STATISTIQUES
&
PROBABILITÉS

CHAPITRE 1

Notions Fondamentales de la Statistique Descriptive

1.1 **Définition de la statistique descriptive:**

La statistique descriptive est un outil scientifique qui permet de recueillir, organiser, classer, résumer et présenter les informations statistiques qualitatives ou quantitatives, concernant l’état ou la modification d’un phénomène.

1.2 Concepts de base de la Statistique Descriptive:

1.2.1 Population – Individus:

- La population est l’ensemble des unités statistique sur lesquels porte l’étude.
- l’individu est chaque élément de la population, les individus de la population peuvent être:

➢ Des êtres humains, ex:

❖ La population algérienne à la date du recensement.

❖ Les étudiants de cette université en 2016

➢ Des objets, ex:

❖ Production d’automobiles pendant l’année 2010

❖ Stock de pièces détachées d’une entreprise en fin de moi.

➢ Des événements, ex:

❖ Les accidents de la route survenus au cours du week-end en Algérie.

❖ L’ensemble des jours de l’année.

1.2.2 Caractères – Modalités:

- Chaque individu d'une population est décrit par un ensemble de caractéristiques appelées caractères.
- les modalités, ce sont les diverses situations de caractère.

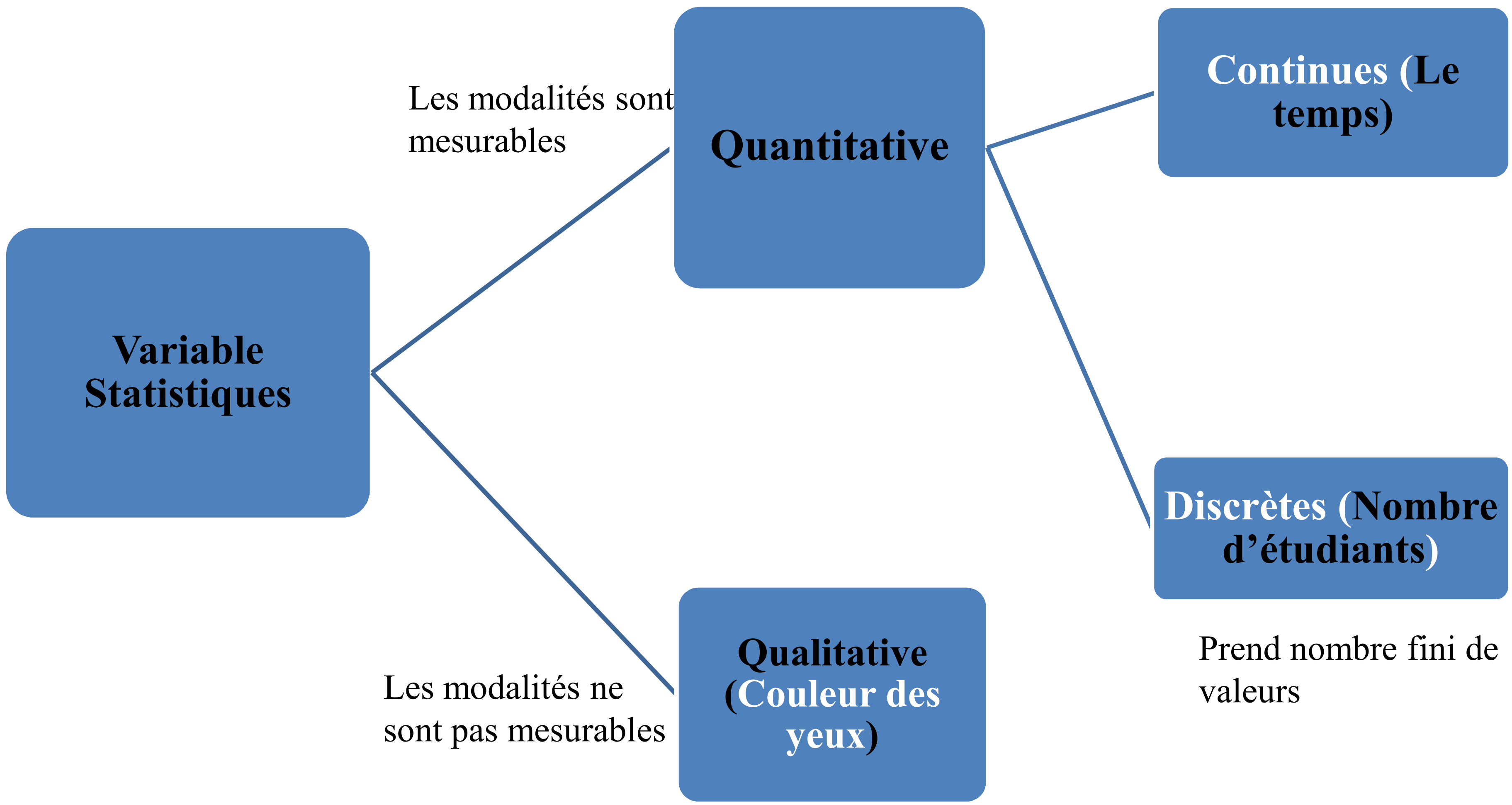
Exemples:

1. Pour décrire la population algérienne, on pourra retenir les caractères: sexe, âge, état matrimonial, lieu de résidence……. Etc.

- ➢ Le caractère « sexe » à deux modalités: masculin et féminin.

2. Personnel d'une entreprise, les caractères sont: le sexe, l'âge, la qualification, le salaire mensuel, l'ancienneté….. Etc.

Différents types de variables statistiques



Effectif n_i:On appelle effectif d'une valeur de caractère, le nombre de fois qu’apparaît cette valeur.

