

UNIVERSITE SCIENTIFIQUE ET MEDICALE DE GRENOBLE

LABORATOIRE DE MATHEMATIQUES PURES

INSTITUT DE RECHERCHE SUR L'ENSEIGNEMENT DES MATHEMATIQUES

PRECIS DE PROBABILITES

**A l'usage des étudiants de Maîtrise d'Enseignement
et des stagiaires de l'I.R.E.M.**

A. BAILLE

1976

TABLE des MATIERES

	Pages.
<u>INTRODUCTION</u>	3
<u>I - ESPACE PROBABILISABLE.</u>	
1 Définition.....	3
2 Conséquences.....	4
3 Exemples de tribus.....	4
4 Exemple fondamental : les boréliens de \mathbf{R}	5
5 Cas particulier d'un espace fini.....	6
<u>II - ESPACE PROBABILISE.</u>	
1 Définition.....	8
2 Conséquences.....	8
3 Terminologie.....	9
4 Cas particulier d'un univers fini.....	10
5 Remarque importante.....	11
6 Exercices conduisant à des dénombrements.....	11
7 Continuité des probabilités.....	12
<u>III - PROBABILITE CONDITIONNELLE - EVENEMENTS INDEPENDANTS.</u>	
1 Probabilité conditionnelle.....	13
2 Indépendance.....	15
3 Produit d'espaces probabilisables. Probabilité-produit.....	17
<u>IV - VARIABLES ALEATOIRES.</u>	
1 Notion d'application mesurable. Probabilité-image.....	19
2 Définition et loi d'une variable aléatoire réelle.....	19
3 Types de lois.....	22
<u>V - NOTIONS SUR L'INTEGRATION.</u>	
1 Applications étagées.....	25
2 Intégration des applications étagées.....	27
3 Applications intégrables.....	29
<u>VI - MOMENTS.</u>	
1 Espérance mathématique.....	32
2 Moments d'ordre quelconque.....	32
3 Variance et écart-type.....	33

4	Inégalité de Bienaymé-Tchébicheff.....	34
---	--	----

VII - VARIABLES ALEATOIRES INDEPENDANTES.

1	Définition.....	36
2	Propriétés.....	36

VIII - FONCTION GENERATRICE. FONCTION CARACTERISTIQUE.

PRODUIT DE CONVOLUTION.

1	Fonction génératrice.....	38
2	Fonction caractéristique.....	39
3	Produit de convolution.....	40

IX - LOIS USUELLES.

1	Lois discrètes.....	41
2	Lois absolument continues.....	44

X - CONVERGENCE.

1	Problème.....	48
2	Convergence en loi.....	48
3	Convergence en probabilité.....	50
4	Convergence presque sûre.....	52

XI - VECTEURS ALEATOIRES.

1	Définition et loi d'un vecteur aléatoire.....	54
2	Types de lois.....	55
3	Lois marginales.....	55
4	Changement de variable.....	57
5	Espérance mathématique. Matrice de covariance et coefficient de corrélation.....	58

	<u>ELEMENTS DE BIBLIOGRAPHIE.....</u>	61-62
--	---------------------------------------	-------

PROBABILITES

INTRODUCTION.

Notion intuitive développée à l'occasion des jeux aux 17 et 18ième siècles. Première étude sérieuse de Bernoulli.

Au 20ième siècle, systématisation, axiomatisation : création d'un MODELE MATHEMATIQUE. On n'étudiera pas l'adéquation de ce modèle, ce serait le rôle de la statistique mathématique (et non de la statistique descriptive, seule enseignée dans le Secondaire).

Notre modèle sera la théorie de Kolmogorov.

I - ESPACE PROBABILISABLE.

1 - Définition. On appelle espace probabilisable le couple (ensemble, tribu de parties de cet ensemble) : (Ω, \mathfrak{a}) .

Ω est l'univers.

\mathfrak{a} est une tribu de parties de Ω , c'est-à-dire un ensemble de parties de Ω ($\mathfrak{a} \subset \mathcal{P}(\Omega)$) répondant aux conditions suivantes :

1) $\Omega \in \mathfrak{a}$;

2) $A \in \mathfrak{a} \Rightarrow \bar{A} \in \mathfrak{a}$ (la tribu est fermée pour la complémentation) ;

3) $(\forall_n \in \mathbb{N}^*, A_n \in \mathfrak{a}) \Rightarrow \bigcup_{n \in \mathbb{N}^*} A_n \in \mathfrak{a}$ (la tribu est fermée pour l'union dénombrable).

2 - Conséquences.

* $\emptyset \in \mathfrak{a}$ car $\emptyset = \overline{\Omega}$.

* \mathfrak{a} est fermée pour l'union finie. En effet, soient

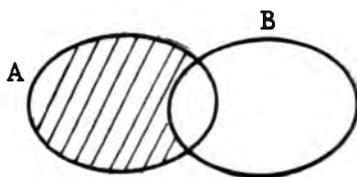
$$A_1, A_2, A_3 \in \mathfrak{a}$$

$$A_1 \cup A_2 \cup A_3 \cup \emptyset \cup \emptyset \dots \in \mathfrak{a}.$$

* \mathfrak{a} est fermée pour l'intersection finie ou dénombrable car

$$\bigcap_{n \in \mathbb{N}^*} A_n = \overline{\bigcup_{n \in \mathbb{N}^*} \overline{A_n}}.$$

* \mathfrak{a} est fermée pour la différence, propre ou non, car



$$A \setminus B = A \cap \overline{B}.$$

* \mathfrak{a} est fermée pour la différence symétrique, car

$$A \Delta B = (A \setminus B) \cup (B \setminus A).$$

(\cup désigne l'union disjointe).

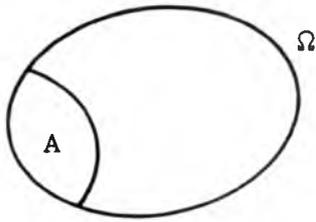
REMARQUE. Un autre système d'axiomes commode pour définir une tribu consiste à conserver les axiomes (2) et (3) ci-dessus et à remplacer l'axiome (1) par :

1') \mathfrak{a} est fermée pour l'union finie et est non vide. On vérifie aisément l'équivalence des deux systèmes d'axiomes.

3 - Exemples de tribus.

* $\mathfrak{a} = \mathcal{P}(\Omega)$.

* Tribu engendrée par un élément A de $\mathcal{P}(\Omega)$, c'est-à-dire



la plus petite tribu qui contient A :

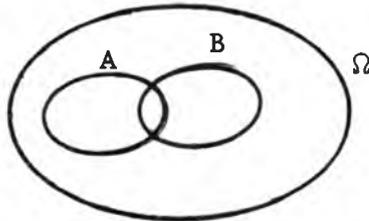
$$\mathfrak{a} = \{\emptyset, \Omega, A, \bar{A}\}.$$

La plus simple tribu est donc $\{\emptyset, \Omega\}$.

* Tribu engendrée par deux éléments de $\mathcal{P}(\Omega)$. Dans le cas général, A et B divisent Ω en quatre parties deux à deux disjointes : $A \setminus B$, $B \setminus A$, et $A \cap B$ et $\overline{A \cup B}$.

Tout élément de la tribu est une union de certaines de ces parties : la tribu compte donc 2^4 , soit 16 éléments.

Si $A \subset B$ ou $A \cap B = \emptyset$, il n'y a plus que 8 éléments dans la tribu.

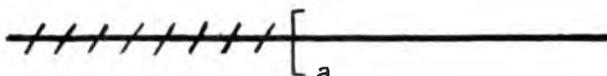


* Plus généralement, on appelle tribu engendrée par un sous-ensemble E de $\mathcal{P}(\Omega)$ l'intersection de toutes les tribus de Ω contenant E . Cet ensemble de tribus n'est pas vide, puisqu'il contient $\mathcal{P}(\Omega)$ et il est aisé de montrer que l'intersection d'un ensemble, fini ou non, de tribus de Ω est une tribu de Ω .

La tribu engendrée par E est donc aussi la plus petite tribu de Ω qui contienne E .

* A étant un sous-ensemble de Ω et \mathfrak{a} une tribu de parties de Ω , on appelle trace de \mathfrak{a} sur A l'ensemble des intersections de A et des éléments de \mathfrak{a} . On montrera aisément que \mathfrak{c} est une tribu de parties de A et que, si $A \in \mathfrak{a}$ \mathfrak{c} est aussi l'ensemble des éléments de \mathfrak{a} inclus dans A .

4 - Exemple fondamental : les boréliens de \mathbb{R} .



C'est la tribu engendrée par les ensembles $] - \infty, a [$. On y trouvera donc :

- les ensembles $[a, +\infty[$, par complémentation ;
- les ensembles $]a, +\infty[$, car

$$]a, +\infty[= \bigcup_{n=1}^{\infty} [a + \frac{1}{n}, +\infty[,$$

et donc les ensembles $] -\infty, a]$, par passage au complémentaire.

Comme intersections d'ensembles ci-dessus, se trouvent dans la tribu tous les intervalles fermés, ouverts, semi-ouverts, tous les ensembles à un élément en particulier et, comme unions de ces derniers, tous les ensembles finis ou dénombrables de \mathbb{R} .

Plus généralement, on trouve dans la tribu tous les ouverts (car tout ouvert de \mathbb{R} est union finie ou dénombrable d'intervalles ouverts) et, par complémentation, tous les fermés de \mathbb{R} .

On désigne cette tribu par \mathcal{B} et l'on montre que ce n'est pas $\mathcal{P}(\mathbb{R})$: $\mathcal{B} \subsetneq \mathcal{P}(\mathbb{R})$.

Généralisation : la tribu des boréliens de \mathbb{R}^n notée $\mathcal{B}_{\mathbb{R}^n}$. Elle est engendrée par les ensembles : $\prod_{i=1}^n]-\infty, a_i[$. On définit ainsi $\mathcal{B}_{\mathbb{C}}$, puisque l'on peut identifier \mathbb{C} à \mathbb{R}^2 .

5 - Cas particulier d'un espace fini.

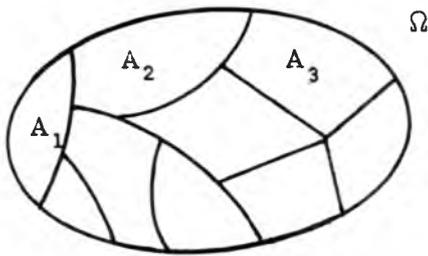
On peut utiliser $\mathcal{P}(\Omega)$: $\text{card } \mathcal{P}(\Omega) = 2^{\text{card } \Omega}$.

De toute façon on peut se libérer de la condition d' "union dénombrable" : la fermeture pour l'union finie suffit. Une tribu $\mathcal{B}(\Omega)$ s'appelle alors algèbre de Boole des parties de Ω ; les conditions sont :

- 1) $\Omega \in \mathcal{B}(\Omega)$ ou (1') $\mathcal{B}(\Omega)$ non vide ;
- 2) $A \in \mathcal{B}(\Omega) \Rightarrow \bar{A} \in \mathcal{B}(\Omega)$;
- 3) $\left. \begin{array}{l} A \in \mathcal{B}(\Omega) \\ B \in \mathcal{B}(\Omega) \end{array} \right\} \Rightarrow A \cup B \in \mathcal{B}(\Omega)$.

$\mathcal{B}(\Omega)$ est donc fermée pour l'union finie, par associativité et de proche en proche.

THEOREME. Si Ω est un ensemble fini et $\mathcal{B}(\Omega)$ une algèbre de Boole de parties de Ω , il existe une partition de Ω en parties (non vides) A_i ($1 \leq i \leq n$) telles que $\mathcal{B}(\Omega)$ est constitué de \emptyset et toutes les unions finies d'ensembles A_i .



Les parties A_i s'appellent les atomes de $\mathcal{B}(\Omega)$.

Soit, en effet, la relation binaire suivante entre éléments de Ω : $\omega_1 \sim \omega_2 \iff$ tout élément de $\mathcal{B}(\Omega)$ qui contient l'un des éléments ω_1 ou ω_2 contient l'autre (ou encore : aucun élément de $\mathcal{B}(\Omega)$ ne sépare ω_1 et ω_2).

C'est une relation d'équivalence, comme on le montre aisément.

Soient, $A_1, A_2, A_3, \dots, A_n$ les classes d'équivalence. Soit $A \in \mathcal{B}(\Omega)$: il ne peut partager un A_i ; soit ($1 \leq i \leq n$) $\Rightarrow ((A_i \subset A) \vee (A_i \cap A = \emptyset))$. Ce qui fait que A est vide ou union d'ensembles A_i . Réciproquement, toutes les unions d'ensembles A_i appartiennent-elles à $\mathcal{B}(\Omega)$? Il suffit de montrer que ($1 \leq i \leq n$) $\Rightarrow (A_i \in \mathcal{B}(\Omega))$.

Considérons l'intersection de tous les éléments de $\mathcal{B}(\Omega)$ qui contiennent A_i (Ω en fait partie, donc cette famille n'est pas vide) :

- c'est un élément de $\mathcal{B}(\Omega)$ qui contient A_i ;
- c'est une union d'ensembles A_j dont A_i fait partie ;
- c'est A_i seul, sinon cette union serait une classe d'équivalence.

Remarquons que, dans la démonstration, intervient seulement le fait que $\mathcal{B}(\Omega)$ est finie, mais non le fait que Ω est fini. Comme, d'autre part, il y a bijection entre $\mathcal{B}(\Omega)$ et $\mathcal{P}(\Omega')$, où $\Omega' = \{A_1, A_2, \dots, A_n\}$, le cardinal d'une algèbre de Boole finie de parties d'un ensemble est toujours une puissance de 2.

