faut vérifier que X est une application mesurable de (E, \mathcal{A}) dans $(\mathbb{R}^n, \mathcal{B}^n)$. Or cela résulte du théorème suivant :

Théorème 1 -

Soient (E, \mathcal{A}) et (E', \mathcal{A}') deux espaces probabilisables, \mathcal{A}' étant la σ -algèbre engendrée par une famille \mathcal{F}' de sous-ensembles de E' , $\mathcal{A}' = \sigma(\mathcal{F}')$. Pour qu'une application α de (E, \mathcal{A}) dans (E', \mathcal{A}') soit mesurable, il faut et il suffit que l'image inverse $\alpha^{-1}(F')$ de tout ensemble F' de \mathcal{F}' soit un évènement de \mathcal{A} .

Pour établir ce théorème, on considère la famille \mathcal{A}'' des sous-ensembles A'' de E' tels que $\alpha^{-1}(A'')$ appartienne à \mathcal{A} . On voit facilement que \mathcal{A}'' vérifie les trois axiomes de σ -algèbres. Comme \mathcal{A}'' contient \mathcal{F}' , elle contient aussi la σ -algèbre engendrée par \mathcal{F}' , c'est-à-dire \mathcal{A}' . Puisque tout évènement A' de \mathcal{A}' appartient à \mathcal{A}'' , on a bien : $\alpha^{-1}(A')$ appartient à \mathcal{A} .

Conséquence.

Etant données n variables aléatoires $X_i(e)$, l'application $X(e) = (X_1(e), \dots, X_n(e))$ est une variable aléatoire vectorielle.

Comme la famille borélienne \mathcal{B}^n est la σ -algèbre engendrée par les pavés $(a_i \leq x_i < b_i, i = 1, 2, \dots, n)$, il faut vérifier que l'image inverse d'un pavé est un évènement de \mathcal{A} .

Considérons l'ensemble A_i des évènements élémentaires vérifiant

$$a_i \leq X_i(e) < b_i$$

Comme $X_i(e)$ est mesurable, A_i est un évènement. L'image inverse du pavé est l'ensemble des e qui appartiennent à la fois à A_1, A_2, \dots, A_n , c'est-à-dire le produit $A_1 A_2 \dots A_n$: c'est donc un évènement de \mathcal{A} .

Généralisation : Fonction aléatoire et processus stochastiques.

On peut également définir des variables aléatoires à une infinité (dénombrable, ou même non dénombrable) de composantes. Par exemple, considérons l'ensemble :

$$E' = \mathbb{R}^{\mathbb{N}}$$

de toutes les fonctions numériques définies sur l'espace euclidien \mathbb{R}^n . E' étant munie d'une σ -algèbre \mathcal{A}' convenable (inutile de dire que la définition d'une σ -algèbre sur un espace de fonctions n'est pas du tout un problème élémentaire), une application mesurable f de (E, \mathcal{A}) dans (E', \mathcal{A}') définira une fonction aléatoire : à tout évènement élémentaire e de E correspond un élément $f(e)$ de E' , c'est-à-dire une fonction numérique $f(x_1, \dots, x_n; e)$ appelée réalisation de la fonction aléatoire f . En un point $x = (x_1, \dots, x_n)$ fixé de \mathbb{R}^n , $f(x; e)$ est une fonction numérique définie sur E : si elle est mesurable (et la σ -algèbre \mathcal{A}' est choisie de manière à ce qu'il en soit ainsi) $f(x; e)$ est donc, à x fixé, une variable aléatoire ordinaire. De même, l'ensemble des valeurs $f(x; e), f(x^1; e) \dots f(x^{(k)}; e)$ des valeurs prises par la fonction aléatoire en k points d'appui fixés $x, x^1 \dots x^{(k)}$ constituent, e variant, une application (mesurable) de E dans \mathbb{R}^k . c'est-à-dire une variable aléatoire à k composantes.

Lorsque l'on prend $E' = \mathbb{R}^{\mathbb{T}}$, où $\mathbb{T} = \mathbb{R}$ est l'ensemble des nombres réels t considérés comme représentant un temps, une fonction aléatoire f s'appelle aussi : processus stochastique. A e fixé, $f(t; e)$, réalisation de f , représente une évolution possible du processus dans le temps. A e variable, $t_1 \dots t_k$ fixés, $f(t_1; e), f(t_2; e) \dots f(t_k; e)$ constituent une variable aléatoire vectorielle à k composantes.

II.- FONCTIONS DE REPARTITION A UNE SEULE VARIABLE.

Soit X une variable aléatoire, c'est-à-dire une application mesurable d'un espace probabilisé (E, \mathcal{A}, P_1) dans $(\mathbb{R}, \mathcal{B})$. On a vu au paragraphe précédent que la fonction définie par

$$P(\mathcal{B}) = P_1 \left[X^{-1}(\mathcal{B}) \right]$$

pour tout ensemble borélien B est une probabilité, et que toutes les propriétés de X peuvent s'étudier dans l'espace probabilisé (R, \mathcal{B}, P) . En vue de simplifier au maximum l'étude de la variable aléatoire X , il est souhaitable de remplacer la fonction d'ensemble $P(B)$ par une fonction numérique ordinaire. Etant donné un nombre b , considérons l'ensemble borélien des nombres x tels que : " $x < b$ ".

Soit $F(b)$ sa probabilité :

$$F(b) = P(x < b)$$

La fonction F ainsi définie s'appelle fonction de répartition de la variable aléatoire X . Par définition, on a :

$$(2) \quad F(x) = P(X < x)$$

La donnée de la fonction de répartition $F(x)$ suffit-elle pour calculer la probabilité de n'importe quel ensemble borélien B , et par suite constitue-t-elle un résumé exhaustif de tout ce qu'il est possible de savoir sur la variable aléatoire X (considérée seule) ? On démontre qu'il en est bien ainsi. Connaissant $F(x)$, on a, en premier lieu :

$$P(X \geq a) = 1 - F(a)$$

et, par suite

$$P(a \leq X < b) = F(b) - F(a) \quad (a \leq b)$$

car l'évènement " $X < b$ " est somme des deux évènements : " $a \leq X < b$ " et " $X < a$ " manifestement incompatibles. La σ -algèbre borélienne \mathcal{B} étant engendrée par les intervalles $(a \leq X < b)$, il est alors possible de démontrer que $P(B)$ peut effectivement se reconstituer à partir de $F(x)$.

Etablissons quelques propriétés fondamentales de $F(x)$.

Propriété 1 - $F(x)$ est une fonction non décroissante de x .

Soient a et b deux nombres avec $a \leq b$. Alors l'évènement " $X < a$ " est contenu dans l'évènement " $X < b$ ". D'après la propriété 3 (ch.I, parag.4) et la définition (2), on a bien $F(a) \leq F(b)$.

Propriété 2 - $F(x)$ tend vers 0 quand x tend vers $-\infty$, et vers +1 quand x tend vers $+\infty$.

Les évènements $B_n : "x < -n"$ forment une suite décroissante. Le produit $\prod_{n=1}^{\infty} B_n$ est l'évènement impossible (car aucun nombre réel x ne vérifie $x < -n$ quel que soit n). D'après le théorème de la continuité (Th.2, parag.4, Ch.I), $P(B_n) = F(-n)$ tend vers 0 pour $n \rightarrow \infty$.

Soit alors $\varepsilon > 0$. Il existe alors N tel que : pour $n \geq N$, $F(-n) \leq \varepsilon$. En particulier $F(-N) \leq \varepsilon$. Comme F est non décroissante, $x \leq -N$ entraîne $F(x) \leq \varepsilon$.

On montre de même que la probabilité $1 - F(x)$ de l'évènement " $X \geq x$ " tend vers 0 quand $x \rightarrow \infty$, d'où résulte que $F(x)$ tend vers +1. Nous avons donc :

$$(3) \quad \left\{ \begin{array}{l} F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0 \\ F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1 \end{array} \right.$$

Propriété 3 - $F(x)$ est continue à gauche, c'est-à-dire : en tout point x , $F(x-h)$ tend vers $F(x)$ lorsque h tend vers 0 en restant positif.

Soit h_n une suite décroissante de nombres positifs tendant vers 0, et A_n l'évènement " $x - h_n \leq X < x$ ". Les A_n constituent une suite décroissante d'évènements dont le produit est impossible. D'après le théorème de la continuité :

$$P(A_n) = F(x) - F(x - h_n) \rightarrow 0 \quad \text{pour } n \rightarrow \infty$$

Alors, quel que soit $\varepsilon > 0$, il existe un n tel que $F(x) - F(x - h_n) \leq \varepsilon$

Comme F est non décroissante, pour $h \leq h_n$ on a $F(x) - F(x - h) \leq \varepsilon$.

Inversement, ces trois propriétés caractérisent les fonctions de répartition : si $F(x)$ vérifie ces trois propriétés, on prendra $P(x < b) = F(b)$ et il sera possible de reconstituer une probabilité P sur (R, \mathcal{B}) . Pour (R, \mathcal{B}, P) la variable aléatoire définie par l'application identique $X(x) = x$ admettra $F(x)$ comme fonction de répartition.

Propriété 4. - $F(x)$ est continue partout sauf au plus en une infinité dénombrable de points de discontinuité.

C'est une propriété classique des fonctions non décroissantes bornées. En tout point x , on pose $F(x+0) = \lim_{h \rightarrow 0} F(x+h)$ lorsque h tend vers 0 en restant positif. (la limite existe, puisque F est non décroissante). Si $F(x+0) = F(x)$ la fonction F est continue en x . Soit D_n l'ensemble des points de discontinuité pour lesquels le saut $F(x+0) - F(x)$ vérifie :

$$\frac{1}{n} \geq F(x+0) - F(x) > \frac{1}{n+1}$$

D_n contient au plus n points (puisque $F(+\infty) - F(-\infty) = 1$), d'où possibilité de dénombrement.

Conséquence

Une fonction de répartition est entièrement définie lorsque l'on connaît sa valeur en tout point de continuité.

En effet, tout intervalle non réduit à un point contient un ensemble de points ayant la puissance du continu et, par suite, comme les points de discontinuité sont dénombrables, une infinité de points de continuité.

Si x_0 est un point quelconque, chacun des intervalles $[x_0 - \frac{1}{n}, x_0]$ contient au moins un point de continuité x_n . Pour n tendant vers l'infini, $x_n \rightarrow x_0$. Comme $F(x)$ est continue à gauche, on a aussi : $F(x_n) \rightarrow F(x_0)$

Cas particulier : Répartitions absolument continues, et discrètes.

1 - Une fonction de répartition $F(x)$ est dite absolument continue s'il existe une fonction intégrable $f(x)$, appelée densité de probabilité, telle que l'on ait, en tout point x :

4)
$$F(x) = \int_{-\infty}^x f(y) dy$$

Comme exemples élémentaires, citons la loi normale ($f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$) et les lois gamma : $f(x) = \frac{b^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-bx}$ (pour $x \geq 0$, et $f(x) = 0$ pour $x < 0$), avec $b > 0$, $\alpha > 0$, dont les lois dites du χ^2 sont un cas particulier. Sur ce dernier exemple (pour $0 < \alpha < 1$), on voit qu'il n'est pas nécessaire que $F(x)$ soit dé-

rivable en tout point. Toute fonction $f(x)$ vérifiant :

$$f(x) \geq 0 \quad \text{pour tout } x$$
$$\int_{-\infty}^{+\infty} f(x) dx = +1$$

définit une densité de probabilité.

2/ - Si une variable aléatoire X ne peut prendre qu'un nombre fini, ou une infinité dénombrable de valeurs numériques x_1, x_2, \dots, x_n (c'est-à-dire : si l'ensemble des valeurs prises par la fonction mesurable $X(e)$ lorsque e parcourt E est un ensemble dénombrable) ou, plus généralement, s'il est presque certain que la valeur de $X(e)$ est égale à x_1 , ou x_2, \dots sa fonction de répartition $F(x)$ est dite discrète.

Si B est un ensemble borélien ne contenant aucun des points x_i , l'évènement " X appartient à B " est presque impossible, et $P(B) = 0$. Chacun des évènements, " $X = x_i$ " possède une probabilité p_i et, d'après les axiomes, on a

$$P(E) = \sum_{i=1}^{\infty} p_i = 1$$

La fonction de répartition $F(x)$ est une fonction de saut. En effet, soient x_{i_1}, x_{i_2}, \dots les x_i vérifiant $x_{i_k} < x$, on a :

$$F(x) = \sum_{k=1}^{\infty} p_{i_k}$$

Par suite, si x_j et x_k sont tels qu'il n'y ait aucun point x_i strictement compris entre x_j et x_k , on a (avec $x_j < x_k$) :

$$F(y) = F(x_k) = C^{\text{te}} \quad \text{pour } x_j < y \leq x_k$$

$F(x)$ est constante sur les intervalles du type (x_j, x_k) et subit en chaque x_i la discontinuité (ou saut) à droite $p_i = F(x_i + 0) - F(x_i)$.

Si l'ensemble des x_i contient des points d'accumulation, l'aspect de $F(x)$ est moins intuitif. En particulier, si les x_i constituent un ensemble dénombrable

partout dense (comme l'ensemble des nombres rationnels) tout intervalle non réduit à un point contient, si petit soit-il, une infinité de points de discontinuité. Exemples simples de répartitions discrètes : loi de Poisson, loi binomiale, etc ...

3/ Plus généralement, on démontre que toute fonction de répartition $F(x)$ est de la forme :

$$F(x) = F_1(x) + F_2(x) + F_3(x)$$

$F_1(x)$ et $F_2(x)$ étant (à un facteur constant près) deux fonctions de répartition l'une discrète, l'autre absolument continue. $F_3(x)$, ou composante singulière, est continue mais non absolument continue, et constante sauf sur un ensemble de mesure nulle. Dans les applications usuelles, cette composante est absente, et F se réduit à ses deux composantes discrète et absolument continue.

III.- FONCTIONS DE REPARTITION DE PLUSIEURS VARIABLES.

Les définitions et propriétés énoncées au paragraphe précédent se généralisent facilement au cas d'une variable aléatoire $X = (X_1, \dots, X_n)$ à n composantes. Comme la σ -algèbre \mathcal{B}^n de \mathbb{R}^n est engendrée par les ensembles de points $x = (x_1 \dots x_n)$ vérifiant $x_i < a_i$ ($i = 1, 2, \dots, n$), on appellera fonction de répartition de X la fonction $F(x) = F(x_1, \dots, x_n)$ définie par :

$$(5) \quad F(x) = F(x_1, x_2, \dots, x_n) = P(X_1 < x_1 \text{ et } X_2 < x_2 \text{ et } \dots \text{ et } X_n < x_n)$$

et cette fonction $F(x)$ suffit à déterminer la probabilité pour que X appartienne à un ensemble borélien B , quelconque de \mathcal{B}^n . En particulier, si B est le pavé $(a_i \leq x_i < b_i, i = 1, 2, \dots, n)$, on établira la formule :

$$(6) \quad P(B) = F(b_1, \dots, b_n) - \sum_{i=1}^n p_i + \sum_{i < j} p_{ij} - \dots + (-1)^n F(a_1, a_2, \dots, a_n)$$

où l'on représente par la notation $p_{ij \dots k}$ la valeur de $F(x_1 \dots x_n)$ au point défini par : $x_l = a_l$ si l'indice l est égal à $i, j \dots$ ou k et $x_l = b_l$ dans le cas contraire.

Répartitions marginales

Si X_1, \dots, X_n sont les composantes d'une variable aléatoire vectorielle, il en est de même de X_1, \dots, X_{n-1} (car $n-1$ fonctions mesurables définissent, nous l'avons vu, une variable aléatoire à $n-1$ composant^s). Cette nouvelle variable possède la fonction de répartition (dite marginale)

$$F_{1,2,\dots,n-1}(x_1, \dots, x_{n-1}) = F(x_1, x_2, \dots, x_{n-1}, +\infty)$$

En termes intuitifs, F représente la répartition d'une masse unité dans R^n , et $F_{1,2,\dots,n-1}$ la projection, parallèlement à l'axe des x_n , de cette répartition dans l'hyperplan des x_1, x_2, \dots, x_{n-1} . Plus généralement, la fonction de répartition F_{i_1, \dots, i_k} des k variables X_{i_1}, \dots, X_{i_k} ($k < n$) s'obtient en faisant $x_i = +\infty$ pour tout i distinct de i_1, \dots, i_k dans l'expression de $F(x_1, \dots, x_n)$.

Les propriétés 1, 2, et 3 du paragraphe précédent deviennent les suivantes :

Propriété 1 - $F(x_1, \dots, x_n)$ est fonction non décroissante de chacun de ses arguments.

Propriété 2 - F est continue à gauche pour chacun de ses arguments.

Propriété 3 - On a $F(+\infty, \dots, +\infty) = 1$ et, pour chaque indice k :

$$\lim_{x_k \rightarrow -\infty} F(x_1, x_2, \dots, x_n) = 0$$

Mais, à la différence des fonctions de répartition d'une seule variable, une fonction de répartition n'est pas caractérisée par ces trois propriétés : il faut leur adjoindre la suivante :

Propriété 4 - Quels que soient les a_i et les b_i ($a_i \leq b_i$) l'expression (6) est non-négative.

L'exemple simple suivant montre que cette condition supplémentaire n'est pas superflue. Soit

$$\begin{aligned} F(x,y) &= 0 && \text{pour } x \leq 0, \text{ ou } y \leq 0, \text{ ou } x + y \leq 1 \\ F(x,y) &= 1 && \text{dans le reste du plan.} \end{aligned}$$

et le carré ($\frac{1}{2} \leq x < 1, \frac{1}{2} \leq y < 1$). On a :

$$F(1,1) - F(\frac{1}{2},1) - F(1,\frac{1}{2}) + F(\frac{1}{2},\frac{1}{2}) = -1$$

Une fonction de répartition F est dite absolument continue, s'il existe une fonction non négative f telle que l'on ait :

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f(x_1, x_2, \dots, x_n) dx_1 \dots dx_n$$

Elle sera dite discrète si les valeurs possibles de la variable aléatoire vectorielle forment un ensemble dénombrable de R^n .

Variables indépendantes. Par définition, on dit que deux variables aléatoires X_1 et X_2 sont indépendantes, si, quels que soient les ensembles boréliens B_1 et B_2 les deux évènements " X_1 appartient à B_1 " et " X_2 appartient à B_2 " sont indépendants, c'est-à-dire si l'on a :

$$P(X_1 \in B_1 \text{ et } X_2 \in B_2) = P(X_1 \in B_1) P(X_2 \in B_2)$$

Le théorème 3 (paragraphe 4, ch.I) montre qu'il est possible de construire des variables aléatoires indépendantes, car, si X_1 et X_2 sont deux variables définies sur les espaces probabilisés (R, \mathcal{B}, P_1) et (R, \mathcal{B}, P_2) et si P est la probabilité unique définie sur (R^2, \mathcal{B}^2) telle que, pour tout ensemble borélien de \mathcal{B}^2 de la forme $B = B_1 \times B_2$ (B_1 et B_2 ensembles boréliens de \mathcal{B}), on ait : $P(B) = P(B_1)P(B_2)$, X_1 et X_2 sont bien indépendantes pour (R^2, \mathcal{B}^2, P) .

En particulier, si B_1 et B_2 sont les évènements " $X_1 < x_1$ " et " $X_2 < x_2$ ", on doit avoir :

$$F(x_1, x_2) = F_1(x_1) F_2(x_2)$$

La réciproque est vraie.

Théorème - Pour que deux variables X_1 et X_2 soient indépendantes, il faut et il suffit que leur fonction de répartition $F(x_1, x_2)$ soit de la forme :

$$(7) \quad F(x_1, x_2) = F_1(x_1) F_2(x_2)$$

F_1 et F_2 étant les fonctions de répartition marginales de X_1 et X_2 .

Pour montrer que la condition (7) est suffisante, on observe qu'il est possible de construire sur $(\mathbb{R}^2, \mathcal{B}^2)$ une probabilité P' telle que

$P'(B_1 \times B_2) = P_1(B_1) P_2(B_2)$ dont la fonction de répartition coïncide avec $F(x_1, x_2)$: comme la donnée de F permet la construction de la probabilité, $P'(B)$ coïncide avec $P(B)$.

