# INTRODUCTION A LA STATISTIQUE

Émile Amzallag
Norbert Piccioli

PREMIERS CYCLES UNIVERSITAIRES

# INTRODUCTION
# A LA
# STATISTIQUE

## Exercices corrigés
## avec rappels détaillés de cours et exemples

A l'usage des étudiants en sciences économiques,
médecine, pharmacie, etc. ainsi que des élèves
des seconds cycles des lycées et des classes préparatoires
aux grandes écoles scientifiques

Avec la collaboration de François Bry

*Hermann Paris* — *Collection Méthodes*

EMILE AMZALLAG, né en 1932 au Maroc, ingénieur de l'Ecole supérieure d'électricité, docteur ès sciences, est maître-assistant à l'Université de Paris VI et chargé de cours à l'Université de Paris V.

NORBERT PICCIOLI, né en 1941 en Algérie, docteur de 3e cycle, est maître-assistant à l'Université de Paris VI.

Tous deux effectuent des recherches sur les propriétés optiques des semi-conducteurs au Laboratoire de physique des solides de l'Université Pierre et Marie Curie.

FRANCOIS BRY, né en 1956 à Paris, a passé la maîtrise de mathématiques en 1978 ; il a fait ses études à l'UER des mathématiques de la décision à l'Université Paris XI Dauphine, où il s'est initié aux mathématiques appliquées à l'économie avec I. Ekeland ; il se consacre à présent aux mathématiques pures.

ISBN 2 7056 5889 0

# Table

# Avant-propos

Cet ouvrage est destiné aux étudiants des premiers cycles universitaires, aux élèves des classes préparatoires aux grandes écoles scientifiques, ainsi qu'à tous ceux qui désirent s'initier au calcul des probabilités et à la statistique. Il intéressera spécialement les étudiants en sciences économiques et ceux du premier cycle des études médicales et dentaires (P.C.E.M) ou de pharmacie.

C'est avant tout un livre d'initiation qui vise à l'acquisition de techniques de base, plutôt qu'à l'étude de théories mathématiques fines ou de problèmes philosophiques posés par les notions de probabilité ou d'induction statistique. Afin de répondre aux difficultés que rencontrent les étudiants pour passer du cours aux applications, l'ouvrage réunit des rappels détaillés de cours visant à familiariser le lecteur avec les notions essentielles, de nombreux exemples d'application, ainsi qu'une centaine d'exercices classés par ordre de difficulté croissante et suivis de corrigés succints, permettant de mettre en pratique et de contrôler les connaissances.

On expose d'abord les grandes lignes de la statistique descriptive, où il s'agit essentiellement de présenter les données sous une forme immédiatement exploitable, en les réduisant à quelques paramètres caractéristiques. Après des rappels d'analyse combinatoire, on introduit les principes généraux du calcul des probabilités, en montrant les possibilités d'utilisation de l'algèbre des ensembles qui est de plus en plus familière aux étudiants. Les différentes lois de probabilité usuelles sont ensuite étudiées et leurs conditions d'application examinées. La dernière partie de l'ouvrage introduit à la statistique inductive qui, grâce à l'assimilation des observations expérimentales aux lois théoriques et à l'application de tests, fournit des éléments de décision.

Les exercices proposés se rapportent à des domaines variés : économie, médecine, jeux, etc... De nombreux problèmes proposés ces dernières années aux concours de P.C.E.M sont donnés avec leurs corrigés.

# 1. Introduction

## I. LA STATISTIQUE

De nombreux domaines de la connaissance pratique s'appuient sur l'étude de collections homogènes d'objets ou de personnes. La statistique est un ensemble de méthodes permettant de dégager les caractéristiques ou la répartition de ces objets en fonction de critères d'étude déterminés.

Ces méthodes tirent leur justification théorique de certaines constructions mathématiques (théorie des probabilités, algèbre linéaire, etc.), mais c'est le domaine d'application qui justifie le choix de la méthode et l'interprétation des résultats obtenus. Il est relativement fréquent que des conclusions erronées soient tirées d'une étude statistique parfaitement cohérente en théorie (c'est souvent le cas des sondages d'opinion en période électorale). Il est donc essentiel de ne pas réduire la statistique à l'application mécanique de formules.

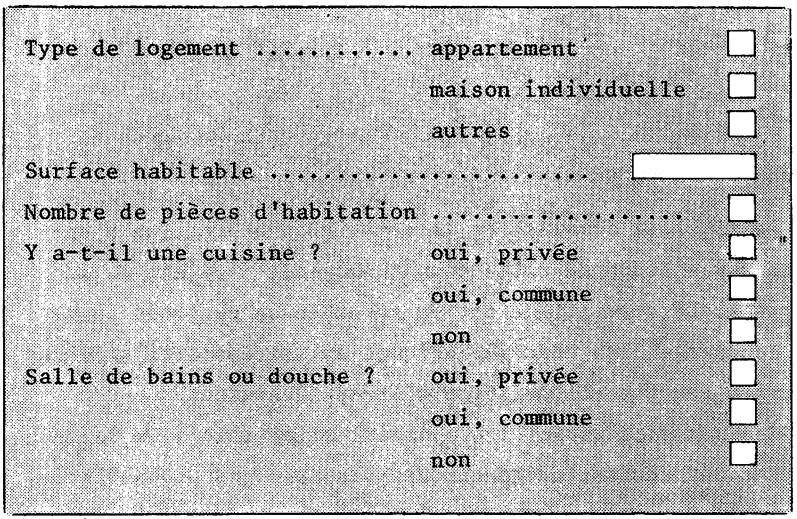
## II. NOTIONS DE BASE

La collection d'objets ou de personnes étudiée est appelée population ou univers.

Un objet ou une personne sur lesquels porte l'étude est appelé individu (individu statistique).

Les critères étudiés constituent des caractères.

On peut préciser ces notions sur l'exemple suivant, extrait d'une feuille de recensement relative aux logements occupés par les ménages dans une commune industrielle du Nord de la France.




*Questionnaire (partiel) relatif aux logements occupés par les ménages.*

La population statistique est constituée ici par l'ensemble des ménages de la commune. Les caractères, qui correspondent aux questions posées sont de deux types :

- on ne peut associer à certains d'entre eux ni une valeur numérique, ni un ordre naturel (par exemple : le type de logement). De tels caractères sont appelés caractères qualitatifs ;

- certains caractères prennent des valeurs numériques (par exemple : le nombre de pièces d'habitation). Ce sont des caractères quantitatifs.

Un caractère continu est un caractère quantitatif qui peut prendre toutes les valeurs numériques d'un intervalle déterminé (par exemple : la surface habitable).

Un caractère discret (ou discontinu) est un caractère qui ne peut prendre que des valeurs numériques isolées dans un intervalle (par exemple : le nombre de pièces d'habitation).

Un tel recensement peut être général (et porter par exemple sur l'ensemble des ménages d'une grande ville) ou partiel (ne porter que sur une partie seulement de ces ménages). D'une manière générale, on appelle échantillon la partie de la population statistique sur laquelle porte l'enquête.

L'effectif ou fréquence absolue associée à une valeur d'un caractère est le nombre de fois où cette valeur du caractère a été observée.

La fréquence relative associée à une valeur d'un caractère est le rapport de la fréquence absolue correspondant à cette valeur du caractère au nombre d'individus de l'échantillon.

Une série statistique, ou distribution statistique, associée à un caractère, est l'ensemble des valeurs du caractère, avec en regard, les fréquences absolues ou relatives correspondantes.

Les statistiques désignent communément les données relatives à une même caractéristique ou encore les résultats obtenus à partir de ces données. Par exemple : les statistiques de l'emploi ou du chômage, les statistiques d'une certaine maladie.

## III. LA METHODE STATISTIQUE

D'une manière générale, la statistique considère des phénomènes qui ne sont pas toujours accessibles à l'expérience. Par suite de la multiplicité des causes, on ne peut comme en physique par exemple, fixer un certain nombre de paramètres et étudier l'évolution du phénomène. La méthode statistique comporte essentiellement trois phases :

- une phase matérielle où il s'agit de rassembler des

données, de les regrouper et de les présenter sous forme de tableaux ou graphes ;

- une phase analytique qui consiste à réduire les données à un nombre limité de paramètres caractéristiques (moments d'ordre 1, 2, 3, ...) susceptibles de décrire la série statistique. L'ensemble de ces deux phases constitue l'objet essentiel de la statistique descriptive (ou déductive) dont les résultats restent limités aux échantillons étudiés ;

- une phase interprétative, qui est à la base de la statistique inductive, et qui permet de déduire des résultats obtenus sur un échantillon des conclusions relatives à l'ensemble de la population d'où est extrait cet échantillon. Ces conclusions doivent tenir compte de la marge d'erreur due au fait que les données sont seulement partielles. Les méthodes utilisées n'ont de sens que si elles sont justifiées par des résultats ultérieurs.

# 2. Tableaux et graphes

Une enquête statistique comporte toujours une phase initiale où il s'agit de collecter des renseignements, suivie d'une phase de dépouillement qui consiste à passer des données brutes à des tableaux ou à des graphes qui se prêtent mieux à l'analyse et l'interprétation.

La manière dont l'enquête est effectuée est évidemment très importante. En particulier, si l'on espère déduire des résultats obtenus sur un échantillon des conclusions relatives à toute la population, il convient de s'assurer que l'échantillon est bien représentatif de cette population, ce qui sera précisé dans la théorie de l'échantillonnage (cf. chap. 7). Cependant l'objet du présent chapitre sera limité à l'étude des différentes manières de présenter une série statistique.

## A. TABLEAUX

### I. TABLEAU DE FREQUENCES A UN CARACTERE

Ce tableau établit la correspondance entre deux séries de nombres, l'une constituée par les valeurs du caractère étudié, l'autre par les effectifs correspondants (ou les fréquences relatives correspondantes).

. *Exemple d'une série quantitative discrète*

L'échantillon est un immeuble de 64 familles, le caractère étudié étant le nombre d'enfants par famille.

| Nombre d'enfants | 0 | 1 | 2 | 3 | 4 | 5 | Total |
|---|---|---|---|---|---|---|---|
| Nombre de familles | 16 | 18 | 14 | 11 | 3 | 2 | 64 |
| Fréquence relative | 0,250 | 0,281 | 0,218 | 0,172 | 0,047 | 0,031 | 1 |

Tableau 2.1

. *Exemple d'une série qualitative*

L'échantillon est l'immeuble de 64 familles de l'exemple précédent, le caractère est la profession du chef de famille, à laquelle on attribue un code d'une manière arbitraire.

La correspondance valeur du caractère - effectif correspondant, définit une fonction dite fonction de distribution, qui sera développée au chapitre 3.

| Profession | Code | Effectif | Fréquence relative |
|---|---|---|---|
| Ouvriers et employés | 1 | 24 | 0,375 |
| Cadres moyens et supérieurs | 2 | 9 | 0,140 ... |
| Commerçants | 3 | 10 | 0,156 ... |
| Fonctionnaires | 4 | 15 | 0,234 ... |
| Professions libérales | 5 | 6 | 0,093 ... |
| | | 64 | 1 |

Tableau 2.2

## II. TABLEAU DE FREQUENCES CUMULEES

La série statistique du tableau 2.1 peut être présentée sous une forme dite cumulée, des deux manières suivantes :

. *Cumul par valeurs inférieures* ou *effectifs cumulés croissants*

| Nombre d'enfants | Effectifs cumulés |
|---|---|
| Moins de 1 enfant | 16 familles |
| " 2 enfants | 16 + 18 = 34 " |
| " 3 " | 34 + 14 = 48 " |
| " 4 " | 48 + 11 = 59 " |
| " 5 " | 59 + 3 = 62 " |
| " 6 " | 62 + 2 = 64 " |

Tableau 2.3

. *Cumul par valeurs supérieures ou effectifs cumulés décroissants*

| Nombre d'enfants | Effectifs cumulés |
|---|---|
| 0 enfant ou plus | 64 familles |
| 1 enfant " " | 64 - 16 = 48 " |
| 2 enfants " | 48 - 18 = 30 " |
| 3 " " " | 30 - 14 = 16 " |
| 4 " " " | 16 - 11 = 5 " |
| 5 " " " | 5 - 3 = 2 " |

Tableau 2.4

La correspondance valeur du caractère - effectif cumulé correspondant, définit une fonction dite de répartition, qui sera également développée au chapitre 3.

## III. TABLEAU DE FREQUENCES A DEUX CARACTERES

Si l'on s'intéresse à deux caractères différents dans un même échantillon, il est possible de représenter l'ensemble des renseignements dans un même tableau (tableau à double entrée).

. *Exemple de tableau de fréquences à deux caractères*

L'échantillon est le même que précédemment. Le caractère X est le nombre de personnes vivant dans un appartement, le caractère Y est le nombre de pièces par appartement. L'intersection d'une ligne et d'une colonne du tableau est le nombre de fois où l'on a observé X personnes vivant dans un appartement de Y pièces. Par exemple, on a observé 7 fois 3 personnes vivant dans un appartement de 3 pièces.

| Y \ X | 2 | 3 | 4 | 5 | 6 | Total nombre d'appartements |
|---|---|---|---|---|---|---|
| 2 | 8 | 5 | 2 | 0 | 0 | 15 |
| 3 | 5 | 7 | 4 | 2 | 0 | 18 |
| 4 | 3 | 6 | 8 | 9 | 5 | 31 |
| Total nombre de familles | 16 | 18 | 14 | 11 | 5 | 64 |

Tableau 2.5

## IV. CAS D'UNE SERIE QUANTITATIVE CONTINUE

Afin de rendre la série statistique plus commode à étudier, il est nécessaire de regrouper les valeurs du caractère en intervalles successifs et contigus, tels que dans chaque intervalle ou classe, on ne distingue pas les valeurs du caractère qui y sont comprises. Les nombres entre lesquels sont comprises ces valeurs constituent les limites de classe. Dans chaque classe, on remplace les valeurs du caractère observées par une valeur unique, celle du milieu de l'intervalle ou centre de classe.

. *Exemple*

Une enquête portant sur la taille des individus d'une certaine collectivité de 80 personnes a permis de dresser le tableau suivant où l'on a adopté un intervalle de classe de 0,05 m.

| Classes | Limites | Centres de classe | Effectifs | Effectifs cumulés croissants | Effectifs cumulés décroissants |
|---|---|---|---|---|---|
| | 1,545 | | | | 80 |
| 1,55-1,59 | | 1,57 | 3 | | |
| | 1,595 | | | 3 | 77 |
| 1,60-1,64 | | 1,62 | 12 | | |
| | 1,645 | | | 15 | 65 |
| 1,65-1,69 | | 1,67 | 18 | | |
| | 1,695 | | | 33 | 47 |
| 1,70-1,74 | | 1,72 | 25 | | |
| | 1,745 | | | 58 | 22 |
| 1,75-1,79 | | 1,77 | 15 | | |
| | 1,795 | | | 73 | 7 |
| 1,80-1,84 | | 1,82 | 5 | | |
| | 1,845 | | | 78 | 2 |
| 1,85-1,89 | | 1,87 | 2 | | |
| | 1,895 | | | 80 | |
| | | | 80 | | |

Tableau 2.6

On peut remarquer que les effectifs cumulés croissants correspondent aux frontières supérieures des classes, et les effectifs cumulés decroissants aux frontières inférieures des classes.

## B. REPRESENTATIONS GRAPHIQUES

Les représentations graphiques ont l'avantage d'offrir une meilleure vue d'ensemble de la série statistique que les tableaux. Elles permettent par simple lecture, de voir les caractéristiques essentielles de la série, et aussi de comparer des séries différentes.

### I. CARACTERE DISCONTINU. DIAGRAMME EN BÂTONS

Lorsque le caractère est discontinu, on utilise le diagramme en bâtons : les valeurs du caractère sont portées en abscisses, les fréquences correspondantes sont représentées par des traits pleins, en ordonnées.

Si l'on joint les sommets des bâtons, on obtient le polygone des fréquences.

*Exemple 1*

Diagramme en bâtons du tableau 2.1 repris en 2.7

| Nombre d'enfants | Fréquences | Fréquences relatives |
|---|---|---|
| 0 | 16 | 0,250 |
| 1 | 18 | 0,281 |
| 2 | 14 | 0,218 |
| 3 | 11 | 0,172 |
| 4 | 3 | 0,047 |
| 5 | 2 | 0,031 |
| | 64 | 1 |

Tableau 2.7

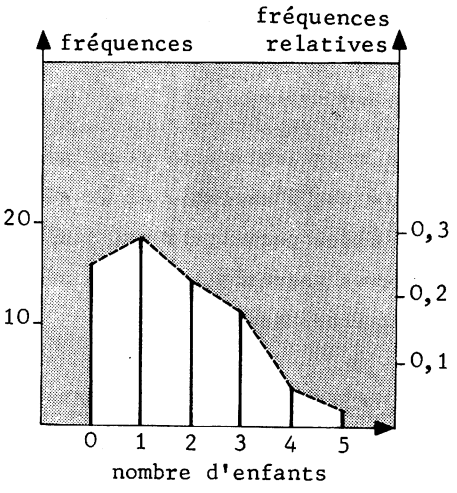



Figure 2.1

*Exemple 2*

Diagramme des fréquences cumulées des tableaux 2.3 et 2.4, repris en 2.8 et 2.9.

| Nombre d'enfants | Fréquences cumulées croissantes |
|---|---|
| Moins de 1 | 16 |
| " 2 | 34 |
| " 3 | 48 |
| " 4 | 59 |
| " 5 | 62 |
| " 6 | 64 |

Tableau 2.8

| Nombre d'enfants | Fréquences cumulées décroissantes |
|---|---|
| 0 ou plus | 64 |
| 1 " | 48 |
| 2 " | 30 |
| 3 " | 16 |
| 4 " | 5 |
| 5 " | 2 |

Tableau 2.9

Le graphe des fréquences cumulées (appelé aussi diagramme intégral) ne met pas en évidence les différences et ne fait pas ressortir la fréquence maximum (fig. 2.2).

Pour chaque valeur du caractère la somme des fréquences cumulées croissante et décroissante est évidemment égale à l'effectif total.

Ces graphes étant constitués par un ensemble discontinu de points (les sommets des bâtons), l'interpolation entre points successifs n'a pas de sens. On peut aussi bien adopter une fréquence constante entre les valeurs discrètes du caractère (fig. 2.3).

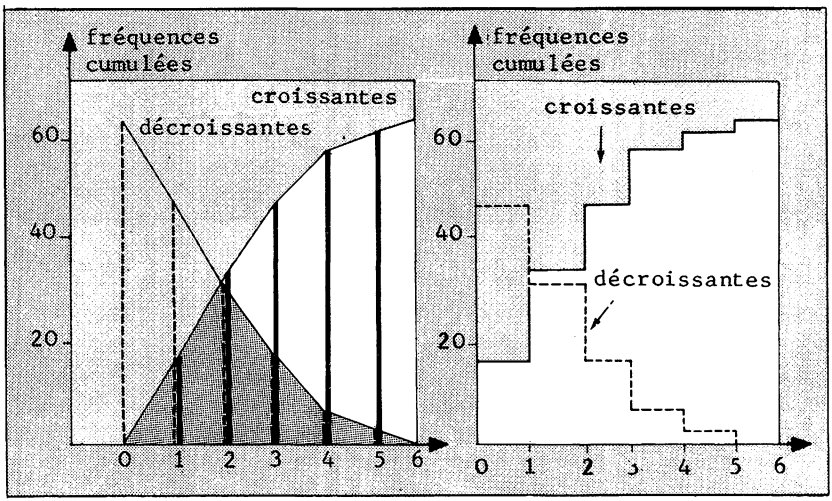



Figure 2.2




Figure 2.3

## II. CARACTERE CONTINU. HISTOGRAMME

Dans le cas d'un catactère continu, on utilise l'histogramme, qui constitue une généralisation du diagramme en bâtons à la notion de classe.

a) Séries à classes égales

Chaque classe est représentée par un rectangle dont la base est égale à l'intervalle de la classe et dont la hauteur

est égale à l'effectif correspondant. L'histogramme est constitué en fait par le contour polygonal enveloppant l'ensemble de ces rectangles.

Le polygone des fréquences absolues (ou des fréquences relatives) s'obtient en joignant les points dont les abscisses sont les milieux des différentes classes et dont les ordonnées sont les effectifs (ou les fréquences relatives) correspondants.

. *Exemple*

Histogramme relatif au tableau 2.6 (repris en 2.10).

| Classes | Limites | Centres de classe | Effectifs | Effectifs cumulés croissants | Effectifs cumulés décroissants |
|---|---|---|---|---|---|
| | 1,545 | | | | 80 |
| 1,55-1,59 | | 1,57 | 3 | | |
| | 1,595 | | | 3 | 77 |
| 1,60-1,64 | | 1,62 | 12 | | |
| | 1,645 | | | 15 | 65 |
| 1,65-1,69 | | 1,67 | 18 | | |
| | 1,695 | | | 33 | 47 |
| 1,70-1,74 | | 1,72 | 25 | | |
| | 1,745 | | | 58 | 22 |
| 1,75-1,79 | | 1,77 | 15 | | |
| | 1,795 | | | 73 | 7 |
| 1,80-1,84 | | 1,82 | 5 | | |
| | 1,845 | | | 78 | 2 |
| 1,85-1,89 | | 1,87 | 2 | | |
| | 1,895 | | | 80 | |
| | | | 80 | | |

Tableau 2.10

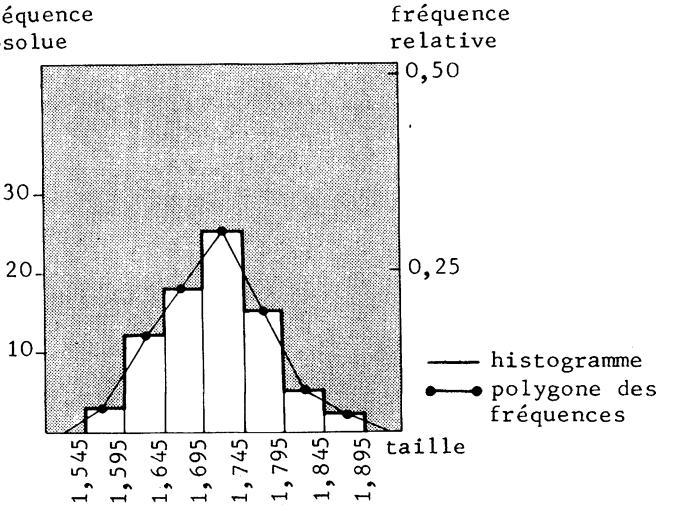



Figure 2.4

. *Aire de l'histogramme*

C'est l'aire comprise entre l'histogramme et l'axe des abscisses. Dans le cas d'un histogramme des fréquences absolues, cette aire est évidemment proportionnelle au produit de l'intervalle de classe par l'effectif total. Dans le cas d'un histogramme des fréquences relatives, si l'intervalle de classe est pris comme unité, l'aire est égale à l'unité, puisque la somme des fréquences relatives est elle-même égale à l'unité. L'aire comprise entre le polygone des fréquences et l'axe des abscisses est égale à celle de l'histogramme, puisque, comme on peut le voir sur la figure 2.4, les surfaces non communes à ces deux aires se compensent deux par deux.

. *Polygone des effectifs cumulés*

En observant que les effectifs cumulés croissants correspondent aux frontières supérieures des différentes classes, et les effectifs cumulés décroissants aux frontières infé-

rieures des classes, on peut, à l'aide du tableau 2.10, construire les polygones des effectifs cumulés suivants :

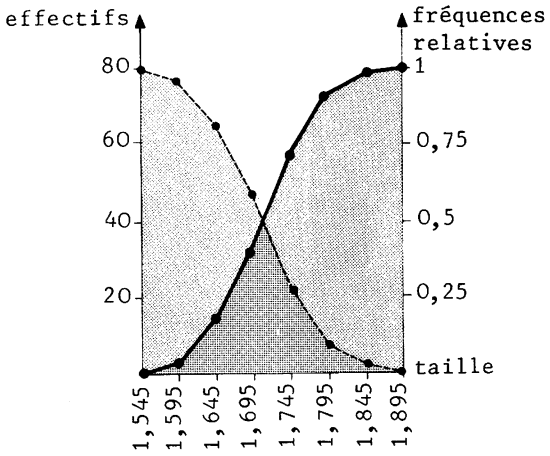



Figure 2.5

On peut remarquer que pour chaque valeur du caractère, la somme des effectifs croissants et des effectifs décroissants est égale à l'effectif total.

Remarque

L'intérêt de ces représentations cumulées apparaîtra un peu plus tard, à l'occasion de l'analyse des séries statistiques.

b) Séries à classes inégales

Si l'on veut que l'aire de l'histogramme soit toujours proportionnelle à l'effectif, il est nécessaire de tenir compte de l'inégalité des classes. On opère alors de la manière suivante : une classe dont l'étendue est égale à n fois l'intervalle de classe fondamental, est représentée avec une ordonnée égale à l'effectif de cette classe divisé par n, de telle sorte que l'aire relative à cette classe soit bien proportionnelle à son effectif.

. *Exemple*

Une enquête portant sur les salaires mensuels perçus dans une certaine entreprise a fourni les renseignements suivants :

| Salaire mensuel en F. | Effectifs |
|---|---|
| Entre 2 000 et 3 000 | 34 |
| " 3 000 et 4 000 | 52 |
| " 4 000 et 5 000 | 60 |
| " 5 000 et 6 000 | 20 |
| " 6 000 et 7 000 | 8 |
| " 7 000 et 11 000 | 6 |
| | 180 |

Tableau 2.11

La dernière classe, de 7 000 à 11 000 F vaut 4 intervalles de classe, il faut donc réaménager sa présentation ainsi :

| Salaire mensuel en F. | Effectifs |
|---|---|
| Entre 7 000 et 8 000 | 1,5 |
| " 8 000 et 9 000 | 1,5 |
| " 9 000 et 10 000 | 1,5 |
| " 10 000 et 11 000 | 1,5 |

Tableau 2.12

Cette répartition de l'effectif dans la dernière classe ne correspond bien sûr à aucune réalité, elle conserve toute-

fois un sens à l'histogramme et aux conclusions qu'on peut en tirer.

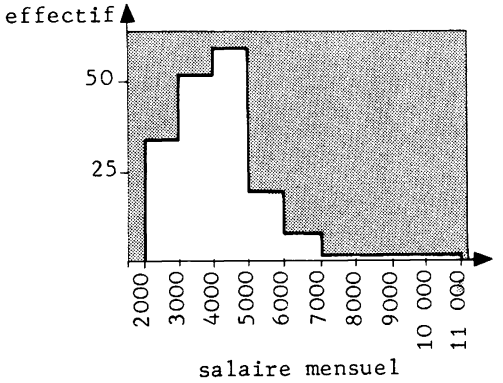


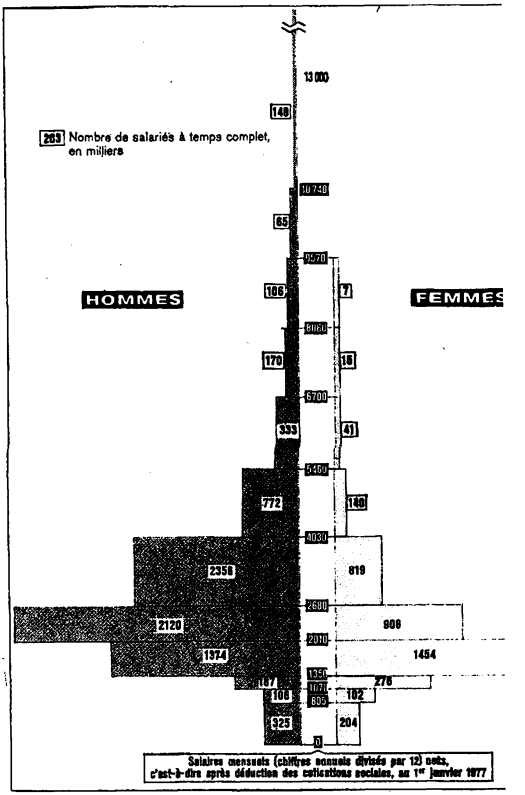

Figure 2.6

. *Autre exemple*

Salaires mensuels perçus à l'échelle nationale dans l'industrie et le commerce

(Le Monde 8/3/77)

Les classes adoptées sont inégales. Les ordonnées (axe horizontal) ne sont pas proportionnelles à l'effectif. Elles sont obtenues en divisant l'effectif (représenté par les nombres encadrés) par la classe correspondante.



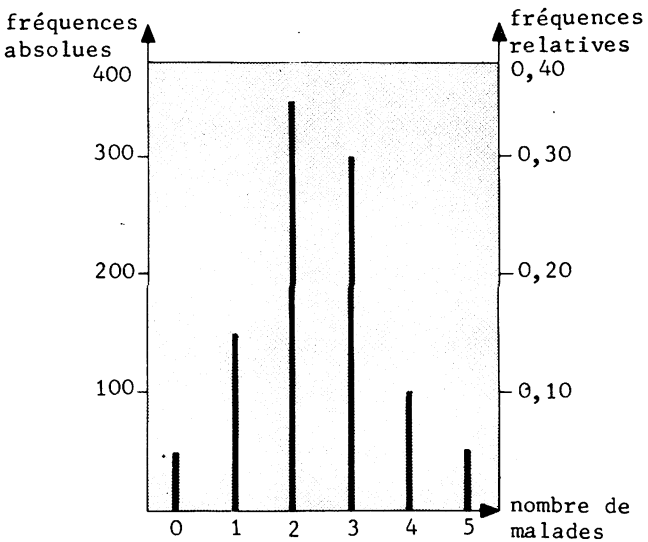
I. On recense dans 1 000 hôpitaux d'un pays européen le cas d'une maladie inconnue. On trouve les résultats suivants :

| Nombre de malades | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Nombre d'hôpitaux | 50 | 150 | 350 | 300 | 100 | 50 |

Représenter graphiquement les données. Calculer les fréquences relatives.

---

SOLUTION



| Nombres de malades | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Nombre d'hôpitaux | 50 | 150 | 350 | 300 | 100 | 50 |
| Fréquences relatives | 0,05 | 0,15 | 0,35 | 0,30 | 0,10 | 0,05 |

On remarque que la somme des fréquences absolues est 1 000 et que la somme des fréquences relatives est 1. ■

---

II. Reprendre la série de l'exercice I précédent.

1°) Construire un tableau donnant le pourcentage d'hôpitaux où le nombre de malades est inférieur à 0, 1, 2, 3, 4, 5 ou 6. Représenter graphiquement les données de ce tableau.

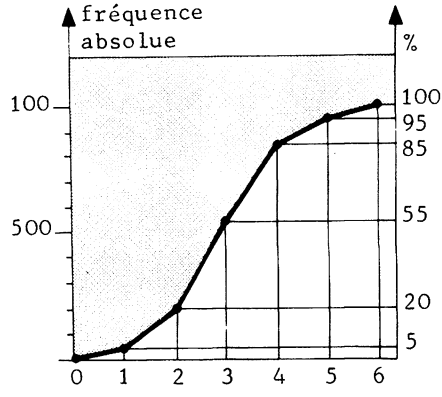
2°) Construire un tableau donnant le pourcentage d'hôpitaux où le nombre de malades est égal ou supérieur à 0, 1, 2, 3, 4, 5 ou 6, en faire la représentation graphique.

---

SOLUTION

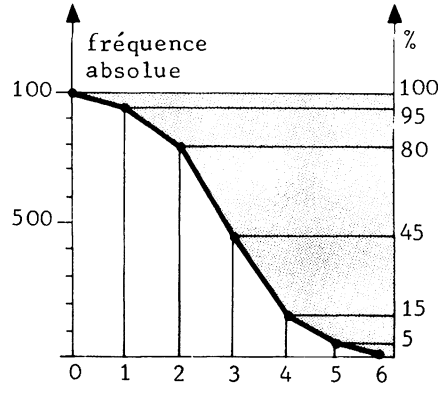
1°)

| Nombre de malades inférieur à | Nombre d'hôpitaux | Pourcentage |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 50 | 5 |
| 2 | 200 | 20 |
| 3 | 550 | 55 |
| 4 | 850 | 85 |
| 5 | 950 | 95 |
| 6 | 1 000 | 100 |




2°)

| Nombre de malades égal ou supérieur à | Nombre d'hôpitaux | Pourcentage |
|---|---|---|
| 0 | 1 000 | 100 |
| 1 | 950 | 95 |
| 2 | 800 | 80 |
| 3 | 450 | 45 |
| 4 | 150 | 15 |
| 5 | 50 | 5 |
| 6 | 0 | 0 |



III. Dans une usine, on a relevé les horaires d'arrivée de 800 personnes

| $t_i$ = classe des temps en heures | [8 h 45 8 h 50] | [8 h 50 8 h 55] | [8 h 55 9 h 00] | [9 h 00 9 h 05] | [9 h 05 9 h 10] |
|---|---|---|---|---|---|
| $n_i$ = nombre de personnes | 4 | 10 | 26 | 110 | 150 |

| $t_i$ = classe des temps en heures | [9 h 10 9 h 15] | [9 h 15 9 h 20] | [9 h 20 9 h 25] | [9 h 25 9 h 30] | [9 h 30 9 h 35] |
|---|---|---|---|---|---|
| $n_i$ = nombre de personnes | 200 | 150 | 100 | 40 | 10 |

1°) Représenter l'histogramme de cette distribution.

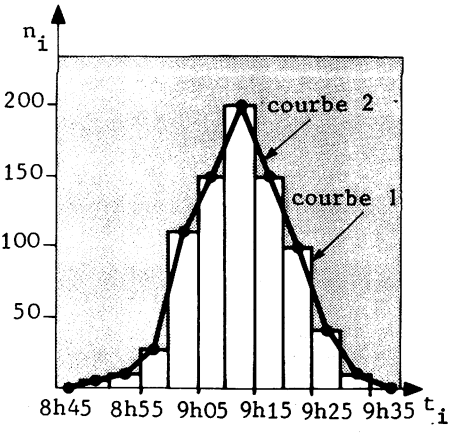
2°) Tracer le polygone de fréquences

---

SOLUTION

1°) Histogramme : courbe 1

2°) *Polygone de fréquences* : courbe 2




■

IV. On effectue l'analyse de sang de 60 personnes qui ont manipulé un gaz toxique. La mesure du taux de leucocytes, par $mm^3$, donne les résultats suivants :

| | |
|---|---|
| $3000 \leqslant X_i < 4000$ | $n_i = 10$ |
| $4000 \leqslant X_i < 10000$ | $n_i = 48$ |
| $10000 \leqslant X_i < 12000$ | $n_i = 12$ |

$n_i$ représente le nombre de gens dont le taux de leucocytes par $mm^3$ est $X_i$. Représenter l'histogramme de cette série quantitative à classes inégales.

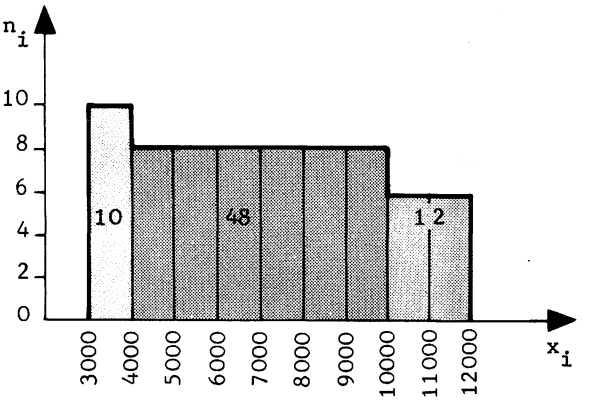
---

SOLUTION

En adoptant un intervalle de classe de 1000, la 2ème classe vaut 6 intervalles de classe et la 3ème en vaut 2. En effectuant une équipartition de l'effectif entre les différents intervalles de classe, on obtient le tableau suivant :

| $X_i$ | $n_i$ |
|---|---|
| $3000 \leqslant X_i < 4000$ | 10 |
| $4000 \leqslant X_i < 5000$ | 8 |
| $5000 \leqslant X_i < 6000$ | 8 |
| $6000 \leqslant X_i < 7000$ | 8 |
| $7000 \leqslant X_i < 8000$ | 8 |
| $8000 \leqslant X_i < 9000$ | 8 |
| $9000 \leqslant X_i < 10000$ | 8 |
| $10000 \leqslant X_i < 11000$ | 6 |
| $11000 \leqslant X_i < 12000$ | 6 |

d'où l'histogramme :



# 3. Analyse d'une distribution de fréquences

Les points de départ de cette analyse ont été exposés dans le chapitre précédent. Ils consistent à regrouper les données, en procédant éventuellement à un découpage en classes (caractère continu), et à présenter sous forme de tableaux ou de graphes, les fonctions de distribution et de répartition correspondantes.

*Fonction de distribution* : on désigne ainsi l'ensemble des couples constitués par les valeurs du caractère et les fréquences absolues ou relatives correspondantes. La représentation graphique de cette fonction - appelée parfois diagramme différentiel - n'est autre que l'histogramme (voir fig. 2.4) dont une propriété essentielle est que les aires des différents rectangles sont proportionnelles aux effectifs correspondants.

*Fonction de répartition* : c'est l'ensemble des couples constitués par une valeur du caractère et

a) soit la somme des effectifs ayant moins que cette valeur du caractère (diagramme intégral croissant)

b) soit la somme des effectifs ayant cette valeur du caractère ou plus (diagramme intégral décroissant).

Les graphes sont les courbes cumulatives respectivement par valeurs inférieures et par valeurs supérieures décrites précédemment (voir tableau 2.10 et fig. 2.5).

La phase suivante est celle de la réduction des données, qui consiste à substituer à la distribution étudiée, quelques paramètres en nombre réduit, dont les valeurs numériques donneront un résumé relativement suffisant de l'information contenue dans la distribution de fréquences. Parmi ces paramètres, on distingue :

a) les paramètres de position (moyenne, médiane, etc.) qui permettent de se rendre compte de l'ordre de grandeur de l'ensemble des observations et de localiser la zone des fréquences maximum ;

b) les paramètres de dispersion (écart moyen, écart type, et qui précisent le degré de dispersion des différentes observations autour d'une valeur centrale.

# A. PARAMETRES DE POSITION

## I. LES MOYENNES

La notion de moyenne est assez commune : il est fréquent de réduire un ensemble fini de nombres à une "valeur moyenne" afin de donner une idée de l'ordre de grandeur des éléments de cet ensemble. Cependant, il existe plusieurs manières de calculer une "valeur moyenne" suivant sa signification.

### 1. La moyenne arithmétique

Symbole de sommation :

Soit $x_1$, $x_2$, $x_3$ ... $x_n$ une suite finie de nombres. Par définition

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \ldots + x_n$$

Les propriétés de ce symbole de sommation sont les suivantes :

- si a est une constante, d'après la définition précédente, on a immédiatement :

$$\sum_{i=1}^{n} a\, x_i = a \sum_{i=1}^{n} x_i$$

- en considérant le cas où $x_1 = x_2 = x_3 = \ldots = x_n$, on en déduit :

$$\sum_{i=1}^{n} a = n\, a$$

- soit $y_1$, $y_2$, $y_3$ ... $y_n$ une autre suite finie de nombres. D'après la même définition, on a

$$\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$$

Définition de la moyenne arithmétique.

Soit $x_1$, $x_2$, $x_3$ ... $x_n$ une suite finie de nombres. La moyenne arithmétique est le rapport :

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad (3.1)$$

Si chaque valeur $x_i$ apparaît $n_i$ fois dans la série, on peut encore écrire

$$\bar{X} = \frac{1}{n} \sum_{i} n_i x_i \qquad (3.2)$$

En remarquant que $n_i/n$ n'est autre que la fréquence relative $f_i$ correspondant à la valeur $x_i$, on a aussi

$$\bar{X} = \sum_{i} f_i x_i \qquad (3.3)$$

. *Exemple*

Cas du tableau 2.1 repris en 3.1

| Nombre d'enfants | 0 | 1 | 2 | 3 | 4 | 5 | Total |
|---|---|---|---|---|---|---|---|
| Nombre de familles | 16 | 18 | 14 | 11 | 3 | 2 | 64 |
| Fréquence relative | 0,250 | 0,281 | 0,218 | 0,172 | 0,047 | 0,031 | 1 |

Tableau 3.1

Le nombre moyen d'enfants par famille est, d'après l'équation (3.2)

$$\bar{X} = \frac{(16\times0)+(18\times1)+(14\times2)+(11\times3)+(3\times4)+(2\times5)}{64} \simeq 1,58 \text{ enfant}$$

ou encore, à l'aide de l'expression (3.3)

$$\bar{X} = (0,25\times0)+(0,281\times1)+(0,218\times2)+(0,172\times3)+(0,047\times4)+(0,031\times5)$$
$$\simeq 1,58 \text{ enfant.}$$

2. Cas de données groupées en classes

On prend pour valeur de $x_i$ les centres de classes.

. *Exemple*

Extrait du tableau 2.6.

| Centres de classes | 1,57 | 1,62 | 1,67 | 1,72 | 1,77 | 1,82 | 1,87 |
|---|---|---|---|---|---|---|---|
| Effectif | 3 | 12 | 18 | 25 | 15 | 5 | 2 |

Tableau 3.2

La taille moyenne d'un individu dans cette collectivité est :

$$\bar{X} = \frac{(3\times1,57)+(12\times1,62)+(18\times1,67)+(25\times1,72)+(15\times1,77)+(5\times1,82)+(2\times1,87)}{80}$$

$$= 1,707 \text{ m}$$

3. Simplification du calcul de la moyenne

a) Changement d'origine

On prend une moyenne provisoire arbitraire $x_o$, qu'on estime être aussi proche que possible de $\bar{X}$, on substitue alors à la variable $x_i$, la variable

$$u_i = x_i - x_o \qquad (3.4)$$

L'expression (3.2) devient

$$\bar{X} = \frac{1}{n} \sum n_i x_i = \frac{1}{n} \sum n_i (u_i + x_o)$$

$$\bar{X} = \bar{u} + x_o \qquad (3.5)$$

Le calcul de $\bar{X}$ revient donc au calcul de $\bar{u}$ qui peut être plus simple (cf. exercices 3 A)

b) Changement d'origine et d'échelle

Si l'intervalle de classe k est constant, on peut le prendre comme nouvelle unité et introduire le changement de variable

$$z_i = \frac{x_i - x_o}{k} \qquad (3.6)$$

On a alors

$$\bar{X} = \frac{1}{n} \sum n_i (x_o + k z_i)$$

$$\bar{X} = x_o + k \bar{z} \qquad (3.7)$$

Le calcul de $\bar{X}$ revient à celui de $\bar{z}$ qui peut encore être plus simple (cf. exercices 3 A).

## 4. Propriétés de la moyenne arithmétique

a) La somme des déviations d'un ensemble de données $x_i$, par rapport à leur valeur moyenne $\bar{X}$, est nulle. En effet :

$$\sum n_i (x_i - \bar{X}) = \sum n_i x_i - n \bar{X} = 0 \qquad (3.8)$$

puisque $\bar{X} = \frac{1}{n} \sum n_i x_i$.

b) La moyenne arithmétique des déviations $x_i - x_o$ est égale à la déviation de la moyenne arithmétique des $x_i$ par rapport à $x_o$.

$$\frac{1}{n} \sum n_i (x_i - x_o) = \frac{1}{n} \sum n_i x_i - \frac{x_o}{n} \sum n_i = \bar{X} - x_o$$

5. La moyenne géométrique

Considérons une population (au sens classique) qui s'accroît suivant une progression géométrique. Ce sera le cas, par exemple, d'une population dont les taux de natalité et de mortalité sont constants pendant la période envisagée.

Soient r le taux de croissance sur une période

$x_o$ l'effectif de la population à la date initiale $t_o$

$x_n$ l'effectif de la population à la date $t_n$, c'est à dire au bout de n périodes.

On a alors

$x_1 = x_o\ r$

$x_2 = x_1\ r = x_o\ r^2$, etc. et par récurrence

$x_n = x_o\ r^n$

Si n est pair, le milieu des n périodes sera la date $t_p$ où $p = n/2$.

On appelle moyenne géométrique des $(x_i)$ le terme correspondant à $t_p$, soit

$$g = x_o\ r^{n/2}$$

On montre que cette moyenne géométrique est donnée par

$$g = (x_o \times x_1 \times x_2 \times \ldots \times x_n)^{1/n+1} \qquad (3.9)$$

En effet cette dernière expression s'écrit

$$g = (x_o \times x_o\ r \times x_o\ r^2 \times \ldots \times x_o\ r^n)^{1/n+1}$$

$$= (x_o^{n+1} \times r^{1+2+\ldots+n})^{1/n+1}$$

et en utilisant la relation $1+2+\ldots+n = \frac{n\ (n+1)}{2}$ on obtient

$$g = (x_o^{n+1} \times r^{n\ (n+1)/2})^{1/n+1} = x_o\ r^{n/2}$$

On peut remarquer, à partir de l'expression (3.9), que le logarithme de g n'est autre que la moyenne arithmétique des logarithmes des $(x_i)$.

## 6. Moyenne harmonique

Soit une suite finie de nombres

$\{x_1, x_2, x_3 \ldots x_n\}$

et l'ensemble des inverses de ces nombres

$$\{\frac{1}{x_1}, \frac{1}{x_2}, \frac{1}{x_3}, \ldots, \frac{1}{x_n}\}$$

Il arrive que ce soit la moyenne arithmétique de ces inverses qui ait une signification, plutôt que celle de l'ensemble initial, par exemple dans le cas de grandeurs inversement proportionnelles.

On appelle moyenne harmonique des $x_i$ (supposés non nuls) l'inverse de la moyenne arithmétique des inverses $(\frac{1}{x_i})$, soit

$$h = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \qquad (3.10)$$

. *Exemple*

Une voiture parcourt un circuit fermé à une vitesse de 10 km/h durant le 1er tour, de 20 km/h durant le second, de 30 km/h durant le 3ème tour. Déterminer la vitesse moyenne de la voiture durant les trois tours.

Soit $\ell$ la longueur du circuit (en km). La dutée totale des 3 parcours est

$$t = \frac{\ell}{10} + \frac{\ell}{20} + \frac{\ell}{30}$$

La vitesse moyenne v recherchée est donc

$$v = \frac{3\,\ell}{t} = \frac{3\,\ell}{\frac{\ell}{10} + \frac{\ell}{20} + \frac{\ell}{30}} = \frac{3}{\frac{1}{10} + \frac{1}{20} + \frac{1}{30}} = 16{,}6 \text{ km/h}$$

On vérifie aisément que tout autre calcul de moyenne conduirait à un résultat erroné.

## 7. Moyenne quadratique

Il arrive que les $(x_i)$ interviennent par l'ensemble de leurs carrés

$$\{x_1^2, x_2^2, x_3^2 \dots x_n^2\}$$

comme par exemple, en thermodynamique où l'on montre que la température absolue d'un gaz est liée aux carrés des vitesses des molécules de ce gaz.

On définit la moyenne quadratique des $(x_i)$ comme étant la racine carrée de la moyenne arithmétique des $(x_i^2)$, soit

$$q = \sqrt{\frac{1}{n} \sum_{i=1}^{n} x_i^2} \qquad (3.11)$$

On montre que les différentes moyennes vérifient les inégalités

$$h \leqslant g \leqslant \bar{x} \leqslant q$$

(les égalités ayant lieu lorsque tous les nombres $(x_i)$ sont identiques).

# II. LA MEDIANE

## 1. Définition

La médiane est la valeur du caractère $M_e$ telle qu'il y ait autant d'individus pour lesquels le caractère est inférieur à $M_e$ que d'individus pour lesquels le caractère est supérieur à $M_e$.

Pour déterminer la médiane, il est donc nécessaire de considérer les effectifs cumulés croissants ou décroissants et de chercher, le cas échéant par interpolation, la valeur du caractère correspondant à 50 % de l'effectif total.

2. Exemple

Dans le cas de la série groupée en classes du tableau 2.6 (repris en 3.3), on détermine d'abord, à partir des effectifs cumulés, la classe contenant la médiane, puis la valeur de la médiane par interpolation linéaire dans cette classe.

| Classes | Limites | Centres de classe | Effectifs | Effectifs cumulés croissants | Effectifs cumulés décroissants |
|---|---|---|---|---|---|
| | 1,545 | | | | 80 |
| 1,55-1,59 | | 1,57 | 3 | | |
| | 1,595 | | | 3 | 77 |
| 1,60-1,64 | | 1,62 | 12 | | |
| | 1,645 | | | 15 | 65 |
| 1,65-1,69 | | 1,67 | 18 | | |
| | 1,695 | | | 33 | 47 |
| 1,70-1,74 | | 1,72 | 25 | | |
| | 1,745 | | | 58 | 22 |
| 1,75-1,79 | | 1,77 | 15 | | |
| | 1,795 | | | 73 | 7 |
| 1,80-1,84 | | 1,82 | 5 | | |
| | 1,845 | | | 78 | 2 |
| 1,85-1,89 | | 1,87 | 2 | | |
| | 1,895 | | | 80 | |
| | | | 80 | | |

Tableau 3.3

La moitié de l'effectif total est 40. Sur les effectifs cumulés croissants, on voit que 33 personnes ont une taille inférieure ou égale à 1,695 m. L'interpolation linéaire dans la classe 1,695-1,745 dont l'effectif est de 25 personnes, donne pour la médiane

$$M_e = 1{,}695 + \frac{0{,}05 \times (40 - 33)}{25} = 1{,}695 + 0{,}014 = 1{,}709 \text{ m}$$

3. Détermination graphique de la médiane

a) à partir de l'histogramme

Etant donnée la signification de l'aire de l'histogramme, la médiane n'est autre que la valeur du caractère qui coupe l'histogramme en deux parties de surfaces égales.

. *Exemple de l'histogramme de la figure 2.4* (reprise dans la fig. 3.1).

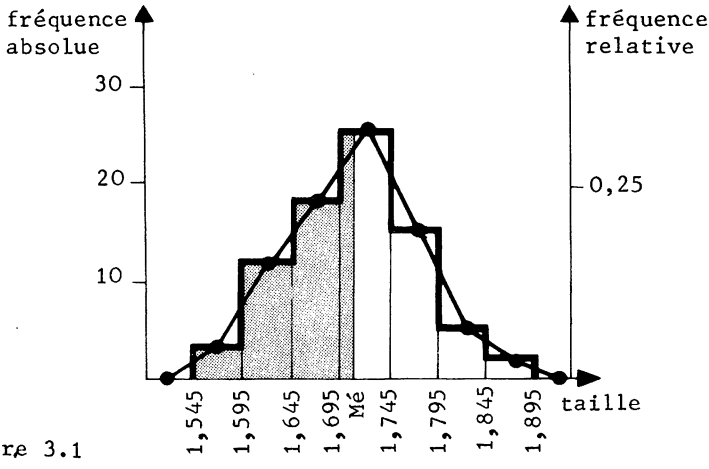



Figure 3.1

b) à partir des polygones des effectifs cumulés

La médiane est représentée par la valeur du caractère correspondant à l'intersection

- soit de la courbe des effectifs cumulés croissants et de la courbe des effectifs cumulés décroissants
- soit de l'une des deux courbes précédentes avec l'horizontale représentant l'effectif moitié.

. *Exemple de la figure 2.5* (reprise dans la fig. 3.2)

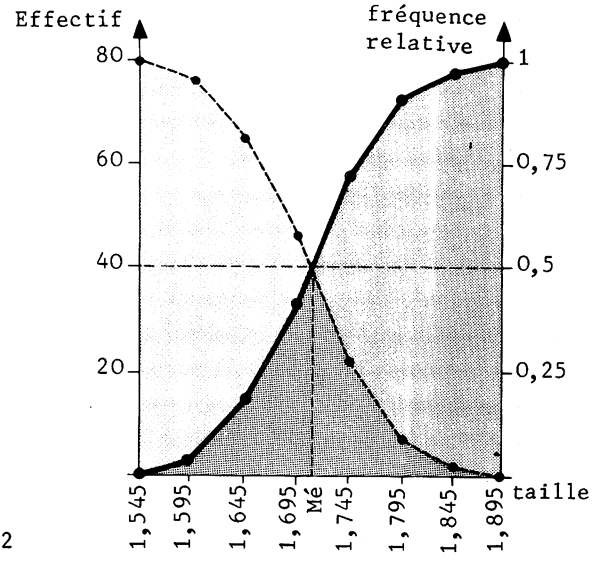



Figure 3.2

## III. LES PERCENTILES

Le kième percentile est la valeur du caractère $C_k$

- telle que l'ensemble des individus dont le caractère est au plus égal à $C_K$ représente les k % de l'effectif total

- telle que l'ensemble des individus dont le caractère est au moins égal à $C_K$ représente les (100 - k) % de l'effectif total

*Parmi les percentiles*, on distingue

les déciles, pour lesquels k = 10, 20, 30, ...

$C_{10} = D_1$ $C_{20} = D_2$ ...

les quartiles, pour lesquels k = 25, 50, 75

$C_{25} = Q_1$ $C_{50} = Q_2$ $C_{75} = Q_3$

la médiane, pour laquelle k = 50, $C_{50} = Me = Q_2 = D_5$

Le calcul des différents percentiles est tout à fait analogue à celui de la médiane.

. *Exemple*

Calcul du 3ème quartile $Q_3$ pour la distribution du tableau 3.3.

Les 75 % de l'effectif total correspondent à 60 personnes. Sur les effectifs cumulés croissants, on voit que 58 personnes ont une taille inférieure ou égale à 1,745 m. Pour les 2 personnes qui manquent, on fait une interpolation linéaire dans la classe 1,745-1,795 dont l'effectif est de 15 personnes. On obtient donc :

$$Q_3 = 1{,}745 + \frac{0{,}05 \times 2}{15} \simeq 1{,}745 + 0{,}006 \simeq 1{,}751 \text{ m}$$

## IV. MODE OU DOMINANTE

C'est la valeur du caractère correspondant à la fréquence maximum.

Une distribution peut présenter plusieurs modes : on dit qu'elle est plurimodale.

. *Exemples*

1. Diagramme en bâtons de la figure 2.1 :
   Le mode est visiblement D = 1 enfant.
2. Histogramme de la figure 2.4 :
   On adopte pour mode le centre de classe de la classe modale D = 1,72 m.

## V. COMPARAISON DES DIFFERENTS PARAMETRES DE POSITION

La moyenne arithmétique est peu sensible aux fluctuations d'échantillonnage. Elle se prête bien aux comparaisons. Des valeurs aberrantes peuvent toutefois la modifier sensiblement.

La médiane est plus sensible aux fluctuations d'échantillonage, elle l'est moins à des valeurs aberrantes. Toutefois, elle se prête moins bien à des calculs algébriques.

Le mode est représentatif de la valeur du caractère le plus courant, le plus typique, mais il peut présenter une certaine ambiguité.

En pratique, il est fréquent que parmi ces trois paramètres, le choix de l'un ne s'impose pas plus que le choix de l'autre. La comparaison des trois permet de se faire une idée plus complète de la distribution. L'interprétation des positions relatives de ces paramètres est parfois plus immédiate que celle des paramètres de dispersion qui seront l'objet de la fin de ce chapitre.

## B. PARAMETRES DE DISPERSION

Les paramètres de position sont insuffisants pour caractériser complètement une série. Par exemple, deux séries différentes ayant la même moyenne, ne se répartissent pas nécessairement de la même manière autour de cette moyenne. Elles sont plus ou moins étalées, ce qui sera décrit par les caractéristiques de dispersion.

Un paramètre de dispersion se rapporte à la différence de deux valeurs du caractère alors qu'un paramètre de position représente une valeur du caractère. On distingue les notions suivantes :

déviation : différence algébrique de deux valeurs du caractère, par exemple $x_i - \overline{X}$ ;

écart : valeur absolue de la différence de deux valeurs du caractère, par exemple $|x_i - \overline{X}|$ ;

intervalle de variation, étendue (ou range) : différence entre les valeurs extrêmes du caractère ;

écart interquartile : différence entre le 3ème et le 1er quartiles, c'est à dire $Q_3 - Q_1$ (voir paragraphe précédent pour la signification des quartiles).

Rappel : la somme des déviations d'un ensemble de données $x_i$ par rapport à leur valeur moyenne $\overline{X}$, est nulle. En effet, d'après l'équation (3.8)

$$\sum_i n_i (x_i - \overline{X}) = \sum_i n_i x_i - n \overline{X} = 0$$

## I. ECART MOYEN ARITHMETIQUE

### . Définition

L'écart moyen arithmétique est la moyenne arithmétique des écarts par rapport à la moyenne arithmétique des valeurs du caractère

$$\overline{E} = \frac{1}{n} \sum_i n_i \, |x_i - \overline{X}| \qquad (3.12)$$

### . *Exemple*

A partir du tableau 3.3, en prenant $\overline{X} = 1{,}707$ m, on obtient le tableau suivant :

| Centres de classes $x_i$ | $n_i$ | $\lvert x_i - \overline{X}\rvert$ | $n_i \lvert x_i - \overline{X}\rvert$ |
|---|---|---|---|
| 1,57 | 3 | 0,137 | 0,411 |
| 1,62 | 12 | 0,087 | 1,044 |
| 1,67 | 18 | 0,037 | 0,666 |
| 1,72 | 25 | 0,013 | 0,325 |
| 1,77 | 15 | 0,063 | 0,945 |
| 1,82 | 5 | 0,113 | 0,565 |
| 1,87 | 2 | 0,163 | 0,326 |
| | 80 | Total | 4,282 |

Tableau 3.4

$$\overline{E} = \frac{1}{n} \sum n_i \, |x_i - \overline{X}| = \frac{1}{80} \times 4{,}282 = 0{,}053 \text{ m}$$

## II. VARIANCE. ECART-TYPE

### 1. Variance

La variance d'une série de valeurs du caractère est la

moyenne arithmétique des carrés des écarts de ces valeurs par rapport à leur moyenne arithmétique.

$$V = \frac{1}{n} \sum n_i \, (x_i - \overline{X})^2 \qquad (3.13)$$

Les carrés des différences évitent l'utilisation de valeurs absolues. Les dimensions de V sont celles du caractère au carré. Il faut donc en extraire la racine carrée pour obtenir un paramètre caractéristique des écarts.

2. Ecart-type

L'écart-type (ou écart quadratique moyen) est la racine carrée de la variance

$$\sigma = \sqrt{V} \qquad (3.14)$$

C'est le plus significatif de tous les paramètres de dispersion.

. *Exemple*

A partir du tableau 3.4, avec $\overline{X}$ = 1,707 m on obtient le tableau suivant :

| Centres de classes $x_i$ | $n_i$ | $(x_i - \overline{X})^2$ | $n_i \, (x_i - \overline{X})^2$ |
|---|---|---|---|
| 1,57 | 3 | 0,01877 | 0,05631 |
| 1,62 | 12 | 0,00757 | 0,09084 |
| 1,67 | 18 | 0,00137 | 0,02466 |
| 1,72 | 25 | 0,00017 | 0,00425 |
| 1,77 | 15 | 0,00397 | 0,05955 |
| 1,82 | 5 | 0,01277 | 0,06385 |
| 1,87 | 2 | 0,02657 | 0,05314 |
| | 80 | Total | 0,35260 |

Tableau 3.5

$$V = \frac{1}{n} \sum n_i (x_i - \bar{X})^2 = \frac{1}{80} \times 0{,}35260 = 0{,}0044$$

$$\sigma = \sqrt{V} = 0{,}066 \text{ m}$$

3. Méthodes rapides de calcul de l'écart-type

a) Autre expression de la variance :

$$V = \overline{X^2} - \bar{X}^2 \qquad (3.15)$$

En effet, la sommation sur les carrés des écarts à la moyenne s'écrit :

$$\sum_{i=1}^{n} (x_i - \bar{X})^2 = \sum x_i^2 + \sum \bar{X}^2 - 2 \sum x_i \bar{X}$$

$$= \sum x_i^2 + n \bar{X}^2 - 2 \bar{X} n \bar{X}$$

$$= \sum x_i^2 - n \bar{X}^2$$

en remplaçant $\sum x_i$ par $n \bar{X}$. Par conséquent, on a pour la variance

$$V = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{X})^2 = \frac{1}{n} \sum x_i^2 - \bar{X}^2 = \overline{X^2} - \bar{X}^2$$

et $$\sigma = \sqrt{\overline{X^2} - \bar{X}^2} \qquad (3.16)$$

b) Changement de variable

Comme pour la moyenne, on peut utiliser suivant les cas, un changement d'origine (équ. 3.4)

$$u_i = x_i - x_o$$

un changement d'échelle, ou les deux à la fois (équ. 3.6)

$$z_i = \frac{x_i - x_o}{k}$$

où k est l'intervalle de classe. On obtient dans le 2ème cas

$$x = k z + x_o$$

$$x^2 = k^2 z^2 + x_o^2 + 2 k x_o z$$

$$\overline{X^2} = k^2\,\overline{z^2} + x_o^2 + 2\,k\,x_o\,\overline{z}$$

$$\overline{X} = k\,\overline{z} + x_o$$

$$\overline{X}^2 = k^2\,\overline{z}^2 + x_o^2 + 2\,k\,x_o\,\overline{z}$$

et pour la variance

$$V = \overline{X^2} - \overline{X}^2 = k^2\,(\overline{z^2} - \overline{z}^2) \qquad (3.17)$$

Le premier cas, du changement d'origine seulement, se déduit du deuxième en faisant k = 1.

## III. MOMENTS D'UNE SERIE STATISTIQUE

### Définition

On appelle moment d'ordre q par rapport à $x_o$, la moyenne arithmétique des puissances qièmes des déviations des valeurs du caractère par rapport à $x_o$ :

$$m_q = \frac{1}{n} \sum n_i\,(x_i - x_o)^q \qquad (3.18)$$

Cas particuliers :

a) $x_o = 0$, $q = 1$, le moment n'est autre que la moyenne arithmétique.

b) $x_o = \overline{X}$, $q = 2$, le moment n'est autre que la variance.

### Remarque

Tout comme les paramètres de dispersion permettent de représenter la distribution statistique plus fidèlement que les seuls paramètres de position, les moments d'ordre q supérieurs à 2 améliorent encore cette représentation. Toutefois, on se limite généralement à q = 1 et q = 2.

I. Les résultats d'un certain processus aléatoire sont des nombres entiers n que l'on a classés suivant l'histogramme ci-contre (fig. 1)

1°) Calculer la valeur moyenne. Quel est le mode ? Quelle est la médiane ?

2°) Tracer le polygone des fréquences et le polygone des fréquences relatives cumulées croissantes. Retrouver la valeur de la médiane.

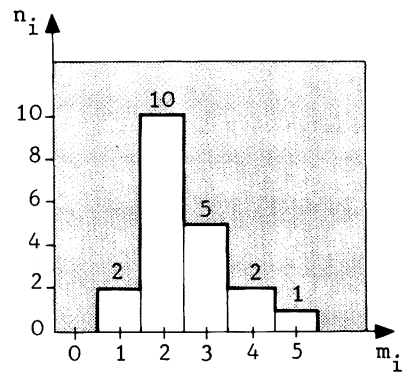



Figure 1

SOLUTION

1°) La moyenne $\bar{m}$ est donnée par la relation

$$\bar{m} = \frac{\sum_i n_i m_i}{\sum_i n_i} \quad \text{d'où}$$

$$\bar{m} = \frac{2\times1+10\times2+5\times3+2\times4+1\times5}{2+10+5+2+1} = \frac{50}{20} = 2,5$$

- Le mode est 2, c'est la valeur de $m_i$ où $n_i$ est maximum.

- La médiane est la valeur de $m_i$ qui partage l'aire S de l'histogramme en deux parties égales (fig. 2, courbe 1). On trouve Me = 2,3

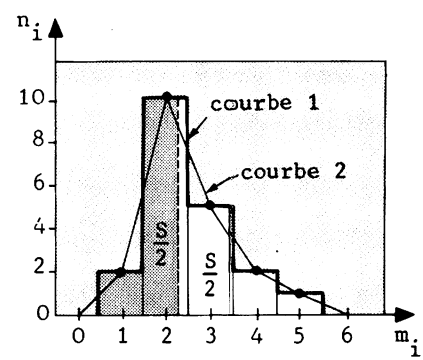



Figure 2

2°) Le polygone des fréquences est représenté sur la figure 2, courbe 2.

Le polygone des fréquences relatives cumulées croissantes est donné par la figure 3.

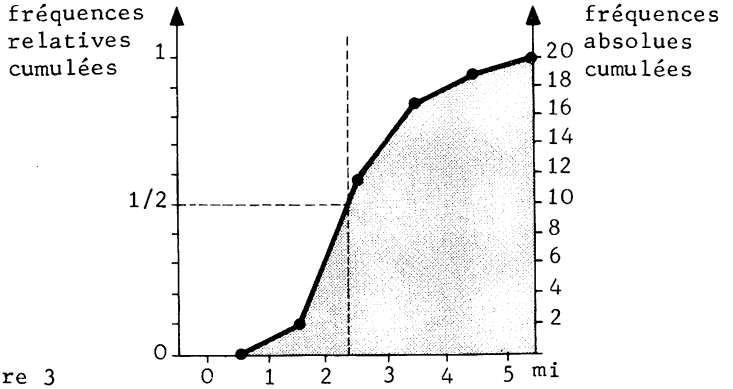



Figure 3

La valeur de la médiane est directement déterminée sur le diagramme des fréquences relatives cumulées, où elle correspond à la valeur du caractère ayant pour ordonnée $\frac{1}{2}$. ■

---

II. Les pesées de 50 nouveaux-nés dans une maternité ont permis d'établir le tableau suivant :

| Classes en kg | Centre de classe | Fréquences absolues | Fréquences absolues cumulées croissantes |
|---|---|---|---|
| 2,0 - 2,4 | 2,2 | 6 | 6 |
| 2,4 - 2,8 | 2,6 | 10 | 16 |
| 2,8 - 3,2 | 3,0 | 20 | 36 |
| 3,2 - 3,6 | 3,4 | 10 | 46 |
| 3,6 - 4,0 | 3,8 | 4 | 50 |

On suppose que dans chaque classe, les poids sont répartis uniformément entre les nouveaux-nés correspondants.

1°) Calculer les quartiles $Q_1$, $Q_2$ et $Q_3$.

2°) Retrouver ces résultats à partir du diagramme des fréquences absolues cumulées croissantes, où l'on fera figurer les fréquences relatives correspondant à $Q_1$, $Q_2$ et $Q_3$.

3°) Déterminer le 4ème centile.

---

SOLUTION

1°) $Q_1$ correspond à la classe qui contient le quart de l'effectif total, soit $\frac{50}{4} = 12,5 = 6 + 6,5$

Donc $Q_1$ appartient à la classe [2,4 - 2,8]. On aura :

$$Q_1 = 2,4 + \frac{0,4 \times 6,5}{10} = 2,66 \text{ kg}$$

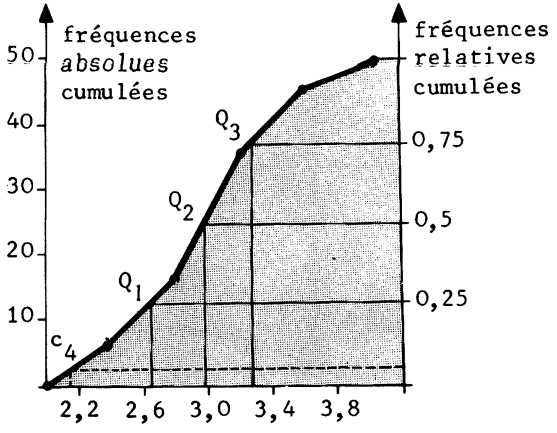
Le 2ème quartile (ou médiane) $Q_2$ correspond à l'effectif cumulé croissant $\frac{50}{2} = 25 = 16 + 9$. Donc $Q_2$ appartient à la classe [2,8 - 3,2]. D'où $Me = Q_2 = 2,8 + \frac{0,4 \times 9}{20} = 2,98$ kg.

