

# **Probabilités et Statistiques**

## **MATH-F-315**

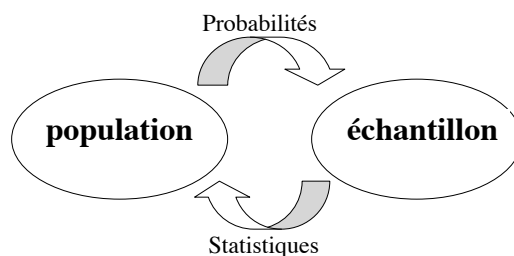
**Simone GUTT**

**2012**

Dans la vie, nous sommes continuellement confrontés à des collections de faits ou données. Les statistiques forment une branche scientifique qui fournit des méthodes pour organiser et résumer les données (statistiques descriptives) et pour utiliser l'information des données afin de tirer des conclusions. Fréquemment, on veut des conclusions à propos d'une population entière constituée de tous les individus ou objets d'un type particulier (par exemple, la population peut être constituée de toutes les tondeuses à gazon fabriquées par la compagnie X en 2004, ou cela peut être la collection de tous les individus ayant reçu un vaccin contre la grippe d'un certain type, ou cela peut être l'ensemble des universités européennes . . . ) Les données dont on dispose consistent fréquemment en un sous-ensemble de la population; on appelle un tel sous-ensemble un échantillon.

L'inférence statistique fournit des méthodes pour tirer des conclusions concernant une population entière à partir d'un échantillon. L'inférence statistique s'est développée à partir du début du 20<sup>eme</sup> siècle.

La théorie des probabilités est une théorie mathématique ayant pour but de décrire et d'interpréter les phénomènes dans lesquels le hasard intervient. Cette théorie permet de comprendre comment les procédures inférentielles sont développées et utilisées et quelles en sont les limites. Dans un problème de probabilité, on suppose connues les propriétés d'une population et on répond à des questions concernant un échantillon prélevé dans cette population. Dans un problème de statistique, ce sont les caractéristiques d'un échantillon qui sont connues et cette information sert à tirer des conclusions à propos de la population.



### Exemple(Science 167, p.277-279)

Le rapport de la masse de la terre à la masse de la lune a été mesuré au cours de différents vols dans l'espace; les résultats mesurés sont:

81,3001 (Mariner 2)

81,3015 (Mariner 4)

81,3006 (Mariner 5)

81,3011 (Mariner 6)

81,2997 (Mariner 7)

81,3005 (Pioneer 6)

81,3021 (Pioneer 7)

Ces nombres diffèrent à cause des erreurs de mesure. En probabilités, on peut supposer que la distribution de toutes les mesures possibles a une forme de cloche centrée en 81,3035 et on peut poser la question:

**“Quelle est la probabilité que sept mesures effectuées soient toutes inférieures à 81,3035?”.**

En statistiques, connaissant les sept résultats plus haut, on peut poser la question:

**“Quel est notre degré d'assurance pour dire que le rapport réel des masses est compris entre 81,2998 et 81,3018?”.**

Dans cet exemple, la population est conceptuelle et n'existe pas vraiment: elle est constituée de toutes les mesures qui pourraient être faites dans des conditions expérimentales semblables.

# Contenu

|          |  |           |
|----------|--|-----------|
| 0.1      | Histogrammes et diagrammes . . . . .                                   | 6         |
| 0.2      | Mesures du “centre” d’un ensemble de données . . . . .                 | 11        |
| 0.3      | Mesures de la dispersion d’un ensemble de données . . . . .            | 12        |
| <b>I</b> | <b>PROBABILITES</b>  | <b>17</b> |
| <b>1</b> | <b>Définitions de base</b>   | <b>19</b> |
| 1.1      | Définitions . . . . .  | 19        |
| 1.2      | Exemples . . . . .   | 26        |
| 1.2.1    | Récréation: Le paradoxe de Simpson . . . . .                           | 28        |
| <b>2</b> | <b>Probabilités conditionnelles</b>                                    | <b>31</b> |
| 2.1      | Définition . . . . .   | 31        |
| 2.2      | Théorèmes . . . . .  | 33        |
| <b>3</b> | <b>Variables aléatoires</b>  | <b>36</b> |
| 3.1      | Variable aléatoire discrète . . . . .                                  | 37        |
| 3.1.1    | La variable binomiale . . . . .  | 41        |
| 3.1.2    | La loi de Poisson . . . . .  | 43        |
| 3.1.3    | Fonction de répartition . . . . .                                      | 44        |
| 3.1.4    | Espérance mathématique-Variance . . . . .                              | 46        |
| 3.1.5    | Fonction d’une variable aléatoire . . . . .                            | 49        |
| 3.2      | Variables aléatoires (absolument) continues . . . . .                  | 50        |
| 3.2.1    | La loi normale standard . . . . .                                      | 56        |
| 3.2.2    | La loi normale . . . . .   | 57        |
| <b>4</b> | <b>Distribution jointe et corrélation de deux variables aléatoires</b> | <b>65</b> |
| 4.1      | Espérance d’une fonction de deux variables aléatoires . . . . .        | 70        |

