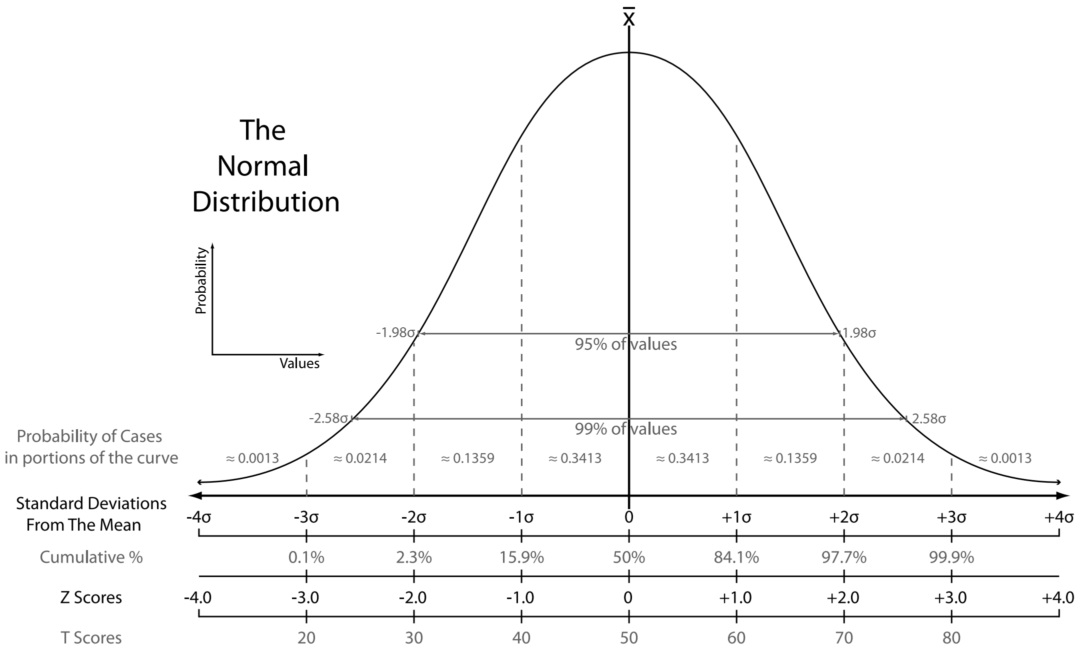
Première année de médecine

Biostatistique

Riad Benchoucha



Ce document est distribué gratuitement
21 octobre 2011
LaTeX

Avant-propos

Ce modeste ouvrage, conçu comme un simple résumé, a pour objectif de présenter les bases de la statistique aux étudiants de première année de médecine, ainsi qu'à tous ceux qui dans d'autres disciplines, s'intéressent à cette science.

C'est en constatant le manque de livres au niveau de notre université, que j'ai décidé d'apporter ma contribution. Tout au long de sa réalisation, j'ai été guidé par le souci de m'adresser à un large public et de présenter les informations de la manière la plus simple et didactique que possible.

Je serais bien évidement heureux que les lecteurs me fassent part de leurs suggestions et critiques afin que je puisse améliorer et parfaire cet ouvrage. Je reste pour cela joignable par e-mail (riad47@gmail.com).

Riad BENCHOUCHA
Étudiant en médecine

Chapitre

1

Dans ce chapitre

Introduction à l'étude de la statistique descriptive

La statistique descriptive est la branche des statistiques qui regroupe les nombreuses techniques utilisées pour *décrire*, c'est-à-dire de résumer ou représenter, un ensemble *relativement important* de données.

1.1 Vocabulaire de base

Population statistique : toute étude statistique porte sur un ensemble appelé **population**. Les éléments de cette population sont les **individus** (ou unité statistique). Les premières études statistiques étaient démographiques et on a gardé ce vocabulaire.

Échantillon : lorsque la population est nombreuse, on travaille sur un ou plusieurs *échantillons*. Pour étendre les résultats collectés à l'ensemble de la population, il est important que l'échantillon soit **représentatif** et non **biaisé**.

Caractère ou variable : lors d'une étude statistique, on s'intéresse à un aspect des individus de la population appelé variable ou caractère.

La variable prend différentes valeurs que l'on appelle **modalités**.

Variable quantitative : si on peut mesurer cet aspect, c'est-à-dire y faire des opérations, la variable est dite de nature quantitative. Exemple : le poids

- Si la variable quantitative ne prend que quelques valeurs, dites isolées, elle est **discrète**. Exemple : le nombre d'enfants
- Si la variable quantitative prend n'importe quelle valeur d'un intervalle, elle est **continue**. Exemple : le poids

Variable qualitative : si l'aspect ne se traduit pas par des nombres, ou n'est pas mesurable, la variable est de nature qualitative. Exemple : la couleur des yeux.