De la même manière

$$Q_3 = 3,2 + \frac{0,4 \times 1,5}{10} = 3,26 \text{ kg}$$

2°) Diagramme des fréquences absolues croissantes.

On remarque que, par rapport aux fréquences relatives cumulées croissantes, $Q_1$, $Q_2$, $Q_3$ correspondront toujours aux valeurs des abscisses dont les ordonnées sont respectivement 0,25, 0,5 et 0,75.



3°) Le 4ème centile $C_4$ correspond à la classe qui contient l'effectif cumulé $\frac{50}{100} \times 4 = 2$, c'est à dire à la première classe [ 2,0 - 2,4] . On en déduit

$$C_4 = 2{,}0 + \frac{0{,}4 \times 2}{6} = 2{,}13 \text{ kg.}$$ ■

## EXERCICES CHAPITRE 3 B. PARAMETRES DE DISPERSION

I. Sur 1 000 électeurs, on observe :

401 électeurs dont l'âge est compris entre 20 et 40 ans

368 " " " " " " 40 et 60 ans

231 " " " " " " 60 et 80 ans

Déterminer la moyenne et l'écart-type de cette série.

SOLUTION

| Age | Centre $x_i$ | $n_i$ | $n_i x_i$ | $n_i x_i^2$ |
|---|---|---|---|---|
| [20 - 40] | 30 | 401 | 12030 | 360900 |
| [40 - 60] | 50 | 368 | 18400 | 920000 |
| [60 - 80] | 70 | 231 | 16170 | 1131900 |
| | | $\sum_i n_i$ = 1000 | $\sum_i n_i x_i$ = 46600 | $\sum_i n_i x_i^2$ = 2412800 |

D'où $\overline{X} = \frac{\sum_i n_i x_i}{\sum_i n_i} = \frac{46600}{1000} = 46{,}6$ ans

Avant de calculer l'écart-type $\sigma$, calculons la variance V

$$V = \overline{x^2} - \overline{x}^2 \quad \text{avec}$$

$$\overline{x^2} = \frac{\sum_i n_i x_i^2}{\sum_i n_i} = \frac{2412800}{1000} = 2412,8$$

et $\overline{X}^2 = 2171,56$ d'où

$$V = 2412,8 - 2171,56 = 241,24$$

On en déduit

$$\sigma = \sqrt{V} = 15,53 \qquad \blacksquare$$

---

II. Soit une grandeur X telle que $X = 2\,X_1 + 3\,X_2$.

Les résultats des mesures sur $X_1$ et $X_2$ ont donné respectivement

$X_1 = 5,\ 10,\ 15$ et $X_2 = 3,\ 5,\ 7$

1°) Calculer la valeur de la moyenne arithmétique de $X_1$, $X_2$ puis de X.

2°) Calculer les variances $V_1 = \sigma_1^2$, $V_2 = \sigma_2^2$ et $V = \sigma^2$.

---

SOLUTION

1°) $\overline{X}_1 = \frac{5 + 10 + 15}{3} = 10 \qquad \overline{X}_2 = \frac{3 + 5 + 7}{3} = 5$

d'où $\overline{X} = 2\,\overline{X}_1 + 3\,\overline{X}_2 = 20 + 15 = 35$

Vérification : les différentes valeurs de X sont X = 19, 35, 51

d'où $\overline{X} = \frac{19 + 35 + 51}{3} = 35$

2°) a) $V_1 = \sigma_1^2 = \overline{X_1^2} - \overline{X}_1^2$ avec $\overline{X_1^2} = \frac{25 + 100 + 225}{3} = \frac{350}{3}$

$$\sigma_1^2 = \frac{350}{3} - 10^2 = \frac{50}{3}$$

b) $V_2 = \sigma_2^2 = \overline{X_2^2} - \overline{X}_2^2$ avec $\overline{X_2^2} = \frac{9 + 25 + 49}{3} = \frac{83}{3}$

$$\sigma_2^2 = \frac{83}{3} - 5^2 = \frac{8}{3}$$

c) $V = \sigma^2 = \overline{X^2} - \overline{X}^2 = \overline{(2\,X_1 + 3\,X_2)^2} - \overline{(2\,X_1 + 3\,X_2)}^2$

$= 4\,\overline{X_1}^2 + 9\,\overline{X_2}^2 + 12\,\overline{X_1 X_2} - 4\,\overline{X_1}^2 - 9\,\overline{X_2}^2 - 12\,\overline{X}_1 . \overline{X}_2$

$\sigma^2 = 4\,\sigma_1^2 + 9\,\sigma_2^2 + 12\,(\overline{X_1 X_2} - \overline{X}_1 . \overline{X}_2)$

$\overline{X_1 X_2} = \dfrac{5 \times 3 + 10 \times 5 + 15 \times 7}{3} = \dfrac{170}{3}$

$\overline{X_1 X_2} - \overline{X}_1 . \overline{X}_2 = \dfrac{170}{3} - 5 \times 10 = \dfrac{20}{3}$

$V = \sigma^2 = 4 \times \dfrac{50}{3} + 9 \times \dfrac{8}{3} + 12 \times \dfrac{20}{3} = \dfrac{512}{3}$ ■

---

III. On a relevé les nombres d'allumettes contenues respectivement dans 20 boîtes, lors d'un contrôle dans une usine de fabrication. Les résultats sont les suivants : 40, 42, 32, 38, 40, 48, 30, 38, 36, 40, 34, 40, 34, 40, 38, 40, 42, 44, 36, 42.

1°) Ranger ces résultats en classes d'intervalle 4 allumettes, borne supérieure exclue.

2°) Tracer l'histogramme de cette distribution.

3°) Calculer la moyenne et l'écart-type de cette série.

4°) Calculer les moments d'ordre 1, d'ordre 2 et d'ordre 3 par rapport à la valeur moyenne $\overline{x} = 39{,}6$.

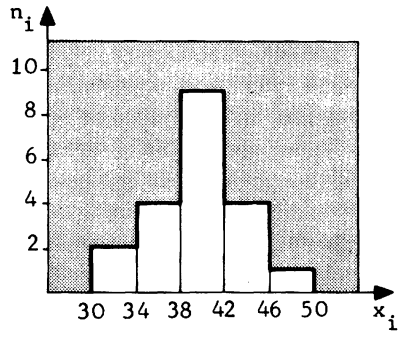
---

SOLUTION

1°)

| Classes $x_i$ | Nombre $n_i$ de boîtes correspondant |
|---|---|
| [ 30 - 34 [ | 2 |
| [ 34 - 38 [ | 4 |
| [ 38 - 42 [ | 9 |
| [ 42 - 46 [ | 4 |
| [ 46 - 50 [ | 1 |

2°) Histogramme de la distribution




3°) Calcul de la moyenne et de l'écart-type.

Choisissons une variable provisoire $t_i = \dfrac{x_i - 40}{4}$

| Classes | Centres de classes $x_i$ | Effectif $n_i$ | variable provisoire $t_i = \dfrac{x_i - 40}{4}$ | $n_i\ t_i$ | $n_i\ t_i^2$ |
|---|---|---|---|---|---|
| 30-34 | 32 | 2 | - 2 | - 4 | + 8 |
| 34-38 | 36 | 4 | - 1 | - 4 | + 4 |
| 38-42 | 40 | 9 | 0 | 0 | 0 |
| 42-46 | 44 | 4 | + 1 | + 4 | + 4 |
| 46-50 | 48 | 1 | + 2 | + 2 | + 4 |
| | | $\sum_i n_i = 20$ | | $\sum_i n_i\ t_i = -2$ | $\sum_i n_i\ t_i^2 = 20$ |

On calcule alors simplement :

$$\bar{t} = \frac{\sum_i n_i\, t_i}{\sum_i n_i} = -\frac{2}{20} = -0{,}1 \text{ et}$$

$$\overline{t^2} = \frac{\sum_i n_i\, t_i^2}{\sum_i n_i} = \frac{20}{20} = +1$$

d'où $V_t = \overline{t^2} - \bar{t}^2 = 0{,}99 = \sigma_t^2$

De l'expression de $t_i$, on déduit $x_i = 4\, t_i + 40$

d'où $\bar{x} = 4\, \bar{t} + 40 = -0{,}4 + 40$

$\bar{x} = 39{,}6$ allumettes.

D'après la relation (3.17) on doit avoir

$$\sigma_x^2 = k^2\, (\overline{t^2} - \bar{t}^2) = k^2\, \sigma_t^2 = 16 \times 0{,}99 = 15{,}84$$

d'où $\sigma_x = \sqrt{15{,}84} = 3{,}98$

4°) D'après l'expression (3.18), le moment d'ordre q par rapport à $x_o$ est égal à

$$m_q = \frac{1}{N} \sum_i n_i\, (x_i - x_o)^q$$

| $n_i$ | $x_i-39{,}6$ | $n_i(x_i-39{,}6)$ | $(x_i-39{,}6)^2$ | $n_i(x_i-39{,}6)^2$ | $(x_i-39{,}6)^3$ | $n_i(x_i-39{,}6)^3$ |
|---|---|---|---|---|---|---|
| 2 | - 7,6 | - 15,2 | 57,76 | 115,52 | - 438,976 | - 877,952 |
| 4 | - 3,6 | - 14,4 | 12,96 | 51,84 | - 46,656 | - 186,624 |
| 9 | + 0,4 | + 3,6 | 0,16 | 1,44 | + 0,064 | + 0,576 |
| 4 | + 4,4 | + 17,6 | 19,36 | 77,44 | + 85,184 | + 340,736 |
| 1 | + 8,4 | + 8,4 | 70,56 | 70,56 | + 592,704 | + 592,704 |
| 20 | | 0 | | 316,80 | | - 130,56 |

a) $m_1 = \frac{1}{N} \sum_i n_i (x_i - 39,6) = 0$

Ceci est normal car $\frac{1}{N} \Sigma n_i x_i = \bar{x}$ et donc $\frac{1}{N} \Sigma n_i x_i = \frac{1}{N} \Sigma n_i \bar{x}$

b) $m_2 = \frac{1}{N} \sum_i n_i (x_i - 39,6)^2 = \frac{316,80}{20} = 15,84$

Ce n'est autre que la variance $\sigma^2$.

c) $m_3 = \frac{1}{N} \sum_i n_i (x_i - 39,6)^3 = - 6,528.$ ■

---

IV. On reprend le tableau 3.2 du cours, soit

| Classes | Centres de classes | Effectif |
|---|---|---|
| 1,55-1,59 | 1,57 | 3 |
| 1,60-1,64 | 1,62 | 12 |
| 1,65-1,69 | 1,67 | 18 |
| 1,70-1,74 | 1,72 | 25 |
| 1,75-1,79 | 1,77 | 15 |
| 1,80-1,84 | 1,82 | 5 |
| 1,85-1,89 | 1,87 | 2 |

En effectuant le changement de variable $z = \frac{x - 1,7}{0,05}$, calculer $\bar{z}$, V (z), $\sigma$ (z) ; en déduire $\bar{x}$, V (x) et $\sigma$ (x)

---

SOLUTION

En faisant $x_o = 1,7$ et $k = 0,05$ qui constitue l'intervalle de classe, on obtient le tableau de la page suivante.

$$\bar{z} = \frac{1}{80} \times 12 = 0,15$$

$$\overline{z^2} = \frac{1}{80} \times 142,8 = 1,785$$

$$V(z) = k^2 (\overline{z^2} - \bar{z}^2) = 0,05^2 [1,785 - (0,15)^2]$$

$$= 0,0044$$

$$\sigma(z) = \sqrt{V(z)} = 0,066 \text{ m}$$

| Centres de classes $x_i$ | $n_i$ | $z_i$ | $n_i z_i$ | $z_i^2$ | $n_i z_i^2$ |
|---|---|---|---|---|---|
| 1,57 | 3 | - 2,6 | - 7,8 | 6,76 | 20,28 |
| 1,62 | 12 | - 1,6 | - 19,2 | 2,56 | 30,72 |
| 1,67 | 18 | - 0,6 | - 10,8 | 0,36 | 6,48 |
| 1,72 | 25 | 0,4 | 10 | 0,16 | 4,00 |
| 1,77 | 15 | 1,4 | 21 | 1,96 | 29,40 |
| 1,82 | 5 | 2,4 | 12 | 5,76 | 28,80 |
| 1,87 | 2 | 3,4 | 6,8 | 11,56 | 23,12 |
| | 80 | | 12 | | 142,80 |

Si $z = \frac{x - 1,7}{0,05}$ alors $x = 0,05\ z + 1,7$

Par suite $\bar{x} = 0,05\ \bar{z} + 1,7 = 0,05 \times 0,15 + 1,7 \simeq 1,708$

et $V(x) = (0,05)^2\ V(z) = 0,11 \cdot 10^{-4}$

d'où $\sigma(x) = 0,33 \cdot 10^{-2}$

# 4. Analyse combinatoire

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

Principe général

Si une expérience complexe résulte de la réalisation dans un certain ordre, d'une première expérience simple pouvant conduire à $n_1$ résultats différents, suivie d'une deuxième expérience simple pouvant conduire à $n_2$ résultats différents, puis d'une troisième expérience et ainsi de suite, le nombre de résultats distincts possibles de l'expérience globale est égal à

$$n = n_1 \times n_2 \times n_3 \times \ldots \qquad (4.1)$$

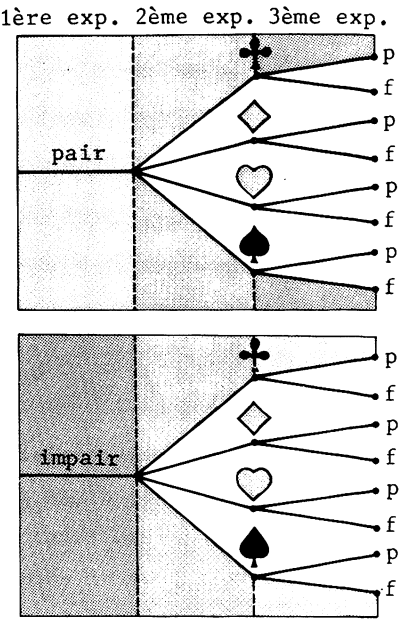
Un moyen pratique pour illustrer cette formule et dénombrer les résultats possibles d'une suite d'expériences consiste à utiliser un diagramme en arbre.

. *Exemple*

On réalise dans l'ordre, les 3 expériences suivantes :

- on lance un dé (résultats possibles : nombre pair, nombre impair)
- on tire au hasard une couleur d'un jeu de cartes (résultats possibles : trèfle, carreau, coeur, pique)
- on lance une pièce (résultats possibles : pile, face).

Dénombrer tous les résultats distincts de l'expérience globale.




Ce diagramme montre que $n_1 = 2$, $n_2 = 4$, $n_3 = 2$ et par conséquent

$$n = 2 \times 4 \times 2 = 16$$

. *Autre exemple*

Un système d'immatriculation comprend 4 chiffres dont le 1er est différent de 0, suivis de 2 lettres distinctes et différentes de I et O. Déterminer le nombre de plaques d'immatriculation possibles.

En assignant une case à chaque chiffre ou lettre, on voit qu'on peut attribuer 9 chiffres différents à la 1ère case, 10 aux 3 cases suivantes, 24 lettres différentes à la 5ème case et 23 seulement à la dernière case, puisque les deux lettres doivent être distinctes. Le nombre de plaques différentes est donc :

$$n = 9 \times 10 \times 10 \times 10 \times 24 \times 23 = 4\ 968\ 000$$

# A. ARRANGEMENTS

## I. DEFINITION

On appelle arrangement de n éléments p à p ($p \leqslant n$), tout ensemble ordonné de p de ces éléments, tous distincts. Un arrangement est donc caractérisé par la nature des éléments ou par leur ordre.

. *Exemple*

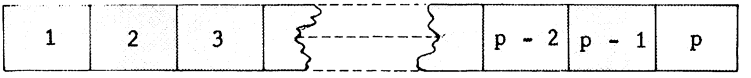
Ensemble de 4 lettres a, b, c, d.

Les groupements abc, abd, bac, ... constituent des arrangements de ces 4 lettres 3 à 3, les groupements ab, ad, ba, ... constituent des arrangements de ces 4 lettres 2 à 2.

★ Calcul de $A_n^p$

On désigne par $A_n^p$ le nombre total d'arrangements distincts de n éléments p à p.

Tout arrangement de p objets peut être construit de la manière suivante : on considère p cases, numérotées de 1 à p ($p \leqslant n$)




Dans la 1ère case, on place un objet, ce qui donne n choix possibles.

Dans la 2ème case, on place un autre objet choisi parmi les (n-1) objets restants, cela donne (n-1) choix possibles.

De la même manière, on obtient (n-2) choix possibles

pour la 3ème case, et ainsi de suite, jusqu'à la pième case pour laquelle il ne reste plus que (n-p+1) choix possibles.

En appliquant le principe général (4.1), on a

$$A_n^p = n\,(n-1)(n-2)\,\ldots\,(n-p+2)(n-p+1) \qquad (4.2)$$

. *Exemple*

Nombre de tiercés dans l'ordre dans une course de 10 chevaux.

$$A_{10}^3 = 10 \cdot 9 \cdot 8 = 720$$

. Notation factorielle

$$n\,! = 1.\ 2.\ 3\ \ldots\ (n-2)(n-1)\ n \qquad (4.3)$$

En particulier $1! = 1$ $\quad 2! = 2$

$3! = 6$ $\quad 4! = 24$, etc.

En appliquant cette notation factorielle à l'expression (4.2) de $A_n^p$, on trouve

$$A_n^p = \frac{n\,(n-1)(n-2)\,\ldots\,(n-p+1) \times (n-p)!}{(n-p)!} = \frac{n!}{(n-p)!} \qquad (4.4)$$

. *Exemple*

$$A_{10}^3 = \frac{10!}{7!} = 10 \cdot 9 \cdot 8 = 720$$

Si n = p, la formule (4.4) ne peut s'appliquer, car on n'a pas défini 0!. Cependant, si n = p, il est clair que $A_n^n = n!$ *d'après le principe général de l'analyse combinatoire.* On pose donc, comme axiome de définition 0! = 1 afin que la relation (4.4) reste valable dans le cas où n = p.

## II. ARRANGEMENTS AVEC REPETITION

Un arrangement de n objets p à p avec répétition est un arrangement où chaque objet peut être répété jusqu'à p fois. Le raisonnement précédent montre que pour chaque case, on

dispose alors de n choix possibles. Le nombre total de tels arrangements est donc

$$\alpha_n^p = n^p \tag{4.5}$$

. *Exemples*

1. Arrangements d'ordre 2 des 3 lettres a, b, c

$\alpha_3^2 = 3^2 = 9$

2. Arrangements d'ordre 3 des 2 lettres a, b

$\alpha_2^3 = 2^3 = 8$

## B. PERMUTATIONS

### I. DEFINITION

Une permutation de n objets est un ensemble ordonné de ces n objets. Les permutations de n objets constituent un cas particulier des arrangements : c'est le cas où n = p. Deux permutations distinctes ne diffèrent donc que par l'ordre des objets.

. *Exemple*

Les permutations possibles des 3 lettres a, b, c sont : abc, bca, cab, bac, acb, cba.

★ Calcul de $P_n$

Le nombre total de permutations $P_n$ se déduit de l'expression du nombre total d'arrangements $A_n^p$, en faisant p = n et en utilisant la convention 0! = 1, soit

$$P_n = A_n^n = n! \qquad (4.6)$$

L'exemple précédent des 3 lettres a, b, c donne

$$P_3 = A_3^3 = 3! = 6$$

. *Autre exemple*

Nombre des configurations possibles à l'arrivée d'une course de 8 chevaux

$$P_8 = 8! = 40\ 320$$

## II. PERMUTATIONS AVEC REPETITION

Il arrive que, parmi les n objets dont on cherche le nombre de permutations, certains d'entre eux, au nombre de r par exemple, soient tous semblables. Auquel cas, rien ne distingue les permutations de ces r objets entre eux.

Pour calculer le nombre de permutations possibles, il faut donc diviser le nombre de permutations des n objets sans répétition, par le nombre de permutations des r objets entre eux, soit

$$P_n \text{ (avec répétition r)} = \frac{P_n}{P_r} = \frac{n!}{r!}$$

★ Généralisation à plusieurs répétitions

On considère n objets, parmi lesquels $r_1$ sont semblables entre eux, $r_2$ sont semblables entre eux, ..., $r_k$ sont semblables entre eux, avec $r_1 + r_2 + \ldots + r_k = n$. On appelle permutation de n objets avec répétitions $(r_1, r_2, \ldots, r_k)$ toute partition de ces n objets en k parties telles que la $i^{ème}$ partie ait $r_i$ éléments $(1 \leqslant i \leqslant k)$.

Le nombre de ces permutations des n objets avec répétitions $(r_1, r_2, \ldots, r_k)$ est :

$$P_n\,(r_1, r_2, \ldots, r_k) = \frac{n!}{r_1!\, r_2! \ldots r_k!} \qquad (4.7)$$

. *Exemple*

Nombre de permutations possibles avec les lettres du mot ETRENNE

$$P_n\,(r_E = 3,\ r_N = 2) = \frac{7!}{3!\ 2!} = 420$$

## III. PERMUTATION CIRCULAIRE

Le rangement de 4 objets sur une rangée fournit 4! = 24

permutations différentes, mais celui de 4 objets sur un cercle fournit seulement 3! = 6 permutations différentes.

*Généralisation*

n objets peuvent être disposés sur un cercle de (n-1)! façons différentes, soit le nombre de permutations $P_n$ divisé par le nombre de manières différentes n de choisir la 1ère place.

## C. COMBINAISONS

### I. DEFINITION

On appelle combinaison de p éléments pris parmi n ($n \geq p$), tout ensemble que l'on peut former en choisissant p de ces éléments, sans considération d'ordre. Deux combinaisons distinctes diffèrent donc par la nature d'au moins un élément.

. *Exemple*

Les combinaisons possibles des 4 lettres a, b, c, d 3 à 3 sont :

abc, abd, bcd, acd.

★ Calcul de $C_n^p$

On désigne par $C_n^p$ le nombre total de combinaisons de n objets p à p.

En remarquant que le nombre d'arrangements de n objets p à p n'est autre que le produit du nombre de combinaisons des n objets p à p, par le nombre de permutations des p éléments de chaque combinaison, soit :

$$A_n^p = C_n^p \times p! \tag{4.8}$$

on en déduit

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

$$= \frac{n!}{(n-p)!\,p!} \tag{4.9}$$

Remarques

1. On note aussi $C_n^p = \binom{n}{p}$

2. En utilisant l'expression (4.9), on peut démontrer les relations suivantes (cf. exercice 4 C I.)

$$C_n^p = C_n^{n-p}$$

$$C_n^p = C_{n-1}^p + C_{n-1}^{p-1} \qquad (4.10)$$

. *Exemples*

1. Nombre de tiercés dans le désordre dans une course de 10 chevaux

$$C_{10}^3 = \frac{10 \cdot 9 \cdot 8}{1 \cdot 2 \cdot 3} = 120$$

2. Nombre de mains différentes de 8 cartes dans un jeu de 32 cartes

$$C_{32}^8 = \frac{32.31.30.29.28.27.26.25}{1.\ 2.\ 3.\ 4.\ 5.\ 6.\ 7.\ 8} = 10\ 518\ 300$$

## II. PERMUTATIONS AVEC REPETITIONS ET COMBINAISONS

Une permutation de n objets avec répétition $r_1 = p$ et $r_2 = n-p$ (où $p \leqslant n$) est une partition de ces n objets en deux ensembles, l'un de p éléments, l'autre de n-p éléments. Se donner une telle permutation revient donc au même que se donner une partie de p éléments parmi n, c'est à dire une combinaison de n éléments pris p à p. On a donc :

$$C_n^p = P_n\ (p,\ n-p) \qquad (4.11)$$

Remarques

. Comme $P_n\ (p,\ n-p) = P_n\ (n-p,\ p)$ par définition, on en déduit que $C_n^p = C_n^{n-p}$.

. Une permutation de n objets à répétition $(r_1, r_2, \ldots, r_k)$

s'appelle aussi une combinaison généralisée. De même que l'on a $C_n^p = \binom{n}{p}$, on note $P_n(r_1, r_2, \ldots, r_k) = \binom{n}{r_1, r_2, \ldots, r_k}$.

## III. BINÔME DE NEWTON

Le binôme de Newton est le produit de n facteurs égaux à (a+b), soit $(a+b)^n$. Le développement de ce binôme est :

$$(a+b)^n = \sum_{p=0}^{n} C_n^p \, a^{n-p} \, b^p \qquad (4.12)$$

(cf. exercice 4 C VI.)

La relation (4.10)

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

permet une détermination pratique de proche en proche des différents coefficients $C_n^p$ au moyen du triangle de Pascal (fig. 4.1).

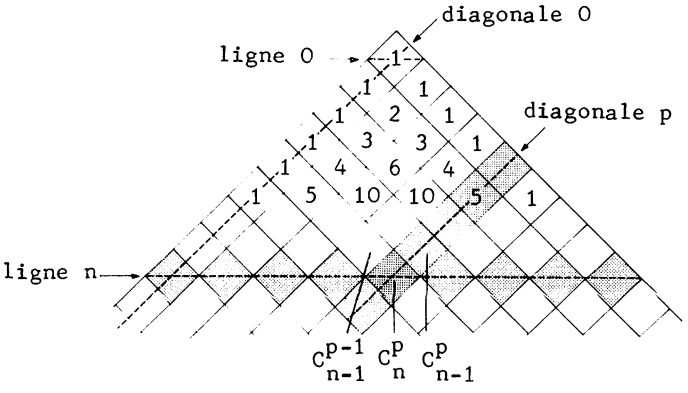



Figure 4.1 : Triangle de Pascal

Cette disposition symétrique du triangle de Pascal permet non seulement de calculer les coefficients du binôme, mais aussi de démontrer concrètement par récurrence des formules relatives à ces coefficients (cf. exercices 4 C V.).

Ces rappels trouveront leur utilisation ultérieurement, à l'occasion de la loi de probabilité dite loi binômiale.

## IV. COMBINAISONS AVEC REPETITION

Supposons que l'on étudie la répartition de n objets en fonction de r critères, et que l'on cherche le nombre de telles répartitions possibles.

Une telle répartition est appelée combinaison avec répétition d'ordre r.

Le nombre de ces combinaisons avec répétition est

$$\left[\begin{matrix} n \\ r \end{matrix}\right] = C^{r}_{n+r-1} \qquad (4.13)$$

En effet soit $x_1$, $x_2$, ..., $x_n$ les objets. Une répartition de ces objets suivant les critères peut être représentée ainsi :

$$x_1\ x_2\ x_3 / x_4 / x_5\ x_6 / . / x_7\ / \ldots / x_{n-1}\ x_n\ /$$

Le nombre de combinaisons avec répétition est donc égal au nombre de manières de séparer les $x_i$ par r frontières. C'est donc le nombre de manières de choisir r objets parmi n+r-1 sans tenir compte de l'ordre.

. *Exemple*

Lors d'un sondage dans une université, on pose à une centaine d'étudiants une question comportant 3 réponses possibles. Quel est le nombre de configurations différentes qu'on peut obtenir ?

Chaque configuration représente une combinaison de 100 réponses avec répétition d'ordre 3. Le nombre de ces combinaisons est donc, d'après (4.13)

$$\left[\begin{matrix} n \\ r \end{matrix}\right] = \left[\begin{matrix} 100 \\ 3 \end{matrix}\right] = C^{3}_{102} = \frac{102 \times 101 \times 100}{1 \times 2 \times 3} = 171\ 700$$

I. De combien de manières peut-on placer 3 dossiers différents dans 15 casiers vides, à raison d'un dossier par casier ?

---

SOLUTION

D'après le principe général (4.1), il y a 15 façons de placer le premier dossier. Celui-ci étant placé, il ne reste plus que 14 casiers vides ; il y a 14 façons de placer le deuxième dossier et enfin 13 façons de placer le troisième.

D'où $N = A_{15}^{3} = 15 \times 14 \times 13 = 2730$ manières différentes ■

---

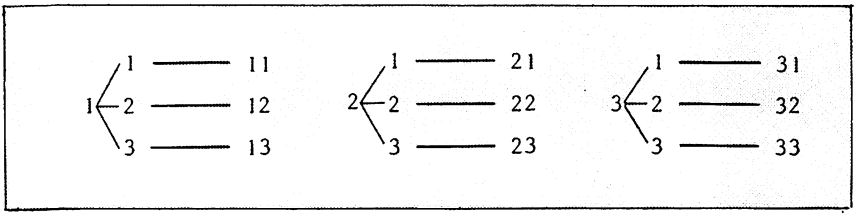
II. 1°) Ecrire tous les arrangements avec répétition d'ordre 2 des trois nombres 1, 2 et 3.

2°) Ecrire tous les arrangements avec répétition d'ordre 3 des deux nombres 4 et 5.

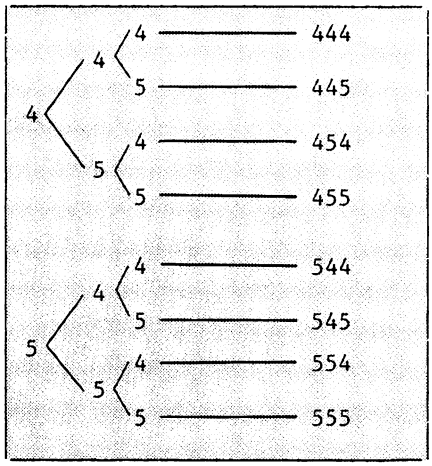
---

SOLUTION

1°) $\alpha_3^2 = 3^2 = 9$



2°) $\alpha_3^2 = 2^3 = 8$




■

III. On considère un jeu forain où 4 souris, numérotées de 1 à 4, se dirigent vers 5 cases A, B, C, D et E, plusieurs souris pouvant choisir la même case. Sur chaque billet, le joueur inscrit une répartition des souris dans les cases et il gagne lorsque son pronostic se réalise.

Combien de billets le joueur doit-il acheter pour être assuré de gagner ?

SOLUTION

Il s'agit d'un arrangement avec répétition de 5 objets pris 4 par 4. Il y a 5 possibilités qui s'offrent à la souris n° 1, de même pour les 3 autres.

Au total il y a $\alpha_5^4 = 5^4 = 625$ séquences possibles.

Le joueur doit donc acheter 625 billets. ■

I. A propos d'une course de chevaux, les rumeurs publiques accordent à 4 chevaux particuliers une chance égale de gagner. Quel est le nombre de quartés différents que l'on peut établir à partir de ces 4 chevaux.

---

SOLUTION

Le nombre de quartés distincts possibles est le nombre de permutations des 4 chevaux favoris

$P_4 = 4 \times 3 \times 2 \times 1 = 4! = 24$ ■

---

II. Afin de tester son sens chromatique, on présente à une personne une série de 5 plaques dont 2 d'une certaine couleur et 3 d'une couleur voisine. Combien de séries différentes peut-on lui présenter ?

---

SOLUTION

Il s'agit de déterminer le nombre de permutations des 5 plaques avec répétition des 2 plaques de la même couleur et des 3 plaques de la couleur voisine, soit

$P_5\ (r_1 = 2, r_2 = 3) = \frac{5!}{2!\ 3!} = 10$ séries. ■

---

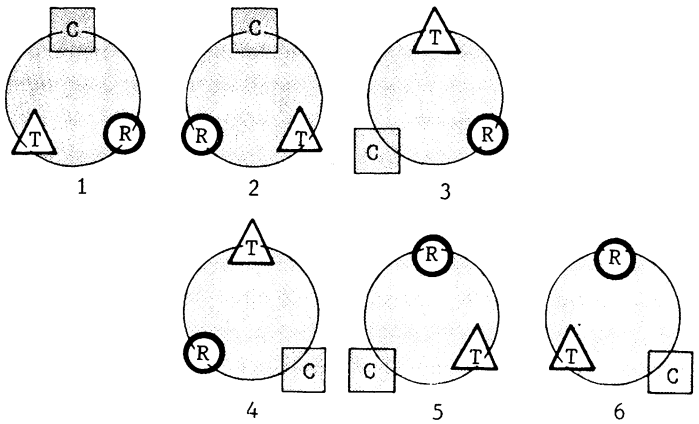
III. Pour réaliser un débat, on réunit trois personnes que l'on installe autour d'une table ronde. De combien de façons différentes pourra-t-on les placer les unes par rapport aux autres ?

---

SOLUTION

Dans ce cas, il ne convient pas de répondre que le nombre de façons différentes est $P_3 = 3! = 6$.

En effet, en désignant par C, T, R les invités du débat, on peut illustrer toutes les possibilités à l'aide des schémas suivants:




On peut alors remarquer que les cas 1 4 et 5 où T est à droite de C, et R à gauche de C, sont identiques. De même, les cas 2 3 et 6 , où R est à droite de C, et T à gauche de C, sont identiques.

On retrouve l'expression du nombre de permutations circulaires de 3 objets

$(n-1)! = 2! = 2$ cas distincts

I. Démontrer les relations suivantes ; en utilisant l'égalité : $C_n^p = \frac{n!}{p!\ (n-p)!}$

(1) $\forall n \in N^*,\ \forall p \in N^*,\ p < n,\ C_n^p = C_{n-1}^{p-1} + C_{n-1}^p$

(2) $\forall n \in N,\ C_n^n = 1$

(3) $\forall n \in N,\ \forall p \in N,\ p \leq n,\ C_n^p = C_n^{n-p}$

(4) Déduire de (2) et (3) que $\forall n \in N\ \ C_n^o = 1$

(5) Déduire de (1) et (3) que $\forall n \in N^*,\ C_{2n}^n = 2\,C_{2\,n-1}^n = 2\,C_{2\,n-1}^{n-1}$

---

SOLUTION

(1) Soient $n \in N^*$, $p \in N^*$, $p < n$. Comme $p < n$, on a $p \leq n-1$

$C_{n-1}^p$, $C_{n-1}^{p-1}$ et $C_n^p$ ont donc un sens.

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

$$= \left\{ \frac{p}{p} \times \frac{(n-1)!}{(p-1)!\ (n-p)!} \right\} + \left\{ \frac{(n-p)}{(n-p)} \times \frac{(n-1)!}{p!\,(n-1-p)!} \right\}$$

$$= \frac{p\ (n-1)!}{p!\ (n-p)!} + \frac{(n-p)(n-1)!}{p!\ (n-p)!} = \frac{(n-1)![\,n-p+p]}{p!\ (n-p)!}$$

$$= C_n^p$$

(2) Soit $n \in N$. On a : $C_n^n = \frac{n!}{n!\ 0!} = 1$ car $0! = 1$ (cf. p. 58)

(3) Soient $n \in N$, $p \in N$, $p \leq n$ :

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

(4) Soit $n \in N$. D'après la question précédente $C_n^n = C_n^{n-n} = C_n^o$. D'après (2) on conclut que $C_n^o = 1$.

(5) Soit $n \in N^\star$. D'après (1) $C_{2n}^n = C_{2n-1}^{n-1} + C_{2n-1}^n$. D'après (3) $C_{2n-1}^{n-1} = C_{2n-1}^{(2n-1)-(n-1)} = C_{2n-1}^n$, d'où le résultat : $C_{2n}^n = 2\, C_{2n-1}^n = 2\, C_{2n-1}^{n-1}$. ∎

---

II. Démontrer les relations suivantes :

(1) $\forall n \in N, \forall p \in N, p < n, \dfrac{C_n^{p+1}}{C_n^p} = \dfrac{n-p}{p+1}$

(2) $\forall n \in N, \forall p \in N, p \leq n, \dfrac{C_n^p}{C_{n+1}^p} = \dfrac{n-p+1}{n+1}$

(3) $\forall n \in N, \forall p \in N, p \leq n, \dfrac{C_{n+1}^{p+1}}{C_n^p} = \dfrac{n+1}{p+1}$

---

SOLUTION

On vérifie tout d'abord que, sous les hypothèses faites, les termes des égalités ont bien un sens.

(1) Soient $n \in N$, $p \in N$, $p < n$.

$$\frac{C_n^{p+1}}{C_n^p} = \frac{n!}{(p+1)!(n-p-1)!} \times \frac{p!(n-p)!}{n!} = \frac{p!}{(p+1)!} \times \frac{(n-p)!}{(n-p-1)!}$$
$$= \frac{n-p}{p+1}$$

(2) Soient $n \in N$, $p \in N$, $p \leq n$.

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

(3) Soient $n \in N$, $p \in N$, $p \leq n$

$$\frac{C_{n+1}^{p+1}}{C_n^p} = \frac{C_{n+1}^{p+1}}{C_{n+1}^p} \times \frac{C_{n+1}^p}{C_n^p} = \frac{n-p+1}{p+1} \times \frac{n+1}{n-p+1} \text{ d'après (1) et (2)}$$

$$= \frac{n+1}{p+1} .$$ ■

---

III. Une entreprise veut engager 4 ingénieurs dans 4 spécialités différentes. Six ingénieurs se présentent.

Combien de choix s'offrent au responsable de l'embauche dans les 3 cas suivants :

1°) Les 6 ingénieurs sont polyvalents (pouvant occuper tous un des 4 postes)

2°) Un seul est polyvalent pour les 4 branches, les 5 autres le sont seulement dans trois branches, les mêmes pour tous les 5.

3°) Parmi les 6 ingénieurs, se trouvent 3 hommes et 3 femmes, tous polyvalents. L'équipe recherchée doit comprendre 2 hommes et 2 femmes.

---

SOLUTION

1°) $n = C_6^4 = \frac{6!}{4!\,2!} = 15$ choix

2°) Puisque celui qui est spécialisé dans les 4 branches doit être obligatoirement pris dans le poste qu'il est le seul à pouvoir assurer, on a $n = 1 \times C_5^3 = \frac{5!}{3!\,2!} = 10$ choix.

3°) Il y a $C_3^2$ façons de former l'équipe masculine et $C_3^2$ façons de former l'équipe féminine, donc au total :
$n = C_3^2 \times C_3^2 = 9$ choix. ■

IV. Le traitement d'un malade nécessite la prise de 2 sirops différents et de 3 sortes de cachets. Le médecin dispose de 3 sortes de sirops et de 4 sortes de cachets qui auraient des effets analogues.

De combien de façons différentes pourra-t-il rédiger son ordonnance, sachant toutefois, qu'un sirop précis ne doit pas être pris en même temps qu'une sorte de cachet précis ?

---

SOLUTION

Soient X et Y le sirop et le cachet qui ne peuvent être pris ensemble. Calculons le nombre de fois où le sirop X est pris avec le cachet Y. Une fois X et Y utilisés, il reste au médecin à choisir un sirop parmi les 2 autres et 2 sortes de cachets parmi les 3 autres, d'où

$$N_1 = 1 \times 1 \times C_2^1 \times C_3^2 = 6$$

Le nombre total d'ordonnances que le médecin peut rédiger, sans prendre garde au choix des divers médicaments, est :

$$N_2 = C_3^2 \times C_4^3 = 12$$

D'où, par différence, le nombre d'ordonnances où le sirop X ne sera pas avec le cachet Y

$$N = N_2 - N_1 = 12 - 6 = 6.$$ ■

---

V. LE TRIANGLE DE PASCAL

La construction du triangle de Pascal (cf. p. 65) résulte de la relation (1) $\forall n \in \mathbb{N}^*$, $\forall p \in \mathbb{N}^*$, $p < n$, $C_n^p = C_{n-1}^{p-1} + C_{n-1}^p$.

On se propose, dans cet exercice, de retrouver quelques propriétés du triangle de Pascal à partir de la relation (1). Les raisonnements n'utiliseront donc pas l'écriture explicite

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

(Les relations qui suivent furent démontrées par B. Pascal dans son Traité du triangle arithmétique, Paris 1665).

(A1) $\forall n \in N^*, \forall p \in N^*, p < n, C_n^p = \sum_{k=0}^{p} C_{n-(k+1)}^{p-k}$

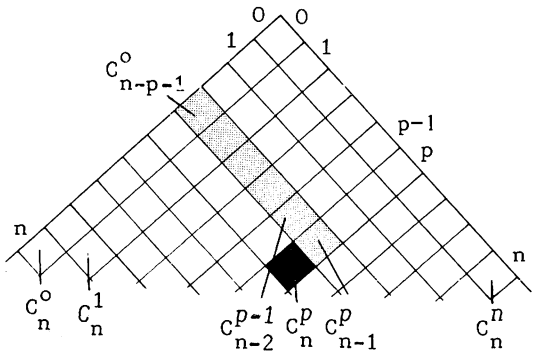
(A2) $\forall n \in N^*, \forall p \in N^*, p < n, C_n^p = \sum_{k=0}^{n-p} C_{p-1+k}^{p-1}$

(A3) $\forall n \in N^*, \forall p \in N^*, p < n, C_n^p - 1 = \sum_{k=0}^{p-1} \sum_{\ell=p-k+1}^{n-k} C_{n-\ell}^{k}$

---

SOLUTION

(A1) Soient $n \in N^*$, $p \in N^*$, $0 < p < n$. $C_n^p$ et $C_{n-(k+1)}^{p-k}$, où $0 \leq k \leq p$, existent alors. Les $C_{n-(k+1)}^{p-k}$, $0 \leq k \leq p$, sont représentés en gris sur le triangle :




D'après la relation (1), on a :

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

$$C_{n-1}^{p-1} = C_{n-2}^{p-1} + C_{n-2}^{p-2}$$

$$C_{n-2}^{p-2} = C_{n-3}^{p-2} + C_{n-3}^{p-3}$$

.....

On en déduit que :

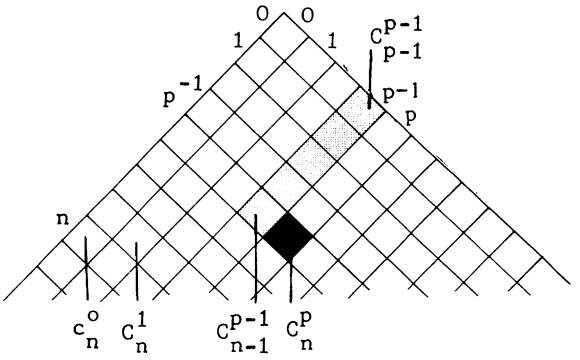
$$C_n^p = C_{n-1}^p + C_{n-2}^{p-1} + C_{n-3}^{p-2} + \ldots + C_{n-(k+1)}^{p-k} + \ldots$$
$$+ C_{n-p+1}^2 + C_{n-p}^1 + C_{n-p}^o$$

Or $C_{n-p}^o = C_{n-(p+1)}^o = 1$

On en déduit donc que :

$$C_n^p = \sum_{k=0}^{p} C_{n-(k+1)}^{p-k}$$

(A2) Soient $n \in N^*$, $p \in N^*$, $0 < p < n$. On vérifie que $C_n^p$ et $C_{p-1+k}^{p-1}$, $0 \leq k \leq p-1$, existent alors. Les $C_{p-1+k}^{p-1}$, $0 \leq k \leq n-p$, sont représentés en gris sur le triangle :




D'après la question (3) de l'exercice 4 C I., $C_n^p = C_n^{n-p}$.
D'après la question précédente :

$$C_n^p = \sum_{k=0}^{p} C_{n-(k+1)}^{p-k} = \sum_{k=0}^{p} C_{n-(k+1)}^{n-(k+1)-(p-k)} = \sum_{k=0}^{p} C_{n-k-1}^{n-p-1}$$

On a donc : $\forall m \in N^*$, $\forall q \in N^*$, $0 < q < m$, $C_m^q = \sum_{k=0}^{q} C_{m-k-1}^{m-q-1}$

En posant $m = n$ et $q = n-p$, on a :

$$C_n^p = C_n^{n-p} = \sum_{k=0}^{n-p} C_{n-k-1}^{p-1}$$

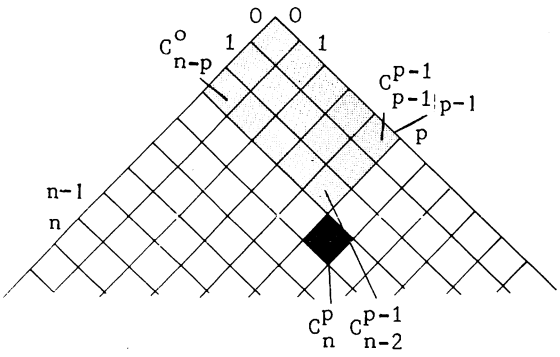
En faisant le changement de variables $\ell = n-p-k$, il vient :

$$C_n^p = \sum_{\ell=0}^{n-p} C_{p-1+\ell}^{p-1}$$

(Une démonstration directe est évidemment possible ; elle serait sensiblement identique à celle de la question précédente).

(A3) Soient $n \in N^*$, $p \in N^*$, $0 < p < n$. On vérifie que sous ces hypothèses $C_n^p$ et $C_{n-\ell}^k$, $0 \leq k \leq p-1$ et $p-k-1 \leq \ell \leq n-k$, existent.

Les $C_{n-\ell}^k$, $0 \leq k \leq p-1$ et $p-k-1 \leq \ell \leq n-k$ sont représentés en gris sur le triangle.




On a pour $0 \leq k \leq p-1$

$$\sum_{\ell=p-k+1}^{n-k} C_{n-\ell}^k = \sum_{j=0}^{n-p-1} C_{k+j}^k$$

en faisant le changement de variable $j = n-k-\ell$.

D'après (A2) on a :

$$C_n^p = \sum_{\ell=0}^{n-p-1} C_{p-1+\ell}^{p-1} + C_{n-1}^{p-1}$$

$$C_{n-1}^{p-1} = \sum_{\ell=0}^{n-p-1} C_{p-2+\ell}^{p-2} + C_{n-2}^{p-2}$$

$$C_{n-2}^{p-2} = \sum_{\ell=0}^{n-p-1} C_{p-3+\ell}^{p-3} + C_{n-3}^{p-3}$$

.....

on en déduit que :

$$C_n^p = \sum_{\ell=0}^{n-p-1} C_{p-1+\ell}^{p-1} + \sum_{\ell=0}^{n-p-1} C_{p-2+\ell}^{p-2} + \sum_{\ell=0}^{n-p-1} C_{p-3+\ell}^{p-3} + \ldots + \sum_{\ell=0}^{n-p-1} C_\ell^0 + C_{n-p+1}^0$$

Or $C_{n-p+1}^0 = 1$, on a donc : $C_n^p - 1 = \sum_{k=0}^{p-1} \sum_{\ell=p-k+1}^{n-k} C_{n-\ell}^k$ . ■

---

## VI. LE BINÔME DE NEWTON

Soient a et b deux nombres.

(1) Démontrer que :

$$\forall n \in N : (a+b)^n = \sum_{k=0}^{n} C_n^k a^{n-k} b^n$$

(2) En déduire que : $\forall n \in N$, $\sum_{k=0}^{n} C_n^k = 2^n$

(3) A partir du calcul de $(1-1)^n$ pour $n \in N^*$ et de (2) calculer

$$\sum_{\substack{0 \leq k \leq n \\ k \text{ pair}}} C_n^k \qquad \text{et} \qquad \sum_{\substack{1 \leq k \leq n \\ k \text{ impair}}} C_n^k$$

SOLUTION

(1) Il est possible de démontrer la formule du binôme de deux manières : par récurrence, en dénombrant le nombre de facteurs.

*Démonstration par récurrence*

a) Il est clair que $(a+b)^0 = C_0^0 a^0 b^0 = 1$

$$(a+b)^1 = C_1^0 a^1 b^0 + C_1^1 a^0 b^1$$

b) Supposons que $(a+b)^n = \sum_{k=0}^{n} C_n^k a^{n-k} b^k$ et montrons que

$$(a+b)^{n+1} = \sum_{k=0}^{n+1} C_{n+1}^k a^{n+1-k} b^k$$

$$(a+b)^{n+1} = (a+b)^n \quad (a+b) = \left\{ \sum_{k=0}^{n} C_n^k a^{n-k} b^k \right\} (a+b)$$

Or il existe des coefficients $A_k$ tels que :

$$(a+b)^{n+1} = \sum_{k=0}^{n+1} A_k a^{n+1-k} b^k$$

et on a : $A_k = C_n^{k-1} + C_n^k = C_{n+1}^k$ pour $1 \leq k \leq n$ (cf. exercice I)

$$A_{n+1} = C_{n+1}^{n+1} = 1$$

$$A_o = C_{n+1}^o = 1$$

d'où le résultat.

*Démonstration directe*

Soit $n \in N^*$. On sait qu'il existe des coefficients $A_k$, $0 \leq k \leq n$, tels que : $(a+b)^n = \sum_{k=0}^{n} A_k a^{n-k} b^k$

pour $k \in N$, $0 \leq k \leq n$, le nombre de produits $a^{n-k} b^k$ dans $(a+b)^n$ est égal au nombre de manières de choisir k fois b parmi n termes, sans tenir compte de l'ordre. On a donc : $A_k = C_n^k$.

(2) D'après la question précédente, on a : $\forall n \in N^*$,

$$(1+1)^n = 2^n = \sum_{k=0}^{n} C_n^k \, 1^{n-k} \, 1^k = \sum_{k=0}^{n} C_n^k .$$

$$(3) \quad (1-1)^n = 0 = \sum_{k=0}^{n} C_n^k \, 1^{n-k} \, (-1)^k$$

$(-1)^k = 1$ si k est pair

$(-1)^k = -1$ si k est impair.

On en déduit donc que :

$$\sum_{\substack{0 \leqslant k \leqslant n \\ k \text{ pair}}} C_n^k = \sum_{\substack{1 \leqslant k \leqslant n \\ k \text{ impair}}} C_n^k$$

D'autre part : $\sum_{k=0}^{n} C_n^k = \sum_{\substack{0 \leqslant k \leqslant n \\ k \text{ pair}}} C_n^k + \sum_{\substack{1 \leqslant k \leqslant n \\ k \text{ impair}}} C_n^k = 2^n$

On a donc : $\sum_{\substack{0 \leqslant k \leqslant n \\ k \text{ pair}}} C_n^k = \sum_{\substack{1 \leqslant k \leqslant n \\ k \text{ impair}}} C_n^k = 2^{n-1}$ ■

---

VII. Soit E un ensemble fini de n éléments.

Quel est le nombre de parties de E ?

---

SOLUTION

Il est clair qu'il y a 1 seule partie vide, et qu'il y a n parties à un seul élément.

Soit $k \in N$, $0 \leqslant k \leqslant n$. Une partie de k éléments est par définition une combinaison de n éléments pris k à k. Il y a donc $C_n^k$ telles parties. Le nombre de parties de E est donc :

$$\text{Card } \mathcal{P}(E) = \sum_{k=0}^{n} C_n^k = 2^n \text{ (cf. exercice VI).}$$ ■

# 5. Calcul des probabilités

## A. LOGIQUE DES EVENEMENTS

### I. INTRODUCTION

La notion de probabilité est tout d'abord d'ordre psychologique. Par exemple, on parle de la probabilité d'obtenir une paire au poker. De même, certains observateurs de la vie politique qualifient de probable telle rencontre internationale "au sommet".

Cependant, ces deux exemples diffèrent dans la mesure où dans le premier cas il s'agit d'une expérience qui peut être répétée plusieurs fois dans les mêmes conditions, alors que dans le second, on ne peut parler d'expérience : une réunion "au sommet" n'est pas régie par des règles précises, sa probabilité s'appuie sur une appréciation subjective de la situation politique.

Une théorie quantitative de la notion de probabilité ne doit considérer nécessairement que des cas où il existe une "probabilité objective", c'est à dire qui ne dépend pas des convictions personnelles.

## II. NOTIONS DE BASE

### Evénement

On peut dire que tout ce qui peut se réaliser ou ne pas se réaliser, à la suite d'une expérience spontanée ou provoquée parfaitement définie, est un événement.

. *Exemples*

En jetant un dé :

- "obtenir un six" est un événement (que l'on peut désigner par exemple par E) ;
- "ne pas obtenir de six" ou "obtenir un non-six" est l'événement contraire du précédent, noté $\overline{E}$ ;
- "obtenir un nombre entier compris entre 1 et 6" est un événement certain ;
- "obtenir un sept" est un événement impossible.

On voit que la notion d'événement est liée à la notion intuitive d'expérience aléatoire.

### Expérience aléatoire

Une expérience aléatoire ou épreuve est un ensemble de conditions précises caractérisant un processus à la suite duquel l'événement est réalisé ou non. On se limite aux cas où l'expérience peut être répétée plusieurs fois, dans les mêmes conditions (cf. 5 B II.).

Dans les exemples précédents, l'expérience aléatoire consiste simplement à jeter le dé.

### Evénement élémentaire

Il s'agit d'un événement qui ne sera réalisé que par un seul résultat de l'expérience aléatoire.

. *Exemples*

- "obtenir un six" en jetant un dé, est un événement élémentaire ;

- "obtenir un nombre pair" en jetant un dé, n'est pas un événement élémentaire, car il peut être réalisé par plusieurs résultats de l'expérience aléatoire qui sont :

"obtenir deux"

"obtenir quatre"

"obtenir six".

Ensemble fondamental associé à une expérience aléatoire

L'ensemble fondamental (ou univers) associé à une expérience aléatoire est l'ensemble des résultats de l'expérience considérée.

. *Exemples*

Expérience aléatoire $\mathscr{A}$ : on jette un dé. Les résultats possibles de $\mathscr{A}$ sont :

$r_i$ : "on obtient le chiffre i" pour i entier compris entre 1 et 6. L'ensemble fondamental est donc :

$$S = \{r_1, r_2, r_3, r_4, r_5, r_6\}$$

A l'événement élémentaire "on obtient le chiffre 1", on peut associer le singleton

$$R_1 = \{r_1\}$$

L'événement "on obtient un nombre pair" peut être représenté par le sous-ensemble

$$R_{pair} = \{r_2, r_4, r_6\}$$

qui est la réunion des événements $R_2$, $R_4$ et $R_6$.

L'événement $R_7$ "on obtient le chiffre 7" ne sera jamais réalisé (événement impossible). On pose donc

$$R_7 = \phi$$

De même, l'événement "on obtient un chiffre compris entre 1 et 6" sera toujours réalisé (événement certain). On pose

$$R_{1 \leq i \leq 6} = S$$

Ainsi, tout événement, élémentaire ou non, peut être considéré comme une partie (sous-ensemble) de l'ensemble fondamental S. Lorsque S est fini ou dénombrable, l'ensemble des événements est l'ensemble des parties de S, soit $\mathcal{P}(S)$.

## III. LOGIQUE DES EVENEMENTS

A une expérience aléatoire $\mathcal{A}$, on peut donc associer un ensemble fondamental S. Un événement peut être considéré comme une partie de S.

$\phi$ est l'événement qui n'est jamais réalisé ou événement impossible. S est l'événement qui est toujours réalisé ou événement certain.

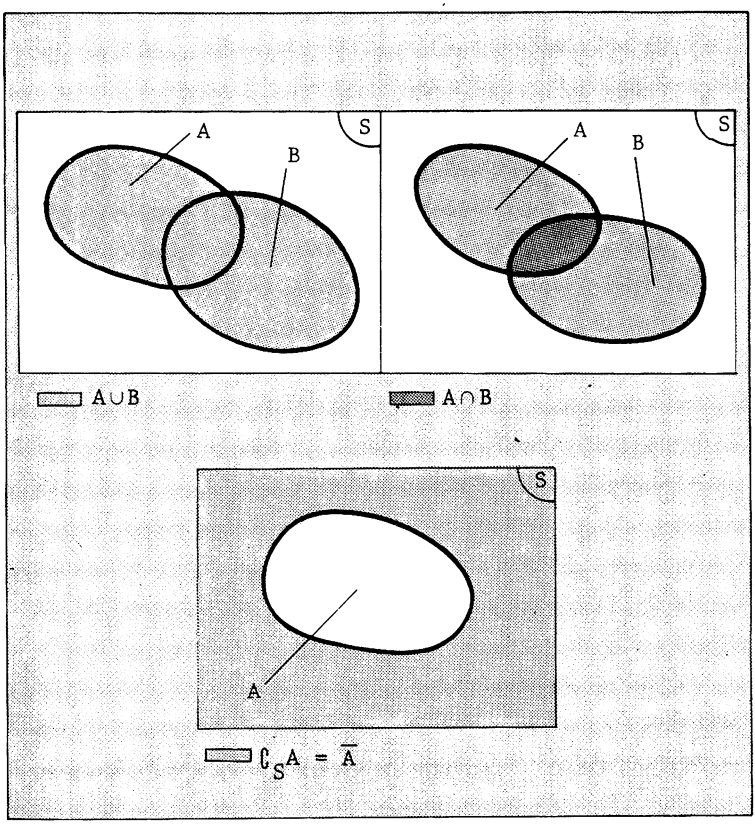
Soient A et B deux événements ($A \subset S$ et $B \subset S$). On peut alors établir les correspondances suivantes :

$A \cup B$ : (union de A et B) désigne un événement qui est réalisé si au moins un des événements A et B est réalisé.

$A \cap B$ : (intersection de A et B) désigne un événement qui est réalisé si A et B sont réalisés.

$C_S A$ : (complémentaire de A dans S) est l'événement qui est réalisé si et seulement si A n'est pas réalisé. On note aussi $C_S A = \overline{A}$ qui désigne l'événement contraire de A. On a en particulier $\overline{\phi} = S$ et $\overline{S} = \phi$

$A \cap B = \phi$ (A et B sont disjoints, leur intersection est l'ensemble vide). Il s'agit de deux événements qui ne peuvent se réaliser simultanément. On dit qu'ils s'excluent mutuellement ou qu'ils sont incompatibles.




. *Exemple*

Expérience aléatoire : tirer une carte d'un jeu de 32 cartes. L'ensemble fondamental de tous les résultats possibles est constitué de 32 éléments. On désigne par :

A : l'événement "tirer un as" auquel on fait correspondre le sous-ensemble des as

R : l'événement "tirer un roi" auquel on fait correspondre le sous-ensemble des rois

C : l'événement "tirer un coeur" auquel on fait correspondre le sous-ensemble des couleurs "coeur".

On peut alors faire les correspondances suivantes :

$A \cup C$ : événement "tirer un as ou un coeur" (y compris l'as de coeur)

$A \cap C$ : "tirer l'as de coeur"

$A \cap R$ : "tirer une carte qui soit à la fois un as et un roi"
$A \cap R = \phi$, événement impossible.

$C_S\ A$ : "tirer une carte autre que as".

## B. PROBABILITE

Deux démarches différentes peuvent conduire à la définition axiomatique d'une probabilité.

### I. PROBABILITE UNIFORME

Considérons l'expérience aléatoire $\mathscr{A}$ : "on jette un dé". L'ensemble fondamental est

$$S = \{r_1, r_2, r_3, r_4, r_5, r_6\}$$

les événements élémentaires sont

$$R_i = \{r_i\} \quad \text{avec } i \in N,\ 1 \leqslant i \leqslant 6$$

On désigne par $p(R_i)$ la probabilité que $R_i$ se réalise. On peut écrire que

$$p(S) = p(R_1) + p(R_2) + p(R_3) + p(R_4) + p(R_5) + p(R_6) = 1$$

puisque l'événement S est certain.

Si l'on suppose que le dé est symétrique et homogène (on dit aussi parfait), chaque face a autant de chances d'apparaître que n'importe quelle autre face. Toutes les probabilités élémentaires $p(R_i)$ sont donc égales entre elles.

On en déduit

$$p(R_i) = p = \frac{1}{6}$$

La probabilité est dite uniforme, les événements élémentaires sont dits équiprobables.

D'une manière générale, si on considère une expérience aléatoire $\mathscr{B}$ et un ensemble fondamental S associé à $\mathscr{B}$, si

on définit une probabilité uniforme sur S et si E est un événement réunion de n événements élémentaires distincts, on a

$$p\ (E) = \frac{\text{Card } E}{\text{Card } S} \qquad (5.1)$$

On exprime cette relation en disant que, dans le cas d'une probabilité uniforme, la probabilité d'un événement est égale au nombre de cas favorables à la réalisation de cet événement (Card E) divisé par le nombre de résultats possibles de l'expérience (Card S).

## II. PROBABILITE ET FREQUENCE

On a vu au paragraphe précédent comment construire une probabilité lorsqu'il est raisonnable de penser que les événements élémentaires sont équiprobables. Considérons maintenant une expérience aléatoire $\mathscr{E}$ quelconque. Soit A un événement, $F_n\ (A)$ et $f_n\ (A)$ ses fréquences absolue et relative de réalisation lors d'une succession de n épreuves. On a (cf. Chap. 1) :

$$f_n\ (A) = \frac{F_n\ (A)}{n} \qquad (5.2)$$

Il est clair que $f_n\ (A)$ dépend de la série des n épreuves : deux séries différentes peuvent conduire à des résultats différents. Il est cependant raisonnable de penser que $f_n\ (A)$ tend vers une limite lorsque le nombre d'épreuves n tend vers l'infini. On dit que la probabilité P (A) de l'événement A est cette limite.

Cette approche étend considérablement le champ des expériences probabilisables, mais ne permet pas de parler de probabilité dans le cas d'une épreuve qui ne peut être répétée, pas plus que dans des situations où le probable est subjectif.

Cette limitation se justifie essentiellement par la théorie : la loi des grands nombres (cf. 6 A IV), établie par J. Bernouilli en 1689 est conforme avec cette "hypothèse raisonnable".

Cependant dans de nombreux domaines d'application des probabilités (notamment en économie), ce choix "fréquentiste" est trop limitatif. (Sur ce sujet, le lecteur pourra consulter l'article de Benjamin Matalon, "Epistémologie des Probabilités" dans le volume "Logique et connaissance scientifique" de l'encyclopédie de la Pléiade, Paris 1967).

## III. DEFINITION D'UNE PROBABILITE

Nous allons maintenant donner une définition axiomatique d'une probabilité (Axiomatique de Kolmogoroff).

Soit S un ensemble fondamental fini ou dénombrable associé à une expérience aléatoire $\mathscr{A}$.

Une probabilité P sur S est une application

$P : \mathscr{P}(S) \longrightarrow [0, 1]$ telle que :

1°) $P(S) = 1$ et $P(\phi) = 0$ (5.3)

2°) $\forall A \in \mathscr{P}(S)$, $\forall B \in \mathscr{P}(S)$, tels que $A \cap B = \phi$

$$P(A \cup B) = P(A) + P(B) \tag{5.4}$$

3° Si $A_0, A_1, \ldots, A_n, \ldots$ est une suite d'événements incompatibles deux à deux (i.e. $\forall n, \forall m, A_n \cap A_m = \phi$) alors :

$$P\left(\bigcup_{n=0}^{\infty} A_n\right) = \sum_{n=0}^{\infty} P(A_n) \tag{5.5}$$

cette propriété s'appelle la σ-additivité.

Remarque

Par récurrence, on déduit du second axiome que si $A_0, A_1, \ldots, A_n$ est une suite finie d'évènements incompatibles deux à deux, alors :

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) \tag{5.6}$$

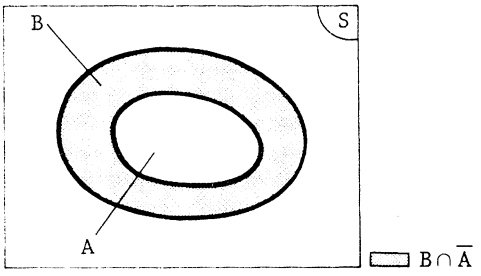
Si S est fini, le troisième axiome est donc redondant.

*Conséquences :*

1°) $\forall A \in \mathscr{P}(S), \forall B \in \mathscr{P}(S), A \subset B \Rightarrow P(A) \leqslant P(B)$ (5.7)

(On dit qu'une probabilité est une application croissante sur $\mathscr{P}(S)$ ).

En effet : $A \subset B \Rightarrow B = (B \cap \bar{A}) \cup A$ et on a : $(B \cap \bar{A}) \cap A = \phi$. D'après l'axiome 2°), $P(B) = P(A) + P(B \cap \bar{A})$ et comme $P(B \cap \bar{A}) \geq 0$, on a bien : $P(A) \leq P(B)$.




2°) Condition de normalisation :

$$\forall A \in \mathscr{P}(S), P(A) = 1 - P(\bar{A}) \quad (5.8)$$

En effet : $S = A \cup \bar{A}$ et $A \cap \bar{A} = \phi$. D'après l'axiome 2) $P(S) = P(A) + P(\bar{A})$. Comme $P(S) = 1$ (axiome 1) on a $P(A) = 1 - P(\bar{A})$.

3°) Cas particulier d'une probabilité uniforme

Remarquons tout d'abord que si la probabilité est uniforme, alors S est nécessairement fini. On a :

$$\forall A \in \mathscr{P}(S), P(A) = \frac{\text{Card } A}{\text{Card } S} \quad (5.9)$$

En effet si $S = \{r_1, r_2, \ldots, r_n\}$ et $A = \{r_1, \ldots, r_k\}$ $(k \leqslant n)$ alors $P(A) = P(\{r_1\}) + \ldots + P(\{r_k\})$

$$= \underbrace{\frac{1}{n} + \ldots + \frac{1}{n}}_{k \text{ fois}} = \frac{k}{n} = \frac{\text{Card } A}{\text{Card } S}$$

La définition axiomatique est donc cohérente avec l'hypothèse du § I.

4°) Si S est fini ou dénombrable, une probabilité P sur S est entièrement déterminée par la donnée d'une famille $p_i$ où $p_i \geq 0$, $\Sigma\, p_i = 1$, et où $p_i$ est la probabilité du $i^{ème}$ événement élémentaire.

. *Exemples*

1°) Expérience aléatoire : tirer une carte d'un jeu de 32 cartes.

On considère les événements suivants :

A : "tirer un as"

R : "tirer un roi"

C : "tirer un coeur"

$A \cup C$ : "tirer un as ou un coeur"

$A \cap C$ : "tirer l'as de coeur"

$A \cap R$ : "tirer une carte qui soit à la fois un as et un roi"

$\overline{A}$ : "tirer une carte autre que as".

Il s'agit ici de probabilité uniforme. En appliquant (5.9) on obtient successivement

$$p(A) = P(R) = \frac{4}{32} = \frac{1}{8} \qquad p(C) = \frac{8}{32} = \frac{1}{4}$$

$$p(A \cup C) = \frac{\text{Card}\,(A \cup C)}{\text{Card}\,S} = \frac{11}{32} \qquad p(A \cap C) = \frac{1}{32}$$

$$p(A \cap R) = \frac{0}{32} = 0 \qquad p(\overline{A}) = \frac{28}{32} = \frac{7}{28}$$

2°) Expérience aléatoire : tirer 2 cartes à la fois d'un jeu de 32 cartes. Probabilité pour que ces 2 cartes soient 2 rois.

Là encore, il s'agit d'une probabilité uniforme. Le nombre de cas favorables n est égal au nombre de manières de combiner 4 rois 2 à 2, soit $C_4^2 = 6$.

Le nombre de cas possibles N est égal au nombre de

manières de combiner 32 cartes 2 à 2, soit $C_{32}^{2} = 496$. On en déduit

$$p\ (E) = \frac{6}{496} \simeq 0{,}012$$

3°) Considérons une suite infinie de jeux de pile ou face. Soit S un ensemble fondamental associé à cette expérience aléatoire : $S = \{(x_i)_{i \in N}\ ,\ x_i = P \text{ ou } x_i = F\}$. Soit A l'événement : "on obtient toujours Pile". On a :

$$A = \{(x_i)_{i \in N}\} \quad \text{où } \forall i \in N \quad x_i = P.$$

Si la pièce est bien équilibrée, la probabilité de n'obtenir que des "Pile", lors des n premiers jets est $\frac{1}{2^n}$ en vertu du principe général de l'analyse combinatoire (cf. Chap. 4). On en déduit que P (A) = 0 car $\frac{1}{2^n} \to 0$ quand $n \to +\infty$. De même la probabilité d'obtenir Pile au premier jet, puis Face, puis Pile, etc. est nulle.

Ces événements ne sont pourtant pas à proprement parler impossibles, on dit qu'ils sont presque impossibles. L'événement B : "on obtient au moins une fois Face" est l'évènement contraire de A. Donc P (B) = 1. On dit que B est un événement presque certain.

