ELEMENTS DE THEORIE DES PROBABILITES

par

J. PARIS

LA PROCURE

14, Boulevard E. -Mélot, NAMUR 161, rue des Tanneurs, BRUXELLES 1

COLLECTION D'INITIATION AUX MATHEMATIQUES MODERNES A L'USAGE DE L'ENSEIGNEMENT MOYEN :

GRAAS CAHIER 1 - Algèbre. Préface du Professeur R. BALLIEU

DEPRIT CAHIER 2 - La Géométrie affine et ses Structures.

PONCELET CAHIER 3 - Ensembles, Relations, Fonctions, Structures algébriques.

SMETS CAHIER 4 - Initiation à la Géométrie moderne; classes de 3e, 2e et 1re.

BALLIEU, BOEL STEENBECKELIERS CAHIER 5 - Algèbre linéaire et analyse numérique épuisé

FOUREZ CAHIER 6 - Approches modernes de la Géométrie analytique

KASSAB CAHIER 7 - Introduction au Calcul Vectorio-matriciel

BALLIEU, BOEL, DELMEZ, MATTHYS, STEENBECKELIERS CAHIER 8 - Eléments d'analyse infinitésimale

WILLIOT CAHIER 9 - G1. Géométrie du plan, Trigonométrie (structures, calculs).

WILLIOT CAHIER 10 - G2. Géométrie de l'espace, Trigonométrie (structures, calculs).

VIATOUR CAHIER 12 - Ensembles. Arithmétique des entiers.

DELMEZ WYBOU CAHIER 13 - Les Géométries analytiques planes (projectives, affines, métrique).

PARIS CAHIER 14 - Eléments de théorie des probabilités

INTRODUCTION

Longtemps confinée à d'élémentaires jeux de hasard, la théorie des probabilités n'a connu son plein développement qu'à partir du dix-neuvième siècle grâce, principalement, aux travaux de l'école russe qui comprenait Chebyshev, Lyapounov et Markov. En axiomatisant la théorie, Kolmogorov l'a rattachée à la théorie de l'intégration, ce qui a conduit certains à ne la considérer que comme une application de celle-ci. Mais l'application est tellement importante qu'elle est devenue une théorie mathématique. De plus, le caractère rigoureux qu'elle a ainsi acquis a attiré sur elle l'attention des meilleurs esprits. On a vu alors se développer des techniques propres aux probabilistes. Si on utilise toujours largement l'algèbre et l'analyse en théorie des probabilités, il est fréquent aujourd'hui de voir les méthodes probabilistes utilisées en analyse. Il faut aussi signaler le développement considérable des applications. Pratiquement, toutes les disciplines des sciences, des sciences humaines et des sciences appliquées utilisent le calcul des probabilités.

Il me semble, dès lors, vraiment indispensable à l'heure actuelle, d'initier convenablement les jeunes aux concepts et aux méthodes de la théorie des probabilités. Le présent texte est destiné à réinformer les enseignants. Ceci m'a conduit à inclure, à côté des éléments qui forment la base de tout manuel de niveau élémentaire, des points de théorie plus avancés. Il me paraît en effet difficile de saisir pleinement une partie de matière sans être au moins initié à certains de ses principaux développements.

Le texte ne comporte finalement que quelques applications. On peut en trouver de nombreuses autres dans les ouvrages cités en référence.

Il me reste à remercier le professeur R. F. BALLIEU pour les remarques qu'il m'a faites à propos des deux premiers chapitres.

CHAPITRE I. - AXIOMATIQUE DE LA THEORIE DES PROBABILITES

§ 1. EPREUVE.

Considérons une expérience effectuée sous un ensemble donné d'hypothèses. Il se peut qu'elle soit de nature telle que les seules hypothèses en fixent univoquement le résultat. Ainsi, si on abandonne une craie dans l'atmosphère terrestre, elle tombe toujours. Il existe cependant de nombreuses expériences où les seules hypothèses ne fixent pas univoquement le résultat de l'expérience. On dit alors qu'il s'agit d'une expérience aléatoire, c'est-à-dire d'une expérience dans laquelle le hasard intervient. Comme la théorie des probabilités tire son origine des jeux de hasard, on parle aussi d'épreuve. Pour qu'une épreuve soit définie, il faut préciser au préalable tout ce qui la caractérise. Toutes les expériences répétées sous des conditions qu'un observateur compétent et impartial juge équivalentes sont considérées comme identiques.

Un trait essentiel des expériences aléatoires est le fait qu'elles peuvent avoir plusieurs résultats possibles. Cela provient de ce que l'expérimentateur ne peut pas contrôler tous les facteurs qui régissent le déroulement d'une expérience aléatoire et en déterminent le résultat. Une fois fixées toutes les hypothèses, la partie la plus importante d'une expérience aléatoire, c'est le résultat. Dans la suite, on désignera par Ω, l'ensemble de tous les résultats possibles d'une expérience aléatoire. Ω est aussi appelé une catégorie d'épreuves. On désignera pas ω l'un quelconque des points de Ω c'est-à-dire l'un quelconque des résultats possibles.

§ 2. EVENEMENT.

A toute expérience aléatoire sont liés des événements. En théorie des probabilités, on ne s'intéresse aux événements que du seul point de vue de leur réalisation ou de leur non réalisation. Mais pour qu'il y ait un sens à parler d'un événement, il doit y avoir un sens à parler de l'événement contraire. Par définition, l'événement contraire d'un événement A n'est réalisé que si et seulement si A ne l'est pas. De même, pour deux événements A, B, il doit y avoir un sens à parler de l'événement A ou B qui est réalisé si au moins l'un des deux événements A, B l'est et de l'événement A et B qui n'est réalisé que si et seulement si A, B le sont tous les deux. On distingue encore l'événement certain qui est toujours réalisé et

l'événement impossible qui ne l'est jamais.

Tout résultat obtenu doit permettre de dire, sans ambiguité, si un événement quelconque fixé à l'avance et lié à l'expérience étudiée a été réalisé ou non. Dans la suite, il sera toujours supposé qu'on se limite à des événements liés à l'expérience étudiée. Il est clair que l'événement "il neige à Moscou" est sans rapport avec l'expérience qui consiste à lancer un dé à Bruxelles.

Considérons l'expérience aléatoire qui consiste à lancer une fois un dé à 6 faces. L'ensemble Ω des résultats possibles peut être symbolisé par $\Omega = \{1,2,3,4,5,6\}$. L'événement "obtenir le point 3" ne peut être réalisé que d'une seule façon à savoir obtenir le résultat 3 de Ω. Un tel événement est appelé événement élémentaire. Dans le même problème, l'événement "obtenir un point pair" ne peut être réalisé qu'en obtenant 2,4,6. Il semble donc naturel d'identifier un événement A avec le sous-ensemble A de Ω formé des points ω de Ω qui réalisent A. Un événement élémentaire est donc identifié avec le sous-ensemble ω (singleton), réduit au seul point ω. On dira donc que l'événement A est réalisé si le résultat ω obtenu appartient à A.

Les concepts et les notations de la théorie des ensembles trouvent une application naturelle en théorie des probabilités. Seulement comme la dernière s'est développée antérieurement à l'autre, elle a conservé une terminologie propre. Il sera commode, dans la suite, d'utiliser conjointement les deux.

L'événement contraire d'un événement A est identifié avec le complémentaire A^c de A; l'événement A ou B est identifié avec $A \cup B$; l'événement A et B est identifié avec $A \cap B$. L'événement certain est identifié avec l'ensemble Ω et l'événement impossible avec l'ensemble vide Φ.

Deux événements A, B sont dits incompatibles quand l'événement $A \cap B$ est impossible. Cette notion correspond à celle de deux sous-ensembles A, B disjoints ($A \cap B = \Phi$). La réunion de deux événements incompatibles s'appelle aussi leur somme. Il est parfois commode dans ce cas d'utiliser la notation $A + B$ au lieu de $A \cup B$.

Si un événement B est réalisé chaque fois que A l'est, nous dirons que A entraîne B ou que A implique B et nous écrirons $A \subseteq B$, traduisant par là, le fait que les résultats qui conduisent à la réalisation de A conduisent également à la réalisation de B. (Attention ! c'est le plus petit événement qui implique le plus grand et non le contraire).

On a les équivalences :

$$A \subseteq B \iff A = A \cap B$$

$$A \subseteq B \iff A^c \supseteq B^c$$

Nous dirons que deux événements A, B sont équivalents si l'on a à la fois $A \subseteq B$ et $B \subseteq A$. Nous écrirons alors $A = B$ et nous

ne distinguerons jamais deux tels événements.

On identifie à la différence $A - B$ l'événement qui est réalisé quand A l'est sans que B le soit. On a :

$$A - B = \{\, \omega : \omega \in A ,\ \omega \notin B \,\} = A \cap B^c$$

Au lieu de A-B, on utilise aussi $A \setminus B$ pour désigner le même sous-ensemble.

On identifie à la différence symétrique $A \triangle B$ l'événement qui est réalisé quand un seul des deux événements A, B est réalisé. On a donc :

$$A \triangle B = (A-B) + (B-A) .$$

Remarques : 1) La relation $\subseteq$ est une relation d'ordre sur l'ensemble des événements. On a en effet :

$$A \subseteq A$$

$$A \subseteq B ,\ B \subseteq A \iff A = B$$

$$A \subseteq B ,\ B \subseteq C \implies A \subseteq C$$

Pour cette relation d'ordre, l'union de deux événements est leur borne supérieure et l'intersection, leur borne inférieure.

2) On étend facilement les notions d'union, d'intersection et de somme au cas d'une famille non vide d'événements. L'événement $\bigcup_{i \in I} A_i$ est réalisé dès que l'un au moins des A_i est réalisé. L'événement $\bigcap_{i \in I} A_i$ n'est réalisé que si et seulement si tous les A_i le sont.

Pour une famille vide, on pose :

$$I \text{ vide} \qquad \bigcup_{i \in I} A_i = \Phi \qquad \bigcap_{i \in I} A_i = \Omega$$

Proposition 1. Etant donné n ($n \geqslant 1$) événements $A_1, A_2, \ldots, A_n$, il est possible de construire n événements $B_1, B_2, \ldots, B_n$, deux à deux incompatibles et tels que :

$$\bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} B_i$$

Démonstration :

Il suffit de poser :

$$B_1 = A_1$$

$$B_i = A_i - \bigcup_{j=1}^{i-1} A_j \qquad i > 1.$$

pour obtenir la suite souhaitée.

Remarque : La proposition traduit simplement un fait important pour certains problèmes. Elle exprime que pour que l'un au moins des événements $A_1, A_2, \ldots, A_n$ soit réalisé, il faut et il suffit qu'il existe un premier événement de cette suite qui soit réalisé. Dire que B_i est réalisé équivaut à dire que A_i est le premier événement de la suite $A_1, A_2, \ldots, A_n$ qui est réalisé.

Exercice. Interpréter en termes d'événements les identités suivantes :

$$A \cup B = (A \triangle B) + (A \cap B)$$

$$(A \triangle B) + (A \triangle B^c) = \Omega$$

$$A - B = A - (A \cap B) = (A \cup B) - B$$

$$A - (B \cup C) = A \cap B^c \cap C^c$$

$$(\bigcup_i A_i)^c = \bigcap_i A_i^c$$

Soit (A_i) une suite monotone croissante d'événements, c'est-à-dire, $\forall i \quad A_i \subseteq A_{i+1}$. On définit la limite de la suite A_i par :

$$\lim_i A_i = \bigcup_i A_i$$

On écrit alors : $(A_i) \uparrow \bigcup_i A_i$

De même, pour une suite $(A'_i)\downarrow$ monotone décroissante d'événements, c'est-à-dire, $\forall i \quad A'_i \supseteq A'_{i+1}$, on définit la limite de la suite (A'_i) par :

$$\lim A'_i = \bigcap_i A'_i$$

On écrit alors : $(A_i) \downarrow \bigcap_i A_i$

Soit (B_j) une suite quelconque d'événements. Alors, la suite (A'_i) où

$$A'_i = \bigcup_{j \geqslant i} B_j$$

est monotone décroissante et $\lim_i A'_i = \bigcap_i A'_i$. On définit :

$$\lim_j \sup B_j = \bigcap_i \bigcup_{j \geqslant i} B_j$$

L'événement $\lim_j \sup B_j$ est réalisé si et seulement si une infinité de B_j sont réalisés.

Avec une suite (B_j) quelconque, la suite (A_i) où

$$A_i = \bigcap_{j \geqslant i} B_j$$

est monotone croissante et $\lim_i A_i = \bigcup_i A_i$. On définit :

$$\lim_j \inf B_j = \bigcup_i \bigcap_{j \geqslant i} B_j$$

L'événement $\lim_j \inf B_j$ est réalisé si et seulement si tous les B_j à l'exception d'un nombre fini d'entre eux sont réalisés.

L'ensemble $\lim_j \sup B_j$ est formé des points $\omega \in \Omega$ qui se trouvent dans une infinité de B_j. Il en résulte que :

$$\lim_j \inf B_j \subseteq \lim_j \sup B_j$$

et on dit que la $\lim_j B_j$ existe quand il y a égalité.

§ 3. TRIBU D'EVENEMENTS.

On appelle anneau de parties sur Ω une famille $\mathcal{F}$ de sous-ensembles de Ω telle que si $A, B \in \mathcal{F}$ alors $A \cup B \in \mathcal{F}$, $A \cap B \in \mathcal{F}$, $A \triangle B \in \mathcal{F}$, $A-B \in \mathcal{F}$. On dit encore qu'un anneau de parties est stable pour les opérations $\cup, \cap, \triangle, -$.

La définition donnée s'explique par le fait qu'un anneau de parties est un anneau, au sens algébrique du terme, pour les deux opérations $\triangle$ et $\cap$. Il est abélien (commutatif) et tous ses éléments sont idempotents (c'est-à-dire $A \cap A = A$ pour tout $A \in \mathcal{F}$).

Un anneau de parties sur Ω ne contient pas nécessairement Ω. Il n'est donc pas nécessairement stable pour l'opération de complémentation. C'est pour cette raison qu'on définit une algèbre de Boole sur Ω comme étant un anneau de parties sur Ω qui contient Ω.

Un anneau de parties sur Ω et une algèbre de Boole sur Ω sont évidemment stables pour les unions finies et les intersections finies.

Quand Ω ne contient qu'un nombre fini de points; tout sous-ensemble de Ω est un événement intéressant dont on voudra définir la probabilité. On prendra alors pour algèbre de Boole des événements, l'ensemble $\mathcal{P}(\Omega)$ formé de tous les sous-ensembles de Ω. Quand Ω contient n éléments, on sait que $\mathcal{P}(\Omega)$ contient les 2^n sous-ensembles possibles de Ω.

Quand Ω contient une infinité de points, on ne va pas essayer de définir la probabilité pour des sous-ensembles compliqués

et non intéressants de Ω (ce ne serait d'ailleurs pas toujours possible). On va se limiter à des familles de sous-ensembles suffisamment simples.

Définition 1 : On appelle tribu sur Ω, une famille $\mathcal{A}$ de sous-ensembles de Ω stable pour le complémentaire, la différence, la différence symétrique, l'union dénombrable et l'intersection dénombrable. Certains auteurs utilisent le terme algèbre dénombrablement addtive ou encore σ-algèbre au lieu du terme tribu.

Une tribu contient toujours Φ, Ω et elle est stable pour les unions finies et les intersections finies. Une tribu est donc une algèbre de Boole mais c'est une structure plus riche.

Remarque : Les conditions données dans la définition d'une tribu sont surabondantes car certaines des opérations s'expriment en fonction des autres. Au titre d'exercice, on peut vérifier que pour qu'une famille de sous-ensembles de Ω soit une tribu sur Ω il suffit qu'elle soit stable pour le complémentaire et l'union dénombrable.

Exemple. Sur un même ensemble Ω, on peut évidemment définir plusieurs tribus. La plus petite est formée de Φ et de Ω. La plus grande est l'ensemble $\mathcal{P}(\Omega)$ de tous les sous-ensembles de Ω. Si $A \subseteq \Omega$, $A \neq \Phi$ alors la famille composée de Φ, A, A^c, Ω est une tribu.

Si $\mathcal{A}$ et $\mathcal{A}'$ sont deux tribus sur Ω, on écrira $\mathcal{A} \subseteq \mathcal{A}'$ pour indiquer que tout sous-ensemble A de Ω qui est un élément de $\mathcal{A}$ est aussi un élément de $\mathcal{A}'$.

Prenons $\Omega = \mathbb{R}$. La classe des intervalles de $\mathbb{R}$ est stable pour l'intersection mais elle n'est pas stable pour l'union car l'union de deux intervalles disjoints n'est pas un intervalle. Prenons alors une classe $\mathcal{C}$ formée des parties de $\mathbb{R}$ qui sont des unions d'un nombre fini d'intervalles. Il existe des tribus qui contiennent la classe $\mathcal{C}$. $\mathcal{P}(\mathbb{R})$ en est une. Prenons l'intersection de toutes les tribus qui contiennent $\mathcal{C}$. Il est facile de voir que cette intersection est une tribu et que c'est la plus petite tribu qui contient $\mathcal{C}$. On l'appelle la tribu des boréliens et on la notera $\mathcal{B}$.

Un borélien est donc soit un intervalle, soit une union d'un nombre fini d'intervalles, soit toute partie de $\mathbb{R}$ obtenue en appliquant un nombre dénombrable de fois les opérations d'ensembles à partir d'intervalles.

Remarquons pour la suite qu'on obtient la même tribu des boréliens en prenant la sous-classe $\mathcal{C}'$ de $\mathcal{C}$ formée des intervalles du type $]-\infty, a]$, $a \in \mathbb{R}$ car tout intervalle s'obtient en appliquant un nombre dénombrable de fois les opérations d'ensembles à des intervalles de ce type. Ainsi, pour $a, b \in \mathbb{R}$, on a

$$]a,b] =]-\infty, b] -]-\infty, a]$$

$$\left[a,b\right] = \bigcap_n \left]a - \frac{1}{n}, b\right] \qquad n \in \mathbb{N}^*$$

La tribu des boréliens jouera un rôle important dans la suite.

Exercice : Vérifier que la famille des sous-ensembles A de $\mathbb{R}$ tels que si $x \in A$, alors $x \pm n \in A$ pour tout $n \in \mathbb{N}$, est une tribu sur $\mathbb{R}$.

Définition 2. On appelle espace mesurable le couple $(\Omega, \mathcal{A})$ formé d'un ensemble Ω non vide et d'une tribu $\mathcal{A}$ sur Ω.

§ 4. AXIOMATIQUE DE KOLMOGOROV (1933).

La notion de probabilité est une notion intuitive souvent mal définie. Pour bâtir une théorie convenable, nous utiliserons l'axiomatique de Kolmogorov. Avant de la présenter, faisons quelques remarques.

La notion de probabilité est une notion du même type que celle de masse. Si on répartit une masse sur un ensemble, la masse portée par un sous-ensemble est positive et inférieure à la masse totale. La masse portée par deux sous-ensembles disjoints égale la somme des masses.

La notion de probabilité doit également traduire le fait suivant : si un événement A est plus souvent réalisé qu'un événement B, la probabilité de A doit être supérieure à celle de B.

Enfin, si j'effectue n expériences au cours desquelles A a été réalisé n_A fois, alors la fréquence de la réalisation de A égale, par définition, $\frac{n_A}{n}$. Ce rapport est évidemment compris entre 0 et 1. Dans la loi des grands nombres, on découvrira la relation qui existe entre la fréquence de la réalisation de A et la probabilité de A.

Nous pouvons maintenant présenter l'axiomatique de Kolmogorov.

Sur un espace mesurable $(\Omega, \mathcal{A})$ formé d'un ensemble Ω, non vide, de résultats et d'une tribu $\mathcal{A}$ d'événements définie sur Ω, on donne :

Axiome 1. A tout événement $A \in \mathcal{A}$, on associe un nombre réel, noté $P[A]$, appelé la probabilité de A et satisfaisant :

$$\forall A \in \mathcal{A} \qquad 0 \leqslant P[A] \leqslant 1.$$

Axiome 2. La probabilité de l'événement certain égale l'unité.

$$P[\Omega] = 1.$$

Axiome 3. Pour toute suite (A_i) d'événements deux à deux incompatibles, la probabilité de la somme des A_i égale la somme des probabilités des A_i.

$$\forall i \in \mathbb{N} \qquad A_i \in \mathcal{A}$$

$$\forall i \neq j \ , \ i \text{ et } j \in \mathbb{N} \qquad A_i \cap A_j = \Phi$$

$$P\left(\bigcup_i A_i\right) = \sum_{i \in \mathbb{N}} P\left[A_i\right]$$

Remarques : 1) Une probabilité est donc une fonction d'ensembles positive, dénombrablement additive, définie sur une tribu. Il est clair que les axiomes impliquent l'additivité finie, c'est-à-dire :

si $A_1, A_2, \ldots, A_n \in \mathcal{A}$ et si $\forall i \neq j$, $A_i \cap A_j = \Phi$, on a :

$$P\left[A_1 \cup A_2 \cup \ldots \cup A_n\right] = P\left[A_1\right] + P\left[A_2\right] + \ldots + P\left[A_n\right]$$

2) Supposons Ω dénombrable (ou fini) et soit $(\omega_i)_{i \in}$ la suite des résultats qui constituent Ω . Prenons : $\mathcal{A} = \mathcal{P}(\Omega)$. Supposons connaître, pour tout $i \in \mathbb{N}$, la probabilité $p_i = P\left[\{\omega_i\}\right]$ de l'événement élémentaire $\{\omega_i\}$. On vérifie immédiatement que la probabilité P ainsi donnée satisfait aux axiomes de Kolmogorov si et seulement si

$$\forall i \in \mathbb{N} \qquad p_i \geqslant 0 \qquad \text{et} \qquad \sum_{i \in \mathbb{N}} p_i = 1.$$

Pour tout $A \in \mathcal{A}$, on a alors :

$$P[A] = \sum_{\omega_i \in A} p_i$$

Définition 3. On appelle espace probabilisé $(\Omega, \mathcal{A}, P)$ un espace mesurable $(\Omega, \mathcal{A})$ sur lequel on a définit une probabilité P satisfaisant les trois axiomes.

Conséquences des axiomes.

1) Si $A \in \mathcal{A}$, alors $P[A] + P[A^c] = 1$.

En effet, on a $A \cap A = \Phi$ et $A \cup A^c = \Omega$

Donc : $P[A] + P[A^c] = 1$.

2) $P[\Phi] = 0$ car $\Phi = \Omega^c$

3) Pour deux événements quelconques $A, B \in \mathcal{A}$ on a l'importante relation de Boole :

$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$

En effet, on a :

$$A \cup B = A \cup (B-A) \quad \text{et} \quad A \cap (B-A) = \Phi$$

Donc :

$$P[A \cup B] = P[A] + P[B-A] \qquad (1)$$

Mais on a aussi :

$$B = (A \cap B) \cup (B-A) \text{ et } (A \cap B) \cap (B-A) = \Phi$$

Donc :

$$P[B] = P[A \cap B] + P[B-A] \qquad (2)$$

En éliminant $P[B-A]$ entre les relations (1) et (2), on obtient le résultat. Notons que les relations (1) et (2) sont aussi importantes pour elles-mêmes.

En particulier, si A et B sont incompatibles, c.à.d. $A \cap B = \Phi$. On retrouve : $P[A \cup B] = P[A] + P[B]$.

4) Si $A, B \in \mathcal{A}$ et $A \subseteq B$, on a $P[A] \leqslant P[B]$

En effet, on a dans ce cas

$$B = A \cup (B-A) \quad \text{et} \quad A \cup (B-A) = \Phi$$

Donc

$$P[B] = P[A] + P[B-A]$$

$$P[B] \geqslant P[A] \quad \text{puisque } P[B-A] \geqslant 0$$

5) Si $A_1, A_2, \ldots, A_n \in \mathcal{A}$, on a : $P\left[\bigcup_1^n A_i\right] \leqslant \sum_{i=1}^n P\left[A_i\right]$

En effet, prenons les ensembles B_i de la proposition 1. On a :

$$\forall i \quad i = 1, 2, \ldots n \qquad B_i \subseteq A_i \quad \text{et} \quad B_i \in \mathcal{A}$$

$$\forall i \neq j \qquad B_i \cap B_j = \Phi \quad \text{et} \quad \bigcup_1^n A_i = \bigcup_1^n B_i$$

Donc :

$$P\left[\bigcup_1^n A_i\right] = \left[P \bigcup_1^n B_i\right] = \sum_{i=1}^n P\left[B_i\right] \leqslant \sum_{i=1}^n P\left[A_i\right]$$

6) Pour deux événements quelconques A, B, on a :

$$\max(P(A), P(B)) \leqslant P[A \cup B] \leqslant P[A] + P[B]$$

$$P[A] + P[B] - 1 \leqslant P[A \cap B] \leqslant \min(P[A], P[B])$$

7) Si $(A_i) \downarrow A$ et si $\forall i \ A_i \in \mathcal{A}$ alors $P(\lim_i A_i) = \lim_i P\left[A_i\right]$

Pour démontrer cette propriété, on peut supposer $A = \Phi$ sinon il suffit de remplacer A_i par $A_i - A$ car alors $(A_i - A) \downarrow \Phi$.

Par l'axiome 3,

$$P\left[A_1\right] = \sum_{i=1}^{+\infty} P\left[A_i - A_{i+1}\right] \qquad (3)$$

$$P\left[A_n\right] = \sum_{i=n}^{+\infty} P\left[A_i - A_{i+1}\right] \qquad (4)$$

La série (4) est le reste de la série (3) qui est convergente. Donc:

$$\lim_n \left[P\ A_n\right] = 0$$

8) Si $(A_i) \uparrow A$ et si $\forall i \quad A_i \in \mathcal{A}$ alors $P(\lim_i A_i) = \lim_i P\left[A_i\right]$

En effet $(A_i^c) \downarrow A^c$. Par la propriété 7, on a

$$P\left[A^c\right] = \lim_i P\left[A_i^c\right]$$
$$1 - P[A] = \lim_i (1 - P\left[A_i\right])$$
$$P[A] = \lim_i P\left[A_i\right]$$

Les relations établies ne permettent pas d'évaluer la probabilité d'un événement. Elles permettent seulement de calculer la probabilité d'un événement quand on connaît la probabilité des événements qui le constituent. La proposition suivante permet d'évaluer la probabilité d'un événement dans de nombreuses applications. Donnons d'abord une définition.

Définition 4. On appelle système contradictoire ou exhaustif d'événements un nombre fini ou une infinité dénombrable d'événements, deux à deux incompatibles, et tels que dans chaque expérience, un et un seul des ces événements soit réalisé.

Pour un système contradictoire, on a donc :
I fini ou dénombrable ; $\forall i \in I, \quad A_i \in \mathcal{A}$; $\forall i, j \in I, i \neq j$, $A_i \cap A_j = \Phi$ et $\bigcup_{i \in I} A_i = \Omega$

Définition 5. Les événments $A_1, A_2, \ldots, A_n \in \mathcal{A}$ sont équiprobables si

$$P\left[A_1\right] = P\left[A_2\right] = \ldots = P\left[A_n\right]$$

Remarquons que $P\left[A_1\right] = P\left[A_2\right]$ n'implique pas nécessairement $A_1 = A_2$. En particulier, $P\left[A_1\right] = 0$ n'implique pas nécessairement $A_1 = \Phi$.

Considérons le cas où on a affaire à un système contradictoire formé d'un nombre fini d'événements $A_1, A_2, \ldots, A_n$. On dit que ces événements forment tous les cas possibles car dans toute expérience, un et un seul des A_i est réalisé. Soit B un événement de qui est réunion de m des événements du système contradictoire. Quitte à écrire les événements $A_1, A_2, \ldots, A_n$ dans un autre ordre,

on peut toujours supposer que ce sont les événements $A_1, A_2, \ldots A_n$ dont la réunion est B. Ces m événements $A_1, A_2, \ldots, A_m$ sont appelés cas favorables à la réalisation de B.

Proposition 2. Si on a un système contradictoire formé d'un nombre fini d'événments équiprobables, alors la probabilité d'un événement, réunion d'événements de ce système contradictoire, égale le rapport du nombre de cas favorables à la réalisation de cet événement au nombre total de cas possibles.

Démonstration :

Par la définition d'un système contradictoire, on a =

$$P\left[\bigcup_{i=1}^{n} A_i\right] = \sum_{i=1}^{n} P\left[A_i\right] = 1$$

Puisque les événements $A_1, A_2, \ldots, A_n$ sont équiprobables, on a :

$$P\left[A_1\right] = P\left[A_2\right] = \ldots = P\left[A_n\right] =$$

Des deux relations, on déduit :

$$P\left[A_i\right] = \frac{1}{n} \qquad i = 1, 2, \ldots, n.$$

Si $B = \bigcup_{i=1}^{m} A_i$, on a :

$$P\left[B\right] = P\left[A_1\right] + P\left[A_2\right] + \ldots + P\left[A_m\right]$$
$$= \frac{m}{n}$$

Remarque : De nombreuses erreurs commises dans l'évaluation de la probabilité d'un événement proviennent d'une application erronée de cette proposition.

Exemples : 1) On considère un dé à 6 faces. Alors $\Omega = \{1,2,3,4,5,6\}$. Prenons $\mathcal{A} = P(\Omega)$. Comme système contradictoire, nous prenons : $A_i = \{i\}$ $i = 1, 2, \ldots, 6$. Supposons le dé parfaitement régulier. De cette considération physique, on déduit : $P\left[A_i\right] = \frac{1}{6}$ $i = 1, 2, \ldots, 6$. Soit B_1 l'événement "obtenir un point impair", alors $B_1 = A_1 \cup A_3 \cup A_5$ et $P\left[B_1\right] = \frac{3}{6} = \frac{1}{2}$. De même, si B_2 est l'événement "obtenir un point pair", on a : $B_2 = A_2 \cup A_3 \cup A_6$ et $p\left[B_2\right] = \frac{3}{6} = \frac{1}{2}$. On remarque que les deux événements B_1 et B_2 forment un nouveau système contradictoire d'événements équiprobables.

2) Reprenons l'exemple 1 en remplaçant l'hypothèse d'un dé parfaitement régulier, par :

$$P\left[A_1\right] = P\left[A_2\right] = P\left[A_3\right] = \frac{1}{9} \quad P\left[A_4\right] = P\left[A_5\right] = P\left[A_6\right] = \frac{2}{9}$$

Alors, $P[B_2] = \frac{5}{9} \neq \frac{1}{2}$.

3) On lance simultanément trois pièces de monnaie parfaitement régulières. Le résultat ne peut jamais fournir que "0 pile", "1 pile", "2 piles" ou "3 piles". Une personne qui ignore le calcul des probabilités en conclut que la probabilité de chacun des quatre événements est $\frac{1}{4}$. C'est faux ! Pourtant ces quatre événements forment bien un système contradictoire.

Avec trois pièces, on a :

$$\Omega = \{ P_1 P_2 P_3,\ P_1P_2F_3,\ P_1F_2P_3,\ P_1F_2F_3,\ F_1P_2P_3,\ F_1P_2F_3,\ F_1F_2P_3,\ F_1F_2F_3 \}$$

où P_i = "obtenir pile avec la $i^{ième}$ pièce"

F_i = "obtenir face avec la $i^{ième}$ pièce" $\qquad i = 1,2,3.$

Prenons comme système contradictoire, le système (A_j) $j = 1,2,\ldots,8$ formé des événements élémentaires écrits dans l'ordre de présentation de Ω . De la considération physique de pièces parfaitement régulières, on déduit que les événements $A_1, A_2, A_3 \ldots A_8$ sont équiprobables. Mais comme trois d'entre eux sont favorables à la réalisation de "1 pile", on a :

$$P["1 \text{ pile}"] = P[A_4 \cup A_6 \cup A_7] = \frac{3}{8}$$

De même ,

$$P["0 \text{ pile}"] = P[A_8] = \frac{1}{8}$$

$$P["2 \text{ piles}"] = P[A_2 \cup A_3 \cup A_5] = \frac{3}{8}$$

$$P["3 \text{ piles}"] = P[A_1] = \frac{1}{8} .$$

On remarque que les événements A_1 ; $A_2 \cup A_3 \cup A_5$; $A_4 \cup A_6 \cup A_7$; A_8 forment un nouveau système contradictoire d'événements, mais les événements de ce nouveau système ne sont plus équiprobables.

CHAPITRE II. VARIABLES ALEATOIRES.

§ 1. DEFINITION ELEMENTAIRE D'UNE VARIABLE ALEATOIRE.

Quand on effectue une expérience aléatoire, on n'est pas toujours intéressé par tous les détails du résultat mais bien par la valeur numérique d'une quantité attachée à l'expérience. Donnons quelques exemples simples de telles quantités.

Exemple 1. Deux joueurs A et B conviennent du jeu suivant : une pièce de monnaie est lancée une fois; si elle donne pile, A donne un franc à B; si elle donne face, B donne un franc à A. Le gain de A est une grandeur numérique dont la valeur est déterminée par le résultat de l'expérience aléatoire. Ici, $\Omega = \{$ pile, face $\}$ et le gain de A est caractérisé entièrement par :

pile $\rightarrow$ -1
face $\rightarrow$ 1

Exemple 2. Deux joueurs A et B conviennent du jeu suivant : une pièce de monnaie est lancée deux fois de suite; pour chaque pile obtenu, A donne un franc à B; pour chaque face obtenu, B donne un franc à A. Le gain est une grandeur numérique dont la valeur est déterminée par le résultat de l'expérience aléatoire. Ici, $\Omega = \{$ (pile, pile), (pile, face), (face, pile), (face, face) $\}$ et le gain de A est caractérisé entièrement par :

(pile, pile) $\rightarrow$ -2
(pile, face) $\rightarrow$ 0
(face, pile) $\rightarrow$ 0
(face, face) $\rightarrow$ +2

Exemple 3. On lance simultanément deux dés à 6 faces. La somme des points obtenus est une grandeur numérique dont la valeur est déterminée par le résultat de l'expérience aléatoire. Ici, $\Omega = \{ (1,1) ; (1,2); \ldots ; (6,6) \}$ et la somme des points est caractérisée entièrement à partir du tableau :

Point marqué par le premier dé \ Point marqué par le second dé	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Dans chacun des exemples, la tribu $\mathcal{A}$ est P (Ω).

Si dans les trois exemples, on suppose les pièces parfaitement régulières et les dés parfaitement réguliers, alors on matérialise cette hypothèse physique en disant que les probabilités des événements élémentaires sont égales. Néanmoins, dans les deux derniers, les probabilités des événements qui donnent une valeur numérique déterminée ne sont plus égales. Ainsi, dans le deuxième exemple, la probabilité d'avoir un gain nul vaut 1/2, celle d'avoir un gain + 2 vaut 1/4. Dans le troisième exemple, la probabilité d'avoir le total 5 est $\frac{4}{36}$, celle d'avoir le total 8 est $\frac{5}{36}$. C'est un élément dont il faudra tenir compte.

Dans chacun des exemples, la valeur numérique était entière. Il est facile de concevoir des exemples où la quantité attachée à une expérience aléatoire est simplement réelle. Il suffit de prendre la durée de vie d'un tube électrique.

Définition 1 (provisoire) . On appelle variable aléatoire (v.a. en abrégé) toute caractéristique numérique X attachée à une expérience aléatoire dont la valeur est déterminée par le résultat de l'expérience et pour laquelle il y a un sens à parler de la probabilité que cette v.a. prenne une valeur inférieure à un nombre réel x arbitraire donné.

On représente les v.a. par des majuscules.
Une v.a. est donc une application de Ω dans $\mathbb{R}$ telle que, pour tout réel x, l'ensemble $\{ \omega : X(\omega) \leqslant x \} = \{ \omega : -\infty < X(\omega) \leqslant x \}$ soit un événement, c'est-à-dire un ensemble de $\mathcal{A}$. Si X est une v.a., il y a un sens à parler de

$$P[\omega : X(\omega) \leqslant x]$$

et cette probabilité est évidemment une fonction de x.

§ 2. DEFINITION MATHEMATIQUE D'UNE VARIABLE ALEATOIRE.

Soit X une application d'un ensemble Ω dans un ensemble Ω'. A cette application X, on peut associer l'application inverse X^{-1} de l'ensemble P (Ω') des parties de Ω' dans l'ensemble P(Ω) des parties de Ω . Cette application inverse possède les propriétés suivantes :

1) $X^{-1}(\Phi) = \Phi \qquad X^{-1}(\Omega') = \Omega$

2) $A' \subseteq \Omega' \qquad X^{-1}(A'^{c}) = (X^{-1}(A'))^{c}$

3) Pour toute famille $\{ A'_i \quad i \in I \}$ de sous-ensembles de Ω', on a :

$$X^{-1}\Big(\bigcup_{i \in I} A'_i \Big) = \bigcup_{i \in I} (X^{-1}(A'_i))$$

$$X^{-1}\Big(\bigcap_{i \in I} A'_i \Big) = \bigcap_{i \in I} (X^{-1}(A'_i))$$

Pour une classe $\mathcal{C}'$ de sous-ensembles de Ω', nous définissons son image inverse par :

$$X^{-1}(\mathcal{C}') = \{ X^{-1}(C') \qquad C' \in \mathcal{C}' \}$$

Des propriétés indiquées de l'application inverse et de cette définition, on déduit immédiatement :

Proposition 1 : L'image inverse $X^{-1}(\mathcal{A}')$ d'une tribu $\mathcal{A}'$ sur Ω' est une tribu sur Ω .