- Mesurées dans une échelle **nominale**, les modalités sont exprimables par des noms et ne sont *pas hiérarchisées*.
  Exemple : la couleur du pelage, les groupes sanguins, les différents nucléotides de l'ADN.
- Mesurées dans une échelle **ordinale**, les modalités traduisent le degré d'un état caractérisant un individu sans que ce degré ne puisse être défini par un nombre qui résulte d'une mesure. Les modalités sont alors *hiérarchisées*.
  Exemple : le stade d'une maladie.

1.2 Série statistique

1.2.1 Modalités et classes

A. Variable discrète :

Dans le cas d'une variable discrète, les différentes modalités du caractère étudié seront notées $x_1, x_2, x_3, \cdots, x_i$.

B. Variable continue :

Dans le cas d'une variable continue, on procède à un **regroupement par classes**.

Amplitude d'une classe : l'amplitude, ou l'**intervalle** d'une classe statistique bornée est sa **largeur**. L'amplitude de la classe $[a, b[$ est $b - a$. On parle aussi d'étendue d'une classe.

Étendu d'une série statistique : l'étendue d'une série statistique est la différence entre les deux valeurs extrêmes prises par un caractère quantitatif, on notera : $E = x_{max} - x_{min}$.

Regroupement par classes En règle générale, on choisit des classes de **même amplitude**. Pour que la distribution en fréquence ait un sens, il faut que chaque classe comprenne un nombre suffisant de valeurs (n_i).

Diverses formules empiriques permettent d'établir le nombre de classes pour un échantillon de taille N. Nous utiliserons celle-çi, soit :

- k le nombre de classes ($k \in \mathbb{N}$) ;
- N l'effectif total ;
- E l'étendue de la série ;
- a l'amplitude de la classe.

$$k = \sqrt{N} \quad ; \quad a = \frac{E}{k}$$

Remarque : $a \geqslant E/k$, arrondir supérieurement.

1.2.2 Effectif

A. Effectif d'une modalité ou d'une classe

L'effectif d'une modalité x_i est égal au nombre d'individus qui prennent cette valeur, on le note n_i. Les différents effectifs seront notés $n_1, n_2, n_3, \ldots\ldots, n_i$.

B. Effectif total

L'effectif total est égal au nombre d'individus de la population, on le note **N**. On a donc :

$$N = n_1 + n_2 + n_3 + \ldots\ldots + n_i = \sum_{k=1}^{i} n_k$$

1.2.3 Fréquence

La fréquence d'une valeur est le rapport de l'effectif de cette valeur sur l'effectif total.
La fréquence de la modalité x_i, notée f_i, est donc égale à :

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

1.2.4 Effectifs ou fréquences cumulés

Effectif cumulé croissant : l'effectif cumulé croissant d'une valeur x_i est la somme des effectifs des valeurs inférieures ou égales à x_i.

Fréquence cumulée croissante : la fréquence cumulée croissante d'une valeur x_i est la somme des fréquences de valeurs inférieures ou égales à x_i.

Effectif cumulé décroissant : l'effectif cumulé décroissant d'une valeur x_i est la somme des effectifs des valeurs supérieures ou égales à x_i.

Fréquence cumulée décroissante : la fréquence cumulée décroissante d'une valeur x_i est la somme des fréquences des valeurs supérieures ou égales à x_i.

1.3 Représentations graphiques

Les représentations graphiques ont l'avantage de renseigner immédiatement sur l'allure générale de la distribution. Elles facilitent l'interprétation des données recueillies.

1.3.1 Caractère quantitatif discret

A. Diagramme en bâtons

Il sert à représenter en général un caractère **quantitatif discret**. La hauteur de chaque bâton représentant une valeur est proportionnelle à l'effectif (ou la fréquence) correspondant(e).

Remarque : si 0 (zéro) est une modalité du caractère, il est d'usage de décaler l'origine.

B. Polygone des fréquences

Le polygone des fréquences est obtenu en joignant les sommets des bâtons par une ligne brisée.

C. Diagramme cumulatif

Le diagramme cumulatif est obtenu en reliant les bâtons cumulés par une ligne en escalier.

1.3.2 Caractère quantitatif continu

A. Histogramme

Il sert à représenter un caractère **quantitatif continu**. L'aire de chaque rectangle est proportionnelle à l'effectif (ou la fréquence) de la valeur correspondante.

B. Polygone des fréquences

Le polygone des fréquences est obtenu en joignant les milieux des sommets des rectangles par une ligne brisée. On joindra les milieux des sommets du premier et dernier rectangle avec, respectivement, les points $(a_0 - \frac{a}{2}; 0)$ et $(a_k + \frac{a}{2}; 0)$.

C. Courbe cumulative

La courbe cumulative est obtenue en reliant les points ayant pour abscisse la borne supérieur de la classe et l'ordonnée n_{ic} ou f_{ic}. On joindra le premier point obtenu avec le point $(a_0; 0)$.

Remarque : une courbe cumulative bien tracée ne comporte pas de cassures et a une forme en s allongé.