Fréquence f_i : On appelle fréquence d'une valeur de caractère, le division de l'effectif de cette valeur par l'effectif total.

$$f_i = \frac{n_i}{N}$$

Pourcentage : C’est la fréquence fois 100

Effectif cumulé

$$\text{Effectif cumulé croissant : } N_i^{\uparrow} = \sum_{k=1}^{i} n_k$$

$$\text{Effectif cumulé décroissant : } N_i^{\downarrow} = N - \sum_{k=1}^{i} n_k$$

Fréquence cumulée

Fréquence cumulée croissant : $F_i^{\uparrow} = \frac{\sum_{k=1}^{i} n_k}{N}$

Fréquence cumulée décroissant : $F_i^{\downarrow} = 1 - \frac{\sum_{k=1}^{i} n_k}{N}$

2.1. Tableaux statistiques:

2.1.1. Tableaux associé un caractère qualitatif:

Modalité (ci)	Effectif (ni)	Fréquence (fi)
C1 C2 . . Cp	n1 n2 . . np	f1 f2 . . fp
Total	$\sum_{i=1}^{p} n_i = N$	$F = \sum_{i=1}^{p} f_i = 1$

2.1.2. Tableaux associé un caractère quantitatif:

➢ Quantitatif discret:

Valeurs de la v.discret	n_i	f_i	$F_i \uparrow$ $F_0 = 0$	$F_i \downarrow$ $F_0 = 1$	$N_i \uparrow$ $N_0 = 0$	$N_i \downarrow$ $N_0 = N$
x1	n1	f1	F1=f1	F1=1-f1	N1=n1	N1=N-n1
x2	n2	f2	F2=f1+f2	F2=1-(f1+f2)	N2=n1+n2	N2=1-(n1+n2)
. .	. .	. .	. .	. .	. .	. .
xp	np	fp	$F_p = \sum_{i=1}^{p} f_i = 1$	$F_p = 1 - \sum_{i=1}^{p} f_i = 0$	$N_p = \sum_{i=1}^{p} n_i = N$	$N_p = N - \sum_{i=1}^{p} n_i = 0$
Total	$N = \sum_{i=1}^{p} n_i$	$\sum_{i=1}^{p} f_i = 1$				

Quantitatif continu:

Dans le cas d'une variable statistique continue, les observations sont nécessairement regroupées par classes.

$[e_i, e_{i+1}[$	Centre	n_i	f_i	$F_i \uparrow$ $F_0 = 0$	$F_i \downarrow$ $F_0 = 1$	$N_i \uparrow$ $N_0 = 0$	$N_i \downarrow$ $N_0 = N$
$[e_1, e_2[$	C_1	n_1	f_1	$F_1=f_1$	$F_1=1-f_1$	$N_1=n_1$	$N_1=N-n_1$
$[e_2, e_3[$	C_2	n_2	f_2	$F_2=f_1+f_2$	$F_2=1-(f_1+f_2)$	$N_2=n_1+n_2$	$N_2=1-(n_1+n_2)$
. .	. .		. .	. .	. .	. .	. .
$[e_{p-1}, e_p[$	C_p	n_p	f_p	$F_p = \sum_{i=1}^{p} f_i = 1$	$F_p = 1 - \sum_{i=1}^{p} f_i = 0$	$N_p = \sum_{i=1}^{p} n_i = N$	$N_p = N - \sum_{i=1}^{p} n_i = 0$
Total	/	$N = \sum_{i=1}^{p} n_i$	$\sum_{i=1}^{p} f_i = 1$				

- ✓ Le centre d'une classe est

$$C_i = \frac{e_i + e_{i+1}}{2}$$

- ✓ L'amplitude d'une classe est: $\mathbf{a_i = e_{i+1} - e_i}$

Représentations graphiques

Qualitative

Diagramme à bonde

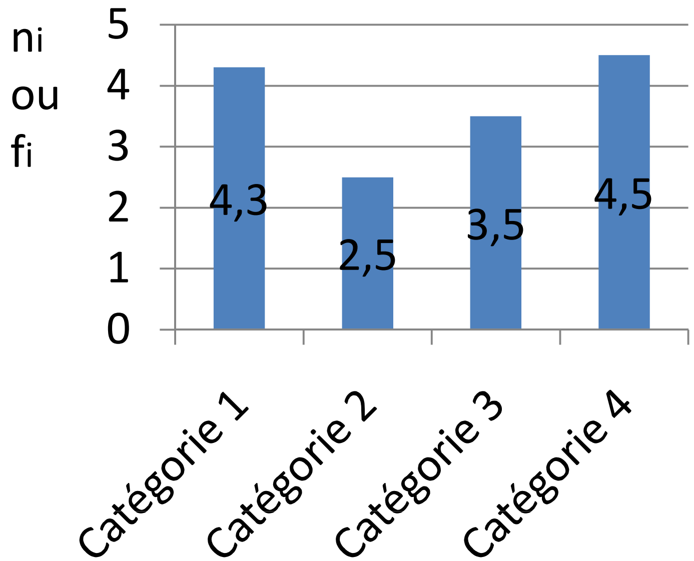
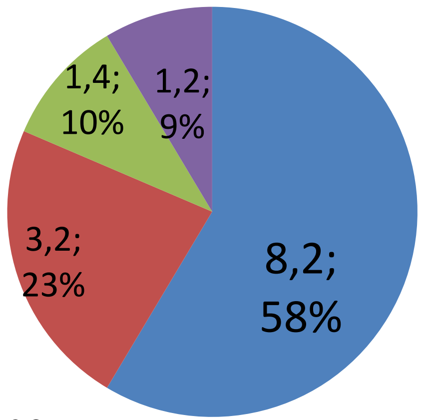