Prenons (Ω , $\mathcal{A}$) et une application X de Ω dans Ω'.
Définissons :

$$\mathcal{D}' = \{ A' : A' \subseteq \Omega' , X^{-1}(A') \in \mathcal{A} \}$$

Il est immédiat que $\mathcal{D}'$ est une tribu sur Ω' qu'on appelle la tribu induite sur Ω', à partir de $\mathcal{A}$, par l'application X.

Proposition 2.

Soit X une application de Ω dans Ω' ,
Soit $\mathcal{C}'$ une classe de sous-ensembles de Ω' ,
Soit $\mathcal{A}'$ la tribu engendrée par $\mathcal{C}'$,
Soit $X^{-1}(\mathcal{C}')$ l'image inverse de $\mathcal{C}'$,
Soit $\mathcal{A}$ la tribu engendrée par $X^{-1}(\mathcal{C}')$,
alors :

$$X^{-1}(\mathcal{A}') = \mathcal{A}.$$

Démonstration :

1) Puisque $\mathcal{C}' \subseteq \mathcal{A}'$, on a : $X^{-1}(\mathcal{C}') \leq X^{-1}(\mathcal{A}')$

Comme $X^{-1}(\mathcal{A}')$ est une tribu qui contient $X^{-1}(\mathcal{C}')$, $X^{-1}(\mathcal{A}')$ contient la tribu $\mathcal{A}$ engendrée par $X^{-1}(\mathcal{C}')$ et donc

$$X^{-1}(\mathcal{A}') \supseteq \mathcal{A}.$$

2) Soit $\mathcal{D}'$ la tribu induite sur Ω', à partir de $\mathcal{A}$, par X. On a $\mathcal{D}' \supseteq \mathcal{C}'$ car pour tout $C' \in \mathcal{C}'$ on a : $X^{-1}(C') \in \mathcal{A}$. Dès lors, $\mathcal{D}'$ contient la tribu $\mathcal{A}'$ engendrée par $\mathcal{C}'$. On en déduit

$$X^{-1}(\mathcal{A}') \subseteq X^{-1}(\mathcal{D}')$$

mais, par définition de $\mathcal{D}'$, on a :

$$X^{-1}(\mathcal{D}') \subseteq \mathcal{A}$$

Donc on a aussi

$$X^{-1}(\mathcal{A}') \subseteq \mathcal{A} \quad .$$

ce qui achève la démonstration.

Définition 2. Une application X de l'espace mesurable $(\Omega, \mathcal{A})$ dans l'espace mesurable $(\Omega', \mathcal{A}')$ est dite mesurable si $X^{-1}(\mathcal{A}') \subseteq \mathcal{A}$.

Remarques : 1) La définition d'une application mesurable revient à dire : une application X de l'espace mesurable $(\Omega, \mathcal{A})$ dans l'espace mesurable $(\Omega', \mathcal{A}')$ est mesurable si la tribu $\mathcal{D}'$ induite sur Ω', à partir de $\mathcal{A}$, par X contient $\mathcal{A}'$.

2) Il est évident qu'une application X reste mesurable lorsqu'on remplace $\mathcal{A}$ par une tribu plus grande mais qu'elle ne le reste pas nécessairement quand on remplace $\mathcal{A}$ par une tribu plus petite. $X^{-1}(\mathcal{A}')$ est la plus petite tribu par laquelle on peut remplacer $\mathcal{A}$ pour que X reste mesurable.

De la proposition 2, on déduit immédiatement :

Proposition 3. Pour que l'application X de l'espace mesurable $(\Omega, \mathcal{A})$ dans l'espace mesurable $(\Omega', \mathcal{A}')$ soit mesurable, il suffit qu'il existe une classe $\mathcal{C}'$ de parties de Ω' engendrant $\mathcal{A}'$ et telle que $X^{-1}(\mathcal{C}') \subseteq \mathcal{A}$.

Prenons maintenant pour espace mesurable $(\Omega', \mathcal{A}')$ l'espace $(\mathbb{R}, \mathcal{B})$ formé de l'ensemble des réels et de la tribu des boréliens.

Définition 3. On appelle v.a. toute application mesurable X de $(\Omega, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B})$.

La définition d'une v.a. ne fait pas appel à la probabilité qu'on peut mettre sur $(\Omega, \mathcal{A})$.

Puisque la classe $\mathcal{C}'$ des intervalles du type $]-\infty, x]$ $x \in \mathbb{R}$ engendre $\mathcal{B}$, il suffit pour avoir une v.a. de vérifier que

$X^{-1}(\mathcal{C}') \subseteq \mathcal{A}$. On retrouve les éléments introduits dans la définition élémentaire.

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et X une v.a. définie sur cet espace. On définit une probabilité P_X sur $(\mathbb{R}, \mathcal{B})$, qu'on appelle la probabilité induite par X à partir de P, par la relation

$$\forall B \in \mathcal{B} \qquad P_X(B) = P(X^{-1}(B))$$

Vérifions que P_X satisfait les axiomes de Kolmogorov.

1) $\quad 0 \leqslant P_X(B) \leqslant 1 \qquad \forall B \in \mathcal{B}$

2) $\quad P_X(\mathbb{R}) = P(X^{-1}(\mathbb{R})) = P(\Omega) = 1$

3) $\forall i \in I, B_i \in \mathcal{B}$ et $\forall i \neq j$, $B_i \cap B_j = \Phi$ et I dénombrable.

$$P_X(\bigcup_i B_i) = P\left[X^{-1}(\bigcup_i B_i)\right] = P\left(\bigcup_i \left[X^{-1}(B_i)\right]\right) = \sum_i P\left[X^{-1}(B_i)\right]$$
$$= \sum_i P_X(B_i)$$

car $X^{-1}(B_i \cap B_j) = X^{-1}(B_i) \cap X^{-1}(B_j) = \Phi$ si $B_i \cap B_j = \Phi$

Quand on utilise une v.a. X c'est généralement dans l'espace $(\mathbb{R}, \mathcal{B}, P_X)$ que l'on travaille. Comme la classe des intervalles $]-\infty, x]$ engendre $\mathcal{B}$, il suffit de connaître la probabilité P_X de ces intervalles. C'est ce qui conduit à définir la fonction de distribution d'une v.a. On dit aussi fonction de répartition.

§ 3. FONCTION DE DISTRIBUTION OU FONCTION DE REPARTITION D'UNE v.a.

Soit X une v.a. définie relativement à $(\Omega, \mathcal{A}, P\%)$ et $(\mathbb{R}, \mathcal{B})$.

Définition 4. On appelle fonction de distribution (notée f.d. en abrégé) d'une v.a. X la fonction F_X définie par :

$$x \in \mathbb{R} \qquad F_X : x \to P[\omega : -\infty < X(\omega) \leqslant x] = P_X(]-\infty, x])$$

Il est plus commode d'utiliser le symbolisme abrégé

$$F_X(x) = P_X(X \leqslant x)$$

mais il ne faudra jamais perdre de vue la signification exacte de l'écriture.

Remarque. Dans un exposé théorique avancé, on peut pratiquement se passer de la f.d. Néanmoins, dans les applications où la structure d'ordre de $\mathbb{R}$ joue un rôle important, elle reste un outil précieux.

Convention d'écriture. On négligera le symbole X associé à F et P quand aucune confusion ne sera possible.

Propriétés des f.d.

Pour toute f.d. d'une v.a. on a :

1) $\forall x \in \mathbb{R} \qquad 0 \leqslant F(x) \leqslant 1$

2) Une f.d. est croissante

Pour $x \leqslant y \quad x, y \in \mathbb{R}$ l'événement $[\omega : X(\omega) \leqslant x]$ implique l'événement $[\omega : X(\omega) \leqslant y]$. Dès lors, on a :

$$F(x) = P[X \leqslant x] \leqslant P[X \leqslant y] = F(y)$$

3) $\lim_{x \to +\infty} F(x) = 1 \qquad \lim_{x \to -\infty} F(x) = 0$

En vertu de la deuxième propriété, F est une fonction monotone croissante. Il suffit donc de vérifier la propriété 3 pour une suite (x_n) monotone de valeurs de x qui tend vers $+\infty$ ou $-\infty$ suivant le cas. La suite des événements

$$[\omega : X(\omega) \leqslant x_n] \uparrow \Omega \quad \text{quand } (x_n) \to +\infty$$

et donc on a :

$$\lim_{x_n \to +\infty} F(x_n) = P(\Omega) = 1.$$

De même, la suite des événements

$$[\omega : X(\omega) \leqslant x_n] \downarrow \Phi \quad \text{quand } (x_n) \to -\infty$$

et donc on a :

$$\lim_{x_n \to -\infty} F(x_n) = P(\Phi) = 0$$

4) Une f.d. est continue à droite, c.à.d. $\lim_{\substack{y \to x \\ >}} F(y) = F(x)$

En vertu de la propriété 2, il suffit encore de vérifier cette propriété pour une suite $(y_n) \downarrow x$. Or, dans ce cas, la suite des événements

$$[\omega : X(\omega) \leqslant y_n] \downarrow [\omega : X(\omega) \leqslant x]$$

ce qui implique :

$$\lim_{\substack{y_n \to x \\ >}} F(y_n) = F(x)$$

Remarques : 1) On peut montrer que si on a une fonction $F : \mathbb{R} \to [0, 1]$ qui satisfait aux quatre propriétés indiquées, alors on peut définir une probabilité P sur $(\mathbb{R}, B)$ en posant $P(]-\infty, x]) = F(x)$.

2) Certains auteurs définissent la f.d. par $F(x) = P[X < x]$. Une telle fonction est alors continue à gauche.

Conséquences. Pour tout $a, b \in \mathbb{R}$, $a \leqslant b$, on a :

$$P[a < X \leqslant b] = P[X \leqslant b] - P(X \leqslant a] = F(b) - F(a)$$
$$P[a \leqslant X \leqslant b] = P[X = a] + F(b) - F(a)$$
$$P[a < X < b] = F(b) - F(a) - P[X = b]$$
$$P[a \leqslant X < b] = F(b) - F(a) - P[X=b] + P[X=a].$$

L'événement $[\omega : X(\omega) = b]$ peut être exprimé comme étant l'intersection d'une suite décroissante d'événements. On a, n étant un entier $\geqslant 1$,

$$[\omega : X(\omega) = b] = \bigcap_n \left[b - \frac{1}{n} < X(\omega) \leqslant b + \frac{1}{n}\right]$$

$$P[X = b] = \lim_{n \to +\infty} P\left[b - \frac{1}{n} < X \leqslant b + \frac{1}{n}\right]$$
$$= \lim_{n \to +\infty} \left[F\left(b + \frac{1}{n}\right) - F\left(b - \frac{1}{n}\right)\right]$$
$$= F(b + 0) - F(b-0)$$

si on désigne par F(b+0) la limite à droite et par F(b-0) la limite à gauche.
On en déduit :

1) Si F est continue au point b, alors F(b+0) = F(b-0) et

$$P[X = b] = 0$$

Ainsi, si une v.a. a une f.d. continue en un point b, la probabilité que la v.a. prenne cette valeur b est nulle.

2) Si F est discontinue au point b, alors $F(b+0) \neq F(b-0)$ et

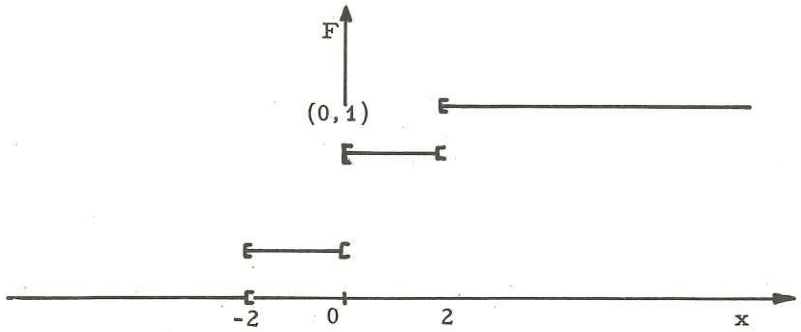
$$P[X = b] = F(b+0) - F(b-0) \neq 0$$

Ainsi, une v.a. peut prendre une valeur b en laquelle sa f.d. est discontinue avec une probabilité strictement positive.

Exemple 4. La v.a. gain de A définie dans l'exemple 2 a une f.d. définie par

$$F(x) = \begin{cases} 0 & x < -2 \\ 1/4 & -2 \leqslant x < 0 \\ 1/4 + 1/2 = 3/4 & 0 \leqslant x < 2 \\ 1/4 + 1/2 + 2/4 = 1 & 2 \leqslant x \end{cases}$$

On voit que la représentation graphique de la f.d. d'une v.a. qui ne prend qu'un nombre fini de valeurs est celle d'une fonction en escalier.



On remarque que la fonction F est continue sauf aux points -2, 0, 2 qui sont les seules valeurs que peut prendre la v.a. dans le problème. Les probabilités afférant à ces valeurs égalent la grandeur du saut de la f.d. en ces valeurs. Elles peuvent s'évaluer en faisant la différence entre la limite à droite et la limite à gauche en ces valeurs.

Le cas où la f.d. est une fonction en escalier correspond à un type bien déterminé de v.a. qu'on appelle v.a. discrète.

§ 4. VARIABLE ALEATOIRE DISCRETE.

On appelle ainsi toute v.a. telle que l'image de Ω par cette v.a. soit une partie finie ou dénombrable de $\mathbb{R}$. Pour caractériser une telle v.a. X il suffit de donner la suite (x_j) des points en lesquels la f.d. de X est discontinue (qui est aussi la suite (x_j) des valeurs que peut prendre la v.a. avec une probabilité strictement positive), ainsi que la suite (p_j) des grandeurs des discontinuités de la f.d. aux points (x_j) correspondants.

Une v.a. discrète X sera symbolisée par

$$X \begin{cases} x_o, x_1, \dots, x_j, \dots \\ p_o, p_1, \dots, p_j, \dots \end{cases}$$

avec $p_j = P\left[X = x_j\right]$

et les conditions : $0 \leqslant p_j \leqslant 1$ $\qquad \sum_{j=0}^{\infty} p_j = 1$

Dans cette catégorie se trouvent les v.a. qui ne prennent qu'une seule valeur a avec probabilité 1. Les constantes sont de telles v.a.

Exemple 5. Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. A tout événement $A \in \mathcal{A}$, on peut associer une v.a., notée I_A, appelée l'indicatrice de A et définie par

$$I_A(\omega) = 1 \quad \text{si } \omega \in A$$
$$I_A(\omega) = 0 \quad \text{si } \omega \notin A.$$

La v.a. indicatrice de A prend donc la valeur 1 quand A est réalisé et la valeur 0 quand A n'est pas réalisé. Elle est entièrement caractérisée par :

$$I_A \begin{cases} 0 & 1 \\ q = P[A^c] & p = P[A] \end{cases} \qquad 0 \leqslant p \leqslant 1 \ ; \ p+q=1$$

L'indicatrice jouit des propriétés suivantes :

$$A \subseteq B \rightleftharpoons I_A \leqslant I_B$$
$$A = B \rightleftharpoons I_A = I_B$$
$$A \cap B = \Phi \rightleftharpoons I_{A \cap B} = 0$$

$$I_\Phi = 0 \ ; \quad I_\Omega = 1 \ ; \quad I_A + I_{A^c} = 1$$

$$1 \leqslant n \leqslant m \begin{cases} I_{\cap A_n} = \prod_n I_{A_n} \\ I_{\cup A_n} = I_{A_1} + (1 - I_{A_1}) I_{A_2} + (1 - I_{A_1})(1 - I_{A_2}) I_{A_3} + \dots \end{cases}$$

$$m \in N$$

En particulier, si $\forall i \neq j \quad A_i \cap A_j = \Phi$, alors

$$I_{\sum A_n} = \sum_n I_{A_n}$$

Exemple 6 : . v.a. de Poisson de paramètre λ caractérisée par

$$X \begin{cases} 0, 1, 2, \dots, j, \dots \\ p_o = e^{-\lambda}, p_1 = \lambda e^{-\lambda}, \quad p_j = \frac{\lambda^j}{j!} e^{-\lambda} \end{cases} \qquad \lambda > 0$$

On vérifie immédiatement que

$$\forall j \in N \qquad p_j \geqslant 0$$
$$\sum_{j=o}^{+\infty} p_j = e^{-\lambda} \sum_{j=o}^{+\infty} \frac{\lambda^j}{j!} = e^{-\lambda} . e^{\lambda} = 1.$$

§ 5. VARIABLE ALEATOIRE CONTINUE.

On appelle ainsi toute v.a. pour laquelle la f.d. est continue et pour laquelle on suppose de plus qu'il existe une fonction $f : \mathbb{R} \to \mathbb{R}$, appelée fonction de fréquence (f.fr. en abrégé) ou densité de probabilité, satisfaisant aux conditions :

1) $\forall x \in \mathbb{R} \qquad f(x) \geqslant 0$

2) $\int_{-\infty}^{+\infty} f(x)\,dx = 1$

3) $\forall x \in \mathbb{R} \qquad F(x) = P[X \leqslant x] = \int_{-\infty}^{x} f(u)\,du$

La fonction f n'est pas nécessairement continue au point x. Cependant, si f est continue au point x, on a :

$$\frac{dF(x)}{dx} = f(x)$$

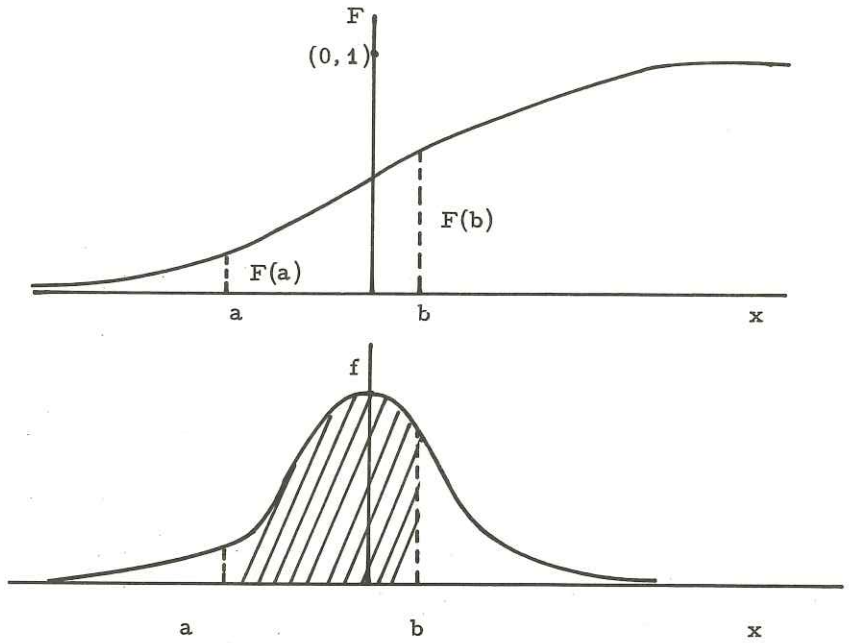
On a :

$$a \in \mathbb{R} \quad , \quad b \in \mathbb{R} \qquad a \leqslant b$$

$$P[a < X \leqslant b] = F(b) - F(a)$$

$$= \int_a^b f(x)\,dx$$

Le résultat est illustré par les deux représentations graphiques suivantes :



On remarque que pour toute v.a. X continue, on a

$$\forall x \in \mathbb{R} \qquad P[X = x] = 0$$

Remarque : L'appellation v.a. continue est une abréviation pour v.a. absolument continue car la f.d. correspondante est non seulement continue mais encore absolument continue. Rappelons qu'une application $F : \mathbb{R} \longrightarrow \mathbb{R}$ est dite une fonction absolument continue,

si $\forall \varepsilon > 0$, il existe un $\delta > 0$ tel que, pour tout système d'intervalles réels disjoints $[a_k, b_k]$ $(k = 1, 2, \ldots, n)$, $a_k < b_k$, la relation

$$\sum_{k=1}^{n} (b_k - a_k) < \delta$$

entraîne

$$\sum_{k=1}^{n} \left| F(b_k) - F(a_k) \right| < \varepsilon.$$

On peut démontrer que toute fonction absolument continue est presque partout différentiable et égale à l'intégrale indéfinie de sa dérivée.

Dans la plupart des applications les valeurs possibles d'une v.a. continue sont tous les points d'un ou de plusieurs intervalles réels.

Exemple 7. La durée de vie de certains tubes électriques est bien représentée par une v.a. continue ayant pour f.fr.

$$f(x) = \begin{cases} \mu\, e^{-\mu x} & \text{pour } x \geqslant 0 \qquad \mu > 0 \\ 0 & \text{pour } x < 0 \end{cases}$$

On en déduit :

$$F(x) = \begin{cases} \int_0^x e^{-\mu t}\,dt = 1 - e^{-\mu x} & \text{pour } x \geqslant 0 \\ 0 & \text{pour } x < 0 \end{cases}$$

§ 6. VARIABLES ALEATOIRES DU TYPE MIXTE.

On rencontre dans les applications des v.a. $X : \Omega \longrightarrow A \subseteq \mathbb{R}$ où A peut être décomposé en une partie dénombrable D de probabilité $P_X(D) \neq 0$ et où sur A-D la v.a. X se comporte comme une v.a. continue. La f.d. F d'une telle v.a. est alors une combinaison linéaire du type

$$F = \lambda F_1 + (1-\lambda) F_2 \qquad < 0\ \lambda\ < 1$$

F_1 étant la f.d. d'une v.a. discrète et F_2 la f.d. d'une v.a. continue.

La f.d. définie par

$$F(x) = \begin{cases} 1 - p\, e^{-(x-a)} & x < a \qquad 0 < p < 1 \\ 0 & x \geqslant a \end{cases}$$

présente une seule discontinuité au point a et on a :

$$P[X = a] = 1 - p = q$$

Pour tout intervalle dont les extrémités sont strictement supérieures à a, on trouve la probabilité de cet intervalle en intégrant la fonction :

$$f(x) = \begin{cases} p\, e^{-(x-a)} & x > a \\ 0 & x < a \end{cases}$$

On voit que si on pose

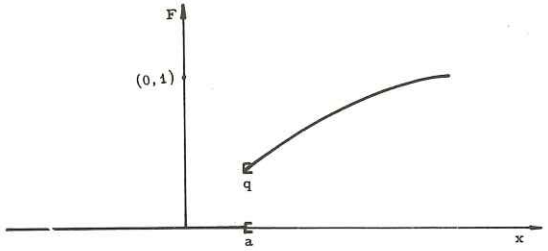
$$F_1(x) = \begin{cases} 0 & x < a \\ 1 & x \geqslant a \end{cases}$$

$$F_2(x) = \begin{cases} 0 & x < a \\ 1 - e^{-(x-a)} & x \geqslant a \end{cases}$$

on trouve $F(x) = q\, F_1(x) + p\, F_2(x)$

On rencontre de telles v.a. dans certains types d'assurances où le montant du sinistre est une v.a. qui prend la valeur 0 avec une probabilité strictement positive et qui se comporte comme une v.a. ayant une f.fr. exponentielle décroissante en dehors du point 0. Le cas où a est strictement positif correspond au cas des assurances avec franchise.

La représentation graphique de la f.d. F est donnée par:



Remarque : Deux v.a. définies sur deux (Ω , $\mathcal{A}$, P) peuvent avoir la même f.d.

1) Je considère une pièce de monnaie parfaitement régulière. Ω = { pile, face } . Je prends $\mathcal{A} = \mathcal{P}(\Omega)$. La v.a. X est définie par X (pile) = + 1 X (face) = -1.
Alors la v.a. X est caractérisée par

$$X \begin{cases} -1 & +1 \\ 1/2 & 1/2 \end{cases}$$

2) Je considère un dé à 6 faces parfaitement régulier. Ω = { 1, 2, 3, 4, 5, 6 } . Je prends $\mathcal{A} = \mathcal{P}(\Omega)$. La v.a. Y est définie par :

Y(1) = Y(3) = Y(5) = -1 Y(2) = Y(4) = Y(6) = +1.

Alors la v.a.Y est caractérisée par

$$Y \begin{cases} -1 & +1 \\ 1/2 & 1/2 \end{cases}$$

Les deux v.a. X et Y ont évidemment même f.d.
La v.a. Z définie par :

$$Z \begin{cases} Z(1) = Z(2) = Z(3) = -1 \\ Z(4) = Z(5) = Z(6) = 1 \end{cases}$$

a encore même f.d. que la v.a. Y.

§ 7. FONCTION D'UNE V.A.

Soit X une v.a. sur (Ω , $\mathcal{A}$, P). Soit g une application mesurable de ($\mathcal{R}$, $\mathcal{B}$) dans ($\mathcal{R}$, $\mathcal{B}$). Alors g o X est une v.a. Y qui sera notée plus facilement Y = g(X). On a, en effet, le schéma

$$(\Omega, \mathcal{A}) \xrightarrow{X} (\mathcal{R}, \mathcal{B}) \xrightarrow{g} (\mathcal{R}, \mathcal{B}) \qquad (Y = g \circ X)$$

$$Y^{-1}(\mathcal{B}) = X^{-1}\, g^{-1}(\mathcal{B}) \subseteq X^{-1}(\mathcal{B}) \subseteq \mathcal{A}$$

Cette propriété est exprimée dans :

Proposition 4: Si X est une application mesurable de (Ω , $\mathcal{A}$) dans (Ω', $\mathcal{A}'$); si X' est une application mesurable de (Ω', $\mathcal{A}'$) dans (Ω'', $\mathcal{A}''$) alors X' o X est une application mesurable de (Ω , $\mathcal{A}$) dans (Ω'', $\mathcal{A}''$).

Pour connaître la f.d. de Y il faut chercher

$$F_Y(y) = P_Y(Y \leqslant y) = P(\omega : Y(\omega) \leqslant y) = P(\omega : g(X(\omega)) \leqslant y)$$

pour $y \in \mathcal{R}$.

Dans les applications, g est le plus souvent continue et monotone, ce qui facilite les choses.

Exemple 8. Soit X une v.a. avec

$$F_X(x) = \begin{cases} 1 - e^{-3x} & x \geqslant 0 \\ 0 & x < 0 \end{cases}$$

Prenons : $Y = \begin{cases} \sqrt{X} & \text{si } X \geqslant 0 \\ 0 & \text{si } X < 0 \end{cases}$

On voit que g n'est ici définie que sur $\mathcal{R}_+$ mais $\mathcal{R}_+$ est tel que $P_X(\mathcal{R}_+) = 1$.

$$F_Y(y) = P_Y[Y \leqslant y] = P_X(\sqrt{X} \leqslant y) = P_X(X \leqslant y^2) \quad y \in \mathcal{R}$$

$$F_Y(y) = \begin{cases} 1 - e^{-3y^2} & y \geqslant 0 \\ 0 & y < 0 \end{cases}$$

Dans le présent cas, on parlera plus simplement de la v.a. $Y = \sqrt{X}$
On remarquera que

$$f_Y(y) = \frac{d\,F_Y(y)}{dy} = \begin{cases} 6\,y\,e^{-3y^2} & y \geqslant 0 \\ 0 & y < 0 \end{cases}$$

Exemple 9. Soit X une v.a. Soit F_X sa f.d. Soient a et b deux constantes avec $a > 0$. Prenons : $Y = a\,X + b$
Nous avons

$$F_Y(y) = P_X\left[a\,X + b \leqslant y \right] = P_X\left[X \leqslant \frac{y-b}{a} \right] = F_X\left(\frac{y-b}{a}\right) \qquad y \in \mathbb{R}$$

Si X a une f.fr. f_X, alors Y a aussi une f.fr. donnée par

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_x\left(\frac{y-b}{a}\right) = \frac{1}{a} f_x\left(\frac{y-b}{a}\right)$$

Si a était strictement négatif, on aurait :

$$F_Y(y) = P_X\,(a\,X + b \leqslant y) = P_X\,(X \geqslant \frac{y-b}{a})$$

$$= P_X\,(X = \frac{y-b}{a}) + 1 - F_X\,(\frac{y-b}{a})$$

$$f_Y\,(y) = -\frac{1}{a}\,f_X\,(\frac{y-b}{a}) = \frac{1}{a}\,f_X\,(\frac{y-b}{a})$$

Ce dernier exemple suggère la formulation générale qui permet d'obtenir la f.fr. de Y à partir de celle de X, quand elle existe, sans passer par la détermination de la f.d. de Y.

Soit X une v.a. ayant f_X pour f.fr. Soit $g : \mathbb{R} \to \mathbb{R}$ continue et strictement croissante (il est facile de vérifier qu'une telle application est mesurable). Alors à toute valeur de x correspond une seule valeur $y = g(x)$ et, réciproquement, à toute valeur y correspond une seule valeur $x = h(y)$ telle que $g(x) = y$. La fonction h est la fonction réciproque de g.

Supposons de plus que la dérivée de g existe et est continue. Alors, la dérivée de h jouit des mêmes propriétés. Soit $Y = g(X)$. On a

$$F_Y(y) = P_Y(Y \leqslant y) = P_X\,(g(X) \leqslant y) = P_X\,[\,X \leqslant h(y)\,]$$

$$= F_X\,(h\,(y)\,)$$

et

$$f_Y(y) = \frac{d\,F_Y}{d\,y} = \frac{d}{dy}\,F_X\,(h(y)) = f_X\,(h(y)).\ h'\,(y)$$

Comme

$$g'(x) = \frac{dy}{dx} = \frac{1}{\frac{dx}{dy}} = \frac{1}{h'\,(y)}$$

on a finalement :

$$f_Y\,(y) = f_X\,(h\,(y)).\ \frac{1}{\frac{dy}{dx}}$$

Si la fonction g jouissait des mêmes propriétés mais était strictement décroissante, on aurait :

$$f_Y(y) = f_X\,(h(y))\ \frac{1}{\left|\frac{dy}{dx}\right|} \qquad (1)$$

Remarque. Cette formule est en fait celle du changement de variable dans une intégrale. Pour la retenir facilement, on peut utiliser un procédé approché commode dans certaines applications.

Soit X une v.a. ayant f_X pour f.fr. On peut écrire la relation

$$P\left[x < X \leqslant x + dx \right] \simeq f_X(x)\,dx$$

Si $Y = g(X)$ a pour f.fr. f_Y, on a aussi :

$$P\left[y < Y \leqslant y + dy \right] \simeq f_Y(y)\,dy$$

Mais en prenant $y = g(x)$, on a :

$$P\left[x < X \leqslant x + dx \right] = P\left[y < Y \leqslant y+dy \right]$$

d'où la relation

$$f_Y(y)\,dy = f_X(x)\,dx$$

$$f_Y(y) = f_X\,(x)\,\frac{1}{\frac{dy}{dx}}$$

Il reste à exprimer le tout en y dans le second membre et à tenir compte du fait que les deux f.fr. sont positives.

Exercice. Retrouver les fonctions de fréquence des exemples 8 et 9 en utilisant la formule (1).

Quand la fonction g n'est pas monotone, on peut parfois décomposer son ensemble de définition en sous-ensembles sur lesquels g est monotone. Ainsi l'application $g : \mathbb{R} \to \mathbb{R}$ définie par $g(x) = x^2$ est strictement croissante pour $x \geqslant 0$ et strictement décroissante pour $x < 0$.

Exemple 10. Prenons une v.a. X ayant pour f.fr.

$$f_X(x) = \frac{1}{\pi\,(1 + x^2)} \qquad -\infty < x < +\infty$$

et prenons $Y = X^2$

Pour $y \geqslant 0$, on a :

$$F_Y(y) = P_Y(Y \leqslant y) = P_X[X^2 \leqslant y] = P_X[-\sqrt{y} \leqslant X \leqslant \sqrt{y}] = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

et par dérivation, pour $y > 0$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} - f_X(-\sqrt{y}) \frac{-1}{2\sqrt{y}} = \frac{1}{\pi\sqrt{y}(1+y)}$$

Pour $y > 0$, on a évidemment $F_Y(y) = 0$ car l'événement $[Y < 0] = [X^2 < 0]$ est vide et donc aussi $f_Y(y) = 0$.

Dans le cas où X et Y sont deux v.a. discrètes et où $Y = g(X)$, il n'est pas nécessaire de chercher la f.d. de Y. Il suffit de chercher les valeurs que peut prendre Y et les probabilités correspondantes.

Exemple 11. Soit X la v.a. caractérisée par

$$X \quad \begin{cases} -2 & 0 & 2 \\ 1/4 & 1/2 & 1/4 \end{cases}$$

alors la v.a. $Y = X^2 + 3$ est caractérisée par :

$$Y \quad \begin{cases} 3 & 7 \\ 1/2 & 1/2 \end{cases}$$

§ 8. FONCTIONS PARTICULIERES DE V.A.

Ce paragraphe a pour but de montrer directement que certaines fonctions de v.a. sont des v.a.

1) Si X est une v.a. et si a est une constante réelle, alors a X est une v.a.

Si $a = 0$ $\quad \{\omega : a X(\omega) \leqslant x\} = \begin{cases} \Phi & \text{si } x < 0 \\ \Omega & \text{si } x \geqslant 0 \end{cases}$

Si $a > 0$ $\quad \{\omega : a X \leqslant x\} = \{\omega : X \leqslant \frac{x}{a}\} \in \mathcal{a}$

Si $a < 0$ $\quad \{\omega : a X \leqslant x\} = \{\omega : X \geqslant \frac{x}{a}\} \in \mathcal{a}$

2) Si X est une v.a., alors X^2 est une v.a.

Si $x < 0$ $\quad \{X^2 \leqslant x\} = \Phi$

Si $x \geqslant 0$ $\quad \{X^2 \leqslant x\} = \{-\sqrt{x} \leqslant X \leqslant \sqrt{x}\} = \{X \leqslant \sqrt{x}\} \cap \{X \geqslant -\sqrt{x}\} \in \mathcal{a}$

3) Si X est une v.a. positive, alors $\sqrt{X}$ est une v.a.

Si $x < 0$ $\quad \{\sqrt{X} \leqslant x\} = \Phi$

Si $x \geqslant 0$ $\quad \{\sqrt{X} \leqslant x\} = \{X \leqslant x^2\} \in \mathcal{a}$

4) Si X et Y sont deux v.a., alors $Z = X + Y$ est une v.a. (Rappelons que Z est la v.a. qui, au point ω, prend la valeur $Z(\omega) = X(\omega) + Y(\omega)$).

Il suffit de montrer que si r parcourt l'ensemble des rationnels, on a :

$$\{\omega : Z(\omega) < z\} = \bigcup_r \{\omega : X(\omega) < r\} \cap \{\omega : Y(\omega) < z - r\}$$

Il est clair que l'ensemble du second membre appartient à $\mathcal{a}$ puisque l'ensemble des rationnels est dénombrable. Il est contenu dans l'ensemble du premier membre. Réciproquement, l'ensemble du premier membre est contenu dans celui du second car si ω est tel que $X(\omega) + Y(\omega) < z$, il suffit de prendre un rationnel r tel que $X(\omega) < r < z - Y(\omega)$ car alors $\omega \in \{\omega : X(\omega) < r\} \cap \{\omega : Y(\omega) < z - r\}$.

5) Si Y est une v.a. telle que $\{\omega : Y(\omega) = 0\} = \Phi$ alors $\frac{1}{Y}$ est une v.a.

On peut écrire :

$$\{\tfrac{1}{Y} \leqslant x\} = (\{\tfrac{1}{Y} \leqslant x\} \cap \{Y > 0\}) \cup (\{\tfrac{1}{Y} \leqslant x\} \cap \{Y < 0\}) .$$

ce qui donne immédiatement le résultat.

6) Si X et Y sont des v.a. alors max (X, Y) est une v.a. et min (X, Y) = - max (-X, -Y) est une v.a.

La v.a. $Z = \max(X, Y)$ est celle qui au point ω prend la valeur $Z(\omega) = \max(X(\omega), Y(\omega))$ et

$$\{\max(X, Y) \leqslant z\} = \{X \leqslant z\} \cap \{Y \leqslant z\}$$

7) Si $A \in \mathcal{A}$, alors I_A est une v.a. car

$$\{ I_A \leqslant x \} = \begin{cases} \Phi & \text{si } x < 0 \\ A^c & \text{si } 0 \leqslant x < 1 \\ \Omega & \text{si } 1 \leqslant x \end{cases}$$

Remarquons que si $A \notin \mathcal{A}$ alors I_A n'est pas une v.a.

8) Si X est une v.a. alors les fonctions X^+ et X^- définies par :

$$X^+ = X\, I_{\{X \geqslant 0\}} = \sup (X, 0)$$

$$X^- = X\, I_{\{X \leqslant 0\}} = \sup (-X, 0)$$

sont des v.a. On a alors : $X = X^+ - X^-$ et $|X| = X^+ + X^-$ est aussi une v.a.