|           |   |            |
|-----------|---|------------|
| 4.2       | Indépendance de deux variables aléatoires . . . . .   | 70         |
| 4.3       | Corrélation de deux variables aléatoires . . . . .  | 71         |
| <b>5</b>  | <b>Quelques théorèmes importants</b>  | <b>73</b>  |
| <b>II</b> | <b>STATISTIQUE</b>  | <b>76</b>  |
| <b>6</b>  | <b>Echantillons-Estimeurs</b>   | <b>80</b>  |
| 6.1       | Echantillons aléatoires simples . . . . .   | 80         |
| 6.2       | Estimateurs . . . . .   | 82         |
| 6.2.1     | Estimation d'une probabilité ou d'une proportion . . .  | 82         |
| 6.2.2     | estimateurs non biaisés, estimateurs de la moyenne et<br>de la variance . . . . .                                   | 85         |
| 6.2.3     | méthode du maximum de vraisemblance . . . . .   | 86         |
| <b>7</b>  | <b>Intervalles de confiance</b>   | <b>88</b>  |
| 7.1       | intervalle de confiance pour une proportion ou une probabilité  | 89         |
| 7.2       | Moyenne d'une grandeur, grand échantillon . . . . .   | 91         |
| 7.3       | Moyenne d'une grandeur distribuée suivant une normale, avec<br>un petit échantillon (variable de Student) . . . . . | 93         |
| <b>8</b>  | <b>Tests d'hypothèses</b>   | <b>96</b>  |
| 8.1       | Tests sur une probabilité ou une proportion . . . . .   | 101        |
| 8.1.1     | Test 1 au niveau $\alpha$ ; $H_0 : p = p_0$ ( $H_1 : p \neq p_0$ ) . . . .  | 102        |
| 8.1.2     | Test 2 au niveau $\alpha$ ; $H_0 : p \geq p_0$ ( $H_1 : p < p_0$ ) . . . .  | 103        |
| 8.1.3     | Test 3 au niveau $\alpha$ ; $H_0 : p \leq p_0$ ( $H_1 : p > p_0$ ) . . . .  | 104        |
| 8.2       | Tests sur la moyenne d'une grandeur . . . . .   | 105        |
| 8.2.1     | exemple de test sur la moyenne d'une population normale   | 105        |
| 8.2.2     | Tests sur la moyenne d'une grandeur . . . . .   | 107        |
| <b>9</b>  | <b>Comparaisons de deux populations</b>   | <b>111</b> |
| 9.1       | Comparaison de deux proportions . . . . .   | 111        |
| 9.1.1     | Le modèle lors d'une comparaison de deux proportions  | 112        |
| 9.1.2     | Intervalle de confiance pour $p_1 - p_2$ . . . . .  | 113        |
| 9.1.3     | Tests d'hypothèse pour $p_1 - p_2$ . . . . .  | 113        |
| 9.2       | Comparaison de deux moyennes . . . . .  | 115        |
| 9.2.1     | Intervalle de confiance au niveau $(1 - \alpha)$ pour $\mu_1 - \mu_2$   | 116        |
| 9.2.2     | Test au niveau $\alpha$ pour $\mu_1 - \mu_2$ . . . . .  | 117        |

|  |            |
|--|------------|
| <b>10 Tests chi-carré (<math>\chi^2</math>)</b>                          | <b>120</b> |
| 10.1 Variables normales, chi-carré et de Student . . . . .               | 120        |
| 10.2 Les tests $\chi^2$ d'ajustement . . . . .                           | 122        |
| 10.2.1 Tests d'ajustement avec paramètres . . . . .                      | 124        |
| 10.3 Les tests $\chi^2$ d'homogénéité . . . . .                          | 129        |
| 10.4 Les tests $\chi^2$ d'indépendance . . . . .                         | 132        |
| <b>11 Le modèle de régression linéaire simple</b>                        | <b>136</b> |
| 11.1 Estimation des paramètres . . . . .                                 | 138        |
| 11.1.1 Méthode des moindres carrés (Gauss 1777–1855) . . .               | 140        |
| 11.1.2 Estimateurs pour $\beta_0$ , $\beta_1$ et $\sigma^2$ . . . . .    | 141        |
| 11.1.3 Distribution des estimateurs dans le cas gaussien . . .           | 142        |
| 11.2 Intervalle de confiance et Tests d'hypothèse pour $\beta_1$ dans le |            |
| cas gaussien . . . . .   | 142        |
| 11.2.1 Intervalle de confiance pour $\beta_1$ . . . . .                  | 142        |
| 11.2.2 Test d'hypothèse pour $\beta_1$ . . . . .                         | 143        |
| 11.3 Corrélation . . . . .   | 143        |
| <b>12 ANOVA à un facteur</b>   | <b>146</b> |

# ANALYSE DES DONNEES

## 0.1 Histogrammes et diagrammes

Soit  $n$  le nombre d'observations dans un ensemble de données. On représente les observations par  $x_1, x_2, \dots, x_n$ . Dans de nombreux cas,  $x_i$  est la  $i^{\text{ème}}$  observation obtenue par l'expérimentateur.

Supposons que chaque  $x_i$  soit un nombre. Comment représenter visuellement les données ?

On divise la droite réelle, entre la plus petite et la plus grande valeur observée, en un certain nombre d'intervalles de longueur égale (un petit nombre d'intervalles si le nombre d'observations est petit, un plus grand nombre d'intervalles s'il y a beaucoup de données). On compte le nombre d'observations dont le résultat appartient à un intervalle : c'est la fréquence de l'intervalle. En dessinant au-dessus de chaque intervalle un rectangle dont la hauteur est proportionnelle au nombre d'observations à valeurs dans l'intervalle (=la fréquence), on obtient l'histogramme des données.

La fréquence relative correspondant à un intervalle est obtenue en divisant la fréquence par le nombre total d'observations ( $n$ ).