Plus généralement, n variables aléatoires seront dites mutuellement indépendantes si l'on a :

$$\forall B_i \in \mathcal{B} \quad P(X_1 \in B_1 \text{ et } \dots \text{ et } X_n \in B_n) = P(X_1 \in B_1) P(X_2 \in B_2) \dots P(X_n \in B_n)$$

Pour qu'il en soit ainsi il faut et il suffit que l'on ait :

$$(8) \quad F(x_1, \dots, x_n) = F_1(x_1) F_2(x_2) \dots F_n(x_n)$$

Lois conditionnelles

Soient X et Y deux variables aléatoires ayant la fonction de répartition $F(x, y)$, soient $F_1(x) = F(x, +\infty)$ et $F_2(y) = F(+\infty, y)$ les deux lois marginales. Relativement à l'évènement " Y appartient à l'ensemble borélien B " de probabilité $P_2(B) \neq 0$, on peut définir la probabilité conditionnelle $P_1(A | B)$ de l'évènement " X appartient à l'ensemble borélien A ".

$$P_1(A | B) = \frac{P(X \in A \text{ et } Y \in B)}{P_2(B)} = \frac{P(A \times B)}{P(E \times B)}$$

Cette probabilité conditionnelle permet de définir une nouvelle variable aléatoire X_B (en général, on écrira encore X) qui représente ce que devient X dans l'hypothèse où Y appartient à B . A X_B correspond une fonction de répartition, appelée loi conditionnelle (relativement à B)

$$F(x | B) = P_1(X < x | B)$$

Si B est un intervalle $(a \leq y < b)$ avec $F_2(b) \neq F_2(a)$, on a :

$$F(x | a \leq y < b) = \frac{F(x, b) - F(x, a)}{F_2(b) - F_2(a)}$$

Si F possède une densité de probabilité $f(x, y)$ définie en tout point, on peut, par passage à la limite, définir la densité de probabilité conditionnelle

$f(x|b)$ dans l'hypothèse où $y = b$ (et bien que l'évènement " $y = b$ " soit presque impossible)

$$f(x|b) = \frac{f(x,b)}{\int_{-\infty}^{\infty} f(x,b) dx}$$

Loi spatiale d'une fonction aléatoire, loi temporelle d'un processus stochastique.

La notion de fonction de répartition se généralise partiellement aux fonctions aléatoires et aux processus stochastiques. Si $f(x,e)$ où x est un point de R^n et e un évènement élémentaire, est une fonction aléatoire, nous avons vu que, pour e fixé, $f(x,e)$ est une fonction numérique ordinaire, ou réalisation particulière de la fonction aléatoire, tandis que si l'on considère k points d'appui fixes x_1, \dots, x_k , les

$$Y_i = f(x_i, e) \quad (i = 1, 2, \dots, k)$$

constituent un système de k variables aléatoires. Ce système peut être caractérisé par sa fonction de répartition, que nous noterons :

$$F(y_1, y_2, \dots, y_k; x_1, x_2, \dots, x_k) = P \left[f(x_1) < y_1, \dots, f(x_k) < y_k \right]$$

L'ensemble de ces fonctions de répartition, pour tous les entiers k et tous les systèmes de points d'appui x_1, \dots, x_k , constitue ce que l'on appelle la loi spatiale de la fonction aléatoire (pour un processus stochastique, les points d'appui x_1, \dots, x_k sont remplacés par des instants t_1, \dots, t_k , et l'on parle plutôt de loi temporelle).

Malheureusement, et contrairement à ce qui se passe pour les variables aléatoires à un nombre fini de composantes, la donnée de la loi spatiale ou temporelle ne suffit pas, en général, pour caractériser complètement une fonction aléatoire ou un processus stochastique.

IV.- CHANGEMENTS DE VARIABLES

Si X est une variable aléatoire, c'est-à-dire une application mesurable $X(e)$ de (E, \mathcal{A}) dans (R, \mathcal{B}) et $\alpha(x)$ une fonction d'une variable, c'est-à-dire une application de R dans R , l'application produit

$$\alpha(X) = \alpha \circ X$$

fait correspondre à l'évènement élémentaire e de E le point $\alpha[X(e)]$ de R . Si la fonction $\alpha(X)$ ainsi construite est mesurable, c'est une variable aléatoire, et la formule :

$$Y = \alpha(X)$$

représente alors un changement de variables aléatoires.

Théorème 1 - Si $\alpha(x)$ est une fonction numérique mesurable, et X une variable aléatoire, $\alpha(X)$ est une variable aléatoire.

Il faut vérifier que $\alpha \circ X$ est une application mesurable de (E, \mathcal{A}) dans (R, \mathcal{B}) . On a pour tout ensemble borélien B

$$(\alpha \circ X)^{-1}(B) = X^{-1}[\alpha^{-1}(B)]$$

$\alpha^{-1}(B)$ est un ensemble borélien, puisque α est mesurable, et $X^{-1}[\alpha^{-1}(B)]$ est un évènement de \mathcal{A} puisque X est une variable aléatoire, donc $\alpha \circ X$ est mesurable.

Ce résultat se généralise aisément au cas des variables à plusieurs composantes, α étant, par exemple, une application mesurable de (R^n, \mathcal{B}^n) dans (R^m, \mathcal{B}^m) .

Énonçons :

Théorème 2 - Soient $\alpha_1, \alpha_2, \dots, \alpha_m$ m fonctions numériques mesurables de n variables x_1, x_2, \dots, x_n , et soient X_1, X_2, \dots, X_n n variables aléatoires. Alors les $Y_i = \alpha_i(X_1, X_2, \dots, X_n)$, $i = 1, 2, \dots, m$ sont m variables aléatoires. (c'est-à-dire une variable vectorielle à m composantes).

Par ailleurs, si (R, \mathcal{B}) et (R^n, \mathcal{B}^n) sont des espaces probabilisables, R et R^n sont aussi des espaces topologiques où les notions de convergence et de continuité sont bien définies. En fait \mathcal{B}^n est la σ -algèbre engendrée par la famille des ensembles ouverts, (ou fermés) de R^n . Les théorèmes suivants permettent de préciser le rapport entre la structure topologique et la structure σ -algébrique de R^n :

Théorème 3 - Toute application continue de R^n dans R^m est mesurable.

Soit α une application continue de R^n dans R^m . Si α est continue, l'image inverse d'un ouvert de R^m est un ouvert de R^n , donc un ensemble borélien

de \mathcal{B}^n . Comme \mathcal{B}^m est la σ -algèbre engendrée par les ouverts de \mathbb{R}^m , le théorème 1 (paragraphe 1) permet de conclure.

Corollaire : si les $\alpha_i(x_1, \dots, x_n)$ sont m fonctions continues de n variables, et si X_1, \dots, X_n sont des variables aléatoires, alors les $Y_i = \alpha_i(X_1, \dots, X_n)$ sont les m composantes d'une variable aléatoire vectorielle.

Théorème 4 - Soit α une application d'un espace probablisable (E, \mathcal{A}) dans $(\mathbb{R}^m, \mathcal{B}^m)$. Si α est limite d'une suite d'applications mesurables, alors α est mesurable. En particulier, toute fonction α qui est limite de fonctions mesurables, α_n , est mesurable.

Pour abrégé, donnons la démonstration dans le cas $m = 1$. Par hypothèse, en tout e de E , on a $\alpha(e) = \lim \alpha_n(e)$ pour n infini. Il faut montrer, par exemple, que l'ensemble :

$$A_\lambda = \left\{ e : \alpha(e) \leq \lambda \right\}$$

des éléments e de E vérifiant $\alpha(e) \leq \lambda$ est un événement de \mathcal{A} , quel que soit le nombre réel λ .

La proposition " e appartient à A_λ ", c'est-à-dire " $\alpha(e) \leq \lambda$ " équivaut (puisque $\alpha(e)$ est limite des $\alpha_n(e)$) à : " Pour tout entier $r \geq 1$, il existe un entier N tel que : pour $n \geq N$, on ait $\alpha_n(e) \leq \lambda + \frac{1}{r}$ ". L'ensemble $B_{n,r}$

$$B_{n,r} = \left\{ e : \alpha_n(e) \leq \lambda + \frac{1}{r} \right\}$$

est un événement de \mathcal{A} puisque $\alpha_n(e)$ est mesurable. Alors l'ensemble A_λ :

$$A_\lambda = \left\{ e : \forall r, \exists N \text{ tel que pour tout } n \geq N \quad \alpha_n(e) \leq \lambda + \frac{1}{r} \right\}$$

qui se met, par une application évidente des règles de la logique, sous la forme :

$$A_\lambda = \prod_{r=1}^{\infty} \sum_{N=1}^{\infty} \prod_{n=N}^{\infty} B_{n,r}$$

est lui-aussi un événement de \mathcal{A} , d'après l'axiome 2 des σ -algèbres.

Conséquence - Soit α une application de \mathbb{R}^n dans \mathbb{R}^m et α_n une suite d'applications

continues, telles que en tout point x de R^n , on ait $\alpha(x) = \lim_{n \rightarrow \infty} \alpha_n(x)$ pour $n \rightarrow \infty$. Alors α est mesurable. En particulier, si une fonction $\alpha(x_1 \dots x_n)$ est, en tout point, limite d'une suite de fonctions continues, alors la fonction α est mesurable, et par suite, si $X_1 \dots X_n$ sont des variables aléatoires $\alpha(X_1 \dots X_n)$ est une variable aléatoire.

C'est une conséquence immédiate des théorèmes 2, 3 et 4.

Exemples.- Si X et Y sont des variables aléatoires, et λ une constante quelconque $X + \lambda$, λX , $X + Y$ et XY sont des variables aléatoires.

Il suffit de remarquer que l'application $(x, y) \rightarrow x + y$ de R^2 dans R , par exemple, est continue.

Changement de variable pour une fonction de répartition.

Soit X une variable aléatoire ayant une fonction de répartition $F(x)$, et $\alpha(x)$ une fonction numérique mesurable. Alors $\alpha(X)$ est une variable aléatoire, et possède une fonction de répartition $F_\alpha(x)$. Pour exprimer $F_\alpha(x)$ en fonction de $F(x)$, il faut, en principe, reconstituer la fonction d'ensemble $P(B) =$ probabilité de l'évènement " X appartient à B ", qui ne dépend que de $F(x)$. On a alors :

$$F_\alpha(x) = P \left[\alpha^{-1}(A_x) \right]$$

où A_x représente l'évènement " $X < x$ ". Si $\alpha^{-1}(A_x)$ possède une structure simple, (par exemple si c'est une somme d'un nombre fini d'intervalles), on pourra exprimer directement F_α en fonction de F .

Exemple - Prenons $\alpha(x) = x^2$, et cherchons la fonction de répartition $F_\alpha(y)$ de la variable aléatoire $Y = X^2$. L'évènement " $X^2 < y$ " est impossible pour $y \leq 0$. Pour $y > 0$, il s'identifie à l'évènement " $-\sqrt{y} < X < +\sqrt{y}$ ", d'où

$$\left\{ \begin{array}{ll} F_\alpha(y) = F(\sqrt{y}) - F(-\sqrt{y} + 0) & (y > 0) \\ F_\alpha(y) = 0 & (y \leq 0) \end{array} \right.$$

On notera bien que $F_\alpha(y)$ n'est pas égal à $F(\sqrt{y})$.

Si $F(x)$ possède une densité de probabilité $f(x)$, $F_\alpha(y)$ en possède une

également, soit $f_{\alpha}(y)$ avec :

$$\left\{ \begin{array}{l} f_{\alpha}(y) = \frac{1}{2\sqrt{y}} f_{\alpha}(\sqrt{y}) + \frac{1}{2\sqrt{-y}} f_{\alpha}(-\sqrt{y}) \quad (\text{pour } y > 0) \\ f_{\alpha}(y) = 0 \quad (\text{pour } y \leq 0) \end{array} \right.$$

V.- INTEGRALES DE LEBESGUE ET DE STIELTJES.

En vue d'établir les notions fondamentales d'espérance mathématique et, au chapitre suivant, de fonction caractéristique, nous devons utiliser l'intégrale de Stieltjes, plus générale que celle de Riemann. Nous nous contenterons, dans ce paragraphe, d'énoncer des définitions et des propriétés, le plus souvent sans démonstration.

Intégrale de LEBESGUE.

Soit (E, \mathcal{A}, P) un espace probabilisé et $f(e)$ une fonction mesurable définie sur E (c'est-à-dire une variable aléatoire). Considérons les ensembles :

$$A_k^n = \left\{ e : \frac{k}{n} \leq f(e) < \frac{k+1}{n} \right\}$$

qui sont des événements de \mathcal{A} , puisque f est mesurable, et, par suite, possèdent des probabilités $P(A_k^n)$. Considérons les deux séries, supposées convergentes :

$$\left\{ \begin{array}{l} J_n = \sum_{k=0}^{+\infty} \frac{k}{n} P(A_k^n) \\ J'_n = \sum_{k=-1}^{-\infty} \frac{k+1}{n} P(A_k^n) \end{array} \right.$$

La fonction f sera dite intégrable, au sens de LEBESGUE, sur E et pour la mesure (probabilité) P , si, lorsque n tend vers l'infini, J_n et J'_n tendent vers des limites définies J et J' . Par définition, la somme $I = J + J'$ sera appelée intégrale de la fonction f , et on écrira (symboliquement) :

$$I = \int f(e) d P(e)$$

Si A est un évènement de \mathcal{A} , on définira de la même manière l'intégrale I_A étendue au sous-ensemble A en remplaçant A_k^n par le produit $A A_k^n$ dans les sommes partielles ci-dessus, et on posera :

$$I_A = \int_A f(e) dP(e)$$

Propriété 1 - Pour qu'une fonction mesurable f soit intégrable, il faut et il suffit que $|f|$ le soit.

En effet, l'intégrale de $|f|$ se construit comme limite de $(J_n - J'_n)$ qui converge vers $J - J'$ si, et seulement si, $J_n \rightarrow J$ et $J'_n \rightarrow J'$.

Propriété 2 - Soit f_n une suite non décroissante de fonctions non négatives mesurables et intégrables sur (E, \mathcal{A}, P) , telles qu'en tout point e de E les $f_n(e)$ admettent une limite $f(e)$. Alors $f(e)$ est mesurable et intégrable, et on a, pour tout A de \mathcal{A} :

$$\lim \int_A f_n(e) dP(e) = \int_A f(e) dP(e)$$

La propriété reste vraie si la convergence des $f_n(e)$ vers une limite $f(e)$ a lieu seulement presque partout, c'est-à-dire en tout point sauf sur un ensemble D tel que $P(D) = 0$.

Propriété 3 - (Théorème de Lebesgue) - Soit f_n une suite de fonctions mesurables et intégrables telles qu'en tout point e la suite $f_n(e)$ ait une limite $f(e)$. S'il existe une fonction g mesurable et intégrable telle que : $|f_n(e)| \leq g(e)$ pour tout n et tout e de E , alors f est mesurable et intégrable et on a, pour tout A de \mathcal{A} :

$$\lim \int_A f_n(e) dP(e) = \int_A f(e) dP(e)$$

La propriété reste vraie si l'expression "pour tout e " est remplacée par "presque partout".

Propriété 4 - (Théorème de Fubini) - Soient $(E_1, \mathcal{A}_1, P_1)$, $(E_2, \mathcal{A}_2, P_2)$ deux espaces probabilisés et $f(e_1, e_2)$ une fonction mesurable sur $(E_1 \times E_2, \mathcal{A}_1 \times \mathcal{A}_2)$, telle que l'intégrale :

$$I = \int_{E_1} \left[\int_{E_2} f(e_1, e_2) d P_2(e_2) \right] d P_1(e_1)$$

soit définie. On peut intervertir l'ordre des intégrations. Autrement dit, l'intégrale

$$\int_{E_2} \left[\int_{E_1} f(e_1, e_2) d P_1(e_1) \right] d P_2(e_2)$$

existe et est égale à I .

Propriété 5 - Soit α une application mesurable de (E_1, \mathcal{A}_1) dans (E_2, \mathcal{A}_2) , et P_1 une probabilité sur (E_1, \mathcal{A}_1) . Alors la formule

$$P_2(A_2) = P_1 \left[\alpha^{-1}(A_2) \right]$$

définit une probabilité sur (E_2, \mathcal{A}_2) . Si f est une fonction mesurable définie sur (E_2, \mathcal{A}_2) , $f \circ \alpha$ est une fonction mesurable définie sur (E_1, \mathcal{A}_1) . Si de plus l'une des deux fonctions (f ou $f \circ \alpha$) est intégrable, l'autre l'est aussi, et l'on a :

$$\int_{E_2} f(e_2) d P_2(e_2) = \int_{E_1} f \left[\alpha(e_1) \right] d P_1(e_1)$$

La première partie de cet énoncé a été établi au paragraphe 1. La deuxième montre qu'il est possible de transformer une intégrale de Lebesgue par un changement d'espace probabilisé.

Intégrale de STIELTJES

Dans l'énoncé de la propriété 5, prenons $E_2 = \mathbb{R}$ et $\mathcal{A}_2 = \mathcal{B}$. Alors l'application α est une variable aléatoire X , possédant une fonction de répartition $F(x)$. Dans un grand nombre de cas, il est possible de remplacer l'intégrale de Lebesgue $\int_{\mathbb{R}} f(x) d P_2(x)$ par une intégrale plus simple, dont la définition ne fait intervenir que la fonction de répartition $F(x)$, et non pas la fonction d'ensemble $P_2(B)$ elle-

même. C'est l'intégrale de Stieltjes qui peut être définie comme suit :

Soit (a, b) un intervalle, et $x_0 = a < x_1 < \dots < x_n = b$ des points de cet intervalle. On considère les sommes :

$$I_n = \sum_{i=1}^n f(\xi_i) \left[F(x_i) - F(x_{i-1}) \right]$$

où $x_{i-1} \leq \xi_i \leq x_i$. Exactement comme dans la définition de l'intégrale de Riemann, on considère les bornes supérieures et inférieures. J_n et J'_n des I_n relativement au choix des ξ_i , et leurs limites J et J' lorsque l'on augmente n faisant tendre uniformément vers 0 chacun des intervalles (x_i, x_{i+1}) . Si $J = J'$, on dit que f est intégrable au sens de Stieltjes, et on pose :

$$J = \int_a^b f(x) dF(x)$$

On démontre que si $|f|$ est intégrable au sens de Stieltjes (relativement à $F(x)$), f est alors intégrable à la fois au sens de Lebesgue (relativement à la probabilité P correspondant à la fonction de répartition $F(x)$) et au sens de Stieltjes, et que les deux intégrales coïncident.

Si J tend vers une limite lorsque a et b tendent vers $-\infty$ et $+\infty$ respectivement, on pose :

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

Propriété 6 - Si $f(x)$ est une fonction continue et bornée, elle est intégrable au sens de Stieltjes sur tout intervalle fini ou infini. Si $f(x)$ est bornée, et possède au plus une infinité dénombrable de points de discontinuité, la conclusion est encore vraie.

Si l'une des limites de l'intervalle d'intégration (a, b) , par exemple a , est un point de discontinuité de $F(x)$, la valeur de l'intégrale peut être différente selon que a appartient ou non à l'intervalle (a, b) . Dans le premier cas, on

note \int_{a-0}^b et \int_{a+0}^b dans le second. On a :

$$\int_{a-0}^b f(x) dF(x) - \int_{a+0}^b f(x) dF(x) = f(a) \left[F(a+0) - F(a) \right]$$

Enfin, avec $f(x) = 1$, on a évidemment :

$$\int_{-\infty}^b dF(x) = F(b)$$

Propriété 7 - Si $F(x)$ est absolument continue et a une densité de probabilité $f(x)$, et si $g(x)$ est intégrable, on a :

$$\int_a^b g(x) dF(x) = \int_a^b g(x) f(x) dx$$

la deuxième intégrale étant une intégrale ordinaire.

Si $F(x)$ est discrète, et $g(x)$ intégrable, on a, en désignant par x_i et p_i les points de discontinuité et les sauts correspondants :

$$\int_a^b g(x) dF(x) = \sum p_i g(x_i)$$

la somme étant étendue aux x_i appartenant à l'intervalle d'intégration.