9) Si X et Y sont des v.a. alors XY est une v.a.
Les propriétés 1, 2, 4 et l'identité algébrique

$$XY = \frac{1}{4}\left[(X + Y)^2 - (X-Y)^2 \right]$$

donnent le résultat :

Exercice : Vérifier de plusieurs façons que :

a) la fonction constante est une v.a.
b) si X et Y sont des v.a. et si $\{ \omega : Y(\omega) = 0 \} = \Phi$ alors $\frac{X}{Y}$ est une v.a.

CHAPITRE III. MOMENTS DES VARIABLES ALEATOIRES.

Dans tout le chapitre, les v.a. sont définies sur un même espace probabilisé $(\Omega, \mathcal{A}, P)$.

§ 1. ESPERANCE OU MOYENNE D'UNE V.A.

Définition 1. Si X est une v.a. discrète caractérisée par (x_j) et (p_j), on appelle espérance mathématique de X (ou simplement espérance de X), notée E[X], ou moyenne de X, notée m_X, le nombre réel défini par :

$$E[X] = m_X = \sum_j x_j\, p_j$$

pour autant que l'on ait :

$$\sum_j |x_j|\, p_j < +\infty$$

Définition 2. Si X est une v.a. continue ayant f_X pour f.fr., on appelle espérance de X, notée E[X], ou moyenne de X, notée m_X, le nombre réel défini par :

$$E[X] = m_X = \int_{-\infty}^{+\infty} x\, f_X(x)\, dx$$

pour autant que l'on ait :

$$\int_{-\infty}^{+\infty} |x|\, f_X(x)\, dx < +\infty$$

Définition 3. Si X est une v.a., si g est une application mesurable de $(\mathcal{R}, \mathcal{B})$ dans $(\mathcal{R}, \mathcal{B})$, on appelle espérance ou moyenne de la v.a. $Y = g(X)$, notée $E[Y] = E(g(X)) = m_Y$, le nombre réel défini de la manière suivante :

a) si X est discrète et caractérisée par (x_j) et (p_j), alors

$$E[Y] = E[g(X)] = m_Y = \sum_j g(x_j)\, p_j$$

pour autant que l'on ait :

$$\sum_j |g(x_j)|\, p_j < +\infty .$$

b) si X est une v.a. continue ayant f_X pour f.fr., alors

$$E[Y] = E[g(X)] = m_Y = \int_{-\infty}^{+\infty} g(x)\, f_X(x)\, dx$$

pour autant que l'on ait :

$$\int_{-\infty}^{+\infty} |g(x)|\ f_X(x)\ dx < +\infty$$

Remarques : 1) Pour chercher $E[g(X)]$, on peut procéder d'une autre façon : Dans le cas discret, on cherche d'abord (y_i) et (q_i) avec $q_i = P[Y = y_i]$ et on définit :

$$E[Y] = E[g(X)] = m_Y = \sum_i y_i\ q_i$$

pour autant que l'on ait :

$$\sum_i |y_i|\ q_i < +\infty.$$

Dans le cas continu, on cherche d'abord f_Y et on définit :

$$E[Y] = E[g(x)] = m_Y = \int_{-\infty}^{+\infty} y\ f_Y(y)\ dy$$

pour autant que l'on ait :

$$\int_{-\infty}^{+\infty} |y|\ f_Y(y)\,dy < +\infty.$$

Par un théorème général de la théorie de l'intégration, on peut montrer que les deux procédés sont équivalents, mais dans les applications, la définition 3 est plus commode.

2) D'après la définition 3, $E[X]$ n'existe que si $E[|X|]$ existe c'est-à-dire si $E[|X|] < +\infty$

Exemple 1. Soit X une v.a. ne prenant que la valeur a avec $P[X = a] = 1$, alors : $E[X] = a.\ 1 = a$.

En particulier, la moyenne d'une constante égale cette constante.

Exemple 2. Soit I_A l'indicatrice d'un événement $A \in \mathcal{A}$.

$$I_A \quad \begin{cases} 1 & 0 \\ P[A] = p & P[A^c] = q \end{cases} \qquad 0 \leqslant p \leqslant 1 \qquad p + q = 1$$

Alors, on a :

$$E\left[I_A\right] = 1 \times P[A] + 0 \times P\left[A^c\right] = P[A]$$

Ainsi, la probabilité d'un événement égale l'espérance mathématique de son indicatrice.

Exemple 3. Soit X une v.a. Binomiale de paramètre (n, p) (en

abrégé Bi (n, p))

$$X \quad \begin{cases} 0\ ,\ 1\ ,\ 2\ ,\ \ldots\ ,\ j\ ,\ \ldots\ n. \\ P[X = j] = C_n^j\ p^j\ q^{n-j} \end{cases}$$

$$E[X] = \sum_{j=o}^{n} j\ C_n^j\ p^j\ q^{n-j} = \sum_{j=1}^{n} j\ C_n^j\ p^j\ q^{n-j} = \sum_{j=1}^{n} j\ \frac{n!}{j!\,n-j!}\ p^j\ q^{n-j}$$

$$= n\,p \sum_{j=1}^{n} \frac{n-1!}{(j-1)!\ (n-j)!}\ p^{j-1}\ q^{(n-1)-(j-1)}$$

$$= np \sum_{i=o}^{n-1} \frac{(n-1)!}{i!\ (n-1-i)!}\ p^i\ q^{(n-1)-i} = np\ (p+q)^{n-1} = n\,p$$

Exemple 4. Soit X une v.a. de Poisson de paramètre $\lambda > 0$

$$E[X] = \sum_{j=o}^{+\infty} j\ \frac{\lambda^j}{j!}\ e^{-\lambda} = \sum_{j=1}^{+\infty} j\ \frac{\lambda^j}{j!}\ e^{-\lambda} = \lambda e^{-\lambda} \sum_{j=1}^{+\infty} \frac{\lambda^{j-1}}{j-1!}$$

$$= \lambda e^{-\lambda} \sum_{i=o}^{+\infty} \frac{\lambda^i}{i!} = \lambda e^{-\lambda}\ e^{\lambda} = \lambda$$

Remarque : Dans ces quatre exemples $E[X]$ existe puisque $E[X] = E[|X|] < +\infty$.

Exemple 5. Soit X une v.a. discrète caractérisée par

$$X \quad \begin{cases} (-1)^j\ \dfrac{2^j}{j} & j = 1, 2, 3, \ldots \\ p_j = \dfrac{1}{2^j} \end{cases}$$

On a bien $\sum_{j=1}^{+\infty} \frac{1}{2^j} = 1$ comme somme des termes d'une progression géométrique.

$$E[X] = \sum_{j=1}^{+\infty} (-1)^j\ \frac{2^j}{j}\ \frac{1}{2^j} = \sum_{j=1}^{+\infty} \frac{(-1)^j}{j} = -\log_e 2$$

par le développement de Taylor de $\mathrm{Log}_e\ (1+x)$ au point $x = 1$. Néanmoins, $E[X]$ n'existe pas, car

$$E[|X|] = \sum_{j=1}^{+\infty} \frac{1}{j} = +\infty$$

Exemple 6. Soit X une v.a. exponentielle ayant pour f.fr.

$$f(x) = \begin{cases} \mu\ e^{-\mu x} & x \geqslant 0 \qquad \mu > 0 \\ 0 & x < 0 \end{cases}$$

$$E[X] = E[|X|] = \int_0^{+\infty} x\,\mu\, e^{-\mu x}\, dx = \frac{1}{\mu}$$

par intégration par parties.

Exemple 7. Soit X une v.a. de Cauchy ayant pour f.fr.

$$f(x) = \frac{1}{\pi\,(1 + x^2)} \qquad -\infty < x < +\infty$$

Une telle v.a. n'a pas d'espérance, car :

$$E[|X|] = \int_{-\infty}^{+\infty} \frac{|x|}{\pi\,(1+x^2)}\, dx = 2\int_0^{+\infty} \frac{x}{\pi\,(1+x^2)}\, dx = \frac{1}{\pi}\int_1^{+\infty} \frac{dt}{t} = +\infty$$

Il est intéressant de remarquer qu'on a :

$$\forall\, a > 0 \qquad \int_{-a}^{+a} \frac{x\, dx}{(1+x^2)} = 0$$

puisqu'il s'agit de l'intégrale d'une fonction impaire prise entre des limites finies symétriques. Dès lors, on a :

$$\lim_{a \to +\infty} \int_{-a}^{+a} \frac{x\, dx}{(1+x^2)} = 0$$

Proposition 1. Si F est la f.d. d'une v.a. X, si E[X] existe et si a > 0, alors on a :

$$\lim_{a \to +\infty} a\ P[|X| > a] = 0$$

ou ce qui est équivalent :

$$\lim_{x \to +\infty} x\,[1 - F(x)] = 0 \qquad \lim_{x \to -\infty} x\,F(x) = 0$$

Démonstration (faite dans le cas continu seulement)

Puisque E[X] existe, on a $E[|X|] < +\infty$ et

$$E[|X|] = \int_{-\infty}^{+\infty} |x|\, f(x)\, dx \geqslant \int_{-\infty}^{-a} |x|\, f(x)\, dx + \int_{+a}^{+\infty} |x|\, f(x)\, dx$$

$$\geqslant a \int_{-\infty}^{-a} f(x)\, dx + a \int_a^{+\infty} f(x)\, dx = a\ P[|X| > a]$$

d'où le résultat puisque $E[|X|] < +\infty$ implique

$$\int_a^{+\infty} f(x)\, dx \to 0 \quad \text{quand } a \to +\infty$$

$$\int_{-\infty}^{-a} f(x)\, dx \to 0 \quad \text{quand } a \to +\infty$$

Proposition 2. Si F est la f.d. d'une v.a. X, si E[X] existe, alors

$$E[X] = \int_0^{+\infty} (1-F(x))\, dx - \int_{-\infty}^{0} F(x)\, dx \qquad (1)$$

Démonstration.

a) X continue ayant f pour f.fr. Intégrons (1) par parties, il vient :

$$E[X] = \Big[x\ (1-F(x))\Big]_0^{+\infty} + \int_0^{+\infty} x\, f(x)\, dx - \Big[x\, F(x)\Big]_{-\infty}^{0} + \int_{-\infty}^{0} x\, f(x)\, dx$$

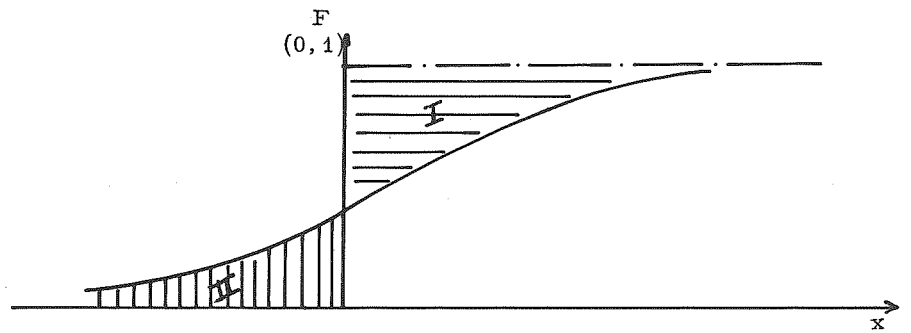
et on a le résultat en appliquant la proposition 1.

b) X est discrète caractérisée par (x_j) et (p_j). Intégrons (1) par parties, il vient

$$E[X] = \Big[x\ (1-F(x))\Big]_0^{+\infty} + \sum_{(x_j > 0)} x_j\, p_j - \Big[x\, F(x)\Big]_{-\infty}^{0} + \sum_{(x_j < 0)} x_j\, p_j$$

d'où le résultat en appliquant la proposition 1.

Graphiquement, la formule (1) montre que E[X] est la différence entre les aires I et II.



Exercice : Si X est une v.a. positive ($P[X \geqslant 0] = 1$) ayant F pour f.d. ($F(x) = 0$ pour $x < 0$) et si $m = E[X]$ existe et est différente de 0, alors :

$$G : \mathbb{R} \to \mathbb{R} \qquad G(x) = \frac{1}{m}\int_0^x (1-F(y))\, dy$$

est la f.d. d'une nouvelle v.a.

Proposition 3. Si X est une v.a., si E[X] existe, si a est une

constante, alors : $E[a\,X] = a\,E[X]$.

Proposition 4. Si X et Y sont deux v.a. et a, b, c des constantes, si $E[X]$ et $E[Y]$ existent, alors

$$E[a\,X + b\,Y + c] = a\,E[X] + b\,E[Y] + c$$

Nous admettons cette proposition qui est un théorème général de la théorie de l'intégration.

Corollaire : Si X est une v.a. et si $E[X] = m_X$ existe, alors

$$E\left[X - E[X]\right] = E\left[X - m_X\right] = 0$$

Des propositions 4 et 5, on déduit immédiatement :

Corollaire : Soient X et Y deux v.a. telles que $E[X]$ et $E[Y]$ existent. Si $P[X \geqslant Y] = 1$ (c'est-à-dire si $P[\,\omega : X(\omega) \geqslant Y(\omega)\,] : 1$) et, en particulier, si $X \geqslant Y$, alors $E[X] \geqslant E[Y]$.

Corollaire : Si $E[X]$ existe, on a $|E[X]| \leqslant E[|X|]$

Remarque : Cas d'une v.a. de type mixte.

Soit X une v.a. de type mixte ayant pour f.d.

$$F(x) := \begin{cases} 1 - 0{,}5\ e^{-(x-3)} & x \geqslant 3 \\ 0 & x < 3 \end{cases}$$

Alors ;

$$E[X] = 3 \times 0{,}5 + \int_3^{+\infty} 0{,}5 \times e^{-(x-3)}\,dx$$

$$= 1{,}5 + 0{,}5 \times 4 = 3{,}5$$

Exercice : Si $E[X]$ existe, il en est de même de $E[X^+]$ et $E[X^-]$.

§ 2. MOMENTS D'UNE V.A.

Définition 4. On appelle moment m_k, d'ordre k ($k \in N$), d'une v.a. X, l'espérance mathématique, quand elle existe, de X^k. On a donc :

$$m_k = E\left[X^k\right] \qquad \text{si} \quad E\left[|X|^k\right] < +\infty$$

En particulier, on a : $m_o = 1$, $m_1 = E[X]$

La relation $|X|^j \leqslant |X|^k + 1$ ($j, k \in N$, $j \leqslant k$) montre que si $E\left[|X|^k\right]$ existe, alors $E\left[X^j\right]$ existe pour tout $j \leqslant k$, puisque

$$E\left[|X|^j\right] \leqslant E\left[|X|^k\right] + 1.$$

Définition 5. Soit X une v.a. dont on suppose que $m_x = E[X]$ existe. On appelle moment centré μ_k, d'ordre k ($k \in N$), l'espérance mathématique quand elle existe de $(X-m)^k$

On a donc :

$$\mu_k = E\left[(X-m)^k\right] \qquad \text{si} \quad E\left[|X-m|^k\right] < +\infty$$

En particulier, on a : $\mu_o = 1$, $\mu_1 = 0$.

Définition 6. On appelle variance d'une v.a. dont on suppose que la moyenne existe, le moment centré μ_2 quand il existe. La variance de X, quand elle existe, est notée σ_X^2 ou var X.

On a donc :

$$\sigma_X^2 = \text{var } X = E\left[(X-m_x)^2\right] \text{ quand } E\left[(X-m_x)^2\right] < +\infty.$$

Définition 7. On appelle écart quadratique ou écart type d'une v.a. X dont la variance existe, la racine carrée arithmétique de la variance. Cet écart est noté σ_X.

La variance d'une v.a. jouit de deux propriétés fondamentales. Prenons $a \in \mathbb{R}$ et cherchons $E\left[(X-a)^2\right]$. On trouve :

$$\begin{aligned} E[(X-a)^2] &= E\left[(X-m_X + m_X - a)^2\right] \\ &= E\left[(X-m_X)^2\right] + 2\,(m_X-a)\,E\left[X-m_X\right] + (m_X-a)^2 \\ &= E\left[(X-m_X)^2\right] + (m_X-a)^2 = \sigma_X^2 + (m_X-a)^2 \end{aligned}$$

De cette dernière relation, on déduit :

1) Si $a \neq m_X$, on a : $\sigma_X^2 < E\left[(X-a)^2\right]$

2) Si $a = 0$, on a : $E\left[X^2\right] = \sigma_X^2 + m_X^2$

Proposition 6 : Si X est une v.a. dont la variance existe, on a :

$$\sigma_X^2 = E\left[X^2\right] - m_X^2 = E\left[X^2\right] - E^2\left[X\right]$$

Proposition 7 : Si X est une v.a. dont la variance existe et si a et b sont des constantes réelles, on a :

$$\text{var } (a\,X + b) = a^2 \text{ var } X.$$

$$\sigma_{aX+b} = |a|\,\sigma_X$$

Démonstration :

$$\text{var } (a\,X+b) = E\left[(aX+b - a\,m_X - b)^2\right] = E\left[a^2(X-m_X)^2\right] = a^2 \text{ var } X.$$

Proposition 8: Si X est une v.a. telle que $P[X=a] = 1$ et, en particulier, si X est une constante, alors var X = 0.

Démonstration.

On a : $P\left[X^2 = a^2\right] = 1$ et donc $E\left[X^2\right] = a^2$. On en déduit : $\sigma_X^2 = a^2 - a^2 = 0$.

Proposition 9 : Si X est une v.a. ayant une variance différente

de 0, alors pour la v.a. $Z = \frac{X-m_X}{\sigma_X}$, on a : $E[Z] = 0$ et $\operatorname{var} Z = 1$.

Démonstration :

$$E[Z] = \frac{1}{\sigma} E\left[X-m_X\right] = 0$$

$$\sigma_Z^2 = \frac{1}{\sigma_X^2}\ \sigma_X^2 = 1$$

Définition 8. Une v.a. ayant une moyenne est dite centrée si cette moyenne est nulle.

Définition 9. Une v.a. ayant une variance est dite réduite si sa moyenne est 0 et si sa variance est 1.

Remarque : Tout moment centré d'ordre k peut s'exprimer à l'aide des moments d'ordre inférieur à k. En particulier, on a :

$$\mu_3 = m_3 - 3\, m_1 m_2 + 2\, m_1^3$$

$$\mu_4 = m_4 - 4\, m_1 m_3 + 6\, m_1^2 m_2 - 3\, m_1^4$$

Définition 10. Pour une v.a. X dont le moment d'ordre k ($k \in N$) existe, on appelle moment factoriel d'ordre k, l'expression

$$m_{[k]} = E\left[X\ (X-1)\ (X-2) \ldots\ (X-k+1)\right]$$

On en déduit :

$$\operatorname{var} X = m_{[2]} + m_X - m_X^2$$

Exemple 8. Soit I_A l'indicatrice d'un événement $A \in \mathcal{A}$

$$\operatorname{var} I_A = (1-p)^2\, p + (0-p)^2\, q = q^2 p + p^2 q = pq(p+q)$$ par la définition 1.

ou $$\operatorname{var} I_A = 1^2 . p - p^2 = p - p^2 = p(1-p) = pq$$ par la proposition 6.

Exemple 9. Soit X une v.a. Bi (n, p), alors

$$E\left[X(X-1)\right] = \sum_{j=o}^{n} j\,(j-1)\, C_n^j\, p^j\, q^{n-j} =$$

$$= n\,(n-1)\, p^2 \sum_{j=2}^{n} \frac{n-2\,!}{j-2\,!\ n-j\,!}\ p^{j-2}\ q^{n-2-(j-2)}$$

$$= n\,(n-1)\, p^2\, (p+q)^{n-2} = n\,(n-1)\, p^2$$

par un procédé entièrement analogue à celui utilisé dans la recherche de la moyenne. On en déduit :

$$\operatorname{var} X = n(n-1)\, p^2 + np - (np)^2 = np - np^2 = np\,(1-p) = n\,p\,q$$

Exemple 10. Soit X une v.a. de Poisson de paramètre $\lambda > 0$.

$$E[X\,(X-1)] = \sum_{j=o}^{+\infty} j(j-1)\, \frac{\lambda^j}{j!}\, e^{-\lambda} = e^{-\lambda}\, \lambda^2 \sum_{j=2}^{+\infty} \frac{\lambda^{j-2}}{j-2\,!} = \lambda^2$$

$$\operatorname{var} X = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Exemple 11. Pour la v.a. X de l'exemple 6, on a :

$$E\left[X^2\right] = \int_o^{\infty} x^2\ \mu e^{-\mu x}\, dx = \frac{2}{\mu^2}$$

$$\operatorname{var} X = \frac{1}{\mu^2}$$

§ 3. INEGALITE DE CHEBYSHEV.

Soit X une v.a. de moyenne m et de variance σ^2. On a :

$$\forall\ a > 0 \qquad P\left[|X - m| \geqslant a\right] \leqslant \frac{\sigma^2}{a^2}$$

ou la formulation équivalente :

$$\forall\ a > 0 \qquad P\left[|X - m| \leqslant a\right] \geqslant 1 - \frac{\sigma^2}{a^2}$$

Démonstration :

Il suffit de démontrer la première partie. Considérons la v.a. Y définie par :

$$Y = \begin{cases} 0 & \text{si } |X-m| < a \\ a^2 & \text{si } |X-m| \geqslant a \end{cases}$$

On trouve immédiatement :

$$E[Y] = a^2\ P\left[|X-m| \geqslant a\right]$$

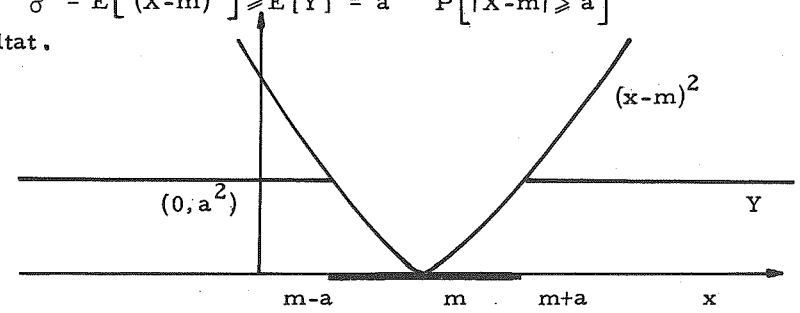
Mais on a aussi :

$$(X-m)^2 \geqslant Y$$

comme le montre la représentation graphique. On en conclut :

$$\sigma^2 = E\left[(X-m)^2\right] \geqslant E[Y] = a^2\ P\left[|X-m| \geqslant a\right]$$

d'où le résultat.



Corollaire. - Si X est une v.a. ayant une variance nulle, on a

$$P[X=m] = 1$$

Démonstration. Puisque $\sigma^2 = 0$, par l'inégalité de Chebyshev, on a:

$$\forall a > 0 \qquad 0 \leqslant P[|X-m| \geqslant a] \leqslant 0$$
$$0 \leqslant P[|X-m| > 0] \leqslant 0$$
Donc $$P[|X-m| > 0] = 0$$
et $$P[|X-m| = 0] = 1.$$

Remarques : Si $\sigma^2 \neq 0$, on peut donner à l'inégalité de Chebyshev une autre forme. Posons $a = k\sigma \qquad k > 0$, alors

$$\forall k > 0 \qquad P\left[|X-m| \geqslant k\sigma\right] \leqslant \frac{1}{k^2}$$

En particulier ,

pour $k = 1$ $\qquad P[|X-m| \geqslant \sigma] \leqslant 1$

pour $k = 2$ $\qquad P[|X-m| \geqslant 2\sigma] \leqslant \frac{1}{4}$

pour $k = 3$ $\qquad P[|X-m| \geqslant 3\sigma] \leqslant \frac{1}{9}$

Ces résultats sont valables quelle que soit la distribution de la v.a. utilisée. On peut les améliorer en précisant la distribution de la v.a. ou en supposant que la v.a. possède des moments d'ordre supérieur au second. Néanmoins, dans la classe des v.a. ayant un moment du second ordre, l'inégalité de Chebyshev est la meilleure qui soit. En effet, la v.a. X caractérisée par

$$X \quad \begin{cases} -k & , \quad 0 & , \quad k \qquad k > 0 \\ \frac{1}{2k^2} & , \quad 1 - \frac{1}{k^2} & , \quad \frac{1}{2k^2} \end{cases}$$

conduit à l'égalité dans l'inégalité de Chebyshev.

Pour cette variable, on a : $E[X] = 0$, $E\left[X^2\right] = \sigma_X^2 = 1$.

Et en prenant dans l'inégalité de Chebyshev le même k que dans la définition de la v.a., on trouve

$$P[|X-m| \geqslant k \ ] = P[|X| \geqslant k] = P[X = k] + P[X = -k] = \frac{1}{k^2}$$

L'inégalité de Chebyshev montre l'importance de la variance d'une v.a. qui mesure la dispersion des valeurs prises par la v.a. autour de sa moyenne. Ainsi, on a :

$$P[|X-m| \leqslant 3\sigma] \geqslant \frac{8}{9}$$

Il y a donc une probabilité supérieure à $\frac{8}{9}$ de trouver une valeur de la v.a. entre $m-3\sigma$ et $m+3\sigma$.

Définition 11. Si la variance d'une v.a. existe et est différente de zéro, on appelle précision de la v.a. l'inverse de la variance.

§ 4. COVARIANCE DE DEUX V.A. ET VARIANCE D'UNE SOMME DE V.A.

Définition 12. La covariance de deux v.a. X, Y dont la variance existe est définie par

$$\text{cov}(X,Y) = E[(X-m_X)(Y-m_Y)]$$

Il est immédiat que cov (X, Y) = cov (Y, X) et cov (X, X) = var X
D'autre part on a :

$$\text{cov}(X,Y) = E[XY - Ym_X - Xm_Y + m_X m_Y]$$
$$\text{cov}(X,Y) = E[XY] - m_X E[Y] - m_Y E[X] + m_X m_Y$$
$$\text{cov}(X,Y) = E[XY] - E[X]\,E[Y]$$

De cette dernière relation, on déduit

Proposition 10 : Si a, b, c, d sont des constantes et X, Y des v.a. dont la variance existe, alors

$$\text{cov}(aX+b,\ cY+d) = ac\ \text{cov}(X,Y)$$

A partir d'ici, on va utiliser des résultats du chapitre 4. Il faut donc lire d'abord le chapitre 4 avant de terminer le chapitre 3.

Proposition 11. Si X et Y sont deux v.a. indépendantes dont la variance existe, on a cov (X, Y) = 0.

La réciproque de cette dernière proposition est inexacte. Donnons-en deux exemples :

1) Soit X une v.a. telle que $E[X] = E\left[X^3\right] = 0$ et $E\left[|X|^3\right] < +\infty$
Alors,

$$\text{cov}(X, X^2) = E\left[X^3\right] - E[X]\,E\left[X^2\right] = 0$$

pourtant X et X^2 ne sont pas indépendantes.

2) Soient W et V deux v.a. de même fonction de distribution, ayant un moment du second ordre. On a donc :
$E(W) = E(V)$ et $E(W^2) = E(V^2)$. Posons : $X = W + V$ et $Y = W - V$.
On a : $E[Y] = 0$ et

$$\text{cov}(X,Y) = E[XY] - E[X].E[Y] = E\left[W^2 - V^2\right] = 0.$$

Pourtant, si W représente le résultat du premier lancer d'un dé et si V représente le résultat du second lancer du même dé, il est impossible que X et Y soient indépendantes puisqu'elles prennent en même temps des valeurs paires ou des valeurs impaires.

Proposition 12. Soient X et Y deux v.a. dont la variance existe. alors :

$$\text{var}(X+Y) = \text{var}\,X + \text{var}\,Y + 2\,\text{cov}(X,Y)$$

Démonstration :

$$\text{var}(X+Y) = E\left[\left\{(X+Y) - (m_X+m_Y)\right\}^2\right]$$

$$= E\left[(X-m_X)^2 + 2\,(X-m_X)\,(Y-m_Y) + (Y-m_Y)^2\right]$$

$$= \text{var } X + 2 \text{ cov } (X,Y) + \text{var } Y.$$

Corollaire : Soient X, Y deux v.a. indépendantes dont la variance existe, alors :

$$\text{var } (X+Y) = \text{var } X + \text{var } Y.$$

La propriété 12 s'étend immédiatement au cas de n variables.

Proposition 13. Soient $X_1, X_2, \ldots, X_n$ des v.a. dont la variance existe, alors on a :

$$\text{var } (X_1+X_2+\ldots+X_n) = \sum_{i=1}^{n} \text{var } X_i + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \text{cov } (X_i, X_j)$$

$$= \sum_{i=1}^{n} \text{var } X_i + 2 \sum_{\substack{j=1 \\ i<j}}^{n} \text{cov } (X_i, X_j)$$

Remarque : La sommation double comporte n (n-1) termes tandis que la sommation $\sum_{\substack{j=1 \\ i<j}}^{n}$ comporte C_n^2 termes.

Corollaire : Soient $X_1, X_2, \ldots X_n$ n v.a. satisfaisant à l'une des trois hypothèses :

a) Les n v.a. sont indépendantes
b) Les n v.a. sont 2 à 2 indépendantes
c) La covariance de tout couple de v.a. est nulle.

alors :

$$\text{var } (X_1 + X_2 + \ldots + X_n) = \sum_{i=1}^{n} \text{var } X_i$$

Exercice : 1) Soient $X_1, X_2, \ldots X_n$ n v.a. de même moyenne m. Si on pose :

$$\bar{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}, \text{ on a}$$

$$\forall\, n \in \mathbb{N}^* \qquad E\left[\bar{X}_n\right] = m$$

2) Soient $X_1, X_2, \ldots, X_n$ n v.a. indépendantes de même variance σ^2. On a :

$$\text{var } \bar{X} = \frac{\sigma^2}{n}.$$

§ 5. COEFFICIENT DE CORRELATION.

Définition 13. On appelle coefficient de corrélation de deux v.a. X, Y, dont on suppose que la variance existe et est différente de zéro, le nombre réel $\rho(X,Y)$ défini par :

$$\rho(X,Y) = \frac{\text{cov } (X,Y)}{\sigma_X\, \sigma_Y} \qquad \sigma_X\, \sigma_Y \neq 0.$$

Il est immédiat que si X, Y sont de plus indépendantes, alors $\rho(X,Y) = 0$.
Nous allons montrer :

Proposition 14 : Si X, Y sont deux v.a. dont la variance existe et est différente de zéro, on a :

$$-1 \leqslant \rho(X,Y) \leqslant 1$$

Démonstration :

Posons : $U = X - E[X]$, $V = Y - E[Y]$.

Nous avons :

$$\forall t \in \mathbb{R} \qquad 0 \leqslant E\left[(U-tV)^2\right] = E[U^2] - 2t\,E[UV] + t^2\,E[V^2]$$

Puisque le trinôme du second degré en t doit être positif, on a :

$$E^2[UV] - E[U^2]\,E[V^2] \leqslant 0$$

C'est l'inégalité de Cauchy-Schwarz pour des v.a. ayant un moment du second ordre. Elle est équivalente à :

$$\text{cov}^2 (X,Y) - \sigma_X^2\, \sigma_Y^2 \leqslant 0$$

ou

$$\rho^2 (X,Y) \leqslant 1.$$

Si $\rho^2(X,Y) = 1$, l'équation du second degré en t a une racine double t_o donnée par :

$$t_o = \frac{E[UV]}{E[V^2]} = \frac{\text{cov } (X,Y)}{\sigma_Y^2}$$

Si $\rho = +1$, on a : $\text{cov } (X,Y) = \sigma_X\, \sigma_Y$ et $t_o = \frac{\sigma_X}{\sigma_Y}$

Si $\rho = -1$, on a : $\text{cov } (X,Y) = -\sigma_X\, \sigma_Y$ et $t_o = \frac{\sigma_X}{\sigma_Y}$

Pour ces valeurs de t_o, on a : $E[(U - t_o V)^2] = 0$. Comme on avait déjà $E[U - t_o V] = 0$, par la définition de U et V, le corollaire de l'inégalité de Chebyshev appliqué à la v.a. $U - t_o V$ donne

$$P\left[U - t_o V = 0\right] = 1$$

On en déduit :

si $\rho = 1$ $\quad P\left[\frac{X-E[X]}{\sigma_X} = \frac{Y-E[Y]}{\sigma_Y}\right] = 1$

si $\rho = -1$ $\quad P\left[\frac{X-E[X]}{\sigma_X} \quad \frac{Y-E[Y]}{\sigma_Y}\right] = 1$

Proposition 15. Si X, Y sont deux v.a. dont le coefficient de corrélation $\rho(X,Y)$ existe et si a, b, c, d sont des constantes $(a > 0, c > 0)$, on a

$$\rho(a X + b,\ c Y + d) = \rho(X, Y)$$

Démonstration :

$$\rho(a X + b,\ c Y + d) = \frac{a\, c\, \mathrm{cov}(X,Y)}{a\, \sigma_X\, c\, \sigma_Y} = \rho(X,Y).$$

CHAPITRE IV INDEPENDANCE

Dans tout ce chapitre, l'espace probabilisé $(\Omega, \mathcal{A}, P)$ sera fixé.

§ 1. INDEPENDANCE DES EVENEMENTS.

Définition 1. On dit que les n sous-tribus $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ de la tribu $\mathcal{A}$ sont indépendantes (sous-entendu relativement à P) si pour tout choix de $A_i \in \mathcal{A}_i$, $i = 1, 2, \ldots, n$, on a :

$$P\left[\bigcap_{i=1}^{n} A_i\right] = \prod_{i=1}^{n} P[A_i]$$

Remarque : De cette définition, on déduit : si pour $i = 1, 2, \ldots, n$ $\mathcal{B}_i$ est une sous-tribu de $\mathcal{A}_i$, alors les sous-tribus $\mathcal{B}_i$ sont aussi indépendantes.

La définition montre également que si $\mathcal{A}_1 \subseteq \mathcal{A}_2 \subseteq \mathcal{A}$, alors $\mathcal{A}_1$ et $\mathcal{A}_2$ ne sont pas indépendantes sauf dans le cas trivial où $\mathcal{A}_1 = \{\Phi, \Omega\}$.

Si les tribus $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ sont indépendantes, alors m $(m \leqslant n)$ d'entre elles le sont aussi.

Définition 2 : On dit que les n événements $A_1, A_2, \ldots, A_n$ de $\mathcal{A}$ sont indépendants (sous-entendu relativement à P) si les n sous-tribus associées

$$\mathcal{A}_i = \{\Phi, A_i, A_i^c, \Omega\}$$

sont indépendantes.

Proposition 1. Pour que deux événements A_1 et A_2 de $\mathcal{A}$ soient indépendants, il faut et il suffit que :

$$P[A_1 \cap A_2] = P[A_1]\ P[A_2] \qquad (1)$$

Démonstration:

C. N. Evident d'après la définition 2.

C. S. Il faut voir que la relation (1) entraîne les autres relations de la définition (2). C'est une vérification facile en utilisant les relations entre les opérations d'ensembles.

$$- P[A_1^c \cap A_2] = P[A_2] - P[A_1 \cap A_2] = P[A_2] - P[A_2]\,P[A_1] = P[A_2]\,P[A_1^c]$$

- $P[A_1 \cap A_2^c] = P[A_1] . P[A_2^c]$ (par symétrie).

- $P[A_1^c \cap A_2^c] = P[(A_1 \cup A_2)^c] = 1 - [P\ A_1 \cup A_2]$.

$$= 1 - P[A_1] - P[A_2] + P[A_1]\ P[A_2]$$

$$= P[A_1^c] - P[A_2]\ P[A_1^c] = P[A_1^c]\ P[A_2^c]$$

- $P[A_i \cap \Omega] = P[A_i] = P[A_i]\ P[\Omega]$ $\qquad i = 1,2.$
- $P[\Phi \cap A_i] = P[\Phi] = P[\Phi]\ P[A_i]$ $\qquad i = 1,2.$
- $P[\Phi \cap \Omega] = P[\Phi] = P[\Phi]\ P[\Omega]$.

Remarque : la définition 2 montre clairement que la seule condition

$$P[A_1 \cap A_2 \cap A_3] = P[A_1].\ P[A_2].\ P[A_3]$$

ne suffit pas pour assurer l'indépendance entre les 3 événements A_1, A_2, A_3. Il faut y ajouter les conditions

$$P[A_i \cap A_j] = P[A_i]\ P[A_j] \qquad i \neq j \qquad i,j = 1,2,3.$$

On en déduit que si trois événements sont indépendants, deux d'entre eux sont également indépendants. Plus généralement, si n événements sont indépendants, m ($m \leqslant n$) d'entre eux sont également indépendants. Mais si trois événements sont deux à deux indépendants, les trois événements ne sont pas nécessairement indépendants. Il se peut même, dans ce cas, que l'événement $(A_1 \cap A_2)$ ne soit pas indépendant de A_3 comme le montre l'exemple suivant. On considère les quatre points de $\mathbb{R}^3$ ayant pour coordonnées $(1,0,0)$; $(0,1,0)$; $(0,0,1)$; $1,1,1)$. On associe à chacun la probabilité $1/4$. Soit A_i l'événement : la $i^{ième}$ coordonnée du point est 1 ($i = 1,2,3$). On a :

$$P[A_i] = \frac{1}{2} \qquad i = 1,2,3$$

$$P[A_i \cap A_j] = \frac{1}{4} = P[A_i].\ P[A_j] \qquad i \neq j\ ,\ i,j = 1,2,3.$$

$$P[A_1 \cap A_2 \cap A_3] = \frac{1}{4} \neq P[A_1]\ P[A_2]\ P[A_3]$$

$$P\left[(A_1 \cap A_2) \cap A_3\right] = \frac{1}{4} \neq P[A_1 \cap A_2]\ P[A_3] = \frac{1}{8}$$

Pour que les n événements $A_1, A_2, \ldots, A_n$ soient indépendants, il suffit que pour $k = 2,3,\ldots,n$ on ait :

$$P\left[A_{i_1} \cap A_{i_2} \cap A_{i_3} \cap \ldots \cap A_{i_k}\right] = P\left[A_{i_1}\right]\ P\left[A_{i_2}\right] \ldots\ P\left[A_{i_k}\right]$$

où $(i_1, i_2, \ldots, i_k)$ désigne une combinaison quelconque des nombres de la suite $1,2,\ldots,n$ pris k à k. Comme il y a C_n^k pareilles combinaisons, pour voir si les événements $A_1, A_2, \ldots, A_n$ sont indépendants, il suffit de vérifier $2^n - n - 1$ conditions de ce type convenablement choisies.