A première vue, il peut sembler paradoxal qu'on ait p (A) = 0 ou p (B) = 1. En fait, cela résulte du choix "fréquentiste" qui exprime la probabilité comme une limite de fréquence (cf. 5 B II).

## C. PROBABILITES TOTALES

Le second axiome de définition d'une probabilité donne la probabilité de la réunion de deux événements lorsqu'ils sont incompatibles. Le théorème des probabilités totales donne une expression de cette probabilité dans le cas général.

*Exemple*

On tire une carte au hasard d'un jeu de 32 cartes. Quelle est la probabilité pour que cette carte soit un as ou un coeur ?

Soient E, A et B les événements :

E : "la carte est un as ou un coeur"

A : "la carte est un as"

B : "la carte est un coeur".

On a : $E = A \cup B$ et : $p(A) = \frac{4}{32}$, $p(B) = \frac{8}{32}$,

$p(E) = \frac{\text{Card } (A \cup B)}{\text{Card } S} = \frac{11}{32}$

*Les ensembles A et B n'étant pas disjoints, pour ne pas* compter deux fois l'as de coeur, il convient de remarquer que :

$$\text{Card } (A \cup B) = \text{Card } A + \text{Card } B - \text{Card } (A \cap B)$$

On en déduit que :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad (5.10)$$

L'expression (5.10) constitue l'énoncé du théorème des probabilités totales. Lorsque $A \cap B = \phi$, on retrouve le second axiome de définition d'une probabilité

$$P(A \cup B) = P(A) + P(B) \qquad (5.11)$$

## I. THEOREME DES PROBABILITES TOTALES

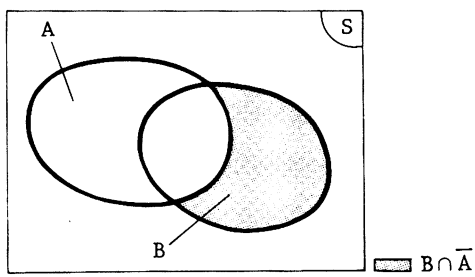
$(\forall A \in \mathcal{P}(S))\ (\forall B \in \mathcal{P}(S))$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad (5.12)$$

Démonstration :

$A \cup B = A \cup (B \cap \bar{A})$ et $A \cap (B \cap \bar{A}) = \phi$

$B = (B \cap \bar{A}) \cup (A \cap B)$ et $(B \cap \bar{A}) \cap (A \cap B) = \phi$




D'après la propriété (5.4) on a :

$$P(A \cup B) = P[A \cup (B \cap \bar{A})] = P(A) + P(B \cap \bar{A})$$

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

et par conséquent

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

## II. GENERALISATION

$(\forall A \in \mathcal{P}(S))\ (\forall B \in \mathcal{P}(S))\ (\forall C \in \mathcal{P}(S))$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C) \qquad (5.13)$$

En effet, en utilisant deux fois le théorème des probabilités totales (5.12) on a

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

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

or, $P[(A \cup B) \cap C] = P[(A \cap C) \cup (B \cap C)]$

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

d'où le résultat (5.13).

D'une manière générale, si P est une probabilité sur un ensemble S et si $A_1$, $A_2$, ..., $A_n$ sont n éléments de S, on a :

$$P\left[\bigcup_{i=1}^{n} A_i\right] = \sum_{i=1}^{n} P(A_i) - \sum_{i} \sum_{j \neq i} P(A_i \cap A_j)$$

$$+ \sum_{i} \sum_{j \neq i} \sum_{\substack{k \neq i \\ k \neq j}} P(A_i \cap A_j \cap A_k) - \ldots$$

$$+ (-1)^{n+1} P\left[\bigcap_{i=1}^{n} A_i\right]$$

Remarque

Si n = 2, on vérifie que l'on retrouve le théorème (5.12) (cf. exercice II).

## D. PROBABILITES COMPOSEES ET THEOREME DE BAYES

### I. DEFINITION D'UNE PROBABILITE COMPOSEE

Soit P une probabilité sur un ensemble S, et B un événement tel que $P(B) \neq 0$. On appelle probabilité d'un événement "A si B" ou A/B (probabilité composée ou conditionnelle) le rapport :

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \tag{5.14}$$

On en déduit

$$P(A \cap B) = P(B) \cdot P(A/B) \tag{5.15}$$

De la même manière, si $P(A) \neq 0$, on peut écrire

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

*Généralisation*

Si $A_1, \ldots, A_n$ sont n événements, alors :

$$P\left[\bigcap_{i=1}^{n} A_i\right] = P(A_1).P(A_2/A_1).P(A_3/A_1 \cap A_2) \ldots \ldots P\left(A_n / \bigcap_{i=1}^{n-1} A_i\right) \tag{5.16}$$

(Cette relation est démontrée dans l'exercice 5 D III)

. *Exemple*

Probabilité de tirer un as, puis un roi d'un jeu de 32 cartes, sans remettre la 1ère carte en jeu (tirage exhaustif)

A : la 1ère carte tirée est un as

B : la 2ème carte tirée est un roi

B/A : la 2ème carte tirée est un roi, sachant que la 1ère (non remise) est un as

$$p\ (A \cap B) = p\ (A).p\ (B/A) = \frac{4}{32} \times \frac{4}{31} = \frac{1}{62}$$

$p\ (B/A) = \frac{4}{31}$ puisqu'il ne reste plus que 31 cartes.

## II. EVENEMENTS INDEPENDANTS

Considérons une expérience aléatoire consistant à jeter un dé deux fois successives.

Soient A et B les événements

A : "on obtient 6 lors du premier jet"

B : "on obtient 6 lors du second jet"

Il est clair que $P\ (B/A) = P\ (B)$. On dit que les événements A et B sont indépendants.

D'une manière générale, on peut remarquer que :

1°) Si $P\ (A) \neq 0$ et $P\ (B) \neq 0$

$P\ (B/A) = P\ (B) \Leftrightarrow P\ (A/B) = P\ (A)$

2°) Si $P\ (A) \neq 0$

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

Définition

Deux événements A et B sont indépendants si

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

. *Exemple*

On reprend l'exemple précédent, mais en remettant la 1ère carte dans le jeu (tirage non exhaustif).

Dans ce cas, $P\ (B/A) = P\ (B) = \frac{4}{32}$

$$P\ (A \cap B) = P\ (A).P\ (B) = \frac{4}{32} \cdot \frac{4}{32} = \frac{1}{64}$$

## III. THEOREME DE BAYES

Ce théorème permet de déterminer la probabilité pour qu'un événement qui est supposé déjà réalisé, soit dû à une

certaine cause plutôt qu'à une autre (d'où le nom de théorème des probabilités des causes que lui a donné Bayes).

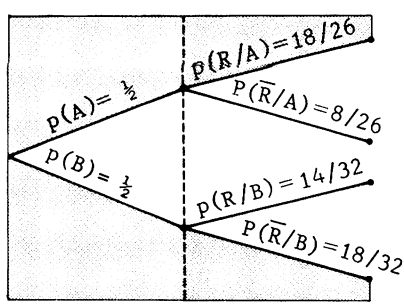
1. Exemple

Le tableau suivant donne, pour les deux classes terminales d'un lycée, le nombre d'élèves ayant été reçus au baccalauréat et l'effectif de chaque classe

| | Classe A | Classe B |
|---|---|---|
| Effectif | 26 | 32 |
| Nombre de reçus | 18 | 14 |

Quelle est la probabilité pour qu'un élève reçu, pris au hasard, provienne de la classe A ?

Le probabilité pour qu'un élève de terminale provienne d'une classe plutôt que d'une autre est $p(A) = p(B) = \frac{1}{2}$. Partant de la classe A, la probabilité pour qu'un élève soit reçu est $p(R/A) = \frac{18}{26}$. En raisonnant de la même manière pour la classe B, on peut construire le diagramme suivant.




La probabilité totale pour qu'un élève de terminale soit reçu est, d'après ce diagramme

$$p\ (A \cap R) + p\ (B \cap R) = p\ (A).p\ (R/A) + p\ (B).p\ (R/B)$$

La probabilité pour qu'un élève de A soit reçu est

$$p\ (A \cap R) = p\ (A).p\ (R/A)$$

Par conséquent, la probabilité pour qu'un élève reçu provienne de A est donnée par

$$p\ (A/R) = \frac{p\ (R/A).p\ (A)}{p\ (R/A).p\ (A) + p\ (R/B)\ p\ (B)} \qquad (5.17)$$

$$= \frac{\frac{18}{26} \times \frac{1}{2}}{\frac{18}{26} \times \frac{1}{2} + \frac{14}{32} \times \frac{1}{2}} = 0{,}613$$

2. Théorème

Soient S un ensemble fondamental et E, $A_1$, $A_2$, ..., $A_n$, n+1 événements tels que :

1) les $A_i$ sont deux à deux disjoints

2) $\forall i = 1, \ldots, n \quad P\ (A_i) \neq 0$

3) $\bigcup_{i=1}^{n} A_i = E$

Alors : $\forall k = 1, \ldots, n$

$$P\ (A_k/E) = \frac{P\ (A_k).P\ (E/A_k)}{\sum_{i=1}^{n} P\ (A_i)\ P\ (E/A_i)} \qquad (5.18)$$

Démonstration

Remarquons tout d'abord que

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

En effet, comme $\bigcup_{i=1}^{n} A_i = E$ on a : $E = \bigcup_{i=1}^{n} [E \cap A_i]$

Comme les $A_i$ sont deux à deux disjoints : $\forall i = 1, \ldots, n$, $\forall j = 1, \ldots, n$, si $i \neq j$ on a :

$$(E \cap A_i) \cap (E \cap A_j) = \phi$$

On a donc :

$$P\ (E) = \sum_{i=1}^{n} P\ (E/A_i).P\ (A_i) \qquad [\text{cf. } (5.6)]$$

$$P\ (A_k/E) = \frac{P\ (A_k \cap E)}{P\ (E)} = \frac{P\ (A_k).P\ (E/A_k)}{\sum_{i=1}^{n} P\ (E/A_i).P\ (A_i)}$$

d'après ce qui précède.

## E. EXEMPLES COMPLEMENTAIRES

### I. LOI BINOMIALE

Une urne contient 100 boules dont 10 blanches. On réalise une série de 5 tirages successifs non exhaustifs d'une boule (c'est à dire que l'on remet la boule après chaque tirage, de manière que la probabilité ne change pas d'un tirage à l'autre). Quelle est la probabilité pour que 3 boules de la série de 5 soient blanches ?

A : tirer une boule blanche $P(A) = \frac{10}{100} = 0,1$

$\overline{A}$ : tirer une boule non blanche $P(\overline{A}) = 1 - 0,1 = 0,9$.

La probabilité de tirer 5 boules dont 3 blanches dans un ordre déterminé tel que A A A $\overline{A}$ $\overline{A}$, est donnée par l'axiome des probabilités composées (5.15)

$$P(A\,A\,A\,\overline{A}\,\overline{A}) = (0,1)^3\,(0,9)^2$$

Mais l'ordre étant indifférent, il existe $C_5^3$ manières de réaliser l'événement précédent. Le théorème des probabilités totales (5.12) fournit donc la probabilité cherchée

$$P(E) = C_5^3\,(0,1)^3\,(0,9)^2$$

*Généralisation*

La probabilité de réaliser k fois l'événement A en une série de n épreuves non exhaustives est

$$P(k) = C_n^k\,p^k\,q^{n-k} \qquad (5.19)$$

où l'on a posé $P(A) = p$ et $P(\overline{A}) = q = 1-p$. On reconnaît le terme général du développement du binôme de Newton $(p+q)^n$ et

l'équ. (4.12), d'où le nom de loi binomiale donnée à l'expression (5.19). C'est une loi à 2 paramètres n et p, notée $\mathscr{B}(n,p)$.

Application au jeu de "pile ou face"

La probabilité de tirer k faces avec une pièce lancée n fois successivement, ou avec n pièces lancées simultanément est

$$P(k) = C_n^k \, p^k \, q^{n-k}$$

avec $p = q = \frac{1}{2}$. Par exemple, la probabilité d'obtenir 3 "face" en 5 lancers est :

$$P(3) = C_5^3 \, (\tfrac{1}{2})^3 \, (\tfrac{1}{2})^2 = \frac{5 \cdot 4 \cdot 3}{1 \cdot 2 \cdot 3} \cdot \frac{1}{8} \cdot \frac{1}{4} = \frac{5}{16}$$

## II. LOI HYPERGEOMETRIQUE

Quelle est la probabilité pour que, au jeu de la belote (32 cartes), un joueur ait une main (de 8 cartes) comportant 5 "pique" et 3 "non-pique" ?

Le problème est différent de l'exemple précédent dans la mesure où le tirage des 8 cartes est ici exhaustif (sans remise).

Le nombre de cas possibles est le nombre de manières de combiner 32 cartes 8 à 8, soit d'après le § 4 C

$$N = C_{32}^8$$

Le nombre de cas favorables peut être obtenu en appliquant le principe fondamental de l'analyse combinatoire (4.1). Les 5 "pique" peuvent être choisis parmi les 8 "pique" existant dans le jeu, de $C_8^5$ manières différentes. Les 3 cartes qui manquent pour constituer une main sont nécessairement des "non-pique", elles peuvent donc être groupées de $C_{24}^3$

façons différentes. Le nombre de cas favorables d'après (4.1) est donc :

$$n = C_8^5 \cdot C_{24}^3$$

Par conséquent, la probabilité cherchée est :

$$P(E) = \frac{n}{N} = \frac{C_8^5 \; C_{24}^3}{C_{32}^8}$$

Généralisation de l'expression précédente

A partir d'un référentiel comptant N éléments dont r d'entre eux ont une caractéristique a, la probabilité d'avoir k fois la caractéristique a, dans une série de n épreuves exhaustives (sans remise) est

$$P(k) = \frac{C_r^k \; C_{N-r}^{n-k}}{C_N^n} \qquad (5.20)$$

La loi (5.20), appelée loi hypergéométrique est une loi à 3 paramètres N, n, r. Elle est notée $\mathcal{H}$ (N,n,r).

I. On considère l'expérience aléatoire suivante :

on jette un dé, si on obtient un chiffre pair, alors on tire une carte d'un jeu de 32 cartes, si on obtient un chiffre impair, on prélève une carte d'un jeu de 52 cartes.

1°) Préciser un ensemble fondamental S associé à cette expérience.

2°) Soient A, B, C, D et E les événements suivants :

A : "on a tiré l'as de trèfle"

B : "on a obtenu un nombre pair"

C : "on a tiré un as"

D : "on a tiré un 2 de coeur et le dé marque 4"

E : "on a tiré une carte 2, 3, 4 ou 5".

Préciser en fonction de S la forme ensembliste de ces événements.

---

SOLUTION

1°) S est l'ensemble des résultats possibles de l'expérience aléatoire. En distinguant les différents résultats possibles du jet du dé, on a :

$$S = \{2, 4, 6\} \times F \cup \{1, 3, 5\} \times G$$

où F désigne l'ensemble des cartes d'un jeu de 32 cartes et G désigne l'ensemble des cartes d'un jeu de 52 cartes car un résultat possible de l'expérience aléatoire est un couple (a,b) où a est un nombre entier compris entre 1 et 6 et b une carte d'un jeu de 32 cartes si a est pair, d'un jeu de 52 cartes si a est impair.

2°) On a alors

$A = \{2, 4, 6\} \times H \cup \{1, 3, 5\} \times H = \{1, 2, 3, 4, 5, 6\} \times H$

où H désigne l'ensemble réduit à l'as de trèfle.

$B = \{2, 4, 6\} \times F$

$C = \{2, 4, 6\} \times I \cup \{1, 3, 5\} \times I = \{1, 2, 3, 4, 5, 6\} \times I$

où I désigne l'ensemble des as.

$D = \phi$ car il est impossible de tirer une carte d'un jeu de 52 cartes qui n'est pas dans un jeu de 32 cartes, lorsqu'on a obtenu un nombre pair.

$E = \{1, 3, 5\} \times J$

où J est l'ensemble des cartes 2, 3, 4 ou 5, car on ne peut tirer une telle carte que si on a obtenu un nombre impair. ■

---

II. Si on ne s'intéresse qu'aux événements A, B, C ou E de l'exercice précédent, est-il possible de considérer un autre ensemble fondamental associé à cette expérience aléatoire ?

---

SOLUTION

Si l'on ne s'intéresse qu'aux événements A, B, C ou E, on ne distingue que la parité du nombre obtenu après le jet du dé et non sa valeur numérique.

On peut donc considérer $S_1$ défini par :

$$S_1 = [\{P\} \times F] \cup [\{I\} \times G]$$

avec P pour pair, et I pour impair.

■

---

III. Soit S un ensemble fondamental, et A, B, C trois événements. Exprimer à partir de A, B ou C et des notations ensemblistes les événements suivants :

$E_1$ : seul A est réalisé

$E_2$ : un événement au moins est réalisé

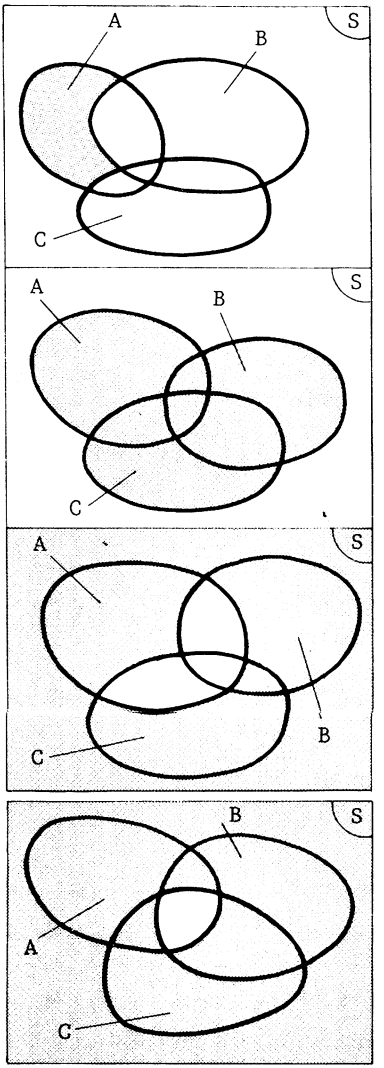
$E_3$ : un événement au plus est réalisé

$E_4$ : les trois événements sont réalisés

$E_5$ : deux événements au plus sont réalisés.

---

SOLUTION




$E_1 = A \cap \overline{B} \cap \overline{C}$

$E_2 = A \cup B \cup C$

$E_3 = [A \cap \overline{B} \cap \overline{C}] \cup [\overline{A} \cap B \cap \overline{C}] \cup [\overline{A} \cap \overline{B} \cap C] \cup [\overline{A} \cap \overline{B} \cap \overline{C}]$

$E_4 = A \cap B \cap C$

$E_5 = \overline{A \cap B \cap C}$

IV. LE JEU DE PASSE-DIX

Le jeu de passe-dix consiste à jeter un dé trois fois successives. On gagne si la somme des points obtenus dépasse 10.

1°) Préciser un ensemble fondamental S qui tient compte de l'ordre où les points sont apparus.

2°) Calculer le nombre de résultats possibles qui totalisent 11, puis 12.

---

SOLUTION

1°) Si l'on tient compte de l'ordre, le résultat (1, 3, 4) est différent de (3, 1, 4). On a donc :

$S = \{1, 2, 3, 4, 5, 6\}^3$

2°) Notons $G_i$ l'ensemble des résultats qui totalisent i points.

$G_{11}$ est formé des résultats :

| | | | | |
|---|---|---|---|---|
| (6,4,1) | (6,3,2) | (5,4,2) | (5,5,1) | (5,3,3) |
| (6,1,4) | (6,2,3) | (5,2,4) | (5,1,5) | (3,5,3) |
| (1,6,4) | (3,2,6) | (4,2,5) | (1,5,5) | (3,3,5) |
| (1,4,6) | (3,6,2) | (4,5,2) | (4,4,3) | |
| (4,6,1) | (2,3,6) | (2,5,4) | (4,3,4) | |
| (4,1,6) | (2,6,3) | (2,4,5) | (3,4,4) | |

Le nombre de manières de gagner avec 11 points est donc

$Card\ G_{11} = 27$

$G_{12}$ est formé des résultats :

| | | | | |
|---|---|---|---|---|
| (6,5,1) | (6,4,2) | (5,4,3) | (5,5,2) | (4,4,4) |
| (6,1,5) | (6,2,4) | (5,3,4) | (5,2,5) | |
| (5,6,1) | (4,6,2) | (4,5,3) | (2,5,5) | |
| (5,1,6) | (4,2,6) | (4,3,5) | (6,3,3) | |
| (1,6,5) | (2,4,6) | (3,4,5) | (3,6,3) | |
| (1,5,6) | (2,6,4) | (3,5,4) | (3,3,6) | |

Le nombre de manières de marquer 12 points est donc

Card $G_{12}$ = 25. ■

---

V. Soit S un ensemble fondamental.

Si A et B sont deux événements, $A \cup B$ est réalisé lorsque A seul est réalisé, B seul est réalisé, ou encore lorsque A et B sont simultanément réalisés (cf. 5 A III).

On appellera $A \Delta B$ l'événement qui n'est réalisé que lorsque A seul est réalisé ou bien B seul est réalisé.

1°) Montrer que $A \Delta B = [A \cap \overline{B}] \cup [B \cap \overline{A}]$

2°) Montrer que $A \Delta B = [A \cup B] \cap [\overline{A \cap B}]$

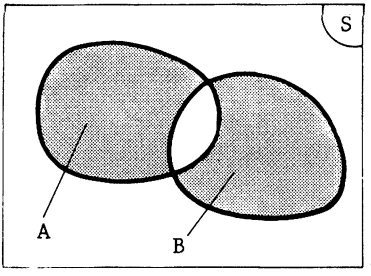
---

SOLUTION

1°) $A \cap \overline{B}$ est l'événement "A est réalisé et B n'est pas réalisé". $\overline{A} \cap B$ est l'événement "B est réalisé et A n'est pas réalisé".

On a donc $A \Delta B = [A \cap \overline{B}] \cup [B \cap \overline{A}]$

2°) $[A \cup B] \cap [\overline{A \cap B}]$ est l'événement : "A ou B sont réalisés, mais A et B ne sont pas simultanément réalisés". C'est donc $A \Delta B$.




▒ $A \Delta B$ ■

I. On considère un lot de 35 boules identiques, numérotées de 1 à 35. Quelle est la probabilité de tirer :

1°) une boule portant un numéro pair ?

2°) une boule portant un numéro impair ?

3°) une boule portant un numéro strictement supérieur à 5 ?

---

SOLUTION

S est l'ensemble des 17 numéros pairs et des 18 numéros impairs. Si on considère les événements :

P : "on a tiré un numéro pair"

I : "on a tiré un numéro impair"

1°) $P(P) = \frac{17}{35}$

2°) $P(I) = \frac{18}{35}$

On a aussi $P(I) = 1 - \frac{17}{35} = \frac{18}{35}$

3°) En appelant x le numéro sorti, on a

$$P(x > 5) = \sum_{i=6}^{35} P(i)$$

$$= 1 - P(x \leq 5)$$

$$= 1 - [P(1) + P(2) + P(3) + P(4) + P(5)]$$

Les boules étant identiques, les P (i) sont toutes égales à $\frac{1}{35}$. Par conséquent

$$P(x > 5) = 1 - 5 \times \frac{1}{35} = \frac{6}{7}$$

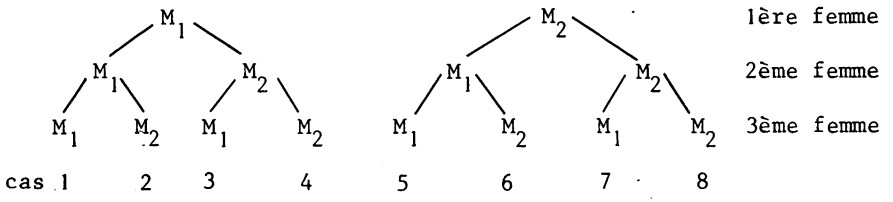
■

II. Dans une même clinique, trois femmes sont sur le point d'accoucher. Deux médecins sont attachés à cette clinique.
1°) Quelle est la probabilité pour que les trois femmes demandent au hasard le même médecin, en même temps ?
2°) Quelle est la probabilité pour que les deux médecins soient appelés ?

---

SOLUTION

Soient $M_1$ et $M_2$ les 2 médecins, chaque femme ayant deux choix possibles, le nombre de cas total est $\alpha_2^3 = 2^3 = 8$




Soient les événements :
1 : le même médecin est demandé
2 : les 2 médecins sont demandés.
1°) Le nombre de cas favorables à l'appel du même médecin est 2 (cas 1 et cas 8) d'où $P(1) = \frac{2}{8} = \frac{1}{4}$.
2°) Deux façons de raisonner
a) $P(2) = 1 - P(1) = \frac{3}{4}$
b) le nombre de cas favorables à l'appel des 2 médecins est 6, d'où $P(2) = \frac{6}{8} = \frac{3}{4}$. ■

---

III. Lors des soldes de fin de série, un fabricant de chemises met en vrac sur une table 200 chemises pratiquement identiques.

Il y a, dans ce lot, des chemises avec 1 ou 2 défauts, ainsi que 100 chemises parfaites. Ces défauts, mineurs, ne sont pas visibles à la présentation.

1°) Quels doivent être les nombres de chemises de chaque catégorie si l'on veut que le premier client, qui prend au hasard une de ces chemises, ait 20 % de chance d'avoir une chemise avec 1 défaut ?

2°) Le client s'aperçoit du défaut. Il remet la chemise dans le tas, sans prendre soin de l'écarter ; quelle est la probabilité qu'il a de prendre une chemise avec 2 défauts ?

---

SOLUTION

On appelle x l'événement "chemise avec x défauts".

1°) Si $N_o$, $N_1$ et $N_2$ sont respectivement les nombres de chemises à 0, 1 et 2 défauts, on doit avoir

$$N_o + N_1 + N_2 = 200 \qquad \text{avec } N_o = 100$$

Les chemises étant apparemment identiques, la probabilité d'avoir une chemise avec 1 défaut est :

$$P\ (x = 1) = \frac{N_1}{200} = 0{,}20$$

On en déduit :

$$N_1 = 0{,}20 \times 200 = 40 \text{ chemises à 1 défaut}$$

Le nombre de chemises avec 2 défauts est donc :

$$N_2 = 200 - (N_o + N_1) = 200 - (100 + 40) = 60 \text{ chemises.}$$

2°) La probabilité de tirer une chemise avec 2 défauts, sachant que le client ne sait plus où se trouve la chemise qu'il avait prise puis remise dans le tas, est :

$$P\ (x = 2) = \frac{N_2}{200} = \frac{60}{200} = 0{,}30$$

On vérifie que $P\ (x = 0) + P\ (x = 1) + P\ (x = 2) = 1$ ■

## IV. LE "PARADOXE" DU CHEVALIER DE MÉRÉ

Le jeu de passe-dix consiste à jeter trois dés, on gagne si la somme des points obtenus dépasse dix (cf. exercice 5 A IV).

Le chevalier de Méré constatait qu'en pratique on gagnait plus souvent avec 11 qu'avec 12. Cela lui semblait paradoxal, car son raisonnement infirmait sa constatation. Voici son raisonnement :

- il y a 6 possibilités de marquer 11 points :
  {6,4,1} ; {6,3,2} ; {5,5,1} ; {5,4,2} ; {5,3,3} ; {4,4,3}
- il y a 6 possibilités de marquer 12 points :
  {6,5,1} ; {6,4,2} ; {6,3,3} ; {5,5,2} ; {5,4,3} ; {4,4,4}

Les probabilités de marquer 11 ou 12 points sont donc égales.

Que penser de ce raisonnement ?

---

SOLUTION

L'erreur du chevalier de Méré est de ne pas distinguer les dés. Il y a en effet une seule manière d'obtenir {4,4,4}, mais six manières d'obtenir 5, 4 et 3 : (5,4,3) ; (5,3,4) ; (3,4,5) ; (3,5,4) ; (4,5,3) et (4,3,5).

On en déduit qu'il y a au total 25 manières de marquer 12 points et 27 manières de marquer 11 points (cf. exercice 5 A IV).

Soit x le nombre de points marqués. On a :

$$P\ (x = 12) = \frac{25}{\alpha_6^3} = \frac{25}{6^3} \simeq 0{,}1157$$

$$P\ (x = 11) = \frac{27}{\alpha_6^3} = \frac{27}{6^3} = 0{,}1250 > P\ (x = 12)$$

Ce raisonnement juste fut trouvé par B. Pascal à qui Méré avait présenté son soi-disant paradoxe. ■

V. Considérons l'épreuve suivante :

- on choisit au hasard un nombre n compris entre 1 et 3 ;
- on choisit ensuite au hasard, n nombres compris entre 0 et 9, mais si le premier de ces nombres est 0, on ne choisit pas d'autre nombre.

1°) Décrire l'ensemble S des résultats possibles. Si N est un nombre, formé de 1, 2 ou 3 chiffres, soit $A_N$ l'événement : "on a obtenu le nombre N".

2°) Calculer les probabilités des événements suivants :

a) $A_o$ ; b) $A_{12}$ ; c) $A_{11}$ ; d) le nombre obtenu est formé d'un seul chiffre ou de chiffres tous différents.

---

SOLUTION

1°) On peut former tous les nombres de 1, 2 ou 3 chiffres. Donc $S = \{0, 1, 2, 3, \ldots, 999\}$.

2°) Remarquons que la probabilité sur S n'est pas uniforme.

a) Il y a en effet 3 manières de former le nombre 0 :

- tirer n = 1 puis tirer 0 parmi $\{0,1,\ldots,9\}$ : A
- tirer n = 2 puis tirer 0 parmi $\{0,1,\ldots,9\}$ : B
- tirer n = 3 puis tirer 0 parmi $\{0,1,\ldots,9\}$ : C

Il est clair que A, B et C sont deux à deux disjoints. Donc :

$$P(A_o) = P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

et $P(A) = P(B) = P(C) = \frac{1}{3} \times \frac{1}{10}$, donc $P(A_o) = \frac{1}{10}$

b) Par contre il n'y a qu'une manière de former un nombre N tel que $1 \leqslant N \leqslant 9$ :

- tirer n = 1 puis choisir N dans $\{1,\ldots,9\}$

Donc $P(A_N) = \frac{1}{3} \times \frac{1}{10} = \frac{1}{30}$ .

Pour former un nombre de 2 chiffres, il faut tirer n = 2 puis choisir un premier chiffre dans $\{1,\ldots,9\}$, et un second dans $\{0,1,\ldots,9\}$. Donc :

$P(A_{12}) = P(A_{11}) = \frac{1}{3} \times \frac{1}{10} \times \frac{1}{10} = \frac{1}{300}$, d'où b) et c).

d) Considérons les événements :

E : "le nombre obtenu est formé d'un seul chiffre ou de plusieurs chiffres tous différents"

A : "le nombre obtenu est formé d'un seul chiffre"

B : "le nombre obtenu est formé de 2 chiffres différents"

C : "le nombre obtenu est formé de 3 chiffres différents".

Il est clair que : $A \cap B = A \cap C = \phi$ et que $(A \cup B) \cap C = \phi$

$$E = A \cup B \cup C$$

D'après le second axiome de définition d'une probabilité :

$$P(E) = P(A) + P(B) + P(C)$$

car $P(E) = P[(A \cup B) \cup C] = P(A \cup B) + P(C)$ car $(A \cup B) \cap C = \phi$

$$= P(A) + P(B) + P(C) \text{ car } A \cap B = \phi$$

$$P(A) = P(0) + P(1) + P(2) + \ldots + P(9)$$

$$= \frac{1}{3}$$

$$P(B) = \frac{1}{3} \times \frac{9}{10} \times \frac{9}{10} = \frac{27}{100}$$

$$P(C) = \frac{1}{3} \times \frac{9}{10} \times \frac{9}{10} \times \frac{8}{10} = \frac{216}{1000}$$

Donc $P(E) = \frac{1}{3} + \frac{27}{100} + \frac{216}{1000} \simeq 0{,}819$ ■

## VI. PROBABILITES GEOMETRIQUES : LE PROBLEME DE BUFFON

*Supposons que l'ensemble des résultats possibles d'une* épreuve $\mathscr{A}$ soit une partie P du plan de surface finie S. Un événement E associé à l'épreuve $\mathscr{A}$ sera une partie de P. Si l'on suppose que la probabilité de E est proportionnelle à sa surface, on dira qu'on a une probabilité géométrique.

Supposons que le plan est recouvert de droites parallèles équidistantes, deux droites successives étant à une distance d (d > 0). On jette au hasard sur ce plan une aiguille de longueur $\ell$ ($\ell > 0$), avec $\ell < d$.

Quelle est la probabilité que l'aiguille rencontre une des droites parallèles ?

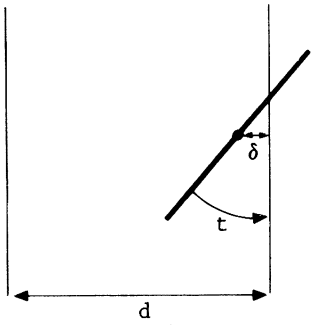
---

SOLUTION

La position de l'aiguille est entièrement déterminée par deux nombres :

- la distance $\delta$ du centre de l'aiguille à la parallèle la plus proche ;
- l'angle t que fait l'aiguille avec cette parallèle,

et on a : $\delta \in [0, \frac{d}{2}]$ et $t \in [0, \pi]$ .

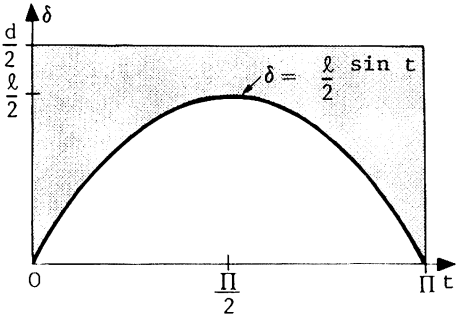



L'ensemble des couples $\delta \in [0, \frac{d}{2}]$ , $t \in [0, \pi]$ est le rectangle de la figure ci-dessous. L'aiguille rencontre une des droites lorsque $\delta \leq \frac{\ell}{2} \sin t$. L'ensemble des couples $(\delta, t)$ qui réalisent l'événement "l'aiguille rencontre une parallèle" est la partie blanche de la figure ci-dessous.

La probabilité cherchée p est donc :

$$p = \frac{1}{\frac{d}{2}\pi} \int_0^{\pi} \frac{\ell}{2} \sin t \, dt$$

soit $p = \frac{2\ell}{\pi d}$ .




[ Le calcul de la probabilité a été possible parce que la fonction sinus est intégrable sur le segment $[0, \pi]$, ou ce qui revient au même parce que la partie blanche de la figure ci-dessus est intégrable.

D'une manière générale, si l'ensemble fondamental S est continu, il sera nécessaire de restreindre la probabilité à une partie de $\mathcal{P}(S)$ (cf. B. Vauquois, Probabilités, Hermann, Paris 1978) ]. ■

I. Soit P une probabilité sur un ensemble S fini ou dénombrable. Montrer les propriétés suivantes :

1°) $\forall A \in \mathcal{P}(S), \forall B \in \mathcal{P}(S) \quad P(A \cap \overline{B}) = P(A) - P(A \cap B)$

2°) $\forall A \in \mathcal{P}(S), \forall B \in \mathcal{P}(S) \quad P(\overline{A} \cup \overline{B}) = 1 - P(A \cap B)$

3°) $\forall A \in \mathcal{P}(S), \forall B \in \mathcal{P}(S) \quad P(A \Delta B) = P(A \cup B) - P(A \cap B)$

($A \Delta B$ est défini dans l'exercice 5 A V)

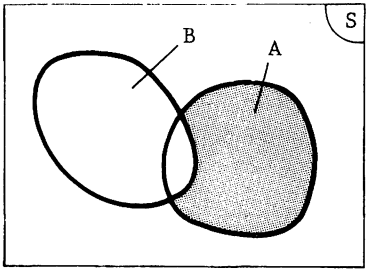
4°) $\forall A \in \mathcal{P}(S), \forall B \in \mathcal{P}(S), \forall C \in \mathcal{P}(S)$

$A \cap B = A \cap C = B \cap C = \phi \Rightarrow P(A \cup B \cup C) = P(A) + P(B) + P(C)$

---

SOLUTION

1°) Montrons que :




$A \cap \overline{B}$

$[A \cap \overline{B}] \cup [A \cap B] = A$

$[A \cap \overline{B}] \cap [A \cap B] = \phi$

$[A \cap \overline{B}] \cup [A \cap B] = A \cap [B \cup \overline{B}] = A$

$[A \cap \overline{B}] \cap [A \cap B] = A \cap B \cap \overline{B} = \phi$ car $B \cap \overline{B} = \phi$

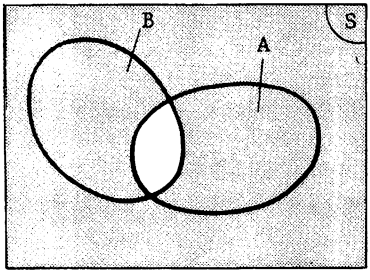
D'après le second axiome de définition d'une probabilité $P(A) = P(A \cap \overline{B}) + P(A \cap B)$, d'où le résultat.

2°) On a : $\overline{A} \cup \overline{B} = \overline{A \cap B}$

On en déduit que :

$P(\overline{A} \cup \overline{B}) = 1 - P(A \cap B)$

(cf. 5 B III, Conditions de normalisation)




$\overline{A} \cup \overline{B}$

3°) D'après l'exercice 5 A V, $A \Delta B = [A \cup B] \cap [\overline{A \cap B}]$

On en déduit que :

$A \cup B = [A \Delta B] \cup [A \cap B]$

et $[A \Delta B] \cap [A \cap B] = \phi$

D'après le second axiome de définition d'une probabilité, on a :

$P(A \cup B) = P(A \Delta B) + P(A \cap B)$, d'où le résultat.

4°) D'après 5 C II) on a :

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

Si $A \cap B = A \cap C = B \cap C = \phi$ alors $A \cap B \cap C = \phi$ et

$P(A \cap B) = P(A \cap C) = P(B \cap C) = P(A \cap B \cap C) = 0$

d'où le résultat. ∎

---

II. GENERALISATION DU THEOREME DES PROBABILITES TOTALES

Soit P une probabilité sur un ensemble S fini et dénombrable. Soient $A_1, A_2, \ldots, A_n$ n événements.

Démontrer par récurrence sur n que :

$$P\left[\bigcup_{i=1}^{n} A_i\right] = \sum_{i=1}^{n} P(A_i) - \sum_{i} \sum_{j \neq i} P(A_i \cap A_j) + \sum_{i} \sum_{j \neq i} \sum_{\substack{k \neq i \\ k \neq j}} P(A_i \cap A_j \cap A_k) + \ldots + (-1)^{n+1} P\left[\bigcap_{i=1}^{n} A_i\right]$$

---

SOLUTION

. Pour n = 2, on a $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$ ce qui n'est rien d'autre que le théorème des probabilités totales.

. Supposons la relation vraie pour n = p et montrons qu'elle est vraie pour n = p+1

$$P\left[\bigcup_{i=1}^{p+1} A_i\right] = P\left[\left(\bigcup_{i=1}^{p} A_i\right) \cup A_{p+1}\right]$$

$$= P\left[\bigcup_{i=1}^{p} A_i\right] + P(A_{p+1}) - P\left[\left(\bigcup_{i=1}^{p} A_i\right) \cap A_{p+1}\right]$$

d'après le théorème des probabilités totales.

Or on a : $\left[\left(\bigcup_{i=1}^{p} A_i\right) \cap A_{p+1}\right] = \bigcup_{i=1}^{p} (A_i \cap A_{p+1})$

D'après l'hypothèse de récurrence :

$$P\left[\bigcup_{i=1}^{p} (A_i \cap A_{p+1})\right] = \sum_{i=1}^{p} P(A_i \cap A_{p+1})$$

$$- \sum_{i} \sum_{j \neq i} P(A_i \cap A_j \cap A_{p+1}) + \ldots$$

$$+ (-1)^{p+1} P\left[\bigcap_{i=1}^{p+1} A_i\right]$$

On a donc :

$$P\left[\bigcup_{i=1}^{p+1} A_i\right] = \sum_{i=1}^{p+1} P(A_i) - \sum_{i} \sum_{j \neq i} P(A_i \cap A_j)$$

$$+ \ldots + (-1)^{p+2} \left[\bigcap_{i=1}^{p+1} A_i\right]$$

■

---

III. Soit P une probabilité sur un ensemble S fini ou dénombrable.

1°) Montrer que si $A \in \mathscr{P}(S)$, $B \in \mathscr{P}(S)$, alors

$P(A \cup B) \leqslant P(A) + P(B)$.

2°) En déduire que si $A_1$, $A_2$, ..., $A_n$ sont n éléments de $\mathcal{P}(S)$, alors :

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

SOLUTION

1°) résulte du théorème des probabilités totales. On a :

$P(A \cap B) \geqslant 0$ et $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Donc $P(A \cup B) \leqslant P(A) + P(B)$

2°) se déduit de 1°) par récurrence sur n. On a vu à la question précédente que la propriété est vraie pour n = 2. Supposons qu'elle est vraie pour n=p, et montrons qu'elle est alors vraie pour n=p+1.

D'après la question 1°) :

$$P\left[\bigcup_{i=1}^{p+1} A_i\right] \leqslant P\left[\bigcup_{i=1}^{p} A_i\right] + P(A_{p+1})$$

Or par hypothèse :

$$P\left[\bigcup_{i=1}^{p} A_i\right] \leqslant \sum_{i=1}^{p} P(A_i)$$

On a donc :

$$P\left[\bigcup_{i=1}^{p+1} A_i\right] \leqslant \sum_{i=1}^{p+1} P(A_i).$$

■

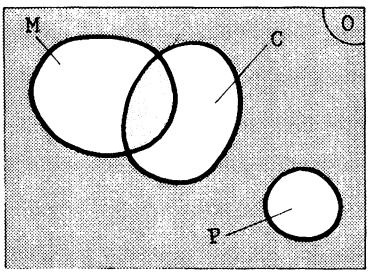
IV. Dans une entreprise de construction, parmi un effectif de 80 ouvriers (ensemble O), 15 sont maçons-carreleurs, 23 sont maçons seulement, 7 sont carreleurs et non maçons, 5 sont plombiers seulement. Quelle est la probabilité pour qu'un ouvrier de cette entreprise soit :

1°) maçon ou plombier ?

2°) maçon ou carreleur ?

SOLUTION

On considère les événements :




M : "l'ouvrier est maçon"

C : "l'ouvrier est carreleur"

P : "l'ouvrier est plombier"

$\square$ $M \cap C$

Card $(M \cap C) = 15$ Card $(C) = 7 + 15 = 22$

Card $M = 23 + 15 = 38$ Card $(P) = 5$

1°) Puisque $M \cap P = \phi$ on a :

$$P(M \cup P) = P(M) + P(P) = \frac{38}{80} + \frac{5}{80} = \frac{43}{80}$$

$P(M \cup P) \simeq 0{,}537$

2°) $P(M \cup C) = P(M) + P(C) - P(M \cap C) = \frac{38}{80} + \frac{22}{80} - \frac{15}{80} = \frac{45}{80}$

$P(M \cup C) \simeq 0{,}562$. ■

V.

1°) Soient A et B deux sous-ensembles d'un ensemble fini E avec a = Card A, b = Card B et c = Card $A \cap B$. Quel est le Cardinal de $A \cup B$ ?

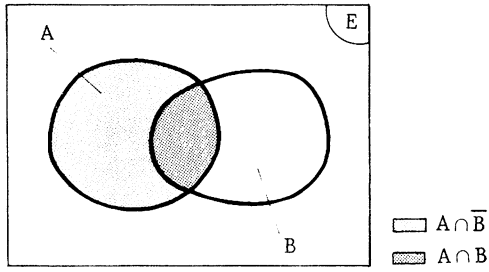
2°) On se servira de ce résultat pour résoudre le problème suivant : à toute personne achetant un de ses produits, un vendeur donne à choisir une enveloppe renfermant un ticket numéroté.

Si ce ticket est bleu ou s'il porte un numéro pair, l'acheteur se voit attribuer un cadeau.

Le vendeur veut que le premier client ait une chance sur deux de gagner. Il fait numéroter, en nombre égal, de 1 à 5, 1 000 tickets dont 10 bleus portent un numéro pair. Combien y aura-t-il d'enveloppes contenant un ticket bleu ?

SOLUTION

1°) Le résultat est mis en évidence dans un diagramme de Venn




$$\text{Card } (A \cup B) = \text{Card } A + \text{Card } B - \text{Card } (A \cap B)$$

$$\text{Card } (A \cup B) = a + b - c$$

2°) Soient A l'ensemble des nombres pairs et B l'ensemble des tickets bleus ; Card $(A \cup B)$ représente le nombre de tickets bleus ou pairs. C'est donc le nombre de cas favorables.

Par conséquent, la probabilité P de gagner, pour le premier client est :

$$P = \frac{\text{Card } (A \cup B)}{1\,000} = 0{,}5$$

d'où

$$\text{Card } (A \cup B) = 500.$$

Les numéros étant répartis uniformément, il y a 200 tickets de chaque numéro. Or, entre 1 et 5, il y a 2 nombres pairs (2 et 4), donc Card (A) = a = 400.

Puisque Card $(A \cap B) = c = 10$ et que Card $(A \cup B) = 500 = a + b - c$, soit $500 = 400 + b - 10$, on en déduit $b = 500 - 400 + 10 = 110$.

Il y a donc 110 enveloppes contenant un ticket bleu. ■

VI. Dans un échantillon de 1 000 patients, on relève 300 personnes malades des poumons (événement P), 600 personnes malades du coeur (événement C) et 200 individus souffrant d'hypertension (événement H).

1°) Calculer Card $(H \cap C)$, sachant que 76 % des patients souffrent d'hypertension ou de maladies cardiaques.

2°) Sachant que Card $(P \cap C) = 60$ et que Card $(P \cap C \cap H) = 0$, calculer Card $(P \cap H)$.

3°) Quelle est la probabilité de trouver un patient souffrant d'hypertension ou d'une maladie pulmonaire ?

---

SOLUTION

1°) $P(H \cup C) = P(H) + P(C) - P(H \cap C) = 0{,}76$

$$P(H) = \frac{200}{1\,000} = 0{,}2$$

$$P(C) = \frac{600}{1\,000} = 0{,}6$$

On en déduit

$$P(H \cap C) = 0{,}8 - 0{,}76 = 0{,}04$$

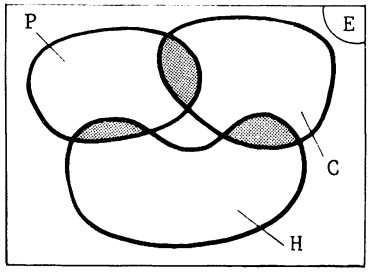
et comme

$$P(H \cap C) = \frac{\text{Card}(H \cap C)}{1\,000}$$

$$\text{Card}(H \cap C) = 40.$$

Il y a donc 40 personnes atteintes d'hypertention et maladies du coeur.

2°)



$$\text{Card}(E) = \text{Card}(P) + \text{Card}(C) + \text{Card}(H) - \text{Card}(P \cap C) - \text{Card}(P \cap H) - \text{Card}(H \cap C)$$

$$1\,000 = 300 + 600 + 200 - 60 - \text{Card}(P \cap H) - 40$$

On en déduit

$$\text{Card}(P \cap H) = 0.$$

3°) $P(P \cup H) = P(P) + P(H) - P(P \cap H) = \frac{300}{1\,000} + \frac{200}{1\,000} = 0{,}5.$

■

I. Soient A et B deux événements indépendants.
Montrer que A et $\overline{B}$ sont indépendants.

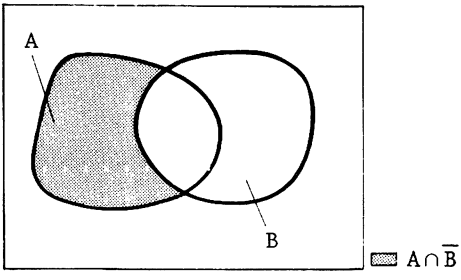
---

SOLUTION

On a à montrer que $P(A \cap \overline{B}) = P(A)\, P(\overline{B})$

Or $A = [A \cap \overline{B}] \cup [A \cap B]$

$[A \cap \overline{B}] \cap [A \cap B] = \phi$




D'après le second axiome de définition d'une probabilité

$$P(A) = P(A \cap \overline{B}) + P(A \cap B)$$

donc $P(A \cap \overline{B}) = P(A) - P(A \cap B) = P(A) - P(A)\, P(B)$

car A et B sont indépendants, d'où :

$$P(A \cap \overline{B}) = P(A)\,[1 - P(B)] = P(A)\, P(\overline{B})$$

(condition de normalisation).

---

II. Considérons une urne contenant $M_1$ boules noires et $M_2$ boules blanches. On tire n boules, avec remise, au hasard.

1°) Préciser un ensemble fondamental S associé à cette expérience, ainsi que la probabilité P sur cet ensemble.

2°) Soit $T_i$ l'événement : "la première boule noire apparaît au $i^{ème}$ tirage". Calculer P $(T_i)$ : pour $1 \leqslant i \leqslant n$.

3°) Soit $E_i$ l'événement : "on obtient i boules blanches lors des n tirages". Calculer P $(E_i)$ pour $1 \leqslant i \leqslant n$.

4°) Déterminer la probabilité d'obtenir au moins une boule noire.

---

SOLUTION

1°) S = {N, B} (N pour noire, B pour blanche)

$$P\ (N) = \frac{M_1}{M}\ ,\ P\ (B) = \frac{M_2}{M} \quad \text{où } M = M_1 + M_2$$

2°) Les tirages se faisant avec remise, les événements "tirer une boule noire au $i^{ème}$ tirage" et "tirer une boule noire au $j^{ème}$ tirage" sont indépendants si $i \neq j$. On a donc :

$$P\ (T_i) = [P\ (B)]^{i-1} \times P\ (N) = \left[\frac{M_2}{M}\right]^{i-1} \times \frac{M_1}{M}$$

3°) de même : $P\ (E_i) = [P\ (B)]^{i} = \left(\frac{M_2}{M}\right)^{i} \times \left(\frac{M_1}{M}\right)^{n-i}$

4°) Soit F l'événement : "on obtient au moins une boule noire". Alors $\overline{F} = E_n$ et d'après la condition de normalisation

$$P\ (F) = 1 - P\ (E_n) = 1 - \left[\frac{M_2}{M}\right]^{n}$$

■

---

III. Soit P une probabilité sur un ensemble S fini ou dénombrable. Si $A_1, \ldots, A_n$ sont n événements, montrer par récurrence sur n que :

$$P\left[\bigcap_{i=1}^{n} A_i\right] = P\ (A_1)\ P\ (A_2/A_1)\ P\ (A_3/A_1 \cap A_2) \ldots P\ (A_n/\bigcap_{i=1}^{n-1} A_i)$$

---

SOLUTION

Si n=2 on a, par définition d'une probabilité composée :

$$P(A_1 \cap A_2) = P(A_1)\ P(A_2/A_1)$$

Supposons la relation vraie pour n=p. Alors par définition d'une probabilité composée

$$P\left(\bigcap_{i=1}^{p+1} A_i\right) = P\left[\left(\bigcap_{i=1}^{p} A_i\right) \cap A_{p+1}\right] = P\left(\bigcap_{i=1}^{p} A_i\right) P\left(A_{p+1} / \bigcap_{i=1}^{p} A_i\right)$$

d'où le résultat. ■

---

IV. Quelle est la probabilité de sortir 421 en lançant 3 dés simultanément ? (jeu du 421)

---

SOLUTION

1ère méthode : Probabilité composée

$$p\ (4 \text{ et } 2 \text{ et } 1 \text{ dans l'ordre}) = p(4) \times p(2) \times p(1) = \left(\frac{1}{6}\right)^3$$

Les 3 dés étant jetés simultanément, l'ordre n'intervient pas, et il y a 3! façons équiprobables d'avoir 4,2,1, d'où :

$$p\ (4 \text{ et } 2 \text{ et } 1 \text{ sans ordre}) = 3!\ \left(\frac{1}{6}\right)^3 = \frac{1}{36} = 0{,}029.$$

2ème méthode : Dénombrement

Le nombre de cas possibles est $\alpha_6^3 = 6^3$

Le nombre de cas favorables est $P_3 = 3!$

d'où $p = \dfrac{3!}{6^3} = \dfrac{1}{36} = 0{,}029$ ■

---

V. Trois dés sont truqués de sorte que, pour chaque dé, la probabilité de sortir 1 as est 2 fois plus grande que celle de sortir n'importe quel autre numéro. Quelle est la probabilité de sortir, sur un lancer des 3 dés, 421 ?

---

SOLUTION

$$P(As) = 2P(2) = 2P(3) = 2P(4) = 2P(5) = 2P(6)$$

avec la condition :

$$P(As) + P(2) + P(3) + P(4) + P(5) + P(6) = 1$$

ce qui entraîne

$$P(As) + 5\frac{P(As)}{2} = 1 = \frac{7}{2}P(As) \text{ d'où } P(As) = \frac{2}{7}$$

La probabilité de sortir 421 (probabilités indépendantes)

$$P(4 \text{ et } 2 \text{ et } 1 \text{ sans ordre}) = 3!\, P(4) \times P(2) \times P(1)$$

$$= 3! \times \frac{1}{7} \times \frac{1}{7} \times \frac{2}{7}$$

$$P(4,2,1) = \frac{12}{343} = 0{,}035. \quad \blacksquare$$

---

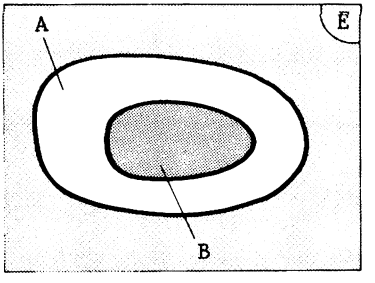
VI. Parmi 1 000 moteurs d'une certaine fabrication, 725 moteurs ont fonctionné sans problème pendant les 3 premières années et 375 les 5 premières années.

Quelle est la probabilité pour qu'un moteur, n'ayant pas eu de problème pendant les 3 premières années, fonctionne encore pendant 2 ans ?

---

SOLUTION

1ère méthode : calcul direct




Soit E l'ensemble des moteurs et considérons les événements :

A : "moteurs ayant fonctionné 3 ans sans problème"

B : "moteurs ayant fonctionné 5 ans sans problème".

Card (E) = 1 000 ; Card (A) = 725 ; Card (B) = 375

= Card (A ∩ B)

Nombre de cas favorables : 375

Nombre de cas possibles : 725

D'où

$$P = \frac{375}{725} = 0,517.$$

2ème méthode :

La probabilité pour qu'un moteur fonctionne 3 ans sans problème est :

$$P(A) = \frac{725}{1\ 000} = 0,725$$

La probabilité pour qu'un moteur fonctionne 5 ans est :

$$P(A \cap B) = \frac{375}{1\ 000} = 0,375$$

Par conséquent, la probabilité pour qu'un moteur n'ayant pas eu de problème pendant les 3 premières années, fonctionne encore 2 ans est :

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{0,375}{0,725} = 0,517.$$ ■

---

VII. On estime à 15 % le nombre de vacanciers français qui choisissent de sortir de France. Parmi ceux-ci, 35 % vont en Espagne et 25 % vont en Italie.

Quelle est la probabilité pour qu'un Français prenne ses vacances :

1°) en Italie ?

2°) en Espagne ?

---

SOLUTION

1°) La probabilité p pour qu'un Français passe ses vacances en Italie est :

$$p = \frac{15}{100} \times \frac{25}{100} = 0{,}0375$$

soit 3,75 % de vacanciers français vont en Italie.

2°) La probabilité q pour qu'un Français passe ses vacances en Espagne est :

$$q = \frac{15}{100} \times \frac{35}{100} = 0{,}0525$$

soit 5,25 % de vacanciers français vont en Espagne. ■

---

VIII. Un révolver à 6 coups contient une balle dans le barillet. On fait tourner le barillet à chaque fois avant de tirer sur une cible qu'on ne peut manquer (Principe du jeu de la roulette russe).

1°) Quelle est la probabilité de toucher la cible au premier essai ?

2°) Quelle est la probabilité de ne pas toucher la cible au bout de N essais ?

3°) Quelle est la probabilité de toucher la cible au $N^{ème}$ essai ?

---

SOLUTION

1°) $p = \frac{n}{N} = \frac{1}{6}$

2°) La probabilité de ne pas toucher la cible après 1 essai est :

$$q = 1-p = 1 - \frac{1}{6} = \frac{5}{6}$$

d'où la probabilité cherchée

$$P = q^N = (\frac{5}{6})^N$$

3°) Les N-1 premiers essais ayant échoué, le $N^{ème}$ devant réussir,

$$P = (\frac{5}{6})^{N-1} \times \frac{1}{6}$$ ■

---

IX. Trois usines A, B et C produisent respectivement 50 %, 30 % et 20 % des moteurs de voitures. Parmi la production de chacune de ces usines, 5 %, 3 % et 2 % des moteurs fabriqués sont défectueux. Calculer la probabilité pour qu'un moteur défectueux provienne de l'usine A.

---

SOLUTION

On considère les événements :

A : le moteur vient de l'usine A (idem pour B et C)

D : le moteur fabriqué est défectueux

$\overline{D}$ : le moteur fabriqué n'est pas défectueux.

1ère méthode : Théorème de Bayes

$$P(A/D) = \frac{P(A) \times P(D/A)}{P(A) \times P(D/A) + P(B) \times P(D/B) + P(C) \times P(D/C)}$$

avec

$$P(A) = 0,50 \quad ; \quad P(B) = 0,30 \quad ; \quad P(C) = 0,20$$

$$P(D/A) = 0,05 \; ; \; P(D/B) = 0,03 \; ; \; P(D/C) = 0,02$$

d'où

$$P(D/A) = \frac{0,50 \times 0,05}{0,50 \times 0,05 + 0,30 \times 0,03 + 0,2 \times 0,02} = \frac{0,025}{0,038} = \frac{2,5}{3,8} \simeq 0,658$$

2ème méthode :

sur 100 moteurs
- 50 viennent de A
  - 2,5 sont défectueux : D
  - 47,5 ne le sont pas : $\overline{D}$
- 30 de B
  - 0,9 D
  - 29,1 $\overline{D}$
- 20 de C
  - 0,4 D
  - 19,6 $\overline{D}$

$$P(A/D) = \frac{\text{Nbre de moteurs défectueux venant de A}}{\text{Nbre total de moteurs défectueux}}$$

$$= \frac{2,5}{2,5 + 0,9 + 0,4} = \frac{2,5}{3,8} = 0,658$$ ■

---

X. Une partie des accidents scolaires est due à des accidents de laboratoires. 25 % des étudiants ne lisent pas les notices de mise en garde qui accompagnent les produits qu'ils manipulent. Parmi ceux qui lisent, 10 % ont tout de même des accidents, par manque de précaution.

Quelle est, pour un étudiant qui ne lit pas la notice, la probabilité d'avoir un accident si la probabilité pour qu'un accidenté n'ait pas lu la notice est de 0,7273 ?

---

SOLUTION

On considère les événements :

A : "l'étudiant a un accident"

$\overline{A}$ : "l'étudiant n'a pas d'accident"

L : "l'étudiant a lu la notice"

$\overline{L}$ : "l'étudiant n'a pas lu la notice".

1ère méthode : Théorème de Bayes

$$P(\overline{L}/A) = \frac{P(\overline{L}) \times P(A/\overline{L})}{P(\overline{L}) \times P(A/\overline{L}) + P(L) \times P(A/L)}$$

On doit calculer $P(A/\overline{L})$ : probabilité d'avoir un accident pour un étudiant qui n'a pas lu la notice

$$0,7273 = \frac{0,25 \times P(A/\overline{L})}{0,25 \times P(A/\overline{L}) + 0,75 \times 0,10}$$

On en déduit :

$$P(A/\overline{L}) = \frac{0,75 \times 0,10 \times 0,7273}{0,25 - 0,25 \times 0,7273} = 0,80.$$

2ème méthode :

sur 100 étudiants
- 75 lisent la notice (L)
  - 7,5 ont un accident (A/L)
  - 67,5 n'ont pas d'accident ($\overline{A}$/L)
- 25 ne la lisent pas ($\overline{L}$)
  - x ont un accident ($A/\overline{L}$)
  - 25-x n'ont pas d'accident ($\overline{A}/\overline{L}$)

$$0,7273 = \frac{x}{x + 7,5} \quad \text{qui donne } x = 20$$

La probabilité cherchée est :

$$P(A/\overline{L}) = \frac{x}{25} = \frac{20}{25} = 0,80.$$

■

---

XI. Une enquête portant sur un plan d'urbanisme donne la répartition suivante des avis des habitants d'une ville comprenant 4 arrondissements. (voir tableau page suivante).

Calculer les probabilités

1°) pour qu'une personne plutôt favorable provienne du quartier 3

2°) pour qu'une personne provenant du quartier 2 soit plutôt défavorable.

| Arrondissements | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Population de l'arrondissement | 15 % | 32 % | 23 % | 30 % |
| Plutôt favorable | 12 % | 31 % | 72 % | 55 % |
| Plutôt défavorable | 60 % | 45 % | 5 % | 31 % |
| Ne se prononcent pas | 28 % | 24 % | 23 % | 14 % |

SOLUTION

On considère les événements

F : "plutôt favorable" D : "plutôt défavorable"

N : "ne se prononce pas" $Q_i$ : "habite le quartier i"

1°) D'après le théorème de Bayes, on a

$$P(Q_3/F) = \frac{P(Q_3)\times P(F/Q_3)}{P(Q_1)\times P(F/Q_1)+P(Q_2)\ P(F/Q_2)+P(Q_3)\ P(F/Q_3)+P(Q_4)\ P(F/Q_4)}$$

ou encore :

$$P(Q_3/F) = \frac{23\%\times 72\%}{[\,15\%\times 12\%] + [\,32\%\times 31\%] + [\,23\%\times 72\%] + [\,30\%\times 55\%]}$$

$$P(Q_3/F) = \frac{23 \times 72}{(15\times 12) + (32\times 31) + (23\times 72) + (30\times 55)}$$

$$P(Q_3/F) = \frac{1656}{4478} \simeq 0{,}34$$

2°) $P(D/Q_2) = 45\ \%$. ■

I. On suppose que la probabilité pour qu'un nouveau-né soit un garçon est de 0,55. Cet événement étant indépendant des individus, quelle est la probabilité pour que, sur 5 nouveaux-nés d'une clinique, il y ait 2 garçons ?

---

SOLUTION

La probabilité élémentaire pour qu'un nouveau-né soit un garçon est :

$$p = 0,55$$

Par suite la probabilité élémentaire pour qu'un nouveau-né soit une fille est :

$$q = 1-p = 0,45.$$

D'où la probabilité pour que, sur 5 nouveaux-nés, il y ait 2 garçons :

$$P(2) = C_5^2 \, p^2 \, q^{5-2}$$

$$P(2) = \frac{5!}{2!\,3!} (0,55)^2 (0,45)^3$$

$$P(2) = 0,28. \quad \blacksquare$$

---

II. Parmi huit équipes de foot-ball dont 6 de première division et 2 de deuxième division, seulement quatre d'entre elles doivent jouer un certain jour. On dispose d'une urne contenant huit tubes à l'intérieur desquels se trouvent les noms des équipes. Un officiel tire 4 tubes au hasard. Quelle est la probabilité d'avoir, parmi les 4 équipes :

1°) 2 équipes de 2ème division P (2) ?

2°) 1 équipe de 2ème division P (1) ?

3°) 0 équipe de 2ème division P (0) ?

4°) Quelle est la relation qui existe entre les 3 probabilités calculées ?

---

SOLUTION

1°) Il y a $C_8^4$ manières possibles de prélever 4 tubes parmi les 8. Le nombre de façons de choisir 2 équipes de 2ème division est $C_2^2$, de même que le nombre de façons de choisir 2 équipes de 1ère division est $C_6^2$. Le nombre de cas favorables est donc $C_2^2 \times C_6^2$. Par conséquent,

$$P\ (2) = \frac{C_2^2 \times C_6^2}{C_8^4} = \frac{1 \times 15}{70} = \frac{3}{14} \simeq 0{,}214$$

2°) Dans ce cas, on doit avoir une équipe de 2ème division parmi 2, et 3 équipes de 1ère division parmi 6, soit

$$P\ (1) = \frac{C_2^1 \times C_6^3}{C_8^4} = \frac{2 \times 20}{70} = \frac{8}{14} \simeq 0{,}572$$

3°) Il faut que les 4 équipes soient de 1ère division, d'où :

$$P\ (0) = \frac{C_2^0 \times C_6^4}{C_8^4} = \frac{1 \times 15}{70} = \frac{3}{14} \simeq 0{,}214$$

4°) On remarque que

$$P\ (0) + P\ (1) + P\ (2) = 1$$

ce qui est normal puisque les trois cas considérés correspondent aux seuls trois cas possibles. ■

---

III. Lors d'un test, on pose à 4 personnes 1 question en leur donnant 3 réponses. Les 4 personnes interrogées n'ont aucune connaissance sur le sujet posé, elles devront donc répondre au hasard.

En distinguant les 3 réponses, quelle est la probabilité pour que :

1°) la même réponse soit donnée par les 4 candidats ?

2°) deux réponses seulement, parmi les 3, soient données ?

3°) en déduire la probabilité pour que les 3 réponses soient énoncées.

---

SOLUTION

Soient $R_1$, $R_2$ et $R_3$ les 3 réponses. Le nombre de configurations possibles est :

$$N = \alpha_3^4 = 3^4 = 81$$

1°) Le nombre de façons d'avoir la même réponse (soit $R_1$, soit $R_2$, soit $R_3$) est :

$$n_1 = C_3^1 = \frac{3!}{2!\ 1!} = 3$$

d'où

$$p_1 = \frac{3}{81} = \frac{1}{27} \simeq 0{,}037$$

2°) Pour que deux des trois réponses soient données, deux possibilités sont à envisager :

a) Deux candidats donnent la même réponse, les 2 autres candidats donnant une autre réponse

( ex : le 1er et le 2ème donnent la réponse $R_1$
le 3ème et le 4ème donnent la réponse $R_2$ )

Le nombre de cas favorables à cette éventualité est :

$$n_2' = C_3^2 \quad \frac{4!}{2!\ 2!}$$

$C_3^2$ correspond au nombre de façons de choisir les couples de réponses et le terme $\frac{4!}{2!\ 2!}$ est égal au nombre de permutations distinctes avec 2 répétitions d'ordre 2, d'où

$$n_2' = C_3^2 \times \frac{4!}{2!\ 2!} = 18$$

b) 3 candidats donnent la même réponse, le 4ème candidat donnant une autre réponse

( ex : le 1er, le 2ème et le 3ème répondent $R_1$
et le 4ème répond $R_2$ )

Le nombre favorable à cette éventualité est :

$$n_2'' = C_3^2 \times \frac{4!}{3!} \times 2 \qquad \text{avec}$$

$C_3^2$ : nombre de façons de choisir 2 réponses parmi les 3 proposées

$\frac{4!}{3!}$ : nombre de permutations avec répétition d'ordre 3

Le facteur 2 exprime le fait que pour un couple de réponses choisi ($R_1$, $R_2$) il y a 2 possibilités suivant que c'est $R_1$ ou $R_2$ qui est donné 3 fois. D'où :

$$n_2'' = C_3^2 \times \frac{4!}{3!} \times 2 = 24$$

Par conséquent, le nombre de cas favorables correspondant à deux réponses énoncées seulement, est :

$$n_2 = n_2' + n_2'' = 18 + 24 = 42$$

d'où

$$p_2 = \frac{42}{81} = \frac{14}{27} = 0{,}519$$

3°) On en déduit que la probabilité pour que les 3 réponses figurent est :

$$p_3 = 1 - (p_1 + p_2) = \frac{12}{27} = 0{,}444.$$

On retrouve, naturellement, le même résultat en remarquant que dans ce cas, une même réponse devra être donnée par 2 candidats Il y a $C_3^1$ façons de choisir cette réponse. Il restera alors à permuter les 4 réponses dont deux sont identiques, d'où

$$n_3 = C_3^1 \times \frac{4!}{2!} = 36 \qquad \text{et} \qquad p_3 = \frac{36}{81} = \frac{12}{27}\,. \qquad \blacksquare$$

IV. Une boîte A contient 12 articles dont 3 sont défectueux ; une autre boîte identique B contient 16 articles dont 5 sont défectueux. On tire d'une boîte choisie au hasard, un article au hasard. Quelle est la probabilité P (E) pour que l'article soit défectueux ?