[Remarquons qu'ici nous avons regardé le cas où un nombre est associé à chaque élément de l'échantillon. Dans de nombreux exemples, si on a un échantillon d'individus ou d'objets, on associe à chacun d'eux plusieurs nombres  $(x_1, y_1, z_1, \dots)(x_2, y_2, z_2, \dots) \dots (x_n, y_n, z_n, \dots)$ .





On parle alors d' analyse multivariée.]

**Exemple 0.1** Les mesures de cadmium dans des coquilles S<sup>t</sup> Jacques prélevées

en Atlantique Nord ont donné les résultats suivants (cf Environmental Concentration and toxicology):

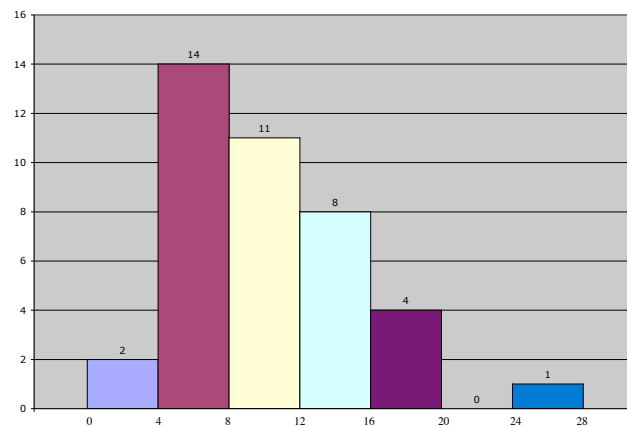
|      |      |      |      |      |      |      |      |      |      |     |     |
|------|------|------|------|------|------|------|------|------|------|-----|-----|
| 5,1  | 14,4 | 14,7 | 10,8 | 6,5  | 5,7  | 7,7  | 14,1 | 9,5  | 3,7  | 8,9 | 7,9 |
| 7,9  | 4,5  | 10,1 | 5,0  | 9,6  | 5,5  | 5,1  | 11,4 | 8,0  | 12,1 | 7,5 | 8,5 |
| 13,1 | 6,4  | 18,0 | 27,0 | 18,9 | 10,8 | 13,1 | 8,4  | 16,9 | 2,7  | 9,6 | 4,5 |
| 12,4 | 5,5  | 12,7 | 17,1 |      |      |      |      |      |      |     |     |

Définissons les intervalles  $0 \leq x < 4$      $4 \leq x < 8 \dots$

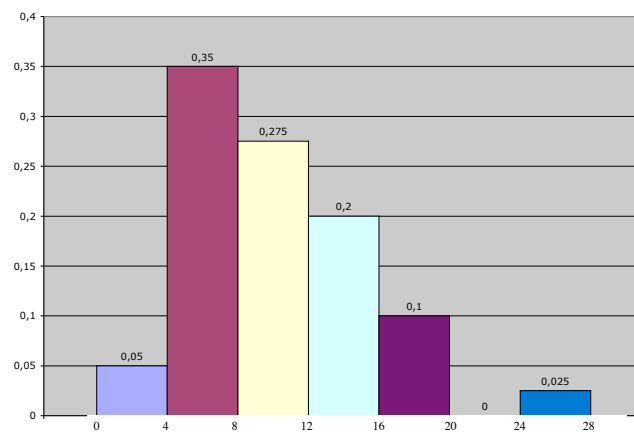
| <u>Tableau de fréquences</u>   | <u>Fréquence</u> | <u>Fréquence relative</u>  |
|--|------------------|----------------------------|
| $[0, 4)$   | 2                | $2/40 = 5/100 = 0,05$      |
| $[4, 8)$      | 14               | $14/40 = 35/100 = 0,35$    |
| $[8, 12)$     | 11               | $11/40 = 275/1000 = 0,275$ |
| $[12, 16)$  | 8                | $8/40 = 2/10 = 0,20$       |
| $[16, 20)$  | 4                | $4/40 = 1/10 = 0,10$       |
| $[20, 24)$   | 0                | $0/40 = 0$                 |
| $[24, 28)$   | 1                | $1/40 = 25/1000 = 0,025$   |
| Total:   | 40               | 1,000                      |



**Histogramme de la distribution des fréquences:**



**Histogramme de la distribution des fréquences relatives:**



(Ces deux histogrammes sont les mêmes sauf l'échelle sur l'axe vertical!)

Une manière simple et rapide d'écrire une liste de données a été introduite par le statisticien John Tukey; c'est la méthode "**Stem-and-leaf**"



Si l'observation est représentée par un nombre à au moins deux chiffres, on écrit une première colonne (la marge) avec les premiers chiffres, de façon à diviser l'ensemble des données en 5 à 20 paquets (ces premiers chiffres dans la marge constituent le "Stem") et on écrit à la droite de chacun de ces chiffres en marge, tous les chiffres supplémentaires d'une observation commençant par ces chiffres (les nouveaux chiffres constituent le "leaf"). On écrit donc, à droite d'une "stem" donnée, toutes les "leaves" lui correspondant.