§ 2. INDEPENDANCE DES VARIABLES ALEATOIRES.

Définition 3. On dit que les n v.a. $X_1, X_2, \ldots, X_n$ définies sur $(\Omega, \mathcal{A}, P)$ sont indépendantes (sous-entendu relativement a P) si les sous-tribus associées $X_i^{-1}(\mathcal{B})$ (qui sont des sous-tribus de $\mathcal{A}$) sont indépendantes.

Ici aussi, si n v.a. sont indépendantes, alors (n-1) d'entre elles sont indépendantes mais la réciproque est inexacte. En particulier, n v.a. deux à deux indépendantes ne sont pas nécessairement indépendantes.

La définition 1 est équivalente à la suivante :

Définition 1' : Les sous-tribus $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ de $\mathcal{A}$ sont indépendantes si pour tout choix de v.a. positives Y_i telles que $Y_i^{-1}(\mathcal{B}) \subseteq \mathcal{A}_i$ (c'est-à-dire pour tout choix de v.a. positives Y_i $\mathcal{A}_i$ mesurables) et dont l'espérance existe, on a :

$$E\left[\prod_{i=1}^{n} Y_i\right] = \prod_{i=1}^{n} E[Y_i]$$

Il est facile de voir que 1' implique 1 en prenant $Y_i = I_{A_i}$ $A_i \in \mathcal{A}_i$. On a en effet :

$$E\left[\prod_{i=1}^{n} I_{A_i}\right] = \left[E\ I_{\bigcap_i A_i}\right] = P\left[\bigcap_1^n A_i\right]$$

$$\prod_{i=1}^{n}\left[E\ I_{A_i}\right] = \prod_{i=1}^{n} P\left[A_i\right]$$

Donc :

$$P\left[\bigcap_1^n A_i\right] = \prod_{i=1}^{n} P[A_i] \qquad \text{pour tout choix de } A_i \in \mathcal{A}_i$$

C'est la théorie de l'intégration qui montre que 1 implique 1'. Nous ne pouvons pas le faire ici.

Proposition 2. Si $X_1, X_2, \ldots, X_n$ sont des v.a. indépendantes, si $g_1, g_2, \ldots, g_n$ sont des applications mesurables de $\mathbb{R}$ dans $\mathbb{R}$, alors les v.a. $g_1(X_1)$, $g_2(X_2), \ldots,$ $g_n(X_n)$ sont indépendantes.

Démonstration : Ce résultat est évident en utilisant la remarque qui suit la définition 1. En effet, on a :

$$(g_i \circ X_i)^{-1}(\mathcal{B}) \subseteq X_i^{-1}(\mathcal{B})$$

et les sous-tribus $X_i^{-1}(\mathcal{B})$ sont indépendantes puisque $X_1, X_2, \ldots, X_n$ le sont.

Remarques : 1) Une constante est indépendante de toute v.a. car la sous-tribu associée est $\{\Phi, \Omega\}$ qui est indépendante de toute tribu.

2) Si X est une v.a. et si g est application mesurable de $\mathbb{R}$ dans $\mathbb{R}$, alors X et g(X) ne sont pas indépendantes car $(g \circ X)^{-1}(\mathcal{B}) \subseteq X^{-1}(\mathcal{B})$.

Proposition 3. Si $X_1, X_2, \ldots, X_n$ sont des v.a. indépendantes et si $E[X_i]$ existe pour $i = 1, 2, \ldots, n$ alors on a :

$$E\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} E[X_i]$$

Démonstration. Il suffit de faire la démonstration pour deux v.a. notées X et Y. Posons : $X = X^+ - X^-$, $Y = Y^+ - Y^-$. Alors X^+ et X^- sont indépendantes de Y^+ et de Y^-, leur espérance existe et elles sont positives. Alors :

$$\begin{aligned} E[XY] &= E\left[(X^+ - X^-)(Y^+ - Y^-)\right] \\ &= E\left[X^+Y^+ - X^-Y^+ - X^+Y^- + X^-Y^-\right] \\ &= E[X^+]\,E[Y^+] - E[X^-]\,E[Y^+] - E[X^+]\,E[Y^-] + E[X^-]E[Y^-] \\ &= E[X]\,E[Y^+] - E[X]\,E[Y^-] \\ &= E[X]\,E[Y] \end{aligned}$$

Définition 4. On dit qu'une suite $(X_i)\ i \in \mathbb{N}$ de v.a. est une suite indépendante de v.a. si pour tout $n \geqslant 2$ $(n \in \mathbb{N})$, les n v.a. $X_1, X_2, \ldots, X_n$ sont indépendantes.

§ 3. APPLICATION.

Considérons une pièce de monnaie parfaitement régulière qu'on lance 2 fois. L'ensemble des résultats possibles est $\Omega = \{ P_1 \cap F_2,\ P_1 \cap P_2,\ F_1 \cap F_2,\ F_1 \cap P_2 \}$ où P_i représente "obtenir pile ou lancer i" et F_i représente "obtenir face ou lancer i" (i = 1, 2). On supposera que les quatre événements élémentaires de Ω ont pour probabilité 1/4.

Un bon exercice pour le lecteur consiste à écrire les $2^4 = 16$ sous-ensembles de $\mathcal{P}(\Omega)$ et à les interpréter dans le langage des événements.

Prenons : $\mathcal{A}_i = \{ \Phi,\ F_i,\ F_i^c = P_i,\ \Omega \}\ i = 1, 2$. Les deux sous-tribus $\mathcal{A}_1$ et $\mathcal{A}_2$ de $\mathcal{P}(\Omega)$ sont indépendantes car

$$P[F_1 \cap F_2] = \frac{1}{4} = P[F_1]\ P[F_2] = \frac{1}{2} \cdot \frac{1}{2}$$

On peut associer aux deux sous-tribus $\mathcal{A}_1$ et $\mathcal{A}_2$ deux v.a. indépendantes, X_1 et X_2, ne prenant que deux valeurs et telles que X_i soit mesurable par rapport à $\mathcal{A}_i$ (i = 1, 2). Par exemple, X_i prend la valeur 1 sur F_i et 0 sur F_i^c (i = 1, 2) c.à.d. X_i prend la valeur 1 (0) si face (pile) sort au lancer i.

On a maintenant le sens précis de l'expression : les deux lancers de la pièce sont indépendants. Cela signifie que les deux sous-tribus $\mathcal{A}_1$ et $\mathcal{A}_2$ sont indépendantes. Le sens qui a été défini ici pour des lancers indépendants sera maintenu dans la suite.

CHAPITRE 5. PROBABILITE CONDITIONNELLE ET FORMULES DE BAYES.

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé fixé.

§ 1. PROBABILITE CONDITIONNELLE.

Définition 1. Soit A un événement fixé de $\mathcal{A}$. Supposons $P[A] > 0$. On appelle probabilité conditionnelle de $B \in \mathcal{A}$ sachant que A est réalisé, notée $P[B/A]$, (à lire P de B si A), l'expression

$$P[B/A] = \frac{P[B \cap A]}{P[A]}$$

La définition 1 n'a pas de sens si $P[A] = 0$.

Proposition 1 : $P[./A]$ est pour $A \in \mathcal{A}$ fixé avec $P[A > 0]$, une nouvelle probabilité sur $\mathcal{A}$.

Démonstration :

Il faut voir que la fonction $P[./A]$ définie sur $\mathcal{A}$ satisfait aux axiomes de Kolmogorov.

1) $\forall B \in \mathcal{A}$ $\quad 0 \leqslant P[B/A] \leqslant 1$

2) $\quad P[\Omega/A] = 1$

3) $(B_i) \in \mathcal{A}$ $\quad i \in \mathbb{N}$ et $i \neq j$ $\quad B_i \cap B_j = \Phi$

$$P\left[\bigcup_i B_i/A\right] = \frac{P[(\bigcup B_i) \cap A]}{P[A]} = \frac{P[\bigcup_i (B_i \cap A)]}{P[A]}$$

$$\geq \frac{\sum_i P(B_i \cap A)}{P[A]} = \sum_i P[B_i/A]$$

De la définition 1, on déduit immédiatement :

Proposition 2 : Si A et B sont deux événements indépendants et si $P[A] > 0$, alors $P[B/A] = P[A]$.

Proposition 3 : Si A et B sont deux événements et si $P[A] > 0$, alors

$$P[A \cap B] = P[A].\ P[B/A]$$

On peut généraliser la proposition 3 sous la forme :

Proposition 4 (Théorème du produit ou théorème des probabilités composées).

Pour n événements $A_1, A_2, \ldots, A_n$ tels que $P[A_1 \cap A_2 \cap \ldots \cap A_{n-1}] > 0$ on a :

$$P[A_1 \cap A_2 \cap A_3 \ldots \cap A_n] = P[A_1]\ P[A_2/A_1]\ P[A_3/A_1 \cap A_2] \ldots P[A_n/A_1 \cap A_2 \cap \ldots \cap A_{n-1}].$$

Démonstration.

Puisque

$$A_1 \cap A_2 \cap \ldots \cap A_{n-1} \subseteq A_1 \cap A_2 \cap \ldots \cap A_{n-2} \subseteq \ldots \subseteq A_1 \cap A_2 \subseteq A_1$$

on a :

$$0 < P[A_1 \cap A_2 \cap \ldots \cap A_{n-1}] \leqslant [P\ A_1 \cap A_2 \cap \ldots\ A_{n-2}] \leqslant \ldots \leqslant P[A_1 \cap A_2] \leqslant P[A_1] \ .$$

On peut procéder par récurrence en supposant la propriété vraie pour les ensembles $A_1, A_2, \ldots, A_{n-1}$. On a, par la proposition 3,

$$P\left[(A_1 \cap A_2 \cap \ldots \cap A_{n-1}) \cap A_n\right] = P[A_1 \cap A_2 \cap \ldots \cap A_{n-1}]\ P\ [A_n/A_1 \cap A_2 \cap \ldots\ A_{n-1}] \ .$$

§ 2. FORMULE DE DECOMPOSITION.

Prenons $A \in \mathcal{A}$ tel que $0 < P[A] < 1$, alors $0 < P[A^c] < 1$.

Pour $B \in \mathcal{A}$, on a :

$$\begin{aligned} P[B] &= P\ [B \cap A] + P\ [B \cap A^c] \\ &= P[A]\ P[B/A] + P[A^c]\ P[B/A^c] \end{aligned}$$

Cette formule est un cas particulier de la formule de décomposition.

Proposition 5. (Formule de décomposition ou théorème des probabilités totales).

Soit $(A_i)\ i \in N$ un système contradictoire d'événements $\in \mathcal{A}$ tels que $0 < P[A_i] < 1 \quad i \in N$. Soit B un événement de $\mathcal{A}$, alors :

$$\begin{aligned} P[B] &= \sum_i P[B \cap A_i] \\ P[B] &= \sum_i P[A_i]\ P[B/A_i] \end{aligned}$$

Cette formule de décomposition est très utile pour l'évaluation des probabilités. Elle permet de plus de rencontrer dans un cas élémentaire la vraie définition de la probabilité conditionnelle.

Considérons une v.a. X sur $(\Omega, \mathcal{A}, P)$ définie de la façon suivante :

X est une v.a. discrète, si $\omega \in A_i$, alors $X(\omega) = \dfrac{P[B \cap A_i]}{P[A_i]}$

et X prend la valeur $\dfrac{P[B \cap A_i]}{P[A_i]}$ avec la probabilité $P[A_i]$.

La v.a. X est ainsi définie sur tout Ω . Elle est mesurable

par rapport à la tribu $\mathcal{A}_1$ engendrée par le système contradictoire (A_i). La tribu $\mathcal{A}_1$ est une sous-tribu de $\mathcal{A}$.

L'espérance mathématique de cette v.a. X existe. Elle est donnée par :

$$E[X] = \sum_i \frac{P[B \cap A_i]}{P[A_i]} \cdot P[A_i] = P[B] \ .$$

La vraie définition d'une probabilité conditionnelle peut maintenant être formulée pour le cas particulier qui nous occupe.

Soit $\mathcal{A}_1$ la sous-tribu de $\mathcal{A}$ engendrée par le système contradictoire (A_i) d'événements de $\mathcal{A}$, alors la probabilité conditionnelle de l'événement $B \in \mathcal{A}$, par rapport à la sous-tribu $\mathcal{A}_1$, est la v.a. X définie plus haut, habituellement notée $P^{\mathcal{A}_1}(B)$.

Remarquons qu'on retrouve bien la définition du § 1.

$$\omega \in A_i \qquad P^{\mathcal{A}_1}(B)\ (\omega) = \frac{P[B \cap A_i]}{P[A_i]}$$

Nous limiterons à ces quelques considérations l'étude fondamentale mais difficile des probabilités conditionnelles.
Indiquons encore une conséquence de la formule de décomposition :

$$\inf_i P[B/A_i] \leqslant P[B] \leqslant \sup_i P[B/A_i]$$

§ 3. FORMULES DE BAYES.

Soient A et B deux événements de $\mathcal{A}$ tels que :

$$0 < P[A] < 1 \qquad\qquad 0 < P\ [B] < 1.$$

On a :

$$P[A \cap B] = P[A]\ P[B/A] = P[B]\ P[A/B]$$

d'où la première formule de Bayes.

$$P[A/B] = \frac{P[A]\ P[B/A]}{P[B]}$$

Si (A_i) est un système contradictoire d'événements de $\mathcal{A}$ tels que $0 < P[A_i] < 1$ et si $B \in \mathcal{A}$ satisfait $0 < P[B] < 1$, on a :

$$P[A_i/B] = \frac{P[A_i]\ P[B/A_i]}{P[B]}$$

En appliquant au dénominateur la formule de décomposition, on obtient la seconde formule de Bayes

$$P\ [A_i/B] = \frac{P[A_i]\ P\ [B/A_i]}{\sum_j P[A_j]\ P[B/A_j]}$$

Remarque : Aucun résultat de la théorie des probabilités n'a suscité plus de controverses. Notons que, puisqu'il est déduit correctement à partir des axiomes, il est parfaitement valable. Si je connais les $P[A_j]$, appelées probabilités a priori des A_j, je peux calculer les probabilités $P[A_j \mid B]$, appelées probabilités a posteriori des A_j . Mais on peut vouloir utiliser le théorème sous la forme suivante : ayant observé que B est réalisé, proposons-nous de savoir dans quelle mesure A_j est la cause de B. C'est de là que vient le nom de théorème des probabilités des causes. Dans la pratique, on ignore les probabilités a priori et on veut utiliser la formule de Bayes pour voir dans quelle mesure la réalisation de B confirme ou infirme certaines hypothèses.

Les formules de Bayes jouent un rôle prépondérant à l'heure actuelle, spécialement dans l'analyse statistique des phénomènes économiques. Elles permettent d'introduire des probabilités subjectives dans le choix des probabilités a priori. Par probabilités subjectives, on entend des probabilités qui, pour un même événement, peuvent varier d'individu à individu. Jusqu'ici nous n'avions rencontré que des probabilités objectives : quel que soit l'événement A, la probabilité $P[A]$ est la même pour tous les individus. Nous ne rencontrerons des probabilités subjectives que dans les applications.

Signalons encore le fait suivant : la première formule de Bayes peut s'écrire :

$$\frac{P[B \mid A]}{P[B]} = \frac{P[A \mid B]}{P[A]}$$

Ainsi si $P[B \mid A] = 2\ P[B]$, on a aussi $P[A \mid B] = 2\ P[A]$. Ceci illustre le fait important suivant : le fait que B soit réalisé, ne modifie en rien la probabilité de A ni celle de $A \cap B$ mais ce qui est modifié dans le concept de probabilité conditionnelle c'est l'unité de mesure.

Signalons encore que Renyi a construit, à partir du concept de probabilité conditionnelle, une axiomatique qui contient celle de Kolmogorov comme cas particulier. On peut trouver des indications sur cette question dans l'ouvrage de Renyi cité dans la bibliographie.

§ 4. APPLICATIONS.

1) Une classe de 20 élèves, appelée classe 1, contient 10 bons élèves, 5 moyens et 5 médiocres. Une seconde classe de 20 élèves, appelée classe 2, contient 5 bons élèves, 5 moyens et 10 médiocres. Un inspecteur, au jugement infaillible, se présente dans l'établissement. Il entre, au hasard, dans l'une des deux classes et interroge un élève pris au hasard dans cette classe. Il fait ensuite de même pour l'autre. Il en conclut que l'élève interrogé dans la première classe visitée était meilleur que celui interrogé dans la seconde. Quelle est la probabilité que le meilleur élève

interrogé soit un élève de la classe 1 ?

Soit A_i l'événement "choisir la classe i" (i = 1, 2). Par hypothèse du choix au hasard, on a $P[A_i] = \frac{1}{2}$. Soit B l'événement "le premier élève interrogé est le meilleur". On a :

$$P[A_1 \mid B] = \frac{P[A_1] \cdot P[B/A_1]}{P[A_1]\ P[B/A_1] + P[A_2]\ P[B/A_2]}$$

Evaluons $P[B/A_1]$. Il y a 20 x 20 choix possibles d'un élève de la classe 1 et d'un élève de la classe 2. Si on a pris un bon élève de la classe 1, alors B est réalisé si on prend l'un quelconque des 5 moyens ou des 10 mauvais de la classe 2. Si on a pris un élève moyen de la classe 1, alors B est réalisé si et seulement si on prend l'un des 10 médiocres de la classe 2. Si on a pris un élève médiocre dans la classe 1, alors l'événement B est impossible. Par dénombrement, on a donc

$$P[B \mid A_1] = \frac{10 + 5 \times 10) + 5 \times 10}{20 \times 20} = \frac{200}{400} = \frac{1}{2}$$

Le même raisonnement donne :

$$P[B \mid A_2] = \frac{75}{400} = \frac{3}{16} .$$

On peut obtenir le même résultat par une décomposition plus fine. Posons C_1 l'événement "choisir un bon élève de la classe 1", C_2 l'événement "choisir un élève moyen de la classe 1", C_3 l'événement "choisir un élève médiocre de la classe 1". On a alors :

$$P[B \mid A_1] = P[C_1]\ P[B \mid A_1 \cap C_1] + P[C_2]\ P[B \mid A_1 \cap C_2] + P[C_3]\ P[B \mid A_1 \cap C_3]$$

$$= \frac{10}{20} \times \frac{15}{20} + \frac{5}{20} \times \frac{10}{20} + \frac{5}{20} \times \frac{0}{20} = \frac{200}{400} = \frac{1}{2}$$

Finalement, on trouve pour résultat cherché :

$$P[A_1 \mid B] = \frac{\frac{1}{2}\ \frac{1}{2}}{\frac{1}{2}\ \frac{1}{2} + \frac{1}{2}\ \frac{3}{16}} = \frac{8}{11}$$

et on en déduit :

$$P[A_2 \mid B] = 1 - P[A_1\ B] = \frac{3}{11}$$

Reprenons le même problème en supposant que l'inspecteur interroge deux élèves différents par classe visitée. Supposons que les deux élèves de la première classe visitée sont supérieurs aux deux élèves de la seconde classe visitée. Si A_i (i = 1, 2) représente toujours l'événement "choisir la classe i" et si B' représente l'événement "les

deux élèves de la première classe visitée sont les meilleurs", on trouve :

$$P[B' \mid A_1] = \frac{C_{10}^2 C_{15}^2 + C_{10}^1 C_5^1 C_{10}^2 + C_5^2 C_{10}^2}{C_{20}^2 C_{20}^2} = \frac{7.425}{(C_{20}^2)^2}$$

$$P[B' \mid A_2] = \frac{C_5^2 C_{10}^2 + C_5^1 C_5^1 C_5^2 + C_5^2 C_5^2}{C_{20}^2 C_{20}^2} = \frac{800}{(C_{20}^2)^2}$$

$$P[A_1 \mid B'] = \frac{7425}{8225}$$

Cet exemple montre combien la répétition des observations modifie la probabilité conditionnelle. Mais il faut remarquer qu'il ne s'agit pas des mêmes événements.

2) Tirage dans une urne de composition inconnue.

Une urne contient N boules blanches ou noires. On effectue n extractions dans cette urne mais on remet dans l'urne la boule extraite lors de l'un des tirages avant de procéder au tirage suivant (tirage ou extraction sans remise). On suppose le procédé aléatoire : c'est-à-dire qu'il donne à tout moment à chaque boule la même probabilité d'être extraite. Sachant qu'on a obtenu une boule blanche à chaque tirage, que peut-on dire au sujet de la composition de l'urne ?

Ainsi formulé, le problème est mal posé puisqu'on ne dit rien sur la façon dont l'urne a été constituée. Il y a N + 1 compositions possibles : (0 blanche, N noire), (1 blanche, N-1 noires),..., (N blanches, 0 noires).

Premier cas : Supposons que les N+1 compositions sont également probables. Soit A_k l'événement "l'urne est composée de k blanches et de N-k noires. Sous l'hypothèse formulée, on a :

$$P[A_k] = \frac{1}{N+1} \quad k = 0, 1, \ldots N.$$

Soit B l'événement "les n boules extraites sont blanches". Alors, on a:

$$P[B \mid A_k] = \left(\frac{k}{N}\right)^n$$

En effet, puisque le procédé d'extraction est aléatoire, on a la probabilité $\frac{k}{N}$ d'extraire, au premier tirage, une boule blanche de l'urne qui contient k blanches et N-k noires. Comme le procédé est avec remise, cette probabilité reste la même à tous les tirages. Enfin, puisque les tirages sont indépendants, on a la probabilité $\left(\frac{k}{N}\right)^n$ d'extraire n blanches de cette urne.

Par la seconde formule de Bayes, on trouve :

$$P[A_k \mid B] = \frac{\frac{1}{N+1}\left(\frac{k}{N}\right)^n}{\sum_{j=0}^{N} \frac{1}{N+1}\left(\frac{j}{N}\right)^n} = \frac{k^n}{\sum_{j=0}^{N} j^n}$$

En particulier, si n = N = 6 , on trouve

$$P[A_6 \mid B] = \frac{6^6}{1^6 + 2^6 + 3^6 + 4^6 + 5^6 + 6^6} = \frac{46.656}{67.171} > \frac{1}{2} .$$

Ainsi, sous les hypothèses formulées, la composition 6 blanches et 0 noire est la plus probable de toutes les compositions possibles.

Second cas : Supposons que l'on constitue l'urne de la façon suivante : on effectue dans une urne qui contient autant de boules blanches que de boules noires, N tirages avec remise. On supposera que les tirages sont indépendants et puisqu'il y a toujours dans l'urne autant de blanches que de noires, on supposera qu'à chaque tirage, la probabilité d'avoir une blanche égale $\frac{1}{2}$ comme celle d'avoir une noire.

Avec les notations du premier cas, on trouve :

$$P[A_k] = C_N^k \left(\frac{1}{2}\right)^N$$

En effet, sous les hypothèses formulées pour la composition de l'urne, il y a 2^N compositions possibles, toutes également probables. Parmi celles-là il y en a C_N^k qui fournissent une composition de k blanches et N-k noires. On peut retrouver facilement le résultat par le schéma de Bernoulli (Chapitre VII. § 2).

Par la seconde formule de Bayes, on trouve :

$$P[A_k \mid B] = \frac{C_N^k \left(\frac{1}{2}\right)^N \left(\frac{k}{N}\right)^n}{\sum_{j=0}^{N} C_N^j \left(\frac{1}{2}\right)^N \left(\frac{j}{N}\right)^n} = \frac{C_N^k \, k^n}{\sum_{j=0}^{N} C_N^j \left(\frac{j}{N}\right)^n}$$

En particulier, si n = N = 6, on trouve :

$$P[A_6 \mid B] = \frac{6^6}{6.1^6 + 15.2^6 + 20.3^6 + 15.4^6 + 6.5^6 + 1.6^6} = \frac{46.656}{217.392}$$

Un calcul analogue donne :

$$P[A_5 \mid B] = \frac{6.5^6}{6.1^6 + 15.2^6 + 20.3^6 + 15.4^6 + 6.5^6 + 1.6^6} = \frac{93.750}{217.392}$$

$$P[A_4 \mid B] = \frac{61.440}{217.392} .$$

Ainsi, sous les hypothèses formulées dans le second cas, on voit que la composition 6 blanches et 0 noire n'est plus celle qui a la plus grande probabilité. On aperçoit ainsi l'importance du choix des probabilités a priori.

3) Loi de sucession de Laplace.

On dispose de N pièces différentes, portant les numéros $1, 2, \ldots, N$, et on suppose que la pièce numéro j ($j = 1, 2, \ldots, N$) a la probabilité p_j de donner face. Soit (C_j) $(1 \leqslant j \leqslant N)$ un système contradictoire d'événements. On choisit la pièce numéro j avec la probabilité $P[C_j]$. On suppose qu'il existe un j $(1 \leqslant j \leqslant N)$ tel que $p_j\, P[C_j] \neq 0$. On effectue n lancers indépendants de la pièce choisie. Sachant qu'ils ont tous donné "face", quelle est la probabilité que les r lancers indépendants de la même pièce, effectués ensuite, donnent encore "face" en supposant que les n+r lancers sont indépendants $(r \geqslant 1)$.

Définissons:

B = l'événement "les n premiers lancers du problème ont donné "face"".

D_r = l'événement "les r lancers ultérieurs du problème ont donné "face"".

Avec les notations antérieures, on peut écrire :

$$B = F_1 \cap F_2 \cap \ldots \cap F_n \quad \text{et} \quad D_r = F_{n+1} \cap F_{n+2} \cap \ldots \cap F_{n+r} .$$

Par la première formule de Bayes, on a :

$$P[D_r \mid B] = \frac{P[D_r \cap B]}{P[B]} = \frac{\sum_{j=1}^{N} P[C_j]\, P[D_r \cap B \mid C_j]}{\sum_{j=1}^{N} P[C_j]\, P[B \mid C_j]}$$

$$= \frac{\sum_{j=1}^{N} P[C_j]\, p_j^{n+r}}{\sum_{j=1}^{N} P[C_j]\, p_j^{n}}$$

Supposons maintenant que

$$P[C_j] = \frac{1}{N} \qquad j = 1, 2, \ldots, N$$

$$p_j = \frac{j}{N} \qquad j = 1, 2, \ldots, N$$

Ces hypothèses donnent au problème traité une formulation équivalente à celle du premier cas de la seconde application. Ceci illustre l'intérêt de la formulation de certains problèmes par des modèles d'urnes.

Insistons sur le fait que dans le contexte du problème, la probabilité d'obtenir face était une v.a. et que les nouvelles hypothèses formulées en font une v.a. prenant chacune des valeurs $\frac{1}{N}, \frac{2}{N}, \ldots, \frac{N}{N}$ avec la probabilité $\frac{1}{N}$ qu'on appelle v.a. discrète uniforme.

On trouve maintenant :

$$P[D_r \mid B] = \frac{\sum_{j=1}^{N} p_j^{n+r}}{\sum_{j=1}^{N} p_j^{n}} .$$

Supposons encore que N soit grand, alors, on peut approximer les sommes par des intégrales. On a en effet :

$$\frac{1}{N} \sum_{j=1}^{N} \left(\frac{j}{N}\right)^n \simeq \int_0^1 x^n\, dx = \frac{1}{n+1}$$

On trouve finalement pour la loi de succession de Laplace :

$$P[D_r \mid B] = \frac{n+1}{n+r+1} \qquad r = 1, 2, \ldots$$

La loi de succession de Laplace correspond au cas r = 1.

Le problème traité a encore reçu la présentation suivante : si n expériences indépendantes ont confirmé une théorie, quelle est la probabilité que les r expériences suivantes, telles que les n+r expériences soient indépendantes, confirment encore la théorie. Il ne faut pas perdre de vue l'ensemble des hypothèses qui ont conduit à la loi de succession de Laplace. Vouloir l'utiliser, sans tenir compte de cette précaution élémentaire, conduit à des résultats absurdes. Considérons, par exemple, un enfant de 8 ans. Puisqu'il a déjà vécu 8 ans, la loi de succession de Laplace lui attribuerait une probabilité $\frac{9}{10}$ d'atteindre son neuvième anniversaire. Son grand-père qui a 98 ans verrait la même loi lui attribuer la probabilité $\frac{99}{100}$ d'atteindre l'âge de 99 ans.

Un autre exemple célèbre d'application erronée est attribué à Laplace lui-même. "Puisque le soleil s'est levé pendant 5.000 ans, c'est-à-dire 1.826.213 jours, je suis prêt à parier 1 franc contre 1.826.214 francs qu'il se levera demain. Si on veut appliquer la loi de Laplace à ce problème il faut se rendre compte du fait suivant : puisque le soleil s'est levé pendant 1.826.213 jours, la probabilité qu'il se lève chacun des 1.826.214 jours à venir est, d'après la loi de succession de Laplace

$$\frac{1.826.213 + 1}{1.826.213 + 1.826.214 + 1} = \frac{1}{2}$$

Ceci implique que la probabilité qu'il ne se lève pas pendant l'un au moins des 1.826.214 jours à venir est $\frac{1}{2}$.

Ce sont des applications de cette nature, traitées sans vérification préalable des hypothèses, qui ont jeté un certain doute sur la validité de l'utilisation des formules de Bayes.

4) Le gardien d'une prison annonce à trois prisonniers notés A, B, C la chose suivante : un des trois prisonniers a été choisi au hasard et sera exécuté, les deux autres seront libérés. Le prisonnier A a étudié la théorie des probabilités. Il conclut qu'il a la probabilité $\frac{1}{3}$ d'être exécuté. Il appelle le gardien et le prie de lui dire lequel de ses deux camarades sera exécuté argumentant qu'il soit déjà que l'un des deux au moins sera libéré. Le gardien refuse de répondre à cette question car il lui semble que si A savait lequel de ses deux compagnons sera libéré alors sa probabilité d'être exécuté deviendrait $\frac{1}{2}$. Montrez que le gardien se trompe et que la probabilité que A soit exécuté reste $\frac{1}{3}$ même si le gardien répond à la question en supposant toutefois que lorsque A doit être exécuté le gardien choisit au hasard B ou C comme devant être libéré (PARZEN, p. 126).

CHAPITRE VI. - FONCTION CARACTERISTIQUE ET CONVOLUTION

§ 1. CONVOLUTION.

Définition 1 : On dit qu'une v.a. discrète X est une v.a. qui ne prend que des valeurs entières et positives (en abrégé v.a. e.p.), si l'image de Ω par X est N ou une partie de N.

Soient X et Y deux v.a.e.p. indépendantes et posons :

$$p_j = P[X = j] \qquad r_j = P[Y = j] \qquad j \in N$$

Considérons la v.a. $Z = X + Y$. Il est clair que E est une v.a.e.p. et il faut chercher :

$$j \in N \qquad s_j = P[Z = j]$$

Mais on a :

$$[Z = j] = \bigcup_{i=o}^{j} \{[X = i] \cap [Y = j - i]\}$$

où les événements du second membre sont deux à deux incompatibles. Dès lors, puisque les deux v.a. X et Y sont indépendantes, on a :

$$P[Z = j] = \sum_{i=o}^{j} P[X=i] \quad P[Y = j-i]$$

ou

$$j \in N \qquad s_j = \sum_{i=o}^{j} p_i \, r_{j-i} \ .$$

Définition 2 : On dit que la suite (s_j) est le produit de convolution de la suite (p_j) et de la suite (r_j) et on écrit

$$(s_j) = (p_j) * (r_j)$$

si

$$\forall j \in N \qquad s_j = \sum_{i=o}^{j} p_i \, r_{j-i} \ .$$

En particulier, si $\forall j \in N$, $r_j = p_j$, on introduit la notation

$$(p_j) * (p_j) = (p_j)^{*2}$$

avec

$$(p_j)^{*1} = (p_j)$$

$$(p_j)^{*0} = (1) \quad \text{où } (1) \text{ représente la suite } 1,0,0,0,\ldots$$

De la définition, on déduit que :

$$(p_j) * (r_j) = (r_j) * (p_j)$$

Si X et Y sont deux v.a. continues indépendantes ayant pour f.fr. f_X et f_Y, alors la f.fr. f_Z de Z = X + Y est donnée par :

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(z-y)\, f_Y(y)\, dy = \int_{-\infty}^{+\infty} f_Y(z-x)\, f_X(x)\, dx$$

Il n'est pas possible de le montrer ici. Nous ne l'utiliserons pas dans la suite.

§ 2. FONCTION CARACTERISTIQUE.

Soient X et Y deux v.a. définies sur $(\Omega, \mathcal{A}, P)$. Alors, Z = X + iY est appelée une v.a. complexe sur $(\Omega, \mathcal{A}, P)$ pour laquelle on définit :

$$E[Z] = E[X] + i\, E[Y]$$

pour autant que E[X] et E[Y] existent. L'espérance des v.a. complexes jouit des propriétés élémentaires de l'espérance des v.a. réelles. Attirons seulement l'attention sur la démonstration de la propriété.

$$|E[Z]| \leqslant E[|Z|]$$

Pour cela, nous passons aux coordonnées polaires et nous posons :

$$Z = R\, e^{i\,\textcircled{H}} \qquad R \text{ et } \textcircled{H} \text{ sont des v.a. réelles}$$

$$E[Z] = \rho\, e^{i\varphi} \qquad \rho \text{ et } \varphi \text{ sont des constantes réelles}$$

On trouve alors :

$$|E[Z]| = \rho = e^{-i\varphi} E\left[R\, e^{i\,\textcircled{H}}\right] = E\left[R\, e^{i(\textcircled{H}-\varphi)}\right]$$
$$= E[R\cos(\textcircled{H}-\varphi)] \leqslant E[R] = E[|Z|]$$

Remarque : Le terme v.a. est réservé pour les v.a. réelles.

Définition 3 : Si X est une v.a. sur $(\Omega, \mathcal{A}, P)$, on appelle fonction caractéristique de X (notée en abrégé f.c.), la fonction φ_X du paramètre réel t définie par

$$t \in \mathbb{R} \qquad \varphi_X(t) = E\left[e^{itX}\right] = E[\cos tX] + i\, E[\sin tX]$$

Il est clair que la définition a un sens car si X est une v.a., il en est de même de cos tX et de sin tX pour tout $t \in \mathbb{R}$. La f.c. existe pour toute v.a. ainsi qu'il résulte de la proposition suivante .

Proposition 1. Si X est une v.a., alors

a) $\varphi_X(0) = 1$ et $\forall t \in \mathbb{R}\ |\varphi_X(t)| \leqslant 1$

Donc la f.c. existe pour toute v.a.

b) φ_X est partout continue

c) Si a, b sont des constantes réelles, alors

$$\varphi_{aX+b}(t) = e^{itb}\, \varphi_X(at)$$

Démonstration

a) Evident puisque $|e^{itX}| \leqslant 1$ donne $|E[e^{itX}]| \leqslant 1$

b) Pour tout $h \in \mathbb{R}$, on a :

$$|\varphi_X(t+h) - \varphi_X(t)| = \left|E\left[e^{i(t+h)X}\right] - E\left[e^{itX}\right]\right|$$
$$= \left|E\left[e^{itX}\,(e^{ihX} - 1)\right]\right|$$
$$E\left[\left|e^{itX}\,(e^{ihX} - 1)\right|\right]$$

Mais comme on a :

$$\left|e^{itX}\,(e^{ihX} - 1)\right| \leqslant \left|e^{ihX} - 1\right|$$

on voit que

$$|\varphi_X(t+h) - \varphi_X(t)| \quad \text{tend vers 0 quand } h \to 0$$

c) $\varphi_{aX+b}(t) = E\left[e^{it(aX+b)}\right] = e^{itb}\, E\left[e^{itaX}\right] = e^{itb}\, \varphi_X(at)$

Définition 4. On dit que deux v.a. complexes sont indépendantes si et seulement si les parties réelles sont indépendantes et les parties imaginaires sont indépendantes.
La notion s'étend facilement au cas de plus de deux v.a.