1.3.3 Caractère qualitatif

A. Diagramme à bande

Le diagramme à bande ou en tuyaux d'orgue sert à représenter un caractère **qualitatif** (souvent **ordinal**), chaque modalité est représentée par une bande de longueur proportionnelle à l'effectif ou la fréquence. Les différentes bandes doivent avoir la même largeur.

Pour comparer des séries dans le temps, il suffit de tracer sur le même diagramme des bandes juxtaposées aux premières, d'égales largeurs que ces dernières mais de couleurs différentes.

Remarque : les modalité étant non mesurables, on ne met pas de flèche sur l'axe des abscisses.

B. Diagramme circulaire

Le diagramme circulaire ou à secteurs sert à représenter en général un caractère **qualitatif nominal**. Chaque secteur angulaire représentant une valeur ou modalité est proportionnel à l'effectif (ou fréquence) correspondant.

Remarque : on représente souvent les caractères qualitatifs nominaux par un diagramme à bandes car il est plus facile à dessiner et permet une comparaison dans le temps de plusieurs séries.

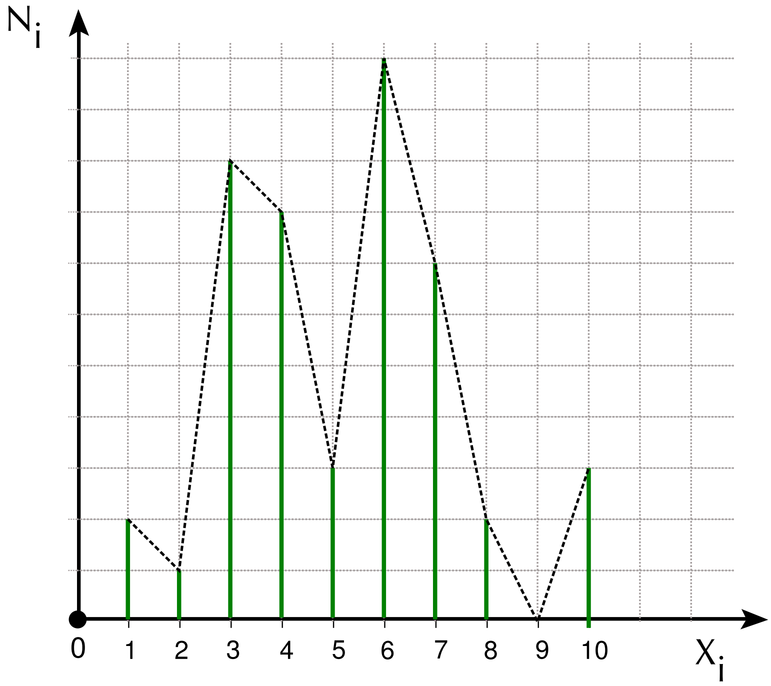


FIGURE 1.1 – En bleu le diagramme en bâtons, en noir le polygone des fréquences.

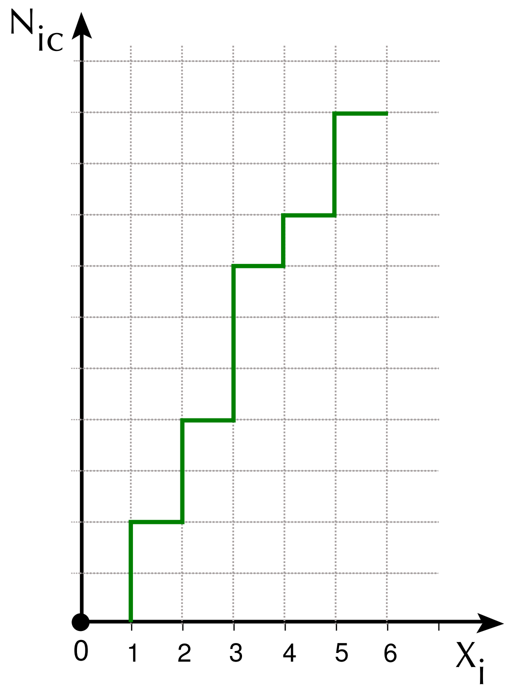


FIGURE 1.2 – En bleu le diagramme cumulatif.

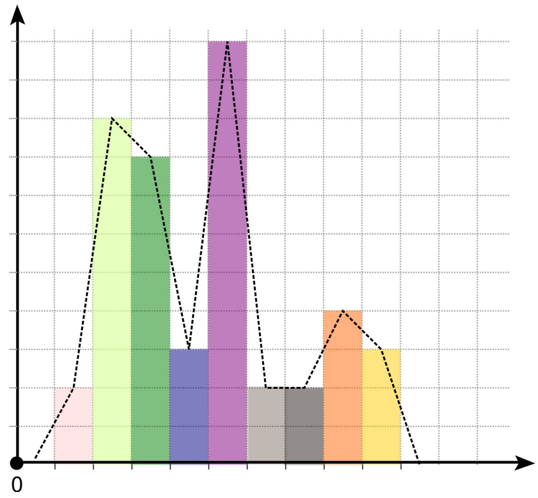


FIGURE 1.3 – En couleurs l'histogramme, en noir le polygone des fréquences.