II - ESPACE PROBABILISE.

1 - Définition. On appelle espace probabilisé le triplet $(\Omega, \mathfrak{a}, P)$ constitué par les deux éléments d'un espace probabilisable et par une probabilité P définie sur (Ω, \mathfrak{a}) c'est-à-dire une application

$$P : \mathfrak{a} \rightarrow \mathbb{R}$$

possédant les propriétés :

1) $\forall A \in \mathfrak{a}, P(A) \geq 0$;

2) $P(\Omega) = 1$;

3) Quels que soient les ensembles $A_n (n \in \mathbb{N}^*)$ de \mathfrak{a} , deux à deux disjoints,

$$\text{on a : } P\left(\bigcup_{n \in \mathbb{N}^*} A_n\right) = \sum_{n \in \mathbb{N}^*} P(A_n).$$

2 - Conséquences.

* $P(\emptyset) = 0$. En effet, soit : $\forall n \in \mathbb{N}^*, A_n = \emptyset$.

$$\text{Alors } P\left(\bigcup_n A_n\right) = P(\emptyset) = \sum P(\emptyset)$$

et cette somme n'existe que si $P(\emptyset) = 0$.

* Quels que soient les ensembles $A_n (1 \leq n \leq N)$ de \mathfrak{a} deux à deux disjoints,

$$\text{on a : } P\left(\bigcup_{1 \leq n \leq N} A_n\right) = \sum_{1 \leq n \leq N} P(A_n).$$

Il suffit de prendre $A_n = \emptyset$ pour $n > N$ et d'appliquer l'axiome (3).

* Soit $A \in \mathfrak{a}$; $\bar{A} \in \mathfrak{a}$ et $P(A \cup \bar{A}) = P(\Omega) = P(A) + P(\bar{A}) = 1$,

$$\text{d'où : } P(\bar{A}) = 1 - P(A).$$

On en déduit :

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

* Si $A \subset B$ ($A \in \mathfrak{a}$ et $B \in \mathfrak{a}$), $P(A) \leq P(B)$.

En effet :

$$B = A \cup (B \setminus A) \text{ et } B \setminus A \in \mathfrak{a}.$$

D'où : $P(B) = P(A) + P(B \setminus A) \geq P(A)$.

* Si A et B sont des éléments de \mathfrak{a} ,

on a : $A \cup B = A \cup (B \setminus A)$,

d'où : $P(A \cup B) = P(A) + P(B \setminus A)$.

D'autre part :

$B = (A \cap B) \cup (B \setminus A)$,

d'où : $P(B \setminus A) = P(B) - P(A \cap B)$.

On en tire :

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Cette formule peut se généraliser : si les $A_i (1 \leq i \leq n)$ sont des éléments de \mathfrak{a} , on a :

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{1 \leq i \leq n} P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) + \dots + (-1)^{n+1} P(A_1 \cap A_2 \cap \dots \cap A_n).$$

Le second membre peut encore s'écrire :

$$\sum_{p=1}^n (-1)^{p+1} \sum_{1 \leq i_1 < i_2 < \dots < i_p \leq n} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_p}).$$

C'est la formule de Poincaré ; on peut la démontrer par récurrence sur n.

REMARQUE.

Puisque : $A \Delta B = (A \setminus B) \cup (B \setminus A)$,

on a : $P(A \Delta B) = P(A) + P(B) - 2P(A \cap B)$.

3 - Terminologie.

Ω est l'univers ;

tout élément de la tribu est un événement ;

\cup est la disjonction ("ou" non exclusif) ;

\cap est la conjonction ("et") ;

\emptyset est l'événement impossible ;

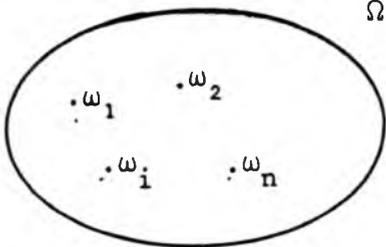
Ω est l'événement certain ;

si A réalise $P(A) = 0$, A est dit presque impossible ;

si A réalise $P(A) = 1$, A est dit presque certain.

4 - Cas particulier d'un univers fini.

$$n = \text{card } \Omega \quad \Omega = \{\omega_1, \omega_2, \dots, \omega_n\}.$$



Ω Soit $a = P(\Omega)$; les nombres $p_i = P(\{\omega_i\})$

doivent réaliser $p_i \geq 0$, et,

puisque : $\Omega = \bigcup_{1 \leq i \leq n} \{\omega_i\}$,

$$P(\Omega) = \boxed{1 = \sum_{1 \leq i \leq n} p_i.}$$

S'étant ainsi donné les p_i , a-t-on défini une probabilité sur $(\Omega, P(\Omega))$?

Tout sous-ensemble A de Ω est l'union disjointe des ensembles $\{\omega_i\}$ tels que ω_i soit élément de A ; d'où, nécessairement :

$$P(A) = \sum_{\{i; \omega_i \in A\}} P(\{\omega_i\}) = \sum_{\{i; \omega_i \in A\}} p_i.$$

On vérifie aisément que l'on a bien défini une probabilité.

Si \mathfrak{a} est une algèbre de Boole distincte de $P(\Omega)$, considérons les atomes A_1, A_2, \dots, A_k ($k < n$) ; on se donne les nombres

$$p_i = P(A_i)$$

tels que :

$$p_i \geq 0 \quad \text{et} \quad \sum_{1 \leq i \leq k} p_i = 1,$$

et l'on définit, de même, pour tout élément A de l'algèbre de Boole :

$$P(A) = \sum_{\{i; A_i \subset A\}} p_i.$$

5 - Remarque importante.

Pour définir une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ ou sur $(\Omega, \mathcal{B}(\Omega))$, il nous a suffi de nous donner convenablement les probabilités d'éléments particuliers de \mathcal{a} : les ensembles à un élément si $\mathcal{a} = \mathcal{P}(\Omega)$, les atomes si $\mathcal{a} = \mathcal{B}(\Omega)$. La probabilité est alors définie de façon unique.

On a ici un exemple de "prolongement" de probabilité.

De même, si Ω et \mathcal{a} sont infinis, il suffit de se donner (convenablement) les probabilités d'éléments particuliers de \mathcal{a} pour définir, et de façon unique, les probabilités de tous les éléments de \mathcal{a} . Mais nous n'étudierons pas ce problème du "prolongement".

6 - Exercices conduisant à des dénombrements.

Beaucoup d'exercices de probabilité (en particulier tous ceux que l'on peut aborder dans l'enseignement secondaire) conduisent à adopter un univers fini Ω : ensemble de tous les événements que l'on peut considérer comme élémentaires. La tribu est alors l'ensemble des parties de Ω , $\mathcal{P}(\Omega)$.

La probabilité, on l'a vu, est alors entièrement déterminée par la donnée des probabilités des événements élémentaires $\{\omega_i\}$.

Les hypothèses du problème permettent éventuellement de considérer que les événements élémentaires $\{\omega_i\}$ jouent des rôles symétriques et, ainsi, de leur attribuer la même probabilité p_i ; on a alors, nécessairement : $p_i = \frac{1}{\text{Card } \Omega}$. On dit que les événements $\{\omega_i\}$ sont équiprobables.

Si A est un événement quelconque, on sait alors que :

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

soit :
$$P(A) = \frac{\text{Card } A}{\text{Card } \Omega}.$$

L'exercice se ramène donc à faire des dénombrements. C'est le cas des exercices d'urnes, de dés, de cartes, etc... Mais il ne faut pas, pour autant, en négliger la partie proprement probabiliste : la caractérisation des événements que l'on peut considérer comme élémentaires, donc la définition de Ω , et la justification du fait qu'ils sont équiprobables, eu égard aux hypothèses du problème.

7 - Continuité des probabilités.

THEOREME. Soit (Ω, a, P) un espace probabilisé. Si $A_n (n \in \mathbb{N}^*)$ est une suite décroissante d'ensembles de a dont la limite est vide (c'est-à-dire tels que $\bigcap_{n \in \mathbb{N}^*} A_n = \emptyset$), $P(A_n)$ tend vers 0.

Quel que soit $\omega \in A_1$, il existe un entier positif unique n tel que :

$$\omega \in A_n, \omega \notin A_{n+1}.$$

A_1 est donc l'union disjointe des ensembles $(A_n \setminus A_{n+1})$,

d'où :
$$P(A_1) = \sum_{n \in \mathbb{N}^*} P(A_n \setminus A_{n+1}).$$

Comme, d'autre part : $n-1$

$$P(A_1) = \sum_{i=1}^{n-1} P(A_i \setminus A_{i+1}) + P(A_n),$$

$P(A_n)$ apparait comme le reste d'une série convergente : cette quantité tend vers 0.

COROLLAIRES.

1°) Si $A_n (n \in \mathbb{N}^*)$ est une suite décroissante d'ensembles de a de limite A (c'est-à-dire tels que $\bigcap_{n \in \mathbb{N}^*} A_n = A$), $P(A_n)$ tend vers $P(A)$.

2°) Si $A_n (n \in \mathbb{N}^*)$ est une suite croissante d'ensembles de a de limite A (c'est-à-dire tels que $\bigcup_{n \in \mathbb{N}^*} A_n = A$), $P(A_n)$ tend vers $P(A)$.

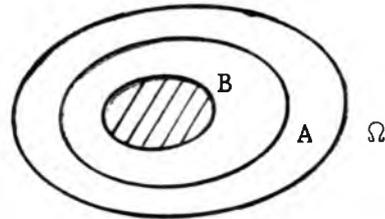
III - PROBABILITE CONDITIONNELLE. EVENEMENTS INDEPENDANTS.

1 - Probabilité conditionnelle.

$(\Omega, \mathfrak{a}, P)$

Soit A un événement de \mathfrak{a} tel que $P(A) \neq 0$.

Si \mathfrak{A} est l'ensemble des éléments de \mathfrak{a} inclus dans A ($\mathfrak{A}, \mathfrak{A}$) est un espace probabilisable (\mathfrak{A} est la trace sur A de la tribu \mathfrak{a}).



$(\mathfrak{A}, \mathfrak{A}, P)$ n'est pas probabilisé sauf si $P(A) = 1$.

On normalise : $(\mathfrak{A}, \mathfrak{A}, P_A)$ probabilisé en prenant $P_A = \frac{P}{P(A)} \cdot \forall B \in \mathfrak{a}$ et $B \subset A$, on a :

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

Plus généralement, soit B, même non inclus dans A, élément de \mathfrak{a} ; $B \cap A \in \mathfrak{A}$.

On définit alors :

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

P_A est la probabilité conditionnelle à l'événement A.

Remarque : $(\Omega, \mathfrak{a}, P_A)$ est un espace probabilisé, mais $P_A(A) = 1$; toute la masse de probabilité est concentrée sur A.

- Interprétation :

On sait que A s'est produit et l'on cherche la probabilité de B. B se réalise "sous la forme" $B \cap A$ seulement. On dit que $P_A(B)$ est la probabilité de réalisation de B "sachant que A s'est produit" ou "conditionnellement à la réalisation de A".

- Notation :

$$P_A(B) = P(B|A) .$$

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

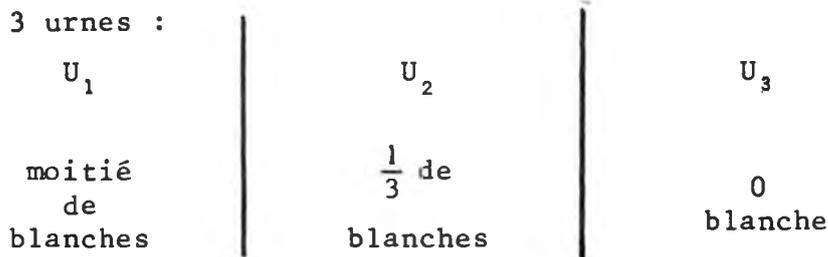
On utilise cette relation pour calculer $P(B|A)$ ou pour calculer $P(B \cap A)$ si l'on connaît, grâce aux hypothèses du problème, $P(B|A)$.

- Théorème de Bayes :

ou théorème de probabilités des causes ;
ou théorème des probabilités à postériori.

Étudions-le sur un exemple.

Soient 3 urnes :



On choisit une urne au hasard et une boule dans cette urne, au hasard.
On demande :

- 1) la probabilité de tirer une blanche ;
- 2) sachant que l'on a tiré une blanche, les probabilités de l'avoir extraite de U_1 , U_2 ou U_3 .

L'univers est l'ensemble de toutes les boules (les événements élémentaires ne sont pas équiprobables).

Soient les événements A_1 , A_2 , A_3 et B :

A_1 : on tire dans l'urne U_1 ;

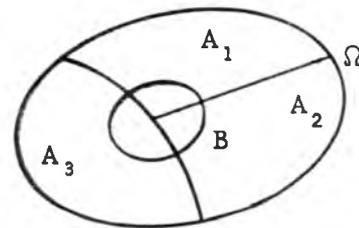
B : on tire une blanche.