Propriété 8 - L'intégrale de Stieltjes est une fonctionnelle bi-linéaire relativement à $F(x)$ et à la fonction intégrable $g(x)$: si $F_1 \dots F_n$ sont des fonctions de répartitions, $g_1 \dots g_m$ des fonctions intégrables pour toutes les F_i , les λ_i et les μ_j des constantes, on a :

$$\int_a^b \left[\sum_{j=1}^m \mu_j g_j(x) \right] d \left[\sum_{i=1}^n \lambda_i F_i(x) \right] = \sum_{j=1}^m \sum_{i=1}^n \lambda_i \mu_j \int_a^b g_j(x) dF_i(x)$$

Exemple : Somme de variables aléatoires indépendantes. Soient X et Y deux variables aléatoires indépendantes, ayant $F(x,y) = F_1(x) F_2(y)$ comme fonction de répartition.

La somme $Z = X + Y$ est une variable aléatoire comme nous l'avons déjà vu. La fonction de répartition $F_3(z)$ de Z , probabilité de l'évènement " $X + Y < z$ " est, d'après le théorème de Fubini :

$$F_3(z) = \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{z-x} dF_2(y) \right] dF_1(x)$$

Soit

$$(9) \quad F_3(z) = \int_{-\infty}^{\infty} F_2(z-x) dF_1(x) = \int_{-\infty}^{\infty} F_1(z-y) dF_2(y)$$

Si F_2 et F_1 ont des densités de probabilité f_1 et f_2 , alors F_3 a aussi une densité f_3 donnée par

$$f_3(z) = \int_{-\infty}^{+\infty} f_2(z-x) f_1(x) dx = \int_{-\infty}^{+\infty} f_1(z-y) f_2(y) dy$$

On écrit, symboliquement :

$$(10) \quad f_3 = f_2 * f_1 = f_1 * f_2$$

L'opération $*$ ainsi associée à l'addition de deux variables aléatoires indépendantes (à loi absolument continue) porte le nom de convolution (on sait qu'elle joue un grand rôle dans les problèmes de physique mathématique où interviennent des équations aux dérivées partielles à coefficients constants). Sous la forme (10), on peut d'ailleurs aussi reconnaître une généralisation du produit de convolution de deux fonctions, à savoir : le produit de convolution de deux mesures (deux probabilités), ici représentées par leurs fonctions de répartition.

Le produit de convolution est associatif et commutatif (tout comme l'addition des variables aléatoires). Il est bilinéaire, en ce sens que :

$$\left(\sum_i \lambda_i f_i \right) * \left(\sum_j \mu_j g_j \right) = \sum_{i,j} \lambda_i \mu_j f_i * g_j$$

Enfin, il existe un élément unité pour la convolution. C'est la mesure de Dirac, $\delta(x)$ bien connue des physiciens, définie par la propriété suivante : quelle que soit la fonction continue $\varphi(x)$, on a :

$$\int \delta(x) \varphi(x) dx = \varphi(0)$$

En calcul des probabilités, il est naturel de représenter la mesure de Dirac par sa fonction de répartition. Nous la noterons $\Theta(x)$:

$$(11) \quad \left\{ \begin{array}{ll} \Theta(x) = 0 & \text{pour } x \leq 0 \\ \Theta(x) = 1 & \text{pour } x > 0 \end{array} \right.$$

Θ est une loi discrète pour laquelle toute la probabilité est concentrée au seul point $x = 0$. Elle représente une variable aléatoire X presque certainement nulle : $P(X = 0) = 1$. Si, dans (9) F_1 est remplacée par Θ , on a $F_3 = F_2$. Autrement dit, si X est presque certainement nulle, alors $X + Y$ a même fonction de répartition que Y : cela est naturel, puisque l'évènement " $X + Y = Y$ " est presque certain.

Ces trois propriétés du produit de convolution (associatif et commutatif, bilinéaire, existence d'un élément unité) définissent la structure appelée algèbre. On dira : l'algèbre de la convolution, ou l'algèbre des lois de probabilités.

VI.- ESPERANCE MATHEMATIQUE ET MOMENTS.

On connaît la signification intuitive de l'espérance mathématique, ou valeur probable, d'une variable aléatoire : c'est une moyenne de toutes les valeurs que X est susceptible de prendre, pondérées par leurs probabilités. Ou encore, si l'on effectue un grand nombre n de répétitions de l'épreuve, c'est la limite vers laquelle tend la moyenne arithmétique des valeurs observées lorsque n augmente indéfiniment (sous ce deuxième aspect, on fait implicitement appel à la loi forte des grands nombres). La définition précise est la suivante :

Définition : Espérance mathématique.

Soit un espace probabilisé (E, \mathcal{A}, P) et une variable aléatoire X , c'est-à-dire une fonction numérique $X(e)$ définie et mesurable sur (E, \mathcal{A}, P) . Si $X(e)$ est intégrable au sens de Lebesgue, son intégrale dans E se note $E(X)$ et s'appelle espérance mathématique de X .

Par définition, on a :

$$(12) \quad E(X) = \int X(e) \, dP(e)$$

D'après la propriété 5 de l'intégrale de Lebesgue, il est possible de calculer directement $E(X)$ dans l'espace (R, \mathcal{B}, P') , où P' est la probabilité déduite de P dans le changement d'espace probabilisable. Dans ce changement d'espace, la fonction $X(e)$ est remplacée par la fonction $X(x) = x$, et l'on a ainsi :

$$(13) \quad E(X) = \int x \, dP'(x)$$

Les deux relations (12) et (13) sont équivalentes. D'après la propriété 4, $X(e)$ est intégrable si et seulement si $|X(e)|$ est intégrable. Autrement dit, $E(X)$ n'existe que si $E(|X|)$ existe également, c'est-à-dire si l'intégrale :

$$E(|X|) = \int |x| \, dP'(x)$$

existe.

Par ailleurs, X possède une fonction de répartition $F(x)$. Si $|x|$ est intégrable pour $F(x)$, au sens de Stieltjes, c'est-à-dire si l'intégrale

$$\int |x| \, dF(x)$$

existe, alors (voir paragraphe précédent) x est aussi intégrable à la fois au sens de Lebesgue et de Stieltjes, et les deux intégrales coïncident. Autrement dit, on a :

$$(14) \quad E(X) = \int_{-\infty}^{+\infty} x \, dF(x)$$

Dans la plupart des applications, la formule (14), jointe à la condition que $|x|$ soit intégrable suffit à définir l'espérance mathématique $E(X)$.

On interprète parfois la relation (14) en disant que l'espérance mathématique $E(X)$ de la variable aléatoire X est égale à la valeur probable, pour la loi $F(x)$, de la variable ordinaire x (cette valeur probable étant définie par l'intégrale du deuxième membre).

Définition : Valeur probable d'une fonction $g(x)$. Etant données une fonction de répartition $F(x)$ et une fonction mesurable $g(x)$ telle que $|g(x)|$ soit intégrable au sens de Stieltjes pour $F(x)$, on appelle valeur probable de la fonction $g(x)$ (relativement à la loi F) l'intégrale :

$$(15) \quad I [g(x)] = \int_{-\infty}^{+\infty} g(x) dF(x)$$

Cette intégrale existe nécessairement du fait que $|g|$ est intégrable. Le théorème suivant précise le rapport entre les notions de valeur probable et d'espérance mathématique.

Théorème 1 - Soit X une variable aléatoire, $F(x)$ sa fonction de répartition, et $g(x)$ une fonction mesurable telle que $|g|$ soit intégrable pour F au sens de Stieltjes.
Alors $g(X)$ est une variable aléatoire qui possède une espérance mathématique égale à la valeur probable de $g(x)$:

$$E [g(X)] = I [g(x)]$$

En effet, puisque $|g|$ est intégrable au sens de Stieltjes, g est intégrable au sens de Lebesgue pour la probabilité P associée dans (R, \mathcal{B}) à la fonction de répartition $F(x)$, et on a :

$$(16) \quad \int g(x) dP(x) = \int g(x) dF(x)$$

Mais $g(x)$, fonction mesurable pour (R, \mathcal{B}, P) définit une variable aléatoire Y , qui n'est autre que $g(X)$ [d'un point de vue un peu formel, il faut dire : X étant l'application identique de (R, \mathcal{B}) , la variable aléatoire $g(X)$ est définie comme l'application produit $Y = g \circ X = g$]. Par définition l'espérance mathématique de Y est l'intégrale de Lebesgue.

$$E(Y) = \int Y(x) dP(x) = \int g(x) dP(x)$$

L'égalité (16) montre que l'on a bien

$$(17) \quad E(Y) = \int g(x) d F(x)$$

Ce théorème permet en pratique de confondre les deux notions d'espérance mathématique de la variable aléatoire $g(X)$ et de valeur probable de la fonction $g(x)$. Mais on n'oubliera pas que $|g|$ doit être intégrable.

Propriétés de l'espérance Mathématique.

1 - L'espérance mathématique définit une opération linéaire : X et Y étant des variables aléatoires, λ et μ des constantes, on a :

$$E(\lambda X + \mu Y) = \lambda E(X) + \mu E(Y)$$

cela résulte des propriétés élémentaires de l'intégrale de Lebesgue.

2 - Si X et Y sont deux variables aléatoires indépendantes, on a

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

et, plus généralement, g et g' étant des fonctions mesurables

$$E \left[g(X) g'(Y) \right] = E \left[g(X) \right] E \left[g'(Y) \right]$$

Comme $F(x,y) = F_1(x) F_2(y)$, la relation (17) permet une démonstration immédiate (mais la propriété reste vraie si les espérances mathématiques sont définies par les intégrales de Lebesgue).

Définition : Moments d'une loi de probabilité $F(x)$ - Soit $F(x)$ une fonction de répartition. Si $|x|^n$ est intégrable pour F , on appelle moment d'ordre n de la loi F la valeur probable de x^n , c'est-à-dire l'intégrale :

$$(18) \quad m_n = \int x^n d F(x)$$

En particulier, le moment du premier ordre m_1 coïncide avec la valeur probable de x , ou espérance mathématique $E(X)$ de la variable aléatoire X associée à F . Posons, pour abréger :

$$m_1 = m$$

On appellera moment centré d'ordre n la valeur probable, si elle existe, de $(x-m)^n$, soit :

$$(19) \quad a_n = \int (x-m)^n d F(x)$$

On sait que le moment centré d'ordre 2 porte le nom de variance σ^2 .

$$\sigma^2 = a_2 = \int (x-m)^2 dF(x)$$

Propriétés de la moyenne m et de la variance σ^2

1 - La valeur probable de l'expression $(x-b)^2$ est minimale pour $m = b$, et ce minimum est égal à la variance σ^2 . De plus, on a :

$$(20) \quad \sigma^2 = m_2 - m^2 = E(X^2) - [E(X)]^2$$

démonstration immédiate à partir de la relation :

$$E[(X-b)^2] = E(X^2) - 2b E(X) + b^2$$

2 - Si X et Y sont deux variables indépendantes, la variance de la somme X + Y est la somme des variances de X et de Y (démonstration immédiate).

3 - Inégalité de Tchebychev. Quel que soit le nombre positif λ l'évènement : " $|X-m| \geq \lambda$ " a une probabilité inférieure ou égale à $\frac{\sigma^2}{\lambda^2}$

$$(21) \quad P(|X-m| \geq \lambda) \leq \frac{\sigma^2}{\lambda^2}$$

En effet, cette probabilité est $\int_D dF(x)$, le domaine D étant défini par : " $|x-m| \geq \lambda$ ". Quand x appartient à D, on a donc $\frac{(x-m)^2}{\lambda^2} \geq 1$. Par suite :

$$\int_D dF(x) \leq \int_D \frac{(x-m)^2}{\lambda^2} dF(x) \leq \int_{-\infty}^{+\infty} \frac{(x-m)^2}{\lambda^2} dF(x) = \frac{\sigma^2}{\lambda^2}$$

Cette inégalité nous permettra dans la suite d'établir très simplement la loi (ordinaire) des grands nombres. Elle souligne la signification de la variance, qui est celle d'un indice de dispersion.

Définition : Moments absolus. Par définition, on appellera moment absolu d'ordre α et on notera M_α la valeur probable, si elle existe, de $|x|^\alpha$, α étant un nombre réel positif quelconque :

$$(22) \quad M_\alpha = \int |x|^\alpha dF(x)$$

Théorème 2 -- Si le moment absolu d'ordre α existe, tous les moments absolus d'ordre $\beta \leq \alpha$ existent également et $(M_\beta)^{1/\beta}$ est une fonction croissante de β .

$$(23) \quad (M_\beta)^{1/\beta} \leq (M_\gamma)^{1/\gamma} \quad \text{si } \beta \leq \gamma \leq \alpha$$

L'existence de M_β s'établit facilement. Si a est un nombre ≥ 1 , les deux intégrales $\int_{-a}^{+a} |x|^\beta dF$ et $\int_{-a}^{+a} |x|^\alpha dF$ existent, puisque $|x|^\beta$ est bornée sur l'intervalle $(-a, +a)$. Comme $a \geq 1$, on a $|x|^\beta \leq |x|^\alpha$ dès que $|x| \geq a$, d'où l'inégalité :

$$\int_a^b |x|^\beta dF \leq \int_a^b |x|^\alpha dF$$

Ces deux intégrales sont des fonctions croissantes de b . La deuxième a une limite lorsque $b \rightarrow +\infty$, puisque M_α existe. Il en est donc de même de la première et M_β existe aussi.

L'importante inégalité (23) peut s'établir comme suit : Soient β et γ deux nombres tels que $\beta + \gamma \leq \alpha$, et $\beta \leq \gamma$. L'expression :

$$\int \left[\lambda |x|^{\frac{\gamma-\beta}{2}} + \mu |x|^{\frac{\gamma+\beta}{2}} \right]^2 dF = \lambda^2 M_{\gamma-\beta} + 2\lambda\mu M_\gamma + \mu^2 M_{\gamma+\beta}$$

est une forme quadratique définie positive en λ et μ , d'où l'inégalité :

$$M_\gamma \leq \sqrt{M_{\gamma-\beta} M_{\gamma+\beta}}$$

En remplaçant γ par $k\gamma$ ($k = 0, 1, \dots, n$) et β par γ , on obtient :

$$\begin{aligned} M_\gamma &\leq \sqrt{M_0 M_{2\gamma}} \\ M_{2\gamma} &\leq \sqrt{M_\gamma M_{3\gamma}} \\ \dots &\dots \dots \dots \\ M_{k\gamma} &\leq \sqrt{M_{(k-1)\gamma} M_{(k+1)\gamma}} \end{aligned}$$

Soit encore (compte tenu de $M_0 = 1$).

$$\left\{ \begin{array}{l} M_Y^2 \leq M_{2Y} \\ M_{2Y}^4 \leq M_Y^2 M_{3Y}^2 \\ \dots \\ M_{kY}^{2k} \leq M_{(k-1)Y}^k M_{(k+1)Y}^k \end{array} \right.$$

En multipliant membre à membre il vient, après simplification :

$$M_{kY}^{k+1} \leq M_{(k+1)Y}^k$$

soit encore

$$(M_{kY})^{\frac{1}{kY}} \leq (M_{(k+1)Y})^{\frac{1}{(k+1)Y}}$$

pourvu que $(k+1)Y \leq \alpha$. Si $\frac{p}{q}$ et $\frac{p'}{q}$ sont deux nombres rationnels avec $p \leq p'$, on obtient (en posant $Y = \frac{1}{q}$ et en itérant en k)

$$(M_{p/q})^{q/p} \leq (M_{p'/q})^{q/p'}$$

La démonstration dans le cas de deux nombres quelconques $\beta \leq \gamma \leq \alpha$ se fait par passage à la limite, en remarquant que M_β est fonction continue de β

Corollaire 1 - Le théorème s'applique aux valeurs probables de $|x-a|^\alpha$, α nombre réel quelconque, et, en particulier, aux moments absolus centrés $E[|X-m|^\alpha]$.

Il suffit de remplacer la variable aléatoire X par la variable aléatoire $X - a$.

Corollaire 2 - Si le moment d'ordre n , m_n , existe, tous les moments m_k d'ordre $k \leq n$ existent aussi. Même énoncé pour les moments centrés et les valeurs probables de $(x-a)^n$.

En effet, si m_n existe, $|x|^n$ est intégrable, donc M_n existe. Par suite, pour $k \leq n$, M_k existe aussi, ce qui signifie que $|x|^k$ est intégrable et entraîne l'existence de m_k .

Corollaire 3.— Pour que $(x-a)^n$ ait une valeur probable quel que soit a , il faut et il suffit qu'il existe un nombre a_0 tel que $(x-a_0)^n$ ait une valeur probable. En particulier, il faut et il suffit que le moment m_n , ou le moment centré d'ordre n existe.

Démonstration immédiate.

Moments d'une loi à plusieurs variables. Si $F(x_1 \dots x_n)$ est une loi de répartition à n variables, on définit le moment $m_{k_1 \dots k_n}$ comme la valeur probable de $(x_1)^{k_1} (x_2)^{k_2} \dots (x_n)^{k_n}$. Les notions de moments centrés, moments absolus, etc... se généralisent d'elles-mêmes.

En particulier, on appellera covariance des deux variables X_i et X_j et on notera σ_{ij} la valeur probable du produit $(x_i - m_i)(x_j - m_j)$, m_i et m_j désignant ici les espérances mathématiques de X_i et X_j . Pour $i = j$, la covariance $\sigma_{ii} = \sigma_i^2$ est égale à la variance de X_i . On établira la propriété suivante : si les u_i sont des constantes, la variable aléatoire :

$$Y = \sum_{i=1}^n u_i X_i$$

possède une variance égale à :

$$(24) \quad \sum_{ij} u_i u_j \sigma_{ij} = \sum_{i=1}^n u_i^2 \sigma_i^2 + 2 \sum_{i < j} u_i u_j \sigma_{ij}$$

Cette variance devant être positive ou nulle, on en déduit que la matrice des covariances σ_{ij} est définie positive. En particulier, on a l'inégalité de Schwartz :

$$(\sigma_{ij})^2 \leq \sigma_i^2 \sigma_j^2$$

Si l'on appelle coefficient de corrélation ρ_{ij} de X_i et X_j le rapport :

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

on a $-1 < \rho_{ij} \leq +1$. On sait que ce coefficient représente, dans une certaine

mesure, le degré de dépendance des deux variables X_i et X_j . En particulier, si X_i et X_j sont indépendantes, on a $\rho_{ij} = 0$. Mais la réciproque n'est pas vraie. On peut avoir $\rho_{ij} = 0$ sans que X_i et X_j soient indépendantes.

Exemple - Soit $F(x,y)$ la fonction de répartition admettant la densité de probabilité $f(x,y)$ uniforme dans un cercle de rayon R :

$$\left\{ \begin{array}{ll} f(x,y) = \frac{1}{\pi R^2} & \text{si } x^2 + y^2 \leq R^2 \\ f(x,y) = 0 & \text{si } x^2 + y^2 > R^2 \end{array} \right.$$

Le coefficient de corrélation est nul. Mais X et Y ne sont pas indépendantes. Cela se comprend : l'évènement $X^2 + Y^2 > R^2$ étant presque impossible, si l'on connaît la valeur y prise par Y , il est presque certain que l'on aura :

$$-\sqrt{R^2 - y^2} \leq X \leq \sqrt{R^2 - y^2}$$

La connaissance de y apporte bien une information sur X . Par contre, on a le résultat suivant :

Théorème 3. - Soient X et Y des variables aléatoires ayant des variances σ_1^2 et σ_2^2 respectivement. Soit ρ leur coefficient de corrélation. Si $\rho = +1$, l'évènement $\frac{X-E(X)}{\sigma_1} = \frac{Y-E(Y)}{\sigma_2}$ est presque certain. Si $\rho = -1$, l'évènement

$\frac{X-E(X)}{\sigma_1} + \frac{Y-E(Y)}{\sigma_2} = 0$ est presque certain.