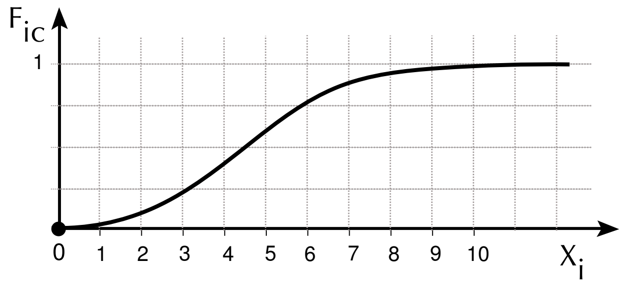


FIGURE 1.4 – Courbe cumulée.

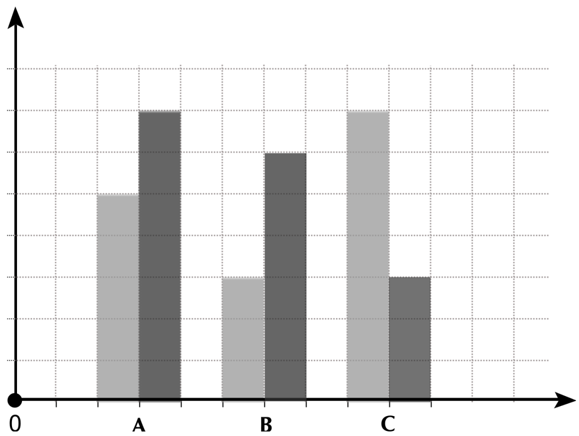


FIGURE 1.5 – Diagramme à bande.

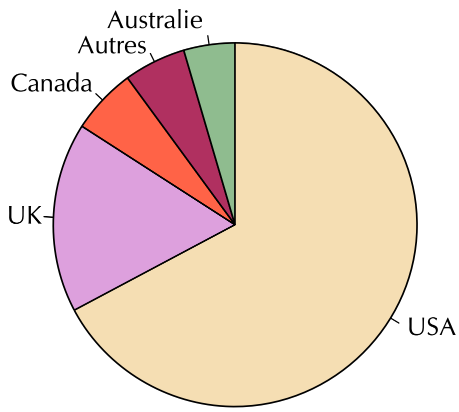


FIGURE 1.6 – Diagramme circulaire.

Chapitre

2

Les paramètres statistiques

Lorsque l'on est en face d'une série statistique comportant un grand nombre de termes, il devient difficile d'analyser directement l'ensemble des données. Nous avons vu que la représentation graphique permettait d'extraire une certaine information. En complément de cette analyse qualitative, le statisticien est amené à simplifier la distribution observée par des caractéristiques ou paramètres, significatives. Nous distinguerons deux types de caractéristiques : celles de la position et celles de dispersion.

2.1 Paramètres de position

Ces paramètres ont pour objectif dans le cas d'un caractère quantitatif de caractériser **l'ordre de grandeur** des observations.

2.1.1 Moyenne arithmétique

Soit un échantillon de n valeurs observées $x_1, x_2, \cdots, x_n$ d'un caractère quantitatif X, on définit sa moyenne arithmétique $\bar{x}$ comme :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Lorsque les valeurs $x_1, x_2, \cdots, x_i$ se répètent réspectivement $n_1, n_2, \cdots, n_i$, il faut les pondérer par les effectifs correspondants :

$$\bar{x} = \frac{1}{N}\sum_{k=1}^{i} n_k \cdot x_k \quad ; \quad N = \sum_{k=1}^{i} n_k$$

- La moyenne de la série x_i associée aux fréquences f_i est :

$$\bar{x} = \sum_{k=1}^{i} f_k \cdot x_k$$

- Pour des données groupées en classes, on peut calculer une valeur approximative de la moyenne en supposant que tous les individus d'une classe se situent au centre de celle-ci. x_i représente alors le **centre de la classe**.

2.1.2 Propriétés de la moyenne arithmétique

- Lorsqu'on augmente chacune des valeurs du caractère du même réel b, la moyenne augmente également de b (linéarité de la moyenne arithmétique) :

$$\overline{x+b} = \bar{x} + b$$

- Lorsqu'on multiplie chacune des valeurs du caractère par un même réel a, la moyenne est multipliée par a (linéarité de la moyenne arithmétique) :

$$\overline{a \times x} = a \cdot \bar{x}$$

- La somme des écarts à la moyenne est nulle :

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0$$

Remarque : On utilisera ces propriétés lors d'un changement de variable statistique (voir section 2.2.6 page 11).

A. Le mode

Le mode Mo d'une série statistique est la valeur du caractère la **plus fréquente** ou dominante dans l'échantillon.

Dans le cas d'une variable continue, la **classe modale** correspond à la classe de fréquence maximale dans la distribution des fréquences. On peut identifier le mode comme la valeur médiane de la classe modale.

Remarque : une série peut être unimodale ou plurimodale.