Proposition 2. Si X et Y sont deux v.a. indépendantes, alors

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

Démonstration : Puisque X et Y sont indépendantes, il en est de même de e^{itX} et e^{itY}. Dès lors, on a :

$$E\left[e^{it(X+Y)}\right] = E\left[e^{itX}\, e^{itY}\right] = E\left[e^{itX}\right] E\left[e^{itY}\right]$$

ce qui est équivalent au résultat à démontrer.

Remarque : Cette propriété de la f.c. permet de remplacer l'opération difficile de convolution, dans l'important problème de l'addition des v.a., par l'opération simple de produit ordinaire des f.c.

Cette remarque acquiert toute son importance grâce au théorème suivant :

Théorème d'unicité.

La connaissance de la f.c. d'une v.a. est équivalente à celle de la f.d. ou, en d'autres termes, deux v.a. ayant des f.c. distinctes ont des f.d. différentes.

Ce théorème est évident dans un sens puisque la f.c. est une espérance mathématique. Dans l'autre sens, il constitue ce que l'on appelle le théorème d'inversion de Fourier dont le rôle est primordial dans de nombreux domaines.
On peut citer le corollaire suivant :

Corollaire : Si φ_X est la f.c. d'une v.a. X et si

$$\int_{-\infty}^{+\infty} \left| \varphi_X(t) \right| \, dt \quad < +\infty$$

alors la v.a. X a une fonction de fréquence donnée par :

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \, \varphi(t) \, dt \, .$$

Nous ne démontrerons pas ce théorème ni son corollaire. Il en sera de même du théorème de continuité qui suit. On peut trouver d'élégantes démonstrations de ces résultats dans le tome II de l'ouvrage de Feller cité dans la bibliographie.

Théorème de continuité.

La condition nécessaire et suffisante pour qu'une suite (F_n) de f.d. d'une suite (X_n) de v.a. converge vers la f.d. F d'une v.a. X, en tout point de continuité de la fonction F, est que la suite (φ_n) des f.c. des X_n converge vers une fonction φ continue à l'origine. La fonction φ est alors continue partout et φ est la f.c. de X.

Nous rencontrerons plus loin des applications de ces deux importants théorèmes.

La proposition suivante montre que par dérivation de la f.c. d'une v.a., on peut trouver les moments qui existent de cette v.a.

Proposition 3 : Si $E[X^k]$ existe, alors la dérivée d'ordre k de φ_X existe et est continue. Elle est donnée par :

$$\varphi_X^{(k)}(t) = E\left[e^{itX} \, (iX)^k \right] \qquad k \in \mathbb{N}$$

et

$$\varphi_X^{(k)}(0) = i^k \, E\left[X^k \right] \qquad k \in \mathbb{N}$$

Démonstration : Il suffit de faire la démonstration pour k=1. Par

la définition de φ_X, on a :

$$\frac{\varphi_X(t+h) - \varphi_X(t)}{h} = E\left[e^{itX} \, \frac{e^{ihX} - 1}{h} \right]$$

et comme

$$\left| e^{itX} \, \frac{e^{ihX} - 1}{h} \right| \leqslant |X|$$

on a le résultat.

Remarque. La démonstration fait implicitement appel à la possibilité d'intervertir, sous les hypothèses données, les signes lim et E.

En utilisant le développement de Taylor et la proposition 3, on a :

Proposition 4 : Si $m_k = E[X^k]$ existe, alors

$$\varphi_X(t) = 1 + it\,m_1 + \ldots + \frac{(it)^k}{k!} \, m_k + 0\left(|t|^k\right)$$

où $0\left(|t|^k\right)$ représente une fonction qui tend vers 0, quand $|t| \to 0$, plus rapidement que $|t|^k$.

§ 3. FONCTION GENERATRICE.

Pour des v.a.e.p., au lieu de la f.c., on utilise souvent la fonction génératrice (notée f.g. en abrégé). Elle est définie par

$$s \in \mathbb{R} \qquad G(s) = \sum_{j=o}^{+\infty} p_j \, s^j \qquad \text{si } p_j = P[X = j]$$

Puisque $\sum_{j=o}^{+\infty} p_j = 1$, la série converge absolument au moins pour $-1 \leqslant s \leqslant 1$. Donc la série dérivée converge au moins pour $-1 < s < 1$. Si E X existe, alors :

$$G'(1) = \sum_{j=o}^{\infty} j \, p_j = E[X]$$

De même, si $E[X^2]$ existe, on a :

$$G''(1) = \sum_{j=o}^{\infty} j\,(j-1)\, p_j = E[X\,(X-1)]$$

et

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

Posons : $q_j = P[X > j]$ et $H(s) = \sum_{j=o}^{+\infty} q_j \, s^j$

On a la relation :

$$(1-s)\, H(s) = 1 - G(s) \qquad -1 < s < 1$$

et si E[X] existe, on a :

$$H(1) = G'(1) = E[X] = \sum_{j=1}^{+\infty} j \, p_j \, .$$

On trouve une nouvelle formule pour évaluer E[X], à savoir :

$$E[X] = \sum_{j=o}^{+\infty} q_j$$

On pouvait d'ailleurs l'écrire directement à partir de la définition de E[X] puisque :

$$\begin{aligned}\sum_{j=1}^{+\infty} j\,p_j = {} & p_1 + p_2 + p_3 + p_4 + \dots \\ & + p_2 + p_3 + p_4 + \dots \\ & + p_3 + p_4 + \dots \\ & + p_4 + \dots \\ & \vdots\end{aligned}$$

et la première ligne est q_o , la seconde q_1, la troisième q_2,... Remarquons que H n'est pas la f. g. d'une v. a. mais que $\frac{H}{E[X]}$ l'est.

Si $\sum_{j=o}^{+\infty} p_j\, s^j$ converge pour un $s_o > 1$, alors le moment de tout ordre de X existe et

$$L(s) = G(e^s) = \sum_{k=o}^{+\infty} m_k \frac{s^k}{k!} \qquad \text{où } m_k = E[X^k]$$

La fonction L s'appelle la fonction génératrice des moments de X puisque :

$$L^{(k)}(0) = m_k$$

La fonction L est liée à la f. c. de X puisque

$$L(is) = \varphi_X(s)$$

Pour une v. a. continue X, la fonction $\psi : \mathbb{R} \to \mathbb{R}$ est appelée la fonction génératrice des moments de X (en abrégé f. g. m), s'il existe un $s_o > 0$ tel que l'intégrale

$$\psi(s) = \int_{-\infty}^{+\infty} e^{s x} f_X(x)\, dx$$

converge pour $|s| \leqslant s_o$. On trouve alors :

$$\psi^{(k)}(0) = E[X^k]$$

Remarque. La f. g. et la f. g. m. jouissent d'une propriété analogue à celle qui figure dans la proposition 2 pour la f. c.

§ 4. SOMME D'UN NOMBRE ALEATOIRES DE V. A.

Soit (X_i) une suite de v. a. e. p., indépendantes, ayant toutes la même f. d. Soit N une v. a. e. p. indépendante de la suite (X_i). Cette dernière partie de phrase signifie que pour tout $k \in \mathbb{N}$, les v. a. $N, X_1, X_2, \dots, X_k$ sont indépendantes.

Posons :

$$S_N = X_1 + X_2 + \dots + X_N \qquad S_o = 0$$

Alors, S_N est une v. a. e. p. et

$$P[S_N = j] = \sum_{n=o}^{+\infty} P[N=n]\; P[X_1 + X_2 + \dots + X_n = j]$$

$$\varphi_{S_N}(t) = \sum_{j=o}^{+\infty} e^{itj} P[S_N = j]$$

$$= \sum_{j=o}^{+\infty} \sum_{n=o}^{+\infty} e^{itj} P[N=n]\; P[X_1+X_2+\dots+X_n=j]$$

et, en intervertissant les signes de sommation, ce qui est légitime, puisque, pour tout $t \in \mathbb{R}$ on a $|\varphi_{S_N}(t)| \leqslant 1$, on trouve

$$\varphi_{S_N}(t) = \sum_{n=o}^{+\infty} P[N=n] \sum_{j=o}^{+\infty} e^{itj} [P\, X_1 + X_2 + \dots + X_n = j]$$

et en vertu de la proposition 2, on a :

$$\varphi_{S_N}(t) = \sum_{n=o}^{+\infty} P[N=n]\; \varphi_X^n(t)$$

Soit G la f. g. de N. On trouve alors

$$\varphi_{S_N}(t) = G \circ \varphi_X(t)$$

Remarquons qu'on peut avoir $P[S_N = 0] > 0$ même si on a $P[X_i = 0] = 0$. Les v. a. de type S_N se rencontrent dans de nombreux problèmes : dans les files d'attente où les personnes arrivent par "grappes". (N est alors le nombre de grappes et X le nombre de personnes par grappes); en biologie (N est le nombre d'espèces animales et X le nombre d'éléments de l'espèce); en théorie des assurances (N est le nombre d'accidents et X le nombre de blessés par accidents).

Supposons que $E[X^2]$ et $E[N^2]$ existent. On peut alors trouver la moyenne et la variance de S_N. On trouve :

$$E[S_N] = \left[\frac{1}{i}\,\varphi'_{S_N}(t)\right]_{t=o} = \frac{1}{i}\left[G'\,\varphi_X(t)\,.\,\varphi'_X(t)\right]_{t=o}$$

$$= \frac{1}{i}\,E[N]\; iE[X] = E[N].\,E[X]$$

C'est le résultat auquel on s'attendait. Il ne faut cependant pas en déduire que $E[S_N^2] = E[N].\,E[X^2]$. On trouve en effet :

$$E[S_N^2] = -\varphi''_{S_N}(0) = -\left[G''(\varphi_X(t))\;\varphi'^2_X(t) + G'(\varphi_X(t))\;\varphi''_X(t)\right]_{t=o}$$

$$= -\left[G''(1)\, i^2\, E^2[X] - G'(1)\, E[X^2]\right]$$

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

$$= E[N]\, E^2[X] + E[N]\left\{E[X^2] - [E^2]X\right\}$$

$$= E[N]\, E^2[X] + E[N]\ \text{var}\ X$$

d'où

$$\text{var}\ S_N = E[N]\, E^2[X] + E[N]\,\text{var}\ X - E^2[N]\, E^2[X]$$

$$= E[N]\ \text{var}\ X + E^2[X]\ \text{var}\ N$$

Dans le cas particulier où N est une v.a. de Poisson de paramètre $\lambda > 0$, on trouve :

$$\varphi_{S_N}(t) = \sum_{n=o}^{+\infty} e^{-\lambda} \frac{\lambda^n}{n!} \varphi_X^n(t) = e^{-\lambda + \lambda \varphi_X(t)}$$

$$= e^{-\lambda[1 - \varphi_X(t)]}$$

Dans ce cas, on dit que la v.a. S_N est une v.a. de Poisson généralisée. La v.a. binomiale négative rentre dans ce cadre comme on le verra plus loin.

CHAPITRE VII. CONVERGENCE DES SUITES DE V.A. LOIS DES GRANDS NOMBRES. THEOREME CENTRAL LIMITE.

Les v.a. utilisées dans ce chapitre sont définies sur un espace probabilisé fixé $(\Omega, \mathcal{A}, P)$. Nous étudierons deux types de convergence pour une suite de v.a. Avec chacun de ces types, on exprimera une loi des grands nombres.

§ 1. CONVERGENCE EN PROBABILITE OU CONVERGENCE STOCHASTIQUE.

Définition 1. On dit que la suite (X_n) $(n \in \mathbb{N})$ de v.a. converge en probabilité ou converge stochastiquement vers 0, et on écrit :

$$\operatorname{plim}_{n \to +\infty} X_n = 0 ,$$

si :

$$\forall \varepsilon > 0 \qquad \lim_{n \to +\infty} P\left[\omega : |X_n(\omega)| \geq \varepsilon\right] = 0$$

Cette formulation est équivalente à la suivante .
On a : $\operatorname{plim}_{n \to +\infty} X_n = 0$ si et seulement si :

$$\forall \varepsilon > 0,\ \forall \delta > 0 \quad \exists n_o(\varepsilon, \delta) \text{ tel que : } \forall\ n > n_o(\varepsilon, \delta) \text{ on ait :}$$

$$P\left[\omega : |X_n(\omega)| \geq \varepsilon\right] \leq \delta$$

Remarque : Dans la suite, on écrira le plus souvent $P\left[|X_n| \geq \varepsilon\right]$ au lieu de $P\left[\omega : |X_n(\omega)| \leq \varepsilon\right]$.

Définition 2 : On dit que la suite (X_n) $(n \in \mathbb{N})$ de v.a. converge en probabilité on converge stochastiquement vers la v.a. X, et on écrit :

$$\operatorname{plim}_{n \to +\infty} X_n = X$$

si la suite $(X_n - X)$ converge en probabilité vers 0.

Remarque : La convergence en probabilité ne demande aucune condition sur l'existence des moments. Si, pour tout $n \in \mathbb{N}$, $E[X_n]$ existe, la convergence en probabilité de la suite (X_n) n'entraîne pas pour autant la convergence de la suite des nombres $E[X_n]$. Considérons, par exemple, la suite (X_n) $(n \in \mathbb{N}^*)$ de v.a. ne prenant que deux valeurs, caractérisées par :

$$X_n = \begin{cases} 0 & , \quad x_n \\ 1 - \frac{1}{n} & , \quad \frac{1}{n} \end{cases}$$

On trouve :

$$E\left[X_n\right] = \frac{x_n}{n}$$

$$P\left[|X_n| \geq \varepsilon\right] = P\left[X_n = x_n\right] = \frac{1}{n} \longrightarrow 0 \text{ quand } n \longrightarrow +\infty.$$

On a donc bien :

$$\operatorname*{plim}_{n \longrightarrow +\infty} X_n = 0$$

Prenons divers exemples pour x_n

a) $x_n = \sqrt{n}$, $E\left[X_n\right] = \frac{1}{\sqrt{n}} \longrightarrow 0 = E[0]$

La suite $E[X_n]$ converge vers $E[0]$.

b) $x_n = n$, $E\left[X_n\right] = 1$

La suite $E[X_n]$ converge, mais elle ne converge pas vers $E[0]$

c) $x_n = (-1)^n n$, $E\left[X_n\right] = (-1)^n$

La suite $E[X_n]$ est une suite oscillante.

d) $x_n = n^2$, $E\left[X_n\right] = n \longrightarrow +\infty$.

La suite $E\left[X_n\right]$ converge, mais elle ne converge pas vers $E[0]$.

Il est aussi intéressant de remarquer que :

$$\forall \omega \in \Omega \qquad \lim_{n \longrightarrow +\infty} X_n(\omega) = a \quad \text{implique} \quad \operatorname*{plim}_{n \longrightarrow +\infty} X_n = a .$$

En effet, l'hypothèse implique :

$$\forall \varepsilon > 0 \ \exists\, n_0(\varepsilon) \text{ tel que } \forall n > n_0, \left\{ \omega : |X_n(\omega) - a| < \varepsilon \right\} = \Omega.$$

Dès lors :

$$\forall n > n_0 \qquad P\left[\omega : |X_n(\omega) - a| \geq \varepsilon \right] = 0$$

Proposition 1. Si (X_n) est une suite de v.a. telle que

$$\lim_{n \longrightarrow +\infty} (\operatorname{var} X_n) = 0$$

alors, la suite $(X_n - E\left[X_n\right])$ converge en probabilité vers 0.

Démonstration :

Par l'inégalité de Chebyshev, on a :

$$\forall \varepsilon > 0 \qquad P\left[|X_n - E\left[X_n\right]| \geq \varepsilon\right] \leq \frac{\operatorname{var} X_n}{\varepsilon^2} \longrightarrow 0. \quad \text{si } n \longrightarrow +\infty$$

Corollaire 1 : Si aux hypothèses de la proposition 1, on ajoute la suivante :

$$\lim_{n \longrightarrow +\infty} E\left[X_n\right] = a .$$

alors, on a :

$$\operatorname*{plim}_{n \longrightarrow +\infty} \left[X_n\right] = a$$

Démonstration :

On a :

$$\left|X_n - a\right| \leq \left|X_n - E\left[X_n\right]\right| + \left|E\left[X_n\right] - a\right|$$

Prenons n assez grand pour avoir $\left|E\left[X_n\right] - a\right| < \frac{\varepsilon}{2}$. Alors :

$$\left\{ \omega : \left|X_n - a\right| \geq \varepsilon \right\} \subseteq \left\{ \omega : \left|X_n - E\left[X_n\right]\right| \geq \frac{\varepsilon}{2} \right\}$$

$$P\left[\left|X_n - a\right| \geq \varepsilon\right] \geq P\left[\left|X_n - E\left[X_n\right]\right| \quad \frac{}{2} \quad 0 \quad \text{si } n \longrightarrow +\infty\right.$$

Application : Loi faible des grands nombres au sens de Bernouilli.

Nous n'aborderons les lois des grands nombres qu'au paragraphe 3. Mais à cause de son caractère élémentaire et de son importance, nous déduirons dès maintenant la loi des grands nombres de Bernoulli.

Soit X une v.a. binomiale de paramètres n et p. Posons :

$$U_n = \frac{X}{n} = \text{v.a. fréquence du succès (Ch. 8 § 2)}$$

Nous avons :

$$E\left[U_n\right] = p \qquad \operatorname{var} U_n = \frac{pq}{n} \leq \frac{1}{4n} \longrightarrow 0 \text{ quand } n \longrightarrow \infty .$$

et dès lors :

$$P\left[|U_n - p| \geq \varepsilon\right] \longrightarrow 0 \text{ quand } n \longrightarrow +\infty .$$

Ainsi, si nous effectuons n expériences identiques indépendantes et si chacune de ces expériences n'a que deux résultats possibles :réaliser l'événement A (succès), ne pas réaliser l'événement A (échec), il y a une probabilité aussi voisine de 1 que l'on veut, que la différence, entre la fréquence de la réalisation de A dans ce schéma d'expériences et la probabilité de A (généralement inconnue), soit aussi petite que l'on veut, pourvu que n soit assez grand.

Proposition 2 : Si (X_n) est une suite de v.a. et si X est une v.a. telle que $\operatorname*{plim}_{n \longrightarrow +\infty} X_n = X$; si g est une application continue de $\mathbb{R}$ dans $\mathbb{R}$, alors :

$$\operatorname*{plim}_{n \longrightarrow +\infty} g\left(X_n\right) = g\left(X\right)$$

Démonstration :

$$\forall \varepsilon > 0 \qquad \exists M = M(\varepsilon) \text{ tel que } P\left[|X| \geq M\right] < \frac{\varepsilon}{2} .$$

Sur l'intervalle fermé $\left[-M, M\right]$, la fonction g est uniformément continue. Il existe donc $\delta(\varepsilon, M)$ tel que, si $|X_n - X| < \delta$, on ait:

$$|g(X_n) - g(X)| < \varepsilon$$

Dès lors, nous avons :

$$\left\{\omega : |g(X_n) - g(X)| < \varepsilon\right\} \supseteq \left\{\omega : |X| \leq M\right\} \cap \left\{\omega : |X_n - X| < \delta\right\}$$

et, en passant aux complémentaires, nous avons :

$$\left\{\omega : |g(X_n) - g(X)| \geq \varepsilon\right\} \subseteq \left\{\omega : |X| > M\right\} \cup \left\{\omega : |X_n - X| \geq \delta\right\}$$

En passant aux probabilités, on voit que

$$P\left\{|g(X_n) - g(X)| \geq \varepsilon\right\} \geq \frac{\varepsilon}{2} + P\left[|X_n - X| \geq \delta\right]$$

ce qui achève la démonstration.

La proposition suivante donne la liaison entre la convergence stochastique de la suite (X_n) de v.a. vers la v.a. X et la convergence de la suite correspondante (F_n) des f.d. des X_n vers la f.d. F de X.

Proposition 3 : Soit (X_n) est une suite de v.a. qui converge en probabilité vers la v.a. X. Si F_n est la f.d. de X_n et si F est la f.d. de X, alors

$$\lim_{n \to +\infty} F_n(x) = F(x)$$

en tout point de continuité x de F.

Démonstration :

On a pour tout $\varepsilon > 0$:

$$\left\{\omega : X_n \leq x\right\} \subseteq \left\{\omega : X \leq x + \varepsilon\right\} \cup \left\{\omega : |X_n - X| \geq \varepsilon\right\}$$

On en déduit :

$$P\left[X_n \leq x\right] \leq P\left[X \leq x + \varepsilon\right] + \delta$$

ou, à l'aide des f.d.,

$$F_n(x) \leq F(x + \varepsilon) + \delta$$

De même, on a :

$$\left\{\omega : X_n > x\right\} \subseteq \left\{\omega : X > x - \varepsilon\right\} \cup \left\{\omega : |X_n - X| > \varepsilon\right\}$$

On en déduit directement

$$1 - F_n(x) \leq 1 - F(x - \varepsilon) + \delta$$

$$F(x - \varepsilon) - \delta \leq F_n(x)$$

Et en combinant les deux relations, on a :

$$F(x - \varepsilon) - \delta \leq F_n(x) \leq F(x + \varepsilon) + \delta$$

Puisque $\varepsilon > 0$ et $\delta > 0$ sont arbitraires, en prenant pour x, $x + \varepsilon$, $x - \varepsilon$ des points de continuité de F et en faisant tendre n vers $+\infty$, on a le résultat.

La proposition 1 et le théorème de continuité donnent :

Corollaire 2. Si $\operatorname{plim}_{n \to +\infty} X_n = X$, si φ_n est la f.c. de X_n et φ la f.c. de X, alors on a :

$$\forall t \in \mathbb{R} \qquad \lim_{n \to +\infty} \varphi_n(t) = \varphi(t)$$

On rencontre souvent dans les applications le cas particulier où $P[X = a] = 1$ et donc $\varphi_X(t) = e^{ita}$. On a dans ce cas :

Corollaire 3 : Si $\operatorname{plim}_{n \to +\infty} X_n = a$, alors $\lim_{n \to +\infty} \varphi(t) = e^{ita}$

Remarque : La réciproque de la proposition 1, n'est pas nécessairement exacte. Ainsi, si (X_n) est une suite de v.a. et même si, pour tout x réel, on a $\lim_{n \to +\infty} F_n(x) = F(x)$, cela n'implique rien sur la convergence en probabilité de la suite (X_n). Pour nous en convaincre, prenons une suite (X_n) de v.a. indépendantes ayant toutes la même f.d. F et soit X une v.a. ayant encore la f.d. F. Il est clair, que pour tout $x \in \mathbb{R}$ on a $\lim_{n \to +\infty} F_n(x) = F(x)$. On n'a cependant pas $\operatorname{plim}_{n \to +\infty} (X_n - X) = 0$ car ceci demande

$$P\left[|X_n - X| \geq \varepsilon\right] \leq \delta$$

ou puisque tous les v.a. ont la même f.d.

$$P\left[|X_1 - X_2| \geq \varepsilon\right] \leq \delta$$

Or on peut démontrer l'inégalité suivante :

Inégalité de symétrie.

Si X_1 et X_2 sont deux v.a. indépendantes ayant la même f.d. Si $a \geq 0$ est choisi de façon telle que :

$$P\left[X_i \geq a\right] \leq 1 - \alpha \quad \text{et} \quad P\left[X_i \leq -a\right] \geq 1 - \alpha \qquad i = 1, 2 ; 0 < \alpha < 1$$

alors on a :

$$P\left[|X_1 - X_2| \geq \varepsilon\right] \geq \alpha P\left[|X_i| > a + \varepsilon\right] \qquad i = 1, 2.$$

Démonstration.

On a

$$\left\{\omega : X_1 > \varepsilon + a \cap X_2 \leq a\right\} \cup \left\{X_1 < -\varepsilon - a \cap X_2 \geq -a\right\} \subseteq \left\{\omega : |X_1 - X_2| \geq \varepsilon\right\}$$

et puisque les v.a. X_1 et X_2 sont indépendantes, on a :

$$P\left[|X_1 - X_2| \geqslant \varepsilon\right] \geqslant P\left[X_1 > \varepsilon + a\right] P\left[X_2 \leqslant a\right] + P\left[X_1 < -\varepsilon - a\right] P\left[X_2 \geqslant -a\right]$$

$$P\left[|X_1 - X_2| \geqslant \varepsilon\right] \geqslant p\ P\left[X_1 > \varepsilon + a\right] + p\ P\left[X_1 < -\varepsilon - a\right]$$

Néanmoins, avec les notations habituelles, on a :

Proposition 4.

Si, $\forall\ x \neq a$, on a : $\lim\limits_{n \to +\infty} F_n(x) = F(x)$

où

$$F(x) = \begin{cases} 1 & \text{pour } x \geqslant a \\ 0 & \text{pour } x < a \end{cases}$$

alors :

$$\operatorname*{plim}_{n \to +\infty} X_n = a.$$

Démonstration :

On a, pour tout $\varepsilon > 0$,

$$P\left[|X_n - a| \geqslant \varepsilon\right] = P\left[X_n \leqslant a - \varepsilon\right] + P\left[X_n \geqslant a + \varepsilon\right]$$

$$= F_n(a - \varepsilon) + 1 - F_n(a + \varepsilon) + P\left[X_n = a + \varepsilon\right]$$

Or, on a :

$$\lim_{n \to +\infty} F_n(a - \varepsilon) = F(a - \varepsilon) = 0$$

$$\lim_{n \to +\infty} (1 - F_n(a + \varepsilon)) = 1 - F(a + \varepsilon) = 0$$

$$\lim_{n \to +\infty} P\left[X_n = a + \varepsilon\right] \leqslant \lim_{n \to +\infty} P\left[X_n > a + \frac{\varepsilon}{2}\right] = \lim_{n \to +\infty} (1 - F_n(a + \frac{\varepsilon}{2})) = 0$$

d'où le résultat.

Remarque : L'hypothèse de la proposition 4 peut encore être formulée à l'aide des f.c. sous la forme :

$$\lim_{n \to +\infty} \varphi_n(t) = e^{ita}$$

Le corollaire 3 et la proposition 4 réunis donnent maintenant:

Proposition 5. Une condition nécessaire et suffisante pour que la suite (X_n) de v.a. converge en probabilité vers la constante a est que la suite correspondante des f.c. converge vers e^{ita}.

§ 2. CONVERGENCE PRESQUE SURE.

Définition 3. On dit qu'une suite (X_n) de v.a. converge presque sûrement vers 0 et on écrit :

$$X_n \xrightarrow{p.s.} 0$$

si :

$$\forall\ \varepsilon > 0 \qquad \lim_{n \to +\infty} P\left[\omega : \sup_{k \geqslant n} |X_k(\omega)| \geqslant \varepsilon\right] = 0$$

Proposition 6 : Si $X_n \xrightarrow{p.s.} 0$, alors $\operatorname*{plim}\limits_{n \to +\infty} X_n = 0$

En effet, on a la relation :

$$\left\{\omega : |X_n| \geqslant \varepsilon\right\} \subseteq \left\{\omega : \sup_{k \geqslant n} |X_k| \geqslant \varepsilon\right\}$$

On peut donner deux expressions équivalentes de la définition 3.

(1) Définissons une suite (Y_n) de v.a. en posant :

$$Y_n = \sup_{k \geqslant n} |X_k|$$

Alors : $X_n \xrightarrow{p.s.} 0$ équivaut à $\operatorname*{plim}\limits_{n \to +\infty} Y_n = 0$

(2) Posons :

$$B_k(\varepsilon) = \left\{\omega : |X_k(\omega)| \geqslant \varepsilon\right\}$$

$$A_n(\varepsilon) = \bigcup_{k \geqslant n} B_k(\varepsilon)$$

La suite $(A_n(\varepsilon))$ et on a :

$$A_n(\varepsilon) = \left\{\omega : Y_n \geqslant \varepsilon\right\} = \left\{\omega : \sup_{k \geqslant n} |X_k(\omega)| \geqslant \varepsilon\right\}$$

Posons

$$A(\varepsilon) = \lim_{n \to +\infty} A_n(\varepsilon) = \bigcap_n A_n(\varepsilon)$$

On a :

$$P[A(\varepsilon)] = P\left[\lim_{n \to +\infty} A_n(\varepsilon)\right] = P\left[\bigcap_n A_n(\varepsilon)\right] = P\left[\bigcap_n \bigcup_{k \geqslant n} B_k(\varepsilon)\right] = P\left[\limsup_{k \to +\infty} B_k(\varepsilon)\right]$$

$$= \lim_{n \to +\infty} P[A_n(\varepsilon)] = \lim_{n \to +\infty} P[Y_n \geqslant \varepsilon] = 0$$

si $X_n \xrightarrow{p.s.} 0$

On peut en déduire :

$$\forall\, \delta > 0 \qquad \exists\, n_o(\varepsilon, \delta) \text{ tel que}$$

$$\forall n > n_o(\varepsilon,\delta) \qquad P\left[A_n(\varepsilon)\right] \leqslant \delta$$

ou

$$P\left[A_n^c(\varepsilon)\right] \geqslant 1 - \delta$$

ou

$$P\left[\left(\bigcup_{k\geqslant n} B_k(\varepsilon)\right)^c\right] = P\left[\bigcap_{k\geqslant n} B_k^c\right] \geqslant 1 - \delta$$

ou

$$P\left[|X_{n_o+1}| < \varepsilon \text{ et } |X_{n_o+2}| < \varepsilon \text{ et } |X_{n_o+3}| < \varepsilon \text{ et } \ldots\right] \geqslant 1 - \delta$$

Cette dernière relation montre bien l'importante différence qui existe entre la convergence en probabilité et la convergence presque sûre. La convergence en probabilité demande que chacun des événements

$$\left\{\omega : |X_{n_o+1}(\omega)| < \varepsilon\right\} ; \left\{\omega : |X_{n_o+2}(\omega)| < \varepsilon\right\}$$

ait une probabilité supérieure à $1-\delta$, mais la convergence presque sûre demande que l'intersection de ces événements ait une probabilité supérieure à $1-\delta$.

Cherchons maintenant la relation qui existe entre le convergence presque sûre de la suite (X_n) de v.a. et la convergence de la suite (X_n) considérée comme une suite de fonctions. On appelle ensemble de convergence vers 0 de la suite (X_n), l'ensemble C défini par

$$C = \left\{\omega : X_n(\omega) \longrightarrow 0\right\}$$

et ensemble de divergence, l'ensemble D défini par

$$D = \left\{\omega : X_n(\omega) \not\longrightarrow 0\right\}$$

On remarque de suite que : $D = \bigcup_{\varepsilon>0} A(\varepsilon)$, mais puisque :

$$\forall\, \varepsilon < \varepsilon \qquad A_n(\varepsilon) \supseteq A_n(\varepsilon')$$

on a :

$$\forall\, \varepsilon < \varepsilon \qquad A(\varepsilon) \supseteq A(\varepsilon')$$

qui montre que $A(\varepsilon)$ croît de façon monotone quand ε tend vers zéro, on peut remplacer l'union non dénombrable par une union dénombrable. On a donc :

$$D = \bigcup_{k\in\mathbb{N}^*} A\left(\frac{1}{k}\right)$$

Avec ces notions, on peut montrer

Proposition 7. $X_n \xrightarrow{p.s.} 0$ si et seulement si $P[D] = 0$

Démonstration :

1) On a : $P[D] \geqslant P[A(\varepsilon)]$ pour tout $\varepsilon > 0$.

Si $P[D] = 0$, alors $P[A(\varepsilon)] = 0$ pour tout $\varepsilon > 0$ et $X_n \xrightarrow{p.s.} 0$

2) La suite $(A(\frac{1}{k})) \uparrow D$ quand $k \longrightarrow +\infty$.

Si $X_n \xrightarrow{p.s.} 0$, alors $P\left[A(\frac{1}{k})\right] = 0$ pour tout $k \in \mathbb{N}^*$ et

$$P[D] = \lim_{k\to+\infty} P\left[A\left(\frac{1}{k}\right)\right] = 0$$

Corollaire 4 : $X_n \xrightarrow{p.s.} 0$ si et seulement si $P\left[\lim_{n\to+\infty} X_n = 0\right] = 1$.

Remarque : En analyse, la convergence presque sûre porte le nom de convergence presque partout. Cette terminologie est justifiée par la proposition 7.

En analyse, la convergence en probabilité porte le nom de convergence en mesure.

Donnons deux critères de convergence presque sûre.

Critère 1 : Si la série $\sum_n P[|X_n| \geqslant \varepsilon]$ est convergente pour tout $\varepsilon > 0$, alors $X_n \xrightarrow{p.s.} 0$.

En effet, avec les notations introduites plus haut, on a :

$$P\left[A(\varepsilon)\right] \leqslant P\left[A_n(\varepsilon)\right] = P\left[\bigcup_{k\geqslant n} B_k(\varepsilon)\right] \leqslant \sum_{k\geqslant n} P\left[B_k(\varepsilon)\right] = \sum_{k\geqslant n} P\left[|X_k| \geqslant \varepsilon\right]$$

Le dernier terme est le reste d'une série convergente, il tend donc vers 0 quand $n \longrightarrow +\infty$.

Critère 2 : Si pour un $\alpha > 0$, $E\left[|X_n|\right]^\alpha$ existe et si la série $\sum_n E\left[|X_n|^\alpha\right]$ converge, alors $X_n \xrightarrow{p.s.} 0$.

On se ramène au critère 1 grâce à l'inégalité de Markov.

$$\forall\, \varepsilon > 0 \qquad P\left[|X_n| \geqslant \varepsilon\right] \leqslant \frac{E\left[|X_n|\right]}{\varepsilon^\alpha}$$

qui est immédiate puisque :

$$E\left[|X_n|^\alpha\right] \geqslant E\left[|X_n|^\alpha I_{B_n(\varepsilon)}\right] \geqslant \varepsilon^\alpha P\left[B_n(\varepsilon)\right] = \varepsilon^\alpha P\left[|X_n| \geqslant \varepsilon.\right]$$

Montrons maintenant par un exemple que les deux notions de convergence introduites ne sont pas équivalentes. Pour ce faire, nous considérons une suite indépendante (X_n) de v.a. caractérisées par :

$$X_n = \begin{cases} 0 & , \quad 1 \\ 1 - \frac{1}{n} & , \quad \frac{1}{n} \end{cases} \qquad n \in \mathbb{N}^*$$

On a $\operatorname{plim}_{n\to+\infty} X_n = 0$, car

$$P\left[\,|X_n| > \varepsilon\,\right] = P\left[X_n = 1\,\right] = \frac{1}{n}$$

Mais :

$$P\left[\sup_{k\geqslant n}|X_k| \geqslant \varepsilon\right] = P\left[A_n(\varepsilon)\,\right] = 1 - P\left[\bigcap_{k\geqslant n} B_k^c\right]$$

et en vertu de l'indépendance de la suite (X_n) on a :

$$P\left[\bigcap_{k\geqslant n} B_k^c\right] = \prod_{k\geqslant n}\left(1 - \frac{1}{k}\right)$$

Comme, par le développement de Taylor, on a :

$$1 - \frac{1}{k} \leqslant e^{-\frac{1}{k}}$$

on en déduit :

$$\prod_{k\geqslant n}\left(1 - \frac{1}{k}\right) \leqslant e^{-\sum_{k\geqslant n}\frac{1}{k}}$$

Comme la série $\sum_{k\geqslant n}\frac{1}{k}$ diverge, on a : $\prod_{k\geqslant n}\left(1 - \frac{1}{k}\right) \longrightarrow 0$ quand $n \longrightarrow +\infty$.

Ainsi, cette suite (X_n) ne converge pas presque sûrement vers 0. Cependant, la sous-suite $(X_{n'})$ où $n' = n^2$ converge presque sûrement vers 0, car maintenant, on a :

$$P\left[\,|X_{n'}| \geqslant \varepsilon\,\right] = \frac{1}{n^2}$$

et la série $\sum \frac{1}{n^2}$ converge . On peut donc appliquer le premier critère de convergence.
Cette dernière partie résulte d'une propriété générale que nous ne démontrerons pas ici.

Définition 4 : On dit que la suite (X_n) converge presque sûrement vers la v.a. X si la suite $(X_n - X)$ converge presque sûrement vers 0.

§ 3. LOI FAIBLE DES GRANDS NOMBRES.

Dans la suite de ce chapitre, on considérera une suite (X_n) de v.a. pour laquelle on définit :

$$S_n = X_1 + X_2 + \ldots + X_n \qquad n \in \mathbb{N}^*$$

$$\overline{X}_n = \frac{S_n}{n} \qquad n \in \mathbb{N}^*$$

Dans la suite de ce chapitre, toutes les suites (X_n) sont définies pour $n \in \mathbb{N}^*$ seulement.

Une loi des grands nombres est la traduction de la convergence vers 0 de la suite $(\overline{X}_n - E[\overline{X}_n])$ suivant l'un des types de convergence introduits. A la convergence en probabilité, correspond

la loi faible des grands nombres. A la convergence presque sûre, correspond la loi forte des grands nombres.