Diagramme circulaire

$$\alpha_{\mathrm{i}} = 360° \times \frac{n_i}{N} = 360° \times f_i$$

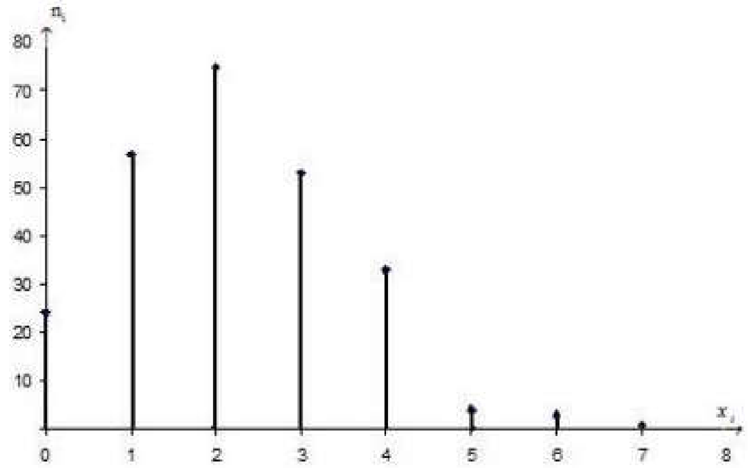


$$\sum \alpha_i = 360°$$

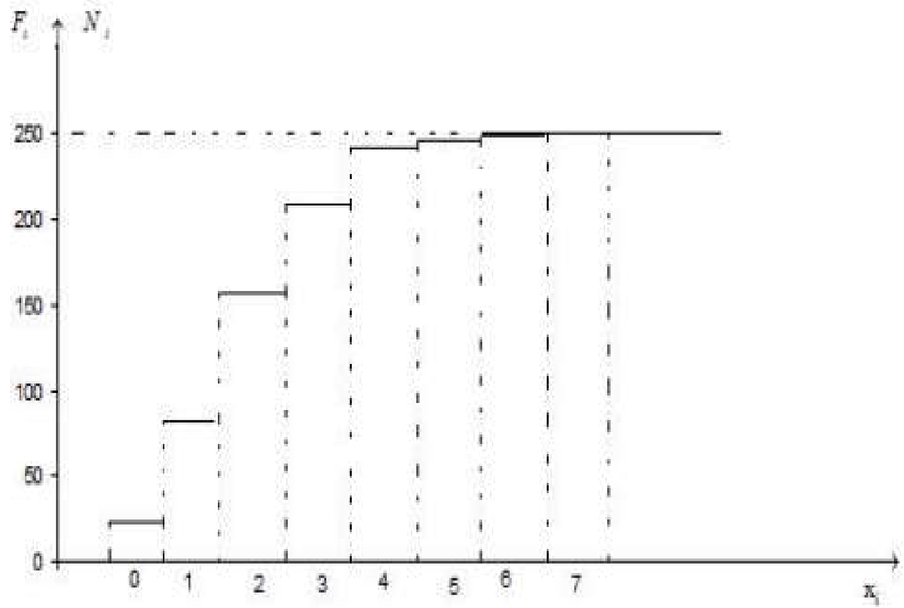
quantitatif

Discrètes

Diagramme en bâton



Diagrammes intégrales (courbe en escaler)



- Fonction repartions dans le cas variable discret:

$$F(x) = \begin{cases} 0 & x \langle x_1 \\ F_i & x_i \leq x \langle x_{i+1} \\ 1 & x \geq x_p \end{cases}$$

- Fonction repartions dans le cas variable continue:

$$F(x) = \begin{cases} 0 & x \langle e_1 \\ F_{i-1} + \dfrac{f_i}{e_i - e_{i-1}}(x - e_{i-1}) & e_{i-1} \leq x \langle e_i \\ 1 & e_p \leq x \end{cases}$$

- Variable quantitative continue:

Le nombre de classes est calculé par la règle de « STURGES » suivante:

$$a = \frac{x_{\max} - x_{\min}}{1 + 3.3 \log N}$$

a: l’amplitude recherchée.

X_{max} : la plus grand valeur de la variable x.

X_{min} : la plus petite valeur de la variable x.

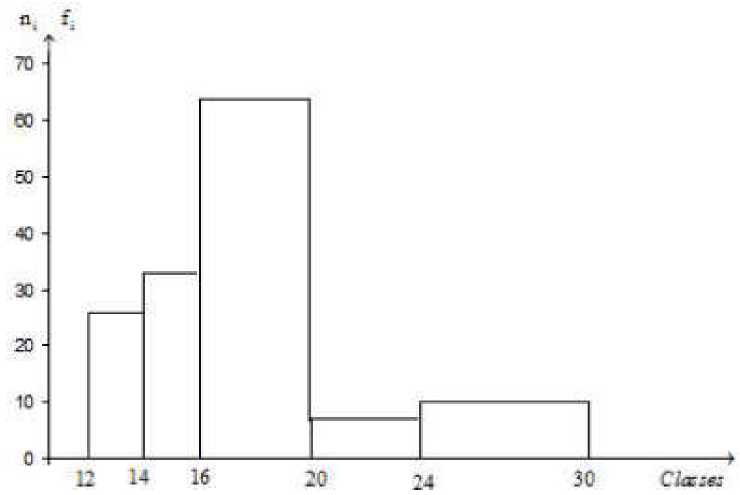
N : le nombre d’observation.

log : le logarithme décimale.