---

SOLUTION

Diagramme en arbre

Soient les événements suivants :

I : "on tire un objet de la boîte I" (I peut être soit A, soit B)

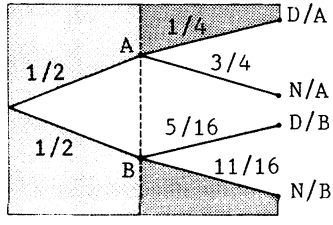
D/I : "l'objet tiré de I est défectueux"

N/I : "l'objet tiré de I n'est pas défectueux"

$$P(A) = P(B) = \frac{1}{2}$$

$$P(D/A) = \frac{3}{12} = \frac{1}{4} \qquad P(N/A) = \frac{9}{12} = \frac{3}{4}$$

$$P(D/B) = \frac{5}{16} \qquad P(N/B) = \frac{11}{16}$$




Par définition des probabilités composées on a :

$$P(A \cap D) = \frac{1}{2} \times \frac{1}{4} = \frac{1}{8}$$

$$P(A \cap N) = \frac{1}{2} \times \frac{3}{4} = \frac{3}{8}$$

$$P(B \cap D) = \frac{1}{2} \times \frac{5}{16} = \frac{5}{32}$$

$$P(N \cap D) = \frac{1}{2} \times \frac{11}{16} = \frac{11}{32}$$

Un article défectueux peut être obtenu soit suivant le pro-

cessus $A \cap D$, soit suivant le processus $B \cap D$. On obtient, les événements étant indépendants

$$P(E) = P[(A \cap D) \cup (B \cap D)] = p(A \cap D) + p(B \cap D)$$
$$= \frac{1}{8} + \frac{5}{32} = \frac{9}{32} .$$

■

# 6. Les lois de probabilité

## A. VARIABLES ALEATOIRES

### I. INTRODUCTION

Il est fréquent que l'on associe une valeur numérique à tout résultat d'une expérience aléatoire. La notion de variable aléatoire est la formalisation mathématique de cette situation.

*Exemple 1*

Considérons une expérience aléatoire $\mathscr{A}$ consistant à jeter n pièces. Un résultat de cette épreuve peut être représenté par une suite de n termes, les uns égaux à F (pour "face"), les autres égaux à P (pour "pile").

L'ensemble fondamental S associé à l'épreuve $\mathscr{A}$ est donc :

$S = \{P, F\}^n = \{(x_1, x_2, \ldots, x_n) / x_i = P \text{ ou } x_i = F \text{ pour } i = 1, \ldots, n\}$

Supposons que le joueur marque un point chaque fois qu'il obtient "pile". A chaque résultat de l'épreuve $\mathscr{A}$, on peut donc associer le nombre de points marqués. On définit ainsi une application X de S dans l'ensemble des entiers naturels N. L'ensemble des valeurs prises par X est

$X(S) = \{0, 1, 2, \ldots, n\}$

*Exemple 2*

Considérons une cible circulaire de rayon $R_o > 0$, et supposons que tous les tirs atteignent la cible et que la probabilité d'atteindre une partie de la cible est proportionnelle à la surface de cette partie. A chaque tir, on associe la distance du point d'impact au centre de la cible. Comme dans l'exemple précédent, à chaque résultat de l'épreuve on associe un nombre, ce qui détermine une application X. L'ensemble des valeurs prises par X est l'intervalle $[0,R_o]$.

Définition

Soit S un ensemble fondamental associé à une épreuve $\mathscr{A}$. On appelle variable aléatoire toute application X définie sur S à valeurs numériques.

Notation

Si X est une variable aléatoire définie sur un ensemble fondamental S relatif à une épreuve $\mathscr{A}$, et si a est un nombre réel, on pose :

$$(X = a) = \{r \in S/X(r) = a\} \qquad (6.1)$$

C'est à dire que $(X = a)$ est l'ensemble des résultats de l'épreuve $\mathscr{A}$ auxquels l'application X associe la valeur a.

. De même :

$$(X \leqslant a) = \{r \in S/X(r) \leqslant a\} \qquad (6.2)$$

On définirait de même $(a \leqslant X \leqslant b)$, $(X < a)$, etc.

. Si E est une partie de R, on pose :

$$(X \in E) = \{r \in S/X(r) \in E\} \qquad (6.3)$$

*Exemple 3*

Considérons la variable aléatoire X du premier exemple. On a vu que l'ensemble des valeurs prises par X est

$$X(S) = \{0,1,2,\ldots,n\}$$

L'événement $(X = 0)$ est l'ensemble des résultats de l'épreuve

tels que le joueur ne marque aucun point. On a donc :

$$(X = 0) = \{(F, F, F, \ldots, F)\}$$

De même :

$$(X = 1) = \{(P,F,\ldots,F), (F,P,F,\ldots,F),\ldots, (F,\ldots,F,P)\}$$

etc.

$$(X = n) = \{(P,P,P,\ldots,P)\}$$

Si la probabilité p d'obtenir "pile" est la même pour toutes les pièces, on aura (cf. 5.19)

$$P(X = i) = C_n^i p^i q^{n-i} \text{ pour } i = 0,1,2,\ldots,n \qquad (6.4)$$

où q = 1-p est la probabilité d'obtenir "face" avec une pièce quelconque.

On a donc construit un nouvel ensemble - l'ensemble X (S) des points que le joueur peut marquer - et transporté la probabilité définie sur S à ce nouvel ensemble.

*Exemple 4*

Considérons la variable aléatoire X du second exemple. Soit r un nombre tel que $0 \leqslant r \leqslant R_o$. (X = r) est l'ensemble des points de la cible dont la distance au centre est r. La surface de cet ensemble est nulle, et par suite P (X = r) = 0. Par contre : $P(X \leqslant r) = \frac{\pi r^2}{\pi R_o^2} = \frac{r^2}{R_o^2}$.

## II. LOI DE PROBABILITE, FONCTION DE REPARTITION, DENSITE DE PROBABILITE

### Définition

On appelle loi de probabilité d'une variable aléatoire X définie sur un ensemble fondamental S, la donnée des probabilités P ($X \in E$) pour tout intervalle E de R.

Plusieurs cas se présentent, suivant que l'ensemble X (S) des valeurs prises par la variable aléatoire X est fini, dénombrable ou continu.

1. Variable aléatoire finie

$$X(S) = \{x_1, x_2, \ldots, x_n\}$$

La loi de probabilité de X est entièrement déterminée par la donnée des $p_i = P(X = x_i)$ pour $i = 1, \ldots, n$. On a :

$$\forall i = 1, \ldots, n,\ p_i \geqslant 0$$

$$\sum_{i=1}^{n} p_i = 1 \qquad (6.5)$$

2. Variable aléatoire dénombrable

$$X(S) = \{x_1, x_2, \ldots, x_n, \ldots\}$$

Comme précédemment la loi de probabilité de X est entièrement déterminée par la donnée des $p_i = P(X = x_i)$ pour $i \in N^*$. On a :

$$\forall i \in N^*,\ p_i \geqslant 0$$

$$\sum_{i=1}^{\infty} p_i = 1$$

Remarque : si X (S) est finie ou dénombrable, X est dite discrète.

3. Variable aléatoire continue

X est dite continue si X (S) est une réunion d'intervalles de R. (c'est le cas du second exemple, où X (S) = [0,R]).

Pour tout nombre x, on a P (X = x) = 0.

On détermine la loi de probabilité de X par la donnée des probabilités P ($X \leqslant x$), pour tout $x \in R$.

Définition

Si X est une variable aléatoire définie sur un ensemble fondamental S, on appelle fonction de répartition de X l'application F définie sur R par :

$$\forall x \in R,\ F(x) = P(X \leqslant x)$$

Définition

Si la fonction de répartition F d'une variable aléatoire continue X est dérivable en tout point $x \in R$, de dérivée f (x), sauf peut-être en un nombre fini de points, et si :

$$\forall x \in R,\ P\ (X \leqslant x) = F\ (x) = \int_{-\infty}^{x} f\ (t)\ dt \qquad (6.6)$$

on dit que X est une variable aléatoire absolument continue. f est appelée la densité de probabilité (ou encore fonction de distribution) de X.

*Exemple 5*

La variable aléatoire X des exemples 2 et 4 est absolument continue. Sa fonction de répartition est :

$$F\ (x) = \begin{cases} 0 & \text{si } x < 0 \\ \dfrac{x^2}{R_o^2} & \text{si } 0 \leq x \leq R_o \\ 1 & \text{si } x > R_o \end{cases}$$

F est dérivable en tout point $x \in R$ sauf en $x = R_o$.

La densité de probabilité de X est :

$$f\ (x) = \begin{cases} 0 & \text{si } x < 0 \\ \dfrac{2x}{R_o^2} & \text{si } 0 \leq x \leq R_o \\ 0 & \text{si } x > R_o \end{cases}$$

Remarques :

1) Si X est une variable aléatoire absolument continue de densité de probabilité f, on a :

$$\text{. } \forall a \in R,\ \forall b \in R,\ a \leqslant b,\ P\ (a \leqslant X \leqslant b) = \int_b^a f(t)dt = F\ (b) - F\ (a) \qquad (6.7)$$

$$\text{. } \int_{-\infty}^{\infty} f\ (t)\ dt = 1 \qquad \text{et} \qquad \forall t \in R,\ f\ (t) \geqslant 0$$

2) Si X est une variable aléatoire finie ou dénombrable

$(X(S) = \{x_1, x_2, \ldots\})$ la fonction de répartition F de X détermine encore la loi de probabilité de X et on a :

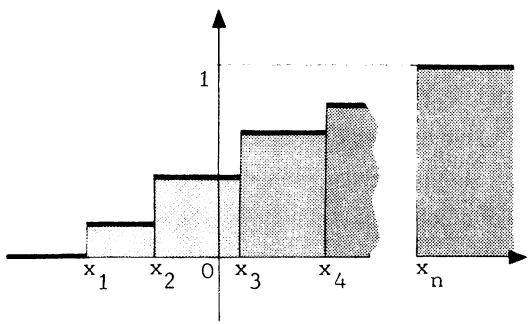
$$P(a \leqslant X \leqslant b) = \sum_{i=\alpha}^{\beta} P(X = x_i),$$

$$\text{si } x_{\alpha-1} < a \leqslant x_\alpha \quad \text{et } x_{\beta-1} < b \leqslant x_\beta$$

Dans ce cas :

$$F(x) = \sum_{i=1}^{k} P(X = x_i), \text{ si } x_k \leqslant x < x_{k+1}$$

et F est une fonction en escalier :




## III. ESPERANCE MATHEMATIQUE ET MOMENTS

La notion de variable aléatoire est la transposition probabiliste de la notion statistique de caractère. Au lieu de distribution de fréquences, on parle de loi de probabilité d'une variable aléatoire.

Les lois de variables aléatoires se représentent comme les distributions de fréquences (cf. chap. 2). Elles s'analysent de la même manière, au moyen de paramètres de position ou de dispersion, ou de moments.

### 1. Variable aléatoire finie

#### Définition

Soit X une variable aléatoire finie sur un ensemble fon-

damental S. Si X (S) = $\{x_1, x_2, \ldots, x_n\}$ on appelle espérance mathématique, ou moyenne, de X le nombre :

$$E(X) = \sum_{i=1}^{n} x_i \, P(X = x_i) \qquad (6.8)$$

## 2. Variable aléatoire dénombrable

### Définition

Soit X une variable aléatoire dénombrable sur un ensemble fondamental S. X (S) = $\{x_1, x_2, \ldots, x_n, \ldots\}$. On appelle espérance mathématique de X la somme de la série

$$(x_i \, P(X = x_i))_{i \in N^*}$$

si cette série est absolument convergente, c'est à dire si

$$\sum_{i=1}^{\infty} |x_i| \, P(X = x_i) < +\infty$$

Dans ce cas :

$$E(X) = \sum_{i=1}^{\infty} x_i \, P(X = x_i)$$

Cela assure que la valeur de E (X) ne dépend pas de l'ordre dans lequel on a numéroté les éléments de S (X). Si l'on avait seulement $\sum_{i=1}^{\infty} x_i \, P(X = x_i) < +\infty$ on montre que l'on pourrait donner n'importe quelle valeur à E (X) en changeant la numérotation des éléments de S (X).

## 3. Variable aléatoire absolument continue

### Définition

Si X est une variable aléatoire absolument continue de densité de probabilité f, on appelle espérance mathématique de X le nombre :

$$E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx \qquad (6.9)$$

lorsque l'intégrale est convergente.

4. Moments d'une variable aléatoire

On appelle moment d'ordre k, $k \in N$, d'une variable aléatoire X, le nombre $m_k$ défini par :

$$m_k = E(X^k).$$

On appelle moment centré d'ordre k de X, $k \in N$, le nombre $\mu_k$ défini par :

$$\mu_k = E\left[ (X - E(X))^k \right] \qquad (6.10)$$

Le moment centré d'ordre 2 est la variance de X :

$$V(X) = \mu_2 = E\left[ (X - E(X))^2 \right] \qquad (6.11)$$

On montre que : (cf. exercice 6 A IV)

$$V(X) = E(X^2) - [E(X)]^2 \qquad (6.12)$$

La racine carrée de la variance est l'écart-type :

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

## IV. INEGALITE DE BIENAYME-TCHEBYCHEFF ET LOI DES GRANDS NOMBRES

L'inégalité de Bienaymé-Tchébycheff permet de calculer la probabilité de l'événement : $(|X - E(X)| > a)$, $a \in R^+$. Soit X une variable aléatoire absolument continue de densité de probabilité f. Soient m l'espérance mathématique de X et $\sigma$ son écart-type.

Si $t \in R$, on a :

$$P(|X - m| > t\sigma) = P(X \leqslant m - t\sigma) + P(X \geqslant m + t\sigma)$$

or

$$P(X \leqslant m - t\sigma) = \int_{-\infty}^{m-t\sigma} f(x)\, dx$$

$$P\ (X \geq m + t\sigma) = \int_{m+t\sigma}^{+\infty} f\ (x)\ dx$$

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - m)^2\ f\ (x)\ dx = \int_{-\infty}^{m-t\sigma} (x - m)^2\ f\ (x)\ dx$$

$$+ \int_{m-t\sigma}^{m+t\sigma} (x - m)^2\ f\ (x)\ dx + \int_{m+t\sigma}^{+\infty} (x - m)^2\ f(x)\ dx$$

donc

$$\sigma^2 \geq \int_{-\infty}^{m-t\sigma} (x-m)^2\ f\ (x)\ dx + \int_{m+t\sigma}^{+\infty} (x-m)^2\ f\ (x)\ dx$$

or

$$\int_{-\infty}^{m-t\sigma} (x-m)^2\ f\ (x)\ dx \geq t^2\ \sigma^2 \int_{-\infty}^{m-t\sigma} f\ (x)\ dx$$

$$\int_{m+t\sigma}^{+\infty} (x-m)^2\ f\ (x)\ dx \geq t^2\ \sigma^2 \int_{m+t\sigma}^{+\infty} f\ (x)\ dx.$$

On a donc : $\sigma^2 \geq t^2\ \sigma^2 \times P\ (|X - m| > t\sigma)$

Il en résulte l'inégalité de Bienaymé-Tchébycheff :

$$P\ (|X - m| > t\sigma) \leq \frac{1}{t^2} \qquad (6.13)$$

Considérons maintenant une suite infinie d'expériences aléatoires $\mathscr{A}_1$, $\mathscr{A}_2$, ..., $\mathscr{A}_n$, ... identiques et indépendantes les unes des autres. A la $i^{ème}$ épreuve $\mathscr{A}_i$, associons une variable aléatoire $X_i$. Ces variables aléatoires ont même espérance mathématique m et même variance $\sigma^2$.

Soit $Y_n$ la variable aléatoire définie pour tout $n \in N$ par :

$$Y_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

On vérifie que $E\ (Y_n) = m$

$$V\ (Y_n) = \frac{\sigma^2}{n}$$

D'après l'inégalité de Bienaymé-Tchébycheff appliquée à $Y_n$ :

$$P\left(|Y_n - m)| > \frac{t\sigma}{\sqrt{n}}\right) < \frac{1}{t^2}$$

en posant :

$$\in = \frac{t\sigma}{\sqrt{n}}$$

on a :

$$P\left(|Y_n - m| > \in\right) < \frac{\sigma^2}{n\epsilon^2} \qquad (6.14)$$

On en déduit la loi (faible) des grands nombres :

$$\lim_{n \to +\infty} P\left(|Y_n - m| > \epsilon\right) = 0 \qquad (6.15)$$

*Exemple*

Soit $X_i$ le nombre de points marqués lors du $i^{ème}$ jet d'un dé. Pour tout

$$i \in N^*, \quad m = E(X_i) = \frac{1}{6}[1+2+3+4+5+6] = 3,5$$

$$\sigma^2 = V(X_i) = E(X_i^2) - [E(X)]^2$$

$$= \frac{1}{6}[1+4+9+16+25+36] - 12,25$$

$$\simeq 2,916$$

Déterminons le nombre n de jets nécessaires pour que l'on ait au moins 8 chances sur 10 que la moyenne $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ des points marqués lors des n jets s'écarte de m de moins de $\frac{1}{10}$.

On a d'après la loi des grands nombres :

$$P\left(|Y_n - m| < \frac{1}{10}\right) \geq 1 - \frac{10^2\,\sigma^2}{n}$$

On souhaite : $1 - \frac{10^2\,\sigma^2}{n} \geq \frac{8}{10}$

c'est à dire : $\frac{10^2\,\sigma^2}{n} \leq \frac{2}{10}$. Compte tenu de ce que $\sigma^2 = 2,916$ on en déduit $n \geq 1458$. ∎

## B. LOI BINOMIALE

### I. DEFINITION

Une variable aléatoire dénombrable X à valeurs dans $\mathbb{N}$, suit une loi binômiale de paramètres n et p si

$$\forall k \in \mathbb{N} \qquad P(X = k) = C_n^k \, p^k \, q^{n-k} \tag{6.16}$$

où $q = 1-p$. Cette loi est notée $\mathcal{B}(n,p)$. On reconnaît dans l'expression (6.16) le terme général du développement du binôme de Newton (cf. 4.12), d'où le nom de loi binômiale donné à cette loi de probabilité (on dit aussi loi de Bernouilli).

On rencontre cette loi à chaque fois où il s'agit de déterminer la probabilité de réaliser k fois un événement A dans une série de n expériences aléatoires $\mathcal{A}$ caractérisées chacune par deux modalités complémentaires de probabilités p et q, telles que $p+q = 1$ (cf. 5.19).

*Exemple*

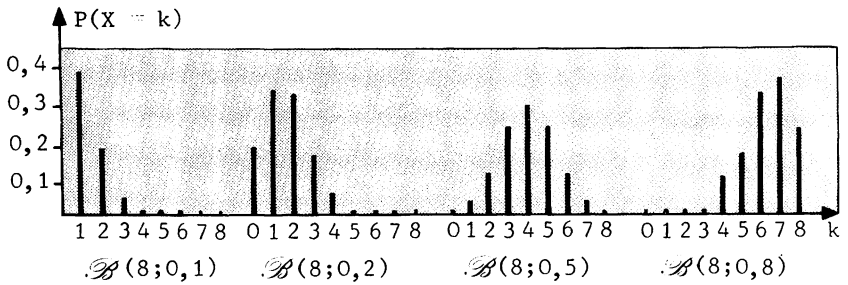
Probabilité d'obtenir k fois as avec un dé lancé n fois successives, ou avec n dés lancés simultanément. On a $p = \frac{1}{6}$, $q = \frac{5}{6}$, d'où d'après (6.16)

$$P(X = k) = C_n^k \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k}$$

## II. DIAGRAMME ET PARAMETRES CARACTERISTIQUES

Le diagramme en bâtons de la loi $\mathscr{B}(n,p)$ dépend des valeurs de n et p. Au fur et à mesure que p augmente, le diagramme passe progressivement d'une forme décroissante dite en L, à une courbe avec un maximum d'abord dissymétrique gauche, puis symétrique pour $p = \frac{1}{2}$, et ensuite dissymétrique droite.

*Exemple*




On peut montrer (cf. exercice 6 B I) que dans le cas d'une loi binômiale, on a :

| | | |
|---|---|---|
| Espérance mathématique | $m = E(X) = np$ | (6.17) |
| Variance | $v = V(X) = npq$ | |

Il existe des tables de la loi binômiale. La table I en fournit un extrait très limité. En pratique, on n'utilise cette loi que pour n réduit à quelques unités. Dans certaines conditions - précisées ci-dessous - où elle est d'un maniement difficile, on préfère la remplacer par la loi de Poisson ou la loi normale qui en sont des formes asymptotiques.

## III. APPROXIMATIONS DE LA LOI BINOMIALE

### a) par la loi de Poisson

Lorsque $n \to +\infty$ et $p \to 0$, et que l'espérance mathématique reste constante $E(X) = np = \lambda$, on peut montrer que :

$$\mathscr{B}(n,p) \longrightarrow \mathscr{P}(\lambda) \qquad \text{(cf. exercice 6 B VIII)}$$

où $\mathscr{P}(\lambda)$ désigne la loi de Poisson (cf. 6 C)

$$P(X = k) = \lambda^k \frac{e^{-\lambda}}{k!} \tag{6.18}$$

En pratique, on considère que $\mathscr{P}(\lambda)$ constitue une très bonne approximation de $\mathscr{B}(n,p)$ lorsque n est assez grand ($n > 50$), p assez petit ($p < 0,1$) et np compris entre 0 et 10. Elle est encore valable pour $10 \leqslant np \leqslant 20$, à condition d'avoir $n \geq 200$. Pour $np > 20$, la loi normale fournit une meilleure approximation.

### b) par la loi normale

Lorsque $n \to \infty$ et que $(x - np)/\sqrt{npq} \to t$, où t est fini et $p+q = 1$

$$\sqrt{npq}\ \mathscr{B}(n,x) \longrightarrow \rho(t)$$

où $\rho(t)$ désigne la densité de probabilité de la loi normale (cf. 6.26)

$$\rho(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \tag{6.19}$$

## C. LOI DE POISSON

### I. DEFINITION

Une variable aléatoire dénombrable X à valeurs dans $\mathbb{N}$, suit une loi de Poisson de paramètre $\lambda$ si

$$\forall k \in \mathbb{N} \qquad P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \tag{6.20}$$

où $\lambda > 0$. Cette loi est notée $\mathcal{P}(\lambda)$, elle ne dépend que du seul paramètre $\lambda$.

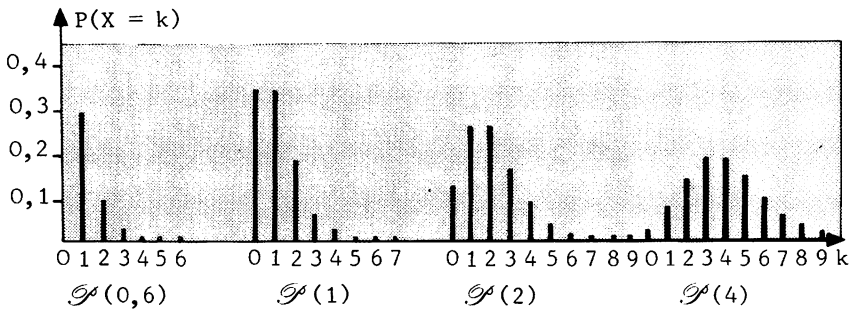
La loi de Poisson intervient généralement lorsque l'événement est très rare sur un grand nombre d'observations, comme par exemple dans le décompte de la fréquence d'apparition d'un événement rare dans un intervalle de temps ou d'espace déterminé (décompte de bactéries pendant un certain temps ou des erreurs typographiques dans un livre, etc.).

Rappelons que la loi de Poisson constitue une approximation de la loi binômiale (cf. 6.18) lorsque dans cette dernière n est assez grand et p est faible de sorte que $0 < np < 10$. Dans ce cas $\lambda$ n'est autre que la moyenne $m = np$.

### II. DIAGRAMME ET PARAMETRES CARACTERISTIQUES

Lorsque $\lambda \leq 1$, le diagramme en bâtons de la loi $\mathcal{P}(\lambda)$ est en forme de L. Pour $\lambda > 1$, il présente une forme "en cloche" d'autant plus symétrique que $\lambda$ est grand.

*Exemple*




On peut montrer que, dans le cas d'une loi de Poisson, on a :

| | |
|---|---|
| **Espérance mathématique** | $m = E(X) = \lambda$ |
| **Variance** | $v = V(X) = \lambda$ |

(6.21)

Il existe des tables donnant les valeurs de $P(X = k)$ pour des valeurs de $\lambda$ variant entre 0 et 20, domaine d'utilisation courante de la loi de Poisson. La table 2 en fournit un extrait très limité.

D'une manière générale, le calcul des différentes valeurs de $P(X = k)$ est grandement facilité par la relation de récurrence simple qui suit. D'après l'expression (6.20) de la loi de Poisson, on a

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$P(X = k+1) = \frac{\lambda^{k+1}}{(k+1)!} e^{-\lambda}$$

soit en divisant membre à membre :

$$P(X = k+1) = P(X = k) \cdot \frac{\lambda}{k+1} \qquad (6.22)$$

Il suffit donc de connaître la probabilité

$$P\ (X = 0) = e^{-\lambda}$$

pour en déduire toutes les autres, de proche en proche, en utilisant la relation (6.22).

*Exemple 1*

4 % des articles d'une certaine fabrication présentent des défauts. Quelle est la probabilité pour que dans une livraison de 75 de ces articles, il y ait 2 articles défectueux ?

La probabilité élémentaire pour qu'un article pris au hasard dans cette fabrication soit défectueux est p = 0,04. En toute rigueur, la probabilité d'avoir 2 articles défectueux dans un lot de 75 est donnée par $\mathcal{B}(75\ ;\ 0{,}04)$ soit :

$$P\ (X = 2) = C_{75}^{2}\ (0{,}04)^{2}\ (0{,}96)^{73}$$

La loi binômiale peut paraître ici d'un maniement difficile. Cependant comme n = 75, p = 0,04 et m = np = 3 < 10, on peut recourir à l'approximation de $\mathcal{B}(75\ ;\ 0{,}04)$ par $\mathcal{P}(3)$. La probabilité cherchée est donc

$$P\ (X = 2) = \frac{3^{2}}{2!}\ e^{-3} = 0{,}224$$

en utilisant la table 2, avec $\lambda = 3$ et $k = 2$.

*Exemple 2*

A partir des tables 1 et 2, étudier l'approximation de $\mathcal{B}(10\ ;\ 0{,}1)$ par $\mathcal{P}(1)$.

Les tables 1 et 2 permettent de dresser le tableau suivant :

| P (X = k) | Loi binômiale | Loi de Poisson |
|---|---|---|
| P (X = 0) | 0,3487 | 0,3679 |
| P (X = 1) | 0,3874 | 0,3679 |
| P (X = 2) | 0,1937 | 0,1839 |
| P (X = 3) | 0,0574 | 0,0613 |
| P (X = 4) | 0,0112 | 0,0153 |

On voit que les écarts sont relativement importants. La table 1 est ici limitée à n = 10, ce qui n'est pas suffisant pour justifier l'approximation. Cependant le calcul montre que ces écarts diminuent au fur et à mesure que n augmente et qu'on s'approche des conditions d'application de l'approximation.

## D. LOI NORMALE

### I. DEFINITION

Soit X une variable aléatoire réelle absolument continue. On dit que X suit une loi normale (ou de Laplace-Gauss) si la densité de probabilité est :

$$\forall x \in R,\ f(x) = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} \qquad (6.23)$$

où e = 2, 718 ... est la base des logarithmes népériens, et m et $\sigma$ sont deux constantes avec $\sigma > 0$. C'est une loi à deux paramètres m et $\sigma$, notée $\mathcal{N}(m,\sigma)$.

Remarque : f est bien une densité de probabilité. En effet :

a) $\forall x \in R \quad f(x) \geqslant 0$ car $\sigma > 0$

b) $S = \int_{-\infty}^{+\infty} f(x)\,dx = 1$ (cf. exercice 6 D I)

$$S = \frac{1}{\sigma\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx$$

ou encore :

$$S = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-t^2/2}\,dt = 1 \qquad (6.24)$$

en effectuant le changement de variable $t = \frac{x-m}{\sigma}$.

Loi normale centrée réduite :

En effectuant le changement de variable $t = \frac{x-m}{\sigma}$, on définit une nouvelle densité de probabilité

$$\forall t \in R,\quad \rho(t) = \frac{1}{\sqrt{2\pi}}\ e^{-t^2/2} \qquad (6.25)$$

Ce changement de variable correspond à un changement d'échelle et à une translation sur l'axe des abscisses. On obtient une densité $\rho$ indépendante de m et $\sigma$, ce qui permet d'utiliser la même courbe pour des variables aléatoires suivant des lois normales de différents paramètres.

## II. DENSITE DE PROBABILITE

Soit T une variable aléatoire absolument continue et suivant $\mathscr{N}(0,1)$. Sa densité de probabilité est donnée par

$$\rho(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \qquad (6.26)$$

Cette fonction est très simple à représenter. Elle est paire, sa dérivée première s'annule pour t = 0 et sa dérivée seconde pour t ± 1 (points d'inflexion). On obtient la courbe de la figure 6.1, dite courbe de Gauss (ou gaussienne).

| t | 0 | 1 | 2 | ∞ |
|---|---|---|---|---|
| $\frac{d^2\rho}{dt^2}$ | - | 0 | + | |
| $\frac{d\rho}{dt}$ | 0 - | | - | |
| $\rho$ | 0,399 | 0,242 | 0,054 | 0 |

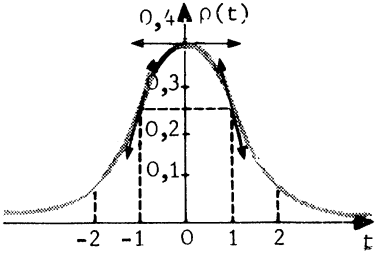



figure 6.1.

D'après la linéarité de l'espérance (cf. exercice 6 A III), on montre que E (T) = 0 et V (T) = 1, et que d'une manière générale les paramètres m et $\sigma$ de la loi normale $\mathscr{N}(m,\sigma)$ définie par (6.23), ne sont autres que l'espérance mathématique et l'écart-type de la variable aléatoire X.

## III. FONCTION DE REPARTITION

Soit F la fonction de répartition d'une variable aléatoire X de densité de probabilité f donnée par (6.23), et soit T la variable aléatoire réduite correspondante $T = \frac{X-m}{\sigma}$, de densité de probabilité $\rho$ donnée par (6.26).

D'après (6.7), on a :

$$F(x) = P(X \leqslant x) = \int_{-\infty}^{x} f(x)\, dx \qquad (6.27)$$

soit en effectuant le changement de variable

$$F(x) = P(X \leqslant x) = \int_{-\infty}^{t=\frac{x-m}{\sigma}} \rho(t)\, dt \qquad (6.28)$$

Par conséquent la fonction de répartition

$$\mathcal{G}(t) = \int_{-\infty}^{t} \rho(t)\, dt \qquad (6.29)$$

pourra être tabulée et servir au calcul des probabilités attachées à n'importe quelle variable aléatoire X distribuée normalement.

Sur le graphe de la figure 6.2., la probabilité $P(X \leq a)$ est représentée par la surface de l'aire foncée a) et la probabilité $P(a \leq X \leq b)$ par celle de l'aire grise b).

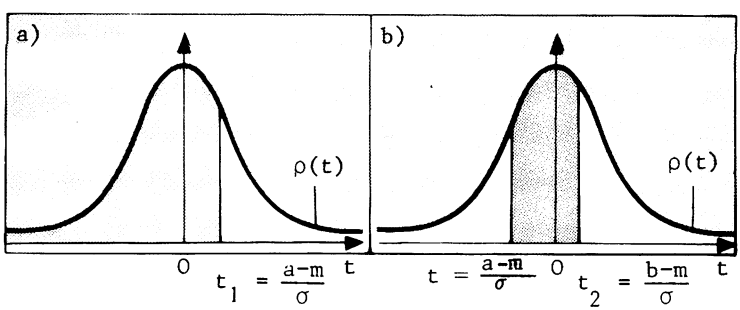



figure 6.2.

On peut remarquer que, d'après (6.29)

$$\mathcal{G}(t) = \int_{-\infty}^{0} \rho(t)\,dt + \int_{0}^{t} \rho(t)\,dt = 0{,}5 + G(t)$$

où

$$G(t) = \int_{0}^{t} \rho(t)\,dt \qquad (6.30)$$

Il existe des tables de valeurs numériques de ces deux fonctions. La table 3 en fournit un exemple pour G (t).

Remarque :

Soit X une variable aléatoire distribuée normalement dans une certaine population, avec une moyenne m et un écart-type σ. L'étude de la concentration de la population autour de la valeur moyenne, permet de mettre en évidence les propriétés suivantes de la loi normale :

A partir de la table 3, on peut remarquer que

$P(m-\sigma \leq X \leq m+\sigma) = P(-1 \leqslant T \leqslant 1) = 2\,G(1) \simeq 0{,}683$

$P(m-2\sigma \leqslant X \leqslant m+2\sigma) = P(-2 \leqslant T \leqslant 2) = 2\,G(2) \simeq 0{,}954$

$P(m-2{,}6\sigma \leqslant X \leqslant m+2{,}6\sigma) = P(-2{,}6 \leqslant T \leqslant 2{,}6) = 2\,G(2{,}6) \simeq 0{,}99$

On peut donc dire que dans une distribution normale,

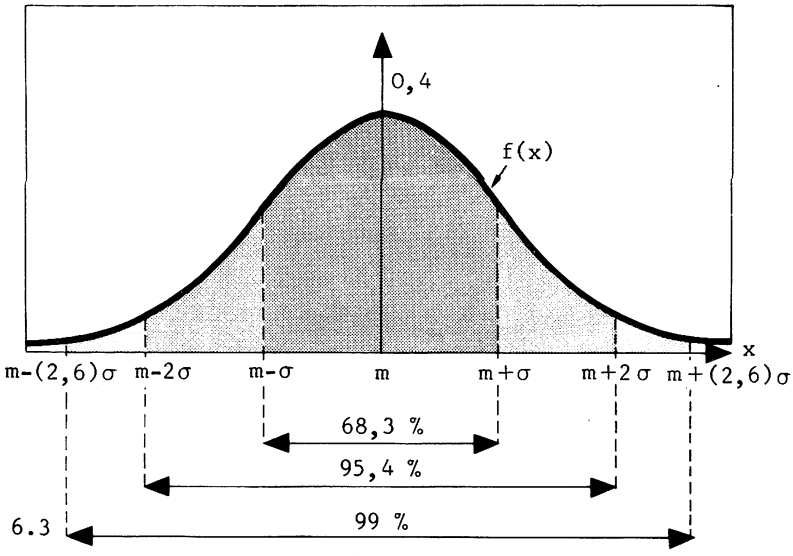



Figure 6.3

68,3 % de la population sont concentrés sur un écart-type de part et d'autre de la moyenne, 95,4 % sur deux écarts-type et 99 % sur 2,6 écarts-type, ce qui est illustré sur la figure 6.3 qui précède.

## IV. EXEMPLE D'APPLICATION DE LA LOI NORMALE

La loi normale occupe une place de choix parmi les différentes lois de probabilité. En effet, un grand nombre de lois de distributions courantes se rapprochent de la loi normale. On montre que, lorsqu'une variable aléatoire peut être considérée comme la somme d'un nombre suffisamment grand de variables aléatoires indépendantes, sa loi de distribution tend vers une loi normale. C'est le cas des erreurs de mesure ou des erreurs de tirs par exemple.

Rappelons que la loi normale constitue une approximation de la loi binômiale (cf. 6.19) lorsque dans cette dernière n est assez grand et p n'est pas très faible, de sorte que np > 10.

### *Exemple 1 : Expérience de Galton*

On considère une planche inclinée sur laquelle sont fixées n rangées horizontales de clous disposés en quinconce (voir figure 6.4); un entonnoir placé au milieu du bord supérieur de la planche déverse des billes qui, après avoir traversé les n rangées de clous, viennent se loger dans des casiers situés à la partie inférieure de la planche.

On peut montrer que la répartition des billes dans les différents casiers suit une loi binômiale qui se rapproche d'une loi normale lorsque le nombre n de rangées de clous est grand. L'expérience permet donc de visualiser une distribution gaussienne.

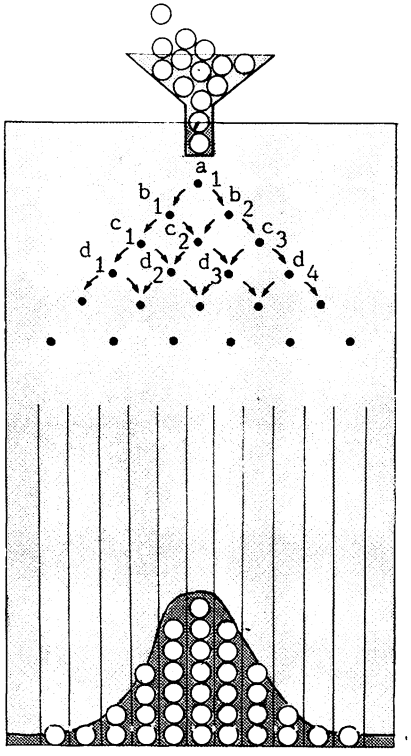



Figure 6.4.

En effet, la probabilité élémentaire pour qu'une telle bille ayant heurté un clou aille à droite ou à gauche est de $\frac{1}{2}$. En considérant par exemple les clous désignés par $a_1$ (1ère rangée), $b_1$, $b_2$ (2ème rangée), $c_1$, $c_2$, $c_3$ (3ème rangée) etc. on peut calculer les probabilités suivantes, selon le trajet de la bille.

ligne b
- arrivée en $b_1$, trajet $a_1b_1$, $p_{b_1} = \frac{1}{2}$
- " en $b_2$, " $a_1b_2$, $p_{b_2} = \frac{1}{2}$

ligne c
- " en $c_1$, " $a_1b_1c_1$, $p_{c_1} = (\frac{1}{2})^2$
- " en $c_2$, " $a_1b_1c_2$, $p = (\frac{1}{2})^2$
- " " $a_1b_2c_2$, $p = (\frac{1}{2})^2$

$p_{c_2} = 2 \times (\frac{1}{2})^2$

- " en $c_3$, " $a_1b_2c_3$, $p_{c_3} = (\frac{1}{2})^2$

De la même manière, on aura pour la ligne d

$$\text{ligne d} \begin{cases} p_{d_1} = (\frac{1}{2})^3 \\ pd_2 = 3\ (\frac{1}{2})^3 \\ p_{d_3} = 3\ (\frac{1}{2})^3 \\ p_{d_4} = (\frac{1}{2})^3 \end{cases} \qquad \text{et ainsi de suite.}$$

Si l'on considère comme variable aléatoire X le numéro du casier dans lequel va tomber la bille, on trouve que la probabilité pour qu'une bille tombe dans le casier k, après avoir franchi les n rangées de clous est donnée par une loi binômiale $\mathscr{B}(n, \frac{1}{2})$, soit

$$P\ (X = k) = C_n^k\ (\tfrac{1}{2})^k\ (\tfrac{1}{2})^{n-k} = C_n^k\ (\tfrac{1}{2})^n$$

La répartition des billes dans les casiers visualise le diagramme en bâtons de la distribution du caractère X. On obtient une courbe "en cloche" symétrique dont la moyenne correspond au casier central et dont l'écart-type est $\sigma = \sqrt{npq} = \frac{\sqrt{n}}{2}$.

En désignant par Y le rang d'un casier par rapport au casier central, et en considérant la variable aléatoire réduite $t = \frac{Y}{\sigma}$, on voit d'après (6.19) que pour n élevé, la répartition des billes dans les casiers tend vers la courbe de Gauss

$$\rho\ (t) = \frac{1}{\sqrt{2\pi}}\ e^{-t^2/2}$$

*Exemple 2*

La taille des élèves d'une école suit une distribution normale avec m = 150 cm et $\sigma$ = 20 cm. Quel est le nombre d'élèves ayant une taille comprise entre 140 et 170 cm, si l'effectif total de l'école est de 1 000 élèves ?

En passant de la variable aléatoire X qui est la taille

de l'élève, à la variable réduite $T = \frac{X - m}{\sigma}$, on obtient successivement

$$P\ (140 \leqslant X \leqslant 170) = P\ (\frac{140 - 150}{20} \leqslant T \leqslant \frac{170 - 150}{20})$$

$$= P\ (-0,5 \leqslant T \leqslant 1)$$

$$= G\ (0,5) + G\ (1)$$

$$= 0,3413 + 0,1915 = 0,533$$

en utilisant la table 3. Le nombre d'élèves cherché est donc :

$0,533 \times 1\ 000 = 533$ élèves

*Exemple 3*

Dans l'exemple précédent, quelle est la taille maximum $x_m$ des 800 élèves les plus petits ?

Ces 800 élèves représentent 80 % de l'effectif total. On a donc :

$$P\ (X \leqslant x_m) = P\ (T \leqslant t_m)$$

$$= 0,5 + G\ (t_m) = 0,8$$

On en déduit

$$G\ (t_m) = 0,3$$

La table 3 fournit $t_m = 0,85$, soit en revenant à la variable aléatoire initiale X

$$x_m = \sigma\ t_m + m = 20 \times 0,85 + 150 = 167 \text{ cm.}$$

## I. VARIABLE DE BERNOUILLI

Soit X une variable aléatoire finie prenant 2 valeurs : 0 et 1. On suppose que $P(X = 1) = p$, où $p \in R$, $0 \leqslant p \leqslant 1$ (on dit que X est une variable de Bernouilli.

1°) Calculer E (X) et V (X).

2°) Calculer les moments centrés d'ordre k ($k \in N^*$) de X.

---

SOLUTION

1°) $E(X) = 0.(1-p) + 1.p = p$

$V(X) = E[(X - E(X))^2] = (-p)^2.(1-p) + (1-p)^2.p = p(1-p)$

2°) Si $k \in N^*$, $E(X^k) = p$

$E[(X - E(X))^k] = E[(X-p)^k] = (-p)^k(1-p)+(1-p)^k p$ ■

---

II. Soit X une variable aléatoire finie prenant les n valeurs $x_1, x_2, \ldots, x_n$ avec les probabilités respectives $p_1, p_2, \ldots, p_n$. ($p_i = P(X = x_i)$).

Soient a et b deux nombres réels, et $Y = aX + b$ une variable aléatoire fonction de X.

Montrer que

1°) $E(Y) = a E(X) + b$

2°) $V(Y) = a^2 V(X)$.

---

SOLUTION

L'ensemble des valeurs de la variable aléatoire Y est $\{a x_1 + b, a x_2 + b, \ldots, a x_n + b\}$
et $P(Y = a x_i + b) = p_i$.

1°) On a donc $E(Y) = \sum_{i=1}^{n} (a x_i + b) p_i$

$$= \sum_{i=1}^{n} a x_i p_i + \sum_{i=1}^{n} b p_i$$

or $\sum_{i=1}^{n} p_i = 1$, $\sum_{i=1}^{n} b p_i = b \sum_{i=1}^{n} p_i$ et $\sum_{i=1}^{n} a x_i p_i = a \sum_{i=1}^{n} x_i p_i$

on a donc $E(Y) = a E(X) + b$.

2°) $V(Y) = E[(Y - E(Y))^2] = \sum_{i=1}^{n} (ax_i + b - aE(X) - b)^2 p_i$

$$= \sum_{i=1}^{n} a^2 (x_i - E(X))^2 p_i$$

on a donc $V(Y) = a^2 V(X)$. ∎

---

III. Soit X une variable aléatoire dénombrable prenant les valeurs $x_1, x_2, \ldots, x_n, \ldots$ avec les probabilités $p_1, p_2, \ldots, p_n, \ldots$ Soient a et b deux réels, $Y = a X + b$. On suppose que $E(X)$ existe.

1°) Montrer que $E(Y) = a E(X) + b$

$$V(Y) = a^2 V(X)$$

2°) Supposons maintenant que X est une variable aléatoire réelle absolument continue de densité de probabilité f.

Si a, b sont deux réels, $Y = a X + b$ montrer que :

$E(Y) = a E(X) + b$ (on dit que l'espérance est linéaire)

$V(Y) = a^2 V(X)$.

SOLUTION

1°) Les calculs sont semblables à ceux de l'exercice II, car si E (X) existe, toutes les séries considérées sont absolument convergentes.

2°) Par définition $E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx$

$E(Y)$ existe car $\int_{-\infty}^{+\infty} (a\,x + b)\, f(x)\, dx$ est absolument convergente comme $\int_{-\infty}^{+\infty} x\, f(x)\, dx$. Par linéarité de l'intégrale on a :

$$E(Y) = \int_{-\infty}^{+\infty} (a\,x + b)\, f(x)\, dx = a \int_{-\infty}^{+\infty} x\, f(x)\, dx + b \int_{-\infty}^{+\infty} f(x)\, dx$$

or $\int_{-\infty}^{+\infty} f(x)\, dx = 1$ car f est une densité de probabilité.

On en déduit que $E(Y) = a\,E(X) + b$.

On a :

$$V(X) = \int_{-\infty}^{+\infty} (x - m)^2\, f(x)\, dx \quad \text{où} \quad m = E(X)$$

$$V(Y) = \int_{-\infty}^{+\infty} (a\,x + b - a\,m - b)^2\, f(x)\, dx = \int_{-\infty}^{+\infty} a^2\, (x - m)^2\, f(x)\, dx$$

$$= a^2 \int_{-\infty}^{+\infty} (x - m)^2\, f(x)\, dx = a^2\, V(X). \quad \blacksquare$$

---

IV. Soit X une variable aléatoire absolument continue de densité de probabilité f. On suppose que E (X) et V (X) existent. Montrer que $V(X) = E(X^2) - [E(X)]^2$

---

SOLUTION

Si $m = E(X)$ :

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

or f étant une densité de probabilité, on a :

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

il vient donc :

$$V(X) = E(X^2) - 2m^2 + m^2 = E(X^2) - [E(X)]^2 \quad \blacksquare$$

(Ce résultat se montre aussi directement, lorsque X est discrète).

---

## V. LOI CONTINUE UNIFORME

Soit $[a,b]$ $(a < b)$ un intervalle de R. On dit que la variable aléatoire X absolument continue est uniformément répartie sur $[a,b]$ si sa densité de probabilité f est constante sur cet intervalle et nulle ailleurs.

1°) Déterminer f en fonction de a et de b.

2°) Calculer E (X) et V (X).

3°) *Déterminer la fonction de répartition F de X.*

---

SOLUTION

1°) Soit k la valeur constante de f sur $[a,b]$. f étant une densité de probabilité :

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

or $\int_{-\infty}^{+\infty} f(x)\,dx = \int_{-\infty}^{a} f(x)\,dx + \int_{a}^{b} (f(x)\,dx$

$+ \int_{b}^{+\infty} f(x)\,dx = \int_{a}^{b} f(x)\,dx$

car f est nulle sur $]-\infty, a[$ et sur $]b, +\infty[$

$\int_{a}^{b} f(x)\,dx = k \int_{a}^{b} dx = k(b-a)$ car $\forall x \in [a,b]$, $f(x) = k$.

On a donc $k = \frac{1}{b-a}$ et par suite $f(x) = \begin{cases} 0 \text{ si } x < a \\ \frac{1}{b-a} \text{ si } a \leqslant x \leqslant b \\ 0 \text{ si } x > b \end{cases}$

2°) $E(X) = \int_{-\infty}^{+\infty} x f(x)\,dx$ si cette intégrale est absolument convergente. Or $\int_{-\infty}^{+\infty} |x| f(x)\,dx = \int_{a}^{b} |x| \frac{dx}{b-a} < +\infty$

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\,dx = \int_{a}^{b} x \frac{dx}{b-a} = \frac{b+a}{2}$$

$$E(X^2) = \int_{a}^{b} \frac{x^2\,dx}{b-a} = \frac{b^3 - a^3}{3(b-a)}$$

donc comme

$$V(X) = E(X^2) - [E(X)]^2 \qquad \text{(cf. exercice IV)}$$

on a :

$$V(X) = \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4} = \frac{(a-b)^2}{12}$$

3°) Soit $x \in \mathbb{R}$, $F(x) = \int_{-\infty}^{x} f(t)\,dt$

donc si $x < a$, $F(x) = 0$

si $x > b$, $F(x) = 1$

si $a \leqslant x \leqslant b$ $\quad F(x) = \int_{a}^{x} f(t)\,dt = \frac{x-a}{b-a}$

on a donc $\quad F(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } x > b. \end{cases}$ ■

## VI. LA LOI DE CAUCHY

Soit f la fonction numérique d'une variable réelle définie par

$$\forall x \in \mathbb{R},\ f(x) = \frac{c}{1 + x^2}\ ,\ \text{où } c \in \mathbb{R}.$$

1°) Déterminer c pour que f soit une densité de probabilité.

2°) Soit X une variable aléatoire absolument continue de densité de probabilité f (on dit alors que X suit une loi de Cauchy). Montrer que X n'admet pas d'espérance mathématique.

---

SOLUTION

1°) Pour que f soit une densité de probabilité, il est nécessaire et suffisant que :

- $\forall x \in \mathbb{R},\ f(x) \geqslant 0$
- $\int_{-\infty}^{+\infty} f(x)\,dx = 1$

La première condition entraîne que $c \geqslant 0$.

Comme f est paire ($f(x) = f(-x)$) on a :

$$\int_{-\infty}^{+\infty} f(x)\,dx = 2\int_{0}^{+\infty} f(x)\,dx$$

et $$\int_{0}^{+\infty} f(x)\,dx = \lim_{a \to +\infty} \int_{0}^{a} f(x)\,dx$$

Or $$\int_{0}^{a} f(x)\,dx = c\int_{0}^{a} \frac{dx}{1 + x^2} = c\ \mathrm{Arctg}\ a \quad (\text{si } a \geqslant 0)$$

On en déduit que :

$$\int_{-\infty}^{+\infty} f(x)\,dx = 2c \lim_{a \to +\infty} \mathrm{Arctg}\ a = 2c\,\frac{\pi}{2} = c\pi$$

On en déduit que $c = \frac{1}{\pi}$

2°) Si E (X) existait, l'intégrale $\int_{-\infty}^{+\infty} \frac{x\,dx}{\pi(1 + x^2)}$ serait définie.

Or si $x \to +\infty$ , $\dfrac{x}{(1 + x^2)} \sim \dfrac{1}{x}$ et $\displaystyle\int_0^{+\infty} \frac{dx}{x}$ n'est pas définie. E (X) n'existe donc pas. ■

I. La distribution d'une variable aléatoire X associé à un événement A est une loi binômiale $\mathscr{B}(n,p)$. Calculer l'espérance mathématique et la variance de X.

---

SOLUTION

Après n épreuves, la probabilité pour que X prenne la valeur k (0, 1, ..., n) est :

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

(k est le nombre de fois où l'événement A s'est réalisé)

p = probabilité de réalisation de l'événement A pour 1 épreuve,

q = 1-p.

$$(1)\quad \sum_{k=0}^{n} P\ (k) = \sum_{k=0}^{n} C_n^k\ p^k\ q^{n-k} = (p+q)^n = 1$$

- Calcul de l'espérance mathématique

La définition de l'espérance mathématique est

$$m = \overline{k} = \sum_{k=0}^{n} k\ C_n^k\ p^k\ q^{n-k}$$

On s'intéresse à l'expression $(p + q)^n$. Si l'on dérive cette expression par rapport à p on obtient :

$$\frac{\partial}{\partial p}\ (p + q)^n = n\ (p + q)^{n-1} = n, \text{ car } p + q = 1$$

On a donc d'après l'équation (1) :

$$\frac{\partial}{\partial p}\ \left[\sum_{k=0}^{n} C_n^k\ p^k\ q^{n-k}\right] = n$$

La dérivée d'une somme de fonctions étant égale à la somme des dérivées, on a :

$$\sum_{k=0}^{n} \frac{\partial}{\partial p} (C_n^k p^k q^{n-k}) = n = \sum_{k=0}^{n} k C_n^k p^{k-1} q^{n-k}$$

En multipliant les 2 derniers termes de l'égalité par p, on a :

$$np = p \sum_{k=0}^{n} k C_n^k p^{k-1} q^{n-k} = \sum_{k=0}^{n} k C_n^k p^k q^{n-k} = m$$

Par suite :

$$m = np$$

- Calcul de la variance

D'après l'exercice 6 A IV, on a :

$$V(X) = E(X^2) - \left[E(X)\right]^2 = \overline{k^2} - \overline{k}^2$$

$$\overline{k^2} = \sum_{k=0}^{n} k^2 C_n^k p^k q^{n-k}$$

On a toujours $(p+q)^n = 1$. On dérive par rapport à p

$$\frac{\partial}{\partial p} (p+q)^n = n (p+q)^{n-1} = \sum_{k=0}^{n} C_n^k k p^{k-1} q^{n-k}$$

On dérive une deuxième fois par rapport à p :

$$\frac{\partial}{\partial p} \left(\frac{\partial}{\partial p} (p+q)^n\right) = \frac{\partial}{\partial p} \left(n (p+q)^{n-1}\right) = n (n-1) (p+q)^{n-2}$$

$$= \sum_{k=0}^{n} C_n^k k (k-1) p^{k-2} q^{n-k}$$

La dernière égalité s'écrit, en multipliant les deux termes par $p^2$

$$n (n-1) p^2 = \sum_{k=0}^{n} C_n^k k (k-1) p^k q^{n-k} = \overline{k (k-1)} = \overline{k^2} - \overline{k}.$$

Par suite

$$\overline{k^2} - \overline{k} = n^2 p^2 - n p^2 \quad \text{qui entraîne} \quad \overline{k^2} - n^2 p^2 = \overline{k} - n p^2$$

Or $n p = \overline{k}$, on a donc $\overline{k^2} - \overline{k}^2 = n p (1 - p)$ soit

$$\overline{k^2} - \overline{k}^2 = V(X) = npq$$

En résumé, pour une loi binômiale $\mathcal{B}(n,p)$, on a :

$$E(X) = np \qquad \text{et} \qquad V(X) = np(1-p).$$ ■

---

II. Exprimer $P_{n+1}$ en fonction de $P_n = C_N^n p^n q^{N-n}$

---

SOLUTION

$$P_{n+1} = C_N^{n+1} p^{n+1} q^{N-(n+1)}$$

Or $$C_N^{n+1} = \frac{N!}{n! \times (n+1) \times (N-n)!} (N-n) = C_N^n \times \frac{(N-n)}{(n+1)}$$

$$p^{n+1} = p \times p^n$$

$$q^{N-(n+1)} = q^{N-n} \times \frac{1}{q}$$

d'où :

$$P_{n+1} = C_N^n p^n q^{N-n} \times \frac{N-n}{n+1} \times \frac{p}{q}$$

d'où

$$P_{n+1} = P_n \cdot \frac{N-n}{n+1} \cdot \frac{p}{q}$$ ■

---

III. Une enquête a permis de constater que sur 200 flacons d'un même produit pharmaceutique, 50 ne pouvaient être utilisés au-delà de 4 mois après leur livraison. Ces 200 flacons sont rangés de façon aléatoire sur une étagère. Trois personnes achètent chacune un flacon dès le 1er jour de livraison. Quelle est la probabilité pour que :

1°) deux de ces 3 personnes aient un flacon ne pouvant être utilisé 4 mois après

2°) aucune de ces personnes n'ait un flacon ne pouvant être utilisé 4 mois après.

N.B. On supposera que les 200 flacons sont disposés de telle sorte qu'ils puissent être pris au hasard de façon équiprobable.

---

SOLUTION

La variable aléatoire X : "nombre de flacons ne pouvant pas être utilisé 4 mois après la livraison" suit une loi binôminale (n,p) de paramètres :

$n = 3$ l'étude se fait sur 3 flacons (3 personnes)

et $p = \frac{50}{200} = \frac{1}{4} = 0{,}25$ en supposant que chaque flacon puisse être choisi de façon équiprobable.

1°) La probabilité pour que deux des 3 personnes ne puissent encore utiliser le produit 4 mois après est :

$$P(X = 1) = C_3^1 p^1 q^2 = 3 \times \frac{1}{4} \times (\frac{3}{4})^2 = \frac{27}{64} \simeq 0{,}42$$

2°) La probabilité pour qu'aucune des 3 personnes ne puisse utiliser le produit 4 mois après est :

$$P(X = 3) = C_3^3 p^3 q^0 = 1 \times (\frac{1}{4})^3 \times 1 = \frac{1}{64} \quad 0{,}016 \qquad \blacksquare$$

---

IV. Une étude statistique a montré que sur 1 800 demandeurs d'emploi, en moyenne 600 recherchent du travail pour la première fois, alors que les 1 200 autres ont été mis au chômage. Quelle est la probabilité pour que sur 6 personnes recherchant un emploi, il n'y ait pas plus de 2 personnes qui aient déjà travaillé ?

---

SOLUTION

La probabilité pour qu'un demandeur d'emploi n'ait jamais travaillé est :

$$q = \frac{600}{1\,800} = \frac{1}{3}$$

La probabilité pour que la personne ait déjà travaillé est :

$$p = 1 - q = \frac{2}{3}$$

La variable aléatoire X est : "nombre de demandeurs d'emploi ayant déjà travaillé". Sur 6 personnes, la probabilité pour qu'il y ait X = k est :

$$P\ (X = k) = C_6^k\ p^k\ q^{6-k}$$

Il faut que $k \leqslant 2$, d'où

$$P(X = k \leq 2) = P\ (X = 0) + P\ (X = 1) + P\ (X = 2)$$

avec $P(X = 0) = C_6^0\ p^0\ q^6 = (\frac{1}{3})^6 = \frac{1}{729}$

$$P\ (X = 1) = C_6^1\ p^1\ q^5 = 6\ .\ (\frac{2}{3})\ (\frac{1}{3})^5 = \frac{12}{729}$$

$$P\ (X = 2) = C_6^2\ p^2\ q^4 = 15\ \ (\frac{2}{3})^2\ (\frac{1}{3})^4 = \frac{60}{729}$$

Par conséquent

$$P\ (X = k \leqslant 2) = \frac{73}{729} \simeq 0{,}10$$ ■

---

V. *On considère une macromolécule formée de N molécules élémentaires* de longueur a, supposées indépendantes. L'ensemble constitue une chaîne articulée. On prend comme origine l'une des extrémités de la chaîne. Celle-ci étant linéaire, chaque élément reste parallèle à l'axe $\overrightarrow{Ox}$ et est libre de s'orienter dans un sens ou dans l'autre.

On posera :

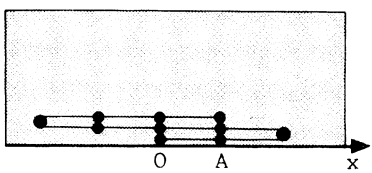
p = probabilité pour que la molécule élémentaire soit orientée vers $x > 0$.

$q = 1-p$ = probabilité pour que la molécule élémentaire soit orientée vers $x < 0$.

$n_1$ = nombre de molécules élémentaires orientées vers $x > 0$

$n_2 = N-n_1$ = nombre de molécules élémentaires orientées vers $x < 0$.

$m = n_1 - n_2$, de sorte que la longueur totale de la chaîne est égale à m.a.




dans le cas de la figure :

$N = 9$ ; $n_1 = 5$ ; $n_2 = 4$

$L = (n_1 - n_2)\, a = \overline{OA} = a$

1°) A quelle loi statistique satisfait la variable aléatoire X : "nombre de molécules élémentaires orientées vers $x > 0$" ?

2°) En déduire la longueur moyenne de la chaîne.

---

SOLUTION

1°) On considère la configuration dans laquelle les $n_1$ premières molécules élémentaires sont orientées ver la droite et les $n_2 = N-n_1$ molécules élémentaires qui restent, vers la gauche. La probabilité de cet état est :

$$p^{n_1} q^{n_2} = p^{n_1} (1-p)^{N-n_1} \qquad \text{avec} \quad p = q = \frac{1}{2}$$

Or il y a $C_N^{n_1}$ configurations où l'on a toujours $n_1$ molécules élémentaires orientées à droite. Par suite la probabilité totale sera :

$$P(X = n_1) = C_N^{n_1} p^{n_1} (1-p)^{N-n_1}$$

La loi de probabilité est donc une loi binômiale $\mathcal{B}(N, \frac{1}{2})$.

2°) On a $n_1 + n_2 = N$

$n_1 - n_2 = m$

On en déduit $n_1 = \frac{N + m}{2}$ et par conséquent

$$\overline{m} = 2\,\overline{n_1} - N$$

Or la valeur moyenne de $n_1$ est $\overline{n_1} = Np = \frac{N}{2}$

Donc $\overline{m} = 0$, la longueur moyenne (distance entre l'origine et l'extrémité) de la chaîne est nulle. ■

---

VI. Au cours d'un jeu de "pile" ou "face", sur 100 jets consécutifs de 6 pièces identiques, on obtient la distribution suivante où on a représenté en ordonnées les fréquences correspondant au nombre de fois "pile" obtenu sur les 100 jets (ce nombre correspond à la variable aléatoire X de l'expérience).

1°) Quelle est la moyenne du nombre de "pile" obtenu par jet ?

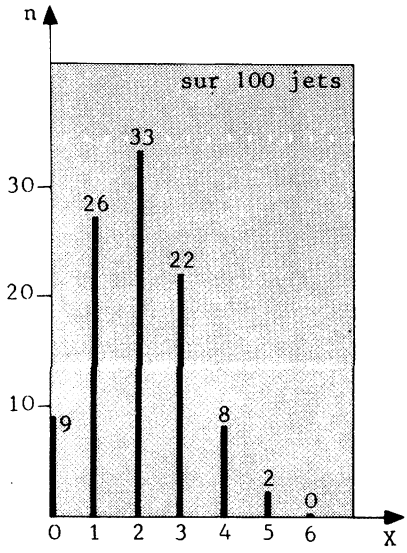
2°) En supposant que X suit une loi de distribution binômiale, déduire la probabilité élémentaire p d'obtenir "pile" sur le jet d'une pièce.

3°) Que peut-on en conclure ?

4°) Comparer les résultats exprérimentaux

a) avec les résultats théoriques déduits de la loi binômiale caractérisée à la 2ème question

b) avec ceux que l'on obtiendrait si les pièces étaient normales.



SOLUTION

1°) Calcul de la valeur moyenne du nombre de "pile" obtenu par jet de 6 pièces

$$m = \frac{0\times9+1\times26+2\times33+3\times22+4\times8+5\times2+6\times0}{100} = 2$$

2°) Si la loi est une distribution binômiale, alors les paramètres N et p sont tels que : $m = N\,p$ avec $N = 6$ et p est la probabilité élémentaire d'avoir "pile" sur le jet d'une pièce, d'où :

$$p = \frac{m}{N} = \frac{2}{6} = \frac{1}{3}$$

3°) Si les pièces étaient normales, la probabilité d'avoir "pile" (ou "face") sur le jet d'une pièce serait $p = \frac{1}{2}$ $(= q)$.

Or ici, on trouve $p = \frac{1}{3}$. On peut donc conclure que ces pièces ne sont pas normales.

4°) a) Les résultats théoriques obtenus à partir de la loi binômiale de paramètres $N = 6$ et $p = \frac{1}{3}$ sont :

$$P(X=0) = C_6^0 (\tfrac{1}{3})^0 (\tfrac{2}{3})^6 = \frac{64}{729}$$

$$P(X=1) = C_6^1 (\tfrac{1}{3})^1 (\tfrac{2}{3})^5 = \frac{192}{729}$$

$$P(X=2) = C_6^2 (\tfrac{1}{3})^2 (\tfrac{2}{3})^4 = \frac{240}{729}$$

$$P(X=3) = C_6^3 (\tfrac{1}{3})^3 (\tfrac{2}{3})^3 = \frac{160}{729}$$

$$P(X=4) = C_6^4 (\tfrac{1}{3})^4 (\tfrac{2}{3})^2 = \frac{60}{729}$$

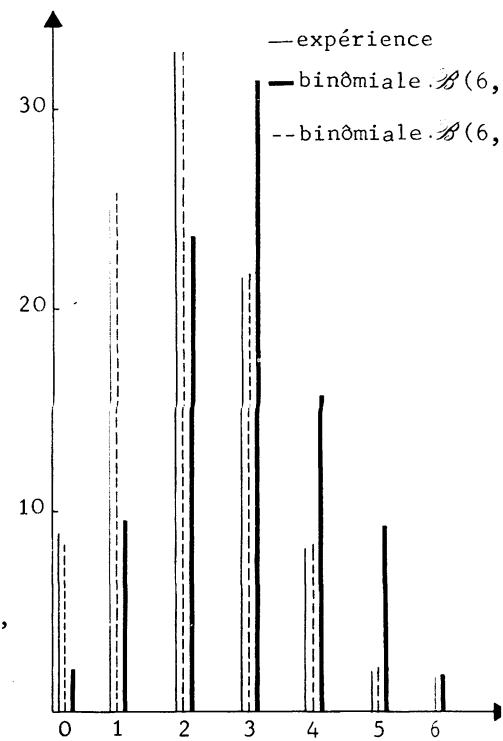
$$P(X=5) = C_6^5 (\tfrac{1}{3})^5 (\tfrac{2}{3}) = \frac{12}{729}$$

$$P(X=6) = C_6^6 (\tfrac{1}{3})^6 (\tfrac{2}{3})^0 = \frac{1}{729}$$

D'où les effectifs théoriques, sur 100 jets

$N(0) = 100P(X=0) = 8{,}78$

$N(1) = 100\,P(X=1) = 26{,}34$



N (2) = 100 P (X=2) = 32,92 N (5) = 100 P (X=5) = 1,65

N (3) = 100 P (X=3) = 21,95 N (6) = 100 P (X=6) = 0,14

N (4) = 100 P (X=4) = 8,23

b) Avec des pièces normales, on aurait obtenu :

$N(0) = 100\, C_6^0 (\frac{1}{2})^0 (\frac{1}{2})^6 = 1,56 = N(6)$.

$N(1) = 100\, C_6^1 (\frac{1}{2}) (\frac{1}{2})^5 = 9,38 = N(5)$

$N(2) = 100\, C_6^2 (\frac{1}{2})^2 (\frac{1}{2})^4 = 23,44 = N(4)$

$N(3) = 100\, C_6^3 (\frac{1}{2})^3 (\frac{1}{2})^3 = 31,25$

Conclusion :

Les résultats expérimentaux sont en très bon accord avec une loi binômiale où $p = \frac{1}{3}$, ce qui implique que les 6 pièces sont identiquement anormales. ■

---

VII. Deux espèces organiques sont caractérisées chacune par deux états génétiques A et a. Pour qu'une malformation apparaisse sur une espèce créée, il faut que, lors du croisement, deux états a fusionnent. On suppose que lorsqu'un type de génotype a été choisi, il élimine alors toute autre combinaison.

1°) Quelle est la probabilité pour que l'espèce générée soit atteinte de la malformation ?

2°) Quelle est la probabilité pour que sur 3 organismes créés lors d'un croisement, il y en ait au moins 1 qui présente l'anomalie ?