En effet, supposons par exemple $\rho = +1$. La formule (24) montre que la variable aléatoire

$$Z = \frac{X-E(X)}{\sigma_1} - \frac{Y-E(Y)}{\sigma_2}$$

a une variance nulle. Alors, quel que soit λ , l'inégalité de Tchebychev (21) donne :

$$P(|Z| \geq \lambda) = 0$$

Prenons $\lambda = \frac{1}{n}$. L'évènement " $|Z| \geq 1$ ou $|Z| \geq \frac{1}{2}$ ou ... ou $|Z| \geq \frac{1}{n}$ ou ..." a une probabilité nulle, d'après l'axiome (3) des probabilités. Son complémentaire qui est " $Z \leq \frac{1}{n}$ quel que soit n" c'est-à-dire " $|Z| = 0$ ", a bien une probabilité unité.

Régression de X et Y. Comme on le sait, le coefficient ρ de corrélation ne donne une mesure précise du degré de dépendance de X et Y que dans le cas où $F(x,y)$ est la loi normale à 2 variables :

$$F(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^x \int_{-\infty}^y \exp \left[-\frac{1}{2} \left(\frac{(u-m_1)^2}{\sigma_1^2} - 2\rho \frac{(u-m_1)(v-m_1)}{\sigma_1\sigma_2} + \frac{(v-m_2)^2}{\sigma_2^2} \right) \right] dudv$$

Dans le cas où $F(x,y)$ admet une densité de probabilité $f(x,y)$, on peut caractériser la loi de x à y fixé, de densité :

$$f(x|y) = \frac{f(x,y)}{\int_{-\infty}^{+\infty} f(x,y) dx}$$

par ses moments, ou moments conditionnels (relativement à l'hypothèse $Y = y$).

En particulier, l'espérance mathématique et la variance de x à y fixé (dite variance résiduelle) sont :

$$m_x(y) = \frac{\int x f(x,y) dx}{\int f(x,y) dx}$$

$$\sigma_x^2(y) = \frac{\int x^2 f(x,y) dx}{\int f(x,y) dx} - [m_x(y)]^2$$

La fonction $m_x(y)$ donnant, en fonction de y, la valeur probable de x à y fixé est dite fonction de régression de X en Y (ou de x en y). On montrera facilement que, parmi toutes les fonctions $g(y)$, elle réalise le minimum de l'intégrale :

$$\iint [x - g(y)]^2 dF(x,y) = \iint [x - g(y)]^2 f(x,y) dx dy$$

La valeur de ce minimum est alors égale à la valeur probable de la variance résiduelle.

$$\iint \sigma_x^2(y) f(x,y) dx dy$$

Dans le cas particulier d'une loi normale à 2 variables, on a

$$(25) \quad \left\{ \begin{array}{l} m_x(y) = m_1 + \rho \frac{\sigma_1}{\sigma_2} (y - m_2) \\ \sigma_x^2(y) = \sigma_1^2 (1 - \rho^2) \end{array} \right.$$

et ces relations montrent bien que ρ épuise, dans ce cas, la notion de dépendance des variables X et Y . Dans le cas général, l'expression (25) de $m_x(y)$ ne représente pas la véritable régression de x en y , mais seulement la meilleure approximation de celle-ci au sens des moindres carrés.

VII.- CONVERGENCE EN LOI

L'impératif général qui nous incite à introduire l'idée d'infinité en calcul des probabilités nous conduit à chercher à donner un sens précis à la proposition "une suite de variables aléatoires X_n tend, ou converge, vers une variable aléatoire X ". Ce n'est qu'en munissant l'ensemble des variables aléatoires d'une structure topologique, ou pseudo topologique définie par une convergence qu'il est possible de fonder ce que l'on appelle l'analyse aléatoire. Parmi les définitions possibles de la convergence (il y en a plusieurs) la plus faible, ou la moins stricte, est ce que l'on appelle la convergence "en loi". C'est une notion qui n'intéresse que les fonctions de répartition (ou lois de probabilités). On dira que la suite X_n de variables aléatoires converge en loi vers la variable aléatoire X , si, en tout point x , la suite des $F_n(x)$ converge vers $F(x)$. En fait, la définition rigoureuse est encore moins stricte et suppose seulement la convergence en tout point x où $F(x)$ est continue. Énonçons :

Définition : convergence en loi - Soient X_n une suite de variables aléatoires de fonctions de répartition $F_n(x)$, et X une variable aléatoire de fonction de répartition $F(x)$. On dira indifféremment que la suite de variables aléatoires X_n converge en loi vers la variable aléatoire X , ou que les lois F_n convergent vers la loi F si en tout point x pour lequel $F(x)$ est continue on a $\lim_{n \rightarrow \infty} F_n(x) = F(x)$ pour $n \rightarrow \infty$.

On prendra garde que la suite $F_n(x)$ peut, en tout point x (ou en tout point de continuité de $F(x)$) converger vers une limite $F(x)$ sans que $F(x)$ soit une loi de probabilité. Dans ce cas, il n'y aura pas convergence en loi.

Exemple : Effet de fuite. - Soit une particule animée d'un mouvement brownien placée au temps $t = 0$ au point $x = 0$ et susceptible de se déplacer sur l'axe des x . Au temps $t > 0$, la probabilité pour que la particule se trouve en un point $X(t)$ tel que $X(t) < x$ est :

$$F(x, t) = \frac{1}{\sqrt{2\pi ct}} \int_{-\infty}^x e^{-\frac{z^2}{2ct}} dz$$

On reconnaît la loi normale de variance ct et de moyenne nulle (c est une constante qui résume les caractéristiques physiques du problème. Elle a la dimension d'une action).

Posons $F_n(x) = F(x, n)$. Lorsque n (donc t) tend vers l'infini, on vérifie facilement qu'en tout point x fixe on a :

$$\lim_{n \rightarrow \infty} F_n(x) = \frac{1}{2}$$

Ainsi $F_n(x)$ converge en tout point vers la fonction constante $F(x) = \frac{1}{2}$. Mais $F(x)$ n'est pas une loi de probabilité, et il n'y a pas convergence en loi, car $F(x)$ ne vérifie pas la propriété 2 du paragraphe 2.

En termes intuitifs, le résultat observé conserve un sens physique ! à la limite où t devient infini, il y a une probabilité nulle pour que la particule reste à distance finie, et une probabilité $\frac{1}{2}$ pour qu'elle se soit enfuie vers $-\infty$ ou vers $+\infty$.

La notion de convergence en loi sera approfondie lorsque la notion de

fonction-caractéristique aura été introduite. Nous nous contenterons ici d'établir deux théorèmes fondamentaux de Helly.

Premier théorème de Helly -- De tout ensemble infini de fonctions de répartition il est possible d'extraire une suite partielle qui converge vers une fonction non décroissante $F(x)$ en tout point x où $F(x)$ est continue, $F(x)$ étant elle-même continue à gauche en tout point.

Comme toute fonction de répartition reste comprise entre 0 et + 1, on aura, par passage à la limite $0 \leq F(x) \leq 1$. La fonction limite $F(x)$ non décroissante étant nécessairement bornée possédera, comme nous l'avons vu, au plus un ensemble dénombrable de points de discontinuité, de sorte que la convergence annoncée aura lieu partout sauf, au plus, sur un ensemble dénombrable.

Il suffit d'établir le théorème de Helly dans le cas où l'ensemble infini de fonctions de répartitions est une suite illimitée (dénombrable) $F_1, F_2 \dots F_n, \dots$. La démonstration va s'appuyer sur le lemme suivant :

Lemme. Pour qu'une suite $F_1, \dots, F_n \dots$ de fonctions de répartition converge vers une fonction non décroissante $F(x)$ en tout point de continuité de $F(x)$ il faut et il suffit qu'il existe un ensemble D dénombrable et partout dense tel que la convergence ait lieu en tout point x de D .

On rappelle que D est partout dense si : quel que soit $\varepsilon > 0$ et quel que soit le nombre réel x_0 il existe au moins un point x appartenant à D tel que $|x - x_0| \leq \varepsilon$.

La nécessité de la condition énoncée résulte immédiatement du fait que les points de discontinuités de $F(x)$ constituent un ensemble dénombrable. Son complémentaire (qui a la puissance du continu) contient certainement un sous-ensemble dénombrable et partout dense.

Montrons que la condition est suffisante. Soit x un point quelconque, et x' , et x'' deux points de D tels que : $x' \leq x \leq x''$. Pour tout n , on a (puisque les F_n sont non décroissantes) :

$$(26) \quad F_n(x') \leq F_n(x) \leq F_n(x'')$$

Passons à la limite. On rappelle que toute suite u_n bornée possède une limite inférieure, notée :

$$\underline{\lim} u_n = \lim_{n \rightarrow \infty} \left[\text{Inf}_{k \geq n} u_k \right] = \text{Sup}_n \left[\text{Inf}_{k \geq n} u_k \right]$$

et une limite supérieure notée

$$\overline{\lim} u_n = \lim_{n \rightarrow \infty} \left[\text{Sup}_{k \geq n} u_k \right] = \text{Inf}_n \left[\text{Sup}_{k \geq n} u_k \right]$$

En effet, la borne inférieure des u_k tels que $k \geq n$ est une fonction décroissante de n et reste bornée : elle a bien une limite pour n infinie, qui coïncide avec la borne supérieure en n de $\text{Inf}_{k \geq n} u_k$. Pour que la suite u_n converge, il faut et il suffit que limites inférieure et supérieure coïncident, et on a alors :

$$\underline{\lim} u_n = \overline{\lim} u_n = \lim u_n$$

Dans le cas général, on a seulement

$$\underline{\lim} u_n \leq \overline{\lim} u_n$$

Par application de ces résultats à (26), il vient :

$$F(x^0) \leq \underline{\lim} F_n(x) \leq \overline{\lim} F_n(x) \leq F(x'')$$

Supposons que x soit un point de continuité de $F(x)$, et soit $\varepsilon > 0$.
Il existe η tel que

$$|y - x| \leq \eta \quad \text{entraîne} \quad |F(x) - F(y)| \leq \varepsilon$$

Comme D est partout dense, on peut prendre x' et x'' tels que :

$$\begin{aligned} x - x' &\leq \eta \\ x' - x'' &\leq \eta \end{aligned}$$

On a alors, puisque F est non décroissante :

$$F(x) - \varepsilon \leq F(x') \leq F(x) \leq F(x'') \leq F(x) + \varepsilon$$

Par suite on a aussi :

$$F(x) - \varepsilon \leq \underline{\lim} F_n(x) \leq \overline{\lim} F_n(x) \leq F(x) + \varepsilon$$

Comme ε est arbitraire, on en déduit

$$F(x) = \underline{\lim} F_n(x) = \overline{\lim} F_n(x) = \lim F_n(x)$$

Le lemme est ainsi démontré.

D'après le lemme, pour établir le théorème de Helly, il faut montrer qu'il existe une fonction $F(x)$ non décroissante et une suite partielle $F_{n_1} \dots F_{n_k} \dots$ extraite de F_n telle que : $F_{n_k}(x) \longrightarrow F(x)$ en tout point x d'un ensemble dénombrable partout dense D (par exemple l'ensemble des nombres rationnels).

a) Soit a_1, a_2, \dots les points de D (une telle indexation est possible, puisque D est dénombrable). En a_1 , les $F_n(a_1)$ constituent une suite infinie et bornée. On peut en extraire une suite partielle $F_{n_1}^1(a_1)$ convergeant vers une certaine limite que nous noterons $G(a_1)$.

En a_2 , de même les $F_n^1(a_2)$ constituent une suite infinie et bornée d'où l'on peut extraire une suite partielle $F_{n_2}^2(a_2)$ convergeant vers une limite notée $G(a_2)$.

Par extractions successives, on obtient ainsi, pour tout k , une suite partielle $F_n^k(x)$ telle que : $F_n^k(a_i)$ converge vers $G(a_i)$ pour $i = 1, 2, \dots, k$.

b) Conformément à un procédé très usuel, considérons la suite diagonale $G_n(x) = F_n^n(x)$, qui est encore une suite partielle extraite de F_n . Quel que soit k , on a :

$$\lim G_n(a_k) = G(a_k) \quad \text{pour } n \longrightarrow \infty$$

En effet, pour $n \geq k$, les $G_n(a_k)$ sont une suite partielle extraite de $F_n^k(a_k)$, laquelle converge vers $G(a_k)$ par construction.

c) On a ainsi extrait de F_n une suite partielle G_n qui converge, en tout point x de D , vers une fonction $G(x)$ définie sur D . Par ailleurs, chaque $G_n(x)$ étant, sur D , fonction non décroissante et bornée de x , il en est de même de la limite $G(x)$. On peut alors prolonger $G(x)$ sur la droite entière, puisque D est partout

dense. Alors, d'après le lemme, les $G_n(x)$ convergent vers $G(x)$ en tout point x où $G(x)$ est continue.

- d) La fonction $G(x)$ n'est pas forcément continue à gauche, mais il est possible de la modifier en posant $F(x) = G(x-0)$: cette modification ne concerne que l'ensemble dénombrable des points de discontinuité, et n'affecte donc pas la convergence des G_n vers F aux points de continuité.

Remarque : le théorème ne dit pas, et on n'a pas démontré, que les G_n convergent en loi. La limite $F(x)$ est bien non décroissante et continue à gauche. Pour que ce soit une fonction de répartition (pour qu'il y ait convergence en loi) il faudrait montrer que l'on a :

$$\begin{aligned} \lim_{x \rightarrow -\infty} F(x) &= 0 && \text{pour } x \rightarrow -\infty \\ \lim_{x \rightarrow +\infty} F(x) &= +1 && \text{pour } x \rightarrow +\infty \end{aligned}$$

Du fait de $0 \leq G_n(x) \leq 1$, on a pour tout x : $0 \leq F(x) \leq 1$, d'où l'on tire seulement :

$$\begin{aligned} \lim_{x \rightarrow -\infty} F(x) &\geq 0 && \text{pour } x \rightarrow -\infty \\ \lim_{x \rightarrow +\infty} F(x) &\leq 1 && \text{pour } x \rightarrow +\infty \end{aligned}$$

Les inégalités peuvent être strictes (effet de fuite).

Deuxième théorème de Helly - Soient $g(x)$ une fonction continue, et $F_n(x)$ une suite de fonctions de répartition convergeant vers une fonction $F(x)$ en tout point de continuité de $F(x)$. Soit (a, b) un intervalle fini tel que a et b soient des points de continuité de $F(x)$. Alors $g(x)$ est intégrable pour $F(x)$ et on a :

$$(27) \quad \lim_{n \rightarrow \infty} \int_a^b g(x) dF_n(x) = \int_a^b g(x) dF(x)$$

Si, de plus, la convergence a lieu en loi, c'est-à-dire si $F(x)$ est elle-même une fonction de répartition, et si $g(x)$ est bornée par un nombre B , alors on a également :

$$(28) \quad \lim_{n \rightarrow \infty} \int_{-\infty}^{+\infty} g(x) dF_n(x) = \int_{-\infty}^{+\infty} g(x) dF(x)$$

a) Démontrons (27). - Étant continue sur l'intervalle fini (a, b) , $g(x)$ y est borné (donc intégrable pour les F_n et F) et uniformément continue : pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que :

$$|x-x'| \leq \eta \implies |g(x) - g(x')| \leq \varepsilon, \text{ quels que soient } x \text{ et } x' \text{ sur } (a, b)$$

Prenons m points $x_k : a = x_1 < x_2 \dots < x_m = b$ tels que $(x_k - x_{k+1}) < \eta$ et que chaque x_k soit un point de continuité de $F(x)$.

Définissons une fonction $h(x)$ par :

$$h(x) = g(x_k) \text{ pour } x_k \leq x < x_{k+1}$$

En tout point x de (a, b) on a par construction :

$$|h(x) - g(x)| \leq \varepsilon$$

D'autre part, les x_k étant des points de continuité de $F(x)$, il existe N tel que :

$$n \geq N \implies |F(x_k) - F_n(x_k)| \leq \frac{\varepsilon}{2m} \text{ pour tous les } x_k$$

Or, on a manifestement :

$$\begin{aligned} \left| \int_a^b g(x) dF - \int_a^b g(x) dF_n \right| &\leq \left| \int_a^b [g(x) - h(x)] dF \right| \\ &+ \left| \int_a^b h(x) dF - \int_a^b h(x) dF_n \right| + \left| \int_a^b [h(x) - g(x)] dF_n \right| \end{aligned}$$

Comme $\int_a^b dF_n \leq 1$ et $\int_a^b dF \leq 1$, la première et la 3ième intégrale sont bornées par ε . Pour majorer le terme intermédiaire, remarquons qu'il est égal à :

$$\begin{aligned} &\left| \sum_{k=0}^{m-1} f(x_k) [F(x_{k+1}) - F(x_k)] - \sum_{k=0}^{m-1} f(x_k) [F_n(x_{k+1}) - F_n(x_k)] \right| \\ &= \left| \sum_{k=0}^{m-1} f(x_k) [F(x_{k+1}) - F_n(x_{k+1})] - \sum_{k=0}^{m-1} f(x_k) [F(x_k) - F_n(x_k)] \right| \end{aligned}$$

Pour $n \geq N$, ce terme est majoré par 2ε , et par suite :

$$\left| \int_a^b g(x) dF - \int_a^b g(x) dF_n \right| \leq 4 \varepsilon$$

ce qui démontre (27)

Corollaire 1 - Si $g(x,u)$ est une fonction continue de l'ensemble des deux variables (x,u) ,

on a aussi :

$$\lim_{n \rightarrow \infty} \int_a^b g(x,u) dF_n(x) = \int_a^b g(x,u) dF(x)$$

et cette convergence est uniforme sur tout intervalle fini de variation de u .

En effet, sur tout domaine rectangulaire fini $(a,b) \times (u_0, u'_0)$, $g(x,u)$ est borné par un nombre B fixe, et uniformément continu : quel que soit ε , il existe η indépendant de u tel que :

$$|x - x'| \leq \eta \implies |g(x) - g(x')| \leq \varepsilon \text{ pour tout } x \text{ de } (a,b) \text{ et tout } u \text{ de } (u_0, u'_0).$$

La démonstration se déroule comme ci-dessus, N étant indépendant de u .

b) Démontrons (28) - Comme F est une fonction de répartition, on peut trouver deux points a et b pour lesquels $F(x)$ est continu et tels que :

$$F(a) \leq \frac{\varepsilon}{B} \quad 1 - F(b) \leq \frac{\varepsilon}{B}$$

D'autre part (a et b étant des points de continuité), il existe N_0 tel que :

$$n \geq N_0 \implies |F(a) - F_n(a)| \leq \frac{\varepsilon}{B} \quad \text{et} \quad |F(b) - F_n(b)| \leq \frac{\varepsilon}{B}$$

Pour $n \geq N_0$, on a alors

$$|F_n(a)| \leq \frac{2\varepsilon}{B} \quad \text{et} \quad |1 - F_n(b)| \leq \frac{2\varepsilon}{B}$$

Majorons la différence :

$$\left| \int_{-\infty}^{+\infty} g(x) dF - \int_{-\infty}^{+\infty} g(x) dF_n \right| \leq \left| \int_{-\infty}^a g(x) (dF - dF_n) \right| \\ + \left| \int_a^b g(x) (dF - dF_n) \right| + \left| \int_b^{\infty} g(x) (dF - dF_n) \right|$$

Le premier et le troisième termes sont majorés chacun par 3ε . Pour le terme central, la première partie de la démonstration montre qu'il est majoré par ε dès que n est supérieur à un certain N . Prenant alors $n \geq \text{Sup}(N_0, N)$, on a bien :

$$\left| \int_{-\infty}^{+\infty} g(x) dF - \int_{-\infty}^{+\infty} g(x) dF_n \right| \leq 7\varepsilon$$

Corollaire 2 - Si la convergence des F_n a lieu en loi et si $g(x,u)$ est une fonction continue de l'ensemble des deux variables x et u bornée par un nombre fixe B , on a :

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{+\infty} g(x,u) dF_n(x) = \int_{-\infty}^{+\infty} g(x,u) dF(x)$$

et cette convergence est uniforme en u sur tout intervalle fini de variation de u .

Même démonstration que pour le premier corollaire. On reprend la démonstration précédente en remarquant que N et N_0 ne dépendent pas de u .