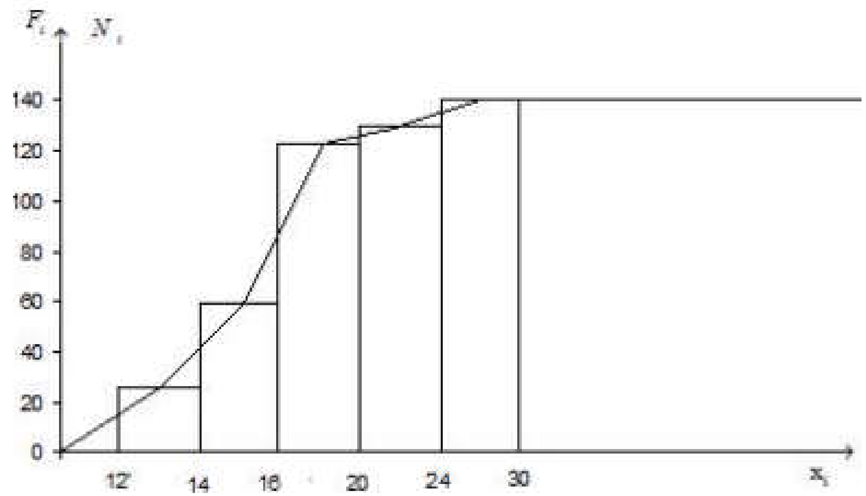
quantitatif

continue

Histogramme



Diagrammes intégrales



Paramètres de tendance centrale (position)

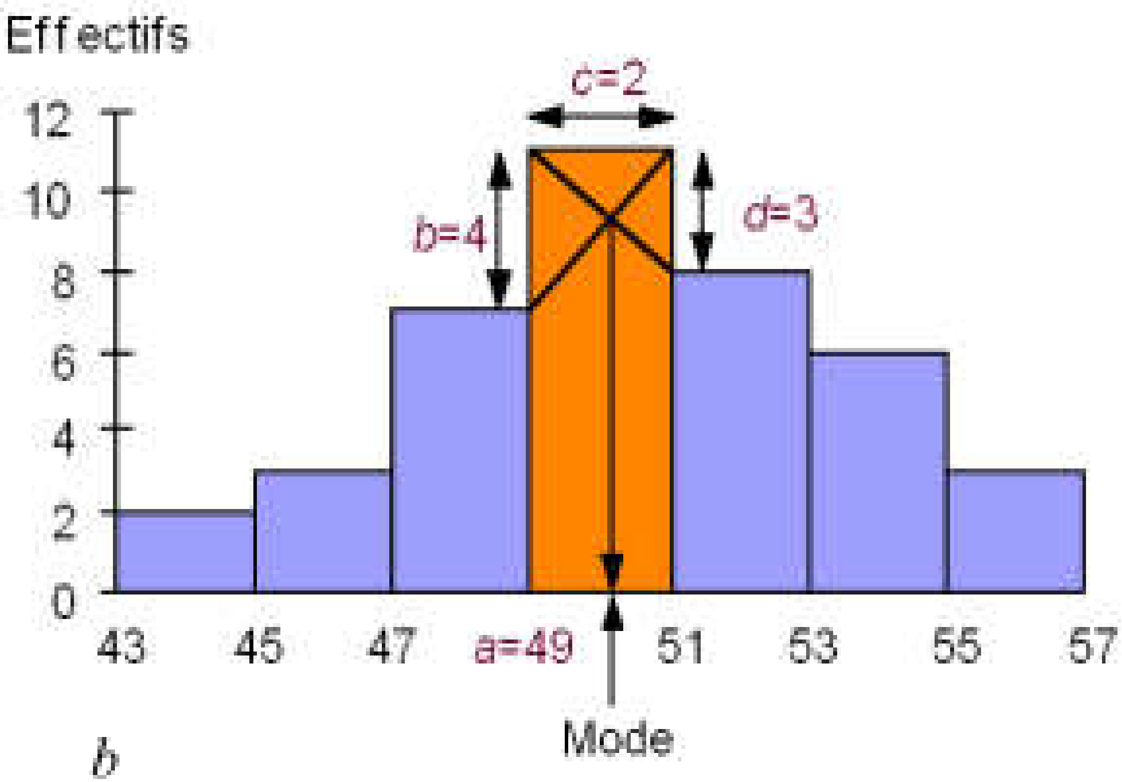
- Le mode
- La moyenne arithmétique
- La moyenne harmonique
- La moyenne géométrique
- La moyenne quadratique
- La médiane
- Les quantiles

Le mode

- Si k vérifie $n_k = \max_{i\in 1,p} (n_i)$ alors on dit que x_k est le mode de la série statistique discret.
- Si k vérifie $n_k = \max_{i\in 1,p} (n_i)$ alors on dit que $[e_{k-1}, e_k[$ est la classe modale de la série statistique continue.
- Détermination du mode:

$$Mo = e_{i-1} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot a_i$$

- Mo: la valeur du mode.
- e_{i-1}: l'extrémité inferieur de la classe modale.
- a_i: l'amplitude de la classe modale.
- $\Delta 1$: la différence entre l'effectif ou la fréquence de la classe précédente et celle de la classe modale.
- $\Delta 2$: la différence entre l'effectif ou la fréquence de la classe modale et celle de la classe suivante.



$$\textbf{mode} = a + c \cdot \frac{b}{b+d}$$

$$\textbf{mode} = 49 + \frac{2 \cdot 4}{4+3} = 50.14.$$

- La moyenne arithmétique pondérée $$\overline{X} = \frac{1}{N}\sum_{i=1}^{p} n_i \cdot x_i$$

- La moyenne harmonique $$\frac{1}{H} = \frac{1}{N}\sum_{i=1}^{p} n_i \frac{1}{x_i}$$

- La moyenne géométrique $$\overline{X} = \frac{1}{N}\sum_{i=1}^{p} n_i \log x_i$$

- La moyenne quadratique $$Q = \sqrt{\frac{1}{N}\sum_{i=1}^{p} n_i x_i^{2}}$$

Médiane

La médiane que l'on not Me, correspond à la valeur de la variable statistique qui partage la population en deux parties égales.

$N(Me)=N/2$ ou $F(Me)=1/2=50\%$

Pour calculé la médiane, on distingue deux cas:

- CAS 1: Soit N, un nombre impair. exemple: 1, 3, 4, 4, 6. $N=5$

On utilise la formule suivante: $(N+1)/2=6/2=3$

La médiane est la 3eme valeur soit 4 (c'est la valeur du milieu).