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{3} .$$

$$P(B|A_1) = \frac{1}{2} \quad , \quad P(B|A_2) = \frac{1}{3} \quad , \quad P(B|A_3) = 0 .$$

$$B = (B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3) \quad (\text{événements 2 à 2 disjoints}).$$

$$\begin{aligned} P(B) &= P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3) \\ &= P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + P(A_3) \cdot P(B|A_3) \\ &= \frac{1}{3} \left[\frac{1}{2} + \frac{1}{3} + 0 \right] = \frac{5}{18} . \end{aligned}$$



Les réponses à la deuxième question sont :

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{\frac{1}{3} \times \frac{1}{2}}{\frac{5}{18}} = \frac{3}{5} .$$

$$P(A_2|B) = \frac{P(A_2 \cap B)}{P(B)} = \frac{\frac{1}{3} \times \frac{1}{3}}{\frac{5}{18}} = \frac{2}{5} .$$

$$P(A_3|B) = \frac{P(A_3 \cap B)}{P(B)} = 0 .$$

On vérifie que la somme de ces trois probabilités est bien égale à 1.

2 - Indépendance.

Si $P(A) \neq 0$, A et B sont indépendants si $P(B|A) = P(B)$,

d'où :

$$P(A \cap B) = P(A) \cdot P(B) .$$

On peut adopter cette relation comme définition de l'indépendance de deux événements.

Si $P(A) = 0$, $P(A \cap B) = 0$, puisque $A \cap B \subset A$.

D'où : un événement presque impossible est indépendant de tout autre.

Il en est de même si $P(A) = 1$.

- Propriété :

Si (A, B) est une paire d'événements indépendants, il en est de même des paires :

$$(\bar{A}, B), (A, \bar{B}), (\bar{A}, \bar{B}) .$$

Par exemple :

$$\begin{aligned} P(\bar{A} \cap B) &= P(B \setminus A) = P(B \setminus (A \cap B)) = P(B) - P(A \cap B) \\ &= P(B) - P(A) \cdot P(B) \\ &= P(B) (1 - P(A)) \\ &= P(B) \cdot P(\bar{A}) . \end{aligned}$$

- Indépendance de plus de 2 événements :

$A_i, i \in I, I$ fini ou infini (pas nécessairement dénombrable).

Les A_i sont indépendants si la probabilité de l'intersection d'un nombre fini quelconque d'entre eux distincts est égale au produit des probabilités de chacun.

Soit $\forall i_1, i_2, \dots, i_n$ distincts dans I .

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

- Propriété :

Si les $A_i (i \in I)$ sont indépendants, les $B_i (i \in I)$, où $B_i = A_i$ ou \bar{A}_i , le sont également.

On sait déjà que la propriété est vraie pour deux événements.

Soit à démontrer, par exemple :

$$P(\bar{A}_1 \cap \bar{A}_2 \cap \dots \cap \bar{A}_p \cap A_{p+1} \cap \dots \cap A_n) = P(\bar{A}_1) P(\bar{A}_2) \dots P(\bar{A}_p) P(A_{p+1}) \dots P(A_n).$$

On opère par récurrence sur p . On a :

$$\begin{aligned} P(\bar{A}_1 \cap A_2 \cap \dots \cap A_n) &= P(A_2 \cap \dots \cap A_n) - P(A_1 \cap A_2 \cap \dots \cap A_n) \\ &= P(A_2) \dots P(A_n) \cdot (1 - P(A_1)) \\ &= P(\bar{A}_1) P(A_2) \dots P(A_n). \end{aligned}$$

$$\begin{aligned} P(\bar{A}_1 \cap \bar{A}_2 \cap A_3 \cap \dots \cap A_n) &= P(\bar{A}_1 \cap A_3 \cap \dots \cap A_n) - P(\bar{A}_1 \cap A_2 \cap \dots \cap A_n) \\ &= P(\bar{A}_1) \cdot P(A_3) \dots P(A_n) - P(\bar{A}_1) \cdot P(A_2) \dots P(A_n) \\ &= P(\bar{A}_1) \cdot (1 - P(A_2)) \cdot P(A_3) \dots P(A_n) \\ &= P(\bar{A}_1) \cdot P(\bar{A}_2) \cdot P(A_3) \dots P(A_n) \end{aligned}$$

etc...

3 - Produit d'espaces probabilisables. Probabilité-produit.

Soient 2 espaces probabilisés : $(\Omega, \mathfrak{a}, P)$ et $(\Omega', \mathfrak{a}', P')$.

L'espace probabilisable produit est défini par : $\Omega \times \Omega'$, produit cartésien de Ω et de Ω' ;

$\mathfrak{a} \times \mathfrak{a}' =$ tribu engendrée par les $A \times A'$, où $A \in \mathfrak{a}$ et $A' \in \mathfrak{a}'$.

Si $\mathfrak{a} = \mathcal{P}(\Omega)$, $\mathfrak{a}' = \mathcal{P}(\Omega')$ et si Ω et Ω' sont finis, $\mathfrak{a} \times \mathfrak{a}' = \mathcal{P}(\Omega \times \Omega')$.

Probabilité-produit : dans le cas fini et avec $\mathcal{P}(\Omega \times \Omega')$, on adopte pour probabilité le produit des probabilités :

$$\pi_{ij} = P(\{(\omega_i, \omega'_j)\}) = P(\{\omega_i\}) \times P'(\{\omega'_j\}) = p_i p'_j .$$

Est-ce bien une probabilité ? Puisque $\Omega \times \Omega'$ est fini, il suffit de vérifier :

- 1) $\pi_{ij} \geq 0$;
- 2) $\sum_{i,j} \pi_{ij} = 1$.

La première condition est remplie et l'on a, si n et n' sont les cardinaux de Ω et de Ω' :

$$\sum_{i,j} \pi_{ij} = \sum_{i=1}^n \sum_{j=1}^{n'} p_i p'_j = \left(\sum_{i=1}^n p_i \right) \times \left(\sum_{j=1}^{n'} p'_j \right) = 1 \times 1 = 1 .$$

D'autre part, si A et A' sont deux sous-ensembles de Ω et de Ω' respectivement, on a :

$$\pi(A \times A') = P(A) \cdot P'(A') .$$

En effet, si l'on pose :

$$I = \{i ; \omega_i \in A\} \quad \text{et} \quad J = \{j ; \omega'_j \in A'\} ,$$

On a :

$$\pi(A \times A') = \sum_{(i,j) \in I \times J} \pi_{ij} = \left(\sum_{i \in I} p_i \right) \times \left(\sum_{j \in J} p'_j \right) = P(A) \cdot P'(A') .$$

Dans le cas général, on cherche à munir l'espace probabilisable produit d'une probabilité π telle que, quels que soient les événements A et A' de (Ω, \mathfrak{a}) et de (Ω', \mathfrak{a}') respectivement, les événements $(A \times \Omega')$ et $(\Omega \times A')$ soient indépendants et de probabilités respectives $P(A)$ et $P'(A')$.

Puisque :

$$(A \times \Omega') \cap (\Omega \times A') = A \times A',$$

on doit donc poser :

$$\pi(A \times A') = P(A) \cdot P'(A').$$

Pour les autres événements de la tribu-produit, on doit effectuer un prolongement.

Enfin, la notion de probabilité-produit peut s'étendre à un produit de plus de deux espaces et, même, à un produit d'une infinité dénombrable d'espaces.

On applique cette notion à la théorie des épreuves répétées.

IV - VARIABLES ALEATOIRES.

1 - Notion d'application mesurable. Probabilité-image.

- Définition :

Etant donné deux espaces probabilisables (Ω, \mathfrak{a}) et (Ω', \mathfrak{a}') , une application φ de Ω dans Ω' est dite mesurable si :

$$\forall A' \in \mathfrak{a}', \varphi^{-1}(A') \in \mathfrak{a}.$$

- Probabilité-image :

Soient $(\Omega, \mathfrak{a}, P)$ un espace probabilisé et φ une application mesurable de (Ω, \mathfrak{a}) dans (Ω', \mathfrak{a}') ; φ définit une probabilité de P' sur (Ω', \mathfrak{a}') de la façon suivante :

$$\forall A' \in \mathfrak{a}', P'(A') = P[\varphi^{-1}(A')].$$

C'est une probabilité :

$$P'(A') \in [0, 1] \text{ de toute évidence ;}$$

$$P'(\Omega') = P(\Omega) = 1, \text{ car } \Omega = \varphi^{-1}(\Omega') ;$$

$$\left\{ \varphi^{-1} \left(\bigcup_{n=1}^{\infty} A'_n \right) = \bigcup_{n=1}^{\infty} \left(\varphi^{-1}(A'_n) \right) ; \right.$$

$$\left. \varphi^{-1}(A'_i) \cap \varphi^{-1}(A'_j) = \varphi^{-1}(A'_i \cap A'_j) = \emptyset \text{ si } A'_i \cap A'_j = \emptyset. \right.$$

Ceci permet de démontrer :

$$P' \left(\bigcup_{n=1}^{\infty} A'_n \right) = \sum_{n=1}^{\infty} P'(A'_n),$$

lorsque les A'_n sont deux à deux disjoints.

On va appliquer ces notions au cas où $\Omega' = \mathbb{R}$ et $\mathfrak{a}' = \mathcal{B}$ (boréliens).

2 - Définition et loi d'une variable aléatoire réelle.

- Définition :

Une variable aléatoire réelle (en abrégé : v.a.r.) définie sur

$(\Omega, \mathfrak{a}, P)$ est une application mesurable de (Ω, \mathfrak{a}) dans $(\mathbb{R}, \mathcal{B})$.

On la note X :

$$X : \Omega \rightarrow \mathbb{R}$$

telle que :

$$\forall B \in \mathcal{B} \quad X^{-1}(B) \in \mathfrak{a}.$$

REMARQUE.

Si $\mathfrak{a} = \mathcal{P}(\Omega)$, toute application est mesurable.

Si $\mathfrak{a} = \{\emptyset, \Omega\}$, seules les applications constantes sont mesurables.

- Propriétés (données sans démonstration).

1) L'ensemble des variables aléatoires réelles définies sur (Ω, \mathfrak{a}) est un espace vectoriel réel.

2) L'ensemble des variables aléatoires réelles définies sur (Ω, \mathfrak{a}) est fermé pour le produit.

- Cas où Ω est fini.

Dans le cas où Ω est fini et muni d'une algèbre de Boole de parties $\mathcal{B}(\Omega)$, l'ensemble des images est fini :

$$X(\Omega) = \{x_1, x_2, \dots, x_i, \dots, x_n\} \quad (n \leq \text{Card } \Omega).$$

Alors, quel que soit le borélien B ,

$$X^{-1}(B) = \bigcup_{\{i; x_i \in B\}} X^{-1}(x_i).$$

En particulier $\{x_i\}$ est un borélien. Une condition nécessaire et suffisante pour que X soit mesurable est donc que, quel que soit x réel, on ait :

$$X^{-1}(x) \in \mathcal{B}(\Omega).$$

Autrement dit, il faut et il suffit que l'application X soit constante sur les atomes de $\mathcal{B}(\Omega)$.

- Loi d'une variable aléatoire réelle.

C'est la probabilité-image P_X sur \mathbb{R} : $\forall B \in \mathcal{B}, P_X(B) = P(X^{-1}(B))$.

On a donc l'espace probabilisé $(\mathbb{R}, \mathcal{B}, P_X)$. Considérons les générateurs de \mathcal{B} , les ensembles $]-\infty, x[$

$$\begin{aligned} P_X(]-\infty, x[) &= P(X^{-1}(]-\infty, x[)) \\ &= P(\{\omega \in \Omega ; X(\omega) < x\}), \end{aligned}$$

quantité que l'on note $P(X < x)$, par abus d'écriture.

$F_X(x) = P(X < x)$ est une probabilité.

C'est la fonction de répartition de la variable aléatoire X .

- Propriétés de F_X .

1) Elle est monotone croissante :

$$x < x' \Rightarrow]-\infty, x[\subset]-\infty, x'[\Rightarrow F_X(x) \leq F_X(x').$$

2) Elle est continue à gauche ; soit h positif :

$$\begin{aligned} F_X(x) - F_X(x - h) &= P(X < x) - P(X < x - h) \\ &= P(x - h \leq X < x) \\ &= P_X([x - h, x[). \end{aligned}$$

Si $h \xrightarrow{>} 0$, $[x - h, x[\downarrow \emptyset$; on en déduit (d'après la propriété de continuité établie au paragraphe 7 du chapitre II).

$$P_X([x - h, x[) \xrightarrow{>} 0.$$

De même :

$$F_X(x + h) - F_X(x) = P(x \leq X < x + h) = P_X([x, x + h[).$$

Si : $h \xrightarrow{>} 0$, $[x, x + h[\downarrow \{x\}$.

$$\text{D'où : } F(x + h) - F(x) \xrightarrow[h \rightarrow 0]{>} P_X(\{x\}) = P(X = x) ;$$

d'où un saut :

$$F_X(x + 0) - F_X(x) = P(X = x).$$

$$3) \quad F_X(x) \xrightarrow{x \rightarrow -\infty} 0 \qquad F_X(x) \xrightarrow{x \rightarrow +\infty} 1.$$

Si : $x \rightarrow -\infty$ $]-\infty, x[\downarrow \emptyset$, donc $F_X(x) \rightarrow 0$.

Si : $x \rightarrow +\infty$, on considère le complémentaire : $F_X(x) \rightarrow 1$.

Réciproquement, une loi de probabilité P sur $(\mathbb{R}, \mathcal{B})$ peut être définie par la donnée des probabilités des ensembles générateurs $]-\infty, x[$, c'est-à-dire par la donnée d'une fonction F ; P doit alors réaliser : $\forall x \in \mathbb{R}, P(]-\infty, x[) = F(x)$.

