Université Frères Mentouri Constantine 1
Faculté des sciences de la nature et de la vie

Département de Biologie Appliquée

Licence : Apiculture

Année Universitaire : 2019/2020

Cours de Statistique

Dr. Habiba BOUHALLOUF

Statistique Descriptive

1. Eléments et données statistiques

1.1 Introduction

Le terme *statistique* est issu du latin *status*, c'est-à-dire *état* et le mot *statisticum* apparaît à la fin du XVII éme siècle et veut dire «qui a trait à l'État» On appelle *statistique* l'ensemble de méthodes scientifiques permettant de collecter, décrire et analyser des données observées. Ces observations consistent généralement en la mesure d'une ou plusieurs caractéristiques communes sur un ensemble de personnes ou d'objets équivalents. La statistique descriptive est la partie de la statistique qui sert à décrire de façon synthétique et parlante un phénomène (mesurer, classer les mesures et présenter ces mesures par quelques indicateurs).
La statistique descriptive traite des propriétés des population plus que des individus particuliers de ces populations.

1.2 Objectifs

L'étudiant sera capable de :

- organiser les données observées.
- Regrouper ses données sous forme de tableaux et de graphes.
- Réduire ses données sous forme de paramètres et d'indicateurs.
- Expliquer les résultats obtenus.

1.3 Variables

En statistique, l'élément statistique est le caractère retenu pour étudier un ensemble donné, on l'appelle *variable*. Une variable est un facteur susceptible de prendre une valeur différente selon les unités statistiques étudiées (la taille, la couleur des yeux, ...). Selon les valeurs que peut prendre une variable, on les différencie en deux catégories : quantitatives et qulitatives.

1.3.1 Variables quantitatives

Sont les variables qu'on peut mesurer, elles sont caractérisées par des valeurs numériques. Elles peuvent être :

Continues : Ce sont les variables qui peuvent prendre toutes les valeurs d'un ensemble ou d'un intervalle d'observations. Ces valeurs appartiennent à l'ensemble des nombres réels (une infinité de valeurs possibles). Ce genre de variable est particulièrement utilisé en médecine et en biologie médicale.
Exemple :
$T = 1,74m$, ici T représente la variable *taille*, 63 représente sa valeur et m est l'unité de mesure (associée à l'instrument de mesure).

Discrètes : Ce sont des variables numériques discontinues qui ne peuvent prendre que quelques valeurs dans un intervalle de nombres entiers donné.
Exemple : soit la variable *rappel vaccin* et sa valeur égale à *3* où l'unité de mesure est *injections*.

Temporelle : Ce sont des variables quantitatives particulières qui utilisent les unités de mesure du temps. Il existe deux types, le type *date* et le type *horaire*.
Exemple : date de naissance d'un bébé : *01/01/2015* (type *date*), heure de consommation d'un produit : *24 h* (type *horaire*).

1.3.2 Variables qualitatives

Ce sont des variables qui ne sont pas mesurables (n'ont pas de valeurs numériques). Leurs valeurs sont des qualités réparties en classes, chaque classe a son propre effectif dénombré. On peut différencier les valeurs qualitatives en trois groupes :

Ordinales : Elles s'expriment en classes qui peuvent être ordonnées selon une échelle de valeurs.
Exemple : variable : *compilation d'une maladie*, classes : *modérée, moyenne, sévére*.

Nominales : Ce sont des variables dont les classes ne sont pas ordonnées.
Exemple : variable : *groupe sanguin*, classes : *A, B, O, AB*.

Binaires : ce type est un cas particulier des variable qualitatives nominale où la variable ne peut prendre que deux valeurs possibles, elle peut être *dichotomiques*, *booléennes* ou *variable de Bernoulli*.
Exemple : variable *sexe*, classe *homme, femme*.

1.4 Vocabulaire statistique

1.4.1 Population

La population correspond à l'ensemble des individus ou d'objets de même nature sur lequel porte l'étude. Vu sa taille très grande, il est généralement dificile, voire impossible, d'observer toutes les données. Au lieu d'examiner la population, on examine une petite partie qu'on appelle *échantillon*. Une population peut être réelle ou fictive.

1.4.2 Modalité statistique

On appelle une *modalité* (ou catégorie) les différentes situations (niveaux) possibles d'une variable qualitative.