B. Quantile d'ordre α

On appelle quantile d'ordre α ($\alpha \in]0;1[$) le nombre noté q_α tel qu'il y ait $n\alpha$ des observations qui lui sont inferieures dans une série ordonnée de taille n.

Certains quantiles ont des noms spéciaux :
- les 100-quantiles sont appelés **centiles** ou percentiles selon un anglicisme fréquent ;
- les 10-quantiles sont appelés **déciles** ;
- les 5-quantiles sont appelés **quintiles** ;
- les 4-quantiles sont appelés **quartiles** ;
- le 2-quantile est appelé **médiane**.

Calcul des quantiles

$$q_\alpha = \begin{cases} \frac{1}{2}(x_{n\alpha} + x_{n\alpha+1}) & si \;\; n\alpha \in \mathbb{N} \\ x_{E[n\alpha+1]} & si \;\; n\alpha \notin \mathbb{N} \end{cases} \tag{2.1}$$

$E\,[n\alpha + 1]$ désignant la partie entière de $n\alpha + 1$.

Interpolation linéaire

Soit f une fonction définie sur $\mathbb{R}$, $[a;b]$ un intervalle de $\mathbb{R}$ et c un nombre réel .

Quand il n'est pas possible de calculer l'image de c par f , on utilise une **interpolation linéaire**, cela consiste à remplacer $f(c)$ par $g(c)$ ou g est la fonction affine telle que $g(a) = f(a)$ et $g(b) = f(b)$.

On remplacera alors la courbe représentative de f sur $[a;b]$ par la droite (AB). On dit que l'on a déterminé $f(c)$ par interpolation linéaire.

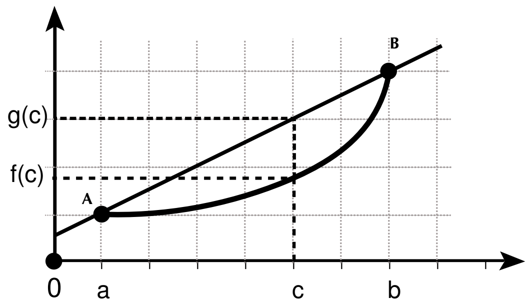


FIGURE 2.1 – Interpolation linéaire.

$$f(c) \simeq f(a) + (c-a)\frac{f(b) - f(a)}{b-a}$$

Calcul des quantiles par interpolation linéaire

On montre grâce au théorème de THALÈS que :

$$\frac{q\alpha - l_1}{l_2 - l_1} = \frac{n\alpha - F_1}{F_2 - F_1}$$

Avec :
- l_1 et l_2 les extrémités de la classe contenant $q\alpha$;
- F_1 la fréquence (ou effectif) cumulée de la classe avant $[l_1;l_2]$;
- F_2 la fréquence (ou effectif) cumulée de la classe $[l_1;l_2]$.

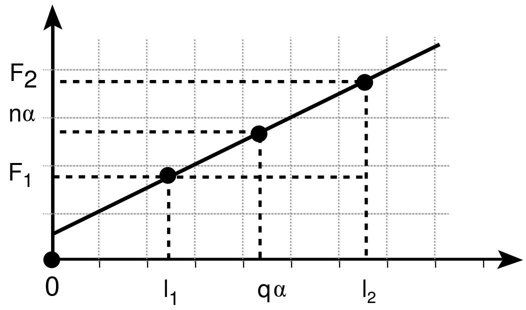


FIGURE 2.2 – Interpolation linéaire.

2.2 Paramètres de dispersion

Ces paramètres ont pour objectif dans le cas d'un caractère quantitatif de caractériser la **variabilité** des données dans l'échantillon.

2.2.1 La variance

Soit un échantillon de n valeurs observées $x_1, x_2, \cdots, x_n$ d'un caractère quantitatif X, $\bar{x}$ sa moyenne arithmétique. On définit la variance, notée σ^2 ou $var(x)$ comme la moyenne arithmétique des carrés des écarts à la moyenne.

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Lorsque les valeurs $x_1, x_2, \cdots, x_i$ se répètent réspectivement $n_1, n_2, \cdots, n_i$, il faut les pondérer par les effectifs correspondants :

$$\sigma^2 = \frac{1}{N}\sum_{k=1}^{i} n_k \cdot (x_k - \bar{x})^2 \quad ; \quad N = \sum_{k=1}^{i} n_k$$

Pour des commodités de calcul, on utilisera la formule développée suivante (théorème de KŒNIG) :

$$\sigma^2 = \frac{1}{N}\sum_{k=1}^{i}(n_k \cdot x_k^2) - \bar{x}^2 \quad ; \quad N = \sum_{k=1}^{i} n_k$$

– Pour des données groupées en classes, x_i représente le centre de la classe.
– De part sa définition, la variance est toujours un nombre positif.
– La dimension de la variance est le carré de celle de la variable.

2.2.2 L'écart type

L'écart-type d'une série (ou écart quadrique) est la racine carrée de sa variance.

$$\sigma(x) = \sqrt{V(x)}$$

L'écart-type s'exprime dans les mêmes unités que la moyenne.