CHAPITRE III

LA FONCTION CARACTERISTIQUE

I. - DEFINITION ET PROPRIETES.

L'exponentielle complexe $e^{iux} = \cos ux + i \sin ux$ est continue et bornée. Alors, quelle que soit la variable aléatoire X de fonction de répartition $F(x)$, la propriété 6 (paragraphe 5, ch.II) de l'intégrale de STIELTJES montre que l'intégrale :

$$\int e^{iux} d F(x) = \int \cos ux d F(x) + i \int \sin ux d F(x)$$

existe toujours et coïncide, comme nous l'avons vu, avec l'espérance mathématique de la variable aléatoire (complexe) e^{iux} .

Définition - On appelle fonction caractéristique d'une variable aléatoire X de fonction de répartition $F(x)$ la fonction $\Phi(u)$ à valeurs complexes définie par :

$$(1) \quad \Phi(u) = E(e^{iuX}) = \int e^{iux} d F(x)$$

Remarque

Si la loi F admet une densité de probabilité $f(x)$, on a :

$$(2) \quad \Phi(u) = \int_{-\infty}^{+\infty} e^{iux} f(x) dx$$

On reconnaît, sur la relation (2), la transformée de Fourier de la fonction $f(x)$. En fait, la définition (1) elle-même constitue une généralisation de la transformation de Fourier : c'est la transformation de Fourier des mesures (probabilités) représentées ici par leurs fonctions de répartition.

Propriété 1 - La fonction caractéristique $\Phi(u)$ est uniformément continue sur toute la droite. De plus, on a :

$$(2) \quad \begin{cases} \Phi(0) = 1 \\ |\Phi(u)| \leq 1 & \text{pour tout } u \\ \Phi(-u) = \overline{\Phi(u)} \end{cases}$$

En effet, de (1) on tire immédiatement

$$\Phi(u+h) - \Phi(u) = \int e^{iux} (e^{ihx} - 1) dF(x)$$

d'où l'inégalité :

$$(3) \quad |\Phi(u+h) - \Phi(u)| \leq \int |e^{iux} (e^{ihx} - 1)| dF(x) = \int |e^{ihx} - 1| dF(x)$$

Soit $\varepsilon > 0$ donné. Comme $\int dF(x) = 1$, on peut trouver un nombre a tel que :

$$\int_{|x| \geq a} dF(x) \leq \frac{\varepsilon}{4}$$

D'autre part, e^{ihx} est uniformément continue en h pour $-a < x < a$, de sorte qu'il existe η tel que : pour $|h| < \eta$ et $|x| < a$, on ait :

$$|e^{ihx} - 1| < \frac{\varepsilon}{2}$$

Alors, pour $h < \eta$, on a aussi :

$$|\Phi(u+h) - \Phi(u)| \leq \int_{-a}^a |e^{ihx} - 1| dF(x) + 2 \int_{|x| > a} dF(x) \leq \varepsilon$$

D'où la continuité uniforme .

La relation $\Phi(0) = 1$ s'obtient immédiatement en faisant $u = 0$ dans la définition (1). En tout point u , on a :

$$|\Phi(u)| = \left| \int e^{iux} dF(x) \right| \leq \int |e^{iux}| dF(x) = \int dF(x) = 1$$

Enfin, en changeant u en $-u$ dans la définition (1), on voit immédiatement que $\Phi(u)$ possède la symétrie hermitique.

$$\Phi(-u) = \overline{\Phi(u)}$$

Propriété 2 - Si $\Phi(u)$ est la fonction caractéristique d'une variable aléatoire X, alors la fonction :

$$(4) \quad \psi(u) = e^{ibu} \Phi(au)$$

est la fonction caractéristique de la variable aléatoire $aX + b$ (a et b étant des nombres réels quelconques.)

En effet, on a :

$$E \left[e^{iu(ax+b)} \right] = e^{iub} E \left[e^{iuaX} \right] = e^{iub} \Phi(au)$$

Propriété 3 - Si X et Y sont deux variables aléatoires indépendantes, admettant $\Phi_1(u)$ et $\Phi_2(u)$ comme fonctions caractéristiques, alors la variable aléatoire $X + Y$ admet $\Phi_1(u) \Phi_2(u)$ comme fonction caractéristique.

En effet, d'après les propriétés de l'espérance mathématique (ch.II, parag.6) on a : X et Y étant indépendantes :

$$E (e^{iuX} e^{iuY}) = E (e^{iuX}) E (e^{iuY}) = \Phi_1(u) \Phi_2(u)$$

On a vu (ch.II, parag.5) que, F_1 et F_2 désignant les fonctions de répartition de X et Y, la somme $X + Y$ admet la fonction de répartition F_3 :

$$F_3(z) = \int F_1(z-x) dF_2(x)$$

qui représente le produit de convolution des probabilités de X et Y.

La propriété qui vient d'être établie peut donc s'énoncer comme suit :

La transformation de Fourier échange les produits convolutifs et multiplicatifs.

Plus exactement, nous avons établi que la transformée du produit de convolution de deux mesures (probabilités) est égale au produit (ordinaire) de leurs transformées. La réciproque résultera du théorème général de réciprocity qui sera énoncé plus loin. Cette propriété de la transformation de Fourier est fondamentale pour la physique mathématique.

Propriété 4 - Si la loi $F(x)$ admet un moment absolu d'ordre n , alors sa fonction caractéristique Φ est n fois dérivable, et on a :

$$(5) \quad m_k = E(x^k) = (-1)^k i^k \Phi^{(k)}(0)$$

En effet, dérivant k fois sous le signe somme l'expression (1), on obtient formellement :

$$\Phi^{(k)}(u) = i^k \int x^k e^{iux} dF(x)$$

Or l'intégrale obtenue est absolument convergente, car $\int |x|^k dF$ est le moment absolu d'ordre $k \leq n$, donc elle représente réellement la $k^{\text{ième}}$ dérivée de $\Phi(u)$. Il suffit de faire $u = 0$ pour obtenir le résultat cherché.

Cette propriété admet une réciproque :

Propriété 5 - Si $\Phi(u)$ admet en $u = 0$ une dérivée d'ordre pair $2n$, le moment absolu d'ordre $2n$ existe ainsi que tous les moments d'ordre $k \leq 2n$, qui sont alors donnés par la formule (5).

En effet, on a :

$$\begin{aligned} \Phi^{(2n)}(0) &= \lim_{u \rightarrow 0} \int \left(\frac{e^{iux} - e^{-iux}}{2u} \right)^{2n} dF(x) \\ &= (-1)^n \lim_{u \rightarrow 0} \int \left(\frac{\sin ux}{u} \right)^{2n} dF(x) \end{aligned}$$

Comme $\left| \frac{\sin ux}{u} \right| \leq 1$ le théorème de Lebesgue donne, pour tout intervalle fini (a, b) :

$$\lim_{u \rightarrow 0} \int_a^b \left(\frac{\sin ux}{u} \right)^{2n} dF(x) = \int_a^b x^{2n} dF(x) \leq |\Phi^{(2n)}(0)|$$

Alors le moment d'ordre $2n$ (qui est aussi le moment absolu d'ordre $2n$) existe, et il suffit d'appliquer la propriété 4.

Ces deux propriétés montrent le lien étroit qui existe entre la régularité à l'infini d'une mesure (probabilité) et la régularité à l'origine de sa transformée de Fourier.

Définition : Caractéristique seconde, et semi-invariant. On appelle caractéristique seconde

$\psi(u)$ d'une loi de probabilité, $\Phi(x)$ le logarithme de sa fonction caractéristique.

$$\psi(u) = \log \Phi(u)$$

On appelle semi-invariant d'ordre k , l'expression

$$c_k = i^k \psi^{(k)}(0)$$

défini si ψ est k fois dérivable. On montre facilement que c_k et m_k existent, ou n'existent pas, simultanément, et que c_k s'exprime en fonction de m_1, m_2, \dots et m_k (et réciproquement). En particulier :

$$\left\{ \begin{array}{l} c_1 = i \psi'(0) = -m_1 \\ c_2 = -\psi''(0) = \sigma^2 \end{array} \right.$$

Exemple : 1 - La loi normale, de densité $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, a comme fonction caractéristique :

$$\Phi(u) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{iux - \frac{x^2}{2}} dx = e^{-\frac{u^2}{2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-iu)^2} dx = e^{-\frac{u^2}{2}}$$

La loi normale la plus générale, de moyenne m et de variance σ^2 a donc, d'après la propriété 2, la fonction caractéristique :

$$\Phi(u) = e^{imu - \frac{\sigma^2}{2} u^2}$$

et la caractéristique seconde remarquablement simple

$$\log \Phi(u) = i m u - \frac{\sigma^2}{2} u^2$$

2 - La loi de Poisson de paramètre θ admet la fonction caractéristique :

$$\Phi(u) = \sum_{n=0}^{\infty} \frac{\theta^n}{n!} e^{-\theta + imu} = e^{\theta(e^{iu} - 1)}$$

3 - La loi gamma de densité $\frac{b^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-bx}$ admet la fonction caractéristique :

$$\Phi(u) = \frac{b^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-bx + iux} dx = \frac{1}{\left[1 - i \frac{u}{b}\right]^\alpha}$$

4 - La loi binomiale de paramètres $p, q = 1 - p$ et n admet la fonction caractéristique :

$$\Phi(u) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} e^{iuk} = (q + p e^{iu})^n$$

II.- FONCTION CARACTERISTIQUE D'UNE LOI A PLUSIEURS VARIABLES.

Les définitions et propriétés énoncées au paragraphe ci-dessus se généralisent aisément au cas de plusieurs variables. Soit $F(x_1, x_2, \dots, x_n)$ la fonction de répartition d'une variable aléatoire $X = (X_1, \dots, X_n)$ à n composantes. On associe à F , comme ci-dessus, sa fonction caractéristique $\Phi(u_1, \dots, u_n)$ définie par :

$$\Phi(u_1, \dots, u_n) = E \left[e^{i \sum_{k=1}^n u_k X_k} \right] = \int e^{i \sum_{k=1}^n u_k x_k} dF(x_1, \dots, x_n)$$

Afin d'abréger les écritures, on désignera par x et u les points de R^n de coordonnées (x_1, \dots, x_n) et (u_1, \dots, u_n) , et par ux le produit scalaire

$$ux = \sum_{k=1}^n u_k x_k$$

Avec ces notations vectorielles, on a simplement

$$(6) \quad \Phi(u) = \int e^{iux} dF(x)$$

Les propriétés suivantes s'obtiennent par généralisation immédiate des propriétés obtenues dans le cas d'une variable.

Propriété 1 - $\Phi(u)$ est uniformément continue dans R^n , et on a :

$$\left\{ \begin{array}{l} \Phi(0) = 1 \\ |\Phi(u)| \leq 1 \\ \Phi(-u) = \overline{\Phi(u)} \end{array} \right.$$

Propriété 2 - Soit $\Phi(u)$ la fonction caractéristique de la variable à n composantes $X = (X_1, \dots, X_n)$. Soit le changement de variables :

$$Y_\ell = \sum_{k=1}^n X_k a_{k\ell} - b_\ell \quad (\ell = 1, 2, \dots, m)$$

définissant une nouvelle variable $Y = (Y_1, \dots, Y_m)$ à m composantes. En notation matricielle, A désignant la matrice des $a_{k\ell}$:

$$Y = X A - b$$

La variable aléatoire vectorielle admet la fonction caractéristique :

$$(7) \quad \psi(u) = e^{-iub} \Phi(Au)$$

En effet, avec des notations indicielles, on a par définition :

$$\psi(u) = E \left[e^{i \sum_{\ell=1}^m u_\ell Y_\ell} \right] = e^{-i \sum_{\ell=1}^m u_\ell b_\ell} E \left[e^{i \sum_{k=1}^n x_k \sum_{\ell=1}^m a_{k\ell} u_\ell} \right]$$

D'où

$$\psi(u) = e^{-i \sum_{\ell=1}^m u_\ell b_\ell} \Phi \left(\sum_{\ell=1}^m a_{k\ell} u_\ell \right)$$

Propriété 3 - Si X et Y sont des variables indépendantes à n composantes, la fonction caractéristique de leur somme est le produit de leurs fonctions caractéristiques.

Propriété 4 - Pourvu que le moment absolu correspondant existe, le moment $E \left[x_1^{k_1} x_2^{k_2} \dots x_n^{k_n} \right]$ s'obtient en faisant $u_1 = \dots = u_n = 0$ dans l'expression :

$$(-i) \sum_{\ell=1}^n k_{\ell} \frac{\partial^{k_1 + \dots + k_n}}{\partial x_1^{k_1} \dots \partial x_n^{k_n}} \overline{\Phi}(u_1, \dots, u_n)$$

Propriété 5 - La fonction caractéristique de la loi marginale $F_{i_1 \dots i_k}(x_{i_1} \dots x_{i_k})$ s'obtient en annulant dans $\overline{\Phi}(u_1, \dots, u_n)$ tous les u_i pour lesquels l'indice i est différent de i_1, \dots, i_k .

Par exemple, si $\overline{\Phi}(u, v)$ est la fonction caractéristique de la loi à deux variables $F(x, y)$, les lois marginales $F_1(x)$ et $F_2(y)$ ont les fonctions caractéristiques suivantes :

$$\left\{ \begin{array}{l} \overline{\Phi}_1(u) = \overline{\Phi}(u, 0) \\ \overline{\Phi}_2(v) = \overline{\Phi}(0, v) \end{array} \right.$$

Démonstration immédiate.

Propriété 6 - Si X_1, \dots, X_n sont n variables aléatoires (non nécessairement indépendantes) et $\overline{\Phi}(u_1, \dots, u_n)$ leur fonction caractéristique, alors $\overline{\Phi}(u, u, \dots, u)$ est la fonction caractéristique de la somme $X_1 + \dots + X_n$.

En effet, on a par définition :

$$E \left[e^{iu \sum_{k=1}^n X_k} \right] = \overline{\Phi}(u, u, \dots, u)$$

La propriété 3 est un cas particulier de la propriété 6, puisque si les variables sont indépendantes $\overline{\Phi}(u_1, \dots, u_n)$ est de la forme :

$$\overline{\Phi}_1(u_1) \overline{\Phi}_2(u_2) \dots \overline{\Phi}_n(u_n)$$

Caractéristique seconde. Le logarithme de la fonction caractéristique sera encore appelé caractéristique seconde. Les m_k et $\sigma_{k\ell}$ désignant les valeurs probables et les covariances (si elles existent), on aura :

$$\text{avec } \psi(u) = \log \overline{\Phi}(u). \quad \begin{cases} \left[\frac{\partial \psi}{\partial u_k} \right]_{u=0} = i m_k \\ \left[\frac{\partial^2 \psi}{\partial u_k \partial u_{\ell}} \right]_{u=0} = - \sigma_{k\ell} \end{cases}$$

Exemple : loi normale à n variables. Sa caractéristique seconde est :

$$\Psi(u) = \log \Phi(u) = i \sum_{k=1}^n u_k m_k - \frac{1}{2} \sum_{ij} u_k u_l \sigma_{kl}$$

III.- PROPRIETE FONDAMENTALE DE LA FONCTION CARACTERISTIQUE

L'importance de la fonction caractéristique en calcul des probabilités découle de la propriété fondamentale suivante :

Théorème Fondamental. La correspondance qui à toute fonction de répartition F associe sa fonction caractéristique Φ est bijective et bicontinue.

Pour démontrer la première partie de l'énoncé (la correspondance est bijective), il faut établir d'une part qu'à toute fonction F correspond une fonction Φ : cela résulte de la définition et du fait que $E(e^{iux})$ existe toujours; de l'autre que toute fonction caractéristique Φ (c'est-à-dire toute fonction Φ qui est la caractéristique d'une loi F) est fonction caractéristique d'une seule loi F: autrement dit, il faut démontrer que la donnée de Φ suffit à déterminer complètement la loi F. Ce point fera l'objet du théorème dit d'unicité.

La deuxième partie de l'énoncé (la correspondance est bicontinue) doit être entendue dans le sens suivant : si une suite F_n de fonctions de répartition converge en loi vers une fonction de répartition F, alors les fonctions caractéristiques Φ_n des F_n convergent vers la fonction caractéristique Φ de F (au sens d'une convergence que nous devons définir). Et, réciproquement, si les Φ_n tendent vers Φ au sens de cette convergence, alors les F_n convergent en loi vers F.

Autrement dit, les problèmes de convergence en loi et de lois limites pourront s'étudier directement sur les fonctions caractéristiques.

Le théorème fondamental s'applique aux lois à plusieurs variables. C'est uniquement pour abrégier les notations que nous l'établirons dans le cas d'une seule variable.

Théorème d'unicité. Une fonction de répartition F est complètement déterminée par sa fonction caractéristique, et on a la formule dite de réciprocity :

$$(8) \quad F(a+h) - F(a-h) = \lim_{T \rightarrow \infty} \frac{1}{\pi} \int_{-T}^T \frac{\sin hu}{u} e^{-iua} \Phi(u) du$$

pourvu que F(x) soit continue en a+h et en a-h.

Il suffit d'établir la formule (8). La première partie du théorème en est une conséquence immédiate puisque F(x) est déterminée dès que l'on connaît sa valeur en tout point de continuité. Posons :

$$I = \frac{1}{\pi} \int_{-T}^T \frac{\sin hu}{u} e^{-iua} \Phi(u) du = \frac{1}{\pi} \int_{-T}^T \frac{\sin hu}{u} e^{-iua} du \int_{-\infty}^{+\infty} e^{iux} dF(x)$$

La fonction $\frac{\sin hu}{u} e^{-iua}$ est bornée en module par h, et l'intégrale I, est absolument convergente : le théorème de Fubini permet d'intervertir l'ordre des intégrations. Soit :

$$I = \frac{1}{\pi} \int_{-\infty}^{+\infty} dF(x) \int_{-T}^T \frac{\sin hu}{u} e^{iu(x-a)} du$$

On en tire :

$$I = \int_{-\infty}^{+\infty} K(x, T) dF(x)$$

avec

$$\begin{aligned} K(x, T) &= \frac{2}{\pi} \int_0^T \frac{\sin hu}{u} \cos[(x-a)u] du \\ &= \frac{1}{\pi} \int_0^T \frac{\sin(x-a+h)u}{u} du - \frac{1}{\pi} \int_0^T \frac{\sin(x-a-h)u}{u} du \end{aligned}$$

Or, la fonction S(h, t) définie par :

$$S(h, T) = \frac{2}{\pi} \int_0^T \frac{\sin hu}{u} du = \frac{2}{\pi} \int_0^{hT} \frac{\sin t}{t} dt$$

est bornée (pour $hT > 0$) et tend vers 1 pour $T \rightarrow \infty$, d'où :

$$\lim_{T \rightarrow \infty} S(h, T) = \begin{cases} 1 & \text{pour } h > 0 \\ 0 & \text{pour } h = 0 \\ -1 & \text{pour } h < 0 \end{cases}$$

Alors la fonction $K(x, T)$, qui s'écrit :

$$K(x, T) = \frac{1}{2} S(x-a+h, T) - \frac{1}{2} S(x-a-h, T)$$

est bornée en module par une constante et vérifie :

$$\lim_{T \rightarrow \infty} K(x, T) = \begin{cases} 0 & \text{pour } x < a-h \\ \frac{1}{2} & \text{pour } x = a-h \\ 1 & \text{pour } a-h < x < a+h \\ \frac{1}{2} & \text{pour } x = a+h \\ 0 & \text{pour } x > a+h \end{cases}$$

Le théorème de Lebesgue, applicable puisque K est bornée par une constante indépendante de T , donne :

$$\lim_{T \rightarrow \infty} \int K(x, T) dF(x) = \int \left[\lim_{T \rightarrow \infty} K(x, T) \right] dF(x) = F(a+h) - F(a-h)$$

Le théorème est donc démontré.