1.4.3 Effectif total

On appelle effectif total le nombre total d'individus dans la population.

1.4.4 Effectif

L'effectif (fréquence absolue) est le nombre des éléments statistiques relatifs à une modalité donnée, noté f.Autrement dire, le nombre d'individus qui correspondent au même caractère.

1.4.5 Effectif cumulé

On appelle effectif cumulé croissant le nombre d'individus qui correspondent au même caractère (modalité) et aux caractères précédents. On l'utilise pour la distribution des variables quantitatives dans la détermination des paramètres statistiques.

1.4.6 Fréquence

On appelle fréquence (fréquence relative), le rapport entre l'effectif d'une valeur et l'effectif total. C'est la part des effectifs d'une modalité.

1.4.7 Série statistique

la série statistique est toujours issue d'un relevé d'élèments dans une population donnée. Quand on regroupe et on ordonne les données, on obtient une série statistique. Elle fait correspondre les différentes modalités et leurs effectifs ou fréquences.

1.4.8 Classe (Intervalle)

On appelle classe un groupement de valeurs d'une variable selon des intervalles qui peuvent être égaux ou inégaux. On l'utilise surtout lorsque la variable édiée est quantitative continue. Pour chaque classe on peut définir :

- Une limite inférieure
- Une limite supérieure
- Intervalle de classe (amplitude)= limite (sup)- limite (inf)
- Centre de classe = [limite (sup) + limite (inf)]/2

1.4.9 Cumul

La signification de ce mot : *cumuler*, c'est additionner des valeurs, des données, ... etc. Concernant une série statistique, pour calculer un effectif cumulé, il suffit d'ajouter à l'effectif (ou la fréquence ou le pourcentage) d'une valeur d'un caractère, le ou les effectifs (ou les fréquences ou les pourcentages) des valeurs précédentes.
Exemple : les notes (sur 20) de 31 élèves d'une classe :

TABLE 1.1 – Effectifs, fréquences et pourcentages cumulés

Note	Effectif (f_i)	f_i cumulé	Fréquence (F_i)	F_i cumulée	pourcentage (%)	% cumulé
2	1	1	0,032	0,032	3,2	3,2
6	3	4	0,097	0,129	9,7	12,9
8	3	7	0,097	0,226	9,7	22,6
9	7	17	0,226	0,452	22,6	45,2
10	6	20	0,194	0,646	19,4	64,6
11	5	25	0,161	0,807	16,1	80,7
12	3	28	0,097	0,904	9,7	90,4
14	2	30	0,065	0,969	6,5	96,9
16	1	31	0,032	1	3,2	100

1.5 Organisation des données

Dans une étude statistique, la première étape consiste à collecter des données portant sur une série des unités statistiques. Dans la deuxième étape, on doit trier ces données en les regroupant sous forme une matrice.

1.5.1 Tri des données

Le tri consiste à ranger les unités statistiques par ordre croissant ou décroissant des valeurs de la variable quatitative étudiée. Si la variable est qualitative, les unités statistiques sont regroupées selon les différentes classes de cette variable.

1.5.2 Regroupement en classe

Lorsqu'on étudie une variable quantitative sur une série ayant un nombre important d'individus, il est nécéssaire de construire une échelle de classification en divisant la série en classes, en intervalles égaux. Dans ce cas la variable sera de type discret.

1.6 Description des données

Elle dépend du type de la variable étudiée. Il existe deux formes de présentation pour décrire une distribution ou une série de données statistiques sont : les tableaux et les représentations graphiques.

1.6.1 Tableaux

Le tableau est utilisable quelle que soit la nature des données, il sert à présenter les données d'une façon exacte. Il permet de présenter de façon complète et précise les données. Les données peuvent être présentées variable par variable et unité par unité dans un tableau brut. Elles peuvent aussi être présentées dans un tableau de fréquences ou les variables sont regroupées en classes.
Exemple : Table 1.2

TABLE 1.2 – série de valeurs (en classes) des tailles d'adolescents de 11 à 14 ans

Tailles (cm)	**Effectif de classe**
$[144, 148[$	18
$[148, 152[$	65
$[152, 156[$	85
$[156, 160[$	71
$[160, 164[$	32
$[164, 168[$	20
$[168, 172[$	4
$[172, 176[$	1