On peut montrer (il s'agit d'un problème de prolongement) que l'on définit ainsi une probabilité P et une seule sur $(\mathbb{R}, \mathcal{B})$, à condition que F possède les propriétés 1, 2 et 3 ci-dessus.

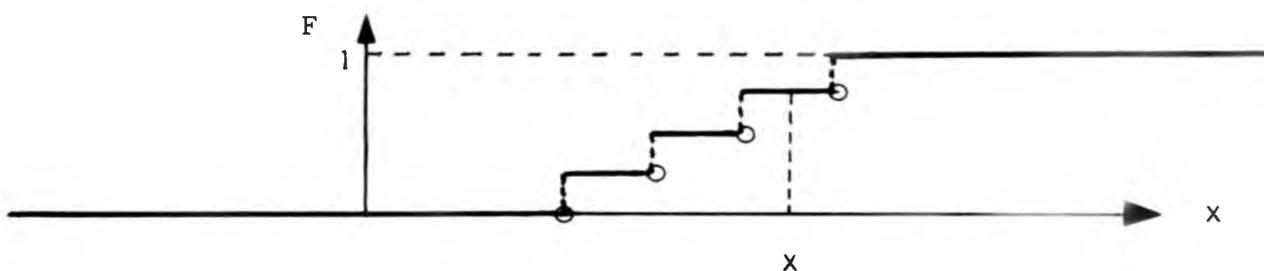
REMARQUE.

Deux variables aléatoires de même loi ne sont pas nécessairement égales (même si elles sont définies sur le même espace probabilisé) ; le fait d'avoir même loi est évidemment une relation d'équivalence (souvent notée $X \sim Y$) entre variables aléatoires définies sur le même espace probabilisé.

3 - types de lois.

* Lois discrètes.

Pour tout x , $F(x)$ est constante dans un voisinage de x ou discontinue en x .



(Les points entourés font partie du graphe).

THEOREME.

L'ensemble des points où F est discontinue est fini ou dénombrable. Cela tient au fait que F est monotone et de variation totale finie.

En effet :

- il y a un saut au plus d'amplitude appartenant à $] \frac{1}{2}, 1]$;
- il y a deux sauts au plus d'amplitude appartenant à $] \frac{1}{3}, \frac{1}{2}]$;
-
- il y a n sauts au plus d'amplitude appartenant à $] \frac{1}{n+1}, \frac{1}{n}]$;
- etc...

On a là un dénombrement des sauts.

Soient $x_i (i \in I)$ les points de discontinuité de F ;

on a : $I = \{1, 2, \dots, n\}$ ou $I = \mathbb{N}^*$

$$p_i = P(X = x_i) = F(x_i + 0) - F(x_i).$$

$$\sum_{i \in I} p_i = 1.$$

On peut avoir d'autres points images, mais :

$$P_X[\mathbb{R} - \{x_i ; i \in I\}] = 0,$$

soit : $P(\{\omega ; \forall i \in I, X(\omega) \neq x_i\}) = 0.$

Si Ω est fini, on a toujours une loi discrète.

* Lois absolument continues.

La fonction F est :

- continue en tout x réel ;
- continûment dérivable sur \mathbb{R} sauf peut-être sur un ensemble de mesure de Lebesgue nulle, (par exemple , un ensemble fini ou dénombrable).

La fonction f dérivée de F, est appelée DENSITE DE PROBABILITE de la loi ou de la variable.

$$f(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h} = \lim_{h \rightarrow 0} \frac{P(x \leq X < x+h)}{h} = \lim_{h \rightarrow 0} \frac{P(x-h \leq X < x)}{h}$$

Une fonction f est une densité de probabilité si et seulement si :

- elle est définie et continue presque partout sur \mathbb{R} ;
- elle est positive ou nulle ;
- elle réalise : $\int_{-\infty}^{+\infty} f(x) dx = 1$.

La fonction F est alors la primitive qui tend vers 0 quand la variable devient infinie négative.

Soit :

$$F(x) = \int_{-\infty}^x f(t) dt$$

* Lois singulières.

Ce sont des lois pour lesquelles la fonction F est continue mais ne remplit pas les conditions ci-dessus (on sait qu'il existe des fonctions continues sur \mathbb{R} et qui ne sont dérivables en aucun point de \mathbb{R}).

THEOREME.

Toute fonction de répartition F est combinaison convexe de trois fonctions de répartition des types ci-dessus :

$$F = \lambda_1 F_1 + \lambda_2 F_2 + \lambda_3 F_3,$$

avec : $\lambda_i \geq 0$, $\lambda_1 + \lambda_2 + \lambda_3 = 1$.

REMARQUE.

Toute combinaison convexe, finie ou dénombrable, de fonctions de répartition est une fonction de répartition.

$$F = \sum_{i \in I} \lambda_i F_i \quad (I = \{1, 2, \dots, n\} \text{ ou } I = \mathbb{N}^*),$$

avec : $\lambda_i \geq 0$, $\sum_{i \in I} \lambda_i = 1$.

On parle de combinaison convexe de lois de probabilité.

V - NOTIONS SUR L'INTEGRATION

Dans tout ce chapitre, on considérera des applications mesurables d'un espace probabilisable (Ω, \mathfrak{a}) dans $(\mathbb{R}, \mathfrak{B})$.

1 - Applications étagées

Définition. On appelle application étagée définie sur (Ω, \mathfrak{a}) , une application X de Ω dans \mathbb{R} telle qu'il existe une partition finie (A_1, A_2, \dots, A_n) de Ω en éléments de \mathfrak{a} et des nombres réels $\alpha_i (1 \leq i \leq n)$ vérifiant :

$$X = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i} .$$

Une telle application est mesurable, puisque, quel que soit le borélien B :

$$X^{-1}(B) = \bigcup_{\{i; \alpha_i \in B\}} A_i \in \mathfrak{a} .$$

On dit que l'ensemble des A_i et l'ensemble des α_i constituent une représentation de l'application étagée X . Il existe de X une et une seule représentation, dite représentation canonique, dans laquelle les α_i sont distincts : ces α_i sont les éléments de $X(\Omega)$ et les A_i correspondants en sont les images réciproques. Ou encore : la représentation canonique est obtenue à partir d'une représentation quelconque en remplaçant par leur union ceux des A_i auxquels correspond le même α_i .

Propriétés. On montrera aisément les deux propriétés suivantes .

- 1) L'ensemble des applications étagées définies sur (Ω, \mathfrak{a}) est un sous-espace vectoriel de l'espace vectoriel des applications mesurables de (Ω, \mathfrak{a}) dans $(\mathbb{R}, \mathfrak{B})$.
- 2) L'ensemble des applications étagées définies sur (Ω, \mathfrak{a}) est fermé pour le produit.

THEOREME. Toute application mesurable X de (Ω, \mathfrak{a}) dans $(\mathbb{R}, \mathfrak{B})$ est limite d'applications étagées X_n définies sur (Ω, \mathfrak{a}) . Si, de plus, X est positive ou nulle, on peut imposer aux applications X_n d'être positives ou nulles et de former une suite croissante.

Démontrons ce théorème en commençant par la deuxième partie.

Soit X une application mesurable positive ou nulle de (Ω, \mathfrak{a}) dans $(\mathbb{R}, \mathfrak{B})$.

Soit n un entier positif ; on découpe l'intervalle $[0, n]$ en $n \cdot 2^n$ intervalles égaux disjoints fermés à gauche et ouverts à droite :

$$\left[\frac{i-1}{2^n}, \frac{i}{2^n} \right[, \quad (1 \leq i \leq n \cdot 2^n)$$

et l'on pose :

$$\forall \omega \in \Omega , \quad X_n(\omega) = \begin{cases} \frac{i-1}{2^n} & \text{si } \frac{i-1}{2^n} \leq X(\omega) < \frac{i}{2^n} ; \\ n & \text{si } X(\omega) \geq n. \end{cases}$$

X_n est une application étagée : on en a une représentation en posant :

$$\begin{cases} A_i = X^{-1} \left(\left[\frac{i-1}{2^n}, \frac{i}{2^n} \right[\right) , & \alpha_i = \frac{i-1}{2^n} , \quad (1 \leq i \leq n \cdot 2^n), \\ A_0 = X^{-1} \left(\left[n, +\infty \right[\right) , & \alpha_0 = n , \end{cases}$$

et en excluant ceux des A_i qui seraient vides ; on remarque que les A_i sont bien des éléments de \mathfrak{a} .

Enfin, on montrera aisément que, pour tout élément ω de Ω , on a :

$$\begin{cases} X_n(\omega) \geq 0 ; \\ X_{n+1}(\omega) \geq X_n(\omega) ; \\ X_n(\omega) \xrightarrow[n \rightarrow \infty]{} X(\omega) . \end{cases}$$

Dans le cas général où X est de signe quelconque, on remarque que :

$$X = X^+ - X^- ,$$

où l'on a posé :

$$X^+ = \text{Sup}(X, 0) \quad \text{et} \quad X^- = - \text{Inf}(X, 0) ,$$

c'est-à-dire :

$$X^+(\omega) = \begin{cases} X(\omega) & \text{si } X(\omega) \geq 0 ; \\ 0 & \text{sinon ;} \end{cases} \quad X^-(\omega) = \begin{cases} 0 & \text{si } X(\omega) \geq 0 ; \\ -X(\omega) & \text{sinon.} \end{cases}$$

On applique la démonstration de la page précédente aux applications mesurables positives ou nulles X^+ et X^- et l'on utilise ensuite le fait que les applications étagées constituent un espace vectoriel.

Remarque : $|X| = X^+ + X^-$.

2 - Intégration des applications étagées

On suppose désormais que les applications sont définies sur un espace probabilisé $(\Omega, \mathfrak{a}, P)$.

Propriété et définition. Etant donné une application étagée X définie sur l'espace probabilisé $(\Omega, \mathfrak{a}, P)$, la quantité $\sum_{i=1}^n \alpha_i P(A_i)$, où les A_i et les α_i ($1 \leq i \leq n$) constituent une représentation de X , ne dépend pas de la représentation choisie ; on l'appelle intégrale de X et on la note $\int X \, dP$.

Pour établir la propriété, il suffit de remarquer que la quantité ci-dessus obtenue à partir d'une représentation quelconque de X est égale à celle que l'on obtient à partir de la représentation canonique de X .

Intégrale sur un événement. Soit A un événement ($A \in \mathfrak{a}$). On appelle intégrale de X sur A et l'on note $\int_A X \, dP$ l'intégrale $\int X \cdot \mathbb{1}_A \, dP$ ($X \cdot \mathbb{1}_A$ est bien une application étagée, puisqu'elle est le produit de deux applications étagées).

Si les A_i et les α_i ($1 \leq i \leq n$) constituent une représentation de X , on a :

$$X = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i},$$

soit
$$X \cdot \mathbb{1}_A = \sum_{i=1}^n \alpha_i (\mathbb{1}_{A_i} \cdot \mathbb{1}_A) = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i \cap A}.$$

D'où
$$\int_A X \, dP = \sum_{i=1}^n \alpha_i P(A_i \cap A).$$

En particulier, si X est l'application constante égale à 1 ($X = \mathbb{1}_\Omega$), on a :

$$\int_A 1 \, dP = P(A)$$

ou, avec une notation plus légère : $\int_A \, dP = P(A)$.

On remarque, enfin, que $\int X \, dP$ n'est autre que $\int_\Omega X \, dP$;

on aura parfois intérêt à adopter cette dernière notation pour éviter des confusions et, même, à écrire :

$$\int_{\Omega} X(\omega) dP(\omega) \quad \text{ou} \quad \int_A X(\omega) dP(\omega) ,$$

ω étant alors dite "variable muette" (voir la fin de ce chapitre, par exemple).

Propriétés de l'intégrale. Dans les énoncés ci-dessous, X et Y désignent des applications étagées définies sur $(\Omega, \mathfrak{a}, P)$.

1) L'intégrale est une forme linéaire sur l'espace vectoriel des applications étagées définies sur $(\Omega, \mathfrak{a}, P)$. On le montrera aisément.

2) Si l'application X est P-presque partout positive ou nulle (c'est-à-dire si l'ensemble sur lequel elle est strictement négative est de probabilité nulle), son intégrale est positive ou nulle.

Cette propriété est évidente.

3) Si $X \geq Y$ P.p.p. , $\int X dP \geq \int Y dP$.

C'est un corollaire des propriétés précédentes ("P.p.p." signifie "P- presque partout").

4) $\int |X+Y| dP \leq \int |X| dP + \int |Y| dP$.

C'est un corollaire des propriétés précédentes.

5) $|\int X dP| \leq \int |X| dP$.

En effet, $|X| \geq X$ et $|X| \geq -X$ et l'on applique la propriété 3).

6) Si A est un événement et si les nombres réels λ et μ réalisent :

$$\forall \omega \in A , \quad \lambda \leq X(\omega) \leq \mu ,$$

on a : $\lambda P(A) \leq \int_A X dP \leq \mu P(A)$.

En effet, $\forall \omega \in \Omega$, on a : $\lambda \cdot \mathbb{1}_A \leq X(\omega) \cdot \mathbb{1}_A \leq \mu \cdot \mathbb{1}_A$.

7) Si $X \geq 0$ P.p.p. et si A et B sont deux événements tels que $A \subset B$, on a :

$$\int_A X dP \leq \int_B X dP .$$

En effet, $X \cdot \mathbb{1}_A \leq X \cdot \mathbb{1}_B$ P.p.p.