Corollaire : Si la fonction caractéristique $\Phi(u)$ est absolument intégrable de $-\infty$ à $+\infty$, alors la fonction de répartition $F(x)$ est absolument continue, et sa dérivée $f(x)$ est continue et égale à :

$$(9) \quad f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-iux} \Phi(u) du$$

En effet, si $x+h$ et $x-h$ sont des points où F est continue, on a :

$$\frac{F(x+h) - F(x-h)}{2h} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\sin hu}{hu} e^{-iux} \Phi(u) du$$

La fonction sous le signe somme est majorée en module par $|\Phi(u)|$, qui est intégrable par hypothèse. Donc, quand h tend vers 0, l'intégrale tend vers la limite $\frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-iux} \Phi(u) du$. Par suite le premier membre tend vers une limite, qui est par définition la dérivée $f(x)$ de $F(x)$ et on a la relation (9). De plus, comme l'intégrale donnant l'expression de $f(x)$ est absolument convergente, $f(x)$ est continue.

Remarque. On reconnaît sur l'équation (9) la formule classique de l'inversion de la transformation de Fourier. La formule (8) généralise cette inversion à la transformation de Fourier des mesures (probabilités) ici représentées par leurs fonctions de répartition.

Premier Théorème limite. Si une suite $F_n(x)$ de fonctions de répartition converge en loi vers une fonction de répartition $F(x)$, alors la suite $\Phi_n(u)$ converge uniformément sur tout intervalle fini vers la fonction caractéristique $\Phi(u)$ de $F(x)$.

En effet, e^{iux} étant bornée sur toute la droite, le deuxième théorème de Helly s'applique et montre que :

$$\lim_{n \rightarrow \infty} \int e^{iux} dF_n(x) = \int e^{iux} dF(x)$$

Le corollaire 2 de ce même théorème montre que cette convergence est uniforme sur tout intervalle fini de variations de u .

Mais c'est la réciproque de ce théorème qui présente le plus grand intérêt.

Théorème Réciproque 1. - Soient $F_n(x)$ une suite de fonctions de répartition, et $\Phi_n(u)$ leurs fonctions caractéristiques. Si la suite des $\Phi_n(u)$ converge vers une fonction $\Phi(u)$ et si cette convergence est uniforme sur tout intervalle fini, alors la suite F_n converge en loi vers une fonction de répartition F dont la fonction caractéristique est $\Phi(u)$

2.- Si l'on suppose simplement que la suite $\Phi_n(u)$ converge vers une fonction $\Phi(u)$ continue en $u = 0$, la conclusion reste vraie.

Ce théorème achève de montrer que la correspondance $F \rightarrow \Phi$ est bicontinue

lorsque l'on prend comme convergences : la convergence en loi pour les fonctions de répartition, et la convergence uniforme sur tout intervalle fini pour les fonctions caractéristiques.

Pour démontrer ce théorème, remarquons qu'il suffit d'établir le deuxième énoncé, car si les $\overline{\Phi}_n(u)$ convergent vers $\overline{\Phi}(u)$ uniformément sur tout intervalle fini, la limite $\overline{\Phi}(u)$ est nécessairement une fonction continue.

Démonstration

a/- D'après le premier théorème de Helly, on peut extraire de la suite F_n une suite partielle $F_{n_1}(x), F_{n_2}(x) \dots$ convergeant vers une fonction $F(x)$ non décroissante et continue à gauche en tout point x où $F(x)$ est continue.

b/- Montrons que $F(x)$ est une fonction de répartition, c'est-à-dire que l'on a :

$$\begin{aligned} \lim_{x \rightarrow -\infty} F(x) &= 0 && \text{pour } x \rightarrow -\infty \\ \lim_{x \rightarrow +\infty} F(x) &= 1 && \text{pour } x \rightarrow +\infty \end{aligned}$$

En effet, désignons par $F(-\infty)$ et $F(+\infty)$ les limites de $F(x)$ pour $x \rightarrow -\infty$ et $x \rightarrow +\infty$ (qui existent, car F est non décroissante et bornée). Comme $0 \leq F_n(x) \leq 1$, on a aussi $0 \leq F(x) \leq 1$ et par suite $F(+\infty) \leq 1$ et $F(-\infty) \geq 0$. Soit

$$\alpha = F(+\infty) - F(-\infty) \quad (\alpha \leq 1)$$

Si l'on montre que $\alpha = 1$, la conclusion sera établie.

Supposons donc que l'on ait $\alpha < 1$, et choisissons ε tel que : $0 < 2\varepsilon < 1 - \alpha$

Comme $\overline{\Phi}_n(0) = 1$, on a aussi $\overline{\Phi}(0) = 1$. $\overline{\Phi}(u)$ étant continue à l'origine, il en est de même de sa partie réelle $\Re \overline{\Phi}(u)$, de sorte qu'il existe $t > 0$, tel que :

$$|u| \leq t \implies 1 - \varepsilon \leq \Re \overline{\Phi}(u)$$

donc aussi :

$$(10) \quad \frac{1}{2t} \left| \int_{-t}^t \overline{\Phi}(u) du \right| \geq 1 - \varepsilon > \alpha + \varepsilon$$

la dernière inégalité étant stricte en raison du choix de ε . Pour parvenir à une contradiction, montrons que l'on a l'inégalité inverse, et pour cela que l'on a, pour n_k assez grand :

$$\left| \frac{1}{2t} \int_{-t}^t \Phi_{n_k}(u) du \right| \leq \alpha + \varepsilon$$

En passant à la limite $n_k \rightarrow \infty$, ce qui est légitime puisque les Φ_{n_k} sont bornés et continus, on obtiendra :

$$\frac{1}{2t} \left| \int_{-t}^t \Phi(u) du \right| \leq \alpha + \varepsilon$$

ce qui montrera que $\alpha = 1$ et que F est bien une fonction de répartition.

Or, on a :

$$\frac{1}{2t} \int_{-t}^t \Phi_{n_k}(u) du = \frac{1}{2t} \int_{-\infty}^{\infty} \left[\int_{-t}^t e^{iux} du \right] dF_{n_k}(x) = \int_{-\infty}^{\infty} \frac{\sin tx}{tx} dF_{n_k}(x)$$

$\frac{\sin tx}{tx}$ est bornée à la fois par 1 et par $\frac{1}{t|x|}$. Ainsi, pour $a > 0$, on a :

$$\left\{ \begin{array}{l} \left| \frac{\sin tx}{tx} \right| \leq 1 \quad \text{pour } |x| \leq a \\ \left| \frac{\sin tx}{tx} \right| \leq \frac{1}{t|x|} \quad \text{pour } |x| > a \end{array} \right.$$

D'où la majoration :

$$\begin{aligned} \left| \int_{-\infty}^{\infty} \frac{\sin tx}{tx} dF_{n_k} \right| &\leq \left| \int_{|x|>a} \frac{\sin tx}{tx} dF_{n_k} \right| + \left| \int_{-a}^a \frac{\sin tx}{tx} dF_{n_k} \right| \\ &\leq \frac{1}{ta} \left[1 - F_{n_k}(a) + F_{n_k}(-a) \right] + F_{n_k}(a) - F_{n_k}(-a) \end{aligned}$$

Soit :

$$\frac{1}{2t} \left| \int_{-t}^t \Phi_{n_k}(u) du \right| \leq \left(1 - \frac{1}{at} \right) \left[F_{n_k}(a) - F_{n_k}(-a) \right] + \frac{1}{at} \leq F_{n_k}(a) - F_{n_k}(-a) + \frac{1}{at}$$

choisissons $a > \frac{2}{\varepsilon t}$ tel que a et $-a$ soient des points de continuité de $F(x)$.

On a :

$$F_{n_k}(a) - F_{n_k}(-a) \leq \left| F_{n_k}(a) - F(a) \right| + \left| F(a) - F(-a) \right| + \left| F_{n_k}(-a) - F(-a) \right|$$

comme a et $-a$ sont des points de continuité, il existe K tel que :

$$n_k \geq K \implies \left| F_{n_k}(a) - F(a) \right| < \frac{\varepsilon}{4} \quad \text{et} \quad \left| F_{n_k}(-a) - F(-a) \right| < \frac{\varepsilon}{4}$$

Alors, pour $n_k \geq K$, on a aussi :

$$F_{n_k}(a) - F_{n_k}(-a) \leq \alpha + \frac{\varepsilon}{2}$$

Compte tenu de $a > \frac{2}{\varepsilon t}$, il vient :

$$\frac{1}{2-t} \left| \int_{-t}^t \Phi_{n_k}(u) du \right| \leq \alpha + \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \alpha + \varepsilon$$

Passant à la limite n_k infini, on a bien l'inégalité complémentaire de (10).
D'où la conclusion : $\alpha = 1$, et $F(x)$ est une fonction de répartition.

c/- Ainsi la suite partielle $F_{n_1}, F_{n_2} \dots$ converge en loi vers $F(x)$. D'après le théorème direct, $\Phi_{n_1}, \Phi_{n_2} \dots$ converge uniformément sur tout intervalle fini vers la fonction caractéristique de $F(x)$, qui coïncide ainsi nécessairement avec $\Phi(u)$.

d/- Il reste à montrer que la suite $F_1, F_2 \dots$ elle-même, et non plus seulement la suite partielle des F_{n_k} , converge en loi vers F . On sait qu'une condition nécessaire et suffisante pour qu'une suite numérique converge vers une limite est que de toute suite partielle (infinie) extraite de celle-ci, on puisse extraire une sous-suite partielle qui converge vers cette limite. Alors, si $F_1, F_2 \dots$ ne convergent pas en loi vers F , il y a une suite partielle extraite de F_n qui ne converge pas vers F . De cette suite partielle, on peut à nouveau (d'après le premier théorème de Helly) extraire une seconde suite, soit G_n , qui converge vers une fonction $G(x)$, distincte de $F(x)$. Mais, d'après les parties b et c de la démonstration précédente, $G(x)$ est une fonction de répartition et admet $\Phi(u)$ comme fonction caractéristique. Le théorème d'unicité donne alors $F(x) = G(x)$. On arrive ainsi à une contradiction.
D'où la conclusion.

Remarque 1. - Dans la démonstration ci-dessus, on a seulement utilisé la continuité en $u = 0$ de la partie réelle de $\overline{\Phi}(u)$. Dans l'énoncé 2, on peut remplacer la condition : $\overline{\Phi}(u)$ continue en $u = 0$, par : Partie réelle de $\overline{\Phi}(u)$ continue en $u = 0$.

2.- Dans l'énoncé 1, on peut remplacer la convergence uniforme sur tout intervalle fini par la convergence uniforme sur un intervalle particulier contenant l'origine. : en effet, cette condition entraîne également la continuité de $\overline{\Phi}(u)$ en $u = 0$.

IV.- FONCTIONS DE TYPE POSITIF.

L'importance des fonctions de type positif est grande en calcul des probabilités. En effet, les fonctions caractéristiques des lois de probabilité sont obligatoirement de type positif, ainsi, comme nous le verrons plus tard, que les fonctions de covariance des fonctions aléatoires ou des processus stochastiques.

Définition. Une fonction $\overline{\Phi}(u)$ de la variable réelle u est dite de type positif si, quels que soient l'entier n , les nombres réels $u_1 \dots u_n$ et les nombres complexes $\lambda_1 \dots \lambda_n$, on a :

$$(11) \quad \sum_{j=1}^n \sum_{k=1}^n \lambda_j \overline{\lambda_k} \overline{\Phi}(u_j - u_k) \geq 0$$

La définition s'applique aussi aux fonctions de plusieurs variables : les u_j sont alors des vecteurs réels, et les λ_k des nombres complexes.

Propriété 1 - Si une fonction $\overline{\Phi}(u)$ est de type positif, on a :

$$(12) \quad \left| \begin{array}{l} \overline{\Phi}(0) \geq 0 \\ \overline{\Phi}(-u) = \overline{\overline{\Phi}(u)} \\ |\overline{\Phi}(u)| \leq \overline{\Phi}(0) \end{array} \right.$$

En effet, avec $n = 1$, $\lambda_1 = 1$ et $u_1 = 0$, (11) donne $\overline{\Phi}(0) \geq 0$.

Avec $n = 2$, $u_1 = 0$, $u_2 = u$, $\lambda_1 = \lambda$, $\lambda_2 = \mu$, il vient :

$$\left[|\lambda|^2 + |\mu|^2 \right] \overline{\Phi}(0) + \overline{\lambda} \mu \overline{\Phi}(u) + \lambda \overline{\mu} \overline{\Phi}(-u) \geq 0$$

En particulier $\bar{\lambda} \mu \Phi(u) + \lambda \bar{\mu} \Phi(-u)$ doit être réel. Avec $\lambda = \mu = 1$, on obtient :

$$\Im(\Phi(u)) + \Im(\Phi(-u)) = 0$$

Avec $\lambda = i$ et $\mu = 1$, on obtient

$$-\Re(\Phi(u)) + \Re(\Phi(-u)) = 0$$

Par suite $\Phi(u)$ possède la symétrie hermitique : $\Phi(-u) = \overline{\Phi(u)}$

Prenant $\lambda = \Phi(u)$ et $\mu = -|\Phi(u)|$, on obtient :

$$2 \Phi(0) |\Phi(u)|^2 - |\Phi(u)|^2 |\Phi(u)| - |\Phi(u)|^2 |\Phi(u)| \geq 0$$

D'où pour $\Phi(u) \neq 0$

$$|\Phi(u)| \leq \Phi(0)$$

inégalité qui est encore vérifiée dans le cas $\Phi(u) = 0$.

Ces trois propriétés sont, comme on l'a vu, vérifiées par les fonctions caractéristiques. Le théorème fondamental suivant va beaucoup plus loin, puisqu'il donne une condition nécessaire et suffisante pour qu'une fonction soit caractéristique.

Théorème de Bochner. Pour qu'une fonction continue $\Phi(u)$ soit une fonction caractéristique, il faut et il suffit que $\Phi(0)$ soit égal à 1 et que $\Phi(u)$ soit de type positif.

a/ La condition est nécessaire. En effet, s'il existe une fonction de répartition $F(x)$ telle que :

$$\Phi(u) = \int e^{iux} dF(x)$$

on a $\Phi(0) = 1$ et

$$\begin{aligned} \sum_{jk} \lambda_j \bar{\lambda}_k \Phi(u_j - u_k) &= \sum_{jk} \lambda_j \bar{\lambda}_k \int e^{ix(u_j - u_k)} dF(x) \\ &= \int \left| \sum_{jk} \lambda_j \bar{\lambda}_k e^{ix(u_j - u_k)} \right|^2 dF(x) \\ &= \int \left| \sum_j \lambda_j e^{iu_j x} \right|^2 dF(x) \geq 0 \end{aligned}$$

b/ La condition est suffisante. La démonstration est beaucoup plus délicate, et ne sera pas donnée ici.

Remarque

Le théorème de Bochner montre que la classe des fonctions caractéristiques s'identifie avec la classe des fonctions continues de type positif, vérifiant de plus la condition de normalisation $\Phi(0) = 1$. En fait, il suffit seulement de vérifier la continuité de $\Phi(u)$ en $u = 0$, car :

Propriété 2 - Pour qu'une fonction de type positif soit continue, il faut et il suffit, qu'elle soit continue en $u = 0$.

Appliquons la relation (11) avec $n = 3$, et

$$\lambda_1 = \lambda \quad \lambda_2 = \mu \quad \lambda_3 = -\mu$$

$$u_1 = u \quad u_2 = -h \quad u_3 = 0$$

Après quelques calculs, on obtient :

$$\begin{aligned} |\lambda|^2 \Phi(0) + \lambda \bar{\mu} [\Phi(u+h) - \Phi(u)] + \bar{\lambda} \mu [\overline{\varphi(u+h) - \varphi(u)}] \\ + \mu^2 [2\varphi(0) - \varphi(h) - \Phi(-h)] \geq 0 \end{aligned}$$

D'une manière générale, pour que l'on ait, quels que soient λ et μ

$$a |\lambda|^2 + b \lambda \bar{\mu} + \bar{b} \bar{\lambda} \mu + c |\mu|^2 \geq 0$$

(a et c réels positifs), on vérifie que la condition

$$|b|^2 \leq a c$$

est nécessaire et suffisante. On a donc

$$(13) \quad \left| \Phi(u+h) - \Phi(u) \right|^2 \leq \Phi(0) [2\Phi(0) - \Phi(h) - \Phi(-h)]$$

Par suite, si $\Phi(u)$ est continue en $u = 0$, elle est nécessairement continue en tout point.

Corollaire

Si $\Phi(u)$ est une fonction caractéristique, on a :

$$|\Phi(u+h) - \Phi(u)|^2 \leq 2 - \Phi(h) - \Phi(-h)$$

Ce n'est pas autre chose que l'inégalité (13) avec $\Phi(0) = 1$

La propriété 2 permet de comprendre pourquoi, dans la réciproque du théorème limite, il était suffisant de supposer $\Phi(u)$ continue en $u = 0$. Les Φ_n étant de type positif, il en était de même de leur limite $\Phi(u)$, et la continuité en $u = 0$ entraînait la continuité en tout point. En effet :

Propriété 3 - Toute limite de fonctions de type positif est de type positif.

Il suffit de passer à la limite dans les inégalités (11).

La propriété 2 suggère que la régularité, en tout point, d'une fonction de type positif est commandée par la régularité en $u = 0$. En effet :

Propriété 4 - Si une fonction de type positif est $2k$ fois dérivable en $u = 0$, elle est $2k$ fois dérivable en tout point.

En effet, si $\Phi(0) = 0$, on a $\Phi(u) \equiv 0$ d'après la propriété 4. Si $\Phi(0) > 0$, on peut supposer $\Phi(0) = 1$, quitte à remplacer $\Phi(u)$ par $\frac{1}{\Phi(0)} \Phi(u)$. Etant dérivable en $u = 0$, Φ y est continue. Le théorème de Bochner $\Phi(0)$ montre que $\Phi(u)$ est fonction caractéristique d'une fonction de répartition $F(x)$:

$$\Phi(u) = \int e^{iux} dF(x)$$

D'après la propriété 5 du paragraphe 1, l'intégrale

$$(-1)^k \int x^{2k} e^{iux} dF(x)$$

est absolument convergente. Elle est donc égale à la dérivée d'ordre $2k$ de $\Phi(u)$.

V.- FONCTIONS GENERATRICES.

Soit X une variable aléatoire discrète, dont les seules valeurs possibles sont les entiers positifs ou nuls, par exemple une variable binomiale ou poissonien-

ne. Dans ce cas, au lieu de la fonction caractéristique $\overline{\Phi}(u)$ on introduit souvent la fonction génératrice $G(s)$ définie par :

$$(14) \quad G(s) = E(s^X) = \sum_{k=0}^{\infty} p_k s^k$$

p_k représentant la probabilité de l'évènement " $X = k$ ". La fonction $G(s)$ est liée à la fonction caractéristique par la relation évidente :

$$(15) \quad G(e^{iu}) = \overline{\Phi}(u)$$

qui a toujours un sens, car la série entière $G(s)$ est absolument et uniformément convergente pour $|s| \leq 1$, en vertu de la relation :

$$\sum_{k=0}^{\infty} p_k = 1$$

En particulier, on a le droit de la dériver terme à terme à l'intérieur du cercle de convergence $|s| < 1$. Énonçons :

Propriété 1 - La fonction génératrice est continue pour $|s| \leq 1$ et indéfiniment dérivable pour $|s| < 1$. La série représentative est uniformément et absolument convergente pour $|s| < 1$, et peut être dérivée terme à terme pour $|s| < 1$.

Propriété 2 - La fonction de répartition d'une variable discrète à valeurs entières positives est déterminée si l'on connaît sa fonction génératrice.

Cela résulte immédiatement de (15) et du théorème d'unicité.

Propriété 3 - Si le moment absolu d'ordre n existe, la fonction génératrice est dérivable n fois en $s = 1$, et tout moment d'ordre $k \leq n$ peut se calculer à partir des valeurs de $G'(1), G''(1), \dots, G^{(k)}(1)$.