1.6.2 Graphiques

L'objectif des graphiques est de faire ressortir une vision systhématique du phénomène étudié en illustrant une tendance générale et en donnant une image globale des résultats :

Histogramme

Les histogrammes sont des surfaces qui permettent la représentation d'une variable quantitative continue.
L'aire de chaque surface (colonne sur l'axe horizontal) est égale à l'effectif correspondant à une classe. La hauteur de la colonne (sur l'axe vertical) indique les valeurs, le nombre de valeurs appartenant à cette classe, la fréquence correspondante apparaît dans la distribution. La somme de toutes les aires est égale à l'effectif total (la taille de l'échantillon).

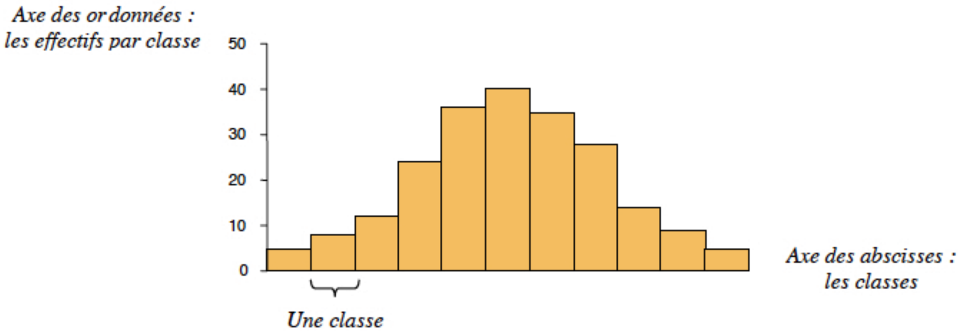

FIGURE 1.1 – Histogramme

Polygone de fréquence

Un polygone statistique est un diagramme formé en reliant les points médians des colonnes d'un histogramme. Ces diagrammes sont utilisés seulement pour présenter des données concernant des variables continues d'un histogramme. Le polygone statistique aplanit les changements abruptes pouvant être visibles dans un histogramme. Il est utile pour montrer la continuité de la variable à l'étude.

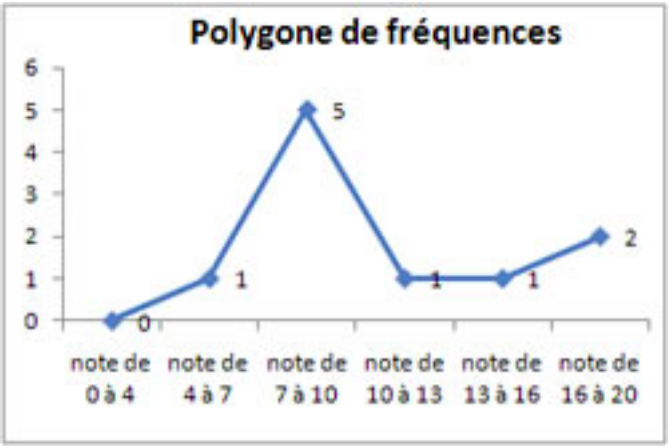

FIGURE 1.2 – Polygone de fréquences

Diagramme en barres

Un diagramme en barres est une représentation graphique réservée surtout pou la distribution d'une variable qualitative à l'aide de rectangles de même largeur. Les valeurs de la variable étudiée sont représentées sur l'axe horizontal, les effectifs sur l'axe vertical.