8) Additivité finie ou dénombrable.

Soient B_j ($j \in J$) des événements deux à deux disjoints, J étant fini ou dénombrable. Posons $B = \bigcup_{j \in J} B_j$.

On a :

$$\int_B X dP = \sum_{j \in J} \int_{B_j} X dP .$$

En effet, si les A_i et les α_i ($1 \leq i \leq n$) constituent une représentation de X , on a :

$$\begin{aligned} \int_B X \, dP &= \sum_{i=1}^n \alpha_i P(A_i \cap B) = \sum_{i=1}^n \alpha_i \left[\sum_{j \in J} P(A_i \cap B_j) \right] \\ &= \sum_{j \in J} \left[\sum_{i=1}^n \alpha_i P(A_i \cap B_j) \right] = \sum_{j \in J} \int_{B_j} X \, dP. \end{aligned}$$

Remarque . On pourra chercher si les propriétés ci-dessus sont encore vraies en y remplaçant les intégrales sur Ω par des intégrales sur un événement quelconque.

3 - Applications intégrables

On considère maintenant des applications mesurables de l'espace probabilisé (Ω, \mathcal{a}, P) dans $(\mathbb{R}, \mathcal{B})$, c'est-à-dire des v.a.r. définies sur (Ω, \mathcal{a}, P) .

Théorème et définition. Soit X une v.a.r. positive ou nulle définie sur (Ω, \mathcal{a}, P) et soit une suite croissante d'applications étagées X_n tendant vers X . L'existence et la valeur de la limite de $\int X_n \, dP$ ne dépendent pas de la suite choisie, pourvu qu'elle soit croissante et tende vers X . Si cette limite existe, on dit que X est intégrable ou sommable et la limite, notée $\int X \, dP$, est appelée intégrale de X .

Définition. On dit qu'une v.a.r. définie sur (Ω, \mathcal{a}, P) est intégrable ou sommable si les v.a.r. X^+ et X^- (définies p.26) sont intégrables. L'intégrale de X , notée $\int X \, dP$, est alors définie par :

$$\int X \, dP = \int X^+ \, dP - \int X^- \, dP.$$

On peut enfin définir l'intégrale de X sur un événement quelconque A : c'est $\int X \cdot \mathbb{1}_A \, dP$ et on la note $\int_A X \, dP$; elle existe dès que X est intégrable.

Propriétés.

0) L'ensemble des v.a.r. intégrables définies sur (Ω, \mathcal{a}, P) est un espace vectoriel réel.

1) à 8) Les propriétés 1) à 8) énoncées au paragraphe précédent pour des applications étagées sont encore vraies pour des v.a.r. intégrables.

Autres propriétés. X désigne une v.a.r. intégrable définie sur $(\Omega, \mathfrak{a}, P)$.

1) Si X est P- presque partout positive ou nulle,

$$\int X \, dP = 0 \quad \text{et} \quad X = 0 \text{ P.p.p.}$$

sont des propositions équivalentes.

2) Si A est un événement presque impossible ($P(A) = 0$),

$$\int_A X \, dP = 0 .$$

3) Si X est strictement positive P- presque partout sur l'événement A et si $\int_A X \, dP = 0$, $P(A) = 0$.

4) Si $\int_A X \, dP = 0$ pour tout événement A, X est nulle P- presque partout.

Enonçons enfin deux théorèmes.

1) Si X et Y sont des v.a.r. définies sur $(\Omega, \mathfrak{a}, P)$, si Y est intégrable et si $|X| \leq |Y|$ P.p.p., X est intégrable.

2) Si X est une v.a.r., l'intégrabilité de X et celle de $|X|$ sont équivalentes.

Changement de variable. Soit X une v.a.r. définie sur $(\Omega, \mathfrak{a}, P)$. Si h est une application mesurable de $(\mathbb{R}, \mathfrak{B})$ dans lui-même, $h \circ X$ est une v.a.r. définie sur $(\Omega, \mathfrak{a}, P)$.

THEOREME. Pour que $(h \circ X)$ soit intégrable sur $(\Omega, \mathfrak{a}, P)$, il faut et il suffit que h le soit sur $(\mathbb{R}, \mathfrak{B}, P_X)$; les intégrales sont alors égales :

$$\int_{\Omega} (h \circ X) \, dP = \int_{\mathbb{R}} h \, dP_X .$$

En particulier, si h est l'application identique sur \mathbb{R} ($h = \text{Id}_{\mathbb{R}}$), on a :

$$\int_{\Omega} X \, dP = \int_{\mathbb{R}} \text{Id}_{\mathbb{R}} \, dP_X ,$$

ce que l'on pourra noter : $\int_{\Omega} X(\omega) dP(\omega) = \int_{\mathbf{R}} x dP_X(x)$.

On se ramènera donc à calculer des intégrales sur \mathbf{R} . On montre, enfin, les formules "pratiques" suivantes :

$$\int_{\mathbf{R}} h(x) dP_X(x) = \begin{cases} \sum_{i \in I} p_i h(x_i) & \text{si la loi de X est discrète ;} \\ \int_{\mathbf{R}} h(x) f(x) dx & \text{si la loi de X est absolument continue} \\ & \text{de densité f.} \end{cases}$$

Si la loi P_X de X est combinaison convexe d'une loi discrète P_1 et d'une loi absolument continue P_2 :

$$P_X = \lambda_1 P_1 + \lambda_2 P_2 \quad (\lambda_1 > 0, \lambda_2 > 0, \lambda_1 + \lambda_2 = 1) ,$$

on a :

$$\int_{\mathbf{R}} h(x) dP_X(x) = \lambda_1 \int_{\mathbf{R}} h(x) dP_1(x) + \lambda_2 \int_{\mathbf{R}} h(x) dP_2(x) .$$

VI - MOMENTS.

1 - Espérance mathématique.

C'est, par définition, le nombre : $\int_{\Omega} X(\omega) d P(\omega)$, s'il existe.

On la note : $E(X)$. On l'appelle aussi "moyenne de X".

Propriétés :

- $E(X + Y) = E(X) + E(Y)$, donc $E(X + a) = E(X) + a$;
- $E(\lambda X) = \lambda E(X)$;
- $E(a X + b) = a E(X) + b$, par conséquent.

Si Ω est fini, l'application X est étagée ; si elle prend les valeurs a_i sur le n atomes A_i de l'algèbre de Boole, on a :

$$E(X) = \sum_{i=1}^n a_i \cdot P(A_i).$$

soit, sur l'ensemble image $X(\Omega) = \{x_1, x_2, \dots, x_j, \dots, x_k\}$ ($k \leq n$) :

$$E(X) = \sum_{j=1}^k x_j P(X = x_j),$$

après avoir réuni les A_i qui ont même image.

Définition : on dit qu'une variable aléatoire est centrée si sa moyenne est nulle. En particulier, si X a pour moyenne m, $(X-m)$ est centrée.

2 - Moments d'ordre quelconque.

On appelle moment d'ordre r ($r \in \mathbb{N}^*$) le nombre (s'il existe) :

$$\mu_r = E(X^r) = \int_{\Omega} X^r(\omega) d P(\omega) = \int_{\mathbb{R}} x^r d P_X(x) = \begin{cases} \sum_{i \in I} p_i x_i^r \\ \text{ou} \\ \int_{-\infty}^{+\infty} x^r f(x) dx. \end{cases}$$

(L'application h considérée au chapitre V est ici $h : x \mapsto x^r$).

On montre que, si le moment d'ordre r existe, tous ceux d'ordre inférieur existent. Il peut donc y avoir une borne supérieure des ordres dont le moment existe. On parlera de variable de carré sommable, de cube sommable, etc...

REMARQUE.

On peut définir les moments d'ordre non entier si l'application X est positive ou nulle.

3 - Variance et écart-type.

Posons $m = E(X)$. On appelle variance de X le moment centré d'ordre deux de X , c'est-à-dire le nombre (s'il existe) défini par :

$$V(X) = E[(X - m)^2] = \left\{ \begin{array}{l} \sum_{i \in I} p_i (x_i - m)^2 \\ \text{ou} \\ \int_{-\infty}^{+\infty} (x - m)^2 f(x) dx . \end{array} \right.$$

La variance de X donne une idée de "l'étalement" de X autour de sa moyenne. On a, en développant :

$$V(X) = E(X^2) - 2m E(X) + m^2 = E(X^2) - [E(X)]^2.$$

Pour que la variance de X existe, il faut et il suffit que X soit de carré sommable.

Propriétés :

- $V(X + a) = V(X)$;
- $V(aX) = a^2 V(X)$;
- $V(aX + b) = a^2 V(X)$, par conséquent.

Écart-type ou écart moyen quadratique. C'est le nombre :

$$\sigma(X) = \sqrt{V(X)}.$$

REMARQUE.

Tous ces moments sont liés à une loi plutôt qu'à une variable ; on parlera de moyenne ou de variance d'une loi donnée : c'est la moyenne ,

ou la variance, de toute variable aléatoire qui suit cette loi.

4 - Inégalité de Bienaymé-Tchébicheff.

Soit X une v. a. r. de carré sommable, c'est-à-dire admettant des moments d'ordre 1 et 2, définie sur $(\Omega, \mathfrak{a}, P)$.

Posons : $m = E(X)$. on a :

$$\begin{aligned}\sigma^2(X) &= E[(X - m)^2] = \int_{\Omega} (X(\omega) - m)^2 d P(\omega) \\ &= \int_{\Omega_1} (X(\omega) - m)^2 d P(\omega) + \int_{\overline{\Omega}_1} (X(\omega) - m)^2 d P(\omega),\end{aligned}$$

avec : $\Omega_1 = \{\omega ; | X(\omega) - m | \geq t \sigma\}$ où $t > 0$; $\sigma = \sqrt{V(X)}$.

$$\text{On a : } \sigma^2(X) \geq \int_{\Omega_1} (X(\omega) - m)^2 d P(\omega) \geq t^2 \sigma^2 \int_{\Omega_1} d P(\omega),$$

$$\text{d'où : } \sigma^2 \geq t^2 \sigma^2 P(\Omega_1).$$

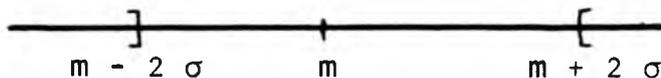
On en déduit ($\sigma \neq 0$) :

$$P(\Omega_1) \leq \frac{1}{t^2},$$

soit : $\forall t > 0$

$$P[|X - m| \geq t \sigma] \leq \frac{1}{t^2}$$

Si t est inférieur ou égal à 1, cette inégalité n'a aucun intérêt. Sinon, prenons $t = 2$, par exemple :



La probabilité pour que X appartienne à l'une ou l'autre des deux demi-droites fermées est inférieure ou égale à $1/4$, ou encore, la probabilité pour que X appartienne à l'intervalle ouvert est supérieure à $3/4$.

L'inégalité peut aussi s'écrire :

$$\forall t > 0, P[|X - m| < t \sigma] \geq 1 - \frac{1}{t^2},$$

ou encore, en posant $\varepsilon = t \sigma$:

$$\forall \varepsilon > 0, P[|X - m| \geq \varepsilon] \leq \frac{\sigma^2}{\varepsilon^2},$$

inégalité qui n'a d'intérêt que si $\varepsilon > \sigma$.

On pourra interpréter ces différentes inégalités en se souvenant que la variance donne une idée de l'"étalement" de X autour de sa moyenne (paramètre de dispersion).

VII - VARIABLES ALEATOIRES INDEPENDANTES.

1 - Définition.

n variables aléatoires X_i définies sur (Ω, \mathcal{a}, P) sont indépendantes si, quels que soient les boréliens $B_i \in \mathcal{B}$ ($1 \leq i \leq n$), les événements $X_i^{-1}(B_i)$ sont indépendants. Il suffit de vérifier que :

$$P \left[\bigcap_{i=1}^n X_i^{-1}(B_i) \right] = \prod_{i=1}^n P \left[X_i^{-1}(B_i) \right] ,$$

en prenant, au besoin, certains B_i égaux à \mathbb{R} .

Des variables en nombre infini sont indépendantes si, par définition, tout sous-ensemble fini de ces variables est constitué de variables indépendantes.

2 - Propriétés.

* Si des v. a. sont indépendantes, des fonctions mesurables de ces variables sont indépendantes : si les X_i ($i \in I$) sont indépendantes et si les f_i ($i \in I$) sont des applications mesurables de $(\mathbb{R}, \mathcal{B})$ dans lui-même, les v. a. $Y_i = f_i(X_i)$ sont indépendantes.

* Si les v. a. X_i ($1 \leq i \leq n$) sont indépendantes, quels que soient les réels x_i ($1 \leq i \leq n$), on a :

$$P \left[\bigcap_{i=1}^n \{X_i < x_i\} \right] = \prod_{i=1}^n P(X_i < x_i) ,$$

en considérant les boréliens $B_i =]-\infty, x_i[$.

On écrit cette égalité sous la forme :

$$F_{\vec{X}}(\vec{x}) = \prod_{i=1}^n F_{X_i}(x_i).$$

$F_{\vec{X}}$ est la fonction de répartition du vecteur aléatoire $\vec{X} = (X_1, X_2, \dots, X_n)$.

Réciproquement, on peut montrer (c'est un problème de prolongement) qu'il suffit que l'égalité ci-dessus soit satisfaite pour tout vecteur \vec{X} de \mathbb{R}^n pour que les v. a. r. X_i soient indépendantes.

Nous reviendrons sur ces questions au chapitre XI.