3°) Dans un lot de 4 croisements, donnant lieu chacun à la création de 3 organismes, quelle est la probabilité d'avoir 1 croisement où les 3 organismes sont anormaux ? Déterminer la valeur moyenne et l'écart-type du nombre de croisements à 3 malformations par lot.

SOLUTION

1°) Parmi les 4 combinaisons AA, Aa, aA et aa, une seule (a,a) est favorable à la malformation, d'où

$$p = \frac{1}{4} = 0{,}25.$$

2°) Si l'on appelle X la variable aléatoire correspondant au nombre d'organismes malformés, alors la probabilité d'obtenir k organismes malformés sur 3 créés est :

$$P\ (X = k) = C_3^k\ p^k\ (1 - p)^{3-k}$$

On en déduit :

$$P\ (X = k \geqslant 1) = 1 - P\ (X = 0) = 1 - C_3^0\ p^0 (1-p)^3 = 1 - (\frac{3}{4})^3 = \frac{37}{64}$$

3°) On s'intéresse maintenant à la statistique portant sur les lots de N = 4 croisements donnant chacun 3 organismes, et l'on considère comme nouvelle variable aléatoire le nombre Y de croisements donnant lieu à 3 malformations ($0 \leqslant Y \leqslant 4$).

La probabilité élémentaire est donc la probabilité d'avoir 3 organismes anormaux créés lors d'un croisement, soit d'après la 2ème question :

$$p' = P\ (X = 3) = C_3^3\ (\frac{1}{4})^3\ (\frac{3}{4})^0 = \frac{1}{64}$$

La probabilité de trouver, dans un lot de 4 croisements, un croisement où les 3 organismes sont anormaux est donc :

$$P\ (Y = 1) = C_4^1\ p'^1\ (1 - p')^3$$

$$P\ (Y = 1) = 4 \times \frac{1}{64} \times (\frac{63}{64})^3 = 0{,}06$$

Le nombre moyen de croisements à 3 malformations par lot de 4 croisements est donc :

$$\overline{x} = N p' = 4\ (\frac{1}{64}) = \frac{1}{16}$$

et l'écart-type

$$\sigma = \sqrt{N p' (1 - p')} = \sqrt{4 \times \frac{1}{64} \times \frac{63}{64}} \simeq 0{,}248$$

■

VIII. Soit X une variable aléatoire suivant une loi $\mathcal{B}(n,p)$.

Montrer que $P(X = k) \longrightarrow \lambda^k \frac{e^{-\lambda}}{k!}$ lorsque $n \to +\infty$, $p \to 0$, $np \to \lambda$

(on rappelle que $[1 - p]^n \to e^{-np}$ lorsque $n \to +\infty$ )

---

SOLUTION

On a $P(X = k) = C_n^k p^k (1 - p)^{n-k}$

$$= ([1 - p]^n \frac{(np)^k}{k!}) (\frac{n(n-1)\ldots(n-k+1)}{n^k} (1-p)^{-k})$$

or $[1 - p]^n \frac{(np)^k}{k!} \longrightarrow e^{-\lambda} \frac{\lambda^k}{k!}$ lorsque $n \to +\infty$, $p \to 0$ et $np \to \lambda$

$\frac{n(n-1)\ldots(n-k+1)}{n^k} (1-p)^{-k} \longrightarrow 1$ puisque $n \to +\infty$, $p \to 0$,

On a donc :

$P(X = k) \longrightarrow \frac{\lambda^k e^{-\lambda}}{k!}$ lorsque $n \to +\infty$, $p \to 0$ et $np \to \lambda$ ■

I. Une variable aléatoire X suit la loi de Poisson

$$P(X = k) = e^{-m} \frac{m^k}{k!}$$

1°) Vérifier que $\sum_{k=0}^{\infty} P(X = k) = 1$.

2°) Calculer l'espérance mathématique E (X).

3°) Calculer la variance V (X).

---

SOLUTION

1°) Nous devons vérifier que $\sum_{k=0}^{\infty} P(X = k) = 1$

$$\sum_{k=0}^{\infty} P(X = k) = \sum_{k=0}^{\infty} e^{-m} \frac{m^k}{k!} = e^{-m} \sum_{k=0}^{\infty} \frac{m^k}{k!}$$

Or $\sum_{k=0}^{\infty} \frac{m^k}{k!} = e^m$ ; on vérifie donc bien que $\sum_{k=0}^{\infty} P(X=k) = 1$.

2°) $E(X) = \sum_{k=0}^{\infty} k \frac{m^k}{k!} e^{-m} = \sum_{k=1}^{\infty} \frac{m^k}{(k-1)!} e^{-m}$ *(le terme k=0 est nul)*

Si l'on fait le changement d'indice k' = k - 1 alors

$$E(X) = \sum_{k'=0}^{\infty} \frac{m^{k'+1}}{k'!} e^{-m} = m \sum_{k'=0}^{\infty} \frac{m^{k'}}{k'!} e^{-m} = m \quad \text{(cf. 1°) )}$$

3°) L'équation (6.12) donne $V(X) = E(X^2) - [E(X)]^2$

$$E(X^2) = \sum_{k=0}^{\infty} k^2 \frac{m^k}{k!} e^{-m} = m \sum_{k=1}^{\infty} k \frac{m^{k-1}}{(k-1)!} e^{-m}$$

(le terme k = 0 est nul).

Posons $k - 1 = k'$

$$E(X^2) = m\sum_{k'=0}^{\infty} (k'+1)\frac{m^{k'}}{k'!}e^{-m} = m\sum_{k'=0}^{\infty} k'\frac{m^{k'}}{k'!}e^{-m} + m$$

D'après 2°)

$$\sum_{k'=0}^{\infty} k'\frac{m^{k'}}{k'!}e^{-m} = m$$

Par conséquent :

$$E(X^2) = m^2 + m.$$

On en déduit :

$$V(X) = m^2 + m - m^2 = m.$$ ■

Résumé : pour une variable aléatoire qui suit une loi de Poisson de paramètre m, on a $E(X) = V(X) = m$.

---

II. L'observation microscopique de plaquettes contenant des quantités égales d'une solution a permis de dresser le tableau suivant

| k = nombre de bactéries | n = nombre de plaquettes |
|---|---|
| 0 | 74 |
| 1 | 22 |
| 2 | 4 |

Comparer cette distribution expérimentale à celle déduite de la loi de Poisson de même valeur moyenne, applicable dans le cas d'une solution homogène. (On donne $e^{-0,3} \simeq 0,74$)

SOLUTION

La valeur moyenne du nombre de bactéries par plaquette est :

$$\lambda = \frac{0 \times 74 + 1 \times 22 + 2 \times 4}{100} = \frac{30}{100} = 0,3$$

Si l'on suppose que la variable aléatoire X ="nombre de bactéries par plaquette" suit une loi de Poisson de paramètre $m = \lambda$ alors

$$P\ (X = k) = \frac{m^k}{k!}\ e^{-m}$$

d'où les effectifs théoriques :

$$n\ (0) = 100 \times P\ (X=0) = 100 \times e^{-0,3} = 74 \text{ plaquettes}$$

$$n\ (1) = 100 \times P\ (X=1) = 100 \times 0,3\ e^{-0,3} = 22,2 \text{ plaquettes}$$

$$n(2) = 100 \times P\ (X=2) = 100\ \frac{0,3^2 e^{-0,3}}{2!} = 3,3 \text{ plaquettes}$$

On remarque que les deux distributions sont très proches. ■

---

III. A partir de la vente de 100 postes de télévision, ayant fonctionné le même nombre d'heures pendant une année, on a pu établir le tableau suivant reliant le nombre $n_i$ de postes vendus au nombre $k_i$ d'interventions du réparateur.

| Nbre d'interventions $k_i$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Nbre de postes $n_i$ | 61 | 30 | 7 | 2 | 0 |

1°) Calculer le nombre moyen d'interventions.

2°) Comparer les valeurs observées de $n_i$ avec celles déduites d'une loi de Poisson ayant même valeur moyenne.

SOLUTION

1°) $\lambda = \frac{\Sigma\, n_i\, k_i}{\Sigma\, n_i} = \frac{(61\times0)+(30\times1)+(7\times2)+(2\times3)+(0\times4)}{100} = \frac{50}{100} = 0{,}5$

2°) Si $P\ (X=k) = \frac{m^k}{k!}\, e^{-m}$ représente la probabilité pour que le réparateur intervienne k fois ($m = \lambda$) la fréquence correspondante sera obtenue en remarquant que $P\ (X=k) = \frac{n_k}{N}$. D'où les résultats théoriques, d'après la table 3 :

$n_0 = 100\ P\ (X=0) = 100 \times 0{,}606 = 60{,}6$ au lieu de 61
$n_1 = 100\ P\ (X=1) = 100 \times 0{,}304 = 30{,}4$ " 30
$n_2 = 100\ P\ (X=2) = 100 \times 0{,}076 = 7{,}6$ " 7
$n_3 = 100\ P\ (X=3) = 100 \times 0{,}013 = 1{,}3$ " 2
$n_4 = 100\ P\ (X=4) = 100 \times 0{,}001 = 0{,}1$ " 0 ∎

---

IV. L'observation de 200 personnes hospitalisées à montré que les résultats d'une série de tests sont tous négatifs pour 10 d'entre elles, et partiellement positifs pour les 190 autres.

1°) Quelle est la probabilité pour que tous les tests d'une personne soient négatifs ?

2°) Quelle est la valeur moyenne du nombre de tests positifs par individu sachant que ce nombre obéit à une loi de Poisson ?

3°) Construire l'histogramme de la distribution.

---

SOLUTION

1°) La probabilité pour qu'une personne ait tous ses tests négatifs est :

$$P = \frac{10}{200} = 0{,}05$$

2°) Si X est la variable aléatoire correspondant au nombre

de tests positifs par individu, alors la probabilité pour que X = k est :

$$P\ (X=k) = \frac{m^k\ e^{-m}}{k!}$$

où m est l'espérance mathématique de X.

Appliquons la relation à X = 0

$$P\ (X=0) = 0{,}05 = \frac{m^0\ e^{-m}}{0!} = e^{-m}$$

d'où

$$-\ m = \text{Log}\ 0{,}05 = \text{Log}\ \frac{5}{100} = \text{Log}\ \frac{1}{20}$$

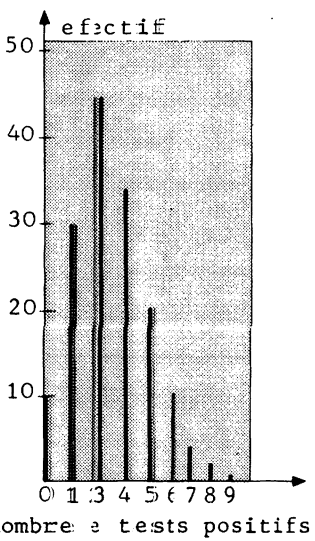
soit

$$m = \text{Log}\ 20 \simeq 2{,}996 \simeq 3.$$

3°) $P\ (X=k) = \frac{m^k}{k!}\ e^{-m}$

et $P\ (X=k+1) = \frac{m}{k+1}\ P\ (X=k)$, d'où

| i | P (X = i) | N(i) = 200 P(X=i) | Eff. cumulé croissant |
|---|---|---|---|
| 0 | 0,05 | 10 | 10 |
| 1 | 0,15 | 30 | 40 |
| 2 | 0,225 | 45 | 85 |
| 3 | 0,225 | 45 | 130 |
| 4 | 0,169 | ∿ 34 | 164 |
| 5 | 0,101 | ∿ 20 | 184 |
| 6 | 0,05 | 10 | 194 |
| 7 | 0,02 | 4 | 198 |
| 8 | 0,0075 | ∿ 2 | ∿ 200 |



V. A la fin d'une chaîne de fabrication de montres, on s'aperçoit au cours d'un dernier contrôle, que certaines montres présentent un défaut relativement mineur. On décide de supprimer ce contrôle systématique et de procéder à l'observation d'une série de boîtes contenant chacune 100 montres. Dans un lot de 200 boîtes, on en trouve 72 qui contiennent une montre défectueuse et 29 qui en contiennent 2.

1°) Quelle est la probabilité pour qu'une boîte ne contienne

a) qu'une seule montre défectueuse ?

b) que deux montres défectueuses ?

2°) En déduire la probabilité élémentaire pour qu'une montre de cette fabrication soit défectueuse, ainsi que le nombre moyen de montres défectueuses par boîte.

3°) Par quelle loi approchée peut-on calculer P (X) où X est le nombre de montres défectueuses par boîte ?

A.N. Calculer P (X=k) pour k = 0, 1, 2, 3, 4, 5.

---

SOLUTION

1°) Sur 200 boîtes, il y en a 72 qui contiennent 1 montre défectueuse et 29 qui en contiennent 2. On a donc :

a) $P_1 = \frac{72}{200} = 0{,}36 = P\ (X = 1)$

b) $P_2 = \frac{29}{200} = 0{,}145 = P\ (X = 2)$

2°) Si p = probabilité élémentaire pour qu'une montre soit défectueuse, q = 1-p est la probabilité pour qu'une montre soit parfaite.

Dans une boîte de 100 montres, on a donc

$$P\ (X = 1) = C^1_{100}\ p^1\ q^{99} = 0{,}36$$

$$P\ (X = 2) = C^2_{100}\ p^2\ q^{98} = 0{,}145$$

d'où

$$100\ p\ q^{99} = 0,36$$
$$4950\ p^2\ q^{98} = 0,145$$

soit en divisant membre à membre

$$\frac{100\ q}{4950\ p} = \frac{2}{99} \cdot \frac{(1-p)}{p} = \frac{0,36}{0,145}$$

On en déduit $p \simeq 0,008$ et $q \simeq 0,992$, et le nombre moyen de montres défectueuses par boîte est :

$$m = Np = 100 \times 0,008 = 0,8$$

3°) p étant très faible, et q voisin de l'unité, on a :

$$\sigma^2 = Npq \simeq Np = m = 0,8$$

On peut donc approximer P (X) par une loi de Poisson de paramètre m = 0,8. On trouve alors

$$P\ (X=0) = e^{-0,8} = 0,449$$
$$P\ (X=1) = \frac{0,8}{1} P\ (X=0) = 0,36 \text{ comparé à } 0,36 \text{ expérimentalement}$$
$$P\ (X=2) = \frac{0,8}{2} P\ (X=1) \simeq 0,144 \text{ comparé à } 0,145$$
$$P\ (X=3) = \frac{0,8}{3} P\ (X=2) \simeq 0,038$$
$$P\ (X=4) = \frac{0,8}{4} P\ (X=3) \simeq 0,008$$
$$P\ (X=5) = \frac{0,8}{5} P\ (X=4) \simeq 0,001$$

■

---

VI. On applique la loi de Poisson au nombre de personnes entrant dans un service de radiologie (variable aléatoire X). En moyenne il entre n personnes en une heure.

1°) Quelle est l'expression de la loi donnant la probabilité pour que k personnes se présentent dans un intervalle de temps T.

2°) Calculer la probabilité pour qu'il y ait 1 seule personne qui entre au bout du temps $T = \frac{2}{n}$.

3°) Calculer la probabilité pour qu'il y ait moins de quatre entrées dans un intervalle de temps $T = \frac{1}{n}$.

4°) Ce service ne peut recevoir plus de 3 personnes par heure sans arriver à la saturation. Quel est le pourcentage de clients qui devront revenir si le nombre moyen d'entrants par heure est 2 ?

On donne $e^{-2} = 0,135$ ; $e^{-1} = 0,368$.

SOLUTION

1°) Pendant le temps T, il entre, en moyenne m = nT personnes

donc $P(X=k) = (nT)^k \frac{e^{-nT}}{k!}$

2°) $P(X=1) = (n \times \frac{2}{n})^1 \times \frac{e^{-(n \times \frac{2}{n})}}{1!} = 2e^{-2} = 0,270$

3°) $P(X=k < 4) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$ avec = m 1.

Par récurrence :

$P(X=0) = e^{-1} = 0,368$ ; $P(X=1) = \frac{1}{1} P(X=0) = 0,368$

$P(X=2) = \frac{1}{2} P(X=1) = 0,184$ ; $P(X=3) = \frac{1}{3} P(X=2) = 0,061$

D'où

$P(X=k < 4) = 0,981.$

4°) $P(X=k) = (2)^k \frac{e^{-2}}{k!}$ . Le pourcentage cherché correspond à $P(X=k > 3)$

$P(X=k > 3) = 1 - P(X=k \leq 3)$

$= 1 - [P(X=0) + P(X=1) + P(X=2) + P(X=3)]$

$= 1 - (0,135 + 0,270 + 0,270 + 0,180) = 0,145$

Il y aura donc 14,5 % de clients qui devront revenir.

■

I. En utilisant le changement de variables $x = \rho \cos \theta$ et $y = \rho \sin \theta$, ce qui entraîne $dx\,dy = \rho\, d\rho\, d\theta$, calculer

1°) $I_1 = \int_{-\infty}^{+\infty} e^{-x^2/2}\, dx$

2°) Montrer que $I_2 = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx = 1$

---

SOLUTION

1°) On calcule $I_1 \times I_1 = \left(\int_{-\infty}^{+\infty} e^{-x^2/2}\, dx\right)\left(\int_{-\infty}^{+\infty} e^{-y^2/2}\, dy\right)$

soit

$$I_1^2 = \iint e^{-\frac{(x^2+y^2)}{2}}\, dx\, dy$$

l'intégrale double étant étendue à tout le plan.

*En effectuant le changement de variable proposé*, on obtient :

$$I_1^2 = \int_0^{2\pi} d\theta \int_0^{\infty} e^{-\rho^2/2}\, \rho\, d\rho$$

d'où

$$I_1^2 = [\theta]_0^{2\pi} \times \left[-e^{-\rho^2/2}\right]_0^{\infty} = 2\pi$$

par conséquent :

$$I_1 = \sqrt{2\pi}$$

2°) Pour calculer $I_2 = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx$, on fait le changement de variable $t = \frac{x-m}{\sigma}$ qui entraîne $dt = \frac{1}{\sigma} dx$, d'où

$$I_2 = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-t^2/2} . \sigma . dt = \frac{1}{\sqrt{2\pi}} I_1$$

On obtient donc

$I_2 = 1$. ■

---

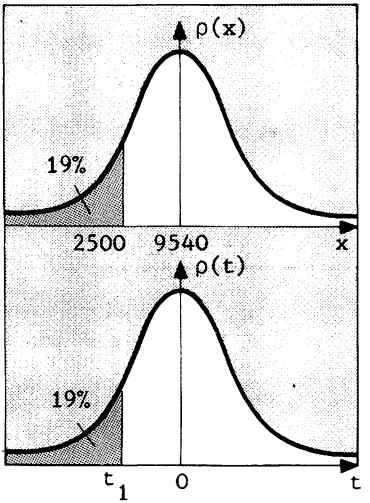
II. En 1961, le professeur Lampman a étudié la pauvreté dans les sociétés industrialisées. Il aboutit à la conclusion que 19 % de la population américaine peut être considérée comme pauvre si le seuil de pauvreté, pour le revenu annuel d'une famille citadine de quatre personnes, est fixé à 2 500 dollars.

1°) En supposant que la répartition des revenus annuels suit une loi normale de valeur moyenne 9 540 dollars, quel est l'écart-type de cette distribution ?

2°) Si seulement 10 % de la population peut être considérée comme riche, entre quelles limites doit se situer le revenu annuel d'une famille citadine pour qu'elle puisse être considérée comme ayant un revenu x moyen tel que $x_1 < x < x_2$, $x_1$ et $x_2$ étant définis par $P(x \leqslant x_1) = 0{,}19$ et $P(x \geqslant x_2) = 0{,}10$?

SOLUTION

1°)




On peut écrire

$$P(x < 2500) = 0,19$$

soit, si l'on pose

$$(1) \quad t = \frac{x - 9\,540}{\sigma}$$

$$P\ (t < t_1) = 0,19$$

avec

$$t_1 = \frac{2\,500 - 9\,540}{\sigma}$$

Comme

$$P\ (t < t_1) = P\ (< t < 0) - P\ (t_1 < t < 0)$$

on déduit

$$P\ (t_1 < t < 0) = 0,50 - 0,19 = 0,31$$

En utilisant la fonction G (t) définie par (6.30), on a

$$G\ (-t_1) = 0,31$$

La table 3 donne alors $-t_1 = 0,88$ ou $t_1 = -0,88$. Par suite, la relation (1) permet d'écrire

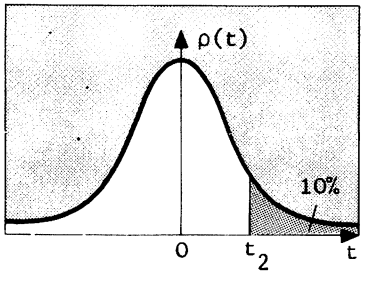
$$\frac{9\,540 - 2\,500}{\sigma} = 0,88$$

d'où

$$\sigma = \frac{9\,540 - 2\,500}{0,88} = 8\,000 \text{ dollars}$$

■

2°)




$$G\ (t_2) = 0,50 - 0,10 = 0,40$$

d'où

$$t_2 = 1,28$$

et comme

$$t_2 = \frac{x_2 - 9\,540}{8\,000},$$

on déduit :

$$x_2 = 8\,000 \times 1,28 + 9\,540$$

soit

$x_2$ = 19 780 dollars

Les limites sont donc

2 500 et 19 780 dollars. ■

---

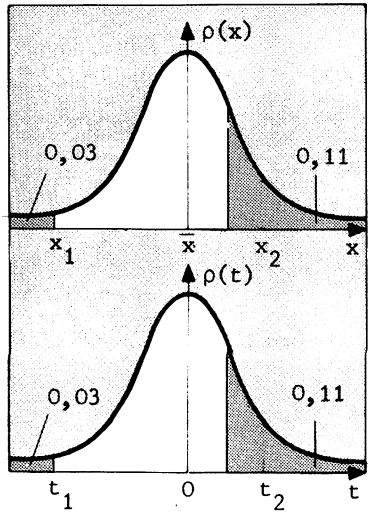
III. En 1965, dans la région parisienne, 11 % des revenus individuels étaient supérieurs à 20 000 F et 3 % des revenus inférieurs à 3 000 F (Enquête de l'I.N.S.E.E.). En supposant que la loi de répartition des revenus suit une loi normale,

1°) Quel est le revenu individuel moyen ?

2°) Quel est le pourcentage d'individus dont le salaire est compris entre 5 000 F et 10 000 F ?

---

SOLUTION




1°) D'après l'enquête, on a

$P\ (x > 20\,000) = 0{,}11$

$P\ (x < 3\,000) = 0{,}03$

En utilisant la loi centrée et réduite obtenue par le changement de variable

$$t = \frac{x - \bar{x}}{\sigma}$$

on a

$P\ (t > t_2) = 0{,}11$

et $P\ (t < t_1) = 0{,}03$

On remarque d'une part que

$P\ (t > t_2) = 0{,}5 - P\ (0 < t < t_2)$

et d'autre part

$P\ (t < t_1) = 0{,}5 - P\ (t_1 < t < 0)$

En utilisant la fonction G (t) définie par (6.30) on a :

$G(t_2) = 0,5 - 0,11 = 0,39$

et $G(-t_1) = 0,5 - 0,03 = 0,47$

d'où, d'après la table 3

$t_2 = 1,23$ et $-t_1 = 1,88$

On a donc

$$\frac{20\,000 - \bar{x}}{\sigma} = 1,23$$

et

$$\frac{3\,000 - \bar{x}}{\sigma} = -1,88$$

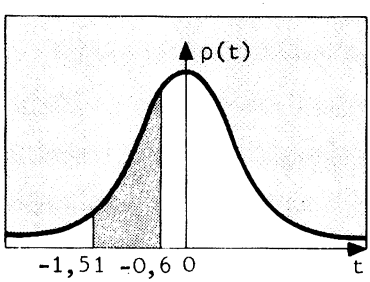
La résolution de ce système donne :

$\bar{x} \simeq 13\,277$ F

et $\sigma \simeq 5\,466$ F

Le revenu individuel moyen est donc égal à 13 277 F.

2°)




On doit calculer

$$P(5\,000 < x < 10\,000) = P(t_1 < t < t_2)$$

avec

$$t_1 = \frac{5\,000 - 13\,277}{5\,466} = -1,51$$

et

$$t_2 = \frac{10\,000 - 13\,277}{5\,466} = -0,60$$

Par symétrie on a :

$P(-1,51 < t < -0,60) = P(0,60 < t < 1,51)$

$= P(0 < t < 1,51) - P(0 < t < 0,6)$

Donc

$$P(5\,000 < x < 10\,000) = G(1,51) - G(0,6)$$

$$= 0,4345 - 0,2257 = 0,2088$$

Il y a donc 20,88 % d'individus qui ont un salaire compris entre 5 000 F et 10 000 F. ■

IV. Une usine est chargée de la fabrication de miroirs pour des appareils optiques de très haute qualité. Pour cela, elle traite des supports de verre de forme parallélépipédique dont les dimensions doivent être :

$x = 80{,}000 \pm 0{,}010$ cm

$y = 50{,}000 \pm 0{,}010$ cm

$z = 6{,}000 \pm 0{,}005$ cm

On contrôle une série de ces supports et on trouve :

$\bar{x} = 80{,}005$ $\qquad \sigma_x = 0{,}005$

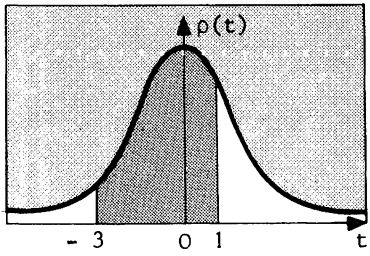
$\bar{y} = 50{,}000$ $\qquad \sigma_y = 0{,}005$

$\bar{z} = 6{,}001$ $\qquad \sigma_z = 0{,}002$

En supposant les distributions de x, y et z normales et indépendantes, déterminer le pourcentage de supports rejetés.

---

SOLUTION




Le pourcentage de supports acceptés est égal à la probabilité d'avoir à la fois

$79{,}99 \leqslant x \leqslant 80{,}01$

$49{,}99 \leqslant y \leqslant 50{,}01$

$5{,}995 \leqslant z \leqslant 6{,}005$

a) En posant

$$t = \frac{x - \bar{x}}{\sigma_x} = \frac{x - 80{,}005}{0{,}005}$$

la 1ère inégalité devient

$$-3 \leqslant t \leqslant 1$$

La probabilité recherchée sur x est donc

$$P(-3 \leqslant t \leqslant 1) = P(-3 \leqslant t \leqslant 0) + P(0 \leqslant t \leqslant 1)$$
$$= G(3) + G(1) = 0{,}4987 + 0{,}3413$$

soit

$$P_x = 0,84$$

b) En posant

$$t = \frac{y - \bar{y}}{\sigma_y} = \frac{y - 50}{0,005}$$

la 2ème inégalité devient

$$-2 \leqslant t \leqslant +2$$

$$P(-2 \leqslant t \leqslant 2) = G(2) + G(2)$$

$$P_y = 0,9544.$$

c) De la même manière, par rapport à z, on aura

$$-3 \leqslant t \leqslant 2$$

$$P_z = G(3) + G(2) = 0,4987 + 0,4772 = 0,9759$$

Le pourcentage de supports acceptés est donc :

$$P_a = 100\ (P_x \times P_y \times P_z) = 78,24\ \%$$

D'où le pourcentage de supports rejetés :

$$P_r = 100 - P_a = 21,76\ \%.$$ ■

---

V. L'I.N.S.E.E. a réalisé une enquête portant sur le nombre d'enfants par ménage depuis 1945. Le résultat est, qu'après onze ans de mariage, sur 10 000 couples :

2 200 sont sans enfant

2 100 ont un enfant

2 400 ont deux enfants

1 600 ont trois enfants

1 700 en ont au moins quatre, la moitié d'entre eux ayant exactement quatre enfants.

1°) Quelle est la probabilité, pour un couple, d'avoir

a) 1 enfant

b) au moins deux enfants

c) au plus quatre enfants ?

2°) On considère 4 couples. On appelle P (n) la probabilité d'avoir, parmi ces 4 couples, n couples sans enfant, après onze ans de mariage.

Donner l'expression de P (n)

Calculer P (0), P (1), P (2), P (3) et P (4).

3°) On effectue plusieurs enquêtes portant chacune sur 1 000 couples, onze ans après leur mariage, et l'on considère comme variable aléatoire, le nombre n de couples sans enfant observé par enquête.

a) Quelle loi de probabilité suit n ? Déterminer le nombre n moyen observé par enquête ainsi que l'écart-type.

b) Par quelle loi de probabilité peut-on approximer cette distribution ?

c) Déterminer $n_o$ tel que la probabilité d'avoir $n > n_o$ soit de 10 %.

---

SOLUTION

1°) On a $P(n) = \frac{n}{N}$, avec n = 2 100 et N = 10 000, d'où

a) $P(1) = \frac{2\ 100}{10\ 000} = 0{,}21$

b) $P(n \geqslant 2) = 1 - P(0) - P(1) = 1 - \frac{2\ 200}{10\ 000} - \frac{2\ 100}{10\ 000}$

$P(n \geqslant 2) = 1 - 0{,}22 - 0{,}21$

$P(n \geqslant 2) = 0{,}57$

c) $P(0) + P(1) + P(2) + P(3) + P(4) = 0{,}22+0{,}21+0{,}24+0{,}16+0{,}085$

$P(4) = 0{,}085$ puisqu'il y a $\frac{1\ 700}{2} = 850$ couples qui ont exactement 4 enfants

$P(n \leqslant 4) = 0{,}915$.

On aurait aussi bien pu écrire

$P(n \leqslant 4) = 1 - P(n > 4)$

avec $P(n > 4) = \frac{850}{10\ 000} = 0{,}085$

d'où $P(n \leqslant 4) = 1 - 0{,}085 = 0{,}915$.

2°) La probabilité, pour un couple, de ne pas avoir d'enfant après onze ans de mariage est p = 0,22, donc q = 1-p = 0,78 représente la probabilité d'avoir au moins 1 enfant.

Les événements étant indépendants et équiprobables, on en déduit :

$$P(n) = C_N^n \, p^n \, q^{N-n} \qquad \text{avec} \qquad N = 4$$

d'où

$$P(0) = C_4^0 \cdot 0{,}22^0 . 0{,}78^4 = 0{,}370$$

$$P(1) = C_4^1 \cdot 0{,}22^1 . 0{,}78^3 = 0{,}418$$

$$P(2) = C_4^2 \cdot 0{,}22^2 . 0{,}78^2 = 0{,}177$$

$$P(3) = C_4^3 \cdot 0{,}22^3 . 0{,}78^1 = 0{,}033$$

$$P(4) = C_4^4 \cdot 0{,}22^4 . 0{,}78^0 = 0{,}002$$

On vérifie que $\sum_{n=0}^{4} P(n) = 1$

3°) a) Sur 1 000 couples on aura

$$P(n) = C_{1\,000}^n \, (0{,}22)^n \cdot (0{,}78)^{1\,000-n} \qquad \text{(loi binômiale)}$$

Par suite

$$\overline{n} = N \times p = 1\,000 \times 0{,}22 = 220$$

et $\sigma = \sqrt{N \times p \times q} = \sqrt{220 \times 0{,}78} \simeq 13$

b) La probabilité élémentaire étant faible, la population *très grande* et $\overline{n} > 20$, *on peut remplacer* la loi binômiale par une loi normale.

c) $P(n \geqslant n_o) = 0{,}10$

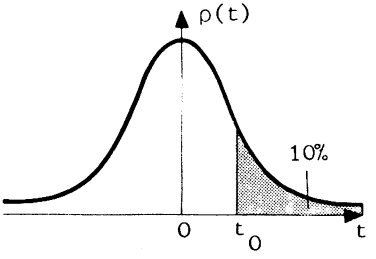
ou $P(t \geqslant t_o) = 0{,}10$

avec $t_o = \dfrac{n_o - \overline{n}}{\sigma} = \dfrac{n_o - 220}{13}$

$P(t \geqslant t_o) = 0{,}50 - P(0 < t < t_o) = 0{,}10$

$P(0 < t < t_o) = G(t_o)$

$= 0{,}50 - 0{,}10 = 0{,}40$



La table 3 donne :

$$t_o = 1{,}28$$

d'où $\dfrac{n_o - 220}{13} = 1{,}28$

et $n_o = 13 \times 1{,}28 + 220 \simeq 236$ couples

■

---

V. Une enquête portant sur la taille de 200 individus a donné les résultats suivants :

| taille en cm | $x_i$ = centre de classe | $n_i$ = effectif |
|---|---|---|
| [ 150 - 155[ | 152,5 | 7 |
| [ 155 - 160[ | 157,5 | 14 |
| [ 160 - 165[ | 162,5 | 24 |
| [ 165 - 170[ | 167,5 | 37 |
| [ 170 - 175[ | 172,5 | 42 |
| [ 175 - 180[ | 177,5 | 35 |
| [ 180 - 185[ | 182,5 | 23 |
| [ 185 - 190[ | 187,5 | 13 |
| [ 190 - 195[ | 192,5 | 5 |

1°) Déterminer la valeur moyenne et l'écart-type de cette distribution.

2°) Calculer les densités de probabilité $\rho\ (x_i)$ déduites d'une loi normale de même valeur moyenne et de même écart-type.

3°) Représenter sur un même diagramme, en choisissant correctement les échelles des ordonnées, la distribution expérimentale $n_i\ (x_i)$ et la distribution théorique $\rho_i\ (x_i)$.

4°) Quel est le pourcentage théorique des tailles comprises entre 152,5 et 167,5 ? Comparer avec le pourcentage expérimental déduit de l'enquête en supposant une répartition linéaire dans les classes.

5°) D'après la formule de Lorents, le poids idéal en kg, pour une taille x en cm est :

$$M = (x - 100) - \left( \frac{x - 150}{4} \right)$$

Calculer, pour un échantillon de 500 individus dont la distribution des tailles a pour moyenne 170 cm et pour écart-type 10 cm, le poids moyen et l'écart-type de la distribution des poids.

6°) Voulant constituer une équipe de rugby à 15 joueurs, on convient de ne garder que les 15 plus lourds. Quelle sera la taille du plus petit joueur ?

On supposera, dans cette question, que les 500 individus ont, par rapport à leur taille, le poids idéal.

---

SOLUTION

1°) Le changement de variable

$$t_i = \frac{x_i - 172,5}{10}$$

permet de dresser le tableau de la page suivante.

On en déduit :

$$\overline{t} = \frac{\Sigma\, n_i\, t_i}{\Sigma\, n_i} = -\,0,0375$$

et $$\overline{t^2} = \frac{\Sigma\, n_i\, t_i^2}{\Sigma\, n_i} = 0,869$$

Donc

$$\sigma_t = \sqrt{\overline{t^2} - \overline{t}^2} = 0,931$$

| $x_i$ | $n_i$ | $t_i$ | $n_i\ t_i$ | $n_i\ t_i^2$ |
|---|---|---|---|---|
| 152,5 | 7 | - 2 | - 14 | 28 |
| 157,5 | 14 | - 1,5 | - 21 | 31,5 |
| 162,5 | 24 | - 1 | - 24 | 24 |
| 167,5 | 37 | - 0,5 | - 18,5 | 9,25 |
| 172,5 | 42 | 0 | 0 | 0 |
| 177,5 | 35 | 0,5 | 17,5 | 8,75 |
| 182,5 | 23 | 1 | 23 | 23 |
| 187,5 | 13 | 1,5 | 19,5 | 29,25 |
| 192,5 | 5 | 2 | 10 | 20 |
| | $\Sigma\ n_i = 200$ | | $\Sigma\ n_i\ t_i = -7,5$ | $\Sigma\ n_i\ t_i^2 = 173,75$ |

et comme

$$x_i = 10\ t_i + 172,5$$

$$\bar{x} = 10\ \bar{t} + 172,5 \qquad \text{et} \qquad \sigma_x = 10\ \sigma_t$$

soit :

$$\bar{x} = 172,125 \text{ cm}$$

$$\sigma = 9,31 \text{ cm}$$

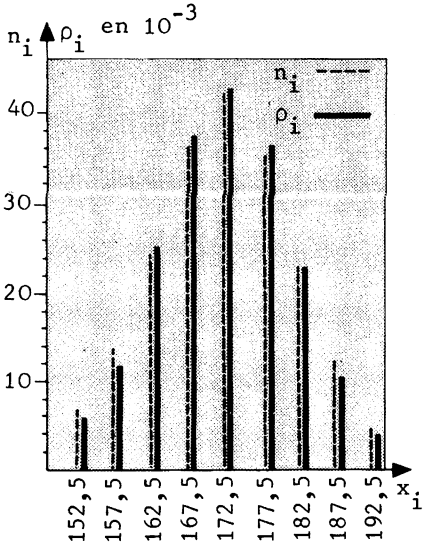
2°) La densité de probabilité est donnée par la formule :

$$\rho\ (x_i) = \frac{1}{\sigma_x\ \sqrt{2\pi}}\ e^{-\frac{(x_i - \bar{x})^2}{2\ \sigma_x^2}}$$

| $x_i$ | $\frac{1}{2}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^2$ | $\rho\ (x_i)$ en $10^{-3}$ | $n_i$ |
|---|---|---|---|
| 152,5 | 2,22 | 5 | 7 |
| 157,5 | 1,23 | 12 | 14 |
| 162,5 | 0,53 | 25 | 24 |
| 167,5 | 0,12 | 38 | 37 |
| 172,5 | 0,0008 | 43 | 42 |
| 177,5 | 0,17 | 36 | 35 |
| 182,5 | 0,62 | 23 | 23 |
| 187,5 | 1,36 | 11 | 13 |
| 192,5 | 2,39 | 4 | 5 |

3°) La figure montre le diagramme en bâtons de la distribution expérimentale $n_i$ (----) et de la distribution théorique $\rho_i$ (——).

On voit que la loi normale constitue une très bonne approximation de la distribution observée.



4°) Le pourcentage théorique correspond à la probabilité

$$P\ (152{,}5 \leqslant x \leqslant 167{,}5) \simeq P\ (-2{,}11 \leqslant t \leqslant -0{,}54)$$

où l'on pose

$$t = \frac{x - 172{,}125}{9{,}31}\ .$$

$$P\ (-2{,}11 \leqslant t \leqslant -0{,}54) = P\ (-2{,}11 \leqslant t \leqslant 0) - P\ (-0{,}54 \leqslant t \leqslant 0)$$
$$= G\ (2{,}11) - G\ (0{,}54)$$
$$= 0{,}4826 - 0{,}2054$$
$$P\ (152{,}5 \leqslant x \leqslant 167{,}5) = 0{,}2772$$

Il y a donc 27,72 % d'individus qui ont une taille comprise entre 152,5 et 167,5 cm.

Calcul du pourcentage expérimental :

Entre 150 et 155, l'effectif est 7. En supposant une distribution linéaire dans chaque classe, on aura entre 152,5 et 155 un effectif de :

$$\frac{7 \times 2{,}5}{5} = 3{,}5.$$

Pour la même raison, on aura entre 165 et 167,5 un effectif

$$\frac{37 \times 2{,}5}{5} = 18{,}5.$$

Par conséquent, l'effectif entre 152,5 et 167,5 est :

$$3{,}5 + 14 + 24 + 18{,}5 = 60$$

D'où le pourcentage cherché

$$\frac{60}{200} = 0{,}30 \quad \textit{soit 30 \% (théoriquement, on a trouvé 27,72 \%)}$$

5°) $M = x - \frac{x}{4} - 100 + \frac{150}{4} = 0{,}75\ x - 62{,}5$

d'où

$$\overline{M} = 0{,}75\ \overline{x} - 62{,}5 = 66{,}59 \text{ kg}$$
$$\sigma_M = 0{,}75\ \sigma_x = 6{,}98 \text{ kg}$$

6°) Si l'on ne garde que 15 joueurs sur 500, cela représente un pourcentage de 3 %. Il s'agit donc de déterminer d'abord $M_{min}$ pour que $P\ (M \geqslant M_{min}) = 0{,}03$

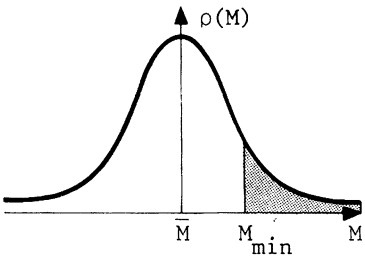
soit pour la loi centrée et réduite

$$P\ (t \geqslant t_m) = 0{,}03$$

$$P\ (t \geqslant t_m) = 0{,}5 - G\ (t_m)$$

Par conséquent

$$G\ (t_m) = 0{,}5 - 0{,}03 = 0{,}47$$




La table 3 donne

$$t_m = 1{,}88$$

Or

$$t_m = \frac{M_{min} - \overline{M}}{\sigma_M}$$

d'où le poids du joueur le plus léger :

$$M_{min} = 1{,}88 \times 6{,}98 + 66{,}59 = 79{,}7 \text{ kg}$$

On en déduit la taille du plus petit joueur :

$$x = \frac{M_{min} + 62{,}5}{0{,}75} = 189{,}6 \text{ cm.}$$

■

# 7. Les tests statistiques

## INTRODUCTION

### Statistique descriptive et statistique inductive

Rappelons que la statistique descriptive est un ensemble de méthodes qui permettent d'ordonner et de classer les données, de les réduire ensuite à un nombre limité de paramètres caractéristiques (moyenne, variance, etc.) susceptible de décrire la distribution du caractère étudié dans une population donnée. Les principes de la représentation et de la réduction des données ont été exposés dans les chapitres 2 et 3.

La statistique inductive est plus ambitieuse dans la mesure où elle recherche les principes permettant de déduire des résultats obtenus sur un échantillon limité, une généralisation à l'ensemble de la population d'où est extrait cet échantillon et qui est généralement inaccessible à l'enquête ou à la mesure. On est alors amené à formuler des hypothèses dont on vérifie la validité à l'aide de certaines épreuves ou tests statistiques. Cela permet de prendre une décision dépendant nécessairement du risque d'erreur adopté dû au fait que les données sont seulement partielles.

L'objet de ce chapitre est de présenter quelques tests statistiques. Nous nous limiterons essentiellement à la résolution de quelques problèmes des types suivants :

. Estimation : estimer les paramètres (moyenne, écart-type) qui caractérisent une population, connaissant les paramètres d'un échantillon extrait de cette population.

. Conformité : déterminer si un échantillon peut être considéré comme représentatif d'une population.

. Homogénéité : déterminer si les différences observées entre deux échantillons sont dues au hasard ou si elles sont significatives (non dues au hasard).

. Ajustement : vérifier si une distribution expérimentale peut être ajustée à une distribution théorique.

# A. ECHANTILLONNAGE

La théorie de l'échantillonnage a pour objet l'étude des relations qui existent entre la distribution d'un caractère dans une population dite population-mère, et les distributions de ce caractère dans tous les différents échantillons prélevés dans cette population.

Pour que ces relations soient valables, il faut que l'échantillon soit prélevé d'une manière aléatoire, c'est à dire que tous les individus de la population aient la même chance d'être prélevés. On y arrive au moyen d'un tirage au sort par exemple, ou encore en utilisant des listes de nombres aléatoires.

L'échantillonnage est dit exhaustif si l'individu n'est pas remis dans la population après avoir été prélevé. Il est dit non-exhaustif dans le cas contraire. Lorsque la population est très grande, on peut considérer que les deux notions sont équivalentes puisqu'un prélèvement exhaustif ne modifie pratiquement pas l'effectif de la population.

## I. DISTRIBUTION DES MOYENNES

Soit X un caractère quantitatif étudié dans une population d'effectif N. La distribution de X dans cette population sera notée $(N, M, \sigma)$ où $M = E(X)$ est la moyenne, et $\sigma = \sigma(X)$ l'écart-type, du caractère X.

Soit $X_i$ le même caractère étudié dans un échantillon i de taille n. La distribution de $X_i$ dans cet échantillon sera notée $(n, m_i, \sigma_i)$, où $m_i = E(X_i)$ et $\sigma_i = \sigma(X_i)$. On suppose que les échantillons ont tous la même taille n.

1) Echantillonnage non-exhaustif

Considérons l'ensemble de tous les échantillons possibles de taille n pouvant être prélevés dans la population-mère, d'une manière non-exhaustive, et soit κ le nombre de ces échantillons.

On appelle distribution d'échantillonnage des moyennes l'ensemble des moyennes des différents échantillons, soit

$$\{m_1, m_2, m_3, \ldots m_i, \ldots m_k\} \qquad (7.1)$$

On introduit ainsi un nouveau caractère m qui associe la valeur $m_i$ à l'échantillon i. La distribution de m est caractérisée par [ k, E (m), σ (m) ].

On peut montrer que :

$$E(m) = E(X) = M \qquad (7.2)$$

$$V(m) = \frac{V(X)}{n} \qquad (7.3)$$

$$\text{d'où } \sigma(m) = \frac{\sigma(X)}{\sqrt{n}} = \frac{\sigma}{\sqrt{n}} \qquad (7.4)$$

Considérons en effet un échantillon aléatoire $(I_1, I_2, \ldots, I_n)$ tiré au hasard de manière non-exhaustive. Si α et β sont deux individus de la population-mère, on a :

$$P(I_\ell = \alpha) = \frac{1}{N} \; ; \; P(I_\ell = \alpha \text{ et } I_b = \beta) = \frac{1}{N^2} \qquad (\ell \neq b)$$

On en déduit que :

$$E(X(I_\ell)) = E(X) \; ; \; E(X(I_\ell)^2) = E(X^2)$$

$$E[X(I_\ell)\, X(I_b)] = E[X(I_\ell)]\; E[X(I_b)]$$

Par suite :

$$E(m) = E\left[\frac{1}{n}\sum_{\ell=1}^{n} X(I_\ell)\right] = \frac{1}{n}\sum_{\ell=1}^{n} E(X(I_\ell)) = E(X) = M$$

$$V(m) = E(m^2) - M^2 = E\left[\frac{1}{n^2}\sum_{\ell=1}^{n} X(I_\ell)^2 + \frac{1}{n^2}\sum_{\ell\neq b}^{n} X(I_\ell)\, X(I_b)\right] - M^2$$

$$= \frac{1}{n} E(X^2) + \left[\frac{n(n-1)}{n^2} - 1\right] M^2 = \frac{1}{n} V(X).$$

2) Echantillonnage exhaustif

En suivant un raisonnement analogue, mais en écrivant cette fois-ci

$$P\ (I_\ell = \alpha \text{ et } I_b = \beta) = \frac{1}{N\ (N-1)}$$

on montre que les expressions de E (m) et $\sigma$ (m) deviennent :

$$E\ (m) = M \tag{7.5}$$

$$\sigma\ (m) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N - n}{N - 1}} \tag{7.6}$$

où N est l'effectif total de la population-mère. On voit que lorsque N est très grand comparé à n, l'expression (7.6) est équivalente à (7.4).

## II. DISTRIBUTION DES FREQUENCES

Supposons que dans une population composée de N éléments, le caractère étudié X ne puisse prendre que les deux valeurs 1 et 0. On désigne par p la proportion d'éléments de caractère 1 et par q la proportion des éléments de caractère 0. (On a $0 < p < 1$ et $q = 1 - p$). La distribution d'un tel caractère dans cette population est caractérisée par une moyenne et un écart-type donnés par :

$$E\ (X) = M = p \tag{7.7}$$

$$\sigma\ (X) = \sigma = \sqrt{pq} \tag{7.8}$$

de sorte qu'elle peut être notée $(N, p, \sqrt{pq})$.

On prélève dans cette population tous les échantillons de taille n et on détermine pour chaque échantillon i la proportion d'éléments dont le caractère a la valeur 1. On définit ainsi un nouveau caractère f qui associe à chaque échantillon i la fréquence $f_i$.

On appelle distribution d'échantillonnage des fréquences l'ensemble des fréquences $f_i$ des différents échantillons

$$\{f_1, f_2, f_3 \ldots f_i \ldots f_k\} \tag{7.9}$$

Cette distribution de f peut être notée [k, E (f), σ (f)] où k est le nombre total d'échantillons, E (f) et σ (f) sont respectivement la moyenne et l'écart-type de f.

1) Echantillonnage non-exhaustif

Les expressions (7.2) et (7.4), en tenant compte de M = p et $\sigma = \sqrt{pq}$ donnés par (7.7) et (7.8) permettent d'écrire :

$$E(f) = p \tag{7.10}$$

$$\sigma(f) = \sqrt{\frac{pq}{n}} \tag{7.11}$$

2) Echantillonnage exhaustif

Les expressions (7.5) et (7.6) fournissent

$$E(f) = p \tag{7.12}$$

$$\sigma(f) = \sqrt{\frac{pq}{n}}\sqrt{\frac{N-n}{N-1}} \tag{7.13}$$

où N est l'effectif de la population-mère.

## III. AUTRES DISTRIBUTIONS D'ECHANTILLONNAGE

On peut définir d'autres distributions d'échantillonnage que les distributions de m et f. Le caractère peut être la médiane, le mode, l'écart-type, etc. ou tout autre paramètre susceptible de varier d'un échantillon à l'autre. Les deux distributions qui suivent - celle de t et celle de $\chi^2$ - sont utilisées par exemple, l'une lorsque les échantillons sont petits (n < 30, cf. 7 B III), l'autre dans des problèmes d'ajustement d'une distribution expérimentale à une distribution théorique (cf. 7 D).

### 1. Distribution de t

Soit une distribution (N, M, σ) d'un caractère X dans une population, qui suit une loi normale $\mathscr{N}$ (M, σ). On consi-

dère tous les échantillons de taille n pouvant être prélevés dans cette population, et caractérisés par $(n, m_i, \sigma_i)$.

On introduit un nouveau caractère t donné par l'écart réduit

$$t = \frac{m - M}{\sigma / \sqrt{n}} \tag{7.14}$$

qui associe à chaque échantillon i l'écart réduit $t_i$. On définit ainsi une nouvelle distribution d'échantillonnage, dite distribution de t

$$\{t_1, t_2, t_3 \cdots t_i, \ldots, t_k\} \tag{7.15}$$

où k est le nombre d'échantillons.

## 2. Distribution de $\chi^2$

On considère encore une population normale $(N, M, \sigma)$ et tous ses échantillons $(n, m_i, \sigma_i)$. On calcule pour chaque échantillon i, le paramètre

$$\chi_i^2 = \frac{\sum_{j=1}^{n} (x_j^i - m_i)^2}{\sigma^2} \tag{7.16}$$

où $x_j^i$ est la valeur du caractère du $j^{\text{ième}}$ individu de l'échantillon i. On définit ainsi une nouvelle distribution d'échantillonnage dite de $\chi^2$

$$\{\chi_1^2, \chi_2^2, \chi_3^3, \cdots \chi_i^2, \cdots \chi_k^2\} \tag{7.17}$$

## B. ESTIMATION

Si l'échantillonnage étudie les relations qui existent entre une population et tous ses échantillons de même taille n, l'estimation se préoccupe de la représentativité de la population par un échantillon. Il s'agit essentiellement d'attribuer une valeur à un paramètre inconnu de la population-mère à partir de la connaissance d'un échantillon extrait de cette population.

On peut chercher à attribuer à ce paramètre une valeur unique (estimation ponctuelle) ou un intervalle susceptible de recouvrir sa valeur inconnue (estimation par un intervalle de confiance).

### I. ESTIMATION PONCTUELLE

Considérons une distribution dans une population-mère (N, M, $\sigma$) et la distribution du même caractère dans un échantillon i (n, $m_i$, $\sigma_i$) extrait de cette population. On suppose que $m_i$ et $\sigma_i$ sont connus, et on cherche M et $\sigma$.

Il est évident qu'en général, l'estimation d'un paramètre inconnu à partir de sa valeur observée sur l'échantillon ne peut constituer qu'une approximation. On considère que certaines conditions sont requises pour qu'un paramètre de l'échantillon puisse servir d'estimateur :

1) lorsque la taille de l'échantillon grandit, il convient que l'estimateur tende vers la vraie valeur du paramètre inconnu. C'est la cas de $m_i$ qui est la moyenne du

caractère dans l'échantillon i. Lorsque le caractère ne peut prendre que les valeurs 1 et 0, c'est aussi le cas de $f_i$ qui est la fréquence d'apparition de X = 1 dans l'échantillon i ;

2) il convient de plus que, sur la série (théorique) de tous les échantillons de taille n, la moyenne des estimateurs soit égale au paramètre de la population-mère (estimation sans biais). Cela est encore vrai pour m et f mais ne l'est pas pour la distribution des $\sigma_i^2$. En effet, si on prend pour estimateur de $\sigma^2$ la quantité

$$S_i^2 = \frac{1}{n} \sum_{j=1}^{n} (x_j^i - m_i)^2$$

où la sommation porte sur les valeurs de X observées dans l'échantillon i dont la moyenne est $m_i$, on montre que $E(S^2) = \frac{n-1}{n} \sigma^2$ et non $\sigma^2$. D'après la 2ème condition, $S_i^2$ n'est pas un estimateur sans biais de $\sigma^2$, et il faudra plutôt prendre

$$S_i^{\star 2} = \frac{n}{n-1} S_i^2 \qquad (7.19)$$

qui satisfait aux deux conditions précédentes.

En pratique, pour estimer la variance $\sigma^2$ de la population-mère, on calculera

$$S_i^{\star 2} = \frac{1}{n-1} \sum_{j=1}^{n} (x_j^i - m_i)^2$$

$$= \frac{1}{n-1} \left[ \sum_{j=1}^{n} x_j^{i2} - n m_i^2 \right] \qquad (7.20)$$

à partir des valeurs du caractère observées sur l'échantillon i. Le dénominateur (n - 1) est appelé nombre de degrés de liberté de l'estimation : c'est le nombre de déviations $(x_j^i - m_i)$ indépendantes observées par échantillon, soit (n - 1), puisque d'après (3.8) il existe une relation entre ces déviations :

$$\sum_{j=1}^{n} (x_j^i - m_i) = 0$$

Remarque

Pour n élevé, l'équ. (7.19) montre que $S_i^{\star}$ est équivalent à $S_i$ qui devient alors un bon estimateur.

## II. ESTIMATION PAR UN INTERVALLE DE CONFIANCE

Soit une distribution (N, M, σ) d'un caractère X dans une population. On suppose que cette distribution suit une loi normale $\mathcal{N}$ (M, σ), ce qui correspond à un cas relativement fréquent et de plus, est très pratique pour les calculs.

D'après les propriétés de la loi normale (cf. 6 D III), on peut dire que 68,3 % de la population sont concentrés sur un intervalle de X recouvrant un écart-type de part et d'autre de la moyenne, 95,4 % sur un intervalle recouvrant deux écarts-type de part et d'autre de la moyenne, etc. (voir fig. 7.1).

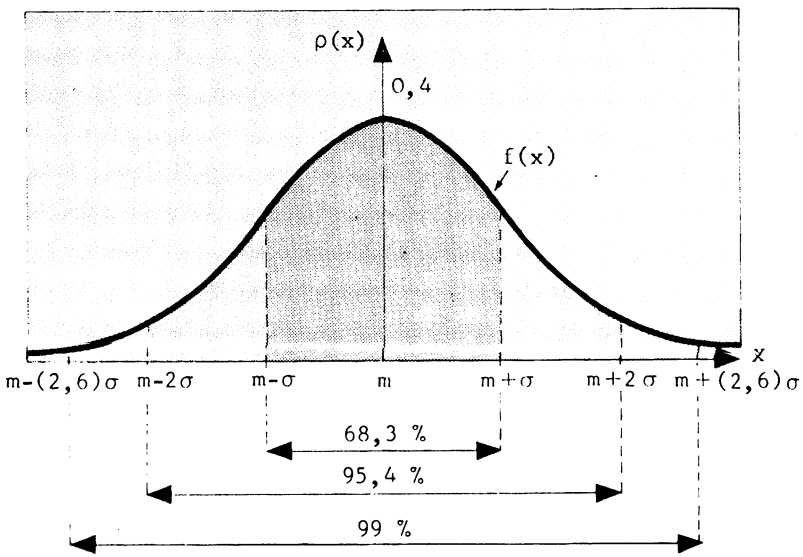



Figure 7.1

En s'intéressant en particulier aux cas des pourcentages de 95 % et 99 %, on a :

$$P\ (M - (1,96)\ \sigma \leq X \leq M + (1,96)\ \sigma) = 0,95$$

$$P\ (M - (2,58)\ \sigma \leq X \leq M + (2,58)\ \sigma) = 0,99$$

Dans le premier cas, par exemple, on peut s'attendre à ce qu'une valeur observée du caractère appartienne à l'intervalle $[M - (1,96)\ \sigma\ ;\ M + (1,96)\ \sigma]$ avec un seuil de confiance de 95 % - on dit aussi avec un risque d'erreur de 5 % -. L'intervalle précédent est appelé intervalle de confiance à 95 % pour la loi normale. De même l'intervalle $[M - (2,58)\ \sigma\ ;\ M + (2,58)\ \sigma]$ constitue l'intervalle de confiance à 99 %, etc.

Dans ce qui suit, nous désignerons le risque d'erreur par $\alpha$, le seuil de confiance par $(1 - \alpha)$ et la valeur absolue de la variable réduite $T = \frac{X - M}{\sigma}$ limitant l'intervalle de confiance par $t_\alpha$. D'une manière générale, on peut donc écrire :

$$P\ [M - t_\alpha\ \sigma \leq X \leq M + t_\alpha\ \sigma] = P[-t_\alpha \leq T \leq t_\alpha] = 1 - \alpha \quad (7.21)$$

Le tableau suivant donne les valeurs de $t_\alpha$ pour quelques risques d'erreur usuels, dans le cas de la loi normale.

| risque $\alpha$ | 0,5 % | 1 % | 5 % | 10 % |
|---|---|---|---|---|
| seuil de confiance $1-\alpha$ | 99,5 % | 99 % | 95 % | 90 % |
| $t_\alpha$ | 2,81 | 2,58 | 1,96 | 1,645 |

Tableau 7.1

## III. NORMALITE DES FLUCTUATIONS D'ECHANTILLONNAGE

Nous nous limiterons ici à l'estimation d'une moyenne et d'une fréquence dans une population à partir des résultats obtenus sur un échantillon extrait de cette population.

## 1. Fluctuations d'échantillonnage d'une moyenne

Soit une distribution (N, M, $\sigma$) d'un caractère X dans une population. On considère tous les échantillons de taille n pouvant être prélevés dans cette population et caractérisés par ($n$, $m_i$, $\sigma_i$).

On introduit le caractère t donné par

$$t = \frac{m - M}{\sigma / \sqrt{n}} \tag{7.22}$$

qui associe à chaque échantillon i, l'écart réduit $t_i$. On définit ainsi la distribution

$$\{t_1, t_2, t_3, \ldots, t_i, \ldots, t_k\} \tag{7.23}$$

des fluctuations de m par rapport à M (fluctuations réduites en divisant par l'écart-type $\sigma/\sqrt{n}$ de la distribution de m, équ. (7.4)). Cette distribution n'est autre que la distribution de t introduite en (7.15).

Pour déterminer un intervalle de confiance pour la moyenne, il convient d'examiner d'abord la normalité de cette distribution. Deux cas peuvent se présenter suivant que la population-mère est normale ou non.

### 1) Cas d'une population normale

La distribution de m est elle-même normale, mais la normalité de la distribution de t dépend, comme nous allons le préciser, de la taille n des échantillons.

La théorie montre que, lorsque la population-mère est distribuée normalement, le caractère $t_i$ suit une loi dite "loi de Student", de densité de probabilité définie par

$$f(t) = \frac{A}{\left(1 + \frac{t^2}{\nu}\right)^{\frac{\nu + 1}{2}}} \tag{7.24}$$

où $\nu$ est le nombre de degrés de liberté et A une constante dépendant uniquement de $\nu$, c'est à dire que :

$$P\ (a \leqslant t_i \leqslant b) = \int_a^b f\ (x)\ dx$$

A chaque valeur de $\nu$ correspond une distribution théorique. La figure 7.2 représente la distribution de Student correspondant à $\nu = 2$ par exemple, comparée à la distribution normale centrée et réduite.

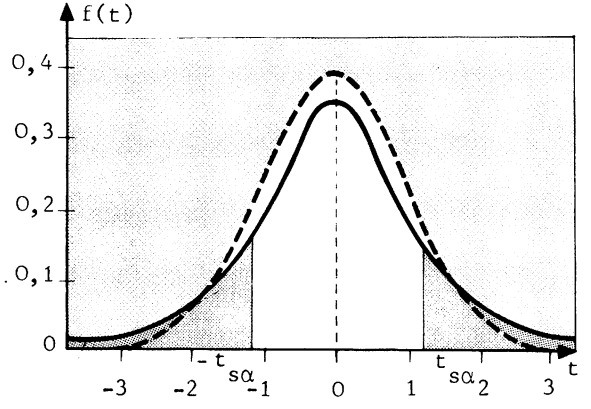



Figure 7.2

La théorie montre également que lorsque $\nu$ ou n est élevé ($n \geqslant 30$) la distribution de Student peut être assimilée à une distribution normale (voir fig. 7.2).

D'une manière générale, si l'on suppose que la population-mère est distribuée normalement, deux cas sont à considérer suivant que $n \geqslant 30$ ou $n < 30$.

a) $n \geqslant 30$, c'est le cas dit des "grands échantillons" : les fluctuations d'échantillonnage sont distribués normalement, les intervalles de confiance sont déterminés par la loi normale (cf. 7 B II).

b) $n < 30$, cas des "petits échantillons" : les fluctuations d'échantillonnage suivent une loi de Student et les intervalles de confiance doivent être déterminés par cette loi. Il existe des tables qui, pour un nombre de degrés de

liberté $\nu$ et un risque $\alpha$ donnés, fournissent les limites de l'intervalle de confiance $\pm t_{s\alpha}$, telles que

$$P[-t_{s\alpha} \leqslant t \leqslant t_{s\alpha}] = 1 - \alpha \qquad (7.25)$$

La table 4 en fournit un exemple, pour quelques valeurs usuelles de $\nu$ et $\alpha$.

Remarque

La table 4 donne directement la valeur de $t_{s\alpha}$ correspondant au risque $\alpha$, à l'encontre de la table 3 de la loi normale qui donne $t_{\alpha}$ pour un seuil de confiance de $(1 - \alpha)/2$.

2) Cas d'une population non normale

Si la population n'est pas distribuée normalement, le théorème dit "de la convergence vers la loi normale" montre que, plus l'échantillon est grand ($n \geqslant 30$) et plus la distribution de m se rapproche de la loi normale. C'est sans doute une des raisons qui expliquent l'importance de la loi normale.

Suivant la taille de l'échantillon, on pourra encore distinguer 2 cas :

a) $n \geqslant 30$

La distribution de m peut être considérée comme normale et la distribution des fluctuations d'échantillonnage aussi. D'une manière générale, pour $n \geqslant 30$, que la population soit normale ou non, les intervalles de confiance seront déterminés par la loi normale.

b) $n < 30$

La distribution de m n'est pas normale. On ne pourra traiter par la loi de Student que les seuls cas où la population peut être supposée normale, ce qui entraîne la normalité de la distribution de m.

## 2. Fluctuations d'échantillonnage d'une fréquence

Soit une population où le caractère X ne peut prendre que les valeurs 1 ou 0, et soient p la proportion des éléments vérifiant X = 1 et q celle des éléments vérifiant X = 0 (p + q = 1).

On considère que la distribution d'échantillonnage de f définie par (7.9) est pratiquement normale si les produits np et nq sont supérieurs à 10, ou à la rigueur à 5. Auquel cas, la normalité de la distribution des fluctuations d'échantillonnage de f est assurée et on peut appliquer la loi normale pour déterminer les intervalles de confiance.

# IV. INTERVALLE DE CONFIANCE D'UNE MOYENNE

On dispose d'un échantillon $(n, m_1, \sigma_1)$. Déterminer un intervalle de confiance centré sur $m_1$ et susceptible de contenir la moyenne M (inconnue) de la population, avec la probabilité $(1 - \alpha)$.

La moyenne $m_1$ est un élément de la distribution d'échantillonnage des moyennes dont nous avons désigné la moyenne et l'écart-type par E (m) et $\sigma$ (m).

## 1. Cas d'un échantillonnage non-exhaustif

*Les paramètres E (m) et $\sigma$ (m) sont donnés par (7.2) et* (7.4), soit :

$$E(m) = M \tag{7.26}$$

$$\sigma(m) = \frac{\sigma}{\sqrt{n}} \tag{7.27}$$

a) Cas où $n \geqslant 30$ (grands échantillons)

La distribution d'échantillonnage de m est normale. En adoptant un risque $\alpha$, on peut écrire d'après (7.21)

$$P\left[E(m) - t_\alpha\, \sigma(m) \leqslant m_1 \leqslant E(m) + t_\alpha\, \sigma(m)\right] = 1 - \alpha \tag{7.28}$$

ou tout simplement

$$E(m) - t_\alpha \sigma(m) \leq m_1 \leq E(m) + t_\alpha \sigma(m)$$

avec un risque d'erreur $\alpha$.

En tenant compte de (7.26) et (7.27), on a

$$M - t_\alpha \frac{\sigma}{\sqrt{n}} \leq m_1 \leq M + t_\alpha \frac{\sigma}{\sqrt{n}} \tag{7.29}$$

Cet encadrement délimite un intervalle dit "du pari" permettant le cas échéant d'estimer $m_1$ connaissant M et $\sigma$.

On en déduit

$$m_1 - t_\alpha \frac{\sigma}{\sqrt{n}} \leq M \leq m_1 + t_\alpha \frac{\sigma}{\sqrt{n}} \tag{7.30}$$

L'encadrement (7.30) délimite l'intervalle de confiance de M et répond au problème de l'estimation de M posé précédemment.

Généralement, $\sigma$ qui est l'écart-type de la population-mère, est inconnu. On le remplace dans l'encadrement par son estimateur $S_i^*$ donné par (7.20). On obtient alors

$$m_1 - t_\alpha \frac{S_i^*}{\sqrt{n}} \leq M \leq m_1 + t_\alpha \frac{S_i^*}{\sqrt{n}} \tag{7.31}$$

Si en particulier n est très proche de N, $S_i^*$ est voisin de $\sigma_1$ qui est l'écart-type de l'échantillon. On peut alors écrire :

$$m_1 - t_\alpha \frac{\sigma_1}{\sqrt{n}} \leq M \leq m_1 + t_\alpha \frac{\sigma_1}{\sqrt{n}} \tag{7.32}$$

b) *Cas où n < 30 (petits échantillons)*

Les encadrements précédents ne sont plus valables. Il faut d'abord supposer que la population-mère est distribuée normalement, et remplacer ensuite la variable t de la loi normale par la variable $t_s$ de Student. L'intervalle de confiance (7.30) devient alors :

$$m_1 - t_{s\alpha} \frac{S_i^*}{\sqrt{n}} \leq M \leq m_1 + t_{s\alpha} \frac{S_i^*}{\sqrt{n}} \tag{7.33}$$

$t_{s\alpha}$ étant donné par la table 4 en fonction du risque $\alpha$ choisi et du nombre de degrés de librrté $\nu = n - 1$.

2. Cas d'un échantillonnage exhaustif

Il faut remplacer dans l'encadrement (7.30) l'écart-type $\frac{\sigma}{\sqrt{n}}$ de la distribution des moyennes, par $\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ donné par (7.6). On trouve

$$m_1 - t\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \leqslant M \leqslant m_1 + t\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \qquad (7.34)$$

où comme précédemment, $t = t_\alpha$ (loi normale) pour $n \geqslant 30$, ou $t = t_{s\alpha}$ (loi de Student) pour $n < 30$.

. *Exemple 1*

Dans une fabrication portant sur 50 000 articles, un sondage sur 400 articles a donné un poids moyen par article de 200 g avec un écart-type de 50 g. Estimer le poids moyen dans la fabrication, au seuil de confiance de 95 %.