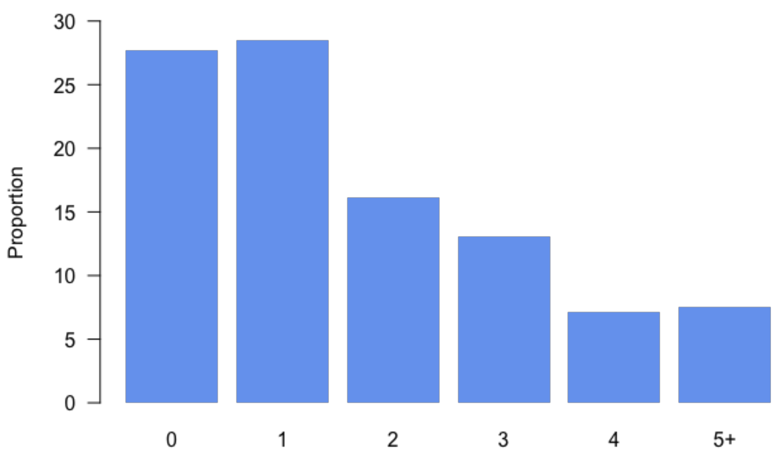

FIGURE 1.3 – Diagramme en barres

Dr. Habiba BOUHALLOUF
Département de biologie appliquée, SNV, UFMC1

Diagramme en bâtons

Un diagramme en bâtons est une représentation graphique de données statistiques à l'aide de segments. Les valeurs de la variable étudiée (quantitative discrète) sont représentées sur l'axe horizontal, les effectifs sur l'axe vertical. À chaque valeur correspond un bâton. Les hauteurs des bâtons sont proportionnelles aux effectifs représentés.

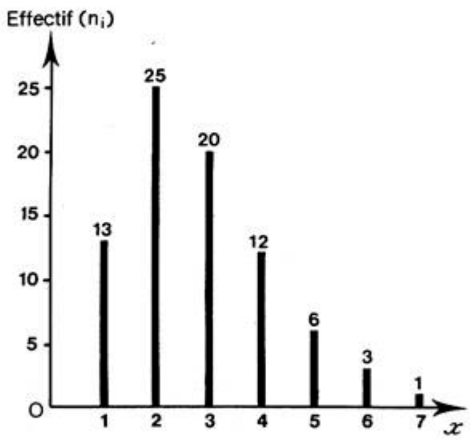

FIGURE 1.4 – Diagramme en bâtons

Diagramme circulaire (Camembert)

Un diagramme circulaire ou à secteurs admet pour support un disque découpé en secteurs dont les aires sont proportionnelles aux pourcentages des différents constituants de la population statistique.
Il est utilisé surtout pour les variables qualitatives nominales quant le nombre de modalités est égale ou inférieur à 5.
Dans un diagramme circulaire, les mesures des angles des secteurs angulaires sont proportionnelles aux effectifs (ou aux fréquences) associés. Une fréquence de 100% correspond à un angle de 360^{o} .

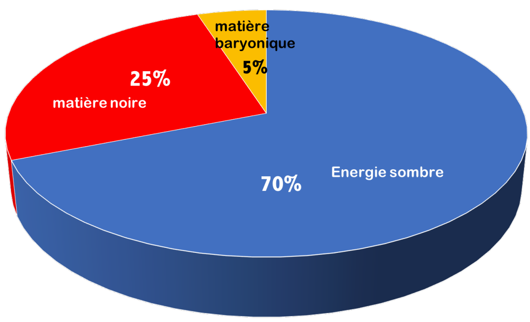

FIGURE 1.5 – Diagramme circulaire

2. Paramètres statistiques

Après avoir collecter les différentes valeurs d'une variable dans un échantillon choisi, il est nécéssaire de résumer les valeurs obtenues en quelques nombres appelés *paramètres* afin de les exprimer et de les utiliser dans la comparaison ou l'estimation,... ect. On distingue deux types de paramètres : paramètres de *position* et paramètres de *dispersion*

2.1 Paramètres de position

Ce type de paramètre sert à exprimer la position d'une distribution en fonction des valeurs associées à la variable étudiée.

2.1.1 Mode

Le mode d'une distribution est la valeur que l'on rencontre le plus fréquemment, c-à-d celui qui a le plus grand effectif.
Le mode peut ne pas exister, et s'il existe, il peut ne pas être unique.

— Quand on a un seul mode, c'est une distribution uni-modale.
— Quand on a deux modes, la distribution est dite bi-modale.
— Quand on a trois modes, la distribution est tri-modale.
— ...

2.1.2 Médiane

La médiane est un paramètre de tendance centrale qui partage une série de nombres rangés par ordre de grandeur croissante, elle peut être la valeur du milieu ou la moyenne arithmétique des valeurs centrales.