La forme la plus utilisée de la loi faible des grands nombres est la suivante :

Proposition 8 : Si (X_n) est une suite indépendante de v.a. ayant la même f.d. et dont la moyenne $m = E[X_n]$ et la variance $\sigma^2 = \operatorname{var} X_n$ existent, alors :

$$\operatorname*{plim}_{n\to+\infty} \overline{X}_n = m .$$

Démonstration : Sous les hypothèses formulées, nous avons vu que :

$$E[\overline{X}_n] = m \qquad \text{et var } \overline{X}_n = \frac{\sigma^2}{n} \qquad \forall n$$

Par la proposition 1, il en résulte immédiatement la thèse.

Corollaire 5 : (Bernoulli).

Si (X_n) est une suite indépendante d'indicatrices caractérisées par

$$X_n = \begin{cases} 1 & 0 \\ p & q \end{cases} \qquad \forall n$$

on a : $$\operatorname*{plim}_{n\to+\infty} \overline{X}_n = p$$

On retrouve ainsi la résultat de l'application de la proposition 1 car au chapitre VIII on indique qu'on peut interpréter une v.a. Bi (n, p) comme une somme de n indicatrices indépendantes de paramètre p.

Les hypothèses de la proposition 8 sont trop fortes, on peut les réduire comme on le verra dans les trois propositions qui suivent.

Proposition 9. Si (X_n) est une suite de v.a. ayant la même moyenne $m = E[X_n]$, $\forall n$, et la même variance $\sigma^2 = \operatorname{var} X_n$, $\forall n$, et si de plus, pour tout $i \neq j$ $\operatorname{cov}(X_i, X_j) = 0$, alors $\operatorname*{plim}_{n\to+\infty} \overline{X}_n = m$

Démonstration : Les hypothèses suffisent encore pour assurer:

$$E[\overline{X}_n] = m \qquad \operatorname{var} \overline{X}_n = \frac{\sigma^2}{n} \qquad \forall n$$

Il suffit encore d'appliquer la proposition 1 puisque $\operatorname{var} \overline{X}_n \longrightarrow 0$ quand $n \longrightarrow +\infty$.

Proposition 10. (Markov). Soit (X_n) une suite de v.a. ayant toutes une variance et telle que pour tout $i \neq j$, $\operatorname{cov}(X_i, X_j) = 0$. Posons :

$$\alpha_n = E[X_n] \qquad , \qquad \sigma_n^2 = \operatorname{var} X_n$$

$$\bar{\alpha}_n = \frac{1}{n} \sum_{i=1}^{n} \alpha_i \quad , \quad \bar{\beta}_n = \frac{1}{n} \sqrt{\sum_{i=1}^{n} \sigma_i^2}$$

et supposons : $\lim_{n \to +\infty} \bar{\alpha}_n = \alpha \qquad -\infty < \alpha < +\infty$

$$\lim_{n \to +\infty} \bar{\beta}_n = 0$$

Alors, on a : $\operatorname{plim} \bar{X}_n = \alpha$.

Démonstration . On a :

$$E\left[\bar{X}_n - \bar{\alpha}_n\right] = 0 \qquad \forall n$$

$$\operatorname{var}\left[\bar{X}_n - \bar{\alpha}_n\right] = \bar{\beta}_n^2$$

En vertu des hypothèses, on déduit :

$$\operatorname*{plim}_{n \to +\infty} \left[\bar{X}_n - \bar{\alpha}_n\right] = 0$$

ce qui équivaut à :

$$\operatorname*{plim}_{n \to +\infty} \bar{X}_n = \alpha \ .$$

Dans les propositions 9 et 10, on a remplacé les hypothèses d'indépendance et de même distribution des v.a. qui figurent dans la proposition 8 par des hypothèses plus faibles mais on a toujours supposé l'existence de la variance.

Dans les démonstrations, on a en fait utilisé chaque fois l'inégalité de Chebyshev. La proposition 11 est une amélioration qui ne suppose plus l'existence de la variance.

Proposition 11 : Si (X_n) est une suite indépendante de v.a. ayant toutes la même f.d. et si $m = E[X_n]$ existe, alors $\operatorname*{plim}_{n \to +\infty} \bar{X}_n = m$.

Démonstration : On utilise les fonctions caractéristiques.

Il revient à montrer :

$$\lim_{n \to +\infty} \varphi_{\bar{X}_n}(t) = e^{itm}$$

Or, on a en vertu de l'indépendance des X_n et des propriétés des f.c.

$$\varphi_{\bar{X}_n}(t) = \varphi_{\frac{S_n}{n}}(t) = \varphi_X^n\left(\frac{t}{n}\right)$$

Mais puisque m existe, on a :

$$\varphi_X^n\left(\frac{t}{n}\right) = \left(1 + i\,m\,\frac{t}{n} + 0\left(\frac{t^2}{n^2}\right)\right)^n$$

et en prenant le logarithme des deux membres, on a :

$$\operatorname{Log}\ \varphi_X^n\left(\frac{t}{n}\right) = n \operatorname{Log}_e\left[1 + i\,m\,\frac{t}{n} + 0\left(\frac{t^2}{n^2}\right)\right]$$

et par le développement de Taylor de $\operatorname{Log}_e(1+x)$ on a :

$$n \operatorname{Log}_e\left[1 + i\,m\,\frac{t}{n} + 0\left(\frac{t^2}{n^2}\right)\right] \to i\,m\,t.$$

§ 4. LOI FORTE DES GRANDS NOMBRES.

Les problèmes rencontrés dans la loi forte des grands nombres sont plus difficiles que ceux rencontrés dans la loi faible des grands nombres. Nous ne montrerons pas ici, que sous les hypothèses de la proposition 11, on a :

$$\bar{X}_n \xrightarrow{\text{p. s.}} m$$

Nous nous limiterons à la proposition suivante :

Proposition 12 (Kolmogorov).

Soit (X_n) une suite de v.a. telle que, pour tout n, $\sigma_n^2 = \operatorname{var} X_n$ existe et telle que, pour tout $i \neq j$, $\operatorname{cov}(X_i, X_j) = 0$. Posons : $\alpha_n = E[X_n]$ et $\bar{\alpha}_n = \frac{1}{n} \sum_{i=1}^{n} \alpha_i$ et supposons que la série

$$\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} \qquad (1)$$

converge. Alors, on a :

$$\bar{X}_n - \bar{\alpha}_n \xrightarrow{\text{p. s.}} 0$$

Remarque : Dans les applications , on rencontre souvent le cas où les variances sont uniformément bornées. Il en est ainsi pour les indicatrices. Dans ce cas, la condition (1) est satisfaite puisque la série $\sum_{1}^{\infty} \frac{1}{n^2}$ converge.

Corollaire 6 : Si (X_n) est une suite d'indicatrices caractérisées par

$$X_n \begin{cases} 1 & , \quad 0 \\ p_n & , \quad q_n \end{cases} \qquad p_n + q_n = 1.$$

et si, pour tout $i \neq j$, on a $\operatorname{cov}(X_i, X_j) = 0$, alors en posant $\bar{\alpha}_n = \frac{1}{n} \sum_{i=1}^{n} p_i$, on a :

$$\bar{X}_n - \frac{1}{n} \sum_{i=1}^{n} p_i \xrightarrow{\text{p. s.}} 0$$

puisque $\operatorname{var} X_n = p_n q_n \leqslant \frac{1}{4}$.

La condition (1) est également satisfaite lorsque toutes les v.a. de la suite (X_n) ont la même variance. En particulier, on a l'analogue de la proposition 8.

Proposition 13. Si (X_n) est une suite indépendante de v.a. ayant la même f.d. et dont la moyenne $m = E[X_n]$ et la variance $\sigma^2 = \text{var } X_n$ existent, alors :

$$\overline{X}_n \xrightarrow{\text{p.s.}} m$$

En particulier, si (X_n) sont des indicatrices, on a :

Corollaire de la proposition 13 (Loi des grands nombres de Borel).

Si (X_n) est une suite indépendante d'indicatrices caractérisées par :

$$\forall n \qquad X_n = \begin{cases} 1 & 0 \\ & \\ p & q \end{cases} \qquad p + q = 1$$

on a :

$$\overline{X}_n \xrightarrow{\text{p.s.}} p$$

Venons maintenant à la démonstration de la proposition 12. Indiquons d'abord que pour démontrer cette proposition, Kolmogorov a introduit les importantes inégalités qui portent son nom et qui généralisent l'inégalité de Chebyshev. Il n'en reste pas moins que la démonstration de la proposition n'est pas encore triviale.

Par les inégalités de Hajek-Renyi, on peut démontrer immédiatement la proposition 12, mais toute la difficulté est reportée sur la démonstration des inégalités. Grâce à une inégalité élémentaire que j'ai obtenue récemment (cfr. J. PARIS : Une généralisation des inégalités de Bimbaum et Marshall et applications à la loi des grands nombres. Annales de la Société Scientifique de Bruxelles, T 83, II, 1969), on peut maintenant obtenir les inégalités de Hajek-Renyi par des moyens élémentaires.

Inégalité de base.

Soient $Y_1, Y_2, \ldots, Y_n$ n v.a. positives. Posons $Y_o = 0$, $A_o = \Phi$ et définissons :

$$A_1 = \left\{ \omega : Y_1 \geqslant 1 \right\}$$

$$A_k = \left\{ \omega : Y_j < 1, \ j = 1, 2, \ldots, k-1 \ ; \ Y_k \geqslant 1 \right\} \qquad k = 2, 3, \ldots, n.$$

$$A = \bigcup_{k=1}^{n} A_k \quad , \quad B_1 = \Omega$$

$$B_k = \left\{ \omega : Y_j < 1, \ j = 1, 2, \ldots, k-1 \right\} \qquad k = 2, 3, \ldots, n.$$

Alors, on a :

$$P\left[\max_{1 \leqslant j \leqslant n} Y_j \geqslant 1 \right] \leqslant \sum_{k=1}^{n} E\left[(Y_k - Y_{k-1}) \, I\,(B_k) \right]$$

Démonstration.

Puisque la probabilité du premier membre est l'espérance de l'indicatrice de A, il suffit de vérifier que :

$$I_A \leqslant \sum_{k=1}^{n} (Y_k - Y_{k-1}) \, I\,(B_k)$$

puisqu'on obtient le résultat à démontrer en prenant l'espérance des deux membres de cette dernière relation.
Pour faire la démonstration, distinguons deux cas :

a) Pour $k = 1, 2, \ldots, n$, $\omega \notin A_k$, alors $I(A) = 0$ et $I(B_k) = 1$ pour $k = 1, 2, \ldots, n$ et on a $0 \leqslant Y_n$.

b) Soit h le plus petit indice tel que $\omega \in A_h$, alors $I(A) = 1$, $I(B_k) = 1$ pour $k = 1, 2, \ldots, h$ et $I(B_k) = 0$ pour $k = h+1, \ldots, n$. On a ainsi : $1 \leqslant Y_h$.

Remarque : L'événement A_k est réalisé lorsque, parmi les v.a. $Y_1, Y_2, \ldots, Y_n$, la v.a. Y_k est la première qui satisfait $Y_k \geqslant 1$. On remarquera que les A_k sont deux à deux incompatibles mais qu'ils ne forment pas un système contradictoire. Néanmoins, si on prolonge la définition des B_k jusqu'à B_{n+1}, on voit que $A_1, A_2, \ldots A_n, B_{n+1}$ forment un système contradictoire. Enfin, on remarquera que :

$$B_k = \bigcap_{j=1}^{k-1} A_j^c$$

Il importe de remarquer que si $Y_k - Y_{k-1} \geqslant 0$, alors on peut remplacer $E\left[(Y_k - Y_{k-1}) \, I\,(B_k)\right]$ par $E\left[Y_k - Y_{k-1}\right]$ puisque

$$(Y_k - Y_{k-1}) \, I_{B_k} \leqslant Y_k - Y_{k-1}$$

Corollaires : Inégalité de Chebyshev.

Si X est une v.a. et si $m = E[X]$ et $\sigma^2 = \text{var } X$ existent, alors on a :

$$\forall \ a > 0 \qquad P\left[|X-m| \geqslant a \right] = P\left[|X-m|^2 \geqslant a^2 \right] \leqslant \frac{\sigma^2}{a^2}$$

Il suffit de poser $Y_1 = \left(\frac{X-m}{a} \right)^2$, $Y_k = 0$ pour $k = 2, \ldots, n$ dans l'inégalité de base.

Inégalités de Kolmogorov.

Si $X_1, X_2, \ldots, X_n$ sont des v.a. telles que $\alpha_k = E[X_k]$ et $\sigma_k^2 = \text{var } X_k$ existent et si pour tout $i \neq j$, $\text{cov}\,(X_i, X_j) = 0$, alors on a :

$$\forall \varepsilon > 0 \quad P\left[\max_{1 \leqslant k \leqslant n} \left| S_k - E[S_k] \right| \geqslant \varepsilon \right] \leqslant \frac{\sum_{k=1}^{n} \sigma_k^2}{\varepsilon^2} = \frac{\text{var } S_n}{\varepsilon^2}$$

Pour la démonstration, on peut sans restriction aucune, supposer $\alpha_k = 0 \quad k = 1, 2, \ldots, n$. On pose alors

$Y_k = \frac{S_k^2}{\varepsilon^2}$ et on remarque que $Y_k - Y_{k-1} \geqslant 0$

On a alors . $P\left[\max_{1\leqslant k\leqslant n} |S_k| \geqslant \varepsilon\right] = P\left[\max_{1\leqslant k\leqslant n} \frac{S_k^2}{\varepsilon^2} \geqslant 1\right] \leqslant \frac{1}{\varepsilon^2} \sum_{k=1}^{n} E\left[S_k^2 - S_{k-1}^2\right]$

$$= \frac{1}{\varepsilon^2} E\left[S_n^2\right] = \frac{1}{\varepsilon^2} \operatorname{var} S_n$$

Remarque : L'inégalité de Chebyshev donnait seulement

$$\forall \varepsilon \qquad P[|S_n - E[S_n]| \geqslant \varepsilon] \leqslant \frac{\operatorname{var} S_n}{\varepsilon^2}$$

Les inégalités de Kolmogorov donnent la même limitation pour $|S_k - E[S_k]|$, $k = 1, 2, \ldots, n$.

Inégalités de Hajek-Renyi.

Soient (X_n) une suite de v.a. telle que, pour tout n, $\alpha_n = E[X_n]$ existe, $\sigma^2 = \operatorname{var} X_n$ existe et telle que, pour tout $i \neq j$, $\operatorname{cov}(X_i, X_j) = 0$. Soit (C_k) une suite décroissante de réels positifs. Alors on a :

$$P\left[\max_{1\leqslant k\leqslant \ell} C_{k+n} |S_{k+n} - E[S_{k+n}]| \geqslant 1\right] \leqslant C_{n+1}^2 \sum_{j=1}^{n+1} \sigma_j^2 + \sum_{j=n+2}^{n+\ell} C_j^2 \sigma_j^2 .$$

Pour la démonstration, on peut sans restriction supposer $\alpha_n = 0$ pour tout n et poser $Y_k = C_{k+n}^2 S_{k+n}^2$. Alors, on a, par l'inégalité de base :

$$P\left[\max_{1\leqslant k\leqslant \ell} C_{k+n} |S_{k+n}| \geqslant 1\right] = P\left[\max_{1\leqslant k\leqslant \ell} C_{k+n}^2 S_{k+n}^2 \geqslant 1\right] \leqslant \sum_{k=1}^{\ell} E\left[(C_{k+n}^2 S_{k+n}^2 - C_{k+n-1}^2 S_{k+n-1}^2) I(B_k)\right] \leqslant \sum_{k=1}^{\ell} C_{k+n}^2 E\left[S_{k+n}^2 - S_{k+n-1}^2\right]$$

puisque la suite (C_k) est décroissante et $S_{k+n}^2 - S_{k+n-1}^2 \geqslant 0$.

puisque $Y_o = 0$, la dernière somme est encore égale à

$$C_{n+1}^2 E[S_{n+1}^2] + \sum_{k=2}^{\ell} C_{k+n}^2 \sigma_{k+n}^2 = C_{n+1}^2 \sum_{j=1}^{n+1} \sigma_j^2 + \sum_{j=n+2}^{n+\ell} C_j^2 \sigma_j^2 .$$

Démonstration de la proposition 12.

Prenons : $C_k = \frac{1}{\varepsilon k}$, $\varepsilon > 0$, $k \in \mathbb{N}^*$, par les inégalités de Hajek-Renyi, on a :

$$P\left[\max_{1\leqslant k\leqslant \ell} \frac{|S_{k+n} - E[S_{k+n}]|}{k+n} \geqslant \varepsilon\right] = P\left[\max_{1\leqslant k\leqslant \ell} |\bar{X}_{k+n} - \bar{\alpha}_{k+n}| \geqslant \varepsilon\right] \leqslant \frac{1}{\varepsilon^2}\left[\frac{1}{(n+1)^2} \sum_{j=1}^{n+1} \sigma_j^2 + \sum_{j=n+2}^{n+\ell} \frac{\sigma_j^2}{j^2}\right]$$

Si on fait d'abord tendre ℓ vers $+\infty$, le second terme de la dernière relation tend vers

$$\sum_{j=n+2}^{+\infty} \frac{\sigma_j^2}{j^2}$$

qui est le reste de la série (1). Si maintenant, on fait tendre n vers $+\infty$, alors les deux termes de la dernière relation tendent vers 0.

§ 5. APPLICATION DE LA LOI DES GRANDS NOMBRES A L'ANALYSE.

Nous allons donner une démonstration élémentaire du fameux théorème d'approximation de Weierstrass. Nous aurons besoin du lemme suivant :

Lemme. Soit (X_n) une suite de v.a. telle que, pour tout n, $\alpha = E[X_n]$ et $\operatorname{var} X_n = \sigma_n^2(\alpha)$. Soit g une application continue et bornée de $\mathbb{R}$ dans $\mathbb{R}$. Si $\lim_{n\to+\infty} \sigma_n^2(\alpha) = 0$, alors on a :

$$\lim_{n\to+\infty} E[g(X_n)] = g(\sigma)$$

et la convergence est uniforme en α lorsque σ_n^2 tend vers 0 uniformément et g est uniformément continue.

Remarque : En vertu des propositions 1 et 2, on a plim $g(X_n) = g(\alpha)$ mais on sait que cela n'implique rien sur la convergence des moments. Seulement, nous n'avons pas utilisé toutes les hypothèses.

Démonstration : Nous avons :

$$|E[g(X_n)] - g(\alpha)| \leqslant E[|g(X_n) - g(\alpha)|]$$

Pour tout $\varepsilon > 0$, $\exists \delta > 0$ tel que

$$|X_n - \alpha| < \delta \quad \text{entraîne} \quad |g(X_n) - g(\alpha)| \leqslant \varepsilon$$

en vertu de la continuité de g. Mais puisque g est bornée, on a aussi :

$$| X_n - \alpha | \geqslant \delta \quad \text{entraîne} \quad | g(X_n) - g(\alpha) | \leqslant 2M$$

Définissons la v.a. Y par

$$Y = \begin{cases} \varepsilon & \text{si} \quad | X_n - \alpha | < \delta \\ 2M & \text{si} \quad | X_n - \alpha | \geqslant \delta \end{cases}$$

Alors, on a :

$$E[|g(X_n) - g(\alpha)|] \leqslant E[Y] = \varepsilon P[|X_n - \alpha| < \delta] + 2M P[|X_n - \alpha| \geqslant \delta]$$

$$\leqslant \varepsilon + 2M \frac{\sigma_n^2(\alpha)}{\delta^2}$$

par l'inégalité de Chebyshev.

Application : Prenons, pour X une v.a. Bi (n, α) et posons : $X_n = \frac{X}{n}$. Alors on a :

$$E[g(X_n)] = \sum_{k=0}^{n} C_n^k \alpha^k (1-\alpha)^{n-k} g\left(\frac{k}{n}\right)$$

Le second membre de cette relation est le polynôme de Bernstein $B_{n,g}(\alpha)$ pour la fonction g et de degré n. Appliqué à ce cas, le lemme donne :

$$\forall \varepsilon > 0 \qquad | B_{n,g}(\alpha) - g(\alpha) | < \varepsilon \qquad 0 \leqslant \alpha \leqslant 1 .$$

On peut donc énoncer :

Théorème d'approximation de Weierstrass.

Si g est une fonction continue sur l'intervalle fermé $[0,1]$, alors les polynômes de Bernstein de g convergent uniformément vers g.

Cette démonstration est due à W. Feller.

§ 6. THEOREME CENTRAL LIMITE.

Ce théorème est un des résultats fondamentaux de la théorie des probabilités. Il a suscité de nombreux travaux et est à la base de nombreuses méthodes de statistique.

La lecture de ce paragraphe suppose comme le paragraphe 8 du chapitre VIII traitant des v.a. normales.

La loi des grands nombres indique que, sous certaines conditions, on peut avoir :

$$\forall \varepsilon > 0 \qquad P[|\overline{X}_n - m| \geqslant \varepsilon] \to 0 \quad \text{quand } n \to +\infty .$$

mais elle ne donne pas un moyen pour évaluer facilement la probabilité $P[|\overline{X}_n - m| \geqslant \varepsilon]$ et voir ainsi rapidement si elle tend vers 0. Le théorème central limite remédie à cet état de choses.

Nous ne donnerons du résultat que sa version la plus simple.

Théorème central limite.

Soit (X_n) une suite indépendante de v.a. ayant la même f.d. et telles que $m = E[X_n]$ et $\sigma^2 = \text{var}[X_n]$ existent, alors on a:

$$\lim_{n \to +\infty} P\left[\frac{S_n - mn}{\sigma \sqrt{n}} \leqslant x\right] = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz.$$

Démonstration: Remarquons d'abord que la v.a. $\frac{S_n - mn}{\sigma\sqrt{n}}$ est réduite. Sans restriction, on peut donc supposer dès le départ $m = 0, \sigma^2 = 1$.

Le théorème veut montrer qu'il y a convergence de la f.d. de la v.a. $\frac{S_n - mn}{\sigma \sqrt{n}}$ vers la f.d. de la v.a. $N(0,1)$. En vertu du théorème de continuité, il suffit de montrer qu'il y a convergence des f.c. correspondantes. Il suffit donc de montrer :

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) \to e^{-\frac{t^2}{2}} \quad \text{quand } n \to +\infty.$$

En vertu des hypothèses , on a :

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) = \varphi_X^n\left(\frac{t}{\sqrt{n}}\right) = \left(1 - \frac{t^2}{2n} + 0\left(\frac{t^2}{n}\right)\right)^n$$

En prenant le logarithme des deux membres et en utilisant le développement de Taylor de $\text{Log}_e(1+x)$ au voisinage de $x = 0$, on a :

$$\text{Log}_e \varphi_{\frac{S_n}{\sqrt{n}}}(t) = n \, \text{Log}_e \varphi_X\left(\frac{t}{\sqrt{n}}\right) \to -\frac{t^2}{2} \quad \text{quand } n \to +\infty.$$

Remarque : Le résultat exprime que, sous les hypothèses indiquées, la v.a. $\frac{S_n - mn}{\sigma\sqrt{n}}$ est approximativement $N(0,1)$. Il en est de même de la v.a. $\frac{\overline{X}_n - m}{\frac{\sigma}{\sqrt{n}}}$.

Les hypothèses formulées sont celles de la version la moins générale de la loi faible des grands nombres (proposition 8). La démonstration qui vient d'être faite rend également ce résultat. On a en effet :

$$\forall \varepsilon > 0 \qquad P[|\bar{X}_n - m| \leqslant \varepsilon] = P\left[\left|\frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}}\right| \leqslant \frac{\varepsilon\sqrt{n}}{\sigma}\right] \to \int_{\frac{-\varepsilon\sqrt{n}}{\sigma}}^{\frac{\varepsilon\sqrt{n}}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

et l'intégrale tend vers 1 quand $n \to +\infty$. Mais le théorème central limite apporte un complément fondamental.

Il importe également de remarquer combien le fait de réduire la v.a. est important. En effet, sous les hypothèses du théorème central limite, c'est la suite $(\bar{X}_n)$ qui suit la loi des grands nombres et pas la suite (S_n). On a, en effet :

$$P\left[\left|S_n - n\,m\right| \geqslant \varepsilon\right] \to 1 - \int_{-\frac{\varepsilon}{\sigma\sqrt{n}}}^{\frac{\varepsilon}{\sigma\sqrt{n}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \to 1 \text{ quand } n \to +\infty.$$

Cependant, la suite $(a_n S_n)$ où $\lim_{n\to+\infty} \sqrt{n}\, a_n = 0$, suit la loi des grands nombres car :

$$P\left[\left|a_n S_n - a_n\, n\, m\right| \geqslant \varepsilon\right] \to 1 - \int_{-\frac{\varepsilon}{\sigma |a_n| \sqrt{n}}}^{\frac{\varepsilon}{\sigma |a_n| \sqrt{n}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz .$$

Application: On veut évaluer une grandeur physique inconnue m. Pour ce faire, on fait n mesures indépendantes aussi précises que possible de cette grandeur. Soient $X_1, X_2, \ldots, X_n$ ces mesures. On suppose que ce sont des v.a. indépendantes ayant la même f.d. et telles que

$$m = E[X_i] \qquad i = 1, 2, \ldots, n .$$

Un problème intéressant est le suivant : combien faut-il effectuer de mesures pour avoir, avec une probabilité donnée, un écart $|\bar{X}_n - m|$ inférieur à une quantité donnée.
Proposons-nous de trouver n tel que :

$$P[|\bar{X}_n - m| \leqslant \frac{m}{100}] \geqslant 0,95$$

Ceci suppose évidemment $m > 0$. On adapte facilement l'écriture si $m \leqslant 0$. Par l'inégalité de Chebyshev , on a :

$$P[|\bar{X}_n - m| \leqslant \frac{m}{100}] \geqslant 1 - \frac{\sigma^2}{n} \frac{10.000}{m^2} \geqslant 0,95$$

On trouve donc n par la relation :

$$\frac{\sigma^2}{n} \frac{10.000}{m^2} \leqslant \frac{5}{100}$$

ou encore

$$n \geqslant 200.000 \left(\frac{\sigma}{m}\right)^2$$

Le nombre n dépend du coefficient de variation $\frac{\sigma}{m}$. Plus σ est pet moins il faudra faire de mesures.
En appliquant le théorème central limite, on a :

$$P[|\bar{X}_n - m| \leqslant \frac{m}{100}] \simeq \int_{-\frac{m\sqrt{n}}{100\,\sigma}}^{\frac{m\sqrt{n}}{100\,\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \geqslant 0,95$$

La table de la loi normale N(0,1) donne

$$\frac{m\sqrt{n}}{100} = 1,96 \simeq 2$$

et donc

$$n \simeq 40.000 \left(\frac{\sigma}{m}\right)^2$$

Cette relation montre que le nombre n fourni par l'inégalité de Chebyshev est trop grand.
Cas particulier du théorème central limite :

Théorème de De Moivre-Laplace.

Si (X_n) est une suite indépendante d'indicatrices de paramètre p, $0 < p < 1$, alors S_n est une v.a. Bi (n, p) et

$$P[S_n \leqslant x] = P\left[\frac{S_n - n\,p}{\sqrt{n\,p\,q}} \leqslant \frac{x - n\,p}{\sqrt{n\,p\,q}}\right] \to \int_{-\infty}^{\frac{x - np}{\sqrt{n\,p\,q}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

Ce théorème affirme qu'une approximation de la valeur de la f.d. d'une v.a. Bi (n, p) au point x est fournie par la valeur de la f.d. de la v.a. N(0,1) au point $\frac{x - np}{\sqrt{n\,p\,q}}$. Puisqu'il s'agit de l'approximation de la f.d. d'une v.a. discrète par la f.d. d'une v.a. continue, on peut montrer par des calculs assez laborieux qu'une meilleure approximation de $P[S_n \leqslant x]$ est fournie par :

$$\int_{-\infty}^{\frac{x + \frac{1}{2} - np}{\sqrt{n\,p\,q}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz.$$

Il n'est plus intéressant aujourd'hui de développer ces calculs. En effet, si n est grand, le terme $\frac{1}{2}$ ne joue aucun rôle et si n est petit, on dispose de tables qui donnent la valeur exacte.
(Havard University Computation Laboratory : "Tables of the Cumulative Binomial Probability Distribution " , Harvard University Press, Cambridge, Massachusetts 1955).
On peut également montrer que si X est Bi(n, p), on a :

$$P[X = x] \simeq \frac{1}{\sqrt{npq}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-np}{\sqrt{npq}}\right)^2}$$

où $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-np}{\sqrt{npq}}\right)^2}$ est la valeur au point $\frac{x-np}{\sqrt{npq}}$ de la f.fr. d'une v.a. N(0,1). Cette quantité a été tabulée.

Application : Proposons-nous de trouver n tel que :

$$P[|\bar{X}_n - p| \leqslant 0,01] \geqslant 0,95$$

Par l'inégalité de Chebyshev, on trouve :

$$P[|\bar{X}_n - p| \leqslant 0,01] \geqslant 1 - \frac{pq}{n(0,01)^2}$$

Dès lors, on trouve :

$$n \geqslant \frac{pq}{(0,01)^2} \quad \frac{100}{5}$$

ou encore

$$n \geqslant \frac{5}{(0,01)^2} = 50.000$$

Par le théorème central limite, on trouve :

$$P[|\bar{X}_n - m| \leqslant \varepsilon] \simeq \int_{-\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}}^{\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 0,95$$

ce qui, dans le cas présent, donne :

$$\frac{0,01\sqrt{n}}{\sqrt{pq}} \simeq 2$$

$$\sqrt{n} \simeq 2 \times \frac{1}{2} \times \frac{1}{0,01}$$

$$n \simeq 10.000 .$$

CHAPITRE VIII. PRINCIPALES VARIABLES ALEATOIRES

§ 1. INDICATRICE D'UN EVENEMENT.

A tout événement $A \in \mathcal{A}$, on peut associer une v.a. discrète prenant la valeur 1 quand A est réalisé et la valeur 0 quand A n'est pas réalisé. Cette v.a. appelée indicatrice de A et notée I_A ou I(A) est caractérisée par :

$$I_A \begin{cases} 1 & 0 \\ p = P[A] & P[A^c] = q = 1-p \end{cases}$$

On a :

$$E[I_A] = 1 \times p = p = P[A]$$

$$\text{var } I_A = pq$$

$$\varphi_{I_A}(t) = q + p e^{it}$$

§ 2. SCHEMA DE BERNOULLI ET V.A. BINOMIALE.

On considère un schéma d'expériences répondant aux hypothèses :

1) on effectue un nombre fixé n d'expériences.

2) chaque expérience se termine de deux façons seulement l'une qu'on appelle conventionnellement "succès", l'autre qu'on appelle conventionnellement "échec".

3) la probabilité p du succès est la même à chaque expérience

4) les expériences sont indépendantes (au sens du chapitre IV)

Un tel schéma d'expériences s'appelle un schéma de Bernoulli. On a une représentation d'un tel schéma quand on extrait des boules d'une urne dont la composition ne varie pas et qui au départ ne contient que deux espèces de boules.

L'ensemble Ω des résultats possibles comporte 2^n points.

Soit X la v.a. que représente le nombre de succès obtenus dans ce schéma. Cette v.a. peut prendre les valeurs 0, 1, 2, ... n. Cherchons $P[X = j]$, $j = 0, 1, \ldots, n$.

Pour obtenir j succès et donc n-j échecs, on peut obtenir d'abord les j succès et ensuite les n-j échecs. Désignons par S_j l'événement qui consiste à obtenir un succès à la $j^{ième}$ expérience

et par E_j l'événement qui consiste à obtenir un échec à la $j^{ième}$ expérience. ($j = 1, 2, \ldots, n$). En vertu des hypothèses, on trouve:

$$P[S_1 \cap S_2 \cap \ldots \cap S_j \cap E_{j+1} \cap \ldots \cap E_n] = p^j q^{n-j}$$

Mais il est immédiat que la séquence obtenue en permutant un S et un E, sans déplacer les indices, fournit le même nombre de succès et a la même probabilité. Or, on obtient toutes les écritures possibles par des permutations avec répétition de n éléments dont j sont du type S et n-j du type E. Il y a donc $\frac{n!}{j!\,n-j!} = C_n^j$ séquences possibles représentant des événements deux à deux incompatibles mais ayant la même probabilité. Ainsi, la probabilité d'obtenir exactement j succès est

$$P[X = j] = C_n^j\, p^j\, q^{n-j} \qquad j = 0, 1, 2, \ldots n .$$

On appelle binomiale de paramètre (n, p) (en abrégé Bi(n, p)) la v.a. discrète caractérisée par :

$$X \begin{cases} 0 & , \quad 1, \ldots , \quad j , & n \\ q^n & , C_n^j\, p^j\, q^{n-j}, & , p^n \end{cases}$$

Il est immédiat que $P[X=j] \geqslant 0$ et

$$\sum_{j=o}^{n} P[X=j] = \sum_{j=o}^{n} C_n^j\, p^j\, q^{n-j} = (p+q)^n = 1.$$

Pour trouver la moyenne et la variance de X, remarquons qu'on peut écrire :

$$X = Y_1 + Y_2 + \ldots + Y_n$$

où Y_j est l'indicatrice de l'expérience j caractérisée par

$$Y_j \quad \begin{cases} 1 & 0 \\ p & q \end{cases} \qquad j = 1, 2, \ldots n$$

et où les v.a. $Y_1, Y_2, \ldots, Y_n$ sont indépendantes.
On en déduit :

$$E[X] = n\,p \qquad \text{var } X = npq$$

$$\varphi_X(t) = (q + p\, e^{it})^n$$

Remarque : Il existe des tables qui pour diverses valeurs de n et p fournissent $P[X = j]$ ou $P[X \leqslant j]$. On a :

$$P[X \leqslant j] = \sum_{i=o}^{j} P[X=i]$$

$$P[X = j] = P[X \leqslant j] - P[X \leqslant j-1]$$

Les tables sont limitées à $0 < p \leqslant 0,5$ car si X est Bi (n, p), alors la v.a. qui représente le nombre d'échecs est Bi (n, q).

A la v.a. X Bi (n, p), on associe la v.a. discrète U appelée la fréquence du succès et définie par :

$$U = \frac{X}{n}$$

est caractérisée par $\begin{cases} 0, \frac{1}{n}, \ldots \frac{j}{n}, \ldots 1 \\ C_n^j p^j q^{n-j} \end{cases}$

car $$P\left[\frac{X}{n} = \frac{j}{n}\right] = P[X = j]$$

De la définition de U, on tire :

$$E[U] = p \qquad \text{var } U = \frac{pq}{n}$$

$$\varphi_U(t) = \left(q + p\, e^{i\frac{t}{n}}\right)^n$$

On peut considérer un schéma plus général où l'hypothèse 3 est remplacée par l'hypothèse :

3') la probabilité du succès à l'expérience j est p_j et celle de l'échec est q_j.

Il devient laborieux d'évaluer $P[X = j]$ dès que n est grand. Ainsi, pour n = 3, on a déjà :

$$P[X=1] = p_1\, q_2\, q_3 + q_1\, p_2\, q_3 + q_1\, q_2\, p_3 .$$

Cependant on a encore la représentation

$$X = Y_1 + Y_2 + \ldots + Y_n$$

où Y_j est caractérisée par $\begin{cases} 1 & 0 \\ p_j & q_j \end{cases}$

On en déduit encore :

$$E[X] = \sum_{j=1}^{n} p_j \quad ; \quad \text{var } X = \sum_{j=1}^{n} p_j\, q_j$$

$$\varphi_X(t) = \prod_{j=1}^{n} (q_j + e^{it}\, p_j)$$

§ 3. VARIABLE ALEATOIRE GEOMETRIQUE.

On considère un schéma répondant aux hypothèses 2, 3 et 4 du schéma de Bernoulli. On s'intéresse à la v.a. U qui représente le nombre d'échecs précédant le premier succès. D'après les hypothèses, cette v.a. est du même type que celle qui représente le nombre d'échecs séparant deux succès. La v.a. U est appelée v.a. géométrique de paramètre p. Elle peut prendre toutes les valeurs de N et

$$P[U = k] = P[E_1 \cap E_2 \cap \ldots \cap E_k \cap S_{k+1}] = q^k p.$$

La v.a. géométrique de paramètre p est caractérisée par

$$U \begin{cases} 0, & 1, & 2, \ldots, & k, & \ldots \\ & & & q^k p & \end{cases} \qquad 0 < p < 1.$$

On a :

$$\sum_{k=o}^{+\infty} q^k p = \frac{p}{1-q} = 1.$$

$$\varphi_U(t) = \sum_{k=o}^{\infty} q^k p\, e^{itk} = \frac{p}{1 - q e^{it}}$$

d'où on déduit :

$$E[X] = \frac{q}{p} \qquad \text{var } X = \frac{q}{p^2}$$

§ 4. VARIABLE ALEATOIRE BINOMIALE NEGATIVE.

On effectue des expériences répondant aux hypothèses 2, 3, 4 du schéma de Bernoulli et on exige la condition :

1') obtenir un nombre fixé c de succès.