Pour $n = 400$, en adoptant l'estimateur $S^\star = \sigma_1$, on obtient à partir de la relation (7.32)

$$200 - 1{,}96 \times \frac{50}{\sqrt{400}} \leqslant M \leqslant 200 + 1{,}96 \times \frac{50}{\sqrt{400}}$$

$$195{,}1 \leqslant M \leqslant 204{,}9$$

. *Exemple 2*

Un dosage de sucre dans une solution effectué sur 8 prélèvements provenant d'une même fabrication, a donné les résultats suivants, exprimés en g/ℓ :

19,5 - 19,7 - 19,8 - 20,2 - 20,2 - 20,3 - 20,4 - 20,8.

Entre quelles limites varie la concentration moyenne de la fabrication, au risque de 5 % ?

L'échantillon étant petit ($n = 8$), il faut utiliser l'estimateur $S^\star$ pour $\sigma$ et recourir à la distribution de Student.

L'intervalle de confiance est alors donné par l'encadrement (7.33)

$$m_1 - t_{s\alpha} \frac{S^\star}{\sqrt{n}} \leqslant M \leqslant m_1 + t_{s\alpha} \frac{S^\star}{\sqrt{n}}$$

en supposant un échantillonage non-exhaustif ou ce qui revient au même un prélèvement exhaustif dans une solution pratiquement inépuisable.

D'après (7.20)

$$S^{\star 2} = \frac{1}{n-1} \left[ \sum_{j=1}^{n} X_j^2 - n m^2 \right]$$

on a successivement

$$\sum X_j = 160,9 \qquad \text{d'où } m = \frac{160,9}{8} = 20,112$$

$$\sum X_j^2 = 3\,237,35$$

$$S^{\star 2} = 0,178 \qquad \text{et} \qquad S^\star = 0,422$$

La table 4, pour $\alpha = 0,05$ et $\nu = 8 - 1 = 7$, donne $t_{s\alpha} = 2,365$. On a donc

$$20,11 - \frac{2,365 \times 0,422}{2,828} \leqslant M \leqslant 20,11 + \frac{2,365 \times 0,422}{2,828}$$

$$19,75 \leqslant M \leqslant 20,46$$

## V. INTERVALLE DE CONFIANCE D'UNE FREQUENCE

On dispose d'un échantillon de taille n où le caractère X étudié ne peut prendre que les valeurs 1 et 0, et où la fréquence d'apparition du caractère X = 1 est $f_1$. Déterminer un intervalle de confiance centré sur $f_1$ et susceptible de recouvrir la fréquence p d'apparition du caractère X = 1 dans la population d'où est extrait l'échantillon, avec la probabilité $(1 - \alpha)$.

La fréquence $f_1$ est un élément de la distribution d'é-

chantillonnage des fréquences (7.9) dont nous avons désigné la moyenne par E (f) et l'écart-type par $\sigma$ (f).

1. Cas d'un échantillonnage non-exhaustif

Les paramètres E (f) et $\sigma$ (f) sont donnés par (7.10) et (7.11) soit

$$E(f) = p \tag{7.35}$$

$$\sigma(f) = \sqrt{\frac{pq}{n}} \tag{7.36}$$

où $q = 1 - p$.

En supposant que les conditions de validité de l'approximation normale sont remplies, c'est à dire que $np, nq > 5$ (cf. 7 B III), on peut écrire

$$p - t_\alpha \sqrt{\frac{pq}{n}} \leqslant f_1 \leqslant p + t_\alpha \sqrt{\frac{pq}{n}} \tag{7.37}$$

où le coefficient $t_\alpha$ est déterminé par la loi normale.

Cet encadrement délimite un intervalle de pari pour $f_1$, permettant le cas échéant d'estimer $f_1$ connaissant p. On en déduit :

$$f_1 - t_\alpha \sqrt{\frac{pq}{n}} \leqslant p \leqslant f_1 + t_\alpha \sqrt{\frac{pq}{n}} \tag{7.38}$$

ce qui délimite l'intervalle de confiance de p et répond au problème posé.

Généralement la valeur de p est inconnue. Une méthode approximative consiste à remplacer p sous le radical par la fréquence $f_1$ observée sur l'échantillon. Cela revient à prendre pour l'écart-type $\sigma(f) = \sqrt{\frac{pq}{n}}$ l'estimateur

$$S = \sqrt{\frac{f_1 (1 - f_1)}{n}} \tag{7.39}$$

(en tenant compte du fait que $q = 1 - p$).

Une autre méthode consiste à remplacer le produit pq sous le radical par sa valeur maximum, qui correspond à $p = q = \frac{1}{2}$. On a alors :

$$S = \sqrt{\frac{1}{4\,n}} \qquad (7.40)$$

Dans les deux cas, l'intervalle de confiance s'écrit :

$$f_1 - t_\alpha\, S \leqslant p \leqslant f_1 + t_\alpha\, S \qquad (7.41)$$

2. Cas d'un échantillonnage exhaustif

Il faut remplacer dans l'encadrement (7.37) et les suivants, l'écart-type $\sigma(f) = \sqrt{\frac{pq}{n}}$ par $\sigma(f) = \sqrt{\frac{pq}{n}}\sqrt{\frac{N-n}{N-1}}$ donné par (7.13).

. *Exemple*

Dans une école de 1 000 élèves, un sondage sur une classe de 35 élèves a permis de constater que 7 d'entre eux avaient une légère infection contagieuse. Estimer la proportion d'enfants atteints dans l'école, au risque de 5 %.

Le caractère étudié ici ne peut prendre que deux valeurs X = 1 pour les élèves atteints, et X = 0 pour les élèves non atteints. La fréquence d'apparition de X = 1 pour l'échantillon est $f_1 = \frac{7}{35} = 0{,}2$.

L'effectif atteint et l'effectif non atteint dans l'échantillon étant supérieurs à 5, on peut appliquer l'approximation normale. En utilisant l'estimateur

$$S = \sqrt{\frac{f_1\,(1 - f_1)}{n}} = 0{,}067$$

dans l'intervalle de confiance (7.41), on obtient

$$0{,}2 - [1{,}96 \times 0{,}067] \leqslant p \leqslant 0{,}2 + [1{,}96 \times 0{,}067]$$

soit $0{,}068 \leqslant p \leqslant 0{,}332$ au seuil de 5 %.

## C. TESTS DE SIGNIFICATION

Jusqu'ici, nous avons étudié la représentativité d'une population par tous ses échantillons (échantillonnage) et par un échantillon (estimation). Les tests de signification ont pour objet d'examiner si les différences observées entre un échantillon et la population-mère ou entre deux échantillons sont dues aux fluctuations d'échantillonnage (c'est-à-dire au hasard) ou si elles sont significatives. En d'autres termes, ces tests permettent de résoudre des problèmes

a) de conformité d'un échantillon à la population

b) d'homogénéité de deux échantillons entre eux.

### I. PRINCIPE DES TESTS D'HYPOTHESE

Considérons le problème suivant : étant donné un échantillon de taille n, dont la moyenne des valeurs d'un certain caractère X est $m_1$, issu d'une population P caractérisée par (N, M, $\sigma$), peut-on considérer que la différence entre $m_1$ et M est significative ?

Pour répondre à cette question, il est nécessaire de disposer d'une méthode permettant de dire par exemple, à partir de quelle différence entre $m_1$ et M, l'écart entre l'échantillon et la population est trop grand pour être attribué aux fluctuations d'échantillonnage.

On est amené généralement à formuler une hypothèse qui consiste à supposer que la différence observée est simplement due aux fluctuations d'échantillonnage, et qui est appelée hypothèse nulle, désignée par $H_o$. On recherche ensuite un

critère de test qui permette de rejeter ou ne pas rejeter l'hypothèse $H_o$, en tenant compte du risque d'erreur ou seuil de signification choisi.

Le critère de test est tout naturellement la déviation réduite :

$$t_o = \frac{m_1 - M}{\sigma(m_i)} \qquad (7.42)$$

c'est-à-dire

$$t_o = \frac{m_1 - M}{\sigma/\sqrt{n}} \, . \qquad (7.43)$$

Sous l'hypothèse $H_o$, ce critère présente des fluctuations d'échantillonnage qui, pour $n \geq 30$, sont distribuées normalement. En considérant par exemple un seuil de 5 %, on est conduit à adopter la règle de décision suivante :

1) Si $t_o$ est extérieur à l'intervalle $[-1{,}96 \; ; \; 1{,}96]$, la probabilité de cette situation étant seulement de 5 % sous l'hypothèse nulle, on rejette $H_o$. On dit que la différence $m_1 - M$ est significative (non due au hasard) au seuil de signification de 5 %, ou encore que l'échantillon n'est pas représentatif de la population, au même seuil.

Dans ce cas, une erreur peut être commise, qui consiste à rejeter l'hypothèse $H_o$ alors que celle-ci est exacte. On dit qu'il s'agit d'un risque d'erreur de 1ère espèce. Il est égal au seuil de signification choisi, et par conséquent adopter un faible seuil revient à limiter la probabilité de rejeter à tort l'hypothèse nulle.

2) Si $t_o$ est intérieur à l'intervalle $[-1{,}96 \; ; \; 1{,}96]$, on n'a pas de raison de rejeter $H_o$. La différence $m_1 - M$ est dite non significative, au seuil de 5 %. On peut accepter $H_o$ et attribuer cette différence au hasard, ou bien on peut ne prendre aucune décision. L'échantillon étudié n'a pas permis de constater une différence significative.

Dans ce cas, le risque d'erreur est dit de 2ème espèce :

il consiste à ne pas rejeter $H_o$ alors que celle-ci est fausse. A l'inverse du risque de 1ère espèce, le risque de 2ème espèce augmente quand on diminue le seuil de signification.

## II. PREMIERE APPLICATION : LES TESTS DE CONFORMITE

### 1. Comparaison d'une moyenne observée à une moyenne théorique

Etant donné un échantillon de taille n, dont les valeurs observées du caractère ont pour moyenne $m_1$, peut-il être considéré comme représentatif de la population P (N, M, $\sigma$) ?

D'une manière pratique, deux possibilités peuvent se présenter suivant que l'écart-type $\sigma$ de la population est connu ou non.

1) Si $\sigma$ est connu, le critère du test $H_o$ est l'écart réduit donné par (7.43)

$$t_o = \frac{m_1 - M}{\sigma / \sqrt{n}} \qquad (7.44)$$

2) Si $\sigma$ n'est pas connu, on utilise son estimateur $S^\star$ donné par (7.20). L'écart réduit devient

$$t_o = \frac{m_1 - M}{S^\star / \sqrt{n}} \qquad (7.45)$$

Rappelons que dans les deux cas, il faut tester $t_o$ à l'aide de l'intervalle de confiance déterminé par

a) la loi normale lorsque $n \geq 30$

b) la loi de Student lorsque $n < 30$.

. *Exemple 1*

40 moteurs représentant un échantillon d'une certaine fabrication ont fonctionné en moyenne pendant 260 jours sans problème. Peut-on considérer cet échantillon comme appartenant à la fabrication habituelle, si dans celle-ci, le caractère (c'est à dire le nombre de jours pendant lesquels un

moteur a fonctionné sans problème) suit une loi normale de moyenne 240 jours et d'écart-type 50 jours ?

Hypothèse nulle $H_o$ : l'échantillon appartient à la fabrication habituelle.

Ecart réduit : $t_o = \frac{260 - 240}{50 / \sqrt{40}} = 2,53$

L'intervalle de confiance est déterminé par la loi normale, puisque $n > 30$.

Intervalle de confiance à 5 % : $I_5 = [-1,96 ; 1,96]$

Intervalle de confiance à 1 % : $I_1 = [-2,58 ; 2,58]$

On a :

$t_o \notin I_5$ et $t_o \in I_1$

On est donc conduit à rejeter $H_o$ au seuil de 5 %, et à ne pas rejeter $H_o$ au seuil de 1 %.

. *Exemple 2*

Dans l'exemple 2 du § 7 B IV, le dosage de sucre sur les 8 prélèvements a donné une moyenne de 20,11 g/ℓ. L'échantillon est-il représentatif de la production au seuil de 5 %, si l'on admet que la concentration en sucre habituelle suit une loi normale de moyenne 19,6 g/ℓ ?

Hypothèse nulle $H_o$ : l'échantillon est représentatif.

L'écart-type $\sigma$ de la population-mère étant inconnu, on le remplace par son estimateur $S^* = 0,422$. L'échantillon étant petit ($n < 30$), il faut recourir à la distribution de Student pour trouver l'intervalle de confiance.

Le critère du test $H_o$ est :

$$t_o = \frac{20,11 - 19,6}{0,422/\sqrt{8}} = 3,43$$

Pour $\alpha = 5$ % et $\nu = n - 1 = 7$, la table 4 fournit $t_{s\alpha} = 2,365$. L'intervalle de confiance est donc $I_5 = [-2,365 ; 2,365]$. Par conséquent $t_o \notin I_5$, l'échantillon n'est pas représentatif de la production.

## 2. Comparaison d'une fréquence observée et d'une fréquence théorique

Le problème est le suivant : étant donné un échantillon de taille n, où la fréquence d'apparition d'un certain caractère est $f_1$, est-il représentatif de la population-mère où la fréquence d'observation de ce caractère est p ?

La distribution d'échantillonnage est ici la distribution de f définie en (7.9). D'après (7.11) son écart-type est donné par

$$\sigma(f) = \sqrt{\frac{pq}{n}} \tag{7.46}$$

Hypothèse nulle : l'échantillon est représentatif. Le critère de test est l'écart $f_1 - p$, d'où l'écart réduit :

$$t_o = \frac{f_1 - p}{\sqrt{\frac{pq}{n}}} \tag{7.47}$$

qui doit être testé à l'aide de l'intervalle de confiance approprié, comme précédemment.

. *Exemple*

Le taux d'écoute d'un certain programme de télévision est supposé constant et égal à 15 %. A la suite d'une nouvelle présentation, un sondage limité à 80 téléspectateurs a révélé que 18 d'entre eux ont suivi ce programme. Peut-on dire que la nouvelle présentation ait influencé le public, au seuil de 5 % ?

Hypothèse $H_o$ : la nouvelle présentation n'a pas influencé le public.

On a ici : $p = 0{,}15 \quad q = 0{,}85 \quad n = 80$

$$f_1 = \frac{18}{80} = 0{,}225$$

D'après (7.47) l'écart réduit est :

$$t_o = \frac{0,225 - 0,15}{\sqrt{\frac{0,15 \times 0,85}{80}}} = 1,87.$$

Les effectifs np et nq étant supérieurs à 5, on peut appliquer l'approximation de la loi normale. Au seuil de 5 %, on a $I_\alpha = [-1,96 ; 1,96]$ , $t_o \in I_\alpha$ , et par conséquent, on ne peut rejeter $H_o$.

3. Test bilatéral et test unilatéral

1) Test bilatéral

Soit $m_1$ la moyenne des valeurs d'un caractère observées sur un échantillon et soit $M_o$ la valeur théorique de la moyenne de ce même caractère dans une population $P_o$.

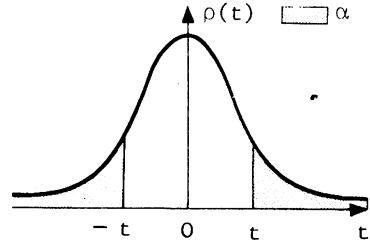
On peut toujours considérer que $m_1$ est un élément d'une distribution d'échantillonnage des moyennes définie sur une population P égale ou différente de $P_o$. Auquel cas, la moyenne E (m) d'une telle distribution a une certaine valeur M.

Dans un test de conformité, comparer $m_1$ à $M_o$ revient en fait à comparer M à $M_o$. L'hypothèse nulle d'une différence $m_1 - M_o$ non significative peut aussi bien s'exprimer par $M = M_o$. Le rejet de l'hypothèse nulle correspond à $M \neq M_o$. On dit que l'on teste l'hypothèse $H_o : M = M_o$ contre l'hypothèse alternative $H_1 : M \neq M_o$.

Dans ce cas, il convient de limiter l'intervalle de confiance $I_\alpha$ aux deux extrémités de la distribution de t, comme nous l'avons fait jusqu'ici (voir fig. 7.3). Le test est dit bilatéral.

Figure 7.3

$P(-t_\alpha \leq T \leq t_\alpha) = 1 - \alpha$



2) Test unilatéral

Test de $H_o : M \leq M_o$ contre $H_1 : M > M_o$.

. *Exemple*

Le caractère X étudié est le total des points obtenus par chaque étudiant à un examen. On a des raisons de croire qu'une certaine promotion est particulièrement douée. Peut-on affirmer cela, au vu des résultats obtenus par cette promotion ?

Soit $M_o$ la moyenne théorique à laquelle on s'attend, au vu des résultats des années précédentes, et soit $m_1$ la moyenne de cette promotion. $m_1$ est un élément de la distribution d'échantillonnage de m, de moyenne $E(m) = M$.

Le test est alors le suivant :

. hypothèse nulle $H_o : M \leq M_o$ (promotion normale)

hypothèse alternative $H_1 : M > M_o$ (promotion douée)

Le jugement est différent de celui du paragraphe précédent dans la mesure où l'on ne doit pas rejeter $H_o$ lorsque $M < M_o$.

Il est évident que lorsque la valeur observée $m_1 < M_o$, il est inutile de recourir à $H_1$, la conclusion est immédiate, on accepte $H_o$ : la promotion n'est pas particulièrement douée. Si $m_1 > M_o$, plus l'écart réduit donné par (7.43)

$$t_o = \frac{m_1 - M_o}{\sigma / \sqrt{n}} \qquad (7.48)$$

est grand, et plus on aura tendance à rejeter $H_o$ et accepter $H_1$. On est donc conduit à limiter l'intervalle de confiance $I_\alpha$ à la seule extrémité droite de la distribution de l'écart t (voir fig. 7.4). On dit que le test est unilatéral.

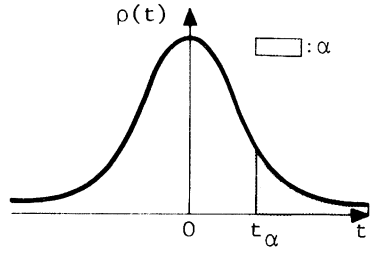



Figure 7.4

$P(T \leq t_\alpha) = 1 - \alpha$

La règle de décision du test devient :

si $t_o \in I_\alpha$ : ne pas rejeter $H_o$

si $t_o \notin I_\alpha$ : rejeter $H_o$ et accepter $H_1$.

3) Test unilatéral

Test de $H_o : M \geq M_o$ contre $H_1 : M < M_o$

. *Exemple*

En reprenant l'exemple précédent, peut-on conclure qu'une promotion est particulièrement faible au vu des résultats obtenus par cette promotion ?

Il est évident que le problème ne peut se poser que si $m_1 < M_o$. Le test est alors le suivant :

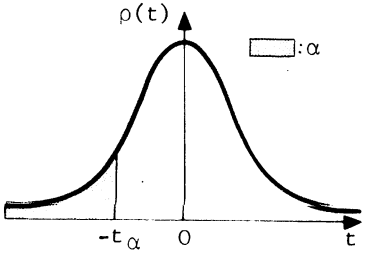
hypothèse nulle $H_o : M \geq M_o$ (promotion normale)

hypothèse alternative $H_1 : M < M_o$ (promotion faible)

Pour les mêmes raisons que précédemment, on est amené à limiter l'intervalle de confiance à la seule extrémité gauche de la distribution de t. (voir fig. 7.5).

Figure 7.5

$P(T \geq -t_\alpha) = 1 - \alpha$




Remarque

Les limites de $I_\alpha$ ne sont pas les mêmes évidemment, suivant que le test est bilatéral ou unilatéral. Le tableau 7.2 donne quelques valeurs de $t_\alpha$ pour une distribution normale de l'écart t.

| risque $\alpha$ | 0,5 % | 1 % | 5 % | 10 % |
|---|---|---|---|---|
| $t_\alpha$ bilatéral | 2,81 | 2,58 | 1,96 | 1,645 |
| $t_\alpha$ unilatéral | 2,58 | 2,33 | 1,645 | 1,28 |

Tableau 7.2

## III. DEUXIEME APPLICATION : LES TESTS D'HOMOGENEITE

### 1. Comparaison de deux moyennes observées

Soient $m_1$ et $m_2$ les moyennes des valeurs d'un caractère observées sur deux échantillons 1 et 2. $m_1$ est un élément de la distribution d'échantillonnage de $m_A$ définie sur une population $P_A$. De même, $m_2$ est un élément de la distribution de $m_B$ définie sur $P_B$. Il s'agit de déterminer si ces deux échantillons proviennent de 2 populations $P_A$ et $P_B$ de même moyenne (on dit alors qu'ils sont homogènes).

Supposons que les deux échantillons sont caractérisés par $(n_1, m_1, \sigma_1)$ et $(n_2, m_2, \sigma_2)$ et les deux populations par $(M_A, \sigma_A)$ et $(M_B, \sigma_B)$.

Hypothèse nulle $H_o$ : $M_A = M_B$

Hypothèse $H_1$ : $M_A \neq M_B$

C'est un test bilatéral, dont le critère est $(m_2 - m_1)$.

a) Cas des grands échantillons ($n_1$ et $n_2 \geqslant 30$)

Les distributions de $m_A$ et de $m_B$ peuvent être considérées comme normales, respectivement de moyennes E $(m_A) = M_A$

et E $(m_B) = m_B$, et d'écart-type $\sigma_A/\sqrt{n_1}$ et $\sigma_B/\sqrt{n_2}$ où $n_1$ et $n_2$ sont les tailles des 2 échantillons (cf. § 7 A I).

Sous l'hypothèse $H_o$, $(m_2 - m_1)$ est distribué normalement avec une moyenne $M_B - M_A = 0$, une variance

$$V_{m_2 - m_1} = V_{m_1} + V_{m_2} = \frac{\sigma_A^2}{n_1} + \frac{\sigma_B^2}{n_2} \qquad (7.49)$$

et un écart-type

$$\sigma_{m_2 - m_1} = \sqrt{\frac{\sigma_A^2}{n_1} + \frac{\sigma_B^2}{n_2}} \qquad (7.50)$$

L'écart réduit à tester est donc

$$t_o = \frac{m_2 - m_1}{\sqrt{\frac{\sigma_A^2}{n_1} + \frac{\sigma_B^2}{n_2}}} \qquad (7.51)$$

Si les écarts-type $\sigma_A$ et $\sigma_B$ sont inconnus, on les remplace par leurs estimateurs $S^{\star}_{A_1}$ et $S^{\star}_{B_2}$ donnés par l'expression (7.20) En particulier, si les échantillons sont très grands, on peut prendre $S^{\star}_{A_1} = \sigma_1$ et $S^{\star}_{B_2} = \sigma_2$

On poursuit alors le test, en utilisant l'intervalle de confiance $I_\alpha$ correspondant au risque $\alpha$ et à une distribution normale, et on applique la règle de décision habituelle.

b) Cas où l'un des échantillons, ou les deux sont petits ($n < 30$)

On suppose pour simplifier que les populations-mères sont normales et qu'elles ont la même variance $\sigma_A^2 = \sigma_B^2 = \sigma^2$. L'estimateur $S^{\star 2}$ de $\sigma^2$ est obtenu en prenant la moyenne de $S^{\star 2}_{A_1}$ et $S^{\star 2}_{B_2}$ pondérés par le nombre de degrés de liberté correspondant, soit

$$S^{\star 2} = \frac{(n_1 - 1)\, S_A^2 + (n_2 - 1)\, S_B^2}{(n_1 - 1) + (n_2 - 1)} \qquad (7.52)$$

Dans ce cas, l'écart réduit (7.51) s'écrit

$$t_o = \frac{m_2 - m_1}{\sqrt{\frac{S^{\star 2}}{n_1} + \frac{S^{\star 2}}{n_2}}} \qquad (7.53)$$

On poursuit le test en utilisant l'intervalle de confiance $I_{\alpha\nu}$ correspondant au risque $\alpha$ et au nombre de degrés de liberté $\nu = n_1 + n_2 - 2$ dans la loi de Student, suivi de la règle de décision habituelle.

. *Exemple*

Deux lycées différents ont obtenu au cours d'une épreuve du baccalauréat les résultats suivants :

| Lycée | Nombre d'élèves | Note moyenne | Ecart-type |
|---|---|---|---|
| A | 65 | 13,2 | 1,8 |
| B | 85 | 12,5 | 1,6 |

Ces résultats ont-ils une différence significative au risque de 1 % ?

Hypothèse nulle : les performances moyennes sont les mêmes, la différence apparente est due au hasard.

En adoptant l'approximation $S^{\star 2}_{A_1} = \sigma^2_1$ et $S^{\star 2}_{B_2} = \sigma^2_2$, on a successivement :

$$m_1 - m_2 = 13,2 - 12,5 = 0,7$$

$$\sigma_{m_1 - m_2} = \sqrt{\frac{1,8^2}{65} + \frac{1,6^2}{85}} = 0,283$$

$$t_o = \frac{0,7}{0,283} = 2,47$$

Les échantillons sont grands ($n > 30$). La table 3 de la loi normale donne pour un intervalle de confiance de 99 %

$t_\alpha = 2{,}58$. On a donc $I_\alpha = [-2{,}58 ; 2{,}58]$, $t_o \in I_\alpha$, et on ne peut rejeter l'hypothèse nulle.

2. Comparaison de deux fréquences observées

Soient $f_1$ et $f_2$ les fréquences d'apparition d'un certain caractère dans deux échantillons 1 et 2. $f_1$ est un élément de la distribution d'échantillonnage de la fréquence $f_A$ définie sur une population $P_A$. De même, $f_2$ est un élément de la distribution de $f_B$ définie sur une population $P_B$. Il s'agit de déterminer si ces deux échantillons proviennent de 2 populations $P_A$ et $P_B$ ayant la même proportion d'éléments possédant ce caractère.

Supposons que les 2 échantillons sont caractérisés par $(n_1, f_1)$ et $(n_2, f_2)$ et les 2 populations par $(p_A, \sigma_A)$ et $(p_B, \sigma_B)$.

Hypothèse nulle $H_o$ : $p_A = p_B = p$

Hypothèse $H_1$ : $p_A \neq p_B$

C'est un test bilatéral, dont le critère est $(f_2 - f_1)$.

Lorsque les effectifs np, nq sont supérieurs à 5, les distributions de $f_A$ et $f_B$ peuvent être considérées comme normales, respectivement de moyennes $E(f_A) = p_A$ et $E(f_B) = p_B$, et d'écarts-type $\sigma_A = \sqrt{\frac{p_A q_A}{n_1}}$ et $\sigma_B = \sqrt{\frac{p_B q_B}{n_2}}$ (cf. § 7 4 II)

Sous l'hypothèse $H_o$, $f_2 - f_1$ est distribué normalement avec une moyenne $p_B - p_A = 0$, une variance

$$V_{f_2 - f_1} = V_{f_1} + V_{f_2}$$

$$= \frac{pq}{n_1} + \frac{pq}{n_2} \qquad (7.54)$$

puisque $p_A = p_B = p$. Par conséquent, l'écart réduit est

$$t_o = \frac{f_2 - f_1}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad (7.55)$$

Une méthode approximative consiste à prendre pour p, qui est généralement inconnu, le pourcentage moyen entre les 2 échantillons soit

$$p = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2} \qquad (7.56)$$

On peut alors tester $t_o$ comme précédemment, à l'aide de $I_\alpha$ déterminé par la loi normale et décider de la validité de $H_o$.

. *Exemple*

Au cours de deux livraisons différentes, on a relevé 48 articles défectueux parmi les 800 constituant la première livraison, et 32 articles défectueux parmi les 400 constituant la deuxième livraison. Les deux pourcentages d'articles défectueux observés diffèrent-ils d'une manière significative, au seuil de 5 % ?

Hypothèse nulle $H_o$ : les 2 pourcentages sont les mêmes, la différence apparente est due au hasard.

On a successivement :

$$f_1 = \frac{48}{800} = 0{,}06 \qquad f_2 = \frac{32}{400} = 0{,}08 \qquad f_2 - f_1 = 0{,}02$$

D'après l'approximation (7.56), p = 0,067. L'écart réduit est donc :

$$t_o = \frac{0{,}02}{\sqrt{0{,}067 \times 0{,}933 \left(\frac{1}{800} + \frac{1}{400}\right)}} = 1{,}31$$

Tous les produits np, nq sont supérieurs à 5, on peut donc appliquer l'approximation normale.

Au seuil de 5 %, on obtient pour la loi normale $I_\alpha = [-1{,}96 \ ; \ 1{,}96]$ . Par conséquent, $t_o \in I_\alpha$, la différence observée n'est pas significative.

## D. TEST DU $\chi^2$

### I. DISTRIBUTION DU $\chi^2$

Considérons une population normale d'écart-type $\sigma$, et tous les échantillons de taille n pouvant être extraits de cette population et caractérisés par $(n, m_i)$ où $m_i$ est la moyenne du caractère X dans l'échantillon i. Pour chaque échantillon i, on calcule le paramètre

$$\chi_i^2 = \frac{\sum_{j=1}^{n} (x_j^i - m_i)^2}{\sigma^2} \qquad (7.57)$$

où $x_j^i$ est la valeur du caractère du $j^{ème}$ individu de l'échantillon i. On définit ainsi la distribution d'échantillonnage de $\chi^2$, introduite en (7.17).

La théorie montre que, lorsque la population-mère est distribuée normalement, le caractère $\chi_i^2$ suit une loi dite "loi du $\chi^2$", de densité de probabilité définie par

$$f(\chi^2) = A\,(\chi^2)^{\nu/2-1}\, e^{-\chi^2/2} \qquad (7.58)$$

où $\nu$ est le nombre de degrés de liberté introduit en 7 B I et utilisé déjà dans la distribution de Student (cf. 7 B III). Ce nombre est égal à (n - 1) et provient de ce que $\chi_i^2$ est la somme de (n - 1) termes indépendants. A est une constante dépendant uniquement de $\nu$.

A chaque valeur de $\nu$ correspond une distribution théorique. La figure 7.6 montre l'allure des courbes pour $\nu$ = 2, 5 et 10. Comme pour la loi normale et la loi de Student, on peut définir des intervalles de confiance. Il existe des

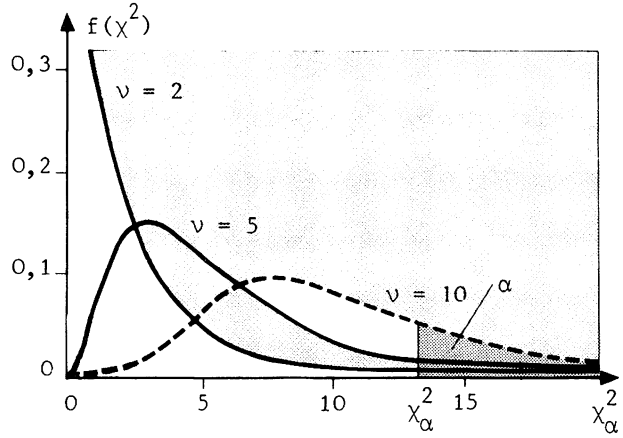



Figure 7.6

tables donnant la valeur de $\chi^2_\alpha$ ayant la probabilité $\alpha$ d'être égalée ou dépassée, en fonction du nombre de degrés de liberté $\nu$. La table 5 en fournit un exemple.

Remarque

La table 5 est limitée à $\nu = 30$. Au-dessus de cette valeur, on utilise le fait que le paramètre $\sqrt{2\,\chi^2}$ est distribué approximativement suivant une loi normale, de moyenne $\sqrt{2\,\nu - 1}$ et d'écart-type égal à 1.

## II. CRITERE DE PEARSON

Soit une population où le caractère X ne peut prendre *que les valeurs 1 et 0, et soient p et q respectivement* les fréquences de X = 1 et de X = 0 dans cette population ($p + q = 1$)

On considère un échantillon où la fréquence observée du caractère X = 1 est $p_1$ et celle de X = 0 est $q_1$ ($p_1 + q_1 = 1$). Nous avons vu que la comparaison de $p_1$ à p pouvait se faire au moyen de l'écart réduit (cf. (7.47) )

$$t_o = \frac{p_1 - p}{\sqrt{\frac{pq}{n}}} \qquad (7.59)$$

où n est la taille de l'échantillon. Comme on ne s'intéresse finalement qu'à la valeur absolue de $t_o$, on peut utiliser aussi bien comme critère de test, le carré de $t_o$, soit

$$t_o^2 = \frac{(p_1 - p)^2}{\frac{pq}{n}} = \frac{(np_1 - np)^2}{npq} \qquad (7.60)$$

Par comparaison avec (7.57), $t_o^2$ apparaît comme un $\chi^2$ de l'échantillon où le caractère est l'effectif vérifiant X = 1. Le produit $np_1$ représente l'effectif observé sur l'échantillon, np est la valeur moyenne théorique de cet effectif, et npq sa variance théorique.

On peut expliciter les rôles symétriques de p (correspondant à X = 1) et q (correspondant à X = 0) en écrivant :

$$t_o^2 = \chi^2 = \frac{(np_1 - np)^2}{np} + \frac{(nq_1 - nq)^2}{nq} \qquad (7.61)$$

ce qui se vérifie aisément.

Le paramètre $\chi^2$ constitue le critère de Pearson. On montre que lorsque n augmente indéfiniment, la distribution d'échantillonnage correspondante tend vers l'expression théorique f ($\chi^2$) donnée par (7.58), avec un nombre de degrés de liberté $\nu = 1$.

Il convient de remarquer qu'ici le nombre de degrés de liberté n'est pas donné par $\nu = n-1$, mais par $\nu = 1$. En effet, d'après la définition des caractères dans (7.61), le nombre d'observations par échantillon est de 2 (c'est à dire $np_1$ et $nq_1$) et comme $n(p_1 + q_1) = n$, le nombre d'observations <u>indépendantes</u> est bien $\nu = 1$.

Le critère $\chi^2$ peut être généralisé au cas où l'observation porterait sur plusieurs caractères indépendants - on dit aussi des classes -. Si ces classes sont au nombre de r, et si l'on désigne par k = 1, 2, 3 ... r le $k^{ième}$ caractère, par $n_k$ l'effectif ayant ce caractère dans l'échantillon i,

par $np_k$ la valeur théorique de cet effectif, on montre que lorsque la taille n des échantillons tend vers l'infini, la distribution

$$\chi_i^2 = \sum_{k=1}^{r} \frac{(n_k - np_k)^2}{np_k} \qquad (7.62)$$

tend vers la distribution théorique f $(\chi^2)$ donnée par (7.58) avec $\nu = r - 1$ (le nombre d'observations par échantillon est égal au nombre de classes r). (Pour une démonstration de ce théorème, cf. par exemple M. Fisz - Probability Theory and Mathematical Statistics, 1967, chez John Wiley and Sons).

En fait, il suffit que n soit assez grand pour que l'on puisse utiliser l'approximation de la loi f $(\chi^2)$. Dans la pratique, on considère qu'une condition essentielle est que les effectifs correspondant aux différentes classes soient supérieurs à 10, à la rigueur à 5. Il est possible toutefois de regrouper des classes pour satisfaire à cette condition.

Le test dit "du $\chi^2$" consiste généralement à tester le $\chi^2$ calculé à partir de l'échantillon, à l'aide d'un intervalle de confiance déterminé sur la loi f $(\chi^2)$. Il permet de résoudre entre autres, les problèmes typiques qui suivent.

## III. TEST DE CONFORMITE

Il s'agit de comparer une distribution d'un caractère observé sur un échantillon donné et une distribution théorique basée sur un modèle susceptible de décrire la probabilité d'observer une valeur du caractère. On dit parfois que l'on cherche à "ajuster" une distribution expérimentale à une distribution théorique.

L'hypothèse nulle consiste à supposer que l'on a concordance des deux distributions. Le critère du test est

$$\chi^2 = \sum_{k=1}^{r} \frac{(n_k - np_k)^2}{np_k} \qquad (7.62)$$

où $n_k$ est l'effectif (observé) ayant le caractère k, $p_k$ est la probabilité d'observer ce caractère et $np_k$ la valeur théorique de cet effectif. Sous l'hypothèse nulle, le $\chi^2$ ainsi calculé devrait être nul. Il sera d'autant plus grand que les deux distributions divergent.

Pour déterminer un intervalle de confiance sur la loi f ($\chi^2$), il est nécessaire de connaître le nombre de degrés de liberté $\nu$. D'une manière générale, $\nu$ est égal au nombre de comparaisons possibles, diminué du nombre de relations entre les effectifs théoriques, soit ici :

$$\nu = r - 1 \qquad (7.63)$$

puisqu'il existe une seule relation : celle qui exprime que la somme des effectifs est égale à la taille de l'échantillon.

Le seuil de signification $\alpha$ étant connu, on utilise la table 5 pour déterminer la valeur de $\chi^2_\alpha$ ayant la probabilité $\alpha$ d'être dépassée et on applique la règle de décision suivante :

1) $\chi^2 \leq \chi^2_\alpha$, l'hypoyhèse $H_o$ est valable
2) $\chi^2 > \chi^2_\alpha$, l'hypothèse $H_o$ est à rejeter.

. *Exemple 1*

On a lancé 200 fois 2 pièces l'une après l'autre et on *a observé les résultats suivants :*

| | PP | FF | PF | FP |
|---|---|---|---|---|
| Nombre de fois | 64 | 35 | 47 | 54 |

Peut-on dire que ces pièces sont normales (bien équilibrées) au seuil de 5 % ? au seuil de 1 % ?

Deux pièces lancées successivement peuvent tomber de 4

façons différentes : PP, FF, PF, FP. La probabilité d'avoir une configuration particulière est donc $\frac{1}{4}$. Pour 200 lancers, on peut espérer avoir $\frac{200}{4} = 50$ fois chaque configuration.

Le $\chi^2$ calculé est donc

$$\chi^2 = \frac{(64-50)^2}{50} + \frac{(35-50)^2}{50} + \frac{(47-50)^2}{50} + \frac{(54-50)^2}{50} = 8,92$$

Le nombre de classes est r = 4, et d'après (7.63) le nombre de degrés de liberté est $\nu = r - 1 = 3$.

Pour un seuil de 5 % et $\nu = 3$, la table 5 donne $\chi^2_{0,05} = 7,81$. On a donc $\chi^2 > \chi^2_{0,05}$. On ne peut retenir l'hypothèse nulle et admettre que les pièces sont normales.

Pour un seuil de 1 % et $\nu = 3$, on obtient $\chi^2_{0,01} = 11,34$. On a donc $\chi^2 < \chi^2_{0,01}$, ce qui ne permet plus de rejeter l'hypothèse $H_o$.

. *Exemple 2*

Les résultats des épreuves d'un examen à l'échelle nationale sont : 60 % de reçus, 25 % admissibles (admis à passer les épreuves orales) et 15 % éliminés.

Un établissement présente 160 élèves et obtient 75 reçus, 53 admissibles et 32 éliminés. Y a-t-il conformité entre ces résultats et ceux valables à l'échelle nationale ?

Pour le calcul de $\chi^2$, on peut utiliser le tableau suivant :

| | $n_k$ | $p_k$ | $np_k$ | $(n_k - np_k)^2$ | $\frac{(n_k - np_k)^2}{np_k}$ |
|---|---|---|---|---|---|
| Reçus | 75 | 0,6 | 96 | 441 | 4,593 |
| Admissibles | 53 | 0,25 | 40 | 169 | 4,225 |
| Eliminés | 32 | 0,15 | 24 | 64 | 2,666 |
| | n=160 | | | | $\chi^2 = 11,484$ |

Ici, r = 3 et $\nu = 2$. Les tables donnent :

$$\chi^2_{0,01} = 9,21 \qquad \chi^2_{0,001} = 13,82$$

D'où

$$\chi^2_{0,01} < \chi^2 < \chi^2_{0,001}$$

Il y a donc conformité au seuil de 1°/$_{oo}$, mais il convient de rejeter $H_o$ au seuil de 1 %.

## IV. TEST D'HOMOGENEITE

Il s'agit de comparer entre elles des distributions relatives à plusieurs échantillons afin de déterminer si les différences observées sont significatives, ou si elles sont dues à des fluctuations d'échantillonnage.

Dans ce cas, les données figurent en général sur un tableau à double entrée, où par exemple, les échantillons sont portés en lignes désignées par i = 1, 2, 3 ... $\ell$, et les classes en colonnes désignées par k = 1, 2, 3 ... r.

Pour chaque case ik du tableau, l'effectif théorique est estimé à l'aide du produit du total des effectifs de la ligne i par le total des effectifs de la colonne k, divisé par l'effectif total, soit

$$t_{ik} = \frac{\sum_{i=1}^{\ell} n_{ik} \sum_{k=1}^{r} n_{ki}}{n} \tag{7.64}$$

Le $\chi^2$ relatif à l'ensemble des données est :

$$\chi^2 = \sum_{i,\,k} \frac{(n_{ik} - t_{ik})^2}{t_{ik}} \tag{7.65}$$

et dans ce cas, $\ell$ étant le nombre d'échantillons et r le nombre de classes, le nombre de degrés de liberté est donné par

$$\nu = (\ell - 1)(r - 1) \tag{7.66}$$

Comme précédemment, on peut alors tester l'hypothèse nulle

qui consiste ici, à supposer que les échantillons sont homogènes.

. *Exemple*

Dans le cadre de l'exemple 2 précédent, deux établissements A et B ont obtenu les résultats qui suivent. Tester aux seuils de 10 % et de 5 % l'hypothèse qu'il n'y a pas de différence significative entre les résultats obtenus par les deux établissements.

| | Reçus | Admissibles | Eliminés | Effectif total |
|---|---|---|---|---|
| Etabl. A | 75 | 53 | 32 | 160 |
| Etabl. B | 140 | 62 | 38 | 240 |
| | T = 215 | T = 115 | T = 70 | T = 400 |

Le tableau à double entrée des $n_{ik}$, où k désigne la classe ($1 \leqslant k \leqslant 3$) et i désigne l'échantillon ($1 \leqslant i \leqslant 2$) est donné dans l'énoncé. L'expression (7.64) permet de dresser le tableau des $t_{ik}$. On a par exemple :

$$t_{11} = \frac{\sum_i n_{i1} \sum_k n_{1k}}{n}$$

$$= \frac{215 \times 160}{400} = 86$$

$$t_{12} = \frac{115 \times 160}{400} = 46, \text{ etc.}$$

| | | classes | | |
|---|---|---|---|---|
| | | k = 1 | k = 2 | k = 3 |
| | | Reçus | Admis-sibles | Eliminés |
| Echantillons i=1 | Etabl. A | 86 | 46 | 28 |
| Echantillons i=2 | Etabl. B | 129 | 69 | 42 |

On en déduit, d'après (7.65) :

$$\chi^2 = \frac{(75 - 86)^2}{86} + \frac{(53 - 46)^2}{46} + \frac{(32 - 28)^2}{28}$$
$$+ \frac{(140 - 129)^2}{129} + \frac{(62 - 69)^2}{69} + \frac{(38 - 42)^2}{42} = 5,07$$

Le nombre de degrés de liberté est d'après (7.66)

$$\nu = (\ell - 1)\,(r - 1) = 1 \times 2 = 2$$

auquel cas la table 5 donne

$$\chi^2_{0,10} = 4,605 \text{ et } \chi^2_{0,05} = 5,99.$$

On a :

$$\chi^2_{0,10} < \chi^2 < \chi^2_{0,05}$$

et par conséquent, il y a lieu de rejeter l'hypothèse d'homogénéité au seuil de 10 %, mais on peut la retenir au seuil de 5 %.

# E. AJUSTEMENT LINEAIRE - CORRELATION

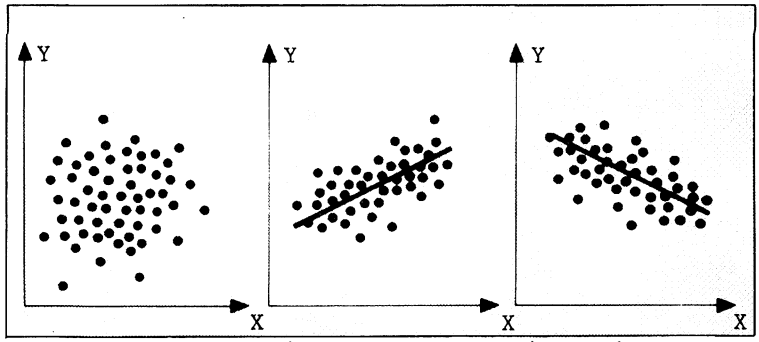
## I. INTRODUCTION

Considérons une population de taille n, et supposons que sur chaque élément de cette population, on effectue deux observations portant sur deux caractères différents associés aux deux variables aléatoires X et Y. Le problème de la corrélation est celui qui consiste à rechercher s'il existe une relation entre les variables X et Y.

A chaque élément i de l'échantillon, on peut associer un couple de valeurs $(X_i, Y_i)$ qui peut être représenté en coordonnées cartésiennes par un point du plan ayant pour abscisse $x = X_i$ et pour ordonnée $y = Y_i$. On obtient ainsi un nuage de n points constituant un diagramme de dispersion.

Ce diagramme peut être par exemple du type a) sur la figure 7.7, auquel cas il est difficile d'imaginer que les variables X et Y puissent être reliées, on dit qu'elles ne sont pas corrélées. Les points peuvent avoir tendance à se rap-

Figure 7.7




a) variables non corrélées  b) corrélation positive  c) corrélation négative

procher d'une même droite, on dit que la corrélation est linéaire. Si Y croît en même temps que X, la corrélation est dite positive (cas de la figure b), si Y décroît lorsque X croît, la corrélation est dite négative (cas de la figure c).

Cependant, les points du diagramme peuvent se rapprocher d'une courbe autre qu'une droite, auquel cas la corrélation est non-linéaire.

D'une manière générale, l'ajustement consiste à rechercher une fonction f (x) dont le graphe se rapproche le plus possible des points du diagramme. Etant donnée une valeur $X_i$ de X, il est évident que f ($X_i$) ne peut expliquer que partiellement la valeur expérimentale de $Y_i$. Il existe donc entre $X_i$ et $Y_i$ une relation de la forme

$$Y_i = f(X_i) + \varepsilon_i \tag{7.67}$$

où $\varepsilon_i$ apparaît comme un écart résiduel qui ne peut être expliqué par le modèle théorique traduit par f (x).

La méthode d'ajustement consiste à déterminer les paramètres de f (x) qui minimisent ces écarts. Celà revient à minimiser la somme des valeurs absolues de ces écarts $\Sigma\ |\varepsilon_i|$, ou encore

$$S = \sum_{i=1}^{n} (Y_i - f(X_i))^2 \tag{7.68}$$

C'est la méthode dite "des moindres carrés".

Nous nous limiterons ici au cas de l'ajustement linéaire, où l'on cherche à déterminer les paramètres a et b de la droite f (x) = a x + b, correspondant au minimum de la somme des carrés des écarts (7.68). Cette droite est appelée droite de régression.

## II. DROITE DE REGRESSION

Il s'agit donc de déterminer a et b pour que la somme

$$S(a,b) = \sum_{i=1}^{n} (Y_i - a X_i - b)^2 \qquad (7.69)$$

soit minimum. Il suffit d'annuler les dérivées partielles de S par rapport à a et à b ; on obtient un système de 2 équations à 2 inconnues conduisant aux expressions suivantes de a et b

$$a = \frac{\sum_{i=1}^{n} X_i Y_i - n \overline{X}\,\overline{Y}}{\sum_{i=1}^{n} X_i^2 - n \overline{X}^2}$$

$$b = \overline{Y} - a \overline{X} \qquad (7.70)$$

où $\overline{X} = E(X)$ et $\overline{Y} = E(Y)$. On voit que la droite de régression passe par le point $(\overline{X}, \overline{Y})$. En fait, quel que soit le diagramme de dispersion (même dans le cas de la fig. 7.7 a), on peut toujours trouver une droite minimisant S. Il convient donc d'étudier la signification de cet ajustement, ou encore sa précision.

## III. COEFFICIENT DE CORRELATION

Il est nécessaire de pouvoir disposer d'un nombre qui mesure la linéarité de la relation entre X et Y. Ce nombre, lié à la précision de l'ajustement, devra être d'autant plus grand que les écarts résiduels $\varepsilon_i$ sont faibles.

D'après (7.69) la somme S des carrés des écarts résiduels est donnée par

$$S = \sum_{i=1}^{n} (Y_i - a X_i - b)^2 \qquad (7.71)$$

On peut montrer que S peut encore s'écrire :

$$S = \sum_{i=1}^{n} \left[ (Y_i - \bar{Y}) - a\,(X_i - \bar{X}) \right]^2 \qquad (7.72)$$

$$= \sum_{i=1}^{n} (Y_i - \bar{Y})^2 - a^2 \sum_{i=1}^{n} (X_i - \bar{X})^2 \qquad (7.73)$$

c'est à dire comme la somme de 2 termes :

a) le 1er caractérisant la dispersion des données en l'absence de relation entre X et Y,

b) le 2ème conduisant à une diminution de cette dispersion lorsqu'on tient compte de la relation entre X et Y.

On définit le coefficient de corrélation (mesurant la précision de l'ajustement) par r (X, Y) tel que $r^2$ (X, Y) soit le rapport du 2ème terme au 1er terme, soit

$$r^2 (X, Y) = \frac{a^2 \sum_i (X_i - \bar{X})^2}{\sum_i (Y_i - \bar{Y})^2} \qquad (7.74)$$

Il est clair que le coefficient r est toujours compris entre -1 et +1. La valeur -1 correspond à une relation linéaire parfaite $Y = aX + b$ avec $a < 0$. De même, la valeur +1 indique une relation parfaite $Y = aX + b$ avec $a > 0$. Une valeur nulle ou voisine de 0 signifie qu'il n'existe aucune relation linéaire entre X et Y. Ainsi, dans le cas de la figure 7.7 a, même si le système d'équations (7.70) donne des valeurs pour a et b, on obtiendra un coefficient de corrélation r = 0, ce qui permettra de conclure à l'absence de toute relation linéaire entre X et Y.

Remarque

Le coefficient de corrélation r ne mesure aucunement une relation de causalité entre X et Y. Un coefficient voisin de l'unité n'implique pas qu'une variable entraîne l'autre, il

exprime simplement que les deux variables varient dans le même sens et que les écarts à la droite de régression sont faibles.

Le coefficient de corrélation r s'exprime également de la manière suivante :

$$r = \frac{\Sigma\ (X_i - \overline{X})\ (Y_i - \overline{Y})}{\sqrt{[\Sigma\ (X_i - \overline{X})^2][\Sigma\ (Y_i - \overline{Y})^2]}} \qquad (7.75)$$

ou encore :

$$r = \frac{\Sigma\ X_i\ Y_i - \dfrac{\Sigma\ X_i\ \Sigma\ Y_i}{n}}{\sqrt{[\Sigma\ X_i^2 - \dfrac{(\Sigma\ X_i)^2}{n}]\,[\Sigma\ Y_i^2 - \dfrac{(\Sigma\ Y_i)^2}{n}]}} \qquad (7.76)$$

qui constitue une formule pratique pour les calculs.

Remarque

En introduisant la covariance de (X, Y) donnée par

$$\text{cov. } (X, Y) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})\ (Y_i - \overline{Y})$$

et en remarquant que le dénominateur de (7.75) n'est autre que $\sigma_X\ \sigma_Y/n$, on a aussi

$$r\ (X, Y) = \frac{\text{cov } (X, Y)}{\sigma_X\ \sigma_Y}$$

. *Exemple*

Au cours des épreuves d'un examen, douze candidats ont obtenu les notes suivantes (sur 10) à deux matières différentes A et B.

| Candidat | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Matière A | 3 | 4 | 4 | 5 | 5 | 6 | 6 | 7 | 7 | 8 | 8 | 9 |
| Matière B | 3 | 3 | 5 | 4 | 5 | 5 | 6 | 5 | 6 | 6 | 8 | 7 |

1) Diagramme de dispersion

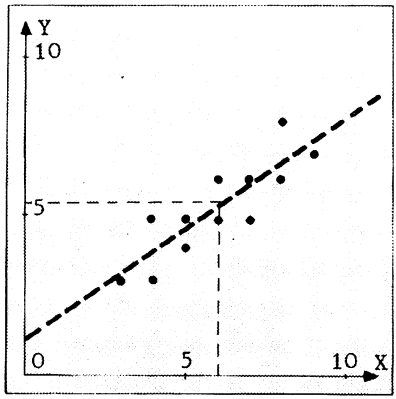



Figure 7.8

2) Coefficient de corrélation

Afin d'utiliser l'expression (7.76) pour le calcul de r, on dresse le tableau suivant :

| $X_i$ | $Y_i$ | $X_i^2$ | $Y_i^2$ | $X_i\ Y_i$ |
|---|---|---|---|---|
| 3 | 3 | 9 | 9 | 9 |
| 4 | 3 | 16 | 9 | 12 |
| 4 | 5 | 16 | 25 | 20 |
| 5 | 4 | 25 | 16 | 20 |
| 5 | 5 | 25 | 25 | 25 |
| 6 | 5 | 36 | 25 | 30 |
| 6 | 6 | 36 | 36 | 36 |

| $X_i$ | $Y_i$ | $X_i^2$ | $Y_i^2$ | $X_iY_i$ |
|---|---|---|---|---|
| 7 | 5 | 49 | 25 | 35 |
| 7 | 6 | 49 | 36 | 42 |
| 8 | 6 | 64 | 36 | 48 |
| 8 | 8 | 64 | 64 | 64 |
| 9 | 7 | 81 | 49 | 63 |
| $\Sigma$ 72 | 63 | 470 | 355 | 404 |

On a donc :

$$r = \frac{404 - \frac{72 \times 63}{12}}{\sqrt{(470 - \frac{72^2}{12})\ (355 - \frac{63^2}{12})}} = 0{,}856$$

Ce coefficient est relativement proche de l'unité. Il existe donc une relation de corrélation (fortement prononcée) entre les deux séries de notes, ce que l'allure du diagramme pouvait laisser prévoir.

3) Droite de régression

On a d'abord

$$\overline{X} = \frac{\sum_i X_i}{n} = \frac{72}{12} = 6$$

$$\overline{Y} = \frac{63}{12} = 5{,}25$$

d'où, en appliquant (7.70)

$$a = \frac{404 - [12 \times 6 \times 5{,}25]}{470 - [12 \times 6^2]} = 0{,}684$$

$$b = 5{,}25 - [0{,}684 \times 6] = 1{,}146$$

La droite de régression est donc donnée par

$$y = 0{,}684\ x + 1{,}146$$

Elle passe bien par le point moyen M (6 ; 5,25) (voir fig. 7.8)

## IV. TESTS D'HYPOTHESE

Considérons une population où les deux caractères X et Y sont distribués normalement et tous les échantillons de taille n susceptibles d'être extraits de cette population. On introduit un nouveau caractère r qui à l'échantillon i associe le coefficient de corrélation $r_i$ déterminé sur cet échantillon. On définit ainsi une distribution d'échantillonnage de r

$$\{r_1, r_2, r_3, \ldots, r_\ell\} \tag{7.77}$$

où $\ell$ désigne le nombre total d'échantillons.

On peut se demander si le coefficient $r_i$ peut servir à l'estimation par intervalle de confiance du coefficient de corrélation $\rho$ de la population-mère. Le caractère r n'étant pas nécessairement distribué normalement, on est amené à distinguer deux cas suivant que l'on a à tester l'hypothèse $\rho = 0$ ou l'hypothèse $\rho \neq 0$.

### 1. Test de l'hypothèse $\rho = 0$

Le problème posé est le suivant : on se demande si le coefficient de corrélation $r_o$ observé au niveau de l'échantillon, est compatible avec l'hypothèse d'absence de corrélation dans la population.

Hypothèse nulle $H_o$ : $\rho = 0$

Hypothèse alternative $H_1$ : $\rho \neq 0$

On peut procéder de deux manières différentes mais finalement équivalentes :

a) sous l'hypothèse $H_o$, la distribution de r est symétrique. On montre que la variable

$$t = \frac{r\sqrt{\nu}}{\sqrt{1 - r^2}} \tag{7.78}$$

où $\nu = n - 2$ est le nombre de degrés de liberté, suit une distribution de Student.

Le critère de test est $t_o = \dfrac{r_o \sqrt{\nu}}{\sqrt{1 - r_o^2}}$

A partir de la table 4 de la loi de Student on peut déterminer l'intervalle de confiance $I_{\alpha\nu}$ correspondant au seuil $\alpha$ et au nombre de degrés de liberté $\nu$, et appliquer la règle de décision suivante :

1) Si $t_o \notin I_{\alpha\nu}$, on rejette l'hypothèse $\rho = 0$

2) Si $t_o \in I_{\alpha\nu}$, on ne peut rejeter cette hypothèse.

b) Le critère de test est $r_o$. Sous l'hypothèse $\rho = 0$, la distribution de r est symétrique, de moyenne 0. La table 6 donne directement les valeurs du coefficient de corrélation déduit de la table 4 de la loi de Student, en utilisant l'équ. (7.78) :

$$r = \frac{t}{t^2 + \nu} \qquad (7.79)$$

Cette table permet de déterminer l'intervalle de confiance $I'_{\alpha\nu}$ pour $r_o$, sous l'hypothèse $\rho = 0$.

1) Si $r_o \notin I'_{\alpha\nu}$, on rejette $H_o$.

2) Si $r_o \in I'_{\alpha\nu}$, on ne peut rejeter $H_o$.

. *Exemple*

Pour un échantillon de taille 22, on a calculé un coefficient de corrélation de 0,30. Peut-on en déduire que le coefficient de corrélation de la population n'est pas nul, au seuil de 5 % ?

Hypothèse $H_o$ : $\rho = 0$, hypothèse $H_1 = \rho > 0$.

Il s'agit visiblement d'un test unilatéral.

Méthode a :

$$t_o = \frac{0{,}30 \sqrt{22 - 2}}{\sqrt{1 - 0{,}30^2}} = 1{,}41$$

La table 4 donne, pour $\nu = 20$ et un test unilatéral à

5 %, une valeur $t_\alpha = 1{,}725$. Par conséquent $t_o < t_\alpha$, on ne peut pas rejeter $H_o$ au seuil de 5 %.

Méthode b :

Le critère de test est $r_o = 0{,}30$.

La table 6, pour $\nu = 20$ et un test unilatéral à 5 %, donne une valeur $r_\alpha = 0{,}36$. Par conséquent, $r_o < r_\alpha$, et on arrive à la même conclusion que par la méthode précédente.

Remarque

Le test de l'hypothèse $\rho = 0$ constitue un test de signification : il s'agit de déterminer si la corrélation observée peut être expliquée par les fluctuations d'échantillonnage ou si elle est significative. Dans le 2ème cas, se pose alors le problème de l'interprétation de la relation entre X et Y, dont la solution n'est plus du domaine de la statistique.

## 2. Test de l'hypothèse $\rho = \rho_o$

Lorsque le test précédent conduit à rejeter l'hypothèse $\rho = 0$, le problème reste de savoir si on peut rejeter l'hypothèse que $\rho$ a une valeur donnée $\rho_o \neq 0$.

Dans ce cas, la distribution d'échantillonnage r est dissymétrique. On utilise alors la transformation dite "de Fisher" qui transforme r en une autre variable aléatoire

$$Z = \frac{1}{2} \text{Log}_e \left[ \frac{1 + r}{1 - r} \right]$$

$$= 1{,}151 \log_{10} \left[ \frac{1 + r}{1 - r} \right] \qquad (7.80)$$

qui suit approximativement une loi normale de moyenne

$$\mu_z = \frac{1}{2} \text{Log}_e \left[ \frac{1 + \rho_o}{1 - \rho_o} \right] \qquad (7.81)$$

et d'écart-type

$$\sigma_z = \frac{1}{\sqrt{n-3}} \tag{7.82}$$

La table 7 donne les valeurs de la variable de Fisher Z pour des valeurs de r comprises entre 0 et 1. Elle fournit de même, la valeur de $\mu_z$ correspondant à $\rho_o$.

La démarche du test est alors la suivante :

- formuler l'hypothèse nulle $H_o$ et l'hypothèse alternative $H_1$
- ayant calculé le coefficient $r_o$ de l'échantillon, au moyen de (7.76) par exemple, déterminer la variable $Z_o$ correspondante, soit à l'aide de la table 7, soit à partir de l'expression (7.80)
- opérer de la même manière pour trouver la moyenne $\mu_z$ correspondant à $\rho_o$
- le critère de test est alors

$$z = \frac{|Z_o - \mu_z|}{\sigma_z} = \frac{|Z_o - \mu_z|}{\sqrt{\frac{1}{n-3}}} \tag{7.83}$$

- déterminer l'intervalle de confiance $I_\alpha$ sur la table 3 de la loi normale
- appliquer la règle de décision habituelle.

Remarque

La transformation de Fisher est générale, elle peut s'appliquer au test de l'hypothèse $\rho = 0$ (test précédent).

. *Exemple 1*

Reprendre l'exemple du paragraphe précédent (échantillon de taille 22, ayant un coefficient $r_o$ de 0,30). Retrouver le résultat en appliquant la transformation de Fisher au test de l'hypothèse $\rho = 0$.

Hypothèse $H_o$ : $\rho = 0$ ; hypothèse $H_1$ : $\rho > 0$.

On a successivement :

$$Z_o = \frac{1}{2} \text{Log}_e \left[\frac{1 + 0,3}{1 - 0,3}\right] = 0,309 \text{ (table 7)}$$

$$\mu_z = 0$$

$$\sigma_z = \sqrt{\frac{1}{19}} \simeq 0,23$$

$$z = \frac{0,309}{0,23} = 1,34$$

La table 3 donne, pour un test unilatéral à 5 %, une valeur $t_\alpha = 1,645$. On a donc $z < t_\alpha$, on retrouve le résultat selon lequel on ne peut rejeter $H_o$ au seuil de 5 %.

. *Exemple 2*

Dans l'exemple du paragraphe 7 E III portant sur la recherche d'une corrélation linéaire entre les notes obtenues à deux matières différentes par un échantillon de 12 étudiants (diagramme de dispersion, fig. 7.8) on a calculé une valeur de $r_o = 0,856$.

Peut-on rejeter l'hypothèse que le coefficient de corrélation de la population soit aussi élevé que $\rho = 0,90$, au seuil de 5 % ?

Hypothèse $H_o$ : $\rho = 0,9$ ; hypothèse $H_1$ : $\rho < 0,9$.
La valeur $r_o = 0,856$ ne figure pas explicitement sur la table 7 de la variable Z de Fisher. On peut interpoler entre 0,85 et 0,86, mais on peut aussi bien appliquer l'expression (7.80) On trouve

$$Z = 1,151 \log_{10}\left[\frac{1 + 0,856}{1 - 0,856}\right] = 1,278$$

Pour $\rho = 0,9$, la table 7 donne $\mu_z = 1,472$

Par ailleurs, $\sigma_z = \sqrt{\frac{1}{n - 3}} = \sqrt{\frac{1}{9}} = 0,333$

d'où $z = \frac{|1,278 - 1,472|}{0,333} = 0,582$

La table 3 de la loi normale donne, pour un test unilatéral

à 5 %, une valeur de $t_\alpha = 1,645$. On a donc $z < t_\alpha$, on ne peut rejeter $H_o$.

. *Exemple 3*

Dans l'exemple précédent, déterminer les limites de confiance à 95 % du coefficient de corrélation de la population.

On a toujours

$Z = 1,278$ et $\sigma_z = 0,333$

L'intervalle de confiance à 95 % de $\mu_z$ est donc

$$1,278 - [1,96 \times 0,333] < \mu_z < 1,278 + [1,96 \times 0,333]$$

$$0,625 < \mu_z < 1,930$$

On en déduit, soit par interpolation sur la table 7, soit à partir de l'expression (7.80)

$$0,554 < \rho < 0,958$$

I. Dans une population de 5 objets, on étudie le caractère X associé au poids de chacun de ces objets. Les poids mesurés sont :

2,5 kg ; 2,53 kg ; 2,6 kg ; 2,62 kg ; 2,7 kg.

1°) Déterminer la valeur moyenne E (X) = M et l'écart-type $\sigma(X)$ de cette distribution de poids.

2°) Quel est le nombre $k_3$ d'échantillons (tirage exhaustif) de taille n = 3 que l'on peut obtenir à partir de la population de ces 5 objets ?

3°) Calculer la moyenne E (m) et l'écart-type $\sigma$ (m) de la distribution d'échantillonnage des moyennes des différents échantillons.

4°) Vérifier que E (m) = M et que $\sigma\ (m) = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}}$ où N est l'effectif total.

(Cas d'une population finie et d'un échantillon exhaustif).

---

SOLUTION

1°) La distribution des poids a pour valeur moyenne

$$M = E\ (X) = \frac{2,5+2,53+2,6+2,62+2,7}{5} = 2,59 \text{ kg}$$

L'écart-type est égal à

$$\sigma\ (X) = \sigma = \left[\overline{(X - M)^2}\right]^{1/2}$$

$$\sigma = \left[\frac{(2,5-2,59)^2+(2,53-2,59)^2+2,6-2,59)^2+(2,62-2,59)^2+(2,7-2,59)^2}{5}\right]^{1/2}$$

$$\sigma = \sqrt{0,005} = 0,0704$$

2°) Le nombre $k_3$ est le nombre de façons de choisir 3 objets parmi les 5, d'où

$$k_3 = C_5^3 = \frac{5!}{3!\ 2!} = 10$$

Les échantillons de poids sont :

(2,5;2,53;2,6) (2,5;2,53;2,62) (2,5;2,53;2,7) (2,5;2,6;2,62)
(2,5;2,6;2,7) (2,5;2,62;2,7) (2,53;2,6;2,62) (2,53;2,6;2,7)
(2,53;2,62;2,7) (2,6;2,62;2,7).

3°) La distribution d'échantillonnage de la moyenne est :

{2,543;2,550;2,577;2,573;2,600;2,607;2,583;2,610;2,617;2,640}

On en déduit successivement :

$$E\ (m) = \frac{\Sigma\ m_i}{k_3} = 2{,}590$$

$$V(m) = \overline{(m - E(m))^2}$$

$$V(m) = \frac{(-0{,}047)^2+(-0{,}040)^2+(-0{,}013)^2+(-0{,}017)^2+(0{,}01)^2}{10}$$
$$+ \frac{(0{,}017)^2+(-0{,}007)^2+(0{,}02)^2+(0{,}027)^2+(0{,}05)^2}{10}$$

$$V\ (m) = 0{,}0008$$

$$\sigma\ (m) = 0{,}0289$$

4°) On voit que

$$E\ (m) = M$$

On peut vérifier que

$$\sigma\ (m) = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}}$$

avec N = 5 et n = 3.

En effet :

$$\sigma\ (m) = \frac{\sqrt{0{,}005}}{\sqrt{3}} \quad \frac{\sqrt{2}}{\sqrt{4}} = 0{,}0289$$

■

II. Reprendre les données du problème I en supposant à présent des échantillons de taille 2 non exhaustifs, c'est à dire que le poids du même objet peut intervenir plusieurs fois.

1°) Quel sera le nombre $k_2$ d'échantillons de taille 2 ? Caractériser ces échantillons en faisant figurer les poids des objets.

2°) Dresser la distribution des valeurs moyennes. En calculer la valeur moyenne E (m) et l'écart-type $\sigma$ (m).

3°) Vérifier que E (m) = M et que $\sigma (m) = \frac{\sigma (X)}{\sqrt{n}}$

---

SOLUTION

1°) Il y a 5 choix possibles pour le premier poids et 5 choix possibles pour le deuxième, d'où $k_2 = 5 \times 5 = 25$. C'est aussi le nombre d'arrangements avec répétitions de 5 objets pris 2 à 2, soit $A_5^2 = 5^2 = 25$.