Par extension, on peut penser aux valeurs qui dévisent la distribution en plus de deux parties :

Quartiles

Les quartiles sont les trois valeurs qui partagent la distribution en quatre parties égales. On les notes Q_1, Q_2 et Q_3 et on les appelle respectivement :

1. **le premier quartile** : est la valeur qui partage la distribution 25% valeurs inférieures et 75% valeurs supérieures.

2. **le deuxième quartile** : est la valeur qui partage la distribution 50% valeurs inférieures et 50% valeurs supérieures.

3. **le troixième quartile** : est la valeur qui partage la distribution 75% valeurs inférieures et 25% valeurs supérieures.

Déciles

Les déciles sont les valeurs qui divisent la distribution en 10 parties égales dont chaque partie comprend 10% des effectifs. On les note D_1, D_2, D_3, ..., D_8, D_9.

Centiles ou Percentiles

Ce sont les valeurs qui partagent la distribution en 100 groupes de tailles égales. Elles sont appelées souvent *les quartiles d'ordre 100* et notées P_1, P_2, ..., P_{98}, P_{99}.
On remarque que :

— $Q_2 \equiv D_5 \equiv P_{50} \equiv \textit{médiane}$

— $Q_1 \equiv P_{25}$

— $Q_3 \equiv P_{75}$

2.1.3 Intervalle interquartile

L'intervalle interquartile est défini par l'interquartile compris entre le premier et le dernier quartile : $[Q_1, Q_3]$.

2.1.4 L'écart interquartile

L'écart interquartile est l'écart entre le premier et le dernier quartile :

$$\Delta Q = |Q_3 - Q_1|. \tag{2.1}$$

2.1.5 Moyenne

La moyenne est un paramètre de tendance centrale résultant de la somme algébrique des valeurs observées dans une série, divisée par le nombre total des sujets. On définit plusieurs types de la moyenne :

Dr. Habiba BOUHALLOUF
Département de biologie appliquée, SNV, UFMC1

Definition 2.1.1 — Moyenne arithmétique. Soit l'ensemble de N éléments notés : $x_1, x_2, x_3, ..., x_N$. La moyenne arithmétique $\bar{x}$ est définie par :

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{x_1 + x_2 + x_3 + \ldots + x_N}{N} \tag{2.2}$$

Si x_1 se produit n_1 fois, x_2, n_2 fois,.., x_k, n_k fois, la moyenne arithmétique $\bar{x}$ devient :

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} n_i x_i = \frac{n_1x_1 + n_2x_2 + n_3x_3 + \ldots + n_kx_k}{n_1 + n_2 + n_3 + \ldots n_k} = \sum_{i=1}^{N} f_i x_i \tag{2.3}$$

où n_i est l'effectif de la valeur x_i, f_i la fréquence relative à la variable i et N est l'effectif total.

Definition 2.1.2 — Moyenne géométique. La moyenne géométrique, notée G d'un ensemble de N valeurs, est la racine $N^{ième}$ du produit de ces valeurs :

$$G = \sqrt[N]{x_1x_2x_3 \ldots x_N} \tag{2.4}$$

Definition 2.1.3 — Moyenne quadratique. La moyenne quadratique, notée MQ d'un ensemble de N valeurs x_i où $i = \overline{1,N}$ est définie par :

$$MQ = \sqrt{\overline{x^2}} = \frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{N} {x_i}^2} \tag{2.5}$$

Definition 2.1.4 — Moyenne harmonique. La moyenne harmonique H d'un ensemble de N valeurs x_i où $i = \overline{1,N}$ est l'inverse de la moyenne arithmétique des inverses des valeurs :

$$H = \frac{1}{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{x_i}} = \frac{N}{\sum_{i=1}^{N}\frac{1}{x_i}} \tag{2.6}$$

Remarque

Voici une relation empirique entre la moyenne, la médiane et le mode :

moyenne - *mode* = 3 (*moyenne* - *médiane*)

2.1.6 Moments

Definition 2.1.5 On appelle le moment d'ordre r la quantité M_r définit ainsi :

$$M_r = \frac{1}{N}\sum_{i=1}^{N} n_i {x_i}^r \tag{2.7}$$

Moment centé

Definition 2.1.6 Le moment centré d'ordre r noté m_r est le moment calculé par rapport à la moyenne $\bar{x}$ est défini par :

$$m_r = \overline{(x-\bar{x})^r} = \frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^r \tag{2.8}$$