* Si les v. a. r. X_i sont discrètes, on écrira plutôt que, quels que soient les réels x_i :

$$P \left[\bigcap_{i=1}^n \{X_i = x_i\} \right] = \prod_{i=1}^n P(X_i = x_i),$$

égalité qui n'a d'intérêt que si les probabilités écrites sont non nulles. Prenons, par exemple, le cas de deux variables X et Y prenant les valeurs respectives :

$$(x_1, x_2, \dots, x_m) \text{ et } (y_1, y_2, \dots, y_n).$$

$$\forall i \in \{1, 2, \dots, m\}, \quad \forall j \in \{1, 2, \dots, n\},$$

$$P \left[\{X = x_i\} \cap \{Y = y_j\} \right] = P(X = x_i) \cdot P(Y = y_j).$$

Ceci revient à faire le produit des deux espaces images.

THEOREME.

Si n variables aléatoires indépendantes sont sommables, leur produit est sommable et l'espérance mathématique du produit est le produit des espérances mathématiques.

Conséquence.

La variance de la somme de n variables aléatoires indépendantes est la somme de leurs variances.

Montrons le pour deux variables X et Y , par exemple :

$$\begin{aligned} V(X + Y) &= E[(X + Y)^2] - [E(X + Y)]^2 \\ &= E[X^2 + 2XY + Y^2] - [E(X) + E(Y)]^2 \\ &= E(X^2) - [E(X)]^2 + E(Y^2) - [E(Y)]^2 + 2[E(XY) - E(X) \cdot E(Y)] \\ &= V(X) + V(Y). \end{aligned}$$

VIII - FONCTION GENERATRICE. FONCTION CARACTERISTIQUE. PRODUIT DE CONVOLUTION

1 - Fonction génératrice.

- Définition :

Soit X une variable aléatoire à valeurs entières positives ou nulles. La fonction génératrice de X est la fonction G_X définie par :

$$G_X : u \mapsto E(u^X) ;$$

$$G_X(u) = E(u^X).$$

X est discrète ; si nous posons :

$$p_n = P(X = n) \quad \left(\sum_{n=0}^{\infty} p_n = 1 \right),$$

on a :
$$G_X(u) = \sum_0^{\infty} p_n u^n \quad (\text{série entière ou polynôme}).$$

On voit immédiatement que, pour qu'une fonction soit génératrice, il faut et il suffit :

- qu'elle soit développable en série entière ;
- que les coefficients du développement soient positifs ou nuls et de somme égale à 1.

On a une série absolument convergente sur le disque fermé de rayon 1, donc de rayon de convergence supérieur ou égal à 1.

- Calcul des moments à partir de la fonction génératrice :

$$G'_X(u) = \sum_0^{\infty} n p_n u^{n-1} .$$

$$u = 1 : G'_X(1) = \sum_0^{\infty} n p_n = E(X). \quad (\text{Si ces quantités existent}).$$

$$G''_X(u) = \sum_0^{\infty} n(n-1) p_n u^{n-2}$$

$$u = 1 : G_X''(1) = \sum_0^{\infty} n(n-1) p_n = E(X(X-1)) \\ = E(X^2) - E(X).$$

(Si ces quantités existent).

On obtient ainsi les moments factoriels successifs :

$$E(X), E[X(X-1)], E[X(X-1)(X-2)], \text{ etc..}$$

2 - Fonction caractéristique.

Soit X une variable aléatoire. La fonction caractéristique de X est la fonction φ_X définie par :

$$\varphi_X(t) = E(e^{itX}),$$

où t est une variable réelle.

La fonction caractéristique de X est déterminée par la loi de X ; on montre qu'elle la caractérise, d'où son nom.

Remarque : Si X est à valeurs entières positives ou nulles, on a :

$$\varphi_X(t) = G_X(e^{it}).$$

Propriétés.

$$1) \varphi_X(0) = 1 .$$

$$2) \forall t \in \mathbb{R} , |\varphi_X(t)| \leq 1 .$$

$$3) \forall t \in \mathbb{R} , \varphi_X(-t) = \overline{\varphi_X(t)} .$$

Toutefois, ces propriétés ne suffisent pas pour qu'une fonction donnée soit une fonction caractéristique.

Enfin, on peut calculer les moments de X à partir de sa fonction caractéristique :

$$\varphi_X^{(r)}(t) = E(i^r X^r e^{itX}) ,$$

$$\varphi_X^{(r)}(0) = i^r E(X^r). \quad (\text{Si ces quantités existent}).$$

3 - Produit de convolution.

Définition. On appelle produit de convolution de deux lois P_1 et P_2 de $(\mathbb{R}, \mathcal{B})$ la loi suivie par la somme de deux v.a. indépendantes de lois respectives P_1 et P_2 .

On note : $P_1 * P_2$.

Cette composition est, de toute évidence, commutative et associative.

Si P_1 et P_2 sont des lois discrètes sur les entiers positifs ou nuls, on a (X suit la loi P_1 , Y suit la loi P_2) :

$$G_{X+Y}(u) = E(u^{X+Y}) = E(u^X u^Y) = E(u^X) \cdot E(u^Y)$$

$$G_{X+Y}(u) = G_X(u) \cdot G_Y(u)$$

La fonction génératrice de la loi $P_1 * P_2$ est le produit des fonctions génératrices des lois P_1 et P_2 .

Plus généralement :

$$\varphi_{X+Y}(t) = E(e^{it(X+Y)}) = E(e^{itX} \cdot e^{itY}) = E(e^{itX}) \cdot E(e^{itY}).$$

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t).$$

La fonction caractéristique de la loi $P_1 * P_2$ est le produit des fonctions caractéristiques des lois P_1 et P_2 .

On remarquera que le produit de convolution de deux lois, défini à partir de variables aléatoires qui suivent ces lois, ne dépend pas du choix de ces variables, pourvu qu'elles soient indépendantes. Il s'agit bien d'une opération sur les lois de probabilité, c'est-à-dire sur des probabilités définies sur $(\mathbb{R}, \mathcal{B})$.

IX - LOIS USUELLES.

1 - Lois discrètes.

* Loi de Dirac.

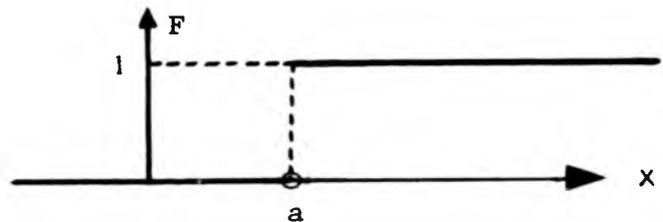
La loi de Dirac au point a , notée D_a , est définie par :

$$P(X = a) = 1,$$

soit : $X(\omega) = a$ sauf sur un ensemble de probabilité nulle.

On écrit $X(\omega) = a$ P.p.p. (se lit "P presque partout").

Fonction de répartition :



On a : $E(X) = a$, $\sigma^2(X) = 0$.

Réciproquement : $\sigma^2 = 0 \Rightarrow$ loi de Dirac, car :

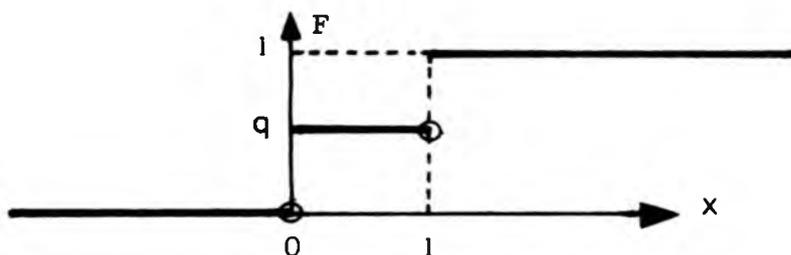
$$\int_{\Omega} (X(\omega) - m)^2 d P(\omega) = 0 \Rightarrow X(\omega) = m \text{ P.p.p.}$$

* Loi de Bernoulli.

Elle est définie par :

$$\begin{cases} P(X = 1) = p ; \\ P(X = 0) = q. \end{cases} \quad \text{avec } p + q = 1$$

On la note : $B(p)$.



$$E(X) = p, E(X^2) = p, E(X^r) = p(r > 0).$$

$$\sigma^2(X) = p - p^2 = pq.$$

$$G_X(u) = pu^1 + qu^0 = pu + q.$$

* Loi binomiale.

On effectue n épreuves indépendantes ; à l'issue de chacune d'elles peut se produire, avec la probabilité p , un certain événement E . On pose : $q = 1 - p$.

Soit X le nombre de réalisations de E à l'issue des n épreuves. On peut montrer, par des méthodes combinatoires, que :

$$P(X = k) = C_n^k p^k q^{n-k} \quad (0 \leq k \leq n).$$

On peut aussi associer à chaque épreuve une variable de Bernoulli définie par :

$$X_i = \begin{cases} 1 & \text{si } E \text{ se réalise à la } i\text{-ième épreuve ;} \\ 0 & \text{sinon.} \end{cases}$$

Toutes les variables X_i suivent $B(p)$ et sont indépendantes. On a :

$$X = \sum_{i=1}^n X_i.$$

D'où :

$$G_X(u) = \prod_{i=1}^n G_{X_i}(u) = (pu + q)^n = \sum_{k=0}^n C_n^k (pu)^k q^{n-k}.$$

On sait que $P(X = k)$ est le coefficient de u^k dans le développement de $G_X(u)$. D'où le résultat déjà donné ci-dessus.

On appelle la loi de X la loi binomiale et on la note $B(n, p)$

$$X = \sum_{i=1}^n X_i \Rightarrow E(X) = np \quad \text{et} \quad \sigma^2(X) = npq.$$

Enfin :

$$B(n, p) * B(m, p) = B(m+n, p).$$

* Loi de Poisson.

On sait que :

$$e^\lambda = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!}, \quad \text{ou} \quad \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} = 1.$$

Si λ est positif, on a une série entière à coefficients positifs de somme 1. On définit alors la loi de Poisson $P(\lambda)$ par :

$$P(X = n) = e^{-\lambda} \frac{\lambda^n}{n!} \quad (n \in \mathbb{N}).$$

$$\begin{aligned} G_X(u) = E(u^X) &= \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} u^n \\ &= e^{-\lambda} e^{\lambda u} = e^{\lambda(u-1)}. \end{aligned}$$

$$G_X'(u) = \lambda e^{\lambda(u-1)}, \quad G_X''(u) = \lambda^2 e^{\lambda(u-1)}, \quad G_X^{(p)} = \lambda^p e^{\lambda(u-1)},$$

d'où : $E(X) = \lambda,$

$$E[X(X-1)] = \lambda^2,$$

$$E(X^2) = \lambda^2 + \lambda,$$

$$\sigma^2(X) = \lambda.$$

Le moment factoriel d'ordre p est :

$$G_X^{(p)}(1) = \lambda^p.$$

Enfin : $P(\lambda) * P(\mu) = P(\lambda + \mu).$

* Loi géométrique.

Soit une suite illimitée d'épreuves indépendantes ; à l'issue de chacune d'elles, peut se produire, avec la probabilité p , un événement E . On pose : $q = 1 - p$.

Soit X le rang de la première réalisation de E . On a :

$$P(X = n) = q^{n-1} p \quad (n \in \mathbb{N}^*).$$

Les probabilités sont en progression géométrique, d'où le nom de la loi.

On a :

$$G_X(u) = \sum_{n=1}^{\infty} p q^{n-1} u^n = \frac{pu}{1-qu}$$

Le rayon de convergence, égal à $\frac{1}{q}$, est inférieur à 1, ce qui justifie les calculs qui suivent.

$$G'_X(u) = \frac{p}{(1-qu)^2}, \quad \text{d'où :} \quad E(X) = G'_X(1) = \frac{1}{p}.$$

$$G''_X(u) = \frac{2pq}{(1-qu)^3}, \quad \text{d'où :} \quad E[X(X-1)] = \frac{2pq}{p^3} = \frac{2q}{p^2},$$

$$\sigma^2(X) = \frac{q}{p^2}.$$

2 - Lois absolument continues.

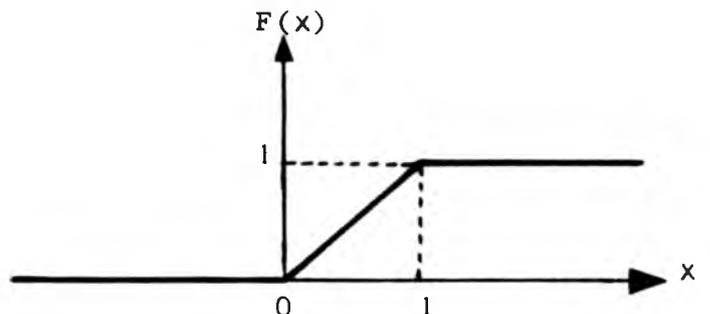
* Loi uniforme sur $[0, 1]$.

La densité est constante sur $]0, 1[$, nulle en dehors.

D'où :

$$f_X(x) = \begin{cases} 1 & \text{si } 0 < x < 1 ; \\ 0 & \text{sinon.} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq 0 ; \\ x & \text{si } 0 \leq x \leq 1 ; \\ 1 & \text{si } x \geq 1. \end{cases}$$



$$E(X) = \int_0^1 x f(x) dx = \frac{1}{2} \quad \text{et} \quad E(X^2) = \int_0^1 x^2 f(x) dx = \frac{1}{3} .$$

$$\text{D'où : } \sigma^2(X) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12} .$$

Dans le cas général d'une loi uniforme sur le segment $[a, b]$ ($a < b$), la densité est égale à $\frac{1}{b-a}$ sur $]a, b[$ et l'on a :

$$E(X) = \frac{a+b}{2} \quad \text{et} \quad \sigma^2(X) = \frac{(b-a)^2}{12} .$$

* Loi normale.