**Exemple 0.2** Les taux d'octane de différentes essences pour moteur ont été mesurés (Technométries, vol 19) et donnent les résultats suivants:

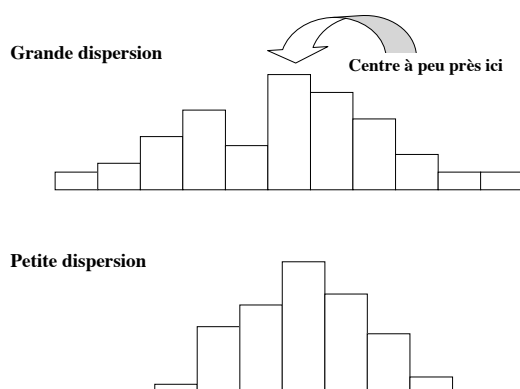
|      |      |      |      |      |      |      |       |      |      |      |      |
|------|------|------|------|------|------|------|-------|------|------|------|------|
| 88,5 | 87,7 | 83,4 | 86,7 | 87,5 | 91,5 | 88,6 | 100,3 | 95,6 | 93,3 | 94,7 | 91,1 |
| 91,0 | 94,2 | 87,8 | 89,9 | 88,3 | 87,6 | 86,7 | 88,2  | 90,8 | 88,3 | 98,8 | 94,2 |
| 92,7 | 84,3 | 93,2 | 91,0 | 90,3 | 93,4 | 88,5 | 90,1  | 89,2 | 88,3 | 85,3 | 87,9 |
| 88,6 | 90,9 | 89,? | 96,1 | 93,3 | 91,8 | 92,3 | 90,4  | 90,1 | 93,0 | 88,7 | 89,9 |
| 89,8 | 89,6 | 87,4 | 88,? | 88,9 | 91,2 | 89,3 | 94,4  | 92,7 | 91,8 | 91,6 | 90,4 |
| 91,1 | 92,6 | 89,8 | 90,6 | 91,  | 90,4 | 89,3 | 89,7  | 90,3 | 91,6 | 90,5 | 93,7 |
| 92,7 | 92,2 | 91,2 | 91,0 | 92,2 | 92,? | 90,0 | 90,7  |      |      |      |      |

Il y a ainsi 80 observations; la plus petite est 83,4, la plus grande 100,3. On choisit comme "stem" les valeurs 83 84 85 ..... 100.

On obtient alors la représentation.

|     |                           |
|-----|---------------------------|
| 83  | 4                         |
| 84  | 3                         |
| 85  | 3                         |
| 86  | 7 7                       |
| 87  | 7 5 8 6 9 4               |
| 88  | 5 6 3 2 3 5 3 6 7 4 9     |
| 89  | 9 2 0 9 8 6 3 8 3 7       |
| 90  | 8 3 1 9 4 1 4 6 4 3 5 0 7 |
| 91  | 5 1 0 0 8 2 8 6 1 1 6 2 0 |
| 92  | 7 3 7 6 7 2 2 2           |
| 93  | 3 2 4 3 0 7               |
| 94  | 7 2 2 4                   |
| 95  | 6                         |
| 96  | 1                         |
| 97  |                           |
| 98  | 8                         |
| 99  |                           |
| 100 | 3                         |

Tout ensemble de mesures a deux propriétés importantes : la valeur “centrale” et la dispersion.



## 0.2 Mesures du “centre” d’un ensemble de données

**Définition 0.1** La moyenne d’échantillon,  $\bar{x}$ , d’un ensemble de nombres  $x_1, \dots, x_n$  est donnée par

$$\bar{x} := \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i.$$

(Remarquons que  $\bar{x}$  représente la valeur moyenne des observations dans un échantillon; on peut penser calculer la moyenne de toutes les valeurs de la population; c’est appelé la moyenne de population et elle est notée  $\mu$ ).

**Définition 0.2** Etant donné un échantillon, on réarrange les observations  $x_1, \dots, x_n$  en ordre croissant. La médiane d’échantillon est donnée par

$$\tilde{x} = \begin{cases} \text{la valeur au milieu de la liste des } x_i \text{ ordonnés si } n \text{ est impair.} \\ \text{la moyenne des 2 valeurs au milieu de la liste des } x_i \text{ ordonnés si } n \text{ est pair.} \end{cases}$$

(On définit de même la médiane  $\tilde{\mu}$  d’une population.)

La médiane est souvent utilisée pour décrire des revenus ou des salaires, car elle n’est pas fort influencée par un petit nombre de très gros salaires!

La médiane divise les données en deux parts de tailles égales; on peut diviser l’ensemble des données en 100 parts égales en utilisant les percentiles: le 99<sup>ème</sup> percentile sépare le 1% des résultats les plus hauts des 99 % restants etc...

Une moyenne tronquée est un compromis entre la moyenne et la médiane: une moyenne tronquée à 10% est obtenue en éliminant les plus petits 10 % et les plus grands 10 % de l’échantillon et en faisant la moyenne des résultats restants.

**Exemple 0.3** Voici 20 observations, écrites dans l’ordre croissant, de la durée de vie, en heures, d’un certain type d’ampoules:

|      |      |      |      |      |      |      |      |      |      |      |
|------|------|------|------|------|------|------|------|------|------|------|
| 612  | 623  | 666  | 744  | 883  | 898  | 964  | 970  | 983  | 1003 | 1016 |
| 1022 | 1029 | 1058 | 1085 | 1088 | 1122 | 1135 | 1197 | 1201 |      |      |

La moyenne est  $\bar{x} = \frac{1}{20}(\sum_{i=1}^{20} x_i) = 965,0$

La médiane est  $\tilde{x} = \frac{1003+1016}{2} = 1009,5$

La moyenne tronquée à 10% est la moyenne de toutes les données dont on a retiré les 2 (=10% de 20) plus petites et les 2 plus grandes. Elle est donc égale à

$$\frac{1}{16}(666 + 744 + 883 + 898 + 964 + 970 + 983 + 1003 + 1016 + 1022 + 1029 + 1058 + 1085 + 1088 + 1122 + 1135) = 979,1$$

### 0.3 Mesures de la dispersion d'un ensemble de données

La quantité  $x_i - \bar{x}$  est appelée la déviation de la  $i^{\text{ème}}$  observation par rapport à la moyenne.