Moment par rapport à une valeur origine

Definition 2.1.7 On définit le moment d'ordre r par rapport à une valeur origine A, noté m'_r comme suit :

$$m'_r = \overline{(x-A)^r} = \frac{1}{N}\sum_{i=1}^{N}(x_i-A)^r \tag{2.9}$$

Remarques

— Si $r = 1$, $M_1 = \bar{x}$ et $m_1 = 0$

— Si $r = 2$, $m_1 = \sigma^2$

2.2 Paramètres de dispersion

Les paramètres de dispersion sont les paramètres qui résument la dispersion des valeurs autour de la valeur centrale :

2.2.1 Extrêmes

Ce sont les deux valeurs extêmes de la distribution, la valeur minimale et la valeur maximale.

2.2.2 Etendue

L'étendue d'une distribution est la différence entre la plus grande et la plus petite valeur (les valeurs extrêmes).

2.2.3 Ecart moyen

Definition 2.2.1 On appelle écart moyen d'une distribution de N valeurs : x_1, x_2, x_3, ..., x_N, la moyenne arithmétique des écarts absolus entre les valeurs d'une variable et la moyenne arithmétique.

$$EM = \overline{|x-\bar{x}|} = \frac{1}{N}\sum_{i=1}^{N}|x_i-\bar{x}| \tag{2.10}$$

Si x_1, x_2, x_3, ..., x_N ont les fréquences d'apparition : n_1, n_2, ..., n_N fois respectivement, l'écart moyen s'écrit ainsi :

$$EM = \frac{1}{N}\sum_{i=1}^{N} n_i |x_i - \bar{x}| \tag{2.11}$$

2.2.4 Variance

Definition 2.2.2 On appelle une variance la moyenne arithmétique des carrés des écarts entre les valeurs d'une variable et la moyenne arithmétique.

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} n_i (x_i - \bar{x})^2 \tag{2.12}$$

2.2.5 Ecart type

Definition 2.2.3 On appelle écart-type (ou écart quadratique moyen) la racine carrée de la variance.

$$\sigma = \frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{N} n_i (x_i - \bar{x})^2} \tag{2.13}$$

II

Probabilités

3. Analyse combinatoire

L'analyse combinatoire comprend un ensemble de méthodes qui permettent de déterminer le nombre de tous les résultats possibles d'une expérience.
La connaissance de ces méthodes est indispensable au calcul des probabilités.

3.1 Principe général

Definition 3.1.1 Si une opération comprend k phases distinctes dont la première pouvant se réaliser de n_1 façons, la deuxième, de n_2 façons, ..., la k^{ieme}, de n_k façons ; alors l'opération peut s'effectuer de :

$$n_1 \;.\; n_2 \;.\; n_3. \;\ldots\; .n_{k-1} \;.\; n_k \text{façons} \tag{3.1}$$

3.2 Arrangement

Definition 3.2.1 L'arrangement est le nombre de possibilités d'ordonner r objets pris parmi les n donnés, il est noté par A_n^r et donné par :

$$A_n^r = \frac{n!}{(n-r)!} \tag{3.2}$$

3.3 Permutation

Definition 3.3.1 La permutation est le nombre de possibilités différentes de ranger n objets distinctes dans un ordre donné et définie par :

$$P_n = n! = n\;(n-1)\;(n-2)\;\ldots\;1 \tag{3.3}$$

3.4 Combinaison

Definition 3.4.1 Une combinaison de n objets différents pris r à r est une sélection de r objets parmi les n donnés sans ordre détérminé. Elle est donnée par :

$$C_n^r = \frac{n!}{r!\,(n-r)!} = \frac{A_n^r}{r!} \tag{3.4}$$

3.5 Disposition avec répétition

Si au moins un des éléments choisis parmi les n objets est répété plusieurs fois, on redéfinit les éléments précédents avec répétition :

3.5.1 Arrangement avec répétition

Definition 3.5.1 En choisissant r objets parmi les n donnés dont chaque objet peut être répété jusqu'à r fois, on peut définir l'arrangement ainsi :

$$A_n^r = n^r \tag{3.5}$$

3.5.2 Permutation avec répétition

Definition 3.5.2 Le nombre total de permutation avec répétition de n objets est donné par :

$$P_n = \frac{n!}{\prod_{i=1}^{k} n_i} = \frac{n!}{n_1!\, n_2! \ldots n_k} \tag{3.6}$$

où n_i est le nombre d'éléments répétés du groupe i avec $(i = \overline{1,k})$.