Définitions : Moments factoriels. On appelle moment factoriel d'ordre k , et on note $M_{[k]}$ la valeur probable, si elle existe, du produit $X(X-1)\dots(X-k+1)$. Si X est une variable discrète à valeurs entières positives, $M_{[k]}$ existe si et seulement si la fonction génératrice est k fois dérivable en $s = 1$, et on a :

$$(16) \quad M_{[k]} = E \left[X(X-1) - (X-k+1) \right] = G^{(k)}(1)$$

De (16) on déduit la moyenne m et la variance σ^2 , lorsqu'elles existent :

$$(17) \quad \begin{cases} m = G'(1) \\ \sigma^2 = G''(1) + m(1-m) \end{cases}$$

De son côté, le théorème limite nous permet d'énoncer :

Propriété 4 - Soient $X_1, X_2 \dots$ une suite de variables discrètes à valeurs entières et $G_1, G_2 \dots$ leurs fonctions génératrices. Pour que les X_k convergent en loi vers une variable X (obligatoirement discrète et à valeurs entières), il faut et il suffit que les $G_k(s)$ convergent vers la fonction génératrice $G(s)$ en tout point s tel que $0 \leq s < 1$, et que l'on ait $G(1) = 1$.

En effet, dans ce cas, la convergence a lieu dans le plan complexe pour $|s| \leq 1$, et, en particulier, pour $|s| = 1$: les $G_k(e^{iu})$ convergent vers $G(e^{iu})$. D'autre part, $G(e^{iu})$, qui est de la forme $\sum p_n e^{inu}$, est continue puisque par hypothèse $\sum p_n = 1$.

Application : Promenade aléatoire (ou probabilité de ruine).

Une particule est placée au temps $t = 0$ au point d'abscisse entière z ($0 < z < a$). Au temps $t = 1, 2, \dots$ elle subit un déplacement de $+1$ ou -1 avec les probabilités p et $q = 1 - p$ respectivement. En $x = 0$ et $x = a$ (a entier) sont placées des barrières absorbantes. On cherche la probabilité $P_n(z)$ pour que la particule soit absorbée au temps $t = n$ en $x = 0$ (parvienne pour la première fois en $x = 0$ au temps $t = n$, sans avoir atteint $x = a$ aux temps $1, 2, \dots, n-1$).

Par un raisonnement probabiliste, on établit la relation de récurrence :

$$(18) \quad P_n(z) = p P_{n-1}(z+1) + q P_{n-1}(z-1)$$

et on introduit la fonction génératrice

$$G_z(s) = \sum_{n=0}^{\infty} P_n(z) s^n$$

Les $P_n(z)$ vérifient les conditions aux limites

$$\left\{ \begin{array}{l} P_n(0) = P_n(a) = 0 \quad \text{pour } n \geq 1 \\ P_0(0) = 1 \\ P_0(z) = 0 \quad \text{pour } z > 0 \end{array} \right.$$

D'où

$$(19) \quad G_0(s) = 1 \quad G_a(s) = 0$$

On multiplie (18) par s^n et on somme de $n=1$ à l'infini.

$$(20) \quad G_z(s) = p s G_{z+1}(s) + q s G_{z-1}(s)$$

Pour résoudre l'équation aux différences finies (20), on cherche une solution de la forme $[\lambda(s)]^z$, $\lambda(s)$ étant racine de l'équation caractéristique :

$$\lambda(s) = p s \lambda^2(s) + q s$$

Soient

$$\left\{ \begin{array}{l} \lambda_1(s) = \frac{1 + \sqrt{1 - 4 p q s^2}}{2 p s} \\ \lambda_2(s) = \frac{1 - \sqrt{1 - 4 p q s^2}}{2 p s} \end{array} \right.$$

La solution de (20) est de la forme :

$$(21) \quad G_z(s) = A(s) [\lambda_1(s)]^z + B(s) [\lambda_2(s)]^z$$

Compte tenu des conditions aux limites (19), on trouvera :

$$G_z(s) = \left(\frac{q}{p}\right)^z \frac{[\lambda_1(s)]^{a-z} - [\lambda_2(s)]^{a-z}}{[\lambda_1(s)]^a - [\lambda_2(s)]^a}$$

Cas où $a = +\infty$ (une seule barrière absorbante en $x = 0$). La solution générale est encore de la forme (21). Mais (pour $0 < s < 1$) $\lambda_1 > 1$ et $\lambda_2 < 1$. Lorsque z augmente, $G_z(s)$ doit rester borné, d'où $A(s) = 0$ et

$$G_z(s) = [\lambda_2(s)]^z$$

On en déduira

$$P_n(z) = \begin{cases} 0 & \text{pour } n - z \text{ impair.} \\ \frac{z}{n} C_{\frac{n-z}{2}}^n (pq)^{n/2} \left(\frac{q}{p}\right)^z & \text{pour } n-z \text{ pair.} \end{cases}$$

VII.- LOIS COMPOSEES.

Dans beaucoup d'applications, une variable aléatoire X se présente comme dépendant elle-même d'un paramètre aléatoire λ . Autrement dit, la valeur du paramètre λ est donnée, dans une première épreuve, comme résultat d'un tirage au sort selon une loi de probabilité $G(\lambda)$. Dans une deuxième épreuve, λ étant fixé, la valeur prise par X est donnée comme résultat d'un tirage au sort selon une loi $F_\lambda(x)$ dépendant de λ . Si $\Phi_\lambda(u)$ est la fonction caractéristique de la loi $F_\lambda(x)$, on doit avoir

$$(23) \quad \Phi(u) = E(e^{iuX}) = \int \Phi_\lambda(u) dG(\lambda)$$

La fonction $\Phi(u)$ ainsi définie est-elle une fonction caractéristique ?

L'intégrale (23) peut être, plus généralement, remplacée par une intégrale de Lebesgue, qui sera toujours définie si $\Phi_\lambda(u)$ est, à u fixé une fonction mesurable de λ (cela résulte du fait que $\Phi_\lambda(u)$ est bornée par 1 quels que soient u et λ). Par ailleurs, comme $\Phi_\lambda(0) = 1$ on a $\Phi(0) = 1$ et on vérifie immédiatement que $\Phi(u)$ est de type positif (les Φ_λ étant de type positif, il suffit d'intégrer les inégalités (11)). Il reste à vérifier que $\Phi(u)$ est continue en $u = 0$.

Théorème - Si les $\Phi_\lambda(u)$ sont uniformément continues en $u = 0$ pour tout intervalle fini de variation de λ , alors la fonction $\Phi(u)$ définie par la relation (23) est une fonction caractéristique.

En effet, on prendra ℓ tel que :

$$\int_{|\lambda| \geq \ell} dG(\lambda) \leq \varepsilon$$

et η tel que :

$$|u| \leq \eta \implies |1 - \Phi_\lambda(u)| \leq \varepsilon \quad \text{quel que soit } \lambda (-l \leq \lambda \leq l)$$

et on aura

$$|1 - \Phi(u)| \leq 2\varepsilon$$

Cette condition est toujours remplie dans les deux cas suivants :

a/ $\Phi_\lambda(u)$ dépend continûment du paramètre λ .

b/ λ est une variable discrète à valeurs entières.

Si $\Phi(u)$ est une fonction caractéristique, le théorème d'unicité montre qu'elle définit la fonction de répartition $F(x)$ de la variable aléatoire X . D'autre part, la formule de réciprocity permet d'écrire :

$$\begin{aligned} F(a+h) - F(a-h) &= \lim_{T \rightarrow \infty} \frac{1}{\pi} \int_{-T}^T \frac{\sin hu}{u} e^{-iua} \Phi(u) du \\ &= \lim_{T \rightarrow \infty} \frac{1}{\pi} \int dG(\lambda) \int_{-T}^T \Phi_\lambda(u) \frac{\sin hu}{u} e^{-iua} du \end{aligned}$$

La fonction sous le signe somme étant bornée en module, le théorème de Lebesgue montre que le passage à la limite peut se faire sous le signe d'intégration. On en déduit :

$$(24) \quad F(x) = \int F_\lambda(x) dG(\lambda)$$

Ainsi, la fonction de répartition $F(x)$ de la variable composée X est égale à la valeur probable, pour la loi $G(\lambda)$ du paramètre aléatoire λ , de la fonction de répartition $F_\lambda(x)$ de X à λ fixé.

La loi $F_\lambda(x)$ de X à λ fixé, qui nous a permis de construire la loi composée $F(x)$ est-elle identique à la loi conditionnelle $F(x|\lambda)$ de X relativement à λ ? En général, il en sera bien ainsi, sous réserve que la loi conditionnelle $F(x|\lambda)$ soit définissable.

La loi à deux variables de X et de λ admet la fonction de répartition :

$$(25) \quad F(x, \lambda) = \int_{-\infty}^{\lambda} F_{\mu}(x) dG(\mu)$$

dont la loi marginale $F(x, +\infty)$ coïncide bien avec (24). La loi conditionnelle de X pour $\mu_1 \leq \lambda < \mu_2$ est alors :

$$F(x | \mu_1 \leq \lambda < \mu_2) = \frac{F(x, \mu_2) - F(x, \mu_1)}{G(\mu_2) - G(\mu_1)}$$

si $G(\mu)$ est une loi discrète, et μ_1 un point de discontinuité, on obtient :

$$F(x | \lambda = \mu_1) = F_{\mu_1}(x) \frac{G(\mu_1 + 0) - G(\mu_1)}{G(\mu_1 + 0) - G(\mu_1)} = F_{\mu_1}(x)$$

De même, si $G(\mu)$ est absolument continue, et si $g(\mu)$ est sa densité, on obtient par passage à la limite (pourvu que $F_{\mu}(x)$ soit une fonction continue de μ)

$$F(x | \lambda = \mu) = \lim_{h \rightarrow 0} \frac{\int_{\mu}^{\mu+h} F_{\mu}(x) g(\mu) d\mu}{\int_{\mu}^{\mu+h} g(\mu) d\mu} = F_{\mu}(x)$$

Ainsi, dans les deux cas les plus intéressants en pratique ($G(\lambda)$ absolument continue ou discrète), la loi conditionnelle $F(x | \lambda)$ coïncide bien avec la fonction de répartition $F_{\lambda}(x)$.

Exemple - Prenons comme variable λ une variable poissonnienne de paramètre θ , c'est-à-dire $P(\lambda = n) = \frac{\theta^n}{n!} e^{-\theta}$, et posons :

$$X = X_1 + X_2 + \dots + X_n \quad \text{si } \lambda = n$$

les X_i étant n variables aléatoires indépendantes admettant la même loi de probabilité. Soit $\Phi(u_i)$ la fonction caractéristique de X_i . Pour $\lambda = n$, celle de X est :

$$\Phi_n(u) = \left[\Phi(u) \right]^n$$

Appliquons (23). La fonction caractéristique $\Psi(u)$ de la loi composée est :

$$\psi(u) = e^{-\theta} \sum_{k=0}^{\infty} \frac{\theta^k}{k!} [\Phi(u)]^k = e^{\theta[\Phi(u)-1]}$$

Ainsi, si $\Phi(u)$ est une fonction caractéristique, il en est de même de $\exp [\theta(\Phi(u)-1)] = \psi(u)$. Nous verrons ultérieurement que les lois de ce type, obtenues par composition poissonienne, sont indéfiniment divisibles.

*

* *

C H A P I T R E I V

CONVERGENCES ET LOIS DES GRANDS NOMBRES

I.- CONVERGENCE EN PROBABILITE ET LOI ORDINAIRE DES GRANDS NOMBRES.

Outre la convergence en loi, déjà définie, qui est la plus faible des convergences utilisées en calcul des probabilités, l'analyse aléatoire introduit d'autres types de convergences, indispensables déjà pour la formulation des différentes formes de la loi des grands nombres. Nous en définirons trois, à savoir :

- la convergence en probabilité,
- la convergence presque sûre,
- la convergence en moyenne quadratique.

La plus simple, et la plus faible, est la convergence en probabilité. On dira que X_n converge vers X en probabilité, si, lorsque n augmente, la probabilité pour que X_n s'écarte d'une quantité donnée de X tend vers 0.

Définition 1 : Convergence en probabilité. On dit qu'une suite X_n de variables aléatoires converge en probabilité vers une variable X si, quel que soit $\alpha > 0$, $P(|X_n - X| \geq \alpha)$ tend vers 0 lorsque n augmente indéfiniment.

Cette définition s'explique ainsi !

Quels que soient α et ε positifs, on peut trouver un entier N tel que :

$$n \geq N \implies P(|X_n - X| \geq \alpha) \leq \varepsilon$$

Théorème 1 - La convergence en probabilité entraîne la convergence en loi.

Soient X_n et X des variables aléatoires, F_n et F leurs fonctions de répartition. On suppose que l'on a :

$$\lim_{n \rightarrow \infty} P[|X_n - X| \geq \alpha] = 0 \quad \text{quel que soit } \alpha > 0$$

et on veut montrer que $F_n(x)$ converge vers $F(x)$ pour tout x où F est continue.

a/ La variable $Y_n = X_n - X$ converge en loi vers une variable Y presque certainement nulle. En effet, soit $G_n(x)$ la fonction de répartition de Y_n . Par hypothèse, pour tout $\varepsilon > 0$, il existe un entier N tel que :

$$n \geq N \implies 1 - G_n(x) + G_n(-x + \varepsilon) \leq \varepsilon$$

Comme on a aussi

$$0 \leq G_n(-x + \varepsilon) \leq G_n(x) \leq 1$$

il en résulte

$$G_n(x) \geq 1 - \varepsilon$$

$$G_n(-x + \varepsilon) \leq \varepsilon$$

Par suite, en tout point $x \neq 0$ $G_n(x)$ tend vers la fonction de répartition $\delta(x)$ déjà introduite (mesure de Dirac).

$$\delta(x) = \begin{cases} 0 & \text{pour } x \leq 0 \\ 1 & \text{pour } x > 0 \end{cases}$$

b/ Montrons que X_n converge en loi vers X . L'évènement " $X_n < x$ " est contenu dans l'évènement " $X < x + \alpha$ ou $X_n - X < -\alpha$ " ($\alpha > 0$ quelconque). En effet la négation du deuxième : " $X \geq x + \alpha$ et $X_n - X \geq -\alpha$ " entraîne la négation du premier : " $X_n \geq x$ ". Par suite, on a

$$P(X_n < x) \leq P(X < x + \alpha) + P(Y_n < -\alpha)$$

c'est-à-dire

$$F_n(x) \leq F(x + \alpha) + G_n(-\alpha)$$

De même, l'évènement : " $X_n \geq x$ " est contenu dans : " $X \geq x - \alpha$ ou $X_n - X \geq \alpha$ ", d'où résulte :

$$1 - F_n(x) \leq 1 - F(x - \alpha) + 1 - G_n(\alpha)$$

ou encore

$$F_n(x) \geq F(x - \alpha) + G_n(\alpha) - 1$$

Prenons pour x un point de continuité de $F(x)$. D'après la partie a/ de la démonstration, il existe N tel que :

$$n \gg N \implies G_n(-\alpha) \leq \varepsilon \quad \text{et} \quad G_n(\alpha) \geq 1 - \varepsilon$$

On a donc aussi :

$$F(x - \alpha) - \varepsilon \leq F_n(x) \leq F(x + \alpha) + \varepsilon$$

Le choix de $\alpha > 0$ n'a pas été précisé jusqu'ici. Comme x est un point de continuité de F , on pouvait prendre pour α un nombre positif tel que :

$$|x - y| \leq \alpha \implies |F(x) - F(y)| \leq \varepsilon$$

Dans ce cas, on obtient

$$F(x) - 2\varepsilon \leq F_n(x) \leq F(x) + 2\varepsilon$$

ce qui achève la démonstration.

Remarque : la réciproque de ce théorème n'est pas exacte. Il n'est pas vrai que la convergence en loi entraîne la convergence en probabilité, et en ce sens on peut dire que la convergence en loi est strictement plus faible que la convergence en probabilité. On s'en convaincra en prenant comme X_n et X des variables indépendantes obéissant à une même loi de probabilité, par exemple la loi normale de moyenne 0 et de variance unité. Il y a convergence en loi, puisque $F_n(x) = F(x)$. Mais $Y = X_n - X$ est une variable normale de moyenne 0 et de variance 2 et

$$P(|X_n - X| \geq \alpha) = 2 \frac{1}{\sqrt{4\pi}} \int_{\alpha}^{\infty} e^{-\frac{x^2}{4}} dx$$

est différent de 0, et ne peut pas tendre vers 0 lorsque $n \rightarrow \infty$

La loi Ordinaire des grands nombres. La notion de convergence en probabilité conduit à la loi ordinaire (ou faible) des grands nombres. Nous démontrerons simplement l'énoncé suivant :

Théorème 2. Soient X_n une suite de variables indépendantes de variances finies σ_n^2 et d'espérances mathématiques m_n . Si l'on a :

$$\lim_{n \rightarrow \infty} \left[\frac{1}{n^2} \sum_{k=1}^n \sigma_k^2 \right] = 0$$

la variable $Y_n = \frac{1}{n} \sum_{k=1}^n (X_k - m_k)$ converge, en probabilité, vers 0.

En effet, la variable Y_n possède la variance

$$S_n^2 = \frac{1}{n^2} \sum_{k=1}^n \sigma_k^2$$

et l'inégalité de Tchebychev permet d'écrire :

$$P(|Y_n| \geq \alpha) \leq \frac{S_n^2}{\alpha^2}$$

Comme S_n^2 tend vers 0, par hypothèse, il en est de même de $P(|Y_n| \geq \alpha)$, ce qui signifie que Y_n converge en probabilité vers 0.

Corollaire 1-(Théorème de Bernoulli). Dans l'alternative répétée, la fréquence empirique converge, en probabilité, vers la probabilité p .

En effet, soit X_n la variable définie par $P(X_n=1)=p$ et $P(X_n=0)=q=1-p$. X_n possède la variance finie $\sigma_n^2=pq$ indépendante de n , et l'espérance $m_n = p$. Prenant $Y_n = \frac{1}{n} \sum_{k=1}^n X_k$, le théorème précédent montre que Y_n converge en probabilité vers p .

Corollaire 2 - Il n'est pas nécessaire, dans le théorème 2, de supposer les X_n indépendantes. Il suffit d'avoir la condition :

$$\lim_{n \rightarrow \infty} E(Y_n^2) = 0$$

qui exprime que les X_n sont faiblement corrélées : lorsque cette condition est vérifiée, l'inégalité de Tchebychev montre que Y_n converge en probabilité vers 0.

Remarque La loi ordinaire des grands nombres indique seulement, dans le corollaire 1, que la probabilité pour que l'écart $|Y_n - p|$ entre fréquence empirique et probabi-

lité dépasse α , (écart individuel relatif à un rang n donné d'avance) est petite si le rang n est grand. Elle ne signifie pas du tout qu'au delà d'un certain rang tous les écarts seront inférieurs à α . Autrement dit, elle ne signifie pas que la fréquence converge vers la probabilité. Cette propriété résultera seulement de la loi forte des grands nombres.

II.- CONVERGENCE PRESQUE SURE, ET LOI FORTE DES GRANDS NOMBRES.

Examinons maintenant un mode beaucoup plus fort de convergence, à savoir la convergence presque sûre (ou presque certaine). On dira que X_n converge presque sûrement vers X si l'évènement " X_n tend vers X " est presque certain. Mais il faut, au préalable, s'assurer que la proposition " X_n tend vers X " définit bien un évènement. Cela résulte du lemme suivant :

Lemme : Soient X_n une suite de variables aléatoires, et X une variable aléatoire sur un espace probabilisable (E, \mathcal{A}) . La proposition " X_n tend vers X lorsque n tend vers l'infini" définit un évènement de \mathcal{A} .

Par hypothèse, les $X_n(e)$ et $X(e)$ sont des fonctions mesurables. Soit A l'ensemble des évènements élémentaires e pour lesquels on a : $X_n(e) \rightarrow X(e)$. Il faut montrer que A appartient à la σ -algèbre \mathcal{A} . Or A est constitué par l'ensemble des e qui vérifient la proposition :

"Quel que soit l'entier k , il existe N tel que $n \geq N \implies |X_n(e) - X(e)| \leq \frac{1}{k}$ "

Considérons l'ensemble

$$B_{k,n} = \left\{ e : \left| X_n(e) - X(e) \right| \leq \frac{1}{k} \right\}$$

qui est un évènement, X_n et X étant mesurables. A se met sous la forme :

$$A = \prod_{k=1}^{\infty} \sum_{N=1}^{\infty} \prod_{n=N}^{\infty} B_{kn}$$

C'est donc un évènement d'après les axiomes des σ -algèbres.