- CAS 2: Soit N, un nombre pair. exemple: 1, 3, 4, 4. $N=4$

On fait deux calculs:

$N/2$ et $(N/2)+1$ $N/2=2$, $N/2+1=4/2+1=3$

On prend la 2eme et la 3eme valeur et on effectue la demi somme: $(3+4)/2=3.5$

La médiane est 3.5.

- **Si la variable statistique discret**

On détermine la médiane, soit à partir des fréquences cumulées des tableaux statistiques, soit à partir du graphique cumulatif.

S'il existe un i tel que $F_i\ 1 < 0.5 < F_i$ alors la médiane est x_i.

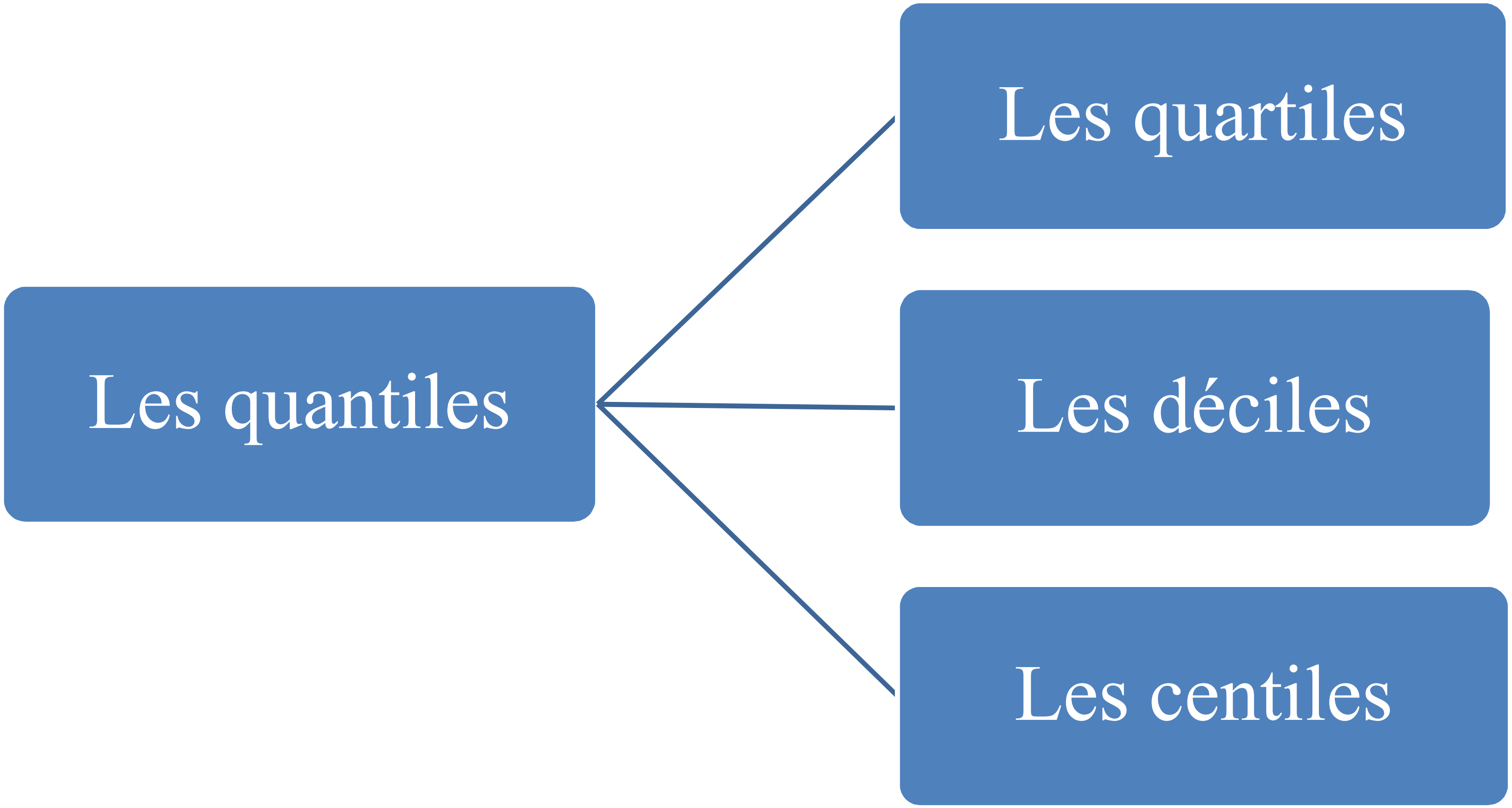
- <u>**Si la variable statistique continue**</u>

Soit $([e_{i\ 1}, e_i[, n_i)_{i=1,p}$ une série statistique continue. La médiane, noté Me est l'abscisse du point du polygone des fréquences cumulées dont l'ordonnée vaut 0,5

$$Me = e_i + a_i \cdot \frac{0.5 - F_i}{f_i}$$

- e_i: l'extrémité inferieur de la classe médiane.
- a_i: l'amplitude de la classe médiane.
- F_i: la fréquence cumulée de la classe qui précède la classe médiane.
- f_i: la fréquence de la classe médiane $[e_{i\ 1}, e_i[$

Les quantiles



- ***<u>Les quartiles</u>*** divisent l'effectif ou la fréquence de la série, en quatre parties égales.

F(Q1)=25% N(Q1)=N/4

F(Q2)=50% N(Q2)=N/2

F(Q3)=75% N(Q3)=3N/4

- ***<u>Les déciles</u>*** partagent la population on dix groupes comprenant chacun 10% des observations.