Remarque : La variance et l'écart-type sont des paramètres de dispersion **absolue** qui mesurent la variation absolue des données indépendamment de l'ordre de grandeur des données.

2.2.3 Coefficient de variation

Le coefficient de variation noté CV_x est un indice de dispersion **relatif** égal à :

$$CV_x = \frac{\sigma_x}{\bar{x}}$$

Exprimé en pour cent, il est indépendant du choix des unités de mesure permettant la comparaison de distributions de fréquence d'unité différente.

2.2.4 Étendu d'une série statistique

L'étendu d'une série statistique est la différence entre les deux valeurs extrêmes prises par un caractère quantitatif, on notera :

$$E = x_{max} - x_{min}$$

2.2.5 L'écart absolu moyen

On appelle écart absolu moyen que l'on désigne par e la moyenne arithmétique des écarts absolus entre les valeurs du caractère et la moyenne arithmétique.

$$e = \frac{1}{N}\sum_{k=1}^{i}|x_k - \bar{x}| \quad ; \quad N = \sum_{k=1}^{i} n_k$$

2.2.6 Écart interquartile

L'intervalle interquartile (écart interquartile) est la différence entre le premier et le troisième quartile.

$$Q_i = Q_3 - Q_1$$

L'intervalle interquartile élimine le premier 25% et le dernier 25% soit la tête et la queue, des observations. On ne garde que les 50% au centre. L'intervalle interquartile est une bonne mesure de la dispersion des données d'une variable.

Changement de variable statistique

Le calcul de la variance ou de la moyenne est parfois laborieux à cause des valeurs élevées de x_i, c'est pourquoi on est souvent ammené à effectuer un changement de la variable statistique.

$$x = a\,y + b$$

On choisira :
- a l'amplitude de la classe ;
- b le mode.

On écrira y, on calculera $\bar{y}$ et ${\sigma_y}^2$, respectivement la moyenne de la variance de y, puis on utilisera ces formules :

$$\bar{x} = a\bar{y} + b$$
$${\sigma_x}^2 = a^2{\sigma_y}^2$$

Chapitre

3

Dans ce chapitre

L'analyse combinatoire

L'analyse combinatoire est l'ensemble des techniques qui servent, en mathématiques, à compter (ou dénombrer) le nombre de dispositions que l'on peut former à partir des éléments d'un ensemble fini de grande cardinalité.

3.1 Principe fondamental de l'analyse combinatoire

Soit ξ une expérience globale composée d'une succession de k sous-expériences. Si l'expérience ξ_1 a n_1 résultats, la ξ_2 : n_2 résultats ... la ξ_k : n_k résultats, alors le nombre total n de résultats possibles de l'expérience globale est :

$$n = \prod_{i=1}^{k} n_i$$

3.2 Arrangements

3.2.1 Arrangements sans répétition

En mathématiques, lorsque nous choisissons p objets parmi n objets discernables et que l'ordre dans lequel les objets sont sélectionnés revêt une importance, nous pouvons les représenter par un k-uplet d'éléments distincts et on en constitue une liste **ordonnée** sans **répétition possible**, c'est-à-dire dans laquelle l'ordre des éléments est pris en compte, si l'on permute deux éléments de la liste, on a une liste différente, et un élément ne peut être présent qu'une seule fois.

Une telle liste ordonnée est appelée un **arrangement** de n éléments p à p ($p \in \mathbb{N}^*, n \in \mathbb{N}^*$ et $p \leqslant n$). Le nombre d'arrangements que l'on peut faire est noté A_n^p et vaut :

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

Exemple : À un examen, cinq candidats tirent les uns après les autres un sujet dans une urne contenant des questions toutes différentes. Le premier tirage se fera sur un ensemble de 50 questions possibles. À chaque tirage suivant, la question qui vient d'être tirée est enlevée de l'urne. Ainsi, en faisant passer les cinq candidats, le tirage se fait d'abord sur 50, puis sur 49, et ainsi de suite jusqu'à 46 qui représente l'ensemble des questions restantes dans l'urne pour le dernier tirage. L'arrangement va consister à additioner à chaque modification possible de cet ensemble de départ la nouvelle probabilité de piocher une question donnée. La solution pour cet exemple est donc un arrangement de 5 (p) à 50 (n).

$$A_{50}^5 = 50(49)(48)(47)(46) = \frac{50!}{45!} = 254\ 251\ 200$$

3.2.2 Arrangements avec répétition

Un arrangement avec répétition de p élements choisis parmi n est une liste ordonnée de ces p élements avec une eventuelle répétition de un ou plusieurs élements. Le nombre d'arrangements avec répétition que l'on peut faire est noté α_n^p et vaut :

$$\alpha_n^p = n^p$$

Exemple : Dans l'exemple précédant, si on remettait la question tirée de nouveau dans l'urne à chaque tirage, ce serait un arrangement avec répétition de 5 (p) à 50 (n), et la solution vaudrait :

$$\alpha_{50}^5 = 50^5 = 312\ 500\ 000$$

3.3 Permutations

3.3.1 Permutations sans répétition

Une permutation de n objets distincts, correspond à toute suite ordonnée de ces n objets distincts ou tout arrangement n à n de ces objets.