3.5.3 Combinaison avec répétition

Definition 3.5.3 On désigne par Γ_n^r, le nombre total de combinaisons avec répétition de n éléments pris r à r tel que :

$$\Gamma_n^r = C_{n+r-1}^r = \frac{(n+r-1)!}{r!(n-1)!} \tag{3.7}$$

3.6 Propriétés

— $0! = 1$ (par convention).

— $n! = n(n-1)!$

— $(n.m)! \neq n!.m!, \quad (n+m)! \neq n!+m!, \quad (n-m)! \neq n!-m!$

— $C_n^r = C_n^{n-r}$

— $C_n^r = C_{n-1}^r + C_{n-1}^{r-1}$

4. Calcul des probabilités

4.1 Événement

Tous les résultats d'une expérience aléatoire constituent un ensemble S appelé l'unvers des résultats possibles. Les éléments de S sont appelés événements élémentaires.

4.2 Définition classique des probabilités

Soit un événement A contient r points et appartient à l'ensemble S. Sa probabilité est donnée par :

$$P(A) = \frac{r}{n} \tag{4.1}$$

n : est le nombre d'éléments de l'ensemble S (nombre de cas possibles).
r : est le nombre de cas favorables pour la réalisation de l'événement A.

4.2.1 Propriétés

1. La probabilité est une valeur positive ne dépassant pas 1,

$$0 \leq P(A) \leq 1 \tag{4.2}$$

Cela implique que :

— Si $P(A) = 0$, on dit que l'événement A est impossible.

— Si $P(A) = 1$, on dit que l'événement A est certain.

— Si $P(A)$ est voisine de 0, on dit que l'événement A est peu probable.

— Si $P(A)$ est voisine de 1, on dit que l'événement A est très probable.

2. La somme de toutes les probabilités d'un événement est égale à 1,

$$\sum_{i=1}^{n} P(A) = 1 \tag{4.3}$$

3. La probabilité de non réalisation de A est égale à $(1 - P(A))$.

4.3 Définition statistique des probabilités

4.3.1 Fréquence d'un événement

Definition 4.3.1 Si on répète N fois une expérience et on suppose qu'on a obtenu k fois la réalisation d'un événement A, on appelle la fréquence de l'événement A, le rapport :

$$f = \frac{k}{N} \tag{4.4}$$

La probabilité d'un événement A est donc la limite des fréquences quant le nombre de répétitions de l'expérience tend vers ∞.

4.3.2 Complémentaire ou contraire d'un événement

L'événement contraire d'un événement A est noté $\bar{A}$. $\bar{A}$ est donc l'événement qui se réalise quand A ne se réalise pas, et réciproquement.

4.3.3 Réunion ou somme de deux événements

La réunion de deux événements A et B est définie par $A \cup B$, c'est l'événement qui se réalise quand l'un au moins des deux événements A ou B se réalise.

4.3.4 Intersection ou produit de deux événements

L'intersection de A et B est définie par $A \cap B$ qui est l'événement qui se réalise quand A et B se réalisent simultatnément.

Par conséquent, on écrit :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{4.5}$$

4.3.5 événements incompatibles

Si $A \cap B = \emptyset$, on dit que A et B sont *incompatibles*. Pour cela, on peut écrire :

$$P(A \cup B) = P(A) + P(B) \tag{4.6}$$

4.4 Probabilités conditionnelles

4.4.1 Evénements dépendants

On dit que l'événement A dépend de l'événement B si la probabilité de la réalisation de l'événement A est modifiée par la réalisation ou la non réalisation de l'événemnt B.

Definition 4.4.1 On désigne la probabilité de la réalisation de l'événement A à condition que l'événement B ait lieu, par $P(A/B)$ qu'on appelle : *probabilité conditionnelle* de l'événement A sachant que B est réalisé.