$F(D_1)=10\%$ $N(D_1)=N/10$

$F(D_2)=20\%$ $N(D_2)=2N/10$

…………... ………………

$F(D_9)=90\%$ $N(D_9)=9N/10$

- ***<u>Les centiles</u>*** partagent la population on cent groupes comprenant chacun 1% des observations.

$F(P_1)=1\%$ $N(P_1)=N/100$

$F(P_2)=2\%$ $N(P_2)=2N/100$

…………... ……………

$F(P_{99})=99\%$ $N(P_{99})=99N/100$

Données quantitatives continues

$$X_k = e_i + a_i \cdot \frac{F(x_i) - F(e_i)}{f_i}$$

- X_k : les quartiles ou les déciles ou les centiles.
- ei: l’extrémité inferieur de la classe.
- ai: l’amplitude de la classe.
- Fi: la fréquence cumulée de la classe qui précède la classe.
- fi: la fréquence de la classe [$e_{i\ 1}$, e_i[

Paramètres de dispersion

- Etendue
- Ecarts inter-quantiles
- La variance
- L’écart type
- L’écarts absolu moyen
- Coefficients de variation

Etendue

1. L'étendu de la série statistique discrète est la différence entre la plus grande et la plus petite valeur observée.

$$\boxed{E = x_{max} \quad x_{min}}$$

2. L'étendu de la série statistique continue $([e_{i\ \ 1}, e_i[, n_i)_{i=1,p}$.

$$\boxed{E = e_{max} \quad e_{min}}$$

Ecarts inter-Quantiles

Les écarts mesurent la dispersion autour des valeurs centrales de la série statistique notamment autour de la médiane

1. *Ecarts inter-quartiles* $I_Q = Q_3 - Q_1$

   il contient 50% des observations

2. *Ecarts inter-déciles* $I_D = D_9 - D_1$

   il contient 80% des observations

3. *Ecarts inter-centiles* $I_P = P_{99} - P_1$

   il contient 98% des observations

Ecart Arithmétique (Ecart Absolus Moyen)

1. Les écarts absolus moyen définir par:

$$\overline{e} = \frac{1}{N}\sum n_i \left| x_i - \overline{X} \right|$$

1. L’écart absolu moyen de la série statistique continue

$$\overline{e} = \frac{1}{N}\sum n_i \left| c_i - \overline{X} \right|$$

Variance

- **<u>La variance</u>** est donnée par les formules suivantes:

$$V = \frac{1}{N}\sum ni\,(xi - \overline{X})^2 = \sum f_i x_i^2 - \overline{X}^2$$

L’écart -type

l'écart - type définir par:

$$\boxed{\sigma_x = \sqrt{Var(x)}}$$

Plus l’écart – type σ est grand, plus les valeurs du caractère sont dispersées autour de la moyenne

Plus il est petit, plus les valeurs du caractère sont groupées autour de la moyenne.

Coefficients de variation

Le coefficients de variation définir par:

$$CV = \frac{\sigma}{\overline{X}}$$

CHAPITRE 2

Calculs des Probabilités

I. Analyses combinatoire

Principe fondamentale de dénombrable

Théorème

Si une expérience peut se réaliser de $n1$ façons, une deuxième de $n2$ façons,…, une $k_{éme}$ de n_k façons, alors la séquence des K opérations peut se réaliser de $n_1 \times n_2 \times \ldots \times n_k$ façons.

On peut résoudre ce problème, en utilisant une autre méthode qui est l'arbre d'étalement.

Exemple1

La répartition d'une population selon le sex et la situation matrimoniale ,est la suivante:

Méthode 1: le dénombrement:

Sexe(F,M) ,donc $n_1=2$, et Situation matrimoniale(M,C,V,D)

donc $n2=4$,Alors la répartition est $n_1 \times n_2 = 2 \times 4 = 8$ façons.

Méthode 2:l'arbre d'étalement:

1. Arrangements sans répétition

- Soit n objets distincts. On appelle un arrangement une manière de sélectionner p objets parmi les n et de les ranger dans des boîtes numérotées de 1 à p.
- Dans la première boîte, on peut mettre chacun des n objets. Dans la seconde boîte, on peut mettre chacun des n 1 objets restants, dans la troisième boîte, on peut mettre chacun des n 2 objets restants et ainsi de suite. Le nombre d'arrangements possibles est donc égal à :
- $1 \leq p \leq n$

$$A_n^p = n \times (n-1) \times (n-2) \times \ldots \times (n-P+1) = \frac{n!}{(n-p)!}$$

2. Arrangements avec répétition

Lorsqu'un objet peut être observé plusieurs fois dans un arrangement, le nombre d'arrangement avec répétition de p objets pris parmi n, est alors :

$$AR_n^p = n^p$$

3. Permutations sans répétition

- Une permutation sans répétition est un classement ordonné de n objets distincts. Considérons par exemple l'ensemble {1, 2, 3}. Il existe 6 manières d'ordonner ces trois chi res :

{ 1, 2, 3} , { 1, 3, 2} , { 2, 1, 3} , { 2, 3, 1} , { 3, 1, 2} , { 3, 2, 1} .