La v.a. qui représente le nombre d'expériences à faire pour obtenir c succès est appelée v.a. binomiale négative de paramètres (c, p) (notée en abrégé : Bin (c, p)). Une v.a. V Bin (c, p) peut prendre les valeurs c, c+1, c+2, ... $c \in N$
Cherchons P [V=j] . L'événement

$$S_1 \cap S_2 \cap \ldots \cap S_{c-1} \cap E_c \cap \qquad \cap E_{j-1} \cap S_j$$

donne exactement c succès. Il faut remarquer que pour avoir exactement c succès, la séquence doit se terminer par un succès. (Il est entendu que lorsqu'on a obtenu les c succès, on arrête les expériences). Une séquence obtenue en permutant un S (autre que S_j) avec un E, sans déplacer les indices, fournit le même nombre de succès et a la même probabilité. On obtient toutes les écritures possibles par des permutations avec répétition de j-1 éléments dont c-1 sont du type S et j-c du type E. On trouve donc

$$P[V=j] = C_{j-1}^{c-1}\, p^c\, q^{j-c}$$

La v.a. V Bin (c, p) est caractérisée par

$$V \begin{cases} c, & c+1 & & , & j, & \ldots \\ & & C_{j-1}^{c-1}\, p^c\, q^{j-c} & & & \end{cases} \qquad 0<p<1$$

On a :

$$\sum_{j=c}^{+\infty} C_{j-1}^{c-1}\, p^c\, q^{j-c} = \sum_{i=o}^{+\infty} C_{c+i-1}^{c-1}\, p^c\, q^i = \left(\frac{p}{1-q}\right)^c = 1.$$

car

$$(1-q)^{-c} = \sum_{i=o}^{+\infty} C_{c+i-1}^{c-1}\, q^i$$

$$\varphi_V(t) = \sum_{j=c}^{+\infty} C_{j-1}^{c-1}\, p^c\, q^{j-c}\, e^{it(j-c)} = \left(\frac{p}{1-q\,e^{it}}\right)^c$$

Par dérivation, on trouve E V et var V. Mais on peut interpréter V comme étant :

$$V = U_1 + U_2 + \ldots + U_c$$

où les $U_1, U_2, \ldots, U_n$ sont des v.a. indépendantes et géométriques de paramètre p. On en déduit :

$$E(V) = \frac{c\,q}{p} \qquad \text{var}(V) = \frac{c\,q}{p^2} .$$

Plus généralement, on définit une v.a. Bin (c, p), sans supposer que c est un entier, en la caractérisant par :

$$V \begin{cases} 0, & 1, & 2, \ldots, & i, & \ldots \\ & & & C_{c+i-1}^{c-1}\, p^c\, q^i & \end{cases} \qquad \begin{matrix} 0<c \\ 0<p<1 \end{matrix}$$

§ 5. VARIABLE ALEATOIRE DE POISSON.

On rencontre une v.a. de Poisson à partir d'un schéma de type suivant. On considère un événement qui peut se réaliser au moins une fois dans l'intervalle de temps]0, T]. Le nombre de fois que cet événement est réalisé dans le sous-intervalle $]t_1, t_2]$ $(0 < t_1 < t_2 \leqslant T)$ est une v.a. On suppose :

1) Les v.a. associées à deux sous-intervalles disjoints sont indépendantes.

2) Les f.d. associées à deux v.a. correspondant à des intervalles de même longueur h sont identiques à la f.d. de la v.a. correspondant à l'intervalle]0, h] .

3) Pour tout intervalle de longueur non nulle, il y a une probabilité strictement positive qu'il se produise au moins un événement mais il n'est pas certain qu'il s'en produira un .

4) Quand h tend vers 0, la probabilité qu'il se produise au moins deux événements dans l'intervalle $]0,h]$ est un infiniment petit rapport à la probabilité qu'il s'en produit exactement un. On peut interpréter cette condition sous la forme : dans tout intervalle suffisamment petit, il ne peut se produire qu'au plus un événement ce qui revient encore à dire qu'il est impossible pour deux événements de se produire simultanément.

5) L'origine n'est pas un instant où un événement s'est produit. Sous ces conditions, on peut montrer que la probabilité pour que le nombre N(t) d'événements qui se produisent dans l'intervalle $]0,t]$ $(0 < t \leqslant T)$ soit égal à k est :

$$P(k,t) = P[N(t) = k] = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \qquad \lambda > 0, \quad k = 0,1,2,\ldots$$

En particulier, on a : $P(0,t) = e^{-\lambda t}$ et $\lambda = -\operatorname{Log} P(0,1)$
On en déduit :

$$P(k,t) = (-1)^k \frac{t^k}{k!} \frac{\partial^k}{\partial t^k} P(0,t)$$

On appelle v.a. de Poisson de paramètre $\lambda > 0$, la v.a. caractérisée par :

$$X \begin{cases} 0, 1, \ldots, k, \ldots \\ e^{-\lambda}, \ldots, e^{-\lambda} \frac{\lambda^k}{k!}, \end{cases}$$

On a : $$\sum_{k=o}^{+\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=o}^{+\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1 .$$

$$\varphi_X(t) = e^{-\lambda} \sum_{k=o}^{\infty} \frac{\lambda^k e^{itk}}{k!} = e^{-\lambda} e^{\lambda e^{it}} = e^{-\lambda(1-e^{it})}$$

et par dérivation de φ_X on trouve

$$E[X] = \lambda \qquad \text{var } X = \lambda .$$

On peut arriver à la v.a. de Poisson de la manière suivante : Dans la matière qui sert à la fabrication d'un produit se trouvent des impuretés. Lorsqu'une impureté se trouve dans un produit, celui-ci est mis au rebut. Supposons qu'il y ait α impuretés dans la quantité de matière nécessaire pour fabriquer 100 produits. A première vue, on est tenté de penser que le pourcentage des rebuts est α. Mais comme un même produit peut contenir plusieurs impuretés, ce pourcentage est généralement inférieur à α

Appelons "succès" le fait pour une impureté quelconque de se trouver dans un produit et supposons que chaque impureté a la même probabilité de figurer dans la matière qui sert à la fabrication de chaque produit. Ainsi, si on fabrique N produits, la probabilité du "succès" est $\frac{1}{N}$. Supposons encore avoir la condition

d'indépendance du schéma de Bernoulli. Alors, s'il y a n impuretés réparties dans les N produits, la probabilité qu'un produit contienne k impuretés est :

$$\pi_k = C_n^k \left(\frac{1}{N}\right)^k \left(1 - \frac{1}{N}\right)^{n-k}$$

Plaçons dans le cas d'une fabrication de longue durée et posons $\frac{\alpha}{100} = \lambda$. On a aussi $\frac{n}{N} = \lambda$ et

$$\pi_k = \frac{n(n-1)\ldots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k} \prod_{j=1}^{k-1} \left(1 - \frac{j}{n}\right).$$

Dans le cas d'une fabrication de longue série, $n \to +\infty$ et

$$\lim_{n \to +\infty} \pi_k = \frac{\lambda^k}{k!} e^{-\lambda} = P_k$$

car $$\lim_{n \to +\infty} \left(1 - \frac{\lambda}{n}\right)^{n-k} = e^{-\lambda}$$

et $$\lim_{n \to +\infty} \prod_{j=1}^{k-1} \left(1 - \frac{j}{n}\right) = 1 .$$

La probabilité pour qu'un produit contienne au moins une impureté est $1 - P_o = 1 - e^{-\lambda}$. Puisque $\lambda = \frac{\alpha}{100}$, le nombre de rebuts par 100 produits est :

$$100 \left(1 - e^{-\frac{\alpha}{100}}\right)$$

Si α est petit, alors $100 \left(1 - e^{-\frac{\alpha}{100}}\right) \simeq \alpha\,\%$.

Si $\alpha = 20$, alors $100 \left(1 - e^{-0,2}\right) \simeq 18,127\ \%$.

Une question se pose: quand on a une matière qui contient beaucoup d'impuretés vaut-il mieux fabriquer des "gros" produits ou des "petits" produits.

Ainsi, si on fabrique des produits qui demandent 5 fois moins de matière, λ sera remplacé par $\lambda' = \frac{\lambda}{5}$ puisque N est remplacé par 5 N. Le nombre de rebuts par 100^5 produits devient:

$$100 \left(1 - e^{-\frac{\alpha}{500}}\right)$$

et si $\alpha = 20$, on trouve seulement 4% environ de rebuts. Il est donc plus intéressant de fabriquer de petits produits.

Cette présentation montre en plus que si p dépend de n dans la loi binomiale et si $\lim_{n \to +\infty} n\,p(n) = \lambda$, alors

$$\lim_{n \to +\infty} C_n^k \; p^k \; q^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}$$

On peut le démontrer en utilisant le théorème de continuité et les f.c. On a en effet :

$$\mathrm{Log}_e \, (q + p\, e^{it})^n = n \, \mathrm{Log}_e \, (1 - p\,(1 - e^{it})) \longrightarrow -\lambda\,(1 - e^{it})$$

Remarque : L'approximation d'une loi binomiale par une loi de Poisson est bonne quand p est voisin de 0 et quand n n'est pas très grand. Si n est grand ou si p est voisin de $\frac{1}{2}$, l'approximation par le théorème central limite est meilleure.

§ 6. VARIABLE ALEATOIRE HYPERGEOMETRIQUE.

Une urne contient N boules dont N_1 sont blanches et N_2 sont noires. On extrait successivement n boules sans jamais remettre la boule déjà extraite dans l'urne. On s'intéresse à la v.a. qui représente le nombre de boules blanches qui figurent parmi les n boules. Pour traiter le problème, nous supposerons que le procédé d'extraction est aléatoire. Ceci conduit à affecter chacune des C_N^n combinaisons possibles de la probabilité $\frac{1}{C_N^n}$.

Remarquons qu'on arrive au même résultat en supposant qu'à chaque tirage la probabilité conditionnelle de tirer l'une quelconque des boules restantes est la même pour toutes les boules. En effet, la probabilité d'obtenir un arrangement donné, pris parmi les A_N^n arrangements possibles, est

$$\frac{1}{N} \cdot \frac{1}{N-1} \cdots \frac{1}{N-n+1} = \frac{1}{A_N^n}$$

La probabilité d'avoir une combinaison donnée est alors $\frac{n!}{A_N^n} = \frac{1}{C_N^n}$

On appelle v.a. hypergéométrique de paramètre (N, N_1, n) la v.a. discrète représentant le nombre de boules blanches figurant parmi les n boules et caractérisée par :

$$X \quad \begin{cases} 0 \;,\; 1 \;,\ldots,\; k,\; \ldots \;,\; n \\ \dfrac{C_{N_1}^k \; C_{N_2}^{n-k}}{C_N^n} \end{cases} \qquad C_{N_i}^j = 0 \text{ si } j > N_i \quad i = 1, 2$$

On vérifie que :

$$\sum_{k=0}^{n} \frac{C_{N_1}^k \; C_{N_2}^{n-k}}{C_N^n} = 1$$

Pour chercher la moyenne et la variance de cette v.a. remarquons encore que

$$X = Y_1 + Y_2 + \ldots + Y_n$$

où Y_j prend la valeur 1 si la $j^{ième}$ boule extraite est blanche et la valeur 0 si elle est noire. On a :

$$P\left[Y_j = 1\right] = \frac{N_1}{N} = p \qquad P\left[Y_j = 0\right] = \frac{N_2}{N} = q \quad j = 1, 2, \ldots, n.$$

$$E\left[Y_j\right] = p \;, \quad \mathrm{var}\; Y_j = pq$$

Seulement, les v.a. $Y_1, Y_2, \ldots, Y_n$ ne sont plus indépendantes.

$i \neq j \quad \mathrm{cov}\; (Y_i, Y_j) = E[Y_i Y_j] - E[Y_i]\, E[Y_j]$

et la v.a. discrète $Y_i Y_j$ est caractérisée par

$$Y_i Y_j \begin{cases} 1 & 0 \\ \dfrac{N_1}{N} \; \dfrac{N_1 - 1}{N-1} & 1 - \dfrac{N_1}{N} \; \dfrac{N_1 - 1}{N-1} \end{cases}$$

$$E\left[Y_i Y_j\right] = \frac{N_1}{N} \; \frac{N_1 - 1}{N-1}$$

$$\mathrm{cov}\; (Y_i, Y_j) = \frac{N_1}{N} \; \frac{N_1 - 1}{N-1} - \left(\frac{N_1}{N}\right)^2 = -\frac{N_1 \, N_2}{N^2 (N-1)}$$

$$\mathrm{var}\; X = npq + n(n-1)\left(-\frac{pq}{N-1}\right)$$

$$= npq\left[1 - \frac{n-1}{N-1}\right] = npq\,\frac{N-n}{N-1}$$

Remarquons que : si $n = 1$, on a : $\mathrm{var}\; X = npq$

si $n = N$, on a : $\mathrm{var}\; X = 0$

si $1 < n < N$, on a : $npq\,\frac{N-n}{N-1} < npq$

Chacun de ces résultats s'interprète aisément.

Ecrivons

$$\frac{C_{N_1}^n \; C_{N_2}^{n-k}}{C_N^n} = C_n^k \; \frac{N_1}{N} \; \frac{N_1 - 1}{N-1} \cdots \frac{N_1 - k + 1}{N-k+1} \; \frac{N_2}{N-k} \cdots \frac{N_2 - n + k + 1}{N-n+1}$$

Supposons maintenant que $N \longrightarrow +\infty$, $N_1 \longrightarrow +\infty$ de façon telle que $\frac{N_1}{N} = p$ reste constant, alors on a :

$$\lim_{\substack{N\to+\infty \\ N_1\to+\infty}} \frac{C_{N_1}^{k}\, C_{N_2}^{n-k}}{C_N^n} = C_n^k\, p^k\, q^{n-k}$$

Cette remarque permet d'approximer la probabilité d'obtenir k boules blanches dans une loi hypergéométrique par la probabilité d'obtenir k boules blanches dans une loi binomiale.

§ 7. VARIABLE ALEATOIRE UNIFORME.

Une v.a. X est uniforme sur l'intervalle $a-h, a+h$, $h>0$, si sa f.fr. donnée par :

$$f(x) = \begin{cases} \frac{1}{2h} & \text{pour } a-h \leqslant x \leqslant a+h \\ 0 & \text{ailleurs} \end{cases}$$

On trouve pour la f.d

$$f(x) = \begin{cases} 0 & \text{pour } x < a-h \\ \frac{x-(a-h)}{2h} & \text{pour } a-h \leqslant x < a+h \\ 1 & \text{pour } a+h \leqslant x \end{cases}$$

$$\varphi_X(t) = \frac{1}{2h}\int_{a-h}^{a+h} e^{itx}\, dx = \frac{1}{2h}\, \frac{e^{it(a+h)} - e^{it(a-h)}}{it}$$

$$= e^{ita}\, \frac{\sin th}{th}$$

Par calcul direct, on trouve :

$$E[X] = a \qquad \text{var } X = \frac{h^2}{3}$$

Par la transformation

$$Y = \frac{X-(a-h)}{2h}$$

on se ramène à une v.a. uniforme sur (0,1) pour laquelle, on a :

$$E[Y] = \frac{1}{2} \qquad \text{var } Y = \frac{1}{12}$$

§ 8. VARIABLE ALEATOIRE NORMALE.

Une v.a. Z est dite normale réduite (en abrégé N (0,1)) si sa f.fr. est

$$f_Z(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}} \qquad -\infty < z < +\infty$$

On acceptera le résultat :

$$\int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}}\, dz = \sqrt{2\pi}$$

qui montre que f est une fonction de fréquence.

$$\varphi_Z(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{itz}\, e^{-\frac{z^2}{2}}\, dz = e^{-\frac{t^2}{2}}\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{(z-it)^2}{2}}\, dz$$

$$= e^{-\frac{t^2}{2}}$$

car, par intégration dans le plan complexe, on a :

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{(z-it)^2}{2}}\, dz = 1.$$

Puisque $\varphi_Z(0) = 1$, on vérifie que f_Z est bien une f.fr.
Il est facile de vérifier directement que $E(Z)$ et $E(Z^2)$ existent.
Par dérivation de φ_Z, on trouve alors

$$E[Z] = 0 \qquad \text{var } Z = 1.$$

Il existe des tables pour les v.a. N(0,1) qui permettent de calculer soit $P(Z \leqslant z)$; soit $P(-z \leqslant Z \leqslant z)$; soit $P(0 \leqslant Z \leqslant z)$.
On voit que :

$$P[|Z| \leqslant 1] = 0,6826$$

$$P[|Z| \leqslant 2] = 0,9544$$

$$P[|Z| \leqslant 3] = 0,9974$$

Une v.a. X est dite normale de moyenne m et de variance σ^2 (en abrégé N (m, σ^2)) si la v.a.

$$Z = \frac{X-m}{\sigma} \qquad \sigma \neq 0$$

est N (0,1).
On vérifie immédiatement que :

$$E[X] = m \qquad \text{var } [X] = \sigma^2$$

$$\varphi_X(t) = \varphi_{\sigma Z+m}(t) = e^{itm - \frac{1}{2}\sigma^2 t^2}$$

Par la formule du changement de variable, on a :

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

On trouve directement si X est N (m, σ^2) , alors

$$P[|X-m| \leqslant \sigma\ ] = 0,6826$$

$$P[|X-m| \leqslant 2\sigma] = 0,9544$$
$$P[|X-m| \leqslant 3\sigma] = 0,9978$$

§ 9. VARIABLE ALEATOIRE GAMMA.

Une v.a. X est appelée une v.a. gamma de paramètre (λ, α) (en abrégé $Ga(\lambda, \alpha)$) sa f. fr. est donnée par :

$$f_X(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & x > 0 \quad \lambda > 0 \quad \alpha > 0 \\ 0 & x \leqslant 0 \end{cases}$$

où $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \qquad \alpha > 0$

Pour tout $\alpha > 0$, on voit en intégrant par parties que

$$\Gamma(\alpha+1) = \alpha \Gamma(\alpha)$$

On en déduit que si α est un entier > 0, on a :

$$\Gamma(\alpha+1) = \alpha !$$

Par le changement de variable $\lambda x = u$, on voit que f_X est bien une f. fr.
On a :

$$\varphi_X(t) = \int_0^\infty e^{itx} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\lambda - it)x} dx$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\lambda - it)^\alpha} = \frac{1}{(1 - \frac{it}{\lambda})^\alpha}$$

Par dérivation de φ_X ou par calcul direct, on trouve

$$E[X] = \int_0^\infty x \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx = \frac{\alpha}{\lambda} \int_0^\infty \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} x^\alpha e^{-\lambda x} = \frac{\alpha}{\lambda}$$

$$E[X^2] = \frac{\alpha(\alpha+1)}{\lambda^2} \qquad \text{var } X = \frac{\alpha}{\lambda^2}$$

Cas particuliers :

1) Une v.a. $Ga(\lambda, 1)$ est appelée v.a. exponentielle de paramètre λ ; Sa f. fr. est donnée par

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{pour } x > 0 \\ 0 & \text{pour } x \leqslant 0 \end{cases}$$

Pour une telle v.a., on a :

$$E[X] = \frac{1}{\lambda} \qquad \text{var } X = \frac{1}{\lambda^2}$$

$$\varphi_X(t) = \frac{1}{1 - \frac{it}{\lambda}}$$

2) Une v.a. $Ga(\frac{1}{2}, \frac{n}{2})$ $n \in \mathbb{N}^*$ est appelée une v.a. chi-carré avec n degrés de liberté (en abrégé $X^2(n)$).
Pour une telle v.a., on a :

$$E[X] = n, \qquad \text{var } X = 2n$$

$$\varphi_X(t) = (1 - 2it)^{-\frac{n}{2}}$$

§ 10. ADDITION DES V.A. INDEPENDANTES.

Les v.a. introduites dans ce chapitre jouissent des propriétés suivantes :

1) Si X est Bi (n_1, p), si Y est Bi (n_2, p) et si X et Y sont indépendantes, alors X + Y est Bi (n_1+n_2, p)

2) Si X est Poisson (λ), si Y est Poisson (μ) et si X et Y sont indépendantes, alors X + Y est Poisson $(\lambda + \mu)$

3) Si X est N (m, σ^2), si Y est N (m', σ'^2) et si X et Y sont indépendantes, alors X + Y est N $(m + m', \sigma^2 + \sigma'^2)$.

4) Si X est Ga (λ, α), si Y est Ga (λ, α') et si X et Y sont indépendantes, alors X + Y est Ga $(\lambda, \alpha + \alpha')$.

Ces propriétés se démontrent de manière identique. Démontrons la troisième, par exemple
On a :

$$\varphi_X(t) = e^{itm - \frac{1}{2}\sigma^2 t^2} \qquad \varphi_Y(t) = e^{it m' - \frac{1}{2}\sigma'^2 t^2}$$

$$\varphi_{X+Y}(t) = \varphi_X(t) . \varphi_Y(t) = e^{it(m+m') - \frac{1}{2}(\sigma^2 + \sigma'^2) t^2}$$

Comme la f. c. caractérise la distribution d'une v.a., on voit que X + Y est N $(m+m', \sigma^2 + \sigma'^2)$.

CHAPITRE IX. APPLICATIONS.

Exercice 1. Une urne contient b boules blanches et r boules rouges. On en extrait simultanément deux boules. Quelle est la probabilité qu'elles soient de couleurs différentes ?

Première solution par dénombrement des cas possibles et des cas favorables.

Si on suppose que les deux boules forment un échantillon ordonné, alors Ω contient A^2_{b+r} résultats possibles supposés équiprobables. Il y a alors $2\,b\,r$ cas favorables à la réalisation de l'événement "obtenir deux boules de couleurs différentes" dont la probabilité p est :

$$p = \frac{2\,b\,r}{A^2_{b+r}} = \frac{2\,b\,r}{(b+r)\,(b+r-1)} = \frac{b\,r}{C^2_{b+r}}$$

La dernière formulation montre qu'on peut aussi travailler avec un échantillon non ordonné.

Seconde solution par la formule de décomposition.

Désignons par B_i l'événement "tirer une boule blanche au tirage numéro i" et par R_i l'événement "tirer une boule rouge au tirage numéro i" (i = 1, 2). On a alors :

$$p = P\left[(B_1 \cap R_2) \cup (R_1 \cap B_2)\right] = \frac{b}{b+r}\ \frac{r}{b+r-1} + \frac{r}{b+r}\ \frac{b}{b+r-1}$$

Exercice 2. Paradoxe du Chevalier de Méré.

Le chevalier de Méré pensait qu'il y avait égalité entre les deux probabilités suivantes :

a) Probabilité d'amener le point 6 au moins une fois en lançant 4 fois un dé à 6 faces

b) Probabilité d'amener le couple (6,6) au moins une fois en lançant 24 fois deux dés à 6 faces.

Le raisonnement du chevalier de Méré était le suivant: en lançant un dé à 6 faces, il y a 6 résultats possibles que nous supposerons équiprobables. En lançant deux dés à 6 faces, il y a 36 résultats possibles que nous supposerons également probables. Obtenir double-six est donc six fois moins probable qu'obtenir 6. Par conséquent, il sera aussi probable d'amener double-six (au moins une fois) en 6 k lancers des deux dés que d'amener six (au moins une fois) en k lancers d'un seul dé.

Comme "la pratique" ne confirmait pas son raisonnement (au grand détriment de sa fortune), le Chevalier de Méré écrivit à Pascal qui trouva l'explication de ce paradoxe. Ce fut le début de la théorie des probabilités.

Venons-en à la solution du problème. En lançant 4 fois un dé à 6 faces, il y a 6^4 résultats possibles du type (a_1, a_2, a_3, a_4) avec $a_i = 1,2,3,4,5,6$ et $i = 1,2,3,4$. Parmi ceux-là, 5^4 ne contiennent pas le chiffre 6. Donc $6^4 - 5^4$ contiennent au moins une fois le chiffre 6. Ainsi, la probabilité d'amener le point 6 au moins une fois en lançant 4 fois un dé à 6 faces est :

$$\frac{6^4 - 5^4}{6^4} = 1 - \left(\frac{5}{6}\right)^4 = 0,517747.$$

En lançant 24 fois deux dés à 6 faces, il y a 36^{24} résultats possibles du type $\left\{(a_1, b_1),\ (a_2, b_2), \ldots . (a_{24},\ b_{24})\right\}$ $a_i, b_j = 1,2,3,4,5,6$ et $i, j = 1, 2, \ldots . , 24$. Parmi ceux-là, 35^{24} ne contiennent pas le résultat (6,6). Donc $36^{24} - 35^{24}$ contiennent au moins une fois le résultat (6,6). Ainsi, la probabilité d'amener double-six en lançant 24 fois deux dés à 6 faces est :

$$\frac{36^{24} - 35^{24}}{36^{24}} = 1 - \left(\frac{35}{36}\right)^{24} = 0,491404 \neq 0,517747$$

Pour la petite historie indiquons que ce qui contrariait le chevalier de Méré était le fait suivant : il avait l'habitude de parier d'amener le point 6 (sous-entendu au moins une fois) en 4 lancers d'un dé et il constatait qu'il gagnait plus souvent qu'il ne perdait. Ceci était normal puisqu'il avait une probabilité de gagner strictement supérieure à $\frac{1}{2}$. Comme ses adversaires avaient fait la même constatation, ils refusaient de jouer. Alors le Chevalier de Méré a proposé le second jeu et il a été fort étonné de constater qu'il perdait plus souvent qu'il ne gagnait. Pourtant la différence entre les deux probabilités n'est que de 0,026343.

On pouvait arriver à la solution par un autre raisonnement. Puisque les lancers sont successifs, on peut les supposer indépendants. Alors la probabilité de ne pas avoir 6 en quatre lancers d'un dé est $(\frac{5}{6})^4$. De même la probabilité de ne pas avoir (6,6) en 24 lancers de deux dés est $(\frac{35}{36})^{24}$. Comme l'événement dont on vient de trouver la probabilité $\frac{35}{36}$ est le complémentaire de celui dont on cherchait la probabilité, on retrouve le résultat.

Exercice 3. Quelle est la probabilité d'obtenir au moins une fois le point 6 en jetant 3 dés ayant respectivement les probabilités p_1, p_2, p_3 d'amener 6 ?

On peut supposer les 3 lancers indépendants. On trouve alors comme réponse : $1 - (1-p_1)\,(1-p_2)\,(1-p_3)$

Exercice 4. Dans un schéma de Bernoulli, quelle est la probabilité que le nombre de succès soit pair ?

La probabilité cherchée est :

$$C_n^0 q^n + C_n^2 p^2 q^{n-2} + \ldots =$$

$$\frac{1}{2}\left[(p+q)^n + (q-p)^n\right] =$$

$$\frac{1}{2}\left[1 + (q-p)^n\right]$$

Une autre façon d'obtenir la solution est d'établir une équation de récurrence dont la probabilité cherchée est la solution. Soit π_n la probabilité que le nombre de succès soit pari en n expériences. On a par la formule de décomposition :

$$\pi_1 = q$$

$$\pi_n = q\,\pi_{n-1} + p\,(1-\pi_{n-1}) = (q-p)\,\pi_{n-1} + p \qquad n = 2,3,\ldots$$

Il reste à résoudre cette équation. Remarquons d'abord que si une suite (p_n) de nombres satisfait l'équation

$$p_n = a\,p_{n-1} \qquad n = 2,3,\ldots \qquad (1)$$

où a est une constante donnée, alors

$$p_n = p_1\,a^{n-1} \qquad n = 1,2,\ldots$$

Par conséquent, l'équation

$$p_n = a\,p_{n-1} + b \qquad n = 2\ 3,\ldots \qquad (2)$$

où a et b sont des constantes données, a pour solution

$$p_n = (p_1 - \frac{b}{1-a})\,a^{n-1} + \frac{b}{1-a} \qquad \text{si } a \neq 1$$

$$p_n = (n-1)\,b + p_1 \qquad \text{si } a = 1$$

En effet, si p_n satisfait l'équation (2), nous posons :

$$p'_n = p_n - \frac{b}{1-a} \qquad a \neq 1$$

et nous obtenons

$$p'_n = a\,p_{n-1} + b - \frac{b}{1-a} = a\,p_{n-1} - \frac{a\,b}{1-a}$$

$$= a\,(p_{n-1} - \frac{b}{1-a}) = a\,p'_{n-1}\ .$$

ce qui montre que la suite (p'_n) satisfait l'équation (1).

Finalement, on trouve :

$$\pi_n = (q - \frac{p}{1-(q-p)})\,(q-p)^{n-1} + \frac{p}{1-(q-p)}$$

$$= \frac{1}{2}\left[1 + (q-p)^n\right]\ .$$

Exercice 5.- Dans un schéma de Bernoulli, on demande la probabilité $P(2n)$ d'obtenir le même nombre de piles que de faces $(n \geqslant 1)$. Montrer que $P(2n)$ est une fonction décroissante de n. Si un joueur reçoit 2n quand il y a n succès en 2 n parties et 0 pour tout autre résultat, montrer que l'espérance de la v.a. représentant ce qu'il reçoit est une fonction croissante de n.

Le nombre de parties est nécessairement pair. On a :

$$P(2n) = C_{2n}^n \left(\frac{1}{2}\right)^{2n}$$

$$\frac{P\,(2n+2)}{P(2n)} = \frac{2n+1}{2n+2} < 1$$

$$E[X_{2n}] = 2n\,P(2n) + 0\,(1 - P(2n)) = 2\,n\,P(2n)$$

$$\frac{(2n+2)\,P\,(2n+2)}{2n\quad P(2n)} = \frac{2n+1}{2n} > 1$$

Exercice 6 : Un joueur joue à pile ou face, avec une pièce parfaitement équilibrée, de la manière suivante : il parie toujours pour pile et, si face sort, il double sa mise pour le lancer suivant. Quelle est l'espérance de son gain ?

Le joueur qui a mise 1 franc en reçoit 2 quand pile sort et 0 quand face sort. Son gain après la première partie est donc 1 franc si pile est sorti et -1 franc si face est sorti.

Si pile sort pour la première fois au $n^{ième}$ lancer, le joueur qui avait mise 1 franc au premier lancer et avait ensuite doublé sa mise reçoit 2^n francs. Son gain à ce moment est donc :

$$2^n - (1 + 2 + 2^2 + \ldots + 2^{n-1}) = 2^n - (2^n - 1) = 1.$$

Puisque la probabilité d'obtenir pile la première fois au $n^{ième}$ lancer est $\frac{1}{2^n}$, l'espérance du gain est

$$\sum_{n=1}^{\infty} \frac{1}{2^n} = 1.$$

Il semble qu'avec cette stratégie le joueur aura trouvé un système pour gagner. Mais ceci n'est valable que si le joueur dispose d'une fortune infinie. Si sa fortune est limitée l'espérance de son gain est nulle comme on le voit facilement. L'espérance de son gain est en effet nulle pour chacune des parties.

Exercice 7. Modèle d'urne de Polya:

Une urne contient b boules blanches et r boules rouges. On extrait au hasard une boule de l'urne. On remet la boule dans l'urne et on ajoute c boules de même couleur que celle de la boule extraite avant de procéder au tirage suivant. On suppose qu'à chaque tirage, chaque boule a la même probabilité d'être extraite.

Désignons par B_i l'événement "tirer une boule blanche au tirage numéro i" et par R_i l'événement "tirer une boule rouge au tirage numéro i"

1) $[P \wedge R_1 \wedge R_2 \wedge R_3] = \frac{r}{b+r} \quad \frac{r+c}{b+r+c} \quad \frac{r+2c}{b+r+2c}$

2) $P[R_2] = P[R_1 \wedge R_2] + P[B_1 \wedge R_2]$

$$= \frac{r}{b+r} \quad \frac{r+c}{b+r+c} + \frac{b}{b+r} \quad \frac{r}{b+r+c}$$

$$= \frac{r}{b+r}\left[\frac{r+c}{b+r+c} + \frac{b}{b+r+c}\right]$$

$$= \frac{r}{b+r}$$

3) $P[R_n] \stackrel{?}{=} \frac{r}{b+r}$

On a :

$$P[R_n] = P[R_{n-1} \wedge R_n] + P[B_{n-1} \wedge R_n]$$

et si on désigne par α le nombre de boules rouges contenues dans l'urne au moment d'effectuer le $(n-1)^{ième}$ tirage, on a :

$$P[R_n] = \frac{\alpha}{b+r+(n-2)c} \quad \frac{\alpha+c}{b+r+(n-1)c} + \frac{b+r+(n-2)c-\alpha}{b+r+(n-2)c} \quad \frac{\alpha}{b+r+(n-1)c}$$

$$= \frac{\alpha}{b+r+(n-2)c} \quad \frac{b+r+(n-1)c}{b+r+(n-1)c}$$

$$= P[R_{n-1}]$$

Par induction, on a le résultat.

4) $P[R_1|R_2] = \frac{P[R_1 \wedge R_2]}{P[R_2]} = \frac{\frac{r}{b+r} \quad \frac{r+c}{b+r+c}}{\frac{r}{b+r}}$

$$= \frac{r+c}{b+r+c}$$

5) Pour $i<j$, on a :

$$P[B_i \wedge B_j] = P[B_1 \wedge B_2] = \frac{b}{b+r} + \frac{b+c}{b+r+c}$$

$$P[B_i \wedge R_j] = P[B_1 \wedge R_2] = \frac{b}{b+r} + \frac{r}{b+r+c}$$

6) De ce qui précède, on déduit immédiatement :

$$P[B_i|R_j] = P[B_j|R_i]$$

7) Si $i_1, i_2, \ldots, i_n$ est une permutation de $1, 2, \ldots n$, on a :

$$P[B_{i_1} \wedge B_{i_2} \wedge \ldots \wedge B_{i_k} \wedge R_{i_{k+1}} \wedge \ldots \wedge R_{i_n}] =$$

$$P[B_1 \wedge B_2 \wedge \ldots \wedge R_{k+1} \wedge \ldots \wedge R_n] =$$

$$\frac{\prod_{i=1}^{k}[b+(i-1)c] \quad \prod_{j=1}^{n-k}[r+(j-1)c]}{\prod_{\ell=1}^{n}[b+r+(\ell-1)c]} =$$

$$\frac{b(b+c)\ldots(b+(k-1)c)\; r(r+c)\ldots(r+(n-k-1)c)}{(b+r)\;(b+r+c)\;\ldots\;(b+r+(n-1)c)}$$

Appelons $\pi_{k,n}$ le nombre ainsi obtenu.

8) Probabilité d'avoir exactement k boules blanches en n tirages.

Soit $P_{k,n}$ cette probabilité. Par la partie 7, on trouve

$$P_{k,n} = C_n^k \quad \pi_{k,n}$$

Si on définit :

$$\pi_{o,n} = \frac{\prod_{j=1}^{n}[r+(j-1)c]}{\prod_{\ell=1}^{n}[b+r+(\ell-1)c]}$$

$$\pi_{n,n} = \frac{\prod_{i=1}^{k}[b+(i-1)c]}{\prod_{\ell=1}^{n}[b+r+(\ell-1)c]}$$

On voit que la suite $(P_{k,n})$ $k = 0, 1, 2, \ldots, n$ définit une distribution de probabilité d'une v.a. discrète prenant les valeurs $0, 1, 2 \ldots k, \ldots n$ et $P[X=k] = P_{k,n}$.

D'après la définition des $P_{k,n}$, il est clair que :

$$\sum_{k=o}^{n} P_{k,n} = 1$$

Une vérification directe est plus laborieuse.

La distribution de la v.a. ainsi définie porte le nom de distribution de Polya.

9) Espérance mathématique et variance de la v.a. X définie à la partie 8 et qui représente le nombre de boules blanches en n tirages.

Une évaluation directe est laborieuse. Il faut en effet chercher :

$$E[X] = \sum_{k=o}^{n} k \; P_{k,n}$$

$$E[X^2] = \sum_{k=o}^{n} k^2 \; P_{k,n}$$

Il est plus facile de procéder comme pour la loi hypergéométrique et de définir l'indicatrice Y_j du $j^{ième}$ tirage par :

$$Y_j = \begin{cases} 1 & \text{si le } j^{ième} \text{ tirage donne une blanche} \\ 0 & \text{si le } j^{ième} \text{ tirage donne une rouge.} \end{cases}$$

D'après la partie 2, on a :

$$P[Y_j=1] = \frac{b}{b+r} \qquad P[Y_j=0] = \frac{r}{b+r}$$

$$E[Y_j] = \frac{b}{b+r} \qquad \text{var } Y_j = \frac{br}{(b+r)^2}$$

$$i \neq j \quad \text{cov}(Y_i, Y_j) = \frac{b}{b+r} \; \frac{b+c}{b+r+c} - \left(\frac{b}{b+r}\right)^2$$

$$= \frac{b\,r\,c}{(b+r)^2\,(b+r+c)}$$

Puisque :

$$X = Y_1 + Y_2 + \ldots + Y_n$$

on trouve :

$$E[X] = \frac{n\,b}{b+r}$$

$$\text{var } X = \frac{nbr}{(b+r)^2} + n(n-1)\frac{brc}{(b+r)^2(b+r+c)}$$

$$= \frac{nbr}{(b+r)^2}\left[1 + (n-1)\frac{c}{b+r+c}\right]$$

$$= \frac{n\,b\,r}{(b+r)^2} \; \frac{b+r+nc}{b+r+c}$$

Si on pose :

$$\frac{b}{b+r} = p \qquad \frac{r}{b+r} = q \qquad \frac{c}{b+r} = \alpha$$

On trouve :

$$E[X] = n\,p$$

$$\text{var } X = n\,p\,q\,\frac{1+n\alpha}{1+\alpha}$$

En particulier, si on pose $c = -1$ (ce qui revient à ne pas remplacer la boule extraite) et $N = b+r$, on retrouve la loi hypergéométrique.