**Remarque 0.1** La moyenne des déviations est nulle

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

**Définition 0.3** La variance d'échantillon de l'ensemble  $\{x_1, \dots, x_n\}$  d'observations numériques, notée  $S^2$ , est définie par

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

! On divise par  $(n-1)$ ; on obtient un nombre un peu plus grand que la moyenne des carrés des déviations.

**Définition 0.4** La déviati on standard (ou écart-type) d'échantillon, notée  $S$ , est la racine positive de la variance d'échantillon.

**Remarque 0.2** On définira la variance  $\sigma^2$  et la déviati on standard  $\sigma$  d'une population. Quand la population est finie et consiste en  $\{y_1, \dots, y_N\}$ , on définit

$$\sigma^2 = \frac{1}{N} \left( \sum_{i=1}^N (y_i - \mu)^2 \right)$$

et  $\sigma$  sa racine carrée positive. Nous verrons que  $\bar{x}$  (la moyenne d'échantillon) servira à estimer  $\mu$  (la moyenne de la population). Comme  $\mu$  n'est pas connu, pour estimer  $\sigma$  on se sert des carrés des déviations avec  $\bar{x}$  mais  $x_i$  est généralement plus proche de  $\bar{x}$  que de  $\mu$  et on compense cela en divisant par  $(n - 1)$  dans  $S^2$  pour estimer  $\sigma^2$ . Nous reviendrons sur ce point de manière précise ultérieurement!

**Proposition 0.1** *La variance échantillon est:*

$$S^2 = \frac{\sum_{i=1}^n x_i^2 - \frac{1}{n}(\sum_{i=1}^n x_i)^2}{n - 1}.$$

Preuve:

$$\begin{aligned} (n - 1)S^2 &= \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x}(n\bar{x}) + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - n\bar{x}^2 = \sum_{i=1}^n x_i^2 - \frac{1}{n}(\sum_{i=1}^n x_i)^2 \end{aligned}$$

**Proposition 0.2** *Si  $C$  est une constante ajoutée (ou soustraite) de chacune des données, la variance ne change pas. Si chacune des données est multipliée par une constante  $D$ , la variance est multipliée par  $D^2$ . En d'autres termes: Soit  $\{x_1, \dots, x_n\}$  un échantillon. Soient  $C$  et  $D$  des constantes.*

1. Si  $y_1 = x_1 + C, \dots, y_n = x_n + C$ , alors  $S_y^2 = S_x^2$ ;
2. Si  $z_1 = Dx_1, \dots, z_n = Dx_n$ , alors  $S_z^2 = D^2 S_x^2$ .

**Exemple 0.4** Le nombre de mutants de la bactérie E.Coli qui résistent au virus T1, observé dans différentes cultures (où une même méthode particulière de culture a été chaque fois appliquée) est : 14, 15, 13, 21, 15, 14, 26, 16, 20, 13.

La moyenne est

$$\bar{x} = \frac{1}{10}(14 + 15 + 13 + 21 + 15 + 14 + 26 + 16 + 20 + 13) = \frac{167}{10} = 16,7$$

La variance est

$$\begin{aligned} S^2 &= \frac{1}{9}(2 \times (3,7)^2 + 2 \times (2,7)^2 + 2 \times (1,7)^2 + (0,7)^2 + (3,3)^2 + (4,3)^2 + (9,3)^2) \\ &= \frac{1}{9}(27,38 + 14,58 + 5,78 + 0,49 + 10,89 + 18,49 + 86,49) = 18,2333 \end{aligned}$$

et

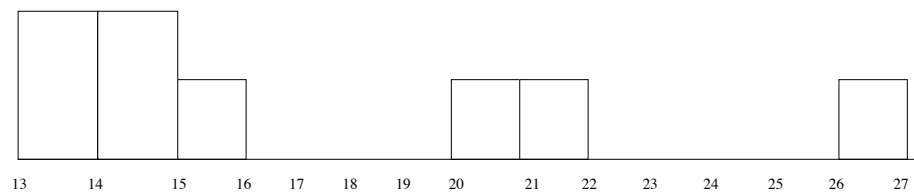
$$\sum_{i=1}^{10} x_i^2 = 2953 \quad \frac{1}{10} \left( \sum_{i=1}^n x_i \right)^2 = 2788,9$$

donc on a aussi

$$\frac{1}{9} \sum_{i=1}^n x_i^2 - \frac{1}{10} \left( \sum_{i=1}^n x_i \right)^2 = S^2$$