- Si on dispose de n objets, chacun des n objets peut être placé à la première place.
- Il reste ensuite n 1 objets qui peuvent être placés à la deuxième place, puis n 2 objets pour la troisième place, et ainsi de suite. Le nombre de permutations possibles de n objets distincts vaut donc

$$n \times (n \quad 1) \times (n \quad 2) \times \cdot \cdot \cdot \times 2 \times 1 = n!$$

$$P_n = n!$$

4. Permutations avec répétition

- Si l'on dispose de n objets appartenant à p groupes de tailles $n_1, n_2, \ldots, n_p$, le nombre de permutations avec répétition est

$$P_n^{n_1, n_2, \ldots, n_p} = \frac{n!}{n_1! \times n_2! \ldots \times n_p!}$$

5. Combinaisons sans répétition

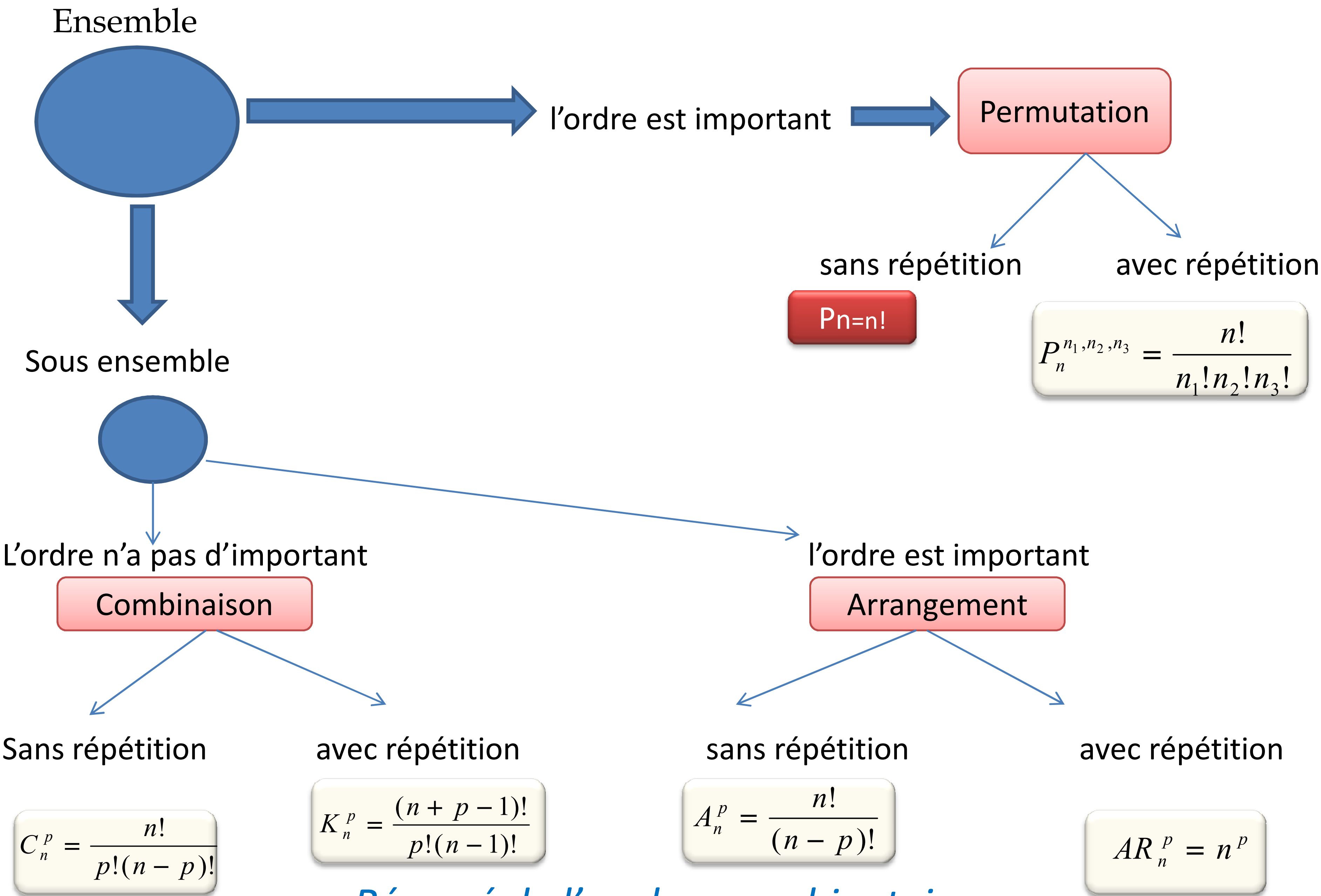
On appelle combinaisons sans répétition de p objets tout ensemble de p objets pris parmi les n objets sans remise. Le nombre de combinaisons de p objets pris parmi n est

$$C_n^p = \frac{n!}{p!(n-p)!}$$

6. Combinaisons avec répétition

On appelle combinaisons avec répétition de p éléments choisis parmi les n éléments. Une disposition non ordonne de p éléments dans laquelle chaque élément peut figurer plus qu'une :

$$K_n^p = C_{n+P-1}^p = \frac{(n+p-1)!}{p!(n-1)!}$$



Résumé de l'analyse combinatoire

Formule de Pascal

Les termes du triangle de Pascal résultent de l'application directe de cette relation.

	0	1	2	3	4	,,,,,,,,	p-1	p
0	0	0	0	0	0		0	0
1	1	1	0					
2	1	2	1	0				
3	1	3	3	1	0			
4	1	4	6	4	1	0		
,,,,,,,,								
n-1							C_{n-1}^{p-1}	C_{n-1}^{p}
n								C_{n}^{p}

4.3. Formule du binôme de Newton

La formule du binôme de Newton correspond à la décomposition des différents termes de la puissance $n^{ième}$ du binôme (a+b).

$$\forall (a,b) \in R^2, n \in N, (a+b)^n = \sum_{p=0}^{n} C_n^p a^{n-p} b^p$$

II. Espace Probabilisable

I. **Expérience aléatoire*:***

Une expérience aléatoire est une expérience dont les résultats (éventualités ou issues possibles) ne sot pas connus à priori.

L'ensemble des issues possibles d'une expérience aléatoire s'appelle l'univers (espace fondamental), ou espace des événements, on le not Ω.

II. Evénements

C'est l'ensemble de tous les résultats caractérisés par une même propriétés lors d'une expérience aléatoire, c'est partie A de Ω.

- Les événements sont représentés par des lettres majuscule: A, B,…, A1,…, Ap,….etc.