Supposons que dans la distribution de Polya, b, r et $b+r=N$ tendent vers $+\infty$ de façon telle que $\frac{b}{N}$ reste constant, c'est-à-dire que $\frac{b}{b+r} = p$ reste constant et donc $\frac{r}{b+r} = q = 1-p$ reste aussi constant.

Supposons de plus que $\lim_{N\to+\infty} \frac{c}{N} = 0$. Cette condition est satisfaite lorsque c reste constant et N tend vers $+\infty$. Alors il est facile de voir que :

$$\lim_{n\to+\infty} P_{k,n} = C_n^k \, p^k \, q^{n-k}$$

Exercice 8. Réalisation de m parmi n événements.

Soient $A_1, A_2, \ldots, A_n$ n événements de $\mathcal{a}$. Proposons-nous de chercher, pour $m \leqslant n$,
la probabilité P_m de la réalisation d'au moins m des événements,
la probabilité $P_{(m)}$ de la réalisation d'exactement m des n événements.

Il est clair que pour $m \geqslant 1$, on a :

$$P_m = P_{(m)} + P_{(m+1)} + \ldots + P_{(n)} .$$

Commençons par chercher P_1 ce qui revient à chercher $P[A_1 \cup A_2 \cup \ldots \cup A_n]$.

Utilisons la méthode des indicatrices, nous trouvons :

$$I\left(\bigcup_{i=1}^{n} A_i\right) = I(A_1) + (1-I(A_1))\,I(A_2) + \ldots + (1-I(A_1))(1-I(A_2))\ldots(1-I(A_{n-1}))I(A_n)$$

$$= \sum_{j=1}^{n} I(A_j) - \sum_{\substack{i,j=1 \\ i<j}}^{n} I(A_i \cap A_j) + \sum_{\substack{i,j,k=1 \\ i<j<k}}^{n} I(A_i \cap A_j \cap A_k) + \ldots$$

$$+(-1)^{n-1} \, I(A_1 \cap A_2 \cap \ldots \cap A_n)$$

Pour trouver P_1, il suffit de prendre l'espérance mathématique des deux membres. Définissons :

$$S_o = 1$$

$$S_1 = \sum_{j=1}^{n} P[A_j]$$

$$S_2 = \sum_{\substack{i,j=1 \\ i<j}}^{n} P[A_i \cap A_j]$$

$$\vdots$$

$$S_k = \sum_{\substack{i_1,i_2,\ldots,i_k=1 \\ i_1<i_2<\ldots<i_k}}^{n} P[A_{i_1} \cap A_{i_1} \cap \ldots \cap A_{i_k}]$$

$$\vdots$$

$$S_n = P[A_1 \cap A_2 \cap \ldots \cap A_n]$$

La somme S_k comporte C_n^k termes.
On trouve finalement la formule de Poincaré.

$$P_1 = S_1 - S_2 + S_3 - S_4 + \ldots + (-1)^{n-1} S_n .$$

On en déduit :

$$P_{(0)} = 1 - P_1 = 1 - S_1 + S_2 - S_3 + S_4 \ldots + (-1)^n S_n .$$

Cherchons maintenant $P_{(m)}$.

Appelons $B_{(m)}$ l'événement qui n'est réalisé que si et seulement si m des n événements $A_1, A_2, \ldots, A_n$ sont réalisés.

L'événement $B_{(m)}$ est la réunion d'événements deux à deux incompatibles du type.

$$A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_m} \cap A^c_{i_{m+1}} \cap \ldots \cap A^c_{i_n}$$

où $(i_1, i_2, \ldots, i_m)$ est l'une des C_n^m combinaisons sans répétition des n nombres $1, 2, \ldots, n$ pris m à m. Par conséquent :

$$I(B_{(m)}) = \Sigma I(A_{i_1}) \ldots I(A_{i_m}) (1 - I(A_{i_{m+1}})) \ldots (1 - I(A_{i_n})).$$

La somme du second membre comporte des termes du type :

$$(-1)^k I(A_{i_1}) \ldots I(A_{i_m}) I(A_{j_1}) \ldots I(A_{j_k})$$

où $(j_1, \ldots, j_k)$ est l'une des C^k_{n-m} combinaisons sans répétition des (n-m) nombres $(i_{m+1}, \ldots, i_n)$ pris k à k. Remarquons que k peut varier de 0 à n-m.

Des termes de cette nature se retrouvent donc $C_n^m \; C^k_{n-m}$ fois.

Pour trouver $P_{(m)}$, il suffit de prendre l'espérance mathématique de $I(B_{(m)})$. On voit donc apparaître dans $P_{(m)}$ des termes du type :

$$P[A_{i_1} \cap \ldots \cap A_{i_m} \cap A_{j_1} \cap \ldots \cap A_{j_k})$$

qui figurent dans S_{m+k}. Comme S_{m+k} compte C_n^{m+k} termes, et puisque :

$$C_n^m \; C^k_{n-m} = C_n^{m+k} \; C^m_{m+k}$$

on trouve finalement :

$$P_{(m)} = \sum_{k=o}^{n-m} (-1)^k \; C^m_{m+k} \; S_{k+m}$$

$$P_m = \sum_{k=o}^{n-m} (-1)^k \; C^{m-1}_{m+k-1} \; S_{k+m}$$

Exercice 9. Le problème de la rencontre.

Supposons avoir n urnes numérotées de 1 à n et n boules numérotées de 1 à n. On distribue aléatoirement les n boules dans les n urnes de façon que chaque urne contienne exactement une boule et on suppose que chacune des n! permutations des n boules a la probabilité $\frac{1}{n!}$. Lorsqu'une boule se trouve dans l'urne qui porte le même numéro qu'elle, on dit qu'il y a rencontre. On demande la probabilité d'avoir une rencontre au moins et la probabilité d'avoir m rencontres exactement pour $m = 0, 1, \ldots, n$.

Ce problème est célèbre. On le rencontre sous de nombreuses formes. Première forme : n personnes déposent leur chapeau au vestiaire avant un spectacle. Un incendie survient, chaque personne se précipite au vestiaire et prend au hasard un chapeau. Quelle est la probabilité qu'une personne au moins retrouve son chapeau? Seconde forme : n couples vont danser. En supposant que chaque homme a la même probabilité de danser avec chacune des femmes, quelle est la probabilité qu'au moins m hommes dansent avec leur partenaire ? Troisième forme: une secrétaire farfelue écrit n lettres. Elle fait ensuite les adresses sur les enveloppes. Elle glisse une lettre dans chaque enveloppe en laissant à chaque lettre la même probabilité de figurer dans chaque enveloppe. Quelle est la probabilité que m personnes exactement recoivent la lettre qui leur est destinée ?

Posons A_k = l'événement "une rencontre a lieu au numéro k". Cet événement est réalisé si la boule numéro k se trouve dans l'urne numéro k et si les n-1 boules restantes sont réparties dans les n-1 urnes restantes. Nous avons :

$$P[A_k] = \frac{(n-1)!}{n!} = \frac{1}{n} \qquad k = 1, 2, \ldots, n.$$

De même, pour $i \neq j$, on a

$$P[A_i \cap A_j] = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)} \qquad i, j = 1, 2, \ldots, n .$$

On trouve alors :

$$S_k = C_n^k \; \frac{(n-k)!}{n!} = \frac{1}{k!} \qquad k = 1, 2, \ldots, n$$

et

$$P_1 = 1 - \frac{1}{2!} + \frac{1}{3!} \; , \ldots + (-1)^{n-1} \frac{1}{n!}$$

$$\simeq 1 - \frac{1}{e} = 0,63212 \qquad \text{lorsque n est grand.}$$

$$P_{(m)} = \sum_{k=o}^{n-m} (-1)^k \; C^m_{m+k} \; \frac{1}{(k+m)!}$$

$$= \frac{1}{m!} \sum_{k=o}^{n-m} (-1)^k \; \frac{1}{k!}$$

$$\simeq \frac{1}{m!} \; e^{-1} \qquad \text{lorsque } n-m \text{ est grand.}$$

Exercice 10. Vérifier que dans le problème de la rencontre, on a :

$$P_{(n-1)} = 0$$

$$P_{(n)} = \frac{1}{n!}$$

Modèles d'urnes.

On peut résoudre de nombreux problèmes en définissant convenablement un modèle du type suivant : une urne contient N boules différentes, numérotées de 1 à N par exemple. On en extrait un ensemble de n boules appelés échantillon de taille n. L'échantillon peut être obtenu par un procédé avec remise si on remet dans l'urne la boule extraite lors de l'un des tirages avant de procéder au tirage suivant. L'échantillon est obtenu par un procédé sans remise si on ne remet pas la boule extraite lors de l'un des tirages avant de procéder au tirage suivant. Il faut encore distinguer si l'échantillon obtenu est ordonné ou s'il n'est pas ordonné.

Une formulation équivalente est fournie par le problème suivant : répartir n boules entre N urnes distinctes. A la notion d'échantillon avec remise correspond la suivante : une même urne peut contenir plus d'une boule. A la notion d'échantillon sans remise correspond la suivante :une même urne peut contenir une boule au plus. A la notion d'échantillon ordonné correspond celle de boules distinctes, à la notion d'échantillon non ordonné correspond celle de boules identiques.

Le tableau suivant donne le nombre de possibilités pour extraire n boules d'une urne qui contient N boules distinctes quand on le lit de haut en bas. Les spécifications appropriées figurent alors sur le côté gauche. Le même tableau donne le nombre de possibilités pour répartir n boules entre N urnes distinctes quand on le lit de bas en haut. Les spécifications appropriées figurent alors sur le côté droit.

Extraire n boules d'une urne
qui contient N boules distinctes

	Avec remise	Sans remise	
Echantillon ordonné	α^n_N	A^n_N	boules distinctes
Echantillon non ordonné	γ^n_N	C^n_N	boules identiques
	Répartition libre	Une boule au plus par urne	

Répartir n boules entre N urnes distinctes.

On a ainsi défini le nombre de résultats possibles de Ω. Il est généralement facile d'écrire l'un de ces résultats. Il est en général fastidieux et inutile de vouloir les écrire tous. Pensons qu'il faudrait écrire 10^{100} résultats pour décrire complètement l'extraction d'un échantillon ordonné de taille 100 pris avec remise d'une urne qui contient 10 boules distinctes. Dans tous les problèmes ramenés à des modèles d'urne où il n'y a qu'un nombre fini de cas possibles, on prend $\mathcal{A} = \mathcal{P}(\Omega)$.

Dans chacune des situations rencontrées dans le tableau on dit que l'échantillon est aléatoire lorsque chacune des dispositions possibles a la même probabilité.

Un premier exemple d'applications des modèles d'urnes est rencontré dans ce que l'on appelle "les statistiques de la physique". Ici le mot "statistique" doit être pris dans le sens de répartition.

En physique statistique, on s'intéresse à la détermination de l'état d'équilibre d'un système physique composé d'un très grand nombre, n, de particules de même nature (électrons, protons, photons, mésons,...). Pour la simplicité de la présentation, on suppose qu'il y a N états possibles pour chaque particule, correspondant, par exemple, aux différents niveaux d'énergie. Chacun de ces niveaux est symbolisé par une urne. L'état du système peut alors être caractérisé par le vecteur $(n_1, n_2, \ldots, n_N)$ où n_j est le nombre de particules à l'état j, symbolisé par le nombre de boules se trouvant dans l'urne numéro j.

L'état d'équilibre du système correspond à la répartition qui a la plus grande probabilité.

En faisant des hypothèses sur la nature des boules et sur le mode de répartition, on trouve différents modèles. Le plus ancien est celui qui porte le nom de "statistique de Maxwell-Boltzmann". On y suppose que les boules sont distinctes, qu'une même urne peut contenir plusieurs boules et que toutes les répartitions sont équiprobables. Elle s'applique, en mécanique statistique, aux molécules de gaz par exemple. Le nombre de répartitions possibles est alors de $\alpha_N^n = N^n$ supposées équiprobables.

Mais pour les particules élémentaires la statistique de Maxwell-Boltzmann ne s'applique pas, car pour aucun type de ces particules, les N^n arrangements ne sont équiprobables. Il y a alors deux modèles appropriés : la "statistique de Bose-Einstein" et la "statistique de Fermi-Dirac".

Dans la "statistique de Bose-Einstein" applicable aux photons, on ne suppose plus les boules comme distinctes, ce qui revient à dire que les particules sont indistinguables mais on suppose encore qu'une même urne peut contenir plusieurs boules. Il est clair que deux répartitions présentant le même nombre de boules dans chaque urne sont considérées comme identiques. Le nombre de répartitions possibles des n boules dans les N urnes est égal dans ce cas au nombre γ_N^n de combinaisons avec répétition de N éléments différents pris n à n. On les suppose équiprobables.

Pour les électrons, protons et neutrons, il faut utiliser la "statistique de Fermi-Dirac" à cause du principe d'exclusion de Pauli suivant lequel il ne peut y avoir qu'au plus un électron dans un état donné. Cela implique $n \leqslant N$. On doit toujours supposer les boules identiques. Il y a alors C_N^n répartitions possibles. On les suppose équiprobables.

Les exercices qui suivent portent sur la réalisation de m parmi n événements, sur les modèles d'urnes, les "statistiques" de la physique et le chapitre des distributions aléatoires.

Exercice 11 : On distribue n boules différentes entre n urnes distinctes. La répartition des boules entre les urnes est libre. On demande la probabilité qu'aucune urne ne reste vide.

La probabilité cherchée est :

$$\frac{A_n^n}{\alpha_n^n} = \frac{n!}{n^n}$$

Le résultat trouvé représente la probabilité de ne pas extraire deux fois la même boule en extrayant, par un procédé avec remise, un échantillon ordonné de taille n d'une urne qui contient n boules distinctes.

Pour n = 6, on trouve : $\frac{6!}{6^6} = 0,01543.$

C'est la probabilité de voir apparaître l'ensemble des 6 résultats possibles en jetant simultanément 6 dés à 6 faces parfaitement réguliers.

Exercice 12. Le problème des anniversaires.

n personnes sont réunies dans une salle. Quelle est la probabilité que deux quelconques d'entre elles n'aient pas leur anniversaire à la même date ?

Pour résoudre ce problème, on supposera que l'année comporte toujours 365 jours et que tous les arrangements possibles ont la même probabilité. Cette dernière phrase constitue vraiment une approximation, puisque les naissances ne sont pas réparties uniformément dans l'année. Il en est d'ailleurs de même des mariages qui, en principe, les précèdent.

Sous ces hypothèses, la probabilité cherchée est :

$$\pi = \frac{A_{365}^n}{\alpha_{365}^n} = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right) \ldots \left(1 - \frac{n-1}{365}\right)$$

Comme le calcul est assez long dès que n dépasse quelques unités, on peut essayer de trouver les formules approchées.

Si n est petit par rapport à 365, on peut négliger dans le produit tous les termes contenant strictement plus d'une fraction dont le dénominateur est 365. On trouve alors pour valeur approchée;

$$\pi \simeq 1 - \frac{1 + 2 + \ldots + (n-1)}{365}$$

et puisque :

$$1 + 2 + \ldots + (n-1) = \frac{n(n-1)}{2}$$

on a pour valeur approchée de la probabilité

$$\pi \simeq 1 - \frac{n(n-1)}{730}$$

Si, au contraire, k est grand par rapport à 365, alors $\frac{k-1}{365}$ est voisin de 1. On peut alors prendre le logarithme du produit et utiliser la relation approchée :

$$\mathrm{Log}_e (1-x) \simeq -x \qquad 0 \leqslant x < 1.$$

On trouve alors :

$$\mathrm{Log}_e \pi \simeq -\frac{n(n-1)}{730}$$

Si n = 23, on trouve $\pi < \frac{1}{2}$. Ainsi, si $n \geqslant 23$, il y a une probabilité strictement supérieure à $\frac{1}{2}$ qu'au moins deux personnes aient leur anniversaire le même jour.

Exercice 13. On répartit n boules différentes entre N urnes distinctes. Sachant qu'une même urne peut contenir jusqu'à n boules, trouver la probabilité p_k qu'une urne déterminée contienne exactement k boules (k = 0, 1, . . . , n) ?

Sous les hypothèses faites, il y a $\alpha_N^n = N^n$ répartitions possibles des n boules entre les N urnes. On peut choisir les k boules parmi les n de C_n^k façons différentes. Ces k boules sont alors déposées dans l'urne choisie. Les (n-k) boules restantes peuvent alors être distribuées librement de $(N-1)^{n-k}$ façons différentes dans les n-1 urnes restantes. La probabilité cherchée est donc :

$$p_k = C_n^k \frac{(N-1)^{n-k}}{N^n} = C_n^k \left(\frac{1}{N}\right)^k \left(1 - \frac{1}{N}\right)^{n-k} \qquad k = 0, 1, \ldots, n.$$

On constate qu'on trouve la probabilité rencontrée dans le schéma binomial. Pour voir qu'il en est bien ainsi, supposons qu'on répartisse les boules l'une après l'autre. Appelons "succès" le fait de déposer une boule dans l'urne choisie. Alors, on a :

$$P\,[\text{"succès" avec la } j^{\text{ième}} \text{ boule}\,] = \frac{1}{N} \qquad j = 1, 2, \ldots, n.$$

et les événements sont manifestement indépendants.

En physique statistique, on a affaire à une "statistique de Maxwell-Boltzmann". La recherche de l'état d'équilibre revient à déterminer k pour que p_k soit maximum.

Posons :

$$\frac{1}{N} = p \qquad \frac{N-1}{N} = q$$

Considérons le rapport :

$$\frac{p_k}{p_{k-1}} = \frac{n - k + 1}{k} \cdot \frac{p}{q}$$

Aussi longtemps que ce rapport sera supérieur à un, les termes p_k croîtront avec k. Dès qu'il deviendra inférieur à un, ils décroîtront avec k. Le terme p_k le plus grand correspondra à l'indice k qui satisfait la double inégalité :

$$p_{k-1} \leqslant p_k \qquad p_k \geqslant p_{k+1}$$

Cet indice k est encore celui qui satisfait les deux inégalités :

$$(n-k+1)\, p \geqslant kq$$

$$(n-k)\, p \leqslant (k+1)q$$

ou encore, puisque p+q=1, celui qui satisfait la double inégalité :

$$np - q \leqslant k \leqslant np + p$$

Ainsi l'indice k cherché est l'unique entier compris entre

np-q et np+p. Comme la différence entre ces deux nombres est l'unité, il se peut que np-q et np+p soient entiers. Alors les termes p_k avec k = np-q et p_{k+1} avec k+1 = np+p correspondent à deux termes égaux dans le développement de $(p+q)^n$.

Exercice 14. On demande la probabilité pour qu'une urne, choisie à l'avance, contienne exactement k boules (k = 0, 1, . . . , n) quand on répartit n boules identiques entre N urnes distinctes, une même urne pouvant contenir jusqu'à n boules.

On peut répartir les n boules entre les N urnes de γ_N^n façons différentes. Après avoir mis k boules dans l'urne choisie, on peut répartir les n-k boules restantes entre les N-1 urnes restantes de γ_{N-1}^{n-k} façons différentes. Ainsi, la probabilité cherchée est :

$$q_k = \frac{\gamma_{N-1}^{n-k}}{\gamma_N^n} = \frac{C_{N+n-k-2}^{n-k}}{C_{N+n-1}^{n}} \qquad k = 0, 1, \ldots, n.$$

En physique statistique, on a affaire ici à une "statistique de Bose-Einstein". La recherche de l'état d'équilibre revient à déterminer k pour que q_k soit maximum.

On traite immédiatement les cas où N ⩽ 2. Si N > 2, on a:

$$q_o > q_1 > q_2 > \cdots > q_n .$$

En effet, le rapport :

$$\frac{q_k}{q_{k+1}} = \frac{N + n - k - 2}{n - k}$$

est strictement supérieur à un dès que N > 2.

Le rapport $\frac{n}{N}$ est considéré comme le nombre moyen de boules par urne. Supposons que $n \to +\infty$, $N \to +\infty$ de façon telle que $\frac{n}{N} \to a$. Alors, on écrit :

$$q_k = \frac{(N-1)\,(n-k+1)\,(n-k+2) \ldots n}{(N+n-k-1)(N+n-k)\ldots(N+n-1)}$$

et en divisant numérateur et dénominateur par N, et en faisant tendre N et n vers $+\infty$, on trouve :

$$q_k \longrightarrow \frac{a^k}{(1+a)^{k+1}} = \left(\frac{a}{1+a}\right)^k \frac{1}{1+a}$$

Si on pose :

$$\frac{1}{1+a} = p \qquad \frac{a}{1+a} = q$$

on voit que :

$$q_k \longrightarrow q^k p$$

qui est le terme général d'une distribution géométrique.

Exercice 15. Même problème que l'exercice précédent mais on suppose qu'une même urne ne peut contenir qu'une boule au plus.

La probabilité cherchée est

$$\frac{C_{N-1}^{n-k}}{C_N^n} \qquad \text{si } k = 0, 1.$$

Exercice 16. On répartit n boules identiques entre N urnes distinctes, une même urne pouvant contenir jusqu'à n boules. Trouver la probabilité que m urnes exactement restent vides.

On peut choisir les m urnes de C_N^m façons différentes. Il faut distribuer les n boules dans N-m urnes de façon qu'aucune urne ne reste vide. Il y a C_{n-1}^{N-m-1} façons de faire cette distribution. La probabilité cherchée est

$$\frac{C_N^n \quad C_{n-1}^{N-m-1}}{C_{N+n-1}^n}$$

Exercice 17. Une urne contient N boules dont N_1 sont blanches et N_2 sont noires. On extrait n boules sans remise comme dans le modèle hypergéométrique. Soit A_j l'événement "obtenir une boule blanche pour la première fois au $j^{ième}$ tirage". On demande la probabilité de A_j.

On remarque que pour réaliser A_j il faut que les j-1 premières boules soient noires, la $j^{ième}$ soit blanche, les n-j autres soient noires ou blanches.

$$P[A_j] = \frac{(j-1)!\ C_{N_2}^{j-1} \quad N_1 \ C_{N-j}^{n-j}\ (n-j)!}{C_N^n\ n!} \qquad j=1, 2, \ldots, \min(n, N_2+1)$$

$$= \frac{A_{N_2}^{j-1} \quad N_1 \ A_{N-j}^{n-j}}{A_N^n}$$

En particulier, si N = n, on trouve :

$$P'[A_j] = \frac{A_{N_2}^{j-1} \quad N_1 \ (N-j)!}{N!} \qquad j=1, 2, \ldots, N_2+1.$$

Si dans cette dernière expression, on suppose que j est petit par rapport à N et N_1 et si $\lim_{\substack{N\to+\infty \\ N_1\to+\infty}} \frac{N_1}{N} = p$ $(0 < p < 1)$, on trouve:

$$\lim P'[A_j] = \lim \frac{N_2(N_2-1)\ldots(N_2-j+2)\ N_1\ 1.2\ldots(N-j)}{1.2\ldots\quad (N-j)\ (N-j+1)\ldots\quad N}$$

$$= (1-p)^{j-1}\ p = q^{j-1}\ p.$$

Exercice 18. En utilisant un argument probabiliste, démontrer que pour a < b, on a :

$$\frac{a}{b} + \frac{a}{b}\ \frac{(b-a)}{b-1} + \frac{a}{b}\ \frac{(b-a)\ (b-a-1)}{(b-1)\ (b-2)} + \ldots + \frac{a}{b}\ \frac{(a-b)(a-b-1)\ldots 2.1}{(b-1)(b-2)\ldots\ (a+1)a} = 1.$$

En effet, si on pose : a = N_1, b = N, on voit de suite que :

$$P'(A_1) = \frac{a}{b} \qquad P'(A_2) = \frac{a}{b}\ \frac{b-a}{b-1} \quad \ldots \quad P'\left[A_{N_2+1}\right] = \frac{a}{b}\ \frac{(a-b)(a-b-1)\ldots 1}{(b-1)(b-2)\ldots(a+1)a}.$$

Comme $A_1, A_2, \ldots, A_{N_2+1}$ forment un système contradictoire, on a le résultat.

Exercice 19. Soient $X_1, X_2, \ldots, X_n$ n v.a. indépendantes ayant la même f.d. F.

1) Soit N la v.a. discrète qui représente le nombre de ces n v.a. qui sont $\leqslant x$ (x fixé $\varepsilon \mathbb{R}$) (De façon précise :)

$N(\omega) = k \rightleftharpoons$ k des n nombres $X_1(\omega), X_2(\omega), \ldots, X_n(\omega)$ sont $\leqslant x$.

Montrer que :

$$P[N=k] = C_n^k\ F(x)^k\ [1 - F(x)]^{n-k}$$

2) Posons $\xi_1 = \min(X_1, X_2, \ldots, X_n)$

$\xi_n = \min(X_1, X_2, \ldots, X_n)$

Montrer que :

$$P[\xi_n \leqslant x] = F(x)^n$$

$$P[\xi_1 \leqslant x] = 1 - [1-F(x)]^n$$

Exercice 20. Supposons avoir une suite indépendante (X_n) de v.a. ayant la même f.d.

1) Si X_n a pour f.fr.

$$f(x) = \begin{cases} e^{-(x-\alpha)} & x \geqslant \alpha \qquad \alpha \in \mathbb{R} \\ 0 & x < \alpha \end{cases}$$

montrer que $\underset{n \to +\infty}{p\lim} \xi_1(x) = \alpha$

En effet, $P[|\xi_1 - \alpha| \geqslant \varepsilon] = [1 - F(\alpha + \varepsilon)]^n \to 0$ quand $n \to +\infty$.

2) Si X_n est uniforme sur $(0, \beta)$, montrer que plim $\xi_n = \beta$

En effet,

$$P[|\xi_n - \beta| \geqslant \varepsilon] = P[\xi_n \leqslant \beta - \varepsilon] = F^n(\beta - \varepsilon) \to 0 \text{ quand } n \to +\infty.$$

3) Montrer que la suite (N_n) où N a été définie à l'exercice précédent satisfait la loi faible et la loi forte des grands nombres.

Exercice 21 : n personnes sont placées le long d'une table. Trouver la probabilité que deux personnes désignées soient séparées par k personnes.

1ère solution : Il y a n! dispositions possibles des personnes. On les suppose équiprobables. Si la première personne occupe la première place, l'autre doit occuper la k+2ième. On peut alors faire glisser les deux personnes d'une place. Ceci nous donne (n-k-1) dispositions favorables. Pour chacune de celles-ci, on peut encore permuter de (n-2)! façons différentes les n-2 personnes restantes. Enfin on peut permuter entre elles les deux personnes.

$$P = \frac{2\,(n-k-1)\,(n-2)\,!}{n\,!}$$

2de solution : Il y a C_n^2 façons possibles de choisir les 2 personnes. On les suppose équiprobables. Parmi celles-là il y en a n-k-1 où les 2 personnes sont séparées par k autres.

$$P = \frac{(n-k-1)}{C_n^2} \quad .$$

APPENDICE

Dans cet appendice, j'ai rassemblé une présentation de l'analyse combinatoire commode pour les applications en probabilité.

Quand on veut connaître le nombre de groupements que l'on peut former en extrayant j éléments d'un ensemble qui en contient n, il faut préciser :

1) si les n éléments sont différents ou non.

2) si un élément quelconque peut intervenir au plus une fois, exactement une fois ou au moins une fois dans un groupement.

3) si l'ordre de présentation des éléments dans un groupement est prépondérant ou non.

Nous étudierons trois types de groupements.

1. Arrangements.

Définition 1. On appelle arrangement sans répétition de n éléments différents pris j à j, tout groupement ordonné, formé de j de ces n éléments, dans lequel chaque élément figure une fois au plus.

Définition 2. On appelle arrangement avec répétition de n éléments différents pris j à j, tout groupement ordonné formé de j de ces n éléments, dans lequel chaque élément peut figurer jusqu'à j fois.

Remarques : la seconde définition n'impose aucune liaison entre n et j, mais la première stipule $j \leq n$.

Deux arrangements sont différents, soit parce qu'il existe au moins un élément qui figure dans l'un des arrangements sans figurer dans l'autre, soit, lorsqu'ils sont formés des mêmes éléments, parce que l'ordre de présentation des éléments est différent.

Exemple : On donne 4 éléments différents notés a, b, c, d. Les 12 arrangements sans répétition de ces 4 éléments pris 2 à 2, sont ab, ac, ad, ba, bc, bd, ca, cb, cd, da, db, dc. Les 16 arrangements avec répétition de ces 4 éléments pris 2 à 2 sont : aa, ab, ac, ad, bb, ba, bc, bd, ca, cb, cc, cd, da, db, dc, dd.

Proposition 1. Le nombre des arrangements sans répétition de n éléments différents pris j à j est donné par la formule :

$$A_n^j = n(n-1)\ (n-2)\ (n-3)\ \ldots\ (n-j+1).$$

qui se lit : le nombre des arrangements sans répétition de n éléments différents pris j à j est égal au produit de j nombres entiers consécutifs décroissants à partir de n.

Démonstration : Nous pouvons choisir le premier élément d'un groupement de n façons différentes ; le second de (n-1) façons différentes. Nous pouvons donc choisir les deux premiers éléments de n(n-1) façons différentes. Nous pouvons alors choisir le troisième élément de (n-2) façons différentes. Ainsi, nous pouvons choisir les trois premiers éléments de n(n-1)(n-2) façons différentes. Finalement, nous pouvons choisir le $j^{ième}$ élément de n-(j-1) façons différentes. Nous avons donc au total n(n-1)(n-2)... (n-j+1) choix possibles pour les j éléments.

Proposition 2. Le nombre des arrangements avec répétition de n éléments différents pris j à j est donné par la formule :

$$\alpha_n^j = n^j$$

Démonstration : Nous pouvons choisir le premier élément de n façons différentes. Mais nous pouvons encore choisir le second de n façons différentes puisque le premier élément choisi peut encore figurer en seconde position dans le groupement. Il en est de même pour chacun des j éléments. Nous avons donc en tout : n. n. n... n = n^j choix possibles.

2. Permutations.

Définition 3. On appelle permutation sans répétition de n éléments différents, tout groupement ordonné formé de ces n éléments.

Définition 4. On appelle permutation avec répétition de n éléments répartis en k groupes, le premier contenant n_1 fois l'élément a, le second contenant n_2 fois l'élément b, ..., le $k^{ième}$ contenant n_k fois l'élément g, tout groupement ordonné formé de ces n éléments.

Remarque : Deux permutations étant toujours formées des mêmes éléments, elles ne peuvent différer que par l'ordre de leurs éléments.

Proposition 3. Le nombre P_n des permutations sans répétition de n éléments différents est égal au produit des n premiers nombres entiers.

Démonstration : Les permutations sans répétition de n éléments différents ne sont rien d'autre que les arrangements sans répétition de n éléments différents pris n à n. Il suffit donc de faire j = n dans la proposition 1.

Notation : Le produit des n premiers nombres entiers est représenté par n ! qui se lit factorielle n. Nous avons donc : $p_n = n!$.

Proposition 4. Le nombre de permutations de la définition 4 est donné par :

$$\frac{n!}{n_1!\ n_2!\ \ldots n_k!}$$

Démonstration : Soit π le nombre de permutations de la définition 4. L'échange de deux lettres a dans une telle permutation ne fournit pas de permutation nouvelle. Affectons maintenant les lettres a d'un indice. Elles deviennent $a_1, a_2, \ldots, a_{n_1}$. Leurs permutations donnent $n_1!$ permutations nouvelles à partir de chaque permutation initiale. Les π permutations deviennent ainsi $\pi\, n_1!$ permutations dans lesquelles nous affectons maintenant les lettres b d'un indice. Nous obtenons ainsi $\pi\, n_1!\ n_2!$ permutations. Finalement, nous obtenons $\pi\, n_1!\ n_2!\ \ldots\ n_k!$ permutations qui sont toutes les permutations sans répétition de n éléments différents. Ainsi, nous avons :

$$\pi\ n_1!\ n_2!\ n_3!\ \ldots\ n_k! = n!$$

$$\pi = \frac{n!}{n_1!\ n_2!\ \ldots\ n_k!}$$

3. Combinaisons.

Définition 5. On appelle combinaison sans répétition de n éléments différents pris j à j, tout groupement non ordonné formé de j de ces n éléments dans lequel chaque élément figure une fois au plus.

Définition 6. On appelle combinaison avec répétition de n éléments différents pris j à j , tout groupement non ordonné formé de j de ces n éléments dans lequel chaque élément peut figurer jusqu'à j fois.

Remarques : Deux combinaisons ne seront différentes que si l'une d'elles contient au moins un élément qui ne figure pas dans l'autre.

Exemple : Soit a, b, c, d quatre éléments différents. Les 6 combinaisons sans répétition de ces 4 éléments pris 2 à 2 sont : ab, ac, ad, bc, bd, cd. Les 10 combinaisons avec répétition de ces 4 éléments pris 2 à 2 sont : aa, ab, ac, ad, bb, bc, bd, cc, cd, dd.

Proposition 5. Le nombre C_n^j des combinaisons sans répétition de n éléments différents pris j à j est égal au produit de n nombres entiers consécutifs décroissants à partir de n divisé par le produit des j premiers nombres entiers, soit :

$$C_n^j = \frac{n\,(n-1)\,(n-2)\ \ldots\ (n-j+1)}{j!}$$

Démonstration : Considérons une combinaison particulière. Elle est formée de j éléments différents. En permutant ces éléments de toutes les façons possibles, on trouve j ! groupements nouveaux qui seront tous les arrangements sans répétition des n éléments différents pris j à j mais ne contenant que les éléments de la combinaison particulière. En procédant de la même façon pour toutes les combinaisons, nous allons trouver tous les arrangements. Nous avons donc la relation :

$$C_n^j \; j! = A_n^j$$

d'où

$$C_n^j = \frac{A_n^j}{j!}$$

Remarque : En multipliant le numérateur et le dénominateur de la dernière relation par (n-j) !, nous ne modifions pas la valeur de C_n^j. Par conséquent :

$$C_n^j = \frac{n!}{j!\;(n-j)!}$$

Proposition 6. Le nombre γ_n^j des combinaisons avec répétition de n éléments différents pris j à j est donné par la relation :

$$\gamma_n^j = C_{n+j-1}^j = \frac{(n+j-1)!}{j!\;(n-1)!}$$

Démonstration : Supposons disposer d'une urne partagée en n parties par (n-1) cloisons. Convenons de réserver chacune des n parties à l'un des n éléments. Supposons disposer encore de j symboles (noté *). Le nombre de symboles qui se trouvent dans une partie de l'urne indique le nombre de fois que l'élément correspondant figure dans la combinaison. Ainsi, le schéma (n = 9, j = 7)

a b c d e f g h i

** * *** *

indique que l'on choisit la combinaison aacfffi. On ne peut obtenir une nouvelle combinaison qu'en permutant les étoiles et les cloisons. Or ces permutations avec répétition sont celles de n-1+j objets formés de n-1 cloisons et j étoiles dont le nombre est

$$\frac{(n+j-1)!}{(n-1)!\;j!}$$

Application : Si j ⩾ n, le nombre des combinaisons avec répétition des n lettres différentes prises j à j dans lesquelles les n lettres interviennent est

$$C_{j-1}^{n-1}$$

En effet, la condition que toutes les lettres interviennent implique que deux cloisons ne peuvent pas être adjacentes. Or

les j symboles laissent entre eux j-1 espaces parmi lesquels n-1 doivent être occupés par des cloisons, ce qui donne C_{j-1}^{n-1} façons de disposer les cloisons entre les symboles .

BIBLIOGRAPHIE

W. FELLER : An Introduction to Probability Theory and its Applications - 2 vol. J. Wiley-London.

Cet ouvrage doit être recommandé pour la qualité de la présentation mathématique, la richesse du contenu et les nombreux exemples. Le premier volume est limité aux variables aléatoires discrètes. Il ne présente guère plus de difficultés que les parties les plus faciles du texte. Le second volume est d'un niveau mathématique plus élevé.

E. PARZEN : Modern Probability Theory and its Applications. J. Wiley. London.

Cet ouvrage a les mêmes qualités que le précédent mais la matière couverte est beaucoup moins vaste.

A. RENYI : Calcul des probabilités avec un appendice sur la théorie de l'information. Dunod. Paris.

Cet excellent ouvrage constitue un cours de probabilité du niveau de la licence. Il contient de nombreux exercices.

J. NEVEU : Bases mathématiques du calcul des probabilités: Masson, Paris (traductions anglaise et allemande).

Un livre de très grande valeur mais de niveau troisième cycle. Suppose une formation mathématique poussée.

M.R. SPIEGEL: Théory and problems of Statistics : Schaum Publishing Co. New-York.

Recueil contenant 875 problèmes élémentaires résolus.

G. CALOT : Exercices de calcul des probabilités. Dunod, Paris.

Recueil d'exercices de difficultés diverses avec solutions.

TABLE DES MATIERES