Le nombre de permutations de n objets distincts est noté P_n et vaut :

$$P_n = n!$$

Remarque : La permutation de n objets constitue un cas particulier d'arrangement sans répétition de p objets pris parmi n lorsque $p = n$.

3.3.2 Permutations avec répétition

On considère n objets parmi lesquels r_1 sont semblables entre eux, r_2 sont semblables entre eux, ..., r_k sont semblables entre eux.

Le nombre de permutations de ces n éléments avec répétitions et noté $P_n(r_1, r_2, ..., r_k)$ et vaut :

$$P_n(r_1, r_2, ..., r_k) = \frac{n!}{\prod_{i=1}^{k} r_i!}$$

3.3.3 Permutations circulaires

Les permutations circulaires sont les différentes dispositions qu'on peut donner aux objets placés sur un cercle. Le nombre de permutations circulaires d'un ensemble à n éléments distincts est le nombre noté PC_n défini par :

$$PC_n = (n-1)!$$

Exemple : Pour trois objets, les permutations abc, bca, cab n'en font qu'une, et de même pour bac, acb, cba puisqu'on peut choisir librement le premier objet.

3.4 Combinaisons

3.4.1 Combinaisons sans répétition

En mathématiques, lorsque nous choisissons p objets parmi n objets **distincts** (discernables), numérotés de 1 à n et que l'**ordre** dans lequel les objets sont placés (ou énumérés) n'a **pas d'importance**, nous pouvons les représenter par un sous-ensemble à k éléments.

Un tel sous-ensemble à k éléments est appelé une **combinaison** de p éléments parmi n. Le nombre de combinaisons que l'on peut faire est noté C_n^p ou $\binom{n}{k}$ et vaut :

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

Voici pourquoi :
- Il y a A_n^p manières de tirer p objets parmi les n en les ordonnants.
- Une fois les p objets tirés, il y a $P_p = p!$ manières de les ordonner.
- Il y a donc $\frac{A_n^p}{p!}$ manières de tirer p objets parmi n sans les ordonner.

A. Propriétés

$$C_n^p = C_n^{n-p}$$
$$C_n^p = C_{n-1}^p + C_{n-1}^{p-1}$$

$$C_n^{p+1} = \frac{n-p}{p+1} \cdot C_n^p$$

3.4.2 Cominaisons avec répétition

Il s'agit de dispositions non ordonnées de p éléments pris parmi n, discernables, avec répétition éventuellement.

$$\binom{n}{k} = C_{n+p-1}^{p}$$

Chapitre

4

Dans ce chapitre

Calcul de probalilités

4.1 Vocabulaire de base

Probabilité : la probabilité (du latin probabilitas) est une évaluation du caractère probable d'un évènement. La probabilité d'un événement est un nombre réel compris entre 0 et 1. Plus ce nombre est grand, plus le risque (ou la chance, selon le point de vue) que l'événement se produise est grand.

Expérience aléatoire : une expérience est dite aléatoire si ses résultats ne sont pas prévisibles avec certitude en fonction des conditions initiales.

Épreuve : on appelle épreuve la réalisation d'une expérience aléatoire.

Espace fondamental (univers) : l'espace fondamental ou univers, noté Ω est l'ensemble de tous les événements élémentaires d'une expérience aléatoire.

Évènement élémentaire (éventualité) : un événement élémentaire ou éventualité, noté ω_i est un sous-ensemble de l'univers constitué d'un seul élément.

Évènement : on appelle évènement la propriété du système qui une fois l'épreuve effectuée est ou n'est pas réalisée. Un évenement est noté par une lettre capitale, par exemple A, il correspond à un sous ensemble de $\Omega(A \subset \Omega)$.

Évènement certain : l'univers Ω est appelé événement certain.

Évènement impossible : l'ensemble vide $\varnothing$ est appelé événement impossible.

4.1.1 Opérations sur les évenements

4.1.2 L'union

L'événement $A \cup B$ est réalisé dès que A ou B est réalisé. Dans un lancer de dé, si l'événement A est « obtenir un nombre pair » et l'événement B « obtenir un multiple de 3 », l'événement $A \cup B$ est l'événement « obtenir un nombre pair OU un multiple de 3 », c'est-à-dire $\{2; 3; 4; 6\}$.

4.1.3 L'intersection

L'événement $A \cap B$ (noté parfois $A \cdot B$) est réalisé dès que A et B sont réalisés dans la même expérience. Dans un lancer de dé, si l'événement A est « obtenir un nombre pair » et l'événement B « obtenir un multiple de 3 », l'événement est l'événement « obtenir un nombre pair ET multiple de 3 », c'est-à-dire $\{6\}$.