Définition : Convergence presque sûre :- On dit qu'une suite de variables X_n converge presque certainement vers une variable X si l'évènement " $X_n \rightarrow X$ " est presque certain, c'est-à-dire si l'on a :

$$P(X_n \rightarrow X) = 1$$

Pour expliciter cette notion fondamentale, donnons quelques critères de convergence presque sûre, les uns nécessaires et suffisants, les autres simplement suffisants.

En premier lieu, remarquons que l'évènement " $X_n \rightarrow X$ " est le produit (avec les notations du lemme)

$$A = \prod_{k=1}^{\infty} B_k$$

chaque évènement B_k étant défini par :

$$B_k = \sum_{N=1}^{\infty} \prod_{n=N}^{\infty} B_{k,n}$$

c'est-à-dire par la proposition "il existe un entier N tel que, pour tout $n \geq N$ on ait $|X_n - X| \leq \frac{1}{k}$ " Comme A est contenu dans chaque B_k , on a :

$$P(A) \leq P(B_k) \quad (\forall k)$$

Donc, si $P(A) = 1$, tous les $P(B_k)$ sont égaux à 1, et la réciproque est immédiate : si $P(B_k) = 1$ quel que soit k , on a $P(A) = 1$, car :

$$P(\bar{A}) = P\left(\sum_{k=1}^{\infty} \bar{B}_k\right) \leq \sum_{k=1}^{\infty} P(\bar{B}_k) = 0$$

Par ailleurs, il est clair que, quel que soit $\varepsilon > 0$, il existe un entier k avec $\frac{1}{k+1} \leq \varepsilon \leq \frac{1}{k}$. Nous pouvons donc énoncer :

Critère 1 - Pour que la suite X_n converge presque certainement vers X , il faut et il suffit que, quel que soit $\varepsilon > 0$, l'évènement "il n'y a qu'un nombre fini d'indices n tels que $|X_n - X| > \varepsilon$ " soit presque certain.

Pour obtenir un critère plus analytique, remarquons qu'il y a convergence presque sûre si, et seulement si, $P(\bar{B}_k) = 0$ quel que soit k . Or l'évènement :

$$\bar{B}_k = \prod_{N=1}^{\infty} C_{k,N}$$

est le produit des évènements

$$C_{k,N} = \bigcap_{n=N}^{\infty} \bar{B}_{k,n}$$

On a les inclusions évidentes : $C_{k,1} \supset C_{k,2} \supset \dots \supset C_{k,N} \supset \dots$

Alors le théorème de continuité du chapitre I, et sa réciproque, montre que l'on a $P(\bar{B}_k) = 0$ si, et seulement si :

$$\lim_{N \rightarrow \infty} P(C_{k,N}) = 0$$

L'évènement $C_{k,N}$: "l'une au moins des inégalités $|X_n - X| > \frac{1}{k}$ a lieu pour un indice $n \geq N$ " est également défini par la proposition : " $\sup_{n \geq N} |X_n - X| > \frac{1}{k}$ ". Enfin, comme ci-dessus, on peut remplacer $\frac{1}{k}$ par $\varepsilon > 0$ quelconque.

Énonçons :

Critère 2 - Pour que la suite X_n converge presque certainement vers X , il faut et il suffit que l'on ait, pour tout $\varepsilon > 0$:

$$\lim_{N \rightarrow \infty} P\left(\sup_{n \geq N} |X_n - X| > \varepsilon\right) = 0$$

Analytiquement, ce critère s'explique ainsi : quels que soient les nombres positifs α et ε , il existe un entier N_0 tel que :

$$N \geq N_0 \implies P\left(\sup_{n \geq N} |X_n - X| > \varepsilon\right) \leq \alpha$$

Passons maintenant à un critère simplement suffisant, mais très commode dans les applications.

Critère 3 -, ou Critère de Borel-Cantelli. - Pour que la suite X_n converge presque certainement vers X , il suffit que, quel que soit $\alpha > 0$, la série :

$$\sum_{n=1}^{\infty} P(|X_n - X| \geq \alpha)$$

soit convergente.

Ce critère est une conséquence immédiate du critère 2. On a, en effet :

$$P\left(\sup_{n \geq N} |X_n - X| \geq \alpha\right) \leq \sum_{n=N}^{\infty} P(|X_n - X| \geq \alpha)$$

Le deuxième membre constitue, par hypothèse, le reste d'une série convergente et tend vers 0 lorsque $N \rightarrow \infty$, et la convergence presque sûre en résulte.

Remarque - Si les variables aléatoires $(X_n - X)$ sont mutuellement indépendantes, le critère précédent est nécessaire et suffisant. En effet, on a alors :

$$P\left(\sup_{n \geq N} |X_n - X| \geq \alpha\right) = 1 - \prod_{n=N}^{\infty} \left[1 - P(|X_n - X| \geq \alpha)\right]$$

Or le produit infini $\prod_{n=1}^{\infty} \left[1 - P(|X_n - X| \geq \alpha)\right]$ converge ou diverge en même temps que la série $\sum_{n=1}^{\infty} P(|X_n - X| \geq \alpha)$. Par suite, on a

$$\lim_{N \rightarrow \infty} \prod_{n=N}^{\infty} \left[1 - P(|X_n - X| \geq \alpha)\right] = 1$$

si, et seulement si cette série est convergente, et le critère de Borel-Cantelli est alors nécessaire et suffisant.

Si la limite X est une constante α , on a, en particulier :

Corollaire 1 - Pour qu'une suite X_n de variables aléatoires indépendantes converge presque sûrement vers une constante α , il faut et il suffit que, quel que soit le nombre $\alpha > 0$, la série :

$$\sum_{n=1}^{\infty} P(|X_n - \alpha| \geq \alpha)$$

soit convergente.

En particulierisant davantage encore, on obtient :

Corollaire 2 - Soit A_n une suite d'évènements indépendants. Pour que l'évènement A :

" un nombre fini seulement d'évènements A_n est réalisé " soit presque certain,

il faut et il suffit que la série $\sum P(A_n)$ soit convergente.

Il suffit, pour le voir, d'introduire les variables aléatoires Z_n définies par :

$$\left\{ \begin{array}{ll} Z_n(e) = 1 & \text{si } e \in A_n \quad (\text{si } A_n \text{ est réalisé}) \\ Z_n(e) = 0 & \text{si } e \in \bar{A}_n \quad (\text{si } A_n \text{ n'est pas réalisé}) \end{array} \right.$$

Pour $0 < \alpha \leq 1$, on a $P(Z_n \geq \alpha) = P(A_n)$. D'après le corollaire 1, Z_n converge presque sûrement vers 0 si, et seulement si, la série $\sum P(A_n)$ est convergente. Or l'évènement " $Z_n \rightarrow 0$ " coïncide manifestement avec A. L'énoncé en résulte.

Théorème 2 - La convergence presque sûre entraîne la convergence en probabilité et la convergence en loi.

En effet, pour $k \geq N$, l'évènement " $|X_k - X| \geq \alpha$ " entraîne l'évènement " $\sup_{n \geq N} |X_n - X| \geq \alpha$ ", et par suite :

$$P(|X_k - X| \geq \alpha) \leq P(\sup_{n \geq N} |X_n - X| \geq \alpha)$$

S'il y a convergence presque sûre, le deuxième membre tend vers 0 lorsque N augmente indéfiniment. Par suite :

$$\lim_{k \rightarrow \infty} P(|X_k - X| \geq \alpha) = 0$$

et les X_k convergent en probabilité vers X, donc aussi en loi d'après le théorème 1.

La réciproque est fautive : la convergence en probabilité n'entraîne pas (est strictement plus faible que) la convergence presque sûre. Cela se comprend : l'évènement " $|X_k - X| < \alpha$ " implique la réalisation d'une seule inégalité, tandis que l'évènement " $\sup_{n \geq N} |X_n - X| < \alpha$ " signifie que toutes les inégalités, en nombre infini, $|X_n - X| < \alpha$, sont réalisées au delà du rang N.

Exemple : Soient X_n une suite de variables indépendantes ayant comme fonctions de répartition :

$$F_n(x) = \begin{cases} 0 & \text{pour } x \leq 0 \\ 1 - \frac{1}{1 + nx} & \text{pour } x > 0 \end{cases}$$

Les X_n convergent en probabilité vers la variable nulle $X = 0$, car

$$P(X_n \geq \alpha) = \frac{1}{1 + n\alpha} \rightarrow 0 \quad \text{pour } n \rightarrow \infty$$

Mais elles ne convergent pas presque sûrement. En effet $P(\sup_{n \geq N} X_n < \alpha)$

est inférieure à $P(X_n < \alpha \text{ et } X_{N+1} < \alpha \text{ et } \dots \text{ et } X_{N+k} < \alpha)$

elle-même égale au produit

$$\left(1 - \frac{1}{N\alpha + 1}\right) \left(1 - \frac{1}{(N+1)\alpha + 1}\right) \dots \left(1 - \frac{1}{(N+k)\alpha + 1}\right)$$

dont la limite est 0 lorsque k augmente indéfiniment. D'où résulte que l'on a

$$P(\sup_{n \geq N} X_n < \alpha) = 0$$

et

$$P(\sup_{n \geq N} X_n \geq \alpha) = 1$$

ce qui est incompatible avec le critère 2 de convergence presque sûre.

LOI FORTE DES GRANDS NOMBRES.

La loi forte des grands nombres donne une signification précise à la proposition : " la fréquence empirique tend vers la probabilité " et explicite les conditions dans lesquelles cette proposition est vraie. Elle constitue donc la justification théorique de tous les procédés d'inférence statistique utilisés dans les applications. Il est très clair qu'en l'absence de la loi forte des grands nombres le calcul des probabilités resterait une pure abstraction, puisqu'il ne serait pas possible de déterminer (avec une précision donnée) la valeur numérique réelle de la probabilité objective d'un événement ou d'un phénomène concret. La signification précise de cette loi est la suivante : " la fréquence empirique converge presque certainement vers la probabilité ". De là provient l'importance de la notion de convergence presque sûre. Nous donnerons seulement la démonstration du théorème de Borel. Des résultats plus généraux et très profonds seront ensuite énoncés sans démonstration.

Théorème de Borel -- Soit p la probabilité d'un événement A , et soit Y_n le nombre de réalisations de A au cours de n épreuves indépendantes successives. Lorsque n augmente indéfiniment, la fréquence $\frac{1}{n} Y_n$ converge presque sûrement vers p .

D'après le lemme de Borel-Cantelli, il suffit de montrer que la série :

$$\sum_{n=1}^{\infty} P\left(\left|\frac{1}{n} Y_n - p\right| \geq \alpha\right)$$

est convergente, quel que soit $\alpha > 0$. Pour cela, nous allons utiliser une inégalité analogue à l'inégalité de Tchebychev, mais où figure le moment d'ordre 4 au lieu de la variance. Si X est une variable aléatoire admettant un moment d'ordre 4, on a :

$$P\left[|X - E(X)| \geq \alpha\right] \leq \frac{E[(X - E(X))^4]}{\alpha^4}$$

En effet, il suffit d'écrire (en posant $E(X) = m$)

$$\int_{|x-m| \geq \alpha} dF(x) \leq \int_{|x-m| \geq \alpha} \frac{(x-m)^4}{\alpha^4} dF(x) \leq \int_{-\infty}^{\infty} \frac{(x-m)^4}{\alpha^4} dF(x) = \frac{E[(x-m)^4]}{\alpha^4}$$

Calculons alors $E\left[\left(\frac{Y_n}{n} - p\right)^4\right]$. Pour cela, on considère la fonction caractéristique de $\frac{Y_n}{n} - p$, soit $\Phi(u)$, et on écrit son développement limité à l'ordre 4 :

$$\begin{aligned} \Phi(u) &= e^{-pi u} (q + p e^{\frac{i u}{n}})^n = \left[q e^{-pi \frac{u}{n}} + p e^{q i \frac{u}{n}} \right]^n \\ &= \left[1 + \frac{1}{2} pq \left(\frac{i u}{n}\right)^2 + \frac{1}{3!} pq(q-p) \left(\frac{i u}{n}\right)^3 + \frac{1}{4!} pq(p^3 + q^3) \left(\frac{i u}{n}\right)^4 + \dots \right]^n \end{aligned}$$

On trouve ainsi :

$$E\left[\left(\frac{1}{n} Y_n - p\right)^4\right] = \frac{pq}{n^4} \left[n(p^3 + q^3) + 3 pq n(n-1) \right] \leq \frac{1}{n^2}$$

On a donc :

$$P\left(\left|\frac{1}{n} Y_n - p\right| \geq \alpha\right) \leq \frac{1}{\alpha^4 n^2}$$

et la série de terme général $P\left(\left|\frac{1}{n} Y_n - p\right| \geq \alpha\right)$ est bien convergente.

Énonçons maintenant, sans démonstration, deux théorèmes de Kolmogorov :

Premier théorème de Kolmogorov

Soit X_n une suite de variables aléatoires indépendantes possédant des variances finies σ_n^2 et des espérances mathématiques m_n . Si la série $\sum_{n=1}^{\infty} \frac{1}{n^2} \sigma_n^2$ est convergente, la variable aléatoire $Y_n = \frac{1}{n} \sum_{k=1}^n (X_k - m_k)$ converge presque certainement vers 0 :

$$P(Y_n \rightarrow 0) = 1$$

Deuxième théorème de Kolmogorov.

Soit X_n une suite de variables aléatoires indépendantes ayant la même loi de probabilité $F(x)$. Pour que la variable $Y_n = \frac{1}{n} \sum_{k=1}^n X_k$ converge presque sûrement vers une constante a , il faut et il suffit que les X_n aient une espérance mathématique m , et on a alors nécessairement $m = a$

On notera l'importance extrême de ce dernier théorème, qui constitue la justification théorique des procédés statistiques usuels. Il correspond, en effet, au problème de l'estimation de la moyenne vraie à partir de la moyenne arithmétique de n échantillons indépendants.

On notera aussi que la condition d'existence de l'espérance mathématique implique que l'intégrale $\int |x| dF$ soit convergente.

La loi du logarithme itéré, que nous énoncerons sans démonstration, est encore plus subtile que la loi forte des grands nombres, et donne des renseignements très précis et très profonds sur le comportement asymptotique des sommes de variables indépendantes.

Considérons une alternative répétée, c'est-à-dire une suite de variables indépendantes X_n de même loi $P(X_n = 1) = p$, $P(X_n = 0) = q = 1 - p$. Le théorème de Borel indique que la fréquence :

$$Y_n = \frac{1}{n} \sum_{k=1}^n X_k$$

converge presque sûrement vers p . A cette variable Y_n , associons la variable ré-

duite Z_n de moyenne nulle et de variance unité :

$$Z_n = \sqrt{\frac{n}{pq}} (Y_n - p)$$

On sait (et on verra dans le chapitre prochain) que Z_n converge en loi vers une variable normale. La loi du logarithme itéré énonce le résultat extraordinairement précis suivant :

Loi du Logarithme itéré - Dans l'alternative indéfiniment répétée, la variable réduite Z_n possède la propriété suivante : si λ est un nombre supérieur à 1, il est presque certain qu'un nombre fini seulement d'inégalités

$$Z_n > \lambda \sqrt{2 \log \log n}$$

sera vérifié. Si, au contraire, $\lambda < 1$, il est presque certain qu'un nombre infini de ces inégalités sera réalisé. Autrement dit, on a :

$$P\left(\lim_{n \rightarrow \infty} \frac{Z_n}{\sqrt{2 \log \log n}} = 1\right) = 1$$

La fonction $\sqrt{2 \log \log n}$ constitue presque certainement la borne supérieure des fluctuations de la variable réduite, au delà d'un rang N fini.

Par raison de symétrie, $-\sqrt{2 \log \log n}$ constitue aussi presque certainement la borne inférieure des fluctuations de Z_n , au delà d'un rang N' fini

$$P\left(\lim_{n \rightarrow \infty} \frac{Z_n}{\sqrt{2 \log \log n}} = 1\right) = 1$$

Énonçons cette même loi, en remplaçant la variable réduite Z_n par la fréquence empirique Y_n :

Si $\lambda > 1$, il est presque certain qu'un nombre fini seulement d'inégalités

$$Y_n > p + \lambda \sqrt{\frac{2pq}{n} \log \log n}$$

sera vérifié. Si $\lambda < 1$, il est presque certain qu'une infinité de telles inégalités sera vérifiée.

III.- CONVERGENCE EN MOYENNE QUADRATIQUE

Examinons maintenant un dernier mode de convergence, qui joue un rôle fondamental dans la théorie des processus stochastiques, notamment dans la définition de la notion d'intégrale stochastique.

Définition : Convergence en moyenne quadratique : On dit qu'une suite X_n de variables aléatoires converge en moyenne quadratique vers une variable aléatoire X si l'on a :

$$\lim_{n \rightarrow \infty} E \left[(X_n - X)^2 \right] = 0$$

Dans les applications, la limite X sera rarement connue a priori. Le critère suivant, où elle n'intervient pas explicitement, sera donc très utile :

Critère de Cauchy. Pour qu'une suite X_n de variables aléatoires possédant des moments d'ordre 2 (ou des variances) finis converge en moyenne quadratique, il faut et il suffit que l'on ait :

$$\lim_{\substack{n \rightarrow \infty \\ m \rightarrow \infty}} E \left[(X_m - X_n)^2 \right] = 0$$

soit, explicitement

$$\forall \varepsilon > 0, \exists N : n \geq N \quad \text{et} \quad m \geq N \quad \Rightarrow \quad E \left[(X_m - X_n)^2 \right] \leq \varepsilon$$

La condition est nécessaire : Cela découle immédiatement de l'inégalité de Schwartz :

$$\begin{aligned} E \left[(X_n - X_m)^2 \right] &= E \left[(X_n - X)^2 \right] + E \left[(X_m - X)^2 \right] - 2 E \left[(X_n - X)(X_m - X) \right] \\ &\leq \left[\sqrt{E \left[(X_n - X)^2 \right]} + \sqrt{E \left[(X_m - X)^2 \right]} \right]^2 \end{aligned}$$

La condition est suffisante : la démonstration est plus délicate, et ne sera pas donnée ici.

Remarque. Si nous posons

$$d(X, Y) = \sqrt{E \left[(X - Y)^2 \right]}$$

la fonction d vérifie deux des axiomes des espaces métriques, car on a :

$$\left\{ \begin{array}{l} d(X, Y) = d(Y, X) \geq 0 \\ d(X, Y) \leq d(X, Z) + d(Z, Y) \end{array} \right.$$

X, Y et Z étant des variables quelconques de variances finies. (l'inégalité triangulaire se déduit immédiatement de l'inégalité de Schwarz). Posons :

$$\|X\| = d(X, 0) = \sqrt{E[(X)^2]}$$

On voit que $\|X\|$ vérifie tous les axiomes des normes (sur un espace vectoriel), sauf l'axiome :

$$\|X\| = 0 \implies X = 0$$

En fait, l'inégalité de Tchebychev montre que $\|X\| = 0$ entraîne simplement $P(X = 0) = 1$, c'est-à-dire que X est presque certainement nulle. Si nous convenons d'identifier deux variables lorsqu'elles sont presque sûrement égales ($P(X = Y) = 1$) alors $\|X\|$ est une norme pour l'espace vectoriel des variables aléatoires de variances finies.

Le critère de Cauchy montre que cet espace vectoriel normé est également complet, c'est-à-dire constitue ce que l'on appelle un espace de Banach. Muni du produit scalaire

$$\langle X, Y \rangle = E[X Y]$$

cet espace de Banach est aussi un espace de Hilbert.

Théorème 3 - La convergence en moyenne quadratique entraîne la convergence en probabilité, donc aussi la convergence en loi

Cela résulte de l'inégalité de Tchebychev :

$$P(|X_n - X| > \alpha) \leq \frac{E[(X_n - X)^2]}{\alpha^2}$$

Ici encore, la réciproque n'est pas vraie : la convergence en probabilité n'entraîne pas (est strictement plus faible que) la convergence en moyenne quadratique.

De même encore :

- La convergence en moyenne quadratique n'entraîne pas la convergence presque sûre.

- La convergence presque sûre n'entraîne pas la convergence en moyenne quadratique.