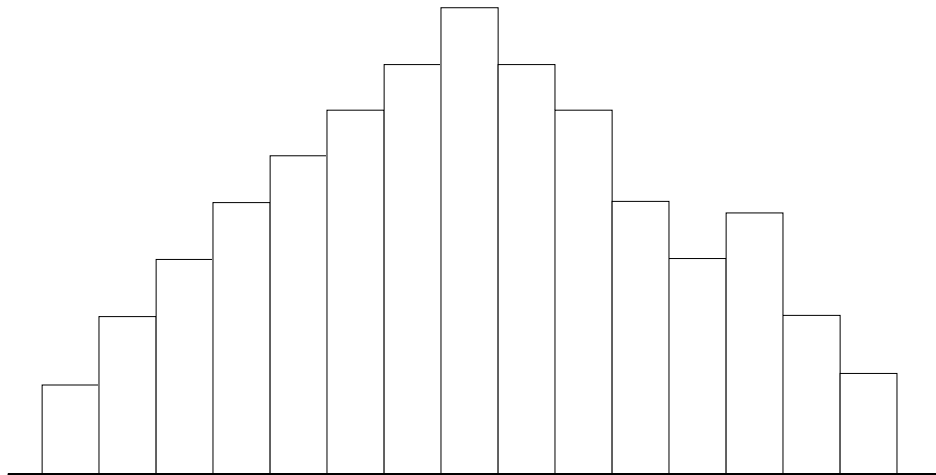
**Histogramme**

**Très dispersé !!**



## Une règle empirique

La moyenne et la déviation standard résument très bien les propriétés d'histogrammes assez symétriques ayant une forme de monticule:



Dans ce cas, environ 68% des données sont à moins d'une déviation standard de la moyenne et 95% des données sont à moins de deux fois la déviation standard de la moyenne (donc dans l'intervalle  $(\bar{x} - 2s, \bar{x} + 2s)$ ).

### Exemple

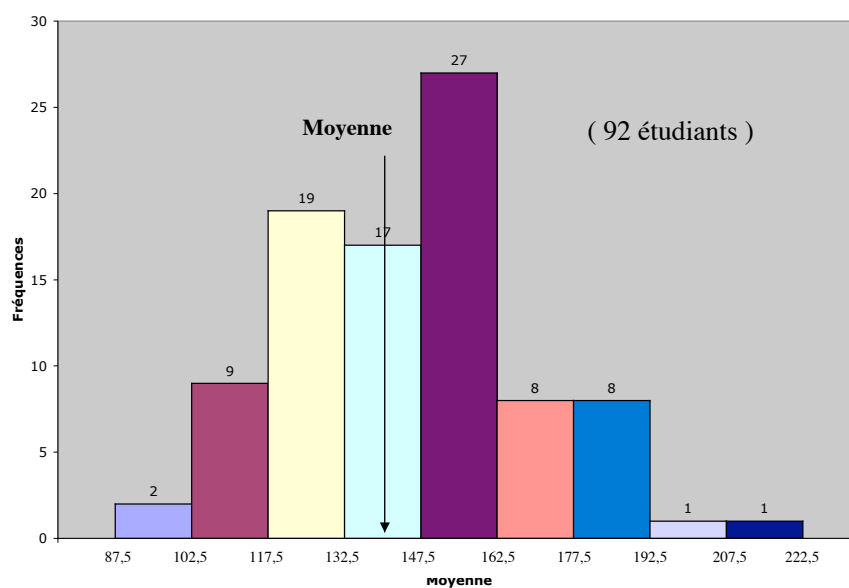
Les poids (en livres!) d'une classe d'étudiants de Penn state sont les suivants:

95,102,108,108,110,110,112,115,115,116,116,118,118,118,120,120,120,121,122,  
123,125,125,125,125,125,125,130,130,130,130,130,131,133,135,135,135,136,138,  
138,140,140,140,140,142,145,145,145,145,145,148,10 x(150),153,10 x(155),157,  
4 x(160),164,165,4 x(170), 2 x(175), 3 x(180),185,4 x(190),195,215



Comme les étudiants ont souvent “arrondi” leur poids à 5 ou 10 prenons les intervalles.

| <u>Intervalles</u> | <u>Fréquence</u> |
|--------------------|------------------|
| 87,5-102,4         | 2                |
| 102,5-117,4        | 9                |
| 117,5-132,4        | 19               |
| 132,5-147,4        | 17               |
| 147,5-162,4        | 27               |
| 162,5-177,4        | 8                |
| 177,5-192,4        | 8                |
| 192,5-207,4        | 1                |
| 207,5-222,4        | 1                |



Moyenne  $\bar{x} = 145,15$  livres      Médiane=145 livres - Ecart-type (= déviation standard)  $s \simeq 23,7$  livres  
 64 % entre  $(\bar{x} - s)$  et  $(\bar{x} + s) = 59$  étudiants sur 92 - 97 % entre  $(\bar{x} - 2s)$  et  $(\bar{x} + 2s) = 89$  étudiants sur 93

**Partie I**

**PROBABILITES**

L'étude des probabilités est une branche des mathématiques qui a plus de 300 ans et dont l'origine est liée à des questions de jeux. Ainsi le Chevalier De Mere posa la question suivante à son ami, le génie Blaise Pascal (1623-1666):

Qu'est-ce qui a le plus de chance de se produire:

- obtenir au moins une fois un 6 en jetant un dé 4 fois ou
- obtenir au moins une fois un double 6 en jetant une paire de dés 24 fois

Réponse ..?

# Chapitre 1

## Définitions de base

### 1.1 Définitions

Une expérience aléatoire est une expérience dont on ne peut prévoir le résultat avec certitude. Il faut avoir précisé ce qu'on entend par résultat; pour cela on définit l'espace fondamental,  $\Omega$ , qui est l'ensemble des résultats possibles. Un élément  $\omega \in \Omega$  est appelé résultat élémentaire.