Évènement contraire : l'événement contraire de A, noté $\bar{A}$ contient tous les éléments de Ω qui ne sont pas dans A. C'est l'événement qui est réalisé dès que A n'est pas réalisé.

Événements incompatibles : lorsque deux événements ont une intersection vide, c'est qu'il ne peuvent pas être réalisés au cours d'une même expérience. On les appelle alors événements incompatibles ou mutuellement exclusifs.

4.1.4 Probabilité sur un ensemble fini

A. Définition mathématique

À chaque événement on associe un nombre positif compris entre 0 et 1, sa probabilité. Àfin d'éviter toute discussion de nature philosophique sur le hasard, la théorie moderne des probabilités repose sur l'axiomatique suivante :

AXIOMATIQUE DE KOLMOGOROV On appelle probabilité sur un ensemble fini Ω est une application P de Ω dans $[0; 1]$ vérifiant :

– $P(\Omega) = 1$;

– pour tout ensemble dénombrable d'évenements incompatibles $A_1, A_2 ... A_i$ on a :

$$P(A_1 \cup A_2 \cup ... \cup A_i) = \sum_{k=1}^{i} P(A_k)$$

B. Equiprobabilité

Si on estime que toutes les éventualités sont équiprobables, et si on note $|\Omega|$, le cardinal de Ω, c'est-à-dire le nombre d'éléments dans Ω, chaque éventualité a une probabilité d'apparition de :

$$P(\{\omega_1\}) = P(\{\omega_2\}) = ... = P(\{\omega_i\}) = \frac{1}{|\Omega|}$$

Dans ce cas, et uniquement dans ce cas la probabilité de l'événement A est donnée par la formule :

$$P(A) = \frac{|A|}{|\Omega|}$$

C. Propriétés élémentaires

Des axiomes on déduit immédiatement les propriétés suivantes :

Propriété 1 : $p(\Omega) = 1$
Propriété 2 : $p(\emptyset) = 0$
Propriété 3 : $p(\overline{A}) = 1 - p(A)$
Propriété 4 : $p(A \cup B) = p(A) + p(B) - p(A \cap B)$
Propriété 5 : $p(A - B) = p(A \cap \bar{B}) = p(A) - p(A \cap B)$
Propriété 7 : $p(A \cup B \cup C) = p(A) + p(B) + p(c) - [p(A \cap B) + p(A \cap C) + p(B \cap C) + p(A \cap B \cap C)]$

D. Rappels utiles sur les opérations appliquées aux ensembles

Lois associatives

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

Lois distributives

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

4.2 Lois de probabilités conditionnelles

4.2.1 Introduction et définitions

Supposons que l'on s'intéresse à la réalisation d'un évenement A, tout en sachant qu'un événement B est réalisé (voir figure 4.2.1). Si A et B sont incompatibles la question est tranchée : A ne se réalisera pas, mais si $A \cap B \neq \emptyset$, il est possible que A se réalise ; cependant, l'univers des possibles n'est plus Ω tout entier, mais est restreint à B. En fait, seule nous intéresse la réalisation de A à l'interieur de B, c'est-à-dire $A \cap B$ par rapport à B.

Ceci justifie la définition suivante :

DÉFINITION

Soit B un événement de probabilité non nulle. On appelle **probabilité conditionnelle A sachant B** (ou encore de A si B) le rapport noté $P(A/B)$:

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

En cas d'équiprobabilité, nous aurons :

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

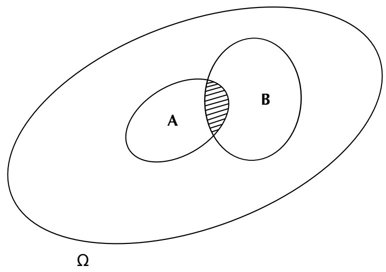


FIGURE 4.1 – Intersection de deux sous ensembles A et B.

A. Probabilité de l'intersection

Pour deux événements nous avons :

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

Pour plus de deux événements, nous aurons :

$$P(A \cap B \cap C) = P(A) \cdot P(B/A) \cdot P(C/A \cap B)$$

4.2.2 Indépendance

A. Indépendance de deux événements

DÉFINITION
A est indépendant de B si $P(A/B) = P(A)$
A et B sont indépendants si et seulement si $P(A/B) = P(A)$ et $P(B/A) = P(B)$
A et B sont indépendants si et seulement si $P(A \cap B) = P(A) \cdot P(B)$

B. Indépendance mutuelle

DÉFINITION
Soit $A_1, A_2, ..., A_n$ des événements, ils sont dits mutuellement indépendants si et seulement s'ils sont indépendants deux à deux.

C. Formules de BAYES

Les formules de BAYES ont pour but d'exprimer $P(A/B)$ en fonction de $P(B/A)$.
Soit B_i un système complet d'événements.

Première formule

$$P(A) = \sum_i P(A/B_i) \cdot P(B_i)$$

Deuxième formule

$$P(B_i/A) = \frac{P(A/B_i) \cdot P(B_i)}{\sum_k P(A/B_k) \cdot P(B_k)}$$

Table des matières