(2,5;2,5) (2,5;2,53) (2,5;2,6) (2,5;2,62) (2,5;2,7)
(2,53;2,5)(2,53;2,53)(2,53;2,6)(2,53;2,62)(2,53;2,7)
(2,6;2,5) (2,6;2,53) (2,6;2,6) (2,6;2,62) (2,6;2,7)
(2,62;2,5)(2,62;2,53)(2,62;2,6)(2,62;2,62)(2,62;2,7)
(2,7;2,5) (2,7;2,53) (2,7;2,6) (2,7;2,62) (2,7;2,7)

2°) La distribution des valeurs moyennes est :

{2,5;2,515;2,55;2,56;2,6;2,515;2,53;2,565;2,575;2,615;2,55;2,565;
2,6;2,61;2,65;2,56;2,575;2,61;2,62;2,66;2,6;2,615;2,65;2,66;2,7 }

On en déduit la valeur moyenne E (m) = 2,59 kg et la variance

$$V(m) = \overline{(m - E(m))^2} = \frac{(2,5-2,59)^2+(2,515-2,59)^2+\ldots+(2,7-2,59)^2}{25}$$

$$= 0,00248$$

d'où $\sigma$ (m) = 0,0498.

3°) On vérifie que l'on a encore E (m) = E (X) = M, mais cette fois $\sigma$ (m) = $\frac{\sigma (X)}{\sqrt{n}}$ avec n = 2. ■

---

III. Sur 10 personnes vaccinées, 8 réagissent positivement. On considère tous les échantillons de taille 9 pris dans cette population de 10 personnes.

1°) Quel est le nombre k d'échantillons en considérant les tirages exhaustifs ?

2°) Quelle est la distribution F des fréquences d'apparition de l'événement : "réaction positive" ?

Trouver sa valeur moyenne $\mu_F$ et sa variance $V_F = \sigma_F^2$

3°) En supposant que la distribution mère suit une loi binômiale, vérifier que

$$E (f) = p$$

et que

$$V (f) = \frac{p (1-p)}{n} \quad \frac{(N - n)}{(N - 1)}$$

p étant la probabilité élémentaire pour qu'une personne réagisse positivement au vaccin.

4°) Quelles seraient la moyenne et la variance de la distribution des fréquences si on avait considéré tous les échantillons non exhaustifs de taille 9 ?

---

SOLUTION

1°) Le tirage est exhaustif. On peut facilement calculer le nombre d'échantillons possibles. C'est le nombre de façons de choisir 9 personnes parmi 10, d'où

$$k = C_{10}^{9} = 10$$

Si on associe l'indice P aux personnes qui réagissent positivement et l'indice N aux personnes qui réagissent négativement,

les 10 échantillons seront constitués par :

2 fois (8 P, 1 N) car il y a $C_2^1 = 2$ façons de choisir 1 N parmi 2 N

et 8 fois (7 P, 2 N) car il y a $C_8^7 = 8$ façons de choisir 7 P parmi 8 P.

2°) La distribution des fréquences est :

$$F = \{\frac{8}{9};\ \frac{8}{9};\ \frac{7}{9};\ \frac{7}{9};\ \frac{7}{9};\ \frac{7}{9};\ \frac{7}{9};\ \frac{7}{9};\ \frac{7}{9};\ \frac{7}{9}\}$$

La valeur moyenne est

$$E\ (f) = \frac{2 \times \frac{8}{9} + 8 \times \frac{7}{9}}{10} = 0{,}8$$

La variance est

$$V\ (f) = \frac{\sum_{i=1}^{10} (f_i - E(f))^2}{10} = 0{,}001975$$

3°) La probabilité p est égale à

$$p = \frac{8}{10} = 0{,}8$$

On vérifie bien que E (f) = p et que

$$\frac{p\ (1-p)}{n} \times \frac{N - n}{N - 1} = \frac{0{,}8 \times 0{,}2}{9} \times \frac{10 - 9}{10 - 1} = 0{,}001975 = V\ (f)$$

4°) Pour un échantillonnage non exhaustif, il est fastidieux de calculer E (f) et $\sigma$ (f) à l'aide des fréquences des différents échantillons.

Toutefois, en appliquant les équations (7.10) et (7.11) on a :

$$E(f) = p = 0{,}8$$

$$V\ (f) = \frac{p\ (1-p)}{n} = \frac{0{,}8 \times 0{,}2}{9} = 0{,}01778$$

■

---

IV. Soit une variable aléatoire X représentant le nombre de globules rouges par $mm^3$ de sang d'individus pris au hasard dans une population donnée. On suppose que X suit une loi

normale. La valeur moyenne E (X) et l'écart-type σ (X) sont :

chez l'homme : $\overline{x_1} = 5\ 000\ 000/mm^3$ ; $\sigma_1 = 255100/mm^3$

chez la femme : $\overline{x_2} = 4\ 500\ 000/mm^3$ ; $\sigma_2 = 255100/mm^3$

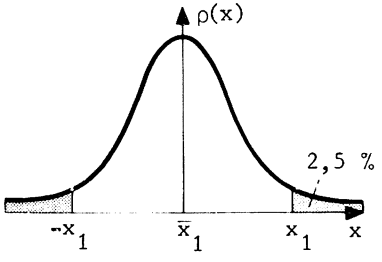
1°) Déterminer, successivement, pour les hommes et pour les femmes, l'encadrement du nombre de globules rouges, en se limitant à un risque d'erreur de 5 %.

2°) On choisit, au hasard, un homme et une femme dans cette population. Quelle est la probabilité pour que le nombre de globules rouges chez l'homme soit inférieur à celui de la femme ?

---

SOLUTION

1°) Pour l'homme




$x_1$ est tel que

$$P(\overline{x_1} \leqslant x \leqslant x_1) = 0{,}50 - 0{,}025 = 0{,}475$$

soit

$$G\left(\frac{x_1 - \overline{x_1}}{\sigma_1}\right) = 0{,}475$$

qui donne, d'après la table 3 :

$$\frac{x_1 - \overline{x_1}}{\sigma_1} = 1{,}96$$

par conséquent

$$x_1 = [1{,}96 \times 255\ 100] + 5\ 000\ 000 \simeq 5\ 500\ 000/mm^3$$

L'encadrement du nombre de globules rouges, chez l'homme, au risque de 5 % est donc :

$$4\ 500\ 000/mm^3 \leqslant x \leqslant 5\ 500\ 000/mm^3$$

Pour la femme, un raisonnement analogue conduit à la relation

$$\frac{x_2 - \overline{x_2}}{\sigma_2} = 1{,}96$$

d'où

$$x_2 = [1{,}96 \times 255\,100] + 4\,500\,000 \simeq 5\,000\,000/mm^3$$

L'encadrement du nombre de globules rouges, chez la femme, au risque de 5 % est :

$$4\,000\,000/mm^3 \leqslant x \leqslant 5\,000\,000/mm^3$$

2°) Si l'on prend comme nouvelle variable aléatoire y la différence des nombres de globules rouges chez l'homme et la femme, on doit alors calculer la probabilité pour que y soit négatif.

Les deux échantillons (hommes et femmes) correspondent à des statistiques indépendantes, la moyenne et l'écart-type de la nouvelle variable aléatoire sont :

$$\overline{y} = \overline{x_1} - \overline{x_2} = 500\,000\ /mm^3$$

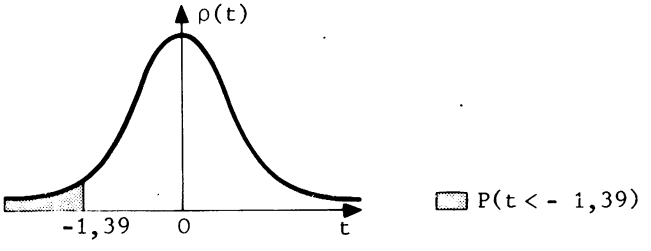
$$\sigma_y = \sqrt{\sigma_1^2 + \sigma_2^2} = 255\,100\ \sqrt{2} \simeq 360\,700\ /mm^3$$

On cherche la probabilité pour que $y < 0$ soit

$$P\ (y < 0) = P\ \left(t < \frac{0 - 500\,000}{360\,700}\right) = P\ (t < -1{,}39)$$

$$P\ (t < -1{,}39) = 0{,}50 - G(1{,}39) = 0{,}50 - 0{,}4177 = 0{,}0823.$$

La probabilité pour que le nombre de globules rouges soit plus petit chez un homme que chez une femme est de 0,0823.




■

I. On relève dans l'analyse du sang de 100 malades, un poids moyen de calcium $m_1$ = 120 mg avec un écart-type $\sigma_1$ = 10 mg. Ces 100 malades représentent un échantillon pris au hasard dans une population N de gens hospitalisés pour des anomalies sanguines ($N \gg 100$).

1°) En supposant que la distribution de poids de calcium est normale, quel est l'encadrement du poids de calcium pour un malade pris dans cet échantillon, si l'on accepte un risque de 5 % ?

2°) Donner l'intervalle de confiance à 95 % relatif au poids moyen de calcium pour l'ensemble des malades.

---

SOLUTION

1°) On appelle X la variable aléatoire qui représente le poids de calcium pour un malade. Cette variable X suit une loi normale. Puisque l'on est dans le cas d'un "grand échantillon" (n = 100), on peut considérer que la distribution est aussi normale dans l'échantillon. On a alors, d'après (7.21) :

$$120 - [1{,}96 \times 10] \leq X \leq 120 + [1{,}96 \times 10]$$

$$100{,}4 \text{ mg} \leq X \leq 139{,}6 \text{ mg}$$

2°) D'après la relation (7.19), on peut estimer l'écart-type par $S_i^* \simeq \sigma_i$. En appliquant (7.32) on a :

$$m_1 - t_\alpha \frac{\sigma_1}{\sqrt{n}} \leq M \leq m_1 + t_\alpha \frac{\sigma_1}{\sqrt{n}}$$

avec $m_1$ = 120 mg ; $\sigma_1$ = 10 mg ; $t_\alpha$ = 1,96 (risque de 5 %) car n = 100 > 30

L'intervalle de confiance du poids moyen de calcium dans la population-mère est :

$$120 - \frac{1,96 \times 10}{\sqrt{100}} \leq M \leq 120 + \frac{1,96 \times 10}{\sqrt{100}}$$

$$118,04 \text{ mg} \leq M \leq 121,96 \text{ mg} \quad \blacksquare$$

---

II. La mesure de la puissance de 5 machines à laver, issues d'une même chaîne de fabrication a donné les résultats suivants (en watts) :

3550 - 3560 - 3580 - 3600 - 3620

Entre quelles limites varie la puissance moyenne de l'ensemble des machines à laver de la série, au risque de 5 % ?

---

SOLUTION

On appelle E(X) = M la valeur moyenne, sur l'ensemble des machines à laver, du caractère X qui est la puissance de ces machines.

L'échantillon est petit (n=5), on applique la relation (7.33)

$$m_1 - t_{s\alpha} \frac{S_1^*}{\sqrt{n}} \leq M \leq m_1 + t_{s\alpha} \frac{S_1^*}{\sqrt{n}}$$

avec

$$m_1 = \frac{3550 + 3560 + 3580 + 3600 + 3620}{5} = 3582 \text{ w.}$$

D'après (7.20) : $S_1^* = \left[\frac{\sum_{j=1}^{5} X_j^2 - n\, m_1^2}{n - 1}\right]^{1/2} = \sqrt{820} = 28,636$

et $t_{s\alpha}$ donné par la table 4 pour $\alpha = 0,05$ et un nombre de degrés de liberté $\nu = 5 - 1 = 4$, est égal à :

$t_{s\alpha} = 2,776$

Par conséquent les limites pour la puissance moyenne sont :

$$3582 - 2,776 \times \frac{\sqrt{820}}{\sqrt{5}} \leq M \leq 3582 + 2,776 \times \frac{\sqrt{820}}{\sqrt{5}}$$

soit

$$3546 \leqslant M \leqslant 3618$$

■

---

III. Une étude sur les salaires mensuels de 50 ouvriers d'une usine a donné une moyenne de 3000 F et un écart-type de 500F. Quel risque prend-on en estimant la moyenne des salaires des 300 ouvriers employés dans l'usine à 3000 F $\pm$ 100 F ?

---

SOLUTION

Le caractère X étudié représente le salaire des ouvriers dans l'usine. Dans l'échantillon des 50 ouvriers, ce caractère a pour moyenne $m_1 = 3000$ F et pour écart-type $\sigma_1 = 500$ F. Pour l'ensemble des 300 ouvriers de l'usine on aura, si $E(X) = M$

(1) $3000 - 100 < M < 3000 + 100$

On est dans le cas d'une population finie et d'un échantillonnage exhaustif, on utilise donc la relation (7.34) :

$$(2) \quad m_1 - t_\alpha \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} < M < m_1 + t_\alpha \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

En comparant (1) et (2) ; on en déduit :

$$100 = t_\alpha \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \text{ avec } N = 300 \text{ et } n = 50.$$

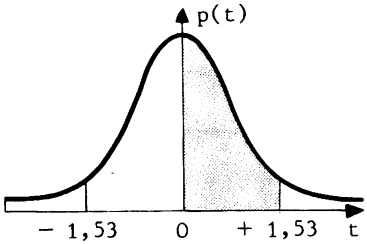
D'après (7.19) on a : $\sigma = \sqrt{\frac{n}{n-1}}\,\sigma_1 = \sqrt{\frac{50}{49}} \times 500.$

par suite :

$$100 = t_\alpha \frac{500}{\sqrt{49}} \sqrt{\frac{300 - 50}{300 - 1}} = 65,314\ t_\alpha$$

soit

$$t_\alpha = 1,53$$




La table 3 de la loi normale donne

$$G(1,53) = 0,4370$$

Le degré de confiance correspond à

$$0,4370 \times 2 = 0,8740$$

D'où le risque cherché :

$$1 - 0,8740 = 0,1260$$

■

IV. Les demi-pensionnaires d'une cantine scolaire ont été intoxiqués. On suppose que la maladie s'est déclarée, chez les élèves, de manière aléatoire.

Un examen, sur 100 enfants ayant mangé à la cantine ce jour-là, a révélé que 20 d'entre eux sont affectés de troubles digestifs. Quelle est la proportion d'individus intoxiqués parmi les 2000 élèves présents ce jour-là à la cantine, au risque de 3 % ?

SOLUTION

On dispose d'un échantillon de 100 élèves où le caractère X étudié ne peut prendre que deux valeurs : X = 1 pour les élèves affectés de troubles digestifs et X = 0 pour ceux qui ne le sont pas. La fréquence d'apparition de X = 1 est $f_1 = \frac{20}{100} = 0,2$.

On applique la relation (7.37) où l'on remplace

$$\sigma(f) = \sqrt{\frac{f_1(1-f_1)}{n}} \quad \text{par} \quad \sigma(f) = \sqrt{\frac{f_1(1-f_1)}{n}} \sqrt{\frac{N-n}{N-1}}.$$

On a donc :

$$f_1 - t_\alpha \sqrt{\frac{f_1(1-f_1)}{n}} \sqrt{\frac{N-n}{N-1}} \leq p \leq f_1 + t_\alpha \sqrt{\frac{f_1(1-f_1)}{n}} \sqrt{\frac{N-n}{N-1}}$$

avec N = 2000, n = 100 et $t_\alpha$ correspondant au risque de 3 %, dans la loi normale.

$$G(t_\alpha) = 0{,}50 - \frac{0{,}03}{2} = 0{,}485$$

ce qui entraîne $t_\alpha = 2{,}17$

d'où

$$0{,}2 - 2{,}17 \sqrt{\frac{0{,}2 \times 0{,}8}{100}} \sqrt{\frac{2000 - 100}{2000 - 1}} \leq p \leq 0{,}2 + 2{,}17 \sqrt{\frac{0{,}2 \times 0{,}8}{100}} \cdot \sqrt{\frac{2000-100}{2000-1}}$$

$0{,}1154 \leq p \leq 0{,}2846$.

Sur les 2000 élèves il y aura donc entre 231 et 569 élèves intoxiqués. ■

---

V. 1) La teneur en glucose dans le sang ou glycémie des sujets d'une population donnée est supposée distribuée suivant une loi normale de valeur moyenne $\mu$ = 1 g et d'écart-type 0,2 g. Quelle est pour un individu, la probabilité d'avoir une glycémie comprise entre 0,70 g et 1,2 g ? Pour 1000 personnes examinées, combien en moyenne auront une glycémie comprise entre les valeurs précédentes (0,70 g et 1,2 g).

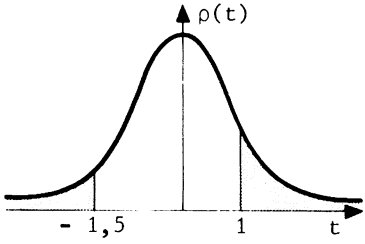
2) La détermination de l'intervalle de confiance (IC) à partir des glycémies obtenues pour un échantillon d'effectif n($n > 30$), a donné pour les sujets d'une deuxième population les valeurs suivantes pour les bornes de cet intervalle : 1,14 g et 1,26 g. Quelle est la valeur moyenne observée $\overline{X}$ des glycémies des sujets composant l'échantillon étudié ?

Au risque d'erreur 5 %, quel est l'effectif de l'échantillon étudié ? On prendra pour valeur de l'estimation de la variance de la distribution des glycémies de cette deuxième population : 0,09 $g^2$. ou $\sqrt{s^2} = 0{,}3$ g.

EXTRAIT D'EXAMEN PARIS-OUEST - 1976.

SOLUTION

1°) $P(0,70\ g \leq x \leq 1,2\ g) = P(\frac{0,70-1}{0,2} \leq t \leq \frac{1,2-1}{0,2}) = P(-1,5 \leq t \leq 1)$




D'après la propriété de a loi normale centrée réduite on a :

$P(-1,5 \leq t \leq 1) = G(1) + G(1,5) =$

$0,3413 + 0,4332$

$P(0,70\ g \leq x \leq 1,2\ g) = 0,7745.$

Pour 1000 personnes examinées, il y aura en moyenne $0,7745 \times 1000 \simeq 775$ personnes qui auront une glycémie comprise entre 0,70 g et 1,2 g.

2°) L'intervalle de confiance (1,14 g, 1,26 g) est centré sur $\overline{X}$. On aura donc $\overline{X} = \frac{1,26\ g + 1,14\ g}{2} = 1,20$ g. Par suite on a :

$1,14 < \overline{X} < 1,26$ qui, comparé à (7.30) donne :

$$m_1 + t_\alpha \frac{\sigma}{\sqrt{n}} = 1,26 \text{ et } m_1 - t_\alpha \frac{\sigma}{\sqrt{n}} = 1,14$$

qui entraîne :

$$2t_\alpha \frac{\sigma}{\sqrt{n}} = 1,26 - 1,14 = 0,12, \text{ soit } n = \frac{t_\alpha . \sigma}{0,06}$$

$t_\alpha$ correspondant au risque de 5 %, donc $t_\alpha = 1,96$ et $\sigma$ est estimé par 0,3 g. Par conséquent : $n = \left(\frac{1,96 \times 0,3}{0,06}\right)^2 \simeq 96$ sujets.

■

VI. Lorsque le dépouillement d'une élection présidentielle sera terminé, on saura que sur N votants, il y aura $N_A$ voix pour le candidat A et $N_B = N - N_A$ pour le candidat B. Mais les stations de radio veulent donner les résultats "dès 20 heures" en se basant sur les résultats de n votants (dont $n_A$ voix pour A et $n_B = n - n_A$ voix pour B), cet échantillonnage provenant de divers bureaux de votes représentatifs où des résultats partiels sont déjà connus.

1°) Quel est le nom de la loi de probabilité suivie par $n_A$ (on considère que l'échantillonnage correspond à un tirage au hasard).

2°) Sachant que pour cette loi la variance de $n_A$ est npq avec : $p = \frac{N_A}{N}$ et que $q = \frac{N_B}{N}$, calculer la variance $\sigma^2$ de la fréquence mesurée : $\alpha = \frac{n_A}{n}$.

3°) En fait, p et q sont des valeurs théoriques que l'on ne peut qu'estimer d'après les résultats de l'échantillonnage. Estimer la variance de $\alpha$.

4°) Quel est l'intervalle de confiance de $\alpha$ au seuil de 5 % (c'est-à-dire qu'il y a 95 chances sur 100 que la vraie valeur $N_A/N$ soit dans cet intervalle).

5°) Application numérique : $\alpha$ = 49 %, quelle valeur de n permettra d'annoncer le nom de l'élu avec 95 chances sur 100 de ne pas se tromper, discuter.

EXTRAIT D'EXAMEN LARIBOISIERE - 1975

---

SOLUTION

1°) L'évènement considéré est : sur n votants, il y a $n_A$ voix pour A et naturellement $n_B = n - n_A$ voix pour B. Si p est la probabilité élémentaire (sur 1 vote) pour que A soit choisi alors q = 1 - p est la probabilité élémentaire pour que B soit choisi.

Par suite

$P(n_A) = C_n^{n_A} p^{n_A}(1-p)^{n-n_A}$. La loi de probabilité est une loi binomiale.

L'échantillon étant représentatif de la population, les probabilités élémentaires d'un vote en faveur de A et en faveur de B sont respectivement égales à

$p = \frac{N_A}{N}$ $q = 1 - p = \frac{N_B}{N}$

2°) $V(n_A) = npq = n \frac{N_A N_B}{N^2}$.

Puisque $\alpha = \frac{1}{n} \times n_A$, on a $V(\alpha) = \frac{1}{n^2} V(n_A)$. Par suite $\sigma^2 = \frac{N_A N_B}{nN^2}$.

3°) Si l'on estime p et q par $\frac{n_A}{n}$ et $\frac{n_B}{n}$ alors :

$$V(n_A) = n \frac{n_A n_B}{n^2} \text{ et}$$

$$V(\alpha) = \sigma^2 = \frac{1}{n^2} V(n_A) = \frac{n_A n_B}{n^3}$$

4°) Si l'on suppose que l'échantillon vérifie $n > 30$ alors l'équation (7.38) donne :

$$\frac{n_A}{n} - t \left[\frac{n_A n_B}{n^3}\right]^{1/2} < p < \frac{n_A}{n} + t \left[\frac{n_A n_B}{n^3}\right]^{1/2}$$

avec t correspondant au risque de 5 %, c'est-à-dire t = 1,96.

5°) A.N : $\alpha = \frac{n_A}{n} = 0,49$ d'où $\frac{n_B}{n} = 1 - 0,49 = 0,51$.

D'après l'échantillon tiré au hasard, c'est le candidat B qui est élu. A l'échelon national, si l'on veut que B soit élu il faut que le pourcentage p de voix en faveur de A soit inférieur à 50 %. *Il suffit donc*

$$\frac{n_A}{n} + t_\alpha \left[\frac{n_A n_B}{n^3}\right]^{1/2} < 0,50.$$

D'où $0,49 + 1,96\sqrt{\frac{0,49 \times 0,51}{n}} < 0,50$ soit

$n > (\frac{1,96}{0,01})^2 (0,49 \times 0,51)$ soit $n > 9600$ voix.

A condition de prendre un échantillon dont la taille est au minimum de 9 600 voix, on pourra conclure que le candidat B a 95 chances sur 100 d'être élu. Il est bien évident que,

si l'on veut diminuer le risque d'erreur ($t_\alpha$ augmente) alors n doit être plus grand. Par exemple avec un risque de 1 % ($t_\alpha$ = 2,58), n devra être plus grand que 16634.

## EXERCICES CHAPITRE 7 C. TESTS DE SIGNIFICATION

I. Le gouvernement d'un pays a décidé de fixer, à l'échelon national, le prix d'un produit. Il tolère une distribution des prix suivant une loi normale de moyenne M = 100 F avec un écart-type de 10 F.

Ne pouvant vérifier les nombreux points de vente, il considère un échantillon de 36 points de vente où la moyenne des prix du produit est de 105 F. Doit-il considérer que ces prix sont en dehors de l'intervalle toléré, avec un seuil de signification de 5 % ?

SOLUTION

On a un échantillon de taille n = 36 dont la moyenne des valeur du caractère X (prix du produit) est $m_1$ = 105 F, alors que dans la population-mère E(X) = M = 100 F et $\sigma$ = 10 F.

- hypothèse nulle $H_o$ :

Les prix pratiqués dans l'échantillon appartiennent à l'intervalle toléré.

- Le critère de test est l'écart réduit déterminé par l'équation (7.44)

$$t_o = \frac{|m_1 - M|}{\frac{\sigma}{\sqrt{n}}} = \frac{|105 - 100|}{\frac{10}{\sqrt{36}}} = 3$$

- L'intervalle $I_\alpha$ correspond à un seuil de signification de 5 %.

Pour une loi normale, $I_\alpha$ est donné par la table 3

$I_\alpha = [-1,96 ; 1,96]$.

Par conséquent

$t_o \notin I_\alpha$

On rejette l'hypothèse nulle, on peut considérer que les prix sont en dehors de l'intervalle toléré.

■

II. Désirant juger le travail d'un ouvrier ajusteur, un chef d'atelier prélève un échantillon de 50 pièces métalliques dans sa production . On associe le caractère X à l'épaisseur de ses pièces. On doit avoir E(X) = 5 mm. Les résultats de la vérification sont portés dans le tableau suivant :

| $n_j$ | $x_j$ en mm |
|---|---|
| 5 | 4,8 |
| 15 | 4,9 |
| 20 | 5,0 |
| 10 | 5,1 |
| n = 50 | |

Cette fabrication est-elle conforme aux exigences, au seuil de 1 %.

SOLUTION

- Hypothèse nulle $H_o$ : La série est conforme aux normes exigées.

- Le critère de test sera l'écart réduit donné par (7.44) où $\sigma$ sera estimé par $S_1^*$ défini par (7.20).

$$t_o = \frac{|m_1 - M|}{\frac{S_1^*}{\sqrt{n}}}$$

avec $m_1 = \frac{[5 \times 4,8]+[15 \times 4,9]+[20 \times 5,0]+[10 \times 5,1]}{50} = 4,97$ mm

$$S_1^* = \left[\frac{\sum_{j=1}^{50} (x_j - m_1)^2}{n-1}\right]^{1/2} = \left[\frac{0,405}{49}\right]^{1/2} = 0,091$$

$$t_o = \frac{|4,97 - 5|}{\frac{0,091}{\sqrt{50}}} = 2,33$$

- L'intervalle $I_\alpha$ pour un seuil de signification de 1 % est

$$I_\alpha = [-2,6, + 2,6].$$

Par suite :

$$t_o \in I_\alpha .$$

On n'a donc aucune raison de considérer cet échantillon comme non conforme aux exigences de fabrication.

■

---

III. 20 % des ampoules provenant d'une certaine fabrication peuvent fonctionner plus de 200 heures. A la suite d'un nouveau traitement appliqué au filament, on constate, sur un échatillon de 100 ampoules, que 30 d'entre elles fonctionnent plus de 200 heures.

L'amélioration apparente est-elle significative au seuil de 5 % ?

SOLUTION

- Hypothèse $H_o$ : Le nouveau traitement n'a pas amélioré la durée de vie des lampes.

- Le critère de test est l'écart $|f_1 - p|$ exprimé en écart-type $\sigma(f)$, défini par (7.47)

$$t_o = \frac{|f_1 - p|}{\sqrt{\frac{pq}{n}}} \quad \text{avec } f_1 = \frac{30}{100} = 0,30 \; ; \; p = 0,20 \; ;$$

$q = 1 - p = 0,80 \; ; \; n = 100$

$$\text{soit } t_o = \frac{0,30 - 0,20}{\sqrt{\frac{0,20 \times 0,80}{100}}} = 2,5.$$

- Pour conclure à une amélioration, il convient d'appliquer un test unilatéral. La variable $t_\alpha$ lié au risque de 5 % est telle que $G(t_\alpha) = 0,50 - 0,05 = 0,45$ d'où $t_\alpha = 1,645$.

On a alors $t_o > t_\alpha$, on doit rejeter l'hypothèse $H_o$ et considérer que le nouveau traitement correspond probablement à une amélioration, au seuil de 5 %.

■

---

IV. Selon les lois de l'hérédité mendélienne on doit s'attendre à trouver une proportion théorique $p = 0,25$ de sourds muets de naissance lorsque les parents sont des consanguins porteurs d'un certain gène récessif.

On considère une population de nouveaux nés issus de tels mariages. Pour évaluer la proportion p' de sourds muets de naissance dans cette population on y prélève par tirage au sort des échantillons d'effectif $n = 300$.

1°) En se plaçant dans le cas de l'hypothèse nulle, $p' = p$, peut-on admettre que pour de tels échantillons la proportion $p_o$ de sourds muets qu'on y observe suit une loi normale ?

Si tel est le cas préciser la valeur des deux paramètres, moyenne et écart type de cette loi.

2°) a) Déterminer pour $p_o$ un intervalle de pari au risque $\alpha = 0,05$.

Peut-on accepter l'hypothèse nulle si pour n = 300 on observe :

b) 72 sourds muets

c) 96 sourds muets ; dans ce dernier cas préciser et interpréter le degré de signification P.

EXTRAIT D'EXAMEN PARIS OUEST - 1976

---

SOLUTION

1°) A l'intérieur d'un échantillon, la variable aléatoire (nombre de sourds muets) suit une loi binomiale d'effectif n = 300, avec une probabilité de réalisation égale, dans le cas de l'hypothèse nulle, à p = 0,25. Pour un effectif aussi grand, la valeur moyenne m = np = 300 × 0,25 = 75 de sourds muets montre que l'on peut approcher la loi binomiale par la loi normale de paramètres

$$m = 75 \text{ et } \sigma = \sqrt{npq} = \sqrt{mq} = \sqrt{75 \times (1-0,25)} = \sqrt{75 \times 0,75} = \sqrt{7,5 \times 7,5}$$

soit $\sigma = 7,5$

2°) a) D'après (7,37) on a :

$$p - t_\alpha\sqrt{\frac{pq}{n}} < p_o < p + t_\alpha\sqrt{\frac{pq}{n}} \text{ avec :}$$

$p = 0,25$ ; $q = 1 - p = 0,75$ ; $n = 300$ ; $t_\alpha = 1,96$

d'où $0,25 - 1,96\sqrt{\frac{0,25 \times 0,75}{300}} < p_o < 0,25 + 1,96\sqrt{\frac{0,25 \times 0,75}{300}}$

$0,201 < p_o < 0,299.$

L'intervalle de pari au risque de 5 % est

SOLUTION

- Hypothèse $H_o$ : Le nouveau traitement n'a pas amélioré la durée de vie des lampes.

- Le critère de test est l'écart $|f_1 - p|$ exprimé en écart-type $\sigma(f)$, défini par (7.47)

$$t_o = \frac{|f_1 - p|}{\sqrt{\frac{pq}{n}}} \quad \text{avec } f_1 = \frac{30}{100} = 0,30 \ ; \ p = 0,20 \ ;$$

$q = 1 - p = 0,80$ ; $n = 100$

$$\text{soit } t_o = \frac{0,30 - 0,20}{\sqrt{\frac{0,20 \times 0,80}{100}}} = 2,5.$$

- Pour conclure à une amélioration, il convient d'appliquer un test unilatéral. La variable $t_\alpha$ lié au risque de 5 % est telle que $G(t_\alpha) = 0,50 - 0,05 = 0,45$ d'où $t_\alpha = 1,645$.

On a alors $t_o > t_\alpha$, on doit rejeter l'hypothèse $H_o$ et considérer que le nouveau traitement correspond probablement à une amélioration, au seuil de 5 %.

■

---

IV. Selon les lois de l'hérédité mendélienne on doit s'attendre à trouver une proportion théorique $p = 0,25$ de sourds muets de naissance lorsque les parents sont des consanguins porteurs d'un certain gène récessif.

On considère une population de nouveaux nés issus de tels mariages. Pour évaluer la proportion p' de sourds muets de naissance dans cette population on y prélève par tirage au sort des échantillons d'effectif $n = 300$.

1°) En se plaçant dans le cas de l'hypothèse nulle, $p' = p$, peut-on admettre que pour de tels échantillons la proportion $p_o$ de sourds muets qu'on y observe suit une loi normale ?

Si tel est le cas préciser la valeur des deux paramètres, moyenne et écart type de cette loi.

2°) a) Déterminer pour $p_o$ un intervalle de pari au risque $\alpha = 0,05$.

Peut-on accepter l'hypothèse nulle si pour n = 300 on observe :

b) 72 sourds muets

c) 96 sourds muets ; dans ce dernier cas préciser et interpréter le degré de signification P.

EXTRAIT D'EXAMEN PARIS OUEST - 1976

---

SOLUTION

1°) A l'intérieur d'un échantillon, la variable aléatoire (nombre de sourds muets) suit une loi binomiale d'effectif n = 300, avec une probabilité de réalisation égale, dans le cas de l'hypothèse nulle, à p = 0,25. Pour un effectif aussi grand, la valeur moyenne m = np = 300 × 0,25 = 75 de sourds muets montre que l'on peut approcher la loi binomiale par la loi normale de paramètres

$m = 75$ et $\sigma = \sqrt{npq} = \sqrt{mq} = \sqrt{75 \times (1-0,25)} = \sqrt{75 \times 0,75} = \sqrt{7,5 \times 7,5}$

soit $\sigma = 7,5$

2°) *a) D'après (7.37) on a :*

$$p - t_\alpha \sqrt{\frac{pq}{n}} < p_o < p + t_\alpha \sqrt{\frac{pq}{n}} \text{ avec :}$$

$p = 0,25$ ; $q = 1 - p = 0,75$ ; $n = 300$ ; $t_\alpha = 1,96$

d'où $0,25 - 1,96 \sqrt{\frac{0,25 \times 0,75}{300}} < p_o < 0,25 + 1,96 \sqrt{\frac{0,25 \times 0,75}{300}}$

$0,201 < p_o < 0,299.$

L'intervalle de pari au risque de 5 % est

$I_\alpha = [0,201 \ ; \ 0,299]$.

b) Si, pour n = 300 nouveaux-nés, il y a 72 sourds muets alors

$$p'_1 = \frac{72}{300} = 0,24.$$

on a alors $p'_1 \in I_\alpha$ on ne doit pas rejeter $H_o$.

c) Pour 96 sourds muets on a $p'_2 = \frac{96}{300} = 0,32$.

Cette fois $p'_2 \notin I_\alpha$, on doit rejeter $H_o$ au risque de 5 % de se tromper.

On peut chercher le degré de signification P pour que $p'_2 \in I_\alpha$. On a

$$p'_2 = 0,32 = 0,25 + t_\alpha \sqrt{\frac{0,25 \times 0,75}{300}}$$

soit $t_\alpha = \dfrac{0,32 - 0,25}{\sqrt{\dfrac{0,25 \times 0,75}{300}}} = 2,8.$

A cette valeur de $t_\alpha$ correspond un degré de signification P = 0,99. On a au plus une chance sur cent de se tromper en rejetant $H_o$ .

■

V. Deux filiales fabriquent des piles électriques de 4,5 volts. Un échantillon $n_1$ = 100 piles de la filiale A a donné une durée de vie moyenne $m_1$ = 84 heures avec un écart-type $\sigma_1$ = 8 heures. Un échantillon $n_2$ = 150 piles prélevées dans la filiale B a donné une durée de vie moyenne $m_2$ = 80 heures avec un écart-type $\sigma_2$ = 5 heures. La différence des moyennes des durées de vie observées dans les deux échantillons, est-elle imputable à une supériorité de fabrication ou tout simplement à des fluctuations d'échantillonnage, au risque de 5 % .

SOLUTION

Hypothèse $H_o$ : Les deux fabrications sont comparables. $n_1$ et $n_2$ étant supérieurs 30, sous l'hypothèse $H_o$, la variable $m_2 - m_1$ est distribuée normalement, avec une valeur moyenne nulle et un écart-type donné par l'équation (7.49)

$$\sigma_{m_2 - m_1} = (\sigma^2_{m_2} + \sigma^2_{m_1})^{1/2} = \left(\frac{\sigma^2_A}{n_1} + \frac{\sigma^2_B}{n_2}\right)^{1/2} .$$

Pour $\sigma_A$ et $\sigma_B$ on peut utiliser les estimateurs suivants :

$$s^2_A = \sigma^2_1 \times \frac{n_A}{n_A - 1} \simeq \sigma^2_1$$

$$s^2_B = \sigma^2_2 \times \frac{n_B}{n_B - 1} \simeq \sigma^2_2 .$$

La variable réduite du test est donc :

$$t_o = \frac{|m_2 - m_1|}{\sigma_{m_2 - m_1}} = \frac{84 - 80}{\left(\frac{8^2}{100} + \frac{5^2}{150}\right)^{1/2}} = 4{,}45.$$

Devant décider d'une supériorité de fabrication, on utilise un test unilatéral.

Au seuil de 5 % : $t_\alpha = 1{,}645$.

Par suite $t_o > 1{,}645$, il faut donc rejeter l'hypothèse nulle et attribuer la différence des durées de vie à une autre cause que le hasard, probablement à une meilleure fabrication dans la filiale A.

■

VI. On a mesuré sur 6 champions cyclistes et 7 champions de natation une variable physiologique, la consommation maximum d'oxygène par minute rapportée au poids corporel (le nombre maximun au cours d'un effort de plus en plus intense, de mil-

lilitres d'oxygène absorbés par kilogramme de poids de corps et par minute). Voici les résultats :

Cyclistes : 73, 71, 69, 72, 74, 70.

Nageurs : 64, 69, 73, 68, 69, 67, 66.

L'estimation de la variance de la population des cyclistes est égale à 3,5, celle de la population des nageurs égale à 8. (On donne les valeurs de la quantité $\sum(X_1 - \overline{X})^2$ respectivement égales à 17,5 pour les cyclistes et 48 pour les nageurs).

On se demande si la consommation maximum d'oxygène diffère dans les deux populations de référence.

1°) On postulera les conditions d'application du test choisi.

2°) On estimera la variance commune dans le cas de l'hypothèse nulle $H_o$.

3°) Pour un risque $\alpha = 0,05$ (risque de première espèce) que peut-on conclure ?

EXTRAIT D'EXAMEN PARIS OUEST - 1976

---

SOLUTION

1°) Les effectifs $n_1 = 6$ et $n_2 = 7$ des 2 échantillons sont petits ( > 30), on choisit donc le test de Student avec pour hypothèse nulle $H_o$ : La consommation maxium est la même pour les 2 populations de référence. $H_1$ : La consommation maximum des cyclistes est plus grande que celle des nageurs (test unilatéral).

Pour simplifier le problème, on suppose que les deux populations sont normales et qu'elles ont la même variance $\sigma^2 = s^2$ donnée par (7.31).

2°) $$s^2 = \frac{(n_1-1)s_A^2 + (n_2-1)s_B^2}{(n_1-1) + (n_2-1)} \quad \text{avec}$$

$$(n_1-1)s_A^2 = n_1\sigma_1^2 \text{ et } (n_2-1)s_B^2 = n_2\sigma_2^2.$$

Calcul de $\sigma_1^2 = \overline{x_1^2} - \bar{x}_1^2$ et $\sigma_2^2 = \overline{x_2^2} - \bar{x}_2^2$

$$x_1 = \frac{73 + 71 + 69 + 72 + 74 + 70}{6} = 71{,}5_{ml/kg.mn} \quad \text{et}$$

$$x_2 = \frac{64 + 69 + 73 + 68 + 69 + 67 + 66}{7} = 68_{ml/kg.mn}$$

de même on aura :

$\overline{x_1^2} = 5115{,}167$ $\qquad$ $\overline{x_2^2} = 4630{,}857$

d'où

$\sigma_1^2 = 5115{,}67 - (71{,}5)^2 = 2{,}917$ et $\sigma_2^2 = 4630{,}857 - 68^2 = 6{,}857$

$(n_1-1)s_A^2 = 2{,}917 \times 6 = 17{,}5$ $\qquad$ $(n_2-1)s_B^2 = 6{,}857 \times 7 \simeq 48$

On a donc $s^2 = \frac{17{,}5 + 48}{5 + 6} = 5{,}95.$

3°) D'après (7.51) l'écart réduit est donné par

$$t_o = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} = \frac{71{,}5 - 68}{\sqrt{\frac{5{,}95}{6} + \frac{5{,}95}{7}}} = 2{,}58.$$

Pour un test unilatéral, à 5 % et $\nu = n_1 + n_2 - 2 = 6 + 7 - 2 = 11$, on a $t = 1{,}796$, d'où $t_o > t$, on doit rejeter $H_o$ et garder $H_1$.

■

---

VII. Une agence de publicité affirme qu'un produit d'entretien est efficace à plus de 98 % pour déboucher éviers et lavabos en 2 heures, quelle que soit la nature de l'obstruction.

Une association pour la défense du consommateur fait une enquête qui révèle que sur 100 lavabos bouchés seulement 90 sont débouchés en 2 heures. Doit-on faire un procès à l'agence de publicité, au risque de 2 % ?

SOLUTION

- Hypothèse $H_o$ : La probabilité d'efficacité du produit est supérieure à 0,98.

On appelle X la variable aléatoire représentant le nombre de lavabos ou éviers débouchés en 2 heures. Sous l'hypothèse $H_o$, on suppose que X suit une loi normale de paramètres $E(X) = m = 0,98 \times 100 = 98$ et $\sigma(X) = \sqrt{98 \times 0,02} = 1,4$.

Il s'agit donc de savoir si 90 est dans l'intervalle de confiance à 2 %

$$t_o = \frac{|90 - 98|}{1,4} \simeq 5,7.$$

Par ailleurs $I_\alpha = [-2,33 ; 2,33]$

d'où $t_o \notin I_\alpha$ .

On doit rejeter l'hypothèse nulle, la publicité est probablement mensongère.

■

VIII. On a étudié la consommation d'essence, sur 100 km, de voitures de même marque et de même cylindrée, choisies au hasard parmi deux chaînes de fabrication. Ces voitures sont conduites par le même conducteur sur le même circuit. Les résultats sont :

| *Nbs de voitures* de la chaîne A $N_1$ | *Consommation* en litres pour 100 km $x_1$ | *Nbs de voitures* de la chaîne B $N_2$ | *Consommation* en litres pour 100 km $x_2$ |
|---|---|---|---|
| 1 | 8 | 0 | 8 |
| 3 | 9 | 4 | 9 |
| 4 | 10 | 6 | 10 |
| 5 | 11 | 4 | 11 |
| 3 | 12 | 2 | 12 |

On suppose que la consommation, pour les deux chaînes de fabrication, suit une loi normale de même écart-type $\sigma$.

L'écart dans les 2 échantillons est-il dû à des fluctuations d'échantillonnage, au risque de 1 % ?

---

SOLUTION

- Hypothèse $H_o$ : L'écart est dû à des fluctuations d'échantillonnage.

Calcul de $E(x_1) = m_1$ , $E(x_2) = m_2$ , $\sigma_1$ et $\sigma_2$.

| $N_1$ | $x_1$ | $N_1x_1$ | $N_1x_1^2$ |
|---|---|---|---|
| 1 | 8 | 8 | 64 |
| 3 | 9 | 27 | 243 |
| 4 | 10 | 40 | 400 |
| 5 | 11 | 55 | 605 |
| 3 | 12 | 36 | 432 |
| 16 | | 166 | 1744 |

| $N_2$ | $x_2$ | $N_2x_2$ | $N_2x_2^2$ |
|---|---|---|---|
| 0 | 8 | 0 | 0 |
| 5 | 9 | 45 | 405 |
| 6 | 10 | 60 | 600 |
| 3 | 11 | 33 | 363 |
| 2 | 12 | 24 | 288 |
| 16 | | 162 | 1656 |

$E(x_1) = m_1 = 10,375$ $\qquad$ $E(x_2) = m_2 = 10,125$

$\sigma_1^2 = 1,359$ $\qquad$ $\sigma_2^2 = 0,984$

*On estime $\sigma$ par $s^*$ défini par (7.52) soit*

$$s^{*2} = \frac{(n_1-1)s_A^2 + (n_2-1)s_B^2}{(n_1-1) + (n_2-1)} \quad \text{avec } s_A^2 = \frac{n_1}{n_1-1}\sigma_1^2 = 1,450 \text{ et}$$

$$s_B^2 = \frac{n_2}{n_2-1}\sigma_2^2 = 1,050 \text{ d'où } s^{*2} = 1,250.$$

La variable réduite est, d'après (7.53)

$$t_o = \frac{m_1 - m_2}{\sqrt{\frac{2s^{*2}}{n_1}}} = \frac{0,250 \times 4}{\sqrt{2,5}} = 0,632$$

Cette valeur doit être comparée à $t_\alpha = 2,75$ (risque 1 % , $\nu = \lceil 16 \times 2 \rceil - 2 = 30$), déduit de la distribution de Student

On trouve que $t_o < t_\alpha$. On ne peut pas rejeter l'hypothèse nulle. L'écart observé est probablement dû aux fluctuations d'échantillonnage.

■

IX. On compare les effets d'un même traitement dans deux hôpitaux différents. Dans un premier hôpital, sur 100 malades traités, 70 montrent des signes de guérison alors que dans le deuxième hôpital, sur 150 malages traités, 100 sont sur le point de guérir. Quelle conclusion peut-on tirer, au risque de 5 % ?

SOLUTION

- L'hypothèse nulle est l'hypothèse qui consiste à dire que les 2 pourcentages sont les mêmes, la différence n'étant due qu'au hasard. Dans ce cas, les deux hôpitaux présentent, pour les malades traités, la même probabilité p de guérison.

- On doit comparer $t_o$ qui est égal d'après (7.55) à $t_o = \dfrac{f_1 - f_2}{\left[pq\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)\right]^{1/2}}$ avec le coefficient $t_\alpha = 1,645$ correspondant au risque de 5 %, pour un test unilatéral, et une loi normale

$$f_1 = \frac{70}{100} = 0,70 \; ; \; f_2 = \frac{100}{150} = 0,67 \; ;$$

p est estimé par l'équation (7.56) $p = \dfrac{n_1 f_1 + n_2 f_2}{n_1 + n_2} = \dfrac{70 + 100}{250} = 0,$

$$t_o = \frac{0,70 - 0,67}{(0,68 \times 0,32(\frac{1}{100} + \frac{1}{150}))^{1/2}} = 0,498.$$

On a donc $t_o < t_\alpha$, par conséquent la différence n'est pas significative. Elle doit être imputée au hasard.

I. On veut savoir si deux dés sont bien équilibrés. On jette 108 fois les deux dés et l'on obtient, en additionnant les 2 chiffres sortants, 12 fois un 6, 15 fois un 9 et 8 fois un 11. Que peut-on conclure sur la nature des dés, au seuil de 5 % ?

---

SOLUTION

Hypothèse $H_o$ : Les deux dés sont bien équilibrés.

La relation (7.61) donne l'expression de $\chi_c^2$ :

$\chi_c^2 = \sum \frac{(n_i - np_i)^2}{np_i}$ avec $\sum n_i = n = 108$ et $p_i$ = probabilité théorique d'avoir l'évènement i sous l'hypothèse $H_o$.

Les évènements à considérer sont, sur un jet de 2 dés, l'obtention d'un 6, d'un 9 ou d'un 11. Il est facile de voir qu'il y a successivement 5 façons d'avoir un 6 avec 2 dés, 4 façons d'avoir un 9 et 2 façons d'avoir un 11. Par suite $p(6) = 5 \times \frac{1}{6} \times \frac{1}{6} = \frac{5}{36}$ ; $p(9) = 4 \times \frac{1}{6} \times \frac{1}{6} = \frac{4}{36}$ et $p(11) = 2 \times \frac{1}{6} \times \frac{1}{6} = \frac{2}{36}$. Par suite

$$\chi_c^2 = \frac{(12 - 108 \times \frac{5}{36})^2}{108 \times \frac{5}{36}} + \frac{(15 - 108 \times \frac{4}{36})^2}{108 \times \frac{4}{36}} + \frac{(8 - 108 \times \frac{2}{36})^2}{108 \times \frac{2}{36}}$$

$$\chi_c^2 = \frac{(12 - 15)^2}{15} + \frac{(15 - 12)^2}{12} + \frac{(8 - 6)^2}{6} = 2{,}02.$$

La table du $\chi^2$ donne pour $\alpha = 0{,}05$ et $\nu = 3 - 1 = 2$

$\chi^2 = 5{,}99$.

Par conséquent $\chi_c^2 < \chi^2$, l'hypothèse $H_o$ reste valable. Les deux dés ne sont probablement pas truqués.

■

II. Appliquer le test du $\chi^2$, au seuil de 1 % à l'exercice II du chapitre 6C.

Doit-on conclure que la solution est homogène ?

---

SOLUTION

On a obtenu :

| k=nombre de bactéries | $n_1$=nombre expérimental de plaquettes | $n_2$=nombre théorique de plaquettes |
|---|---|---|
| 0 | 74 | 74 |
| 1 | 22 } 26 | 22,2 } 25,5 |
| 2 | 4 } | 3,3 } |

Hypothèse $H_o$ : la solution est homogène.

Calcul de $\chi^2_c$ : Pour obtenir des valeurs de $np_i$ supérieures à 5, on groupe les deux dernières classes

$$\chi^2_c = \frac{(74 - 74)^2}{74} + \frac{(26 - 25,5)^2}{25,5} = 0,0098.$$

La valeur de $\chi^2$ est donnée pour $\alpha = 0,01$ et $\nu = 2 - 1 = 1$ (après regroupement) par $\chi^2 = 6,63$.

On a donc $\chi^2_c < \chi^2$.

*On garde l'hypothèse $H_o$, la solution est probablement* homogène.

■

---

III. On pose 50 questions à un candidat qui ne doit répondre que par oui ou par non. A partir de combien de réponses exactes pourra-t-on considérer que ce candidat n'a pas répondu au hasard, au seuil de 2,5 % ?

---

SOLUTION .

Hypothèse $H_o$ : Le candidat répond au hasard.

Avec l'hypothèse $H_o$, il donnera alors

$50 \times \frac{1}{2} = 25$ bonnes réponses.

Si n est le nombre de réponses exactes données réellement par le candidat alors

$$\chi^2_c = \frac{(n - 25)^2}{25} + \frac{[(50-n) - 25]^2}{25} = 2\frac{(n - 25)^2}{25}$$

Il faut comparer $\chi^2_c$ à $\chi^2$ déterminé par le risque $\alpha = 0{,}025$ et le nombre de degrés de liberté $\nu = k - 1 = 2 - 1 = 1$.

La table 5 donne $\chi^2 = 5{,}02$. Il faut donc que $\chi^2_c \geq 5{,}02$ si l'on veut rejeter l'hypothèse $H_o$, soit

$$2\frac{(n - 25)^2}{25} \geq 5{,}02$$

avec naturellement la condition

$n > 25$ soit

$$(n - 25)^2 \geq \frac{5{,}02 \times 25}{2} = 62{,}75$$

d'où

$n - 25 \geq 7{,}92 \simeq 8$

Par suite $n \geq 25 + 8 = 33$.

A partir de 33 réponses exactes, au seuil de 2,5 %, on pourra dire que le candidat ne répond pas au hasard.

■

---

IV. Un référendum, à l'échelon national, a donné 55 % de oui, 40 % de non et 5 % de bulletins blancs ou nuls.

On recherche une ville test pour les élections à venir. On en trouve une qui, sur 10 000 votants, a donné les résultats suivants :

On en déduit, d'après (7,33)

$$\chi_c^2 = \frac{(45 - 42,5)^2}{42,5} + \frac{(5 - 7,5)^2}{7,5} + \frac{(40 - 42,5)^2}{42,5} + \frac{(10 - 7,5)^2}{7,5} =$$

$$\chi_c^2 = (0,147 + 0,833) \times 2 = 1,96.$$

Le nombre de degrés de liberté est

$\nu = (k-1)(\ell-1) = 1 \times 1 = 1$ d'où $\chi^2_{0,05} = 3,84$

$\chi_c^2 < \chi^2$, on peut donc accepter l'hypothèse $H_o$ au seuil de 5 %.

■

VI. Des promoteurs veulent implanter un grand centre commercial. Ils hésitent entre trois localités A, B et C qui ont respectivement 50000, 20000, 30000 habitants. Ils décident de procéder à un sondage sur 10 % de la population de chaque localité. A la fin de leur enquête, ils dressent le tableau suivant :

| i \ j | | 1 | 2 | 3 | |
|---|---|---|---|---|---|
| | réponse / localité | Favorable | Défavorable | sans opinion | Total des gens intérrogés |
| 1 | A | 3000 | 1000 | 1000 | $\sum_j n_{1j} =$ 5000 |
| 2 | B | 1000 | 700 | 300 | $\sum_j n_{2j} =$ 2000 |
| 3 | C | 2000 | 800 | 200 | $\sum_j n_{3j} =$ 3000 |
| | | $\sum_i n_{i1} = 6000$ | $\sum_i n_{i2} = 2500$ | $\sum_i n_{i3} = 1500$ | 10 000 |

Les trois localités ont-elles répondu me manière équivalente ?

---

SOLUTION

Hypothèse $H_o$ : Les trois localités ont répondu de manière équivalente. La relation (7.64) donne le tableau suivant

| i \ j | 1 | 2 | 3 |
|---|---|---|---|
| 1 | $\frac{5000 \times 6000}{10000} = 3000$ | $\frac{5000 \times 2500}{10000} = 1250$ | $\frac{5000 \times 1500}{10000} = 750$ |
| 2 | $\frac{2000 \times 6000}{10000} = 1200$ | $\frac{2000 \times 2500}{10000} = 500$ | $\frac{2000 \times 1500}{10000} = 300$ |
| 3 | $\frac{3000 \times 6000}{10000} = 1800$ | $\frac{3000 \times 2500}{10000} = 750$ | $\frac{3000 \times 1500}{10000} = 450$ |

D'où le calcul de $\chi^2_c$ déduit de la relation (7,65)

$$\chi^2_c = \frac{(3000 - 3000)^2}{3000} + \frac{(1000 - 1250)^2}{1250} + \frac{(1000 - 750)^2}{750} +$$

$$+ \frac{(1000 - 1200)^2}{1200} + \frac{(700 - 500)^2}{500} + \frac{(300 - 300)^2}{300} +$$

$$+ \frac{(2000 - 1800)^2}{1800} + \frac{(800 - 750)^2}{750} + \frac{(200 - 450)^2}{450} > 50$$

Le nombre de degrés de liberté est

$\nu = (k-1)(\ell-1) = (3-1)(3-1) = 4$ ce qui entraine $\chi^2_c > \chi^2$ quel que soit le risque pris. Les trois localités n'apprécient pas de manière identique l'installation d'un centre commercial.

VII. On observe, chez les cobayes, le nombre de réactions allergiques à une teinture de goudron administrée par injection intradermique. Cette teinture est administrée 4 fois à des intervalles de temps suffisamment espacés pour que les 4 administrations soient considérées sans effet les unes sur les autres.

Après chaque administration, la réaction de l'animal peut être positive ou non (on suppose que la probabilité d'une réaction positive est la même pour tous les cobayes). Pour chaque cobaye, on compte le nombre de réactions positives. On obtient les résultats suivants :

| Nombre de réactions positives par cobaye $x_j$ | Nombre de cobayes observés $n_j$ |
|---|---|
| 0 | 52 |
| 1 | 84 |
| 2 | 42 |
| 3 | 16 |
| 4 | 6 |
| Total | 200 |

A) Etude du caractère dichotomique : réaction positive, réaction négative.

1°) Calculer le nombre total d'injections, le nombre total et la fréquence expérimentale des réactions positives.
2°) Estimer ponctuellement la probabilité d'observer une réaction positive à la suite d'une seule administration d'une teinture.
3°) Construire l'intervalle de confiance de cette probabilité caractérisé par le niveau de probabilité P = 0,90.

B) Etude chez les cobayes.

1°) On veut ajuster un modèle binômial aux données expérimentales. Quels sont ses paramètres ?

2°) Calculer les probabilités $p_j$ d'observer les différentes valeurs $x_j$ (j=0,1,2,3,4). En déduire les effectifs théoriques de la distribution binômiale.

3°) L'ajustement binômial est-il valable ?

4°) Au début du problème, on a supposé que les 4 administrations successives de la teinture étaient sans effet les unes sur les autres au cours du temps. Que peut-on dire de cette hypothèse compte-tenu du résultat de la 3ème question ?

| $x_j$ | $n_j$ | $p_j$ | Effectifs théoriques (distribution binômiale) |
|---|---|---|---|
| 0 | 52 | | |
| 1 | 84 | 0,4116 | 82,32 |
| 2 | 42 | 0,2646 | 52,92 |
| 3 | 16 | 0,0756 | 15,12 |
| 4 | 6 | | |
| Total | 200 | | 200,00 |

N.B. : résultats partiels

EXTRAIT D'EXAMEN LYON - 1976

---

SOLUTION

A) Etude du caractère dichotomique

1°) - Le nombre total d'injections est : $n = 200 \times 4 = 800$

- Le nombre total de réactions positives est :

$n_+ = 1 \times 84 + 2 \times 42 + 3 \times 16 + 4 \times 6 = 240.$

- La fréquence expérimentale des réactions positives est

$f_+ = \frac{n_+}{n} = \frac{240}{800} = 0,30.$

2°) On peut estimer ponctuellement la probabilité d'observer une réaction positive à la suite d'une seule injection par la fréquence $f_+$

$p \simeq 0,30$.

3°) La relation (7,38) donne :

$$f_+ - t_\alpha\sqrt{\frac{pq}{n}} < p < f_+ + t_\alpha\sqrt{\frac{pq}{n}}.$$

Si l'on estime, sous le radical, p par $f_+$ et q par $1 - f_+$ on obtient :

$$f_+ - t_\alpha\sqrt{\frac{f_+(1-f_+)}{n}} < p < f_+ + t_\alpha\sqrt{\frac{f_+(1-f_+)}{n}}$$ avec t correspondant au risque de 0,10, soit, en supposant que la variable aléatoire suit une loi normale ; $t_\alpha = 1,645$.

Par suite :

$$0,30 - 1,645\sqrt{\frac{0,30 \times 0,70}{800}} < p < 0,30 + 1,645\sqrt{\frac{0,30 \times 0,70}{800}}$$

$$0,273 < p < 0,327$$

B) Etude chez les cobayes :

1°) Les paramètres de la loi binômiale pour la variable aléatoire x (= nombre de réactions positives d'un cobaye) sont : $n = 4$ et $p = 0,30$ (cf. A 2°)) d'où $m = np = 1,2$ et $\sigma = \sqrt{npq} = \sqrt{mq} = \sqrt{1,2 \times 0,7} = 0,916$.

2°) $P(x_i) = C_4^{x_i}(p)^{x_i}(1-p)^{4-x_i}$ on déduit :

$P(0) = 0,2401$ ; $P(1) = 0,4116$ ; $P(2) = 0,2646$ ; $P(3) = 0,0756$ ; $P(4) = 0,0081$; d'où les effectifs : $n_i = 200 \times P(x_i)$, soit

$n_o = 48,02$ ; $n_1 = 82,32$ ; $n_2 = 59,92$ ; $n_3 = 15,12$ ; $n_4 = 1,62$.

3°) Hypothèse $H_o$ = l'ajustement binômial est valable.

Hypothèse $H_1$ = l'ajustement binômial n'est pas valable.

La valeur de $\chi_c^2$ est donnée par (7,62) :

$$\chi_c^2 = \sum_{i=0}^{4} \frac{(n_i - np_i)^2}{np_i}$$ avec la condition $np_i \geq 5$.

Or $n_4 = 1,62$. On groupe donc les deux dernières classes. D'où

$$\chi^2_c = \frac{(52 - 48,02)^2}{48,02} + \frac{(84 - 82,32)^2}{82,32} + \frac{(42 - 52,92)^2}{52,92} + \frac{(22 - 16,74)^2}{16,74} = 4,27.$$

Si l'on prend le seuil de signification à 95 % , alors $\chi^2$ doit être pris pour $\alpha = 0,05$ et $\nu = 4 - 1 - 1 =$ (nombre de classes - 1) - nombre de paramètres estimés.

Ici le paramètre estimé est $p \simeq f_+$

$\chi^2 = 5,99 > \chi^2_c$.

On garde donc l'hypothèse $H_o$.

4°) Puisque l'ajustement binômial qui est probablement valable, nécessite que les 4 administrations successives de la teinture soient sans effet les unes sur les autres, on peut donc conclure que cette dernière hypothèse est probablement valable.

■

VIII. D'après les résultats au référendum du 27/4/69 les 3 communes suivantes peuvent-elles être considérées comme homogènes, en ce qui concerne leur électorat :

| | OUI | NON | ABSTENTION | TOTAL |
|---|---|---|---|---|
| Neuilly | 15000 | 16000 | 9000 | 40000 |
| Corbeil | 6000 | 10000 | 4000 | 20000 |
| Mazamet | 4000 | 4000 | 2000 | 10000 |
| | 25000 | 30000 | 15000 | 70000 |

EXTRAIT D'EXAMEN KREMLIN-BICETRE - 1975

SOLUTION

Hypothèse $H_o$ : Les 3 communes sont homogènes

Sous l'hypothèse $H_o$,

$p(\text{oui}) = \frac{25000}{70000} = \frac{5}{14}$ $\quad p(\text{non}) = \frac{30000}{70000} = \frac{6}{14}$ d'où $p(\text{ABS}) = \frac{3}{14}$.

On en déduit donc le tableau des $np_i$ pour chaque commune.

| | OUI | | NON | | ABSTENTION | | TOTAL |
|---|---|---|---|---|---|---|---|
| | $n_i$ | np(oui) | $n_i$ | np(non) | $n_i$ | np(abs) | |
| Neuilly | 15000 | 14286 | 16000 | 17143 | 9000 | 8571 | 40000 |
| Corbeil | 6000 | 7143 | 10000 | 8571 | 4000 | 4286 | 20000 |
| Mazamet | 4000 | 3571 | 4000 | 4286 | 2000 | 2143 | 10000 |
| | 25000 | 25000 | 30000 | 30000 | 15000 | 15000 | 70000 |

On en tire $\chi_c^2 = \frac{(15000 - 14286)^2}{14286} + \frac{(16000 - 17143)^2}{17143} +$

$+ \frac{(9000 - 8571)^2}{8571} + \frac{(6000 - 7143)^2}{7143} + \frac{(10000 - 8571)^2}{8571} +$

$+ \frac{(4000 - 4286)^2}{4286} + \frac{(4000 - 3571)^2}{3571} + \frac{(4000 - 4286)^2}{4286} +$

$+ \frac{(2000 - 2143)^2}{2143} \simeq 654.$

On voit donc que $\chi_c^2 > \chi^2$ quel que soit le risque choisi : on doit donc rejeter l'hypothèse $H_o$. Les différences observées sont significatives : l'électorat des 3 communes est différent.

I. Désirant savoir si l'accélération que subit un corps est nulle ou constante, un expérimentateur mesure la distance que parcourt cet objet en fonction du temps. Il trouve :

| $t_i$ (s) | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $\ell_i$ (m) | 30 | 110 | 220 | 375 |

1°) Tracer le diagramme de dispersion

2°) Déterminer le coefficient de corrélation dans l'hypothèse

a) d'un mouvement rectiligne uniforme $\ell = at$ (accélération nul

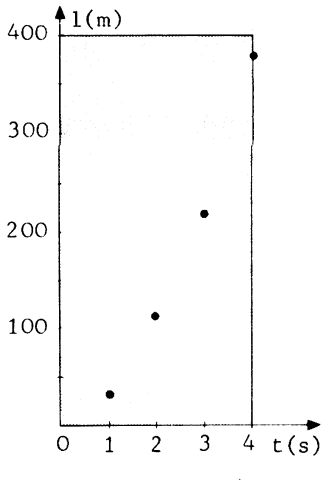
b) d'un mouvement uniformément accéléré $\ell = \alpha t^2$ (accélération constante)

3°) Représenter, dans les cas a) et b), la droite de régression.

---

SOLUTION

1°)




2°)

a) On suppose l'accélération nulle. Il faut ajuster alors les données à une droite $\ell = at + b$.

Afin d'utiliser l'expression (7.76) du coefficient de corrélation soit :

$$r = \frac{\sum_{i=1}^{4} t_i \ell_i - \frac{\sum_{i=1}^{4} t_i \sum_{i=1}^{4} \ell_i}{n}}{\left[\left(\sum_{i=1}^{4} t_i^2 - \frac{(\sum_{i=1}^{4} t_i)^2}{n}\right)\left(\sum_{i=1}^{4} \ell_i^2 - \frac{(\sum_{i=1}^{4} \ell_i)^2}{n}\right)\right]^{\frac{1}{2}}}$$

on dresse le tableau suivant :

| $t_i$ | $\ell_i$ | $t_i^2$ | $\ell_i^2$ | $t_i\ell_i$ |
|---|---|---|---|---|
| 1 | 30 | 1 | 900 | 30 |
| 2 | 110 | 4 | 12100 | 220 |
| 3 | 220 | 9 | 48400 | 660 |
| 4 | 375 | 16 | 140625 | 1500 |
| 10 | 735 | 30 | 202025 | 2410 |

$$r_a = \frac{2410 - \frac{10 \times 735}{4}}{\left[(30 - \frac{(10)^2}{4})(202025 - \frac{(735)^2}{4})\right]^{\frac{1}{2}}} = 0,989$$

b) Si l'on suppose l'accélération constante, la relation qui existe entre la distance parcourue $\ell$ et le temps mis pour la parcourir t, est une relation du type $\ell = \alpha t^2$

Si l'on pose X = Log t et Y = Log $\ell$

on a :

Log $\ell$ = Log( $t^2$) = Log $\alpha$ + 2 Log t, soit

Y = 2X + b.

On utilise toujours la relation (7.76), mais cette fois on dresse le tableau par rapport à $X_i$ et $Y_i$

| $t_i$ | $\ell_i$ | $X_i$=Log$t_i$ | $Y_i$=Log$\ell_i$ | $X_i^2$ | $Y_i^2$ | $X_iY_i$ |
|---|---|---|---|---|---|---|
| 1 | 30 | 0 | 3,401 | 0 | 11,568 | 0 |
| 2 | 110 | 0,693 | 4,700 | 0,480 | 22,095 | 3,258 |
| 3 | 220 | 1,099 | 5,394 | 1,207 | 29,091 | 5,926 |
| 4 | 375 | 1,386 | 5,927 | 1,922 | 35,128 | 8,216 |
| | | 3,178 | 19,422 | 3,609 | 97,882 | 17,400 |

$$r_b = \frac{17,4 - \frac{3,178 \times 19,422}{4}}{\left[(3,609 - \frac{(3,178)^2}{4})(97,882 - \frac{(19,422)^2}{4})\right]^{\frac{1}{2}}} = 1,000$$

En comparant les deux coefficients de corrélation, on peut rejeter l'hypothèse d'un mouvement uniforme (accélération nulle).