1. Evénements élémentaires:

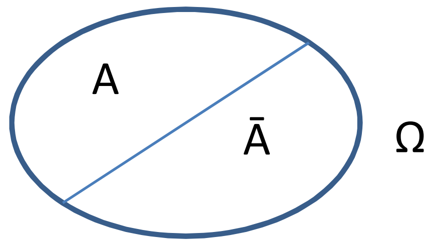
c'est une partie de Ω qui ne contient qu'un seul élément, exemple: obtenir le chiffre 6 dans le cas de lancer un dé.

2. **Evénements composés:**

C'est un ensemble des événements élémentaires, <u>exemple</u>: dans l'expérience de lancer un dé, l'événement d'avoir un chiffre pair est un événement composé de 3 événement élémentaires sont {2}, {4}, {6}.

3. **Evénements contraire ou complémentaire:**

De A est $\bar{A}$ qui contient toutes les éventualités de Ω qui ne sont pas dans A.



4. **Evénements compatibles et événements incompatibles:**

A et B deux événements incompatibles (disjoints ou distincts), s'ils ne se réalisent pas ensemble, c'est-à-dire: $A \cap B = \Phi$

➢ Relations et opérations sur les événements:

1. Intersection d'événements:

L'intersection des événements A et B est constitué des issues appartenant à la fois à A et B, c'est événement noté $A \cap B$ est appelé l'événement «A et B»

2. Réunion d'événements:

La réunion des événements A et B est constitué des issues appartenant à A ou B, c'est événement noté AUB est appelé l'événement «A ou B»

3. Inclusion d'événements:

Un événements A inclus dans l'événement B, si la réalisation de A implique celle de B, on dit que A inclus dans B $(A \subset B)$

4. Système complet d'événements:

n événements: $A_1, A_2, \ldots, A_n$, constituent un système complet d'événements si et seulement si $A_1, A_2, \ldots, A_n$ forment une partition de Ω.

III. Tribu des événements:

Une famille $\mathcal{A}$ de parties de Ω est une tribu sur Ω (algèbre de événements), si elle vérifie les 3 conditions suivantes:

1. $\phi, \Omega \in \mathrm{A}$
2. $\text{si } A \in \mathcal{A} \Rightarrow \overline{A} \in \mathcal{A}$
3. $A_1, A_2, \ldots, A_n \in \mathcal{A} \Rightarrow A_1 \cup A_2 \cup \ldots \cup A_n \in \mathcal{A}$

IV. Espace probabilité:

1. Définition de la probabilité:

P(A) représente la chance qu'il a de se réaliser :

A: un événement

P(A): la probabilité de réaliser l'événement A.

$$P(A) = \frac{n}{N} = \frac{\text{Nombre de cas favorables}}{\textit{Nombre} \text{ de cas possibles}}$$

➢ Propriétés de la probabilité:

1. $P(\Omega)=1$
2. $P(\Phi)=0$
3. $1 \leq P(A) \leq 0$
4. $P(A)=1-P(\bar{A})$
5. Pour tout couple (A, B) d'événements incompatibles : $P(AUB)=P(A)+P(B)$
   $(P(A \cap B)=\Phi)$
6. Pour tout couple (A, B) d'événements compatibles:

   $P(AUB)=P(A)+P(B)-P(A \cap B)$

   $P(A \cap B)=P(A)+P(B)-P(AUB)$
7. A, B, C trois événements compatibles:

- ✓ $P(AUBUC)=P(A)+P(B)+P(C)-P(A \cap B)- P(A \cap C)- P(B \cap C)+ P(A \cap B \cap C)$
- ✓ $P(A \cap B \cap C)=P(A)+P(B)+P(C)-P(AUB)- P(AUC)- P(BUC)+ P(AUBUC)$

Probabilité sur un ensemble à événements élémentaires équiprobables:

- Les événements A et B sont dits équiprobables si et seulement si $P(A)=P(B)$

Probabilité conditionnelle:

1. Définition:

Soient A et B deux événements de l'ensemble fondamental Ω, avec B de probabilité non nulle ($P(B)\neq 0$), on appelle probabilité conditionnelle de A sachant B, notée $P(A|B)$, la probabilité de réalisation de l'événement A sachant que l'événement B s'est réalisé, sa valeur est donnée par:

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

Avec $P(B)\neq 0$

2. Probabilités composées:

A partir de la définition de la probabilité conditionnelle, on peut en déduire un résultat intéressant:

$$P(A\cap B)=P(A|B).P(B)=P(B|A).P(A) \quad \text{avec } P(B)\neq 0 \text{ et } P(A)\neq 0$$

Ce résultat peut être généralisé au cas suivant:

$$P(A\cap B\cap C)=P(A|B\cap C).P(B\cap C)=P(A|B\cap C).\,P(B|C).P(C)$$

3. Probabilité totales:

Si B_1 , B_2 ,…,B_n forment un système complet de Ω

$\forall A \subset \Omega, (A \cap B_1),...,(A \cap B_n)$ forment un système complet de A.

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + ... P(A \cap B_n)$$

$$P(A) = \sum_{k=1}^{n} P(A \cap B_k)$$

4. Formule de Bayes:

Soit B_1 , B_2,…, B_n un système complet de l'ensemble fondamental Ω associé à une expérience aléatoire, et soit l'événement A dans Ω, on a:

$$P(B_i | A) = \frac{P(A|B_i).P(B_i)}{P(B_1).P(A|B_1) + P(B_2).P(A|B_2) + ... + P(B_n).P(A|B_n)}$$

➢ Indépendance stochastique:

1. Indépendance de deux événements:

On dit que l'événement A est indépendant de l'événement B si :

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

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

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

2. Indépendance de plusieurs événements:

A , B, C sont 3 événements indépendants si et seulement si :

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

$P(A \cap C) = P(A).P(C)$

$P(B \cap C) = P(B).P(C)$

2) $P(A \cap B \cap C) = P(A).P(B).P(C)$