! Le choix de  $\Omega$  dépend des idées qu'on a (et de ce qui nous intéresse) à propos de l'expérience aléatoire. Par exemple, si l'expérience consiste à jeter une pièce de monnaie, on peut prendre pour espace fondamental:

$$\Omega_1 = \{\text{Pile, Face}\}$$

$$\Omega_2 = \{\text{Pile, Face, Tranche}\}$$

$$\Omega_3 = \mathbb{R}^3 \text{ ou partie de } \mathbb{R}^3 = \{(x, y, z) \text{ coordonnées du centre de gravité de la pièce retombée}\}$$

....

Un événement est aléatoire si, une fois l'expérience effectuée, on peut dire si cet événement a été réalisé ou non. [Par exemple, si on lance un dé, un événement aléatoire peut-être: obtenir un 6 ou obtenir un résultat pair ou obtenir un résultat  $\geq 4 \dots$  ].

Un événement aléatoire  $A$  peut-être identifié à la partie de  $\Omega$  dont les éléments réalisent  $A$ . [Par exemple, si on lance un dé et qu'on est intéressé par le nombre obtenu,  $\Omega = \{1, 2, 3, 4, 5, 6\}$ ; l'événement  $A =$  obtenir un résultat pair

peut être identifié au sous-ensemble  $\{2, 4, 6\} \dots$ .

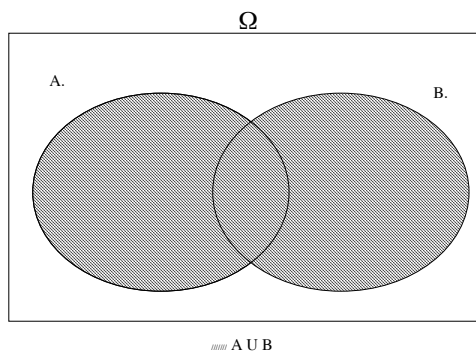
Un événement est donc un sous-ensemble de  $\Omega$ .

! L'ensemble de toutes les parties de  $\Omega$  est en général trop grand pour que l'on puisse le "probabiliser" de manière intéressante, en particulier si  $\Omega = \mathbb{R}$ . On regarde une classe  $\mathcal{A}$  de parties de  $\Omega$  et on réservera le nom d'événement aléatoire (ou événement tout court) à une partie de  $\Omega$  qui appartient à la classe  $\mathcal{A}$ .

La classe des événements,  $\mathcal{A}$ , est une algèbre de Boole de parties de  $\Omega$ .

**Définition 1.1** Une classe  $\mathcal{A}$  de parties d'un ensemble  $\Omega$  est une algèbre de Boole ssi elle vérifie les 4 propriétés suivantes:

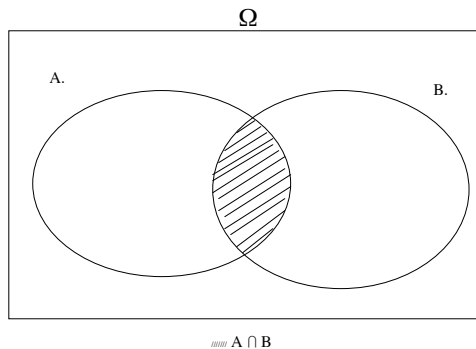
1) Si  $A$  et  $B \in \mathcal{A}$ , alors  $A \cup B \in \mathcal{A}$



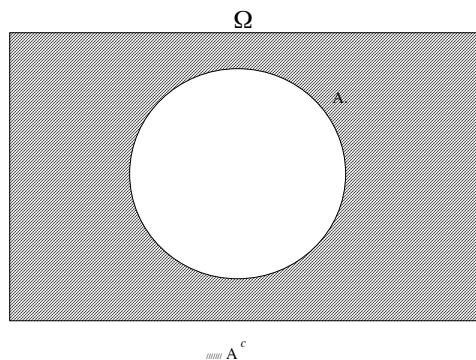
( $A \cup B$  est l'événement constitué de tous les résultats qui sont dans  $A$ , dans  $B$  ou dans les deux événements.  $A \cup B$  représente  $A$  ou  $B$ ; l'événement  $A$  ou l'événement  $B$  ou les deux ont lieu).

2) Si  $A$  et  $B \in \mathcal{A}$ , alors  $A \cap B \in \mathcal{A}$

( $A \cap B$  est l'événement constitué de tous les résultats qui sont à la fois dans  $A$  et dans  $B$ ;  $A \cap B$  représente  $A$  et  $B$



3) Si  $A \in \mathcal{A}$ , alors le complémentaire de  $A$ , noté  $A^c$ ,  $\in \mathcal{A}$



( $A^c$  est l'événement constitué de tous les résultats qui ne sont pas dans  $A$ ,  $A^c = \Omega \setminus A$ ;  $A^c$  représente non  $A$ )

4)  $\Omega \in \mathcal{A}$

(donc l'ensemble de tous les résultats possibles est un événement).

**Remarque 1.1** L'ensemble vide, noté  $\emptyset$ , est toujours dans la classe  $\mathcal{A}$  car  $\emptyset = \Omega^c$ .

**Remarque 1.2** Si  $\Omega$  est un ensemble fini, on peut prendre (et on prendra!) comme classe des événements la classe de toutes les parties de  $\Omega$ , donc  $\mathcal{A} = \mathcal{P}(\Omega)$ , et tout sous-ensemble de  $\Omega$  peut-être vu comme un événement; il en est de même si  $\Omega$  est dénombrable.

**Remarque 1.3** On verra plus tard que les unions dénombrables d'événements sont des événements (en étendant -ce qui est toujours possible-  $\mathcal{A}$ ).