3°) Il faut déterminer dans les deux cas a) et b) la droite de regression

a) En supposant $\ell = at + b$, on obtient, en utilisant (7.70)

$$a = \frac{\sum_{i=1}^{4} t_i \ell_i - 4\left(\frac{\sum_{i=1}^{4} t_i}{4}\right)\left(\frac{\sum_{i=1}^{4} \ell_i}{4}\right)}{\sum_{i=1}^{4} (t_i)^2 - 4\left(\frac{\sum_{i=1}^{4} t_i}{4}\right)^2} = \frac{2410 - 4 \times \frac{10}{4} \times \frac{735}{4}}{30 - 4(\frac{10}{4})^2} = 114,5$$

$$b = \frac{\sum_{i=1}^{4} l_i}{4} - a \frac{\sum_{i=1}^{4} t_i}{4} = \frac{735}{4} - 114,5 \times \frac{10}{4} = -102,5.$$

D'où $\ell = 114,5\ t - 102,5$

b) Si le mouvement est uniformément accéléré alors la droite de régression sera obtenue par rapport aux variables Log l et Log t soit d'après 2°) $Y = aX + b$ avec

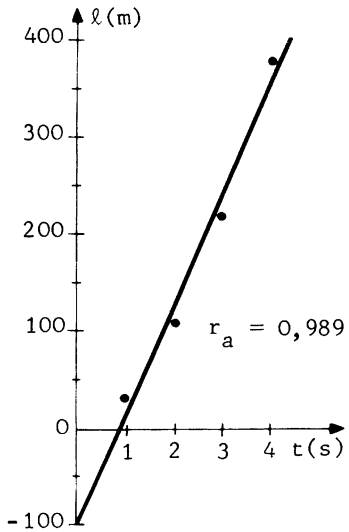
$$a = \frac{\sum_{i=1}^{4} X_i Y_i - 4\left(\frac{\sum_{i=1}^{4} X_i}{4}\right)\left(\frac{\sum_{i=1}^{4} Y_i}{4}\right)}{\sum_{i=1}^{4} (X_i)^2 - 4\left(\frac{\sum_{i=1}^{4} X_i}{Y}\right)^2} = \frac{17,4 - 4 \times \frac{3,178}{4} \times \frac{19,422}{4}}{3,609 - 4(\frac{3,178}{4})^2} = 1,816$$

$$b = \frac{\sum_{i=1}^{4} Y_i}{4} - a \frac{\sum_{i=1}^{4} X_i}{4} = \frac{19,422}{4} - 1,816 \times \frac{3,178}{4} = -3,413$$

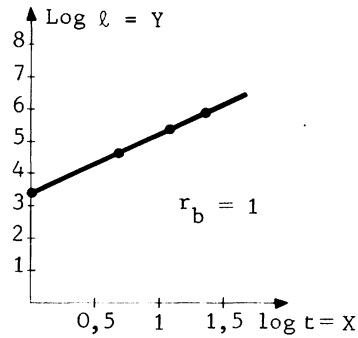
D'où $Y = 1,816X + 3,413$.

Représentations graphiques

a) Accélération nulle




b) Accélération constante




II. On s'intéresse aux variations de la valeur du taux de cholestérol sanguin avec le poids corporel chez des sujets sains.

1°) Dans la population des employés d'une administration, on tire au sort 100 personnes ; pour chacune de ces personnes on note :

son poids corporel,

son taux de cholestérol sanguin.

La corrélation observée entre le poids corporel et le taux de cholestérol sanguin vaut + 0,30.

Montrez que cette valeur diffère significativement de 0 au risque d'erreur 5 % ; précisez le degré de signification de cette différence.

2°) D'après ce résultat, le taux de cholestérol sanguin moyen des personnes de la population pesant 90 kg est-il

inférieur

égal

supérieur

à celui des personnes de la population pesant 70 kgs ?
Précisez à quelle population s'applique ce résultat.

EXTRAIT D'EXAMEN - COCHIN 1976

---

SOLUTION

1°) L'échantillon choisi a pour taille n = 100.

Les caractères étudiés sont : le taux de cholestérol sanguin et le poids corporel, donc leur nombre est 2.

On en déduit le nombre de degrés de liberté $\nu = 100 - 2 = 98$
Parmi les 2 méthodes possibles (cf. exemples du chapitre 7), on ne peut pas utiliser la table 4 car elle s'arrête à $\nu = 30$. On utilisera le critère de test "$r_o$" et le critère de test "z".

Hypothèse $H_o$ : $\rho = 0$, hypothèse $H_1$ : $\rho \neq 0$

a) Critère de test : $r_o = \dfrac{t}{\sqrt{t^2 + \nu}}$ (7.79)

En interpolant entre les valeurs $\nu = 90$ et $\nu = 100$, la table 6 donne pour $\alpha = 0,05$,

$$r_\alpha = 0,2050 - \frac{(0,2050 - 0,1946) \times 8}{10} \text{ soit } r_\alpha = 0,1967.$$

Par conséquent $r_o = 30 > r_\alpha$, on rejette donc $H_o$.

b) Critère de test : $z = \dfrac{|z - \mu_z|}{\sigma_z}$ (7.83).

D'après (7.80) $z = \frac{1}{2}\text{Log}(\frac{1 + r}{1 - r}) = \frac{1}{2}\text{Log}(\frac{1 + 0,3}{1 - 0,3}) = 0,3095.$

Pour l'hypothèse $H_o$ $(\rho = 0)$ $\mu_z = \frac{1}{2}\text{Log}(\frac{1 + \rho}{1 - \rho}) = 0.$

D'autre part, d'après (7.82) $\sigma_z = \frac{1}{\sqrt{100 - 3}} = 0,1015.$

Le critère de test est alors donné par l'équation (7.83) soit

$$z = \frac{|z - \mu_z|}{\sigma_z} = \frac{0,3095}{0,1015} \simeq 3,049.$$

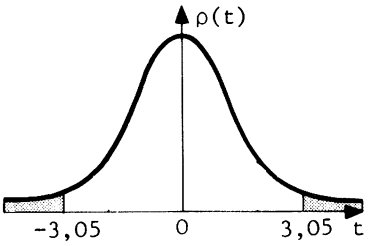
La table 3 donne, pour un risque de 5 % soit pour $G(t_\alpha) = 0,50 - \frac{0,05}{2} = 0,475$ une valeur $t_\alpha = 1,96$.

On a $z > t_\alpha$, on rejette donc $H_o$.

La valeur observée diffère significativement de 0 au risque de 5 %.

Le degré de signification s est déterminé par

$s = P(|t| > 3,049)$




soit $s = 2\,[0,5 - G(3,049)]$

avec $G(3,049) \simeq 0,4989$ d'après la table 3

$s = 2[0,5 - 0,4989] = 22 \times 10^{-4}$

2°) D'après le résultat $r_o = + 0,30$ est significatif ; cela signifie que le taux de cholestérol varie dans le même sens que le poids des sujets. Par suite, le taux de cholestérol *sanguin moyen des personnes pesant 90 kg est supérieur à* celui des personnes pesant 70 kg.

Il faut remarquer que l'échantillon a été pris dans une population des employés d'une administration. C'est donc à ce type de population que s'applique ce résultat.

■

III. On considère un échantillon de 10 personnes, pris au hasard dans une population et on mesure, pour chaque individu,

deux caractères différents : la longueur des bras et la taille. On trouve :

| individu | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| X=longueur des bras (cm) | 68 | 69 | 70 | 72 | 72 | 74 | 75 | 75 | 80 | 80 |
| Y = taille(m) | 1,60 | 1,68 | 1,70 | 1,68 | 1,75 | 1,80 | 1,80 | 1,85 | 1,90 | 1,75 |

1°) Que pouvez-vous dire à propos du coefficient de corrélation $\rho$ de la population, au seuil de 5 % ?

2°) Déterminer l'intervalle de confiance à 95 % du coefficient $\rho$ de la population.

---

SOLUTION

1°) Les donnés du tableau permettent de calculer facilement $E(X) = \overline{X}$ et $E(Y) = \overline{Y}$

On trouve $\overline{X} = 73,5$ cm et $\overline{Y} = 1,751$ m.

On utilise la relation (7.75) pour calculer r soit

$$r = \frac{\sum_{i=1}^{10} (X_i - \overline{X})(Y_i - \overline{Y})}{\left( \left( \sum_{i=1}^{10} (X_i - \overline{X})^2 \right) \left( \sum_{i=1}^{10} (Y_i - \overline{Y})^2 \right) \right)^{1/2}}$$

(voir le tableau de la page suivante)

On trouve donc que $r = \dfrac{2,645}{(156,5 \times 0,0723)^{1/2}} = 0,7864.$

Effectuons le test de l'hypothèse $\rho = 0$.

Hypothèse nulle $H_o$ : $\rho = 0$ Hypothèse alternative $H_1$ : $\rho > 0$

| $X_i$ | $Y_i$ | $(X_i-\bar{X})$ | $(Y_i-\bar{Y})$ | $(X_i-\bar{X})(Y_i-\bar{Y})$ | $(X_i-\bar{X})^2$ | $(Y_i-\bar{Y})^2$ |
|---|---|---|---|---|---|---|
| 68 | 1,60 | -5,5 | -0,151 | 0,8305 | 30,25 | 0,022[illegible] |
| 69 | 1,68 | -4,5 | -0,071 | 0,3195 | 20,25 | 0,005[illegible] |
| 70 | 1,70 | -3,5 | -0,051 | 0,1785 | 12,25 | 0,002[illegible] |
| 72 | 1,68 | -1,5 | -0,071 | 0,1065 | 2,25 | 0,005[illegible] |
| 72 | 1,75 | -1,5 | -0,001 | 0,0015 | 2,25 | $10^{-6}$ |
| 74 | 1,80 | 0,5 | 0,049 | 0,0245 | 0,25 | 0,002[illegible] |
| 75 | 1,80 | 1,5 | 0,049 | 0,0735 | 2,25 | 0,002[illegible] |
| 75 | 1,85 | 1,5 | 0,099 | 0,1485 | 2,25 | 0,009[illegible] |
| 80 | 1,90 | 6,5 | 0,149 | 0,9685 | 42,25 | 0,022[illegible] |
| 80 | 1,75 | 6,5 | -0,001 | -0,0065 | 42,25 | $10^{-6}$ |
| | | | | 2,645 | 156,5 | 0,072[illegible] |

On utilise comme critère $t_o = \dfrac{r_o \sqrt{\nu}}{\sqrt{1 - r_o^2}}$

avec $r_o = r = 0,7864$

$\nu = 10 - 2 = 8$ (n = 10, nombre de caractères étudiés = 2).

On en déduit $t_o = 3,6$.

La table 4 donne, pour $\nu = 8$ et pour un test unilatéral à 5 % une valeur $t_\alpha = 1,86$.

Par conséquent $t_o > t_\alpha$, on rejette l'hypothèse $H_o$.

2°) On fait la transformation de Fisher, soit d'après (7.80)

$$z = \frac{1}{2}\text{Log}\left(\frac{1 + r}{1 - r}\right) = \frac{1}{2}\text{Log}\left(\frac{1 + 0,7864}{1 - 0,7864}\right) = 1,0619.$$

Avec $\sigma_z = \dfrac{1}{\sqrt{n - 3}} = \dfrac{1}{\sqrt{7}} = 0,3780.$

L'intervalle de confiance à 95 % de $\mu_z$ est

$$1,0619 - 1,96 \times 0,3780 < \mu_z < 1,0619 + 1,96 \times 0,3780$$

soit $0,3210 < \mu_z < 1,8028$.

Pour trouver l'intervalle de confiance de $\rho$, on utilise la formule (7.81) : $\mu_z = \frac{1}{2}\text{Log}\left(\frac{1 + \rho}{1 - \rho}\right)$.

On trouve alors $\rho = \dfrac{e^{2\mu_z} - 1}{e^{2\mu_z} + 1}$.

Par conséquent, l'intervalle de confiance à 95 % du coefficient de corrélation $\rho$ de la population est :

$0,3104 < \rho < 0,9471$.

■

# Naissance du calcul des probabilités

Le calcul des probabilités est né tardivement, au milieu du dix-septième siècle. Très vite, ce qui est remarquable, il sera utilisé pour des études sociales. Ainsi, dès que la théorie semble suffisamment efficace, elle est appliquée là même où sa mise en oeuvre est la plus délicate.

I - Genèse

Bien que la notion de probabilité apparaisse dès le milieu du seizième siècle, puis chez Galilée (1564-1642) (cf. [4] p. 279), c'est la résolution par Blaise Pascal (1623-1662) du "problème des parties" (juste partage des mises entre deux partenaires interrompant le jeu avant sa fin) qui marque la véritable naissance du calcul des probabilités. Durant l'été 1654, Pascal correspond avec P. Fermat (1601-1665) et lui soumet ses raisonnements. S'il ne s'agit à proprement parler que de dénombrement, la solution de Pascal dégage les notions fondamentales de résultat possible et d'équiprobabilité. Sans l'énoncer explicitement, Pascal utilise la notion d'espérance mathématique. Il donnera ensuite une nouvelle démonstration de sa solution, utilisant ses résultats d'analyse combinatoire - qu'il fonde par son Traité du triangle arithmétique. A ses débuts, le calcul des probabilités n'est pas autre chose qu'une application de l'analyse combinatoire. Un grand pas est cependant franchi : les notions de combinatoire et de probabilité prennent des sens nouveaux.

Malgré quelques propriétés établies auparavant $\binom{n}{2} = \frac{n.(n-1)}{2}$ au troisième siècle après J.C., et $\binom{n}{p} = \frac{n.(n-1)\ldots(n-p+1)}{p.(p.1)\ldots 1}$ au troisième siècle (cf. [1] p. 64), l'analyse combinatoire n'apparait qu'avec le Traité du triangle arithmétique. Pour les philosophes du dix-septième siècle, - les mathématiques, comme la physique, sont encore, et resteront longtemps, une branche de la philosophie -, la combinatoire est essentiellement un mode de reflexion. Suivant Descartes dans ses tentatives pour introduire la rigueur et la justesse du raisonnement mathématique aux autres domaines de la connaissance, Leibniz (1646-1716) cherchera à construire une "combinatoire algébrique de la pensée". Cette notion de combinatoire et le pas qui est franchi par le Traité du triangle arithmétique s'expliquent par les objectifs des philosophes classiques.

L'Europe du dix-septième siècle est profondément religieuse, mais il n'y a plus unité de la foi. Les schismes et les controverses théologiques font que la théologie - ou une théologie - n'est plus la référence inéluctable de toute connaissance. La foi est cependant présente en l'homme "roi et prêtre", roi pour dominer la nature, prêtre pour louer Dieu. Il s'agit autant de déchiffrer le "langage mathématique que parle la nature" (Galilée) que de dominer. La croyance en une harmonie du monde créée par Dieu et accessible à l'intelligence humaine est universelle, bien que sous des formes différentes. Cette harmonie du monde n'est pas l'ordre d'une mécanique bien réglée. Les révolutions des astres, les cycles de la nature sont vus comme des chants de louange à Dieu, chants dont la musique serait perceptible à travers l'expression mathématique des régularités de ces cycles. Képler (1571-1630) donne un bon exemple de cette cosmogonie lorsqu'il pose une harmonie entre les planètes du système solaire et la Trinité, pour défendre le

système de Copernic. L'homme comprend la nature par sa propre raison : on est loin de la scolastique, et cette liberté n'est certainement pas étrangère à la Réforme. Un des fruits de cette liberté est que le philosophe se consacre maintenant à des réalisations pratiques utiles, tant par goût que par nécessité de subvenir à ses besoins car, rançon de cette liberté, les grands philosophes n'enseignent pratiquement jamais. C'est ainsi que les réflexions sur les jeux du hasard ne sont pas futiles au temps de Pascal. A travers un problème tel que celui des parties ses contemporains cherchent à percer les harmonies naturelles du hasard. Il s'agit de découvrir les harmonies du monde et non d'en construire une représentation. Le dix-septième siècle voit la naissance de la méthode scientifique dont Descartes disait qu'elle enseigne... à dénombrer exactement toutes les circonstances de ce que l'on cherche. On trouve chez de nombreux philosophes l'idée qu'en s'obstinant à combiner les pensées, fût-ce de manière anarchique, on est assuré d'obtenir un résultat intéressant. Le passage d'une telle notion de la combinatoire au calcul combinatoire représente donc une énorme restriction des objectifs, mais cela permet la création d'un outil efficace.

De même, la genèse du calcul des probabilités est manifeste par la restriction de la notion de probabilité. Lorsque Pascal, Fermat, Roberval et d'autres cherchent à résoudre un problème posé par un jeu du hasard, ils ne s'attachent pas à montrer qu'une opinion est plus ou moins probable, mais à calculer la proportion de chances de réalisation d'un évènement. Il n'est pas sans signification que Pascal se soit opposé aux doctrines des casuistes de la Compagnie de Jésus, selon lesquels une opinion peut se justifier dès lors qu'elle n'est pas totalement improbable (cf. Les Provinciales).

En 1657 Huyghens (1629-1695) publie un traité de calcul des

probabilités, Du calcul dans les jeux de hasard, où les notions de probabilité et d'espérance mathématique, dans le cas d'une variable aléatoire finie, sont explicitement définies. Huyghens n'utilise cependant pas l'analyse combinatoire dans les résolutions de problèmes qu'il expose.

## II - Premières applications du calcul des probabilités L'analyse au service du calcul des probabilités.

Durant le dix-huitième siècle, le calcul des probabilités se développe suivant deux axes : l'application de ce calcul au domaine social d'une part, l'introduction de l'analyse dans ce calcul d'autre part. Le calcul des probabilités se libère des jeux de hasard et cette libération favorise son extension.

1) Premières applications

Les premières applications de la théorie des probabilités furent des prolongements d'études d'arithmétique politique apparaissant à la fin du dix-septième siècle, qui sont de véritables études économiques développant les premières méthodes de statistique descriptive. Par arithmétique politique, on entend alors celle dont les opérations ont pour but des recherches utiles à l'art de gouverner les peuples : nombre d'habitants d'un pays, quantité de nourriture qu'ils doivent consommer, travail qu'ils peuvent accomplir, temps qu'ils ont à vivre, fertilité des terres, fréquence des naufrages, etc... (cf. Condorcet, article arithmétique politique de l'Encyclopédie). Nicolas Bernouilli (1687-1759) s'appuiera sur de tels recensements pour déterminer au bout de combien de temps on peut supposer, avec une probabilité fixée, qu'un absent d'un âge donné est mort, afin de pouvoir partager son héritage (cf. l'article Absent de l'Encyclopédie). Dans sa Dissertation sur l'usage de l'art de la conjecture dans le droit, de 1709,

N. Bernouilli expose également comment évaluer une rente viagère, quelle part peut réclamer un fils dans la succession de son père, lorsque ce dernier laisse une femme enceinte, suivant les probabilités qu'elle ait un ou plusieurs enfants, ainsi que la détermination du prix des assurances.

Les tentatives de Condorcet (1743-1794), en vue de la fondation d'une science sociale mathématique, constituent un cas particulier intéressant d'utilisation du calcul des probabilités. Au contraire de ses contemporains probabilistes pour qui les applications ne visent qu'à préciser les méthodes du calcul et à montrer son intérêt, Condorcet voit dans cette théorie l'outil permettant l'introduction des mathématiques dans l'étude de phénomènes sociaux, tel le choix collectif (cf. [2]). C'est ainsi qu'il étend la notion d'arithmétique politique, distinguant trois aspects : la recherche des données statistiques (les faits) et la réduction de ces données, leur interprétation, (les conséquences) et, finalement, la détermination de la probabilité de ces faits et de ces conséquences (cf. Condorcet, article Arithmétique politique de l'Enclyclopédie).

Le choix du domaine social comme terrain privilégié d'application du calcul des probabilités peut s'expliquer par l'évolution économique du dix-huitième siècle. Le système capitaliste naissant nécessite une connaissance macro-économique du marché *du travail, le développement du commerce et l'apparition de compagnies* d'assurances favorisent des études statistiques. Les travaux d'arithmétique politique se multiplient donc. Les probabilistes trouvent ainsi rassemblées des données susceptibles d'interprétation. Par ailleurs il ne faut pas négliger l'intérêt que portent les mathématiciens du dix-huitième siècle, comme les philosophes et les physiciens, aux questions politiques, économiques ou sociales.

Pour les philosophes des lumières, qui posent le prin-

cipe de l'unité de la nature, il ne suffit pas d'étendre les domaines de la connaissance, mais il faut construire un système, réduire à un petit nombre de règles ou notions générales chaque Science ou chaque Art en particulier,... renfermer en un système qui soit un les branches infiniment variées de la science humaine (cf. d'Alembert, Discours préliminaire à l'Encyclopédie). Car, pour d'Alembert comme pour Condorcet et leurs contemporains, pour peu qu'on ait réfléchi sur la liaison que les découvertes ont entr'elles, il est facile de s'apercevoir que les Sciences et les Arts se prêtent mutuellement des secours, et qu'il y a par conséquent une chaine qui les unit (cf. d'Alembert, Dicours préliminaire à l'Encyclopédie). Rapprocher divers domaines de la connaissance, appliquer une science à une autre sont donc des éléments de la méthode scientifique du dix-huitième siècle. Il est de l'intérêt de la vérité que [les sciences] se réunissent toutes, parce qu'il n'en est pas une seule qui ne tienne à toutes les autres parties du système scientifique par une dépendance plus ou moins immédiate. Il n'en est pas une où l'on puisse rompre la chaîne sans nuire aux deux portions que l'on aurait séparées (cf. Condorcet, Fragment sur l'Atlandide, cité par Roshdi Rashed [2] p. 15). Les applications du calcul des probabilités au domaine social s'inscrivent donc dans cette recherche, ou construction, d'une continuité entre les divers domaines de la connaissance. Ces applications furent cependant diversement acceptées. Pour certains, une connaissance n'est certaine que lorsqu'elle se plie aux modes de déduction mathématiques. Les mathématiques sont donc l'outil par excellence de l'organisation des sciences et de leurs extensions. Chez d'Alembert, ce point de vue n'est pas exempt d'un rigorisme quelque peu stérile. Pour d'autres, il existe deux ordres de vérité : mathématique et physique (cf. le Discours préliminaire à l'Histoire naturelle de Buffon).

Ces derniers limitent les possibilités d'application des mathématiques. Dans ce contexte, le projet de Condorcet présente une grande originalité. Il ne se donne pas en effet pour objectif de montrer les possibilités d'application du calcul des probabilités, mais se sert de ce calcul comme d'un outil permettant la constitution d'une science sociale. Il se préoccupe donc plutôt des fondements philosophiques des concepts de la théorie que du développement de la théorie elle-même.

Le calcul des probabilités fut également appliqué aux erreurs d'observations. L'estimation d'une grandeur à l'aide d'observations, écrit Gauss (1777-1855) avec une erreur plus ou moins grande, peut être comparée à un jeu de hasard où l'on ne peut que perdre, et où à chaque erreur correspond une perte (Gauss, Werke, cité par M. Loève, [4] p. 285). Legendre (1752-1833) et Gauss développeront la méthode des moindres carrés, Laplace (1749-1827) une méthode voisine.

2) L'analyse au service du calcul des probabilités

Ces applications du calcul des probabilités furent possibles grâce aux travaux de Jacques Bernouilli (1654-1705). Il fut le premier, à notre connaissance, à avoir réclamé l'application de ce calcul au domaine social. Mais c'est surtout son théorème, la loi des grands nombres découverte en 1689 et publiée dans l'Art de dresser des conjectures en 1713, qui marque la naissance du calcul des probabilités comme théorie mathématique. Si $S_n$ est le nombre de réalisations d'un évènement E au cours de n épreuves indépendantes et identiques, si p est la probabilité de réalisation de l'évènement E, alors la loi des grands nombres de J. Bernouilli énonce que, pour tout $\varepsilon > 0$,

$$\lim_{n \to +\infty} P\left(\left|\frac{S_n}{n} - p\right| < \varepsilon\right) = 1$$

En d'autres termes, lorsque le nombre d'épreuves augmente

les fréquences $\frac{s_n}{n}$ de réalisation de l'évènement E sont presque sûrement égales à la probabilité p de réalisation de cet évènement. Ce théorème fut à l'origine de travaux et de recherches qui se sont poursuivis durant le dix-huitième et le dix-neuvième siècle. Son importance vient également de ce qu'il fonde le point de vue fréquentiste en probabilité et se trouve ainsi à l'origine du développement de la statistique. On peut distinguer entre deux types d'interprétations : d'une part une estimation subjective du nombre de réalisations d'un évènement, d'autre part une estimation de la probabilité d'un évènement par la limite des fréquences empiriques observées. Cette seconde interprétation s'appuie donc sur la loi des grands nombres. Condorcet, qui fut l'un des premiers à se pencher sur cette distinction, présente ainsi ces deux sources de probabilités : l'une renferme les probabilités tirées de la considération de la nature même, et du nombre des causes ou des raisons qui peuvent influer sur la vérité de la proposition dont il s'agit; l'autre n'est fondée que sur l'exprérience du passé, qui peut nous faire tirer avec confiance des conjectures pour l'avenir, lors du moins que nous sommes assurés que les mêmes causes qui ont produit le passé existent encore, et sont prêtes à produire l'avenir ([2] p. 123). Le débat entre les types d'interprétations de la probabilité se poursuivent encore, bien que les mathématiciens *se rallient généralement au point de vue fréquentiste.*

A la suite de J. Bernouilli, et en utilisant son théorème, de Moivre (1667-1754) trouve la loi normale. Dans son traité Doctrine of chances, publié en 1718, fondamental à son époque, il explicite les concepts d'indépendance et de probabilité conditionnelle. Il démontre également le théorème de convergence vers la loi normale dans un cas particulier. Ce théorème montre que pour tout réel t

$$\lim_{n\to+\infty} P\left(\frac{s_n - np}{\sqrt{np(1-p)}} < t\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2}\, dx$$

(pour $0 < p < 1$).

Il sera démontré dans le cas général par Laplace. Le théorème actuel de convergence vers la loi normale ne sera correctement démontrée qu'à la fin du dix-neuvième siècle, à la suite des travaux de Tchebychef (1821-1894), Markov (1856-1922) et Liapounov (1857-1918). Auparavant, Laplace avait tenté d'étendre son théorème de convergence à des sommes de variables aléatoires indépendantes, et Poisson (1781-1840) l'avait étendu à des suites d'évènements indépendants.

A la suite de J. Bernouilli et de Moivre, Laplace développe les méthodes de l'analyse au service du calcul des probabilités : La méthode la plus générale et la plus directe pour résoudre les questions de probabilités consiste à les faire dépendre d'équations aux différences finies. (Théorie analytique des probabilités). En 1771, Laplace publie ses Recherches sur le calcul intégral aux différences infiniment petites et aux différences finies et un Mémoires sur les suites récurro-réccurentes et leurs usages dans la théorie du hasard, où il montre comment résoudre un problème tel que celui des parties par les équations aux différences finies.

3) Le problème de Bayes

Un dernier point important des recherches des probabilités du dix-huitième siècle est le problème de Bayes. Ce problème a intéressé tant les théoriciens de l'introduction de l'analyse dans le calcul des probabilités, comme Laplace, que les artisans des applications de la théorie, comme Condorcet. A la fin du dix-huitième siècle, Bayes (1702-1761) a cherché à calculer la probabilité des causes : "Donné" le nombre des réalisations et des non réalisations d'un évènement inconnu, "demandée" la

chance que la probabilité de réalisation de cet évènement dans une seule épreuve se trouve entre deux degrés de probabilités que l'on peut nommer (cf. Bayes, An essay towards solving a problem in the doctrine of chances, cité par Roshdi Rashed in [2] p. 55). Il s'agit donc d'approcher la probabilité d'un évènement à partir de ses fréquences de réalisation dans une suite d'épreuves. Ce problème est symétrique de la loi des grands nombres de J. Bernouilli, qui permet d'approcher les fréquences à partir de la probabilité d'un évènement. En 1774, Laplace publie un Mémoire sur la probabilité des causes où il énonce le principe de Bayes sous une forme voisine de la formulation actuelle : si un évènement E a été réalisé m fois dans n épreuves indépendantes et identiques, et si on suppose a priori la distribution de la probabilité p de l'évènement E uniforme sur $a,b$, alors la probabilité conditionnelle est égale à :

$$P(a \leq p \leq b/E \text{ a été réalisé m fois dans n épreuves}) = \frac{\int_a^b x^m(1-x)^{n-m}dx}{\int_0^1 x^m(1-x)^{n-m}dx}$$

Pour pouvoir être utilisé dans des problèmes pratiques, ce résultat suppose la connaissance de probabilité a priori. A la suite de Bayes, Laplace et Condorcet considèrent que l'on peut regarder ces probabilités comme égales lorsqu'on n'a aucune raison de penser que l'une est plus grande que les autres. Lorsqu'on ne voit aucune raison qui rende l'un plus probable que l'autre, parce que, quand bien même il y aurait une inégale possibilité entre eux, comme nous ignorons de quel côté est la plus grande, cette incertitude nous fait regarder l'un comme aussi probable que l'autre (cf. Laplace, Mémoire...). C'est le principe d'indifférence de Laplace.

Le débat entre subjectivistes et fréquentistes s'est très tôt cristallisé autour de l'utilisation de la formule de Bayes et du principe d'indifférence. Jusqu'à la fin du dix-neuvième siècle, les statisticiens ont cherché à éviter l'emploi du théorème de Bayes.

A la fin du dix-huitième siècle, les fondements du calcul des probabilités sont établis. Certes la théorie n'est pas encore affranchie du débat philosophique sur l'interprétation de la probabilité. Mais les probabilités mathématiques vont se developper sans égard pour ce débat.

F.B.

Bibliographie sommaire

[1] N. Bourbaki — Eléments d'histoire des mathématiques
Paris, 1974

[2] M. de Condorcet — Mathématique et société, choix de textes et commentaires par Roshdi Rashed
Paris, 1974

[3] Ch. Gouraud — Histoire du calcul des probabilités des origines à nos jours
Paris, 1848

[4] M. Loève — Calcul des probabilités in J. Dieudonné Abrégé d'histoire des mathématiques
Paris, 1978

[5] J. Montucla — Histoire des mathématiques
Paris, 1802

[6] H. Burrow Arton — The Enlightenment et ses adversaires
(Histoire de la philosophie, Encyclopédie de la Pléiade)
Paris, 1973

[7] J. Deprun — Philosophies et problématique des lumières
(Histoire de la philosophie, Encyclopédie de la Pléiade)
Paris, 1973

[8] G. Rodis Lewis — Descartes et anticartésiens français (Histoire de la philosophie, Encyclopédie de la Pléiade) Paris, 1973

[9] J. Ehrard — L'idée de nature en France à l'aube des lumières Paris, 1970

[10] A. Koyré — Etudes galiléennes Paris, 1966

[11] B. Matalon — Epistémologie des probabilités (Logique et connaissance scientifique, Encyclopédie de la Pléiade) Paris, 1976

# Tables

Table 1 : Loi binômiale $P(X = k) = C_n^k \, p^k \, q^{n-k}$.

| n | k | p | | | | |
|---|---|---|---|---|---|---|
| | | 0,10 | 0,20 | 0,30 | 0,40 | 0,50 |
| 2 | 0 | 0,8100 | 0,6400 | 0,4900 | 0,3600 | 0,2500 |
| | 1 | 0,1800 | 0,3200 | 0,4200 | 0,4800 | 0,5000 |
| | 2 | 0,0100 | 0,0400 | 0,0900 | 0,1600 | 0,2500 |
| 3 | 0 | 0,7290 | 0,5120 | 0,3430 | 0,2160 | 0,1250 |
| | 1 | 0,2430 | 0,3840 | 0,4410 | 0,4320 | 0,3750 |
| | 2 | 0,0270 | 0,0960 | 0,1890 | 0,2880 | 0,3750 |
| | 3 | 0,0010 | 0,0080 | 0,0270 | 0,0640 | 0,1250 |
| 4 | 0 | 0,6561 | 0,4096 | 0,2401 | 0,1296 | 0,0625 |
| | 1 | 0,2916 | 0,4096 | 0,4116 | 0,3456 | 0,2500 |
| | 2 | 0,0486 | 0,1536 | 0,2646 | 0,3456 | 0,3750 |
| | 3 | 0,0036 | 0,0256 | 0,0750 | 0,1536 | 0,2500 |
| | 4 | 0,0001 | 0,0016 | 0,0081 | 0,0256 | 0,0625 |
| 5 | 0 | 0,5905 | 0,3277 | 0,1681 | 0,0778 | 0,0312 |
| | 1 | 0,3280 | 0,4096 | 0,3602 | 0,2592 | 0,1562 |
| | 2 | 0,0729 | 0,2048 | 0,3087 | 0,3456 | 0,3125 |
| | 3 | 0,0081 | 0,0512 | 0,1323 | 0,2304 | 0,3125 |
| | 4 | 0,0004 | 0,0064 | 0,0284 | 0,0768 | 0,1562 |
| | 5 | 0,0000 | 0,0003 | 0,0024 | 0,0102 | 0,0312 |
| 6 | 0 | 0,5314 | 0,2621 | 0,1176 | 0,0467 | 0,0156 |
| | 1 | 0,3543 | 0,3932 | 0,3025 | 0,1866 | 0,0938 |
| | 2 | 0,0984 | 0,2458 | 0,3241 | 0,3110 | 0,2344 |
| | 3 | 0,0146 | 0,0819 | 0,1852 | 0,2765 | 0,3125 |
| | 4 | 0,0012 | 0,0154 | 0,0595 | 0,1382 | 0,2344 |
| | 5 | 0,0001 | 0,0015 | 0,0102 | 0,0369 | 0,0938 |
| | 6 | 0,0000 | 0,0001 | 0,0007 | 0,0041 | 0,0156 |
| 7 | 0 | 0,4783 | 0,2097 | 0,0824 | 0,0280 | 0,0078 |
| | 1 | 0,3720 | 0,3670 | 0,2471 | 0,1306 | 0,0547 |
| | 2 | 0,1240 | 0,2753 | 0,3177 | 0,2613 | 0,1641 |
| | 3 | 0,0230 | 0,1147 | 0,2269 | 0,2903 | 0,2734 |
| | 4 | 0,0026 | 0,0287 | 0,0972 | 0,1935 | 0,2734 |
| | 5 | 0,0002 | 0,0043 | 0,0250 | 0,0774 | 0,1641 |
| | 6 | 0,0000 | 0,0004 | 0,0036 | 0,0172 | 0,0547 |
| | 7 | 0,0000 | 0,0000 | 0,0062 | 0,0016 | 0,0078 |

Table 1 : Loi binômiale $P(X = k) = C_n^k \, p^k \, q^{n-k}$.

| n | k | p | | | | |
|---|---|---|---|---|---|---|
| | | 0,10 | 0,20 | 0,30 | 0,40 | 0,50 |
| 8 | 0 | 0,4305 | 0,1678 | 0,0576 | 0,0168 | 0,0039 |
| | 1 | 0,3826 | 0,3355 | 0,1977 | 0,0896 | 0,0312 |
| | 2 | 0,1488 | 0,2936 | 0,2965 | 0,2090 | 0,1094 |
| | 3 | 0,0331 | 0,1468 | 0,2541 | 0,2787 | 0,2188 |
| | 4 | 0,0046 | 0,0459 | 0,1361 | 0,2322 | 0,2734 |
| | 5 | 0,0004 | 0,0092 | 0,0467 | 0,1239 | 0,2188 |
| | 6 | 0,0000 | 0,0011 | 0,0100 | 0,0413 | 0,1094 |
| | 7 | 0,0000 | 0,0001 | 0,0012 | 0,0079 | 0,0312 |
| | 8 | 0,0000 | 0,0000 | 0,0001 | 0,0007 | 0,0039 |
| 9 | 0 | 0,3874 | 0,1342 | 0,0404 | 0,0101 | 0,0020 |
| | 1 | 0,3874 | 0,3020 | 0,1556 | 0,0605 | 0,0176 |
| | 2 | 0,1722 | 0,3020 | 0,2668 | 0,1612 | 0,0703 |
| | 3 | 0,0446 | 0,1762 | 0,2668 | 0,2508 | 0,1641 |
| | 4 | 0,0074 | 0,0661 | 0,1715 | 0,2508 | 0,2461 |
| | 5 | 0,0008 | 0,0165 | 0,0735 | 0,1672 | 0,2461 |
| | 6 | 0,0001 | 0,0028 | 0,0210 | 0,0743 | 0,1641 |
| | 7 | 0,0000 | 0,0003 | 0,0039 | 0,0212 | 0,0703 |
| | 8 | 0,0000 | 0,0000 | 0,0004 | 0,0035 | 0,0176 |
| | 9 | 0,0000 | 0,0000 | 0,0000 | 0,0003 | 0,0020 |
| 10 | 0 | 0,3487 | 0,1074 | 0,0282 | 0,0060 | 0,0010 |
| | 1 | 0,3874 | 0,2684 | 0,1211 | 0,0403 | 0,0098 |
| | 2 | 0,1937 | 0,3020 | 0,2335 | 0,1209 | 0,0439 |
| | 3 | 0,0574 | 0,2013 | 0,2668 | 0,2150 | 0,1172 |
| | 4 | 0,0112 | 0,0881 | 0,2001 | 0,2508 | 0,2051 |
| | 5 | 0,0015 | 0,0264 | 0,1029 | 0,2007 | 0,2461 |
| | 6 | 0,0001 | 0,0055 | 0,0368 | 0,1115 | 0,2051 |
| | 7 | 0,0000 | 0,0008 | 0,0090 | 0,0425 | 0,1172 |
| | 8 | 0,0000 | 0,0001 | 0,0014 | 0,0106 | 0,0439 |
| | 9 | 0,0000 | 0,0000 | 0,0001 | 0,0016 | 0,0098 |
| | 10 | 0,0000 | 0,0000 | 0,0000 | 0,0001 | 0,0010 |

Table 2 : Loi de Poisson $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$

| k | λ 0,1 | 0,2 | 0,3 | 0,4 | 0,5 | 0,6 | 0,7 | 0,8 | 0,9 | 1,0 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0,9048 | 0,8187 | 0,7408 | 0,6703 | 0,6065 | 0,5488 | 0,4966 | 0,4493 | 0,4066 | 0,3679 |
| 1 | 0,0905 | 0,1637 | 0,2222 | 0,2681 | 0,3033 | 0,3293 | 0,3476 | 0,3595 | 0,3659 | 0,3679 |
| 2 | 0,0045 | 0,0164 | 0,0333 | 0,0536 | 0,0758 | 0,0988 | 0,1217 | 0,1438 | 0,1647 | 0,1839 |
| 3 | 0,0002 | 0,0011 | 0,0033 | 0,0072 | 0,0126 | 0,0198 | 0,0284 | 0,0383 | 0,0494 | 0,0613 |
| 4 | 0,0000 | 0,0001 | 0,0002 | 0,0007 | 0,0016 | 0,0030 | 0,0050 | 0,0077 | 0,0111 | 0,0153 |
| 5 | 0,0000 | 0,000 | 0,0000 | 0,0001 | 0,0002 | 0,0004 | 0,0007 | 0,0012 | 0,0020 | 0,0031 |
| 6 | 0,0000 | 0,000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0001 | 0,0002 | 0,0003 | 0,0005 |
| 7 | 0,0000 | 0,000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0001 |

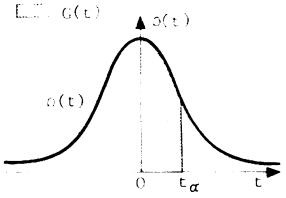
| k | λ 1,1 | 1,2 | 1,3 | 1,4 | 1,5 | 1,6 | 1,7 | 1,8 | 1,9 | 2,0 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0,3329 | 0,3012 | 0,2725 | 0,2466 | 0,2231 | 0,2019 | 0,1827 | 0,1653 | 0,1496 | 0,1353 |
| 1 | 0,3662 | 0,3614 | 0,3543 | 0,3452 | 0,3347 | 0,3230 | 0,3106 | 0,2975 | 0,2842 | 0,2707 |
| 2 | 0,2014 | 0,2169 | 0,2303 | 0,2417 | 0,2510 | 0,2584 | 0,2640 | 0,2678 | 0,2700 | 0,2707 |
| 3 | 0,0738 | 0,0867 | 0,0998 | 0,1128 | 0,1255 | 0,1378 | 0,1496 | 0,1607 | 0,1710 | 0,1804 |
| 4 | 0,0203 | 0,0260 | 0,0324 | 0,0395 | 0,0471 | 0,0551 | 0,0636 | 0,0723 | 0,0812 | 0,0902 |
| 5 | 0,0045 | 0,0062 | 0,0084 | 0,0111 | 0,0141 | 0,0176 | 0,0216 | 0,0260 | 0,0309 | 0,0361 |
| 6 | 0,0008 | 0,0012 | 0,0018 | 0,0026 | 0,0035 | 0,0047 | 0,0061 | 0,0078 | 0,0098 | 0,0120 |
| 7 | 0,0001 | 0,0002 | 0,0003 | 0,0005 | 0,0008 | 0,0011 | 0,0015 | 0,0020 | 0,0027 | 0,0034 |
| 8 | 0,0000 | 0,0000 | 0,0001 | 0,0001 | 0,0001 | 0,0002 | 0,0003 | 0,0005 | 0,0006 | 0,0009 |
| 9 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0001 | 0,0001 | 0,0001 | 0,0002 |

| k | λ 2,1 | 2,2 | 2,3 | 2,4 | 2,5 | 2,6 | 2,7 | 2,8 | 2,9 | 3,0 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0,1225 | 0,1108 | 0,1003 | 0,0907 | 0,0821 | 0,0743 | 0,0672 | 0,0608 | 0,0550 | 0,0498 |
| 1 | 0,2572 | 0,2438 | 0,2306 | 0,2177 | 0,2052 | 0,1931 | 0,1815 | 0,1703 | 0,1596 | 0,1494 |
| 2 | 0,2700 | 0,2681 | 0,2652 | 0,2613 | 0,2565 | 0,2510 | 0,2450 | 0,2384 | 0,2314 | 0,2240 |
| 3 | 0,1890 | 0,1966 | 0,2033 | 0,2090 | 0,2138 | 0,2176 | 0,2205 | 0,2225 | 0,2237 | 0,2240 |
| 4 | 0,0992 | 0,1082 | 0,1169 | 0,1254 | 0,1336 | 0,1414 | 0,1488 | 0,1557 | 0,1622 | 0,1680 |
| 5 | 0,0417 | 0,0476 | 0,0538 | 0,0602 | 0,0668 | 0,0735 | 0,0804 | 0,0872 | 0,0940 | 0,1008 |
| 6 | 0,0146 | 0,0174 | 0,0206 | 0,0241 | 0,0278 | 0,0319 | 0,0362 | 0,0407 | 0,0455 | 0,0504 |
| 7 | 0,0044 | 0,0055 | 0,0068 | 0,0083 | 0,0099 | 0,0118 | 0,0139 | 0,0163 | 0,0188 | 0,0216 |
| 8 | 0,0011 | 0,0015 | 0,0019 | 0,0025 | 0,0031 | 0,0038 | 0,0047 | 0,0057 | 0,0068 | 0,0081 |
| 9 | 0,0003 | 0,0004 | 0,0005 | 0,0007 | 0,0009 | 0,0011 | 0,0014 | 0,0018 | 0,0022 | 0,0027 |
| 10 | 0,0001 | 0,0001 | 0,0001 | 0,0002 | 0,0002 | 0,0003 | 0,0004 | 0,0005 | 0,0006 | 0,0008 |
| 11 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0001 | 0,0001 | 0,0001 | 0,0002 | 0,0002 |
| 12 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0001 |

| k | λ 3,1 | 3,2 | 3,3 | 3,4 | 3,5 | 3,6 | 3,7 | 3,8 | 3,9 | 4,0 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0,0450 | 0,0408 | 0,0369 | 0,0334 | 0,0302 | 0,0273 | 0,0247 | 0,0224 | 0,0202 | 0,0183 |
| 1 | 0,1397 | 0,1304 | 0,1217 | 0,1135 | 0,1057 | 0,0984 | 0,0915 | 0,0850 | 0,0789 | 0,0733 |
| 2 | 0,2165 | 0,2087 | 0,2008 | 0,1929 | 0,1850 | 0,1771 | 0,1692 | 0,1615 | 0,1539 | 0,1465 |
| 3 | 0,2237 | 0,2226 | 0,2209 | 0,2186 | 0,2158 | 0,2125 | 0,2087 | 0,2046 | 0,2001 | 0,1954 |
| 4 | 0,1734 | 0,1781 | 0,1823 | 0,1858 | 0,1888 | 0,1912 | 0,1931 | 0,1944 | 0,1951 | 0,1954 |
| 5 | 0,1075 | 0,1140 | 0,1203 | 0,1264 | 0,1322 | 0,1377 | 0,1429 | 0,1477 | 0,1522 | 0,1563 |
| 6 | 0,0555 | 0,0608 | 0,0662 | 0,0716 | 0,0771 | 0,0826 | 0,0881 | 0,0936 | 0,0989 | 0,1042 |
| 7 | 0,0246 | 0,0278 | 0,0312 | 0,0348 | 0,0385 | 0,0425 | 0,0466 | 0,0508 | 0,0551 | 0,0595 |
| 8 | 0,0095 | 0,0111 | 0,0129 | 0,0148 | 0,0169 | 0,0191 | 0,0215 | 0,0241 | 0,0269 | 0,0298 |
| 9 | 0,0033 | 0,0040 | 0,0047 | 0,0056 | 0,0066 | 0,0076 | 0,0089 | 0,0102 | 0,0116 | 0,0132 |
| 10 | 0,0010 | 0,0013 | 0,0016 | 0,0019 | 0,0023 | 0,0028 | 0,0033 | 0,0039 | 0,0045 | 0,0053 |
| 11 | 0,0003 | 0,0004 | 0,0005 | 0,0006 | 0,0007 | 0,0009 | 0,0011 | 0,0013 | 0,0016 | 0,0019 |
| 12 | 0,0001 | 0,0001 | 0,0001 | 0,0002 | 0,0002 | 0,0003 | 0,0003 | 0,0004 | 0,0005 | 0,0006 |
| 13 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0001 | 0,0001 | 0,0001 | 0,0001 | 0,0002 | 0,0002 |
| 14 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0001 |

Table 3 : Fonction de répartition de la loi normale

$$P(0 < T < t) = G(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt$$




La table donne aussi la valeur de $t_\alpha$ telle que

$P(0 < T < t_\alpha) = \frac{1 - \alpha}{2}$

où $1 - \alpha$ est le seuil de confiance

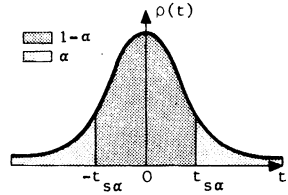
| t | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0,0 | 0,0000 | 0,0040 | 0,0080 | 0,0120 | 0,0160 | 0,0199 | 0,0239 | 0,0279 | 0,0319 | 0,0359 |
| 0,1 | 0,0398 | 0,0438 | 0,0478 | 0,0517 | 0,0557 | 0,0596 | 0,0636 | 0,0675 | 0,0714 | 0,0753 |
| 0,2 | 0,0793 | 0,0832 | 0,0871 | 0,0910 | 0,0948 | 0,0987 | 0,1026 | 0,1064 | 0,1103 | 0,1141 |
| 0,3 | 0,1179 | 0,1217 | 0,1255 | 0,1293 | 0,1331 | 0,1368 | 0,1406 | 0,1443 | 0,1480 | 0,1517 |
| 0,4 | 0,1554 | 0,1591 | 0,1628 | 0,1664 | 0,1700 | 0,1736 | 0,1772 | 0,1808 | 0,1844 | 0,1879 |
| 0,5 | 0,1915 | 0,1950 | 0,1985 | 0,2019 | 0,2054 | 0,2088 | 0,2123 | 0,2157 | 0,2190 | 0,2224 |
| 0,6 | 0,2257 | 0,2291 | 0,2324 | 0,2357 | 0,2389 | 0,2422 | 0,2454 | 0,2486 | 0,2517 | 0,2549 |
| 0,7 | 0,2580 | 0,2611 | 0,2642 | 0,2673 | 0,2704 | 0,2734 | 0,2764 | 0,2794 | 0,2823 | 0,2852 |
| 0,8 | 0,2881 | 0,2910 | 0,2939 | 0,2967 | 0,2995 | 0,3023 | 0,3051 | 0,3078 | 0,3106 | 0,3133 |
| 0,9 | 0,3159 | 0,3186 | 0,3212 | 0,3238 | 0,3264 | 0,3289 | 0,3315 | 0,3340 | 0,3365 | 0,3389 |
| 1,0 | 0,3413 | 0,3438 | 0,3461 | 0,3485 | 0,3508 | 0,3531 | 0,3554 | 0,3577 | 0,3599 | 0,3621 |
| 1,1 | 0,3643 | 0,3665 | 0,3686 | 0,3708 | 0,3729 | 0,3749 | 0,3770 | 0,3790 | 0,3810 | 0,3830 |
| 1,2 | 0,3849 | 0,3869 | 0,3888 | 0,3907 | 0,3925 | 0,3944 | 0,3962 | 0,3980 | 0,3997 | 0,4015 |
| 1,3 | 0,4032 | 0,4049 | 0,4066 | 0,4082 | 0,4099 | 0,4115 | 0,4131 | 0,4147 | 0,4162 | 0,4177 |
| 1,4 | 0,4192 | 0,4207 | 0,4222 | 0,4236 | 0,4251 | 0,4265 | 0,4279 | 0,4292 | 0,4306 | 0,4319 |
| 1,5 | 0,4332 | 0,4345 | 0,4357 | 0,4370 | 0,4382 | 0,4394 | 0,4406 | 0,4418 | 0,4429 | 0,4441 |
| 1,6 | 0,4452 | 0,4463 | 0,4474 | 0,4484 | 0,4495 | 0,4505 | 0,4515 | 0,4525 | 0,4535 | 0,4545 |
| 1,7 | 0,4554 | 0,4564 | 0,4573 | 0,4582 | 0,4591 | 0,4599 | 0,4608 | 0,4616 | 0,4625 | 0,4633 |
| 1,8 | 0,4641 | 0,4649 | 0,4656 | 0,4664 | 0,4671 | 0,4678 | 0,4686 | 0,4693 | 0,4699 | 0,4706 |
| 1,9 | 0,4713 | 0,4719 | 0,4726 | 0,4732 | 0,4738 | 0,4744 | 0,4750 | 0,4756 | 0,4761 | 0,4767 |
| 2,0 | 0,4772 | 0,4778 | 0,4783 | 0,4788 | 0,4793 | 0,4798 | 0,4803 | 0,4808 | 0,4812 | 0,4817 |
| 2,1 | 0,4821 | 0,4826 | 0,4830 | 0,4834 | 0,4838 | 0,4842 | 0,4846 | 0,4850 | 0,4854 | 0,4857 |
| 2,2 | 0,4861 | 0,4864 | 0,4868 | 0,4871 | 0,4875 | 0,4878 | 0,4881 | 0,4884 | 0,4887 | 0,4890 |
| 2,3 | 0,4893 | 0,4896 | 0,4998 | 0,4901 | 0,4904 | 0,4906 | 0,4909 | 0,4911 | 0,4913 | 0,4916 |
| 2,4 | 0,4918 | 0,4920 | 0,4922 | 0,4925 | 0,4927 | 0,4929 | 0,4931 | 0,4932 | 0,4934 | 0,4936 |
| 2,5 | 0,4938 | 0,4940 | 0,4941 | 0,4943 | 0,4945 | 0,4946 | 0,4948 | 0,4949 | 0,4951 | 0,4952 |
| 2,6 | 0,4953 | 0,4955 | 0,4956 | 0,4957 | 0,4959 | 0,4960 | 0,4961 | 0,4962 | 0,4963 | 0,4964 |
| 2,7 | 0,4965 | 0,4966 | 0,4967 | 0,4968 | 0,4969 | 0,4970 | 0,4971 | 0,4972 | 0,4973 | 0,4974 |
| 2,8 | 0,4974 | 0,4975 | 0,4976 | 0,4977 | 0,4977 | 0,4978 | 0,4979 | 0,4979 | 0,4980 | 0,4981 |
| 2,9 | 0,4981 | 0,4982 | 0,4983 | 0,4983 | 0,4984 | 0,4984 | 0,4985 | 0,4985 | 0,4986 | 0,4986 |
| 3,0 | 0,4987 | 0,4987 | 0,4987 | 0,4988 | 0,4988 | 0,4989 | 0,4989 | 0,4989 | 0,4990 | 0,4990 |
| 3,1 | 0,4990 | 0,4991 | 0,4991 | 0,4991 | 0,4992 | 0,4992 | 0,4992 | 0,4992 | 0,4993 | 0,4993 |
| 3,2 | 0,4993 | 0,4993 | 0,4994 | 0,4994 | 0,4994 | 0,4994 | 0,4994 | 0,4995 | 0,4995 | 0,4995 |
| 3,3 | 0,4995 | 0,4995 | 0,4995 | 0,4996 | 0,4996 | 0,4996 | 0,4996 | 0,4996 | 0,4996 | 0,4997 |
| 3,4 | 0,4997 | 0,4997 | 0,4997 | 0,4997 | 0,4997 | 0,4997 | 0,4997 | 0,4997 | 0,4997 | 0,4998 |
| 3,5 | 0,4998 | 0,4998 | 0,4998 | 0,4998 | 0,4998 | 0,4998 | 0,4998 | 0,4998 | 0,4998 | 0,4998 |
| 3,6 | 0,4998 | 0,4998 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 |
| 3,7 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 |
| 3,8 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 | 0,4999 |
| 3,9 | 0,5000 | 0,5000 | 0,5000 | 0,5000 | 0,5000 | 0,5000 | 0,5000 | 0,5000 | 0,5000 | 0,5000 |
| 4,0 | 0,5000 | 0,5000 | 0,5000 | 0,5000 | 0,5000 | 0,5000 | 0,5000 | 0,5000 | 0,5000 | 0,5000 |

Table 4 : Distribution de Student

Table donnant la valeur de $t_{s\alpha}$ telle que

$P(-t_{s\alpha} \leq t_s \leq t_{s\alpha}) = 1 - \alpha$

où $\alpha$ est le risque

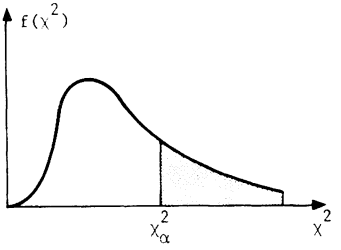



| ν \ α | 0,90 | 0,50 | 0,40 | 0,30 | 0,20 | 0,10 | 0,05 | 0,01 | 0,001 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0,158 | 1,000 | 1,376 | 1,963 | 3,078 | 6,314 | 12,706 | 63,657 | 636,619 |
| 2 | 0,142 | 0,816 | 1,061 | 1,386 | 1,886 | 2,920 | 4,303 | 9,925 | 31,958 |
| 3 | 0,137 | 0,765 | 0,978 | 1,250 | 1,638 | 2,353 | 3,182 | 5,841 | 12,929 |
| 4 | 0,134 | 0,741 | 0,941 | 1,190 | 1,533 | 2,132 | 2,776 | 4,604 | 8,610 |
| 5 | 0,132 | 0,727 | 0,920 | 1,156 | 1,476 | 2,015 | 2,571 | 4,032 | 6,869 |
| 6 | 0,131 | 0,718 | 0,906 | 1,134 | 1,440 | 1,943 | 2,447 | 3,707 | 5,949 |
| 7 | 0,130 | 0,711 | 0,896 | 1,119 | 1,415 | 1,895 | 2,365 | 3,499 | 5,408 |
| 8 | 0,130 | 0,706 | 0,889 | 1,108 | 1,397 | 1,860 | 2,306 | 3,355 | 5,041 |
| 9 | 0,129 | 0,703 | 0,883 | 1,100 | 1,383 | 1,833 | 2,262 | 3,250 | 4,781 |
| 10 | 0,129 | 0,700 | 0,879 | 1,083 | 1,372 | 1,812 | 2,228 | 3,169 | 4,587 |
| 11 | 0,129 | 0,697 | 0,876 | 1,088 | 1,363 | 1,796 | 2,201 | 3,106 | 4,437 |
| 12 | 0,128 | 0,695 | 0,873 | 1,083 | 1,356 | 1,782 | 2,179 | 3,055 | 4,318 |
| 13 | 0,128 | 0,694 | 0,870 | 1,079 | 1,350 | 1,771 | 2,160 | 3,012 | 4,221 |
| 14 | 0,128 | 0,692 | 0,868 | 1,076 | 1,345 | 1,761 | 2,145 | 2,977 | 4,140 |
| 15 | 0,128 | 0,691 | 0,866 | 1,074 | 1,341 | 1,753 | 2,131 | 2,947 | 4,073 |
| 16 | 0,128 | 0,690 | 0,865 | 1,071 | 1,337 | 1,746 | 2,120 | 2,921 | 4,015 |
| 17 | 0,128 | 0,689 | 0,863 | 1,069 | 1,333 | 1,740 | 2,110 | 2,898 | 3,965 |
| 18 | 0,127 | 0,688 | 0,862 | 1,067 | 1,330 | 1,734 | 2,101 | 2,878 | 3,922 |
| 19 | 0,127 | 0,688 | 0,861 | 1,066 | 1,328 | 1,729 | 2,093 | 2,861 | 3,883 |
| 20 | 0,127 | 0,687 | 0,860 | 1,064 | 1,325 | 1,725 | 2,086 | 2,845 | 3,850 |
| 21 | 0,127 | 0,686 | 0,859 | 1,063 | 1,323 | 1,721 | 2,080 | 2,831 | 3,819 |
| 22 | 0,127 | 0,686 | 0,858 | 1,061 | 1,321 | 1,717 | 2,074 | 2,819 | 3,792 |
| 23 | 0,127 | 0,685 | 0,858 | 1,060 | 1,319 | 1,714 | 2,069 | 2,807 | 3,767 |
| 24 | 0,127 | 0,685 | 0,857 | 1,059 | 1,318 | 1,711 | 2,064 | 2,797 | 3,745 |
| 25 | 0,127 | 0,684 | 0,856 | 1,058 | 1,316 | 1,708 | 2,060 | 2,787 | 3,725 |
| 26 | 0,127 | 0,684 | 0,856 | 1,058 | 1,315 | 1,706 | 2,056 | 2,779 | 3,707 |
| 27 | 0,127 | 0,684 | 0,855 | 1,057 | 1,314 | 1,703 | 2,052 | 2,771 | 3,690 |
| 28 | 0,127 | 0,683 | 0,855 | 1,056 | 1,313 | 1,701 | 2,048 | 2,763 | 3,674 |
| 29 | 0,127 | 0,683 | 0,854 | 1,055 | 1,311 | 1,699 | 2,045 | 2,756 | 3,659 |
| 30 | 0,127 | 0,683 | 0,854 | 1,055 | 1,310 | 1,697 | 2,042 | 2,750 | 3,646 |

Table 5 : Distribution du $\chi^2$

Table donnant la valeur de $\chi^2$ ayant la probabilité $\alpha$ d'être dépassée

$P(\chi^2 \geq \chi^2_\alpha) = \alpha$




| ν \ α | 0,990 | 0,975 | 0,950 | 0,900 | 0,100 | 0,050 | 0,025 | 0,010 | 0,001 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0,0002 | 0,0010 | 0,0039 | 0,0158 | 2,71 | 3,84 | 5,02 | 6,63 | 10,83 |
| 2 | 0,02 | 0,05 | 0,10 | 0,21 | 4,61 | 5,99 | 7,38 | 9,21 | 13,82 |
| 3 | 0,12 | 0,22 | 0,35 | 0,58 | 6,25 | 7,81 | 9,35 | 11,34 | 16,27 |
| 4 | 0,30 | 0,48 | 0,71 | 1,06 | 7,78 | 9,49 | 11,14 | 13,28 | 18,47 |
| 5 | 0,55 | 0,83 | 1,15 | 1,61 | 9,24 | 11,07 | 12,83 | 15,09 | 20,52 |
| 6 | 0,87 | 1,24 | 1,64 | 2,20 | 10,64 | 12,59 | 14,45 | 16,81 | 22,46 |
| 7 | 1,24 | 1,69 | 2,17 | 2,83 | 12,02 | 14,07 | 16,01 | 18,47 | 24,32 |
| 8 | 1,65 | 2,18 | 2,73 | 3,49 | 13,36 | 15,51 | 17,53 | 20,09 | 26,13 |
| 9 | 2,09 | 2,70 | 3,33 | 4,17 | 14,68 | 16,92 | 19,02 | 21,67 | 27,88 |
| 10 | 2,56 | 3,25 | 3,94 | 4,87 | 15,99 | 18,31 | 20,48 | 23,21 | 29,59 |
| 11 | 3,05 | 3,82 | 4,57 | 5,58 | 17,27 | 19,67 | 21,92 | 24,72 | 31,26 |
| 12 | 3,57 | 4,40 | 5,23 | 6,30 | 18,55 | 21,03 | 23,34 | 26,22 | 32,91 |
| 13 | 4,11 | 5,01 | 5,89 | 7,04 | 19,81 | 22,36 | 24,74 | 27,69 | 34,53 |
| 14 | 4,66 | 5,63 | 6,57 | 7,79 | 21,06 | 23,68 | 26,12 | 29,14 | 36,12 |
| 15 | 5,23 | 6,26 | 7,26 | 8,55 | 22,31 | 25,00 | 27,49 | 30,58 | 37,70 |
| 16 | 5,81 | 6,91 | 7,96 | 9,31 | 23,54 | 26,30 | 28,84 | 32,00 | 39,25 |
| 17 | 6,41 | 7,56 | 8,67 | 10,08 | 24,77 | 27,59 | 30,19 | 33,41 | 40,79 |
| 18 | 7,01 | 8,23 | 9,39 | 10,86 | 25,99 | 28,87 | 31,53 | 34,80 | 42,31 |
| 19 | 7,63 | 8,91 | 10,12 | 11,65 | 27,20 | 30,14 | 32,85 | 36,19 | 43,82 |
| 20 | 8,26 | 9,59 | 10,85 | 12,44 | 28,41 | 31,41 | 34,17 | 37,57 | 45,32 |
| 21 | 8,90 | 10,28 | 11,59 | 13,24 | 29,61 | 32,67 | 35,48 | 38,93 | 46,80 |
| 22 | 9,54 | 10,98 | 12,34 | 14,04 | 30,81 | 33,92 | 36,78 | 40,29 | 48,27 |
| 23 | 10,20 | 11,69 | 13,09 | 14,85 | 32,01 | 35,17 | 38,08 | 41,64 | 49,73 |
| 24 | 10,86 | 12,40 | 13,85 | 15,66 | 33,20 | 36,41 | 39,37 | 42,98 | 51,18 |
| 25 | 11,52 | 13,12 | 14,61 | 16,47 | 34,38 | 37,65 | 40,65 | 44,31 | 52,62 |
| 26 | 12,20 | 12,84 | 15,38 | 17,29 | 35,56 | 38,88 | 41,92 | 45,64 | 54,05 |
| 27 | 12,88 | 13,57 | 16,15 | 18,11 | 36,74 | 40,11 | 43,19 | 46,96 | 55,48 |
| 28 | 13,57 | 15,31 | 16,93 | 18,94 | 37,92 | 41,34 | 44,46 | 48,28 | 56,89 |
| 29 | 14,26 | 16,05 | 17,71 | 19,77 | 39,09 | 42,56 | 45,72 | 49,59 | 58,30 |
| 30 | 14,95 | 16,79 | 18,49 | 20,60 | 40,26 | 43,77 | 46,98 | 50,89 | 59,70 |

Table 6 : Coefficient de corrélation

Table donnant la valeur de $r_\alpha$ telle que

$P(-r_\alpha < r < r_\alpha) = 1 - \alpha$

| $\nu$ \ $\alpha$ | 0,10 | 0,05 | 0,02 | 0,01 |
|---|---|---|---|---|
| 1 | 0,9877 | 0,9969 | 0,9995 | 0,9999 |
| 2 | 0,9000 | 0,9500 | 0,9800 | 0,9900 |
| 3 | 0,8054 | 0,8783 | 0,9343 | 0,9587 |
| 4 | 0,7293 | 0,8114 | 0,8822 | 0,9172 |
| 5 | 0,6694 | 0,7545 | 0,8329 | 0,8745 |
| 6 | 0,6215 | 0,7067 | 0,7887 | 0,8343 |
| 7 | 0,5822 | 0,6664 | 0,7498 | 0,7977 |
| 8 | 0,5494 | 0,6319 | 0,7155 | 0,7646 |
| 9 | 0,5214 | 0,6021 | 0,6851 | 0,7348 |
| 10 | 0,4973 | 0,5760 | 0,6581 | 0,7079 |
| 11 | 0,4762 | 0,5529 | 0,6339 | 0,6835 |
| 12 | 0,4575 | 0,5324 | 0,6120 | 0,6614 |
| 13 | 0,4409 | 0,5139 | 0,5923 | 0,6411 |
| 14 | 0,4259 | 0,4973 | 0,5742 | 0,6226 |
| 15 | 0,4124 | 0,4821 | 0,5577 | 0,6055 |
| 16 | 0,4000 | 0,4683 | 0,5425 | 0,5897 |
| 17 | 0,3887 | 0,4555 | 0,5285 | 0,5751 |
| 18 | 0,3783 | 0,4438 | 0,5155 | 0,5614 |
| 19 | 0,3687 | 0,4329 | 0,5034 | 0,5487 |
| 20 | 0,3598 | 0,4227 | 0,4921 | 0,5368 |
| 25 | 0,3233 | 0,3809 | 0,4451 | 0,4869 |
| 30 | 0,2960 | 0,3494 | 0,4093 | 0,4487 |
| 35 | 0,2746 | 0,3246 | 0,3810 | 0,4182 |
| 40 | 0,2573 | 0,3044 | 0,3578 | 0,3932 |
| 45 | 0,2428 | 0,2875 | 0,3384 | 0,3721 |
| 50 | 0,2306 | 0,2732 | 0,3218 | 0,3541 |
| 60 | 0,2108 | 0,2500 | 0,2948 | 0,3248 |
| 70 | 0,1954 | 0,2319 | 0,2737 | 0,3017 |
| 80 | 0,1829 | 0,2172 | 0,2565 | 0,2830 |
| 90 | 0,1726 | 0,2050 | 0,2422 | 0,2673 |
| 100 | 0,1638 | 0,1946 | 0,2301 | 0,2540 |

Table 7 : Valeurs de la variable Z de Fisher correspondant au coefficient de corrélation r, $Z = \frac{1}{2} \text{Log}_e(\frac{1 + r}{1 - r})$.

| r | 0,00 | 0,01 | 0,02 | 0,03 | 0,04 |
|---|---|---|---|---|---|
| 0,0 | 0,00000 | 0,01000 | 0,02000 | 0,03001 | 0,04002 |
| 0,1 | 0,10034 | 0,11045 | 0,12058 | 0,13074 | 0,14093 |
| 0,2 | 0,20273 | 0,21317 | 0,22366 | 0,23419 | 0,24477 |
| 0,3 | 0,30952 | 0,32055 | 0,33165 | 0,34283 | 0,35409 |
| 0,4 | 0,42365 | 0,43561 | 0,44769 | 0,45990 | 0,47223 |
| 0,5 | 0,54931 | 0,56273 | 0,57634 | 0,59014 | 0,60415 |
| 0,6 | 0,69315 | 0,70892 | 0,72500 | 0,74142 | 0,75817 |
| 0,7 | 0,86730 | 0,88718 | 0,90764 | 0,92873 | 0,95048 |
| 0,8 | 1,09861 | 1,12703 | 1,15682 | 1,18813 | 1,22117 |
| 0,9 | 1,47222 | 1,52752 | 1,58902 | 1,65839 | 1,73805 |

| r | 0,05 | 0,06 | 0,07 | 0,08 | 0,09 |
|---|---|---|---|---|---|
| 0,0 | 0,05004 | 0,06007 | 0,07012 | 0,08017 | 0,09024 |
| 0,1 | 0,15114 | 0,16139 | 0,17167 | 0,18198 | 0,19234 |
| 0,2 | 0,25541 | 0,26611 | 0,27686 | 0,28768 | 0,29857 |
| 0,3 | 0,36544 | 0,37689 | 0,38842 | 0,40006 | 0,41180 |
| 0,4 | 0,48470 | 0,49731 | 0,51007 | 0,52298 | 0,53606 |
| 0,5 | 0,61838 | 0,63283 | 0,64752 | 0,66246 | 0,67767 |
| 0,6 | 0,77530 | 0,79281 | 0,81074 | 0,82911 | 0,84795 |
| 0,7 | 0,97295 | 0,99621 | 1,02033 | 1,04537 | 1,07143 |
| 0,8 | 1,25615 | 1,29334 | 1,33308 | 1,37577 | 1,42192 |
| 0,9 | 1,83178 | 1,94591 | 2,09229 | 2,29756 | 2,64665 |

## INDEX DES SYMBOLES

R et N désignent respectivement l'ensemble des nombres réels et l'ensemble des entiers naturels

INDEX

Imprimé en France, Boisseau, Toulouse
Dépôt légal : quatrième trimestre 1978
Numéro d'édition 5889
Hermann, éditeurs des sciences et des arts