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \tag{4.7}$$

4.4.2 Evénements indépendants

Definition 4.4.2 On dit que l'événement A est indépendant de l'événement B si la probabilité de la réalisation de l'événement A ne dépend pas du fait que l'événement B soit réalisé ou non. Cela permet d'écrire :

$$P(A/B) = P(A) \tag{4.8}$$

$$P(B/A) = P(B) \tag{4.9}$$

$$P(A \cap B) = P(A).P(B) \tag{4.10}$$

4.5 Propriétés

1. soit $\bar{A}$ l'événement contraire de l'événement A, on a :

$$P(A) + P(\bar{A}) = 1 \tag{4.11}$$

2. Si A et B sont deux événements quelconques, alors :

$$P(A \cap B) = P(B \cap A) \Rightarrow P(A/B).P(B) = P(B/A).P(A) \tag{4.12}$$

3. Si A et B sont incompatibles, alors :

$$P(A \cup B) = P(A) + P(B) \tag{4.13}$$

4. Si A et B sont indépendants, alors :

$$P(A \cup B) = P(A) + P(B) - P(A).P(B) \tag{4.14}$$

5. Si A dépend de B, alors

$$P(A \cup B) = P(A) + P(B) - P(A/B).P(B) \tag{4.15}$$

6. Si A, B et C sont trois événements quelconques, alors :

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \tag{4.16}$$

4.6 Espérance mathématique

Si x désigne une variable aléatoire discrète pouvant prendre des valeurs : $x_1, x_2, ..., x_n$ avec $P_1, P_2, ..., P_n$ comme probabilité respectivement, de telle manière que : $P_1 + P_2 + ... + P_n = 1$, l'espérance mathématique, notée $E(x)$ de x est définie par :

$$E(x) = P_1\, x_1 + P_2\, x_2 + ... + P_n\, x_n = \sum_{i=1}^{n} P_i\, x_i \tag{4.17}$$

4.7 Théorème de Bayes

On s'intéresse à la modification des probabilités d'événements suite à la connaissance des faits. Il s'agit d'exprimer la probabilité conditionnelle $P(A/B)$:

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

En changeant la formule du numérateur tel que : $P(A \cap B) = P(B/A).P(A)$, on aura :

$$P(A/B) = \frac{P(B/A).P(A)}{P(B)} \tag{4.18}$$

En général, on ne connait pas B. On peut l'exprimer en fonction de A :

$$P(B) = P(A \cap B) + P(\bar{A} \cap B) \tag{4.19}$$

Les événements $(A \cap B)$ et $(\bar{A} \cap B)$ sont incompatibles. La probabilité d'avoir l'un et l'autre est la somme des probabilités :

$$P(B) = P(B/A).P(A) + P(B/\bar{A}).P(\bar{A}) \tag{4.20}$$

Finalement, la formule du théorème de *Bayes* s'écrit ainsi :

$$P(A/B) = \frac{P(B/A).P(A)}{P(B/A).P(A) + P(B/\bar{A}).P(\bar{A})} \tag{4.21}$$

en généralisant cette formule, il vient :

$$P(A_i/B) = \frac{P(B/A_i).P(A_i)}{P(B/A_i).P(A_i) + P(B/\bar{A}_i).P(\bar{A}_i)} \tag{4.22}$$

avec i varie de 1 à n.

4.8 Lexique

A $\equiv$ L'événement associé à A s'est réalisé.

$\bar{A}$ $\equiv$ L'événement associé à A ne s'est réalisé pas.

$A \cap B$ $\equiv$ L'événement associé à A et B se sont réalisés au *même temps*.

$A \cup B$ $\equiv$ L'un au moins des événements associés à A et B s'est réalisé.

$A \cap \bar{B}$ $\equiv$ L'événement associé à A s'est réalisé **et** l'événement associé à B ne s'est réalisé pas.

$A \cup \bar{B}$ $\equiv$ L'événement associé à A s'est réalisé **ou** l'événement associé à B ne s'est réalisé pas.

$\overline{A \cup B}$ $\equiv$ Aucun des deux événements associés à A et B ne s'est réalisé pas.

$\overline{A \cap B}$ $\equiv$ Les événements associés à A et B ne se sont réalisés pas au *même temps*.

Dr. Habiba BOUHALLOUF
Département de biologie appliquée, SNV, UFMC1