Admettons :

$$\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi} .$$

La fonction f définie par :

$$f(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} ,$$

est donc une densité de probabilité. La fonction de répartition correspondante est :

$$F(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{t^2}{2}} dt .$$

Elle est tabulée. On a :

$$F(u) + F(-u) = 1 .$$

On montre que cette loi admet pour fonction caractéristique la fonction :

$$t \longmapsto e^{-\frac{t^2}{2}} .$$

Si U est une v. a. r. suivant cette loi, on a :

$$E(U) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u e^{-\frac{u^2}{2}} du = 0 .$$

$$E(U^2) = \sigma^2(U) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u^2 e^{-\frac{u^2}{2}} du = 1 \quad (\text{grâce à une intégration par parties}).$$

Plus généralement :

$$\begin{cases} E(U^{2p+1}) = 0, \\ E(U^{2p}) = 1 \cdot 3 \cdot 5 \dots (2p-1). \end{cases}$$

La loi est dite centrée, car $m = 0$; elle est dite réduite car $\sigma^2 = 1$.

On la note $N(0, 1)$: loi normale centrée réduite.

Soit U une v. a. r. de loi $N(0, 1)$; quelle est la loi de la v. a. r. X définie par :

$$X = aU + b \quad (a \in \mathbb{R}^*, b \in \mathbb{R}) ?$$

$$F_X(x) = P(X < x)$$

$$= P(aU + b < x) = \begin{cases} P\left(U < \frac{x-b}{a}\right) = F_U\left(\frac{x-b}{a}\right), & \text{si } a > 0 ; \\ P\left(U > \frac{x-b}{a}\right) = 1 - F_U\left(\frac{x-b}{a}\right), & \text{si } a < 0 . \end{cases}$$

$$f_X(x) = \begin{cases} \frac{1}{a} f_U\left(\frac{x-b}{a}\right), & \text{si } a > 0 ; \\ -\frac{1}{a} f_U\left(\frac{x-b}{a}\right), & \text{si } a < 0 . \end{cases}$$

$$\begin{aligned} \text{D'où : } f_X(x) &= \frac{1}{|a|} f_U\left(\frac{x-b}{a}\right) \\ &= \frac{1}{|a|\sqrt{2\pi}} e^{-\frac{(x-b)^2}{2a^2}} . \end{aligned}$$

$$\text{On a : } E(X) = a E(U) + b = b,$$

$$\sigma^2(X) = E[(X-b)^2] = E(a^2 U^2) = a^2 E(U^2) = a^2.$$

Posons : $m = b$ et $\sigma = |a|$.

La loi de X est dite normale et on la note $N(m, \sigma^2)$. Elle a une densité f_X définie par :

$$f_X(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

Réciproquement, si X suit la loi $N(m, \sigma^2)$, $\frac{X-m}{\sigma}$ suit la loi $N(0, 1)$.

Plus généralement, si X suit la loi $N(m, \sigma^2)$, $Y = aX + b$ suit $N(am + b, a^2 \sigma^2)$.

On le montre aisément par la méthode déjà utilisée ci-dessus.

* Fonction caractéristique et propriété de convolution.

Si X suit la loi $N(m, \sigma^2)$, $U = \frac{X-m}{\sigma}$ suit la loi $N(0, 1)$.

d'où :

$$\begin{aligned} \varphi_X(t) &= E\left(e^{itX}\right) = E\left(e^{it(m + \sigma U)}\right) = e^{itm} \cdot E\left(e^{it\sigma U}\right) = e^{itm} \cdot \varphi_U(\sigma t) \\ \varphi_X(t) &= e^{itm} - \frac{\sigma^2 t^2}{2} \end{aligned}$$

D'où la propriété de convolution :

$$N(m, \sigma^2) * N(m', \sigma'^2) = N(m + m', \sigma^2 + \sigma'^2).$$

X - CONVERGENCE.

1 - Problème.

Soit une suite de variables aléatoires X_n ($n \geq 1$) définies sur $(\Omega, \mathfrak{a}, P)$. Dans quelles conditions dira-t-on que X_n tend vers X , X étant une variable aléatoire également définie sur $(\Omega, \mathfrak{a}, P)$?

Nous aurons plusieurs types de convergence.

Rappelons que X_n et X sont des applications de Ω dans \mathbb{R} .

2 - Convergence en loi.

- Définition.

On dit que X_n tend, en loi, vers X et l'on note :

$$X_n \xrightarrow[n \rightarrow \infty]{L} X,$$

si
$$F_{X_n}(x) \xrightarrow[n \rightarrow \infty]{} F_X(x),$$

en tout x réel où F_X est continue.

En fait, on devrait plutôt écrire :

$$P_{X_n} \longrightarrow P_X ;$$

il s'agit d'une convergence sur les lois plutôt que sur les variables aléatoires.

On montre que, pour que X_n tende en loi vers X , il faut et il suffit que :

$$\forall t \in \mathbb{R} , \quad \varphi_{X_n}(t) \longrightarrow \varphi_X(t).$$

- Exemple 1.

Soit :
$$Y_n = \sup_{1 \leq i \leq n} X_i ,$$

où les X_i sont des v. a. réelles indépendantes de loi uniforme sur $[0, 1]$. Quelle est la loi de Y_n ? Quelle est la loi limite ?

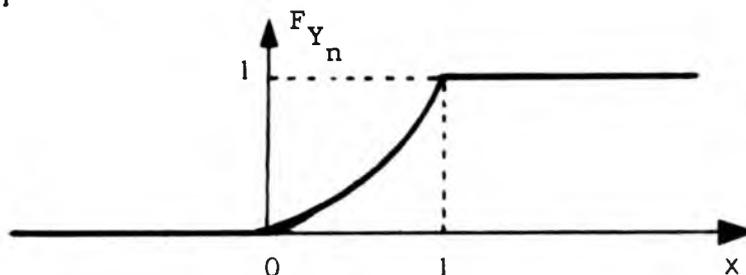
Y_n prend ses valeurs sur $[0, 1]$.

$$\text{Si } x \in]0, 1[, \quad F_{Y_n}(x) = P(Y_n < x) = P(\sup X_i < x),$$

$$= P\left(\bigcap_1^n \{X_i < x\}\right)$$

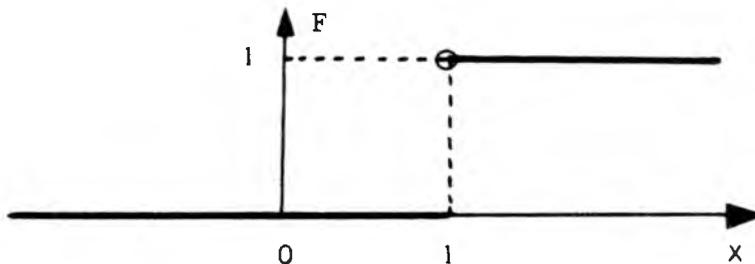
$$= \prod_1^n P(X_i < x) = x^n.$$

$$F_{Y_n}(x) = \begin{cases} 0 & \text{si } x \leq 0 ; \\ x^n & \text{si } 0 \leq x \leq 1 ; \\ 1 & \text{si } x \geq 1. \end{cases}$$



La fonction limite F vérifie donc :

$$F(x) = \begin{cases} 0 & \text{si } x < 1 ; \\ 1 & \text{si } x \geq 1. \end{cases}$$



La loi limite est la loi de Dirac en 1 :

$$P_{Y_n} \xrightarrow[n \rightarrow \infty]{} D_1.$$

On écrit :

$$Y_n \xrightarrow[n \rightarrow \infty]{L} 1.$$

On remarquera que la limite F de F_{Y_n} n'est pas exactement la fonction de répartition de la loi D_1 ; mais ces deux fonctions ne diffèrent qu'en leur point de discontinuité.

- Exemple 2.

Quelle est la limite de la loi binomiale $B(n, p)$ lorsque

p tend vers 0, n vers l'infini, le produit np ayant une limite finie λ non nulle ?

On cherche la limite de chacune des masses de probabilité.

$$P(X_n = k) = C_n^k p^k q^{n-k} \quad (k \leq n).$$

On peut prendre k quelconque, en ne commençant l'étude qu'à partir de $n = k$.

$$P(X = k) = \frac{n(n-1) \dots (n-k+1)}{k!} p^k (1-p)^{n-k},$$

$$P(X = k) \sim \frac{n^k}{k!} p^k (1-p)^{n-k} \rightarrow e^{-\lambda} \frac{\lambda^k}{k!},$$

car $(1-p)^n$ tend vers $e^{-\lambda}$ ($\log(1-p)^n = n \log(1-p) \sim -np \rightarrow -\lambda$).

La loi limite est la loi de Poisson $P(\lambda)$.

3 - Convergence en probabilité.

- Définition.

On dit que X_n tend vers X en probabilité et l'on note :

$$X_n \xrightarrow[n \rightarrow \infty]{P} X,$$

$$\text{si, } \forall \epsilon > 0, \quad P(|X_n - X| > \epsilon) \xrightarrow[n \rightarrow \infty]{} 0,$$

c'est-à-dire :

$$P\{\omega ; |X_n(\omega) - X(\omega)| > \epsilon\} \xrightarrow[n \rightarrow \infty]{} 0,$$

ou encore :

$\forall \epsilon > 0, \forall \alpha > 0, \exists N$ tel que $n > N$ implique :

$$P\{\omega ; |X_n(\omega) - X(\omega)| > \epsilon\} < \alpha.$$

THEOREME.

La convergence en probabilité entraîne la convergence en loi.
C'est-à-dire :

$$(X_n \xrightarrow{P} X) \Rightarrow (X_n \xrightarrow{L} X).$$

THEOREME.

Si la loi de X_n tend vers la loi de Dirac D_a , X_n tend, en probabilité, vers toute variable aléatoire de la loi D_a (définie sur le même espace probabilisé).

Montrons-le :

$$X_n \xrightarrow{L} a.$$

Considérons une variable aléatoire X de loi D_a :

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq a ; \\ 1 & \text{si } x > a. \end{cases}$$

On a :

$$F_{X_n}(x) \rightarrow \begin{cases} 0 & \text{si } x < a ; \\ 1 & \text{si } x > a. \end{cases}$$

$$\begin{aligned} P(|X_n - X| > \epsilon) &= P(|X_n - a| > \epsilon) = P(X_n < a - \epsilon) + P(X_n > a + \epsilon), \\ &= F_{X_n}(a - \epsilon) + [1 - F_{X_n}(a + \epsilon + 0)] \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

La propriété est établie.

- Loi faible des grands nombres.

Soit une suite de v. a. réelles X_n indépendantes de même loi,

de carré sommable ; on pose :

$$Y_n = \frac{X_1 + X_2 + \dots + X_n}{n} .$$

Montrons que :

$$Y_n \xrightarrow[n \rightarrow \infty]{P} E(X) .$$

$$E(Y_n) = \frac{\sum_{i=1}^n E(X_i)}{n} = \frac{n E(X)}{n} = E(X) .$$

$$\begin{aligned} \sigma^2(Y_n) &= \frac{1}{n^2} \sigma^2(X_1 + \dots + X_n) = \frac{1}{n^2} [\sigma^2(X_1) + \dots + \sigma^2(X_n)] \\ &= \frac{1}{n^2} n \sigma^2(X) = \frac{\sigma^2(X)}{n} . \end{aligned}$$

D'après l'inégalité de Bienaymé-Tchébicheff :

$$\forall \varepsilon > 0, \quad P(|Y_n - E(X)| \geq \varepsilon) \leq \frac{\sigma^2(Y_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \xrightarrow[n \rightarrow \infty]{} 0 ,$$

donc :
$$Y_n \xrightarrow{P} E(X) .$$

4 - Convergence presque sûre.

- Définition.

On dit que X_n tend vers X presque sûrement et l'on note :

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

si :
$$X_n(\omega) \xrightarrow[n \rightarrow \infty]{} X(\omega) ,$$

sauf sur un ensemble de Ω de probabilité nulle. C'est un mode de convergence plus puissant que les autres, car on montre que : la convergence presque sûre implique la convergence en probabilité.

- Loi forte des grands nombres.

On reprend les hypothèses de la loi faible des grands nombres, mais en supposant seulement que les X_i sont sommables. On a :

$$Y_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} E(X).$$

XI - VECTEURS ALEATOIRES

1 - Définition et loi d'un vecteur aléatoire.

Définition. Etant donné un espace probabilisé $(\Omega, \mathfrak{a}, P)$, on appelle vecteur aléatoire de dimension n défini sur cet espace une application mesurable de (Ω, \mathfrak{a}) dans $(\mathbb{R}^n, B_{\mathbb{R}^n})$.

Une telle application \vec{X} définit une probabilité-image $P_{\vec{X}}$ sur $(\mathbb{R}^n, B_{\mathbb{R}^n})$:

$$\forall B \in B_{\mathbb{R}^n}, \quad P_{\vec{X}}(B) = P[\vec{X}^{-1}(B)].$$

D'où la notion de fonction de répartition de \vec{X} , application $F_{\vec{X}}$ de \mathbb{R}^n dans $[0, 1]$ définie par :

$$\vec{x} = (x_1, \dots, x_n) \longmapsto P_{\vec{X}} \left[\prod_{i=1}^n] - \infty, x_i[\right] = P \left[\bigcap_{i=1}^n \{X_i < x_i\} \right].$$

On pourra écrire :

$$F_{\vec{X}}(\vec{x}) = P(\vec{X} < \vec{x}),$$

en établissant sur \mathbb{R}^n une relation d'ordre évidente (qui n'est pas d'ordre total).

La fonction de répartition d'un vecteur aléatoire possède des propriétés analogues à celles d'une variable aléatoire (pp. 21-22).

THEOREME. Les composantes d'un vecteur aléatoire défini sur $(\Omega, \mathfrak{a}, P)$ sont des v.a.r. définies sur $(\Omega, \mathfrak{a}, P)$.

En effet, soit $\vec{X} = (X_1, \dots, X_n)$ un vecteur aléatoire défini sur $(\Omega, \mathfrak{a}, P)$.

Si B_1 est un borélien de \mathbb{R} , on a :

$$X_1^{-1}(B_1) = \vec{X}^{-1}(B_1 \times \mathbb{R} \times \dots \times \mathbb{R}) \in \mathfrak{a},$$

car $B_1 \times \mathbb{R} \times \dots \times \mathbb{R}$ est un borélien de \mathbb{R}^n .

Plus généralement, tout "sous-vecteur" d'un vecteur aléatoire défini sur $(\Omega, \mathfrak{a}, P)$ est un vecteur aléatoire défini sur $(\Omega, \mathfrak{a}, P)$.

Nous admettrons la réciproque : une application de Ω dans \mathbb{R}^n dont les composantes sont des v.a.r. définies sur $(\Omega, \mathfrak{a}, P)$ est un vecteur aléatoire défini sur $(\Omega, \mathfrak{a}, P)$.

2 - Types de lois.

Lois discrètes. Il existe un ensemble fini ou dénombrable d'éléments \vec{x}^j de \mathbb{R}^n ($j \in J$) tel que :

$$P(\vec{X} = \vec{x}^j) = p_j, \quad \text{avec} \quad \sum_{j \in J} p_j = 1.$$

La fonction de répartition admet des discontinuités en ces points.

Lois absolument continues. Il existe une application f de \mathbb{R}^n dans \mathbb{R}^+ telle que :

$$\forall B \in \mathcal{B}_{\mathbb{R}^n}, \quad P(\vec{X} \in B) = \int_B f(\vec{x}) d\vec{x}$$

(il s'agit d'une intégrale de Lebesgue).

f s'appelle la densité de probabilité de \vec{X} .

Propriété. Si la densité existe, la fonction de répartition F est continue et, en tout point où f est continue, on a :

$$f(x_1, \dots, x_n) = \frac{\partial^n F}{\partial x_1 \dots \partial x_n}.$$

Il suffit de revenir à la définition ci-dessus, en prenant

$$B = \prod_{i=1}^n]-\infty, x_i[, \text{ pour établir ce résultat.}$$

On pourrait aussi définir des lois singulières.

3 - Lois marginales.

Définition. On appelle lois marginales d'un vecteur aléatoire les lois de ses composantes ou, plus généralement, de ses "sous-vecteurs".

Soit $\vec{X} = (X_1, \dots, X_n)$ un vecteur aléatoire.

Cherchons, par exemple, la loi de X_1 .

Si la loi de \vec{X} est discrète, on a :

$$P(X_1 = x_1) = \sum_{\{j; x_1^j = x_1\}} p_j.$$

Si la loi de \vec{X} est absolument continue de densité f , on a :

$$F_{X_1}(x_1) = P(X_1 < x_1) = \int_B f(\vec{t}) d\vec{t} ,$$

où $B =]-\infty, x_1[\times \mathbb{R} \times \dots \times \mathbb{R}$. D'où

$$F_{X_1}(x_1) = \int_{-\infty}^{x_1} dt_1 \int_{\mathbb{R}^{n-1}} f(\vec{t}) dt_2 \dots dt_n .$$

D'où, par dérivation :

$$f_{X_1}(x_1) = \int_{\mathbb{R}^{n-1}} f(\vec{x}) dx_2 \dots dx_n .$$

On généralise aisément ce résultat à toutes les lois marginales.

Indépendance des composantes. On a vu (chapitre VII) que, pour que les composantes du vecteur aléatoire $\vec{X} = (X_1, \dots, X_n)$ soient indépendantes, il faut et il suffit que l'on ait :

$$F_{\vec{X}}(\vec{x}) = \prod_{i=1}^n F_{X_i}(x_i) .$$

Si la loi de \vec{X} est discrète, on a

$$P(\vec{X} = \vec{x}) = \prod_{i=1}^n P(X_i = x_i) ;$$

si elle est absolument continue, on a :

$$f_{\vec{X}}(\vec{x}) = \prod_{i=1}^n f_{X_i}(x_i) ,$$

et ces conditions sont suffisantes, comme on l'établira aisément.

THEOREME. Pour que les composantes d'un vecteur aléatoire \vec{X} de loi absolument continue soient indépendantes, il faut et il suffit que sa densité f se factorise, c'est-à-dire qu'il existe des fonctions g_i telles que :

$$\forall \vec{x} \in \mathbb{R}^n , f(\vec{x}) = \prod_{i=1}^n g_i(x_i) .$$

Montrons-le en prenant, par exemple, $n = 3$.

Puisque $\int_{\mathbf{R}^3} f(\vec{x}) d\vec{x} = 1$, on a : $k_1 k_2 k_3 = 1$, où l'on a posé :

$$k_i = \int_{\mathbf{R}} g_i(x) dx \quad (i = 1, 2, 3).$$

On a :

$$f_{X_1}(x_1) = g_1(x_1) \int_{\mathbf{R}^2} g_2(x_2) \cdot g_3(x_3) dx_2 dx_3 = k_2 k_3 g_1(x_1).$$

De même,

$$f_{X_2}(x_2) = k_3 k_1 g_2(x_2) \text{ et } f_{X_3}(x_3) = k_1 k_2 g_3(x_3).$$

D'où :

$$f(\vec{x}) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdot f_{X_3}(x_3).$$

On peut établir un résultat analogue pour une loi discrète : il faut et il suffit que $P(\vec{X} = \vec{x})$ se factorise.

On peut, enfin, généraliser ces théorèmes en cherchant des conditions nécessaires et suffisantes pour que des "sous-vecteurs" d'un vecteur aléatoire soient indépendants.

4 - Changement de variable.

Soit \vec{X} un vecteur aléatoire de \mathbf{R}^n de loi absolument continue dont la densité $f_{\vec{X}}$ est nulle en dehors d'un borélien Δ de \mathbf{R}^n .

Soit φ une application injective mesurable et continûment dérivable de Δ dans \mathbf{R}^n . Posons $\Delta' = \varphi(\Delta)$ et supposons que Δ' est un borélien de \mathbf{R}^n . φ établit donc une bijection entre Δ et Δ' .

Quelle est la densité du vecteur aléatoire \vec{Y} défini par :

$$\vec{Y} = \varphi(\vec{X}) ?$$

On peut, bien sûr, prendre cette densité nulle en dehors de Δ' .

Soit B' un borélien de \mathbf{R}^n ; supposons-le contenu dans Δ' , quitte à le remplacer par $B' \cap \Delta'$.

On a :

$$P(\vec{Y} \in B') = P(\vec{X} \in \varphi^{-1}(B')) = \int_{\varphi^{-1}(B')} f_{\vec{X}}(\vec{x}) d\vec{x} ,$$

d'où, par changement de variable dans cette intégrale :

$$P(\vec{Y} \in B') = \int_{B'} f_{\vec{X}}(\vec{x}(\vec{y})) \left| \frac{D(\vec{x})}{D(\vec{y})} \right| d\vec{y} ,$$

où $\frac{D(\vec{x})}{D(\vec{y})}$ désigne le jacobien de \vec{x} par rapport à \vec{y} , c'est-à-dire l'inverse du jacobien de φ .

$$D'où : f_{\vec{Y}}(\vec{y}) = \begin{cases} f_{\vec{X}}(\vec{x}(\vec{y})) \left| \frac{D(\vec{x})}{D(\vec{y})} \right| & \text{si } \vec{y} \in \Delta' ; \\ 0 & \text{sinon.} \end{cases}$$

Ce résultat est d'un usage courant dans les exercices.

5 - Espérance mathématique. Matrice de covariance et coefficient de corrélation.

Espérance mathématique. Si \vec{X} est un vecteur aléatoire défini sur $(\Omega, \mathfrak{a}, P)$, son espérance est, par définition, le vecteur (s'il existe) :

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

c'est-à-dire le vecteur de composantes : $\int_{\Omega} X_i(\omega) dP(\omega) = E(X_i)$.

$$E(\vec{X}) = (E(X_1), \dots, E(X_n)) .$$

Les propriétés de linéarité de l'espérance d'une v.a.r. se conservent.

Matrice de covariance. La matrice de covariance d'un vecteur aléatoire \vec{X} de \mathbb{R}^n est la matrice $\Lambda_{\vec{X}}$ carrée d'ordre n dont le terme

général λ_{ij} est défini par :

$$\lambda_{ij} = E \left[(X_i - E(X_i)) \cdot (X_j - E(X_j)) \right] = E(X_i X_j) - E(X_i) \cdot E(X_j).$$

C'est une matrice symétrique ; elle existe si et seulement si les composantes de \vec{X} sont toutes de carré sommable, comme on peut le montrer.

Les termes diagonaux de $\Lambda_{\vec{X}}$ sont les variances des composantes de \vec{X} ; les autres sont appelés "covariances".

Si les composantes de \vec{X} sont indépendantes, les covariances sont nulles et la matrice $\Lambda_{\vec{X}}$ est diagonale ; la réciproque est fautive.

Coefficient de corrélation. Soient X et Y deux variables aléatoires définies sur le même espace probabilisé, de carré sommable et de variances non nulles (aucune des deux n'est de Dirac).

On appelle coefficient de corrélation de X et Y le nombre réel :

$$\rho_{X,Y} = \frac{E(XY) - E(X) \cdot E(Y)}{\sigma(X) \cdot \sigma(Y)}.$$

Il est nul si X et Y sont indépendantes, mais la réciproque est fautive.

$$\text{D'autre part, } \rho_{aX+b, cY+d} = \frac{|ac|}{ac} \rho_{X,Y}.$$

THEOREME. Le coefficient de corrélation de deux variables aléatoires X et Y est un élément du segment $[-1, +1]$; il est égal à 1 ou à (-1) si et seulement s'il existe une liaison affine entre X et Y.

Désignons par X' et Y' les variables aléatoires centrées déduites de X et Y. Le numérateur de $\rho_{X,Y}$ n'est autre que la covariance de X et Y, c'est-à-dire $E(X'Y')$.

On a donc :

$$\rho_{X,Y}^2 = \frac{[E(X'Y')]^2}{E(X'^2) \cdot E(Y'^2)}.$$

$$\forall \lambda \in \mathbb{R}, \quad E[(\lambda X' + Y')^2] \geq 0,$$

$$\text{soit : } \forall \lambda \in \mathbb{R}, \quad \lambda^2 E(X'^2) + 2\lambda E(X'Y') + E(Y'^2) \geq 0.$$

Le discriminant de ce trinôme en λ est donc négatif ou nul.

d'où : $|\rho_{X,Y}| \leq 1$.

Pour qu'il soit nul, c'est-à-dire pour que $|\rho_{X,Y}| = 1$, il faut et il suffit qu'il existe une valeur λ_0 de λ telle que :

$$E[(\lambda_0 X' + Y')^2] = 0, \text{ donc telle que :}$$

$$\lambda_0 X' + Y' = 0 \quad \text{P-p.p.}$$

d'où :

$$\lambda_0 X + Y = \lambda_0 E(X) + E(Y) \quad \text{P-p.p.}$$

Il existe une liaison affine entre X et Y.

ELEMENTS de BIBLIOGRAPHIE

J. BASS.

Eléments de calcul des probabilités théorique et appliqué.
Paris - Masson. 1967. (248 p.).

F. BRODEAU et G. ROMIER.

Mathématiques pour l'informatique. 4. Probabilités.
Paris - Armand Colin (Collection U). 1973. (199 p.).

L. GUERBER et P.L. HENNEQUIN.

Initiation aux probabilités.
Paris - A.P.M. - 1968. (228 p.).

D. HERAULT.

Eléments de théorie moderne des probabilités.
Paris - Dunod - 1967. (242 p.).

P. JAFFARD.

Initiation aux méthodes de la statistique et du calcul des
probabilités.
Paris - Masson - 1973. (327 p.).

A. KAUFMANN.

Cours moderne de calcul des probabilités.
Paris - A. Michel - 1965. (458 p.).

A. KRIEF et S.LEVY.

Calcul des probabilités. Exercices.
Paris - Hermann. 1972. (250 p.).

G. LETAC.

Problèmes de probabilités.
Paris - P.U.F. - 1970. (119 p.).

S. LIPSCHUTZ.

Probabilités. Cours et problèmes. 500 exercices résolus.
New-York - Mac Graw Hill
Paris - Ediscience - 1973. (153 p.).

M. METIVIER.

Notions fondamentales de la théorie des probabilités.
Paris - Dunod - 1968. (304 p.).

A. RENYI.

Calcul des probabilités.
Paris - Dunod - 1966. (620 p.).

J. ZIDANI

Probabilités et Statistique par les méthodes actives.
P.U.G. - 1972 (235 p.).

* *
*