**Remarque 1.4** Si  $\Omega = \mathbb{R}$ , on prendra pour  $\mathcal{A}$  l'ensemble des unions finies ou dénombrables d'intervalles ouverts, fermés, semi-ouverts ou semi-fermés de  $\mathbb{R}$ .

### On veut définir ce qu'est la probabilité d'un événement

Si  $\Omega$  est un ensemble fini et si tous les résultats élémentaires sont "également probables", alors la probabilité d'un événement  $A \in \mathcal{P}(\Omega)$  est donnée par

$$\frac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}}$$

! Cette définition est tautologique puisqu'on utilise une notion d'également probables pour définir la probabilité. Cette notion ne peut donner des résultats que pour un ensemble fini.

### Exemples d'espaces fondamentaux, événements et probabilités

| <u>Expériences aléatoires</u>                                  | <u>Espace fondamental <math>\Omega</math></u>      |
|--|--|
| 1. Lancer d'une pièce une fois                                 | $\Omega = \{P, F\} = \{0, 1\}$                     |
| 2. Lancer d'une pièce jusqu'au 1 <sup>er</sup> pile            | $\Omega = \{P, FP, FFP, F^3P, \dots\}$             |
| 3. Choix d'une carte dans un jeu de 52 cartes                  | $\Omega = \{As, \spadesuit, 2, \clubsuit, \dots\}$ |
| 4. Chronométrage de la durée de vie d'une ampoule              | $\Omega = \mathbb{R}^+$                            |
| 5. Sondage avant un référendum d'un échantillon de 100 votants | $\Omega = \{\text{Oui}, \text{Non}\}^{100}$        |

## Evénements

1. Quatre événements  $\emptyset \{P\} \{F\} \Omega$ .
2. Nombre infini d'événements par ex:  
 $A_1 = \text{"nombre pair de lancer"} = \{FP, FFFP, F^5P, \dots\}$  ou  
 $A_2 = \text{"plus de 10 lancers"} = \{F^{10}P, F^{11}P, F^{12}P, \dots\}$ .
3. Il y a 52 résultats élémentaires, donc  $2^{52}$  événements.
4. Nombre infini d'événements, par exemple  $A_1 = \text{"vie} \geq x" = [x, \infty)$ .
5.  $2^{2^{100}}$  événements.

(Rappel: si  $\Omega$  contient  $n$  éléments, il y a  $2^n$  parties de  $\Omega$ ).

## Probabilités

1.

$$\begin{cases} \text{Proba}(F) = 1/2 (\text{si la pièce n'est pas truquée ; par symétrie}) \\ \text{Proba}(\emptyset) = 0 \\ \text{Proba}(P) = 1/2 \\ \text{Proba}(\Omega) = 1 \end{cases}$$

3.  $\text{Proba}(\text{choisir une dame}) = \frac{4}{52} = \frac{1}{13}$

**Exemple 1.1** En supposant que les dates d'anniversaire, pour une année de 365 jours (oublions les années bisextiles!) soient toutes également probables, quelle est la probabilité pour qu'au moins deux individus dans un groupe de 23 personnes aient le même anniversaire?

Combien de façons d'avoir 23 anniversaires distincts:

$$365 \times 364 \times 363 \times \dots \times (365 - 22)$$

Parmi combien de possibilités:

$$(365)^{23}$$

$P[\text{au moins 2 personnes parmi les 23 aient le même anniversaire}] =$

$$\frac{(365)^{23} - 365 \times 364 \times \dots \times (365 - 22)}{(365)^{23}} \approx 0,51$$



(Pour 2 personnes, c'est  $\frac{1}{365} \approx 0,003$ ; pour 22 personnes, c'est  $\approx 0,48$ ; pour 366 personnes, c'est 1 !!)

Que fait-on si  $\Omega$  est infini ou si les résultats élémentaires ne sont pas “également probables”? Regardons la **notion fréquentielle de probabilité**. Considérons une expérience aléatoire  $E$ , l'espace fondamental  $\Omega$  des résultats possibles et un événement  $A (\in \mathcal{A} \subset \mathcal{P}(\Omega))$ . Répétons l'expérience  $n$  fois et mesurons le nombre  $n(A)$  de fois que l'événement  $A$  se réalise (donc que le résultat de l'expérience soit dans  $A$ ); on a bien sûr que  $n - n(A)$  est le nombre de fois que l'événement  $A^c$  se réalise. **Constataction expérimentale:** la fréquence relative  $\frac{n(A)}{n}$  tend, lorsque  $n$  tend vers l'infini, vers une constante qui dépend du choix de l'événement  $A$  et que l'on note  $P(A)$ .

**Remarque 1.5** Ce nombre  $P(A)$  est intuitivement la probabilité de  $A$ . On raccroche ici la notion de probabilité à l'expérience, mais faire tendre  $n$  vers l'infini est impossible expérimentalement. Il nous faudra donc une définition plus mathématique, qui tiendra compte des propriétés de cette notion expérimentale de probabilité.

**Propriétés:**

- Pour tout événement  $A$   $0 \leq P(A) \leq 1$  ;  $P(\Omega) = 1, P(A) + P(A^c) = 1$
- Si  $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$
- Si  $\Omega$  est fini  $\Omega = \{\omega_1, \dots, \omega_k\}$ , on peut écrire  $\Omega = \{\omega_1\} \cup \dots \cup \{\omega_k\}$  (Union disjointe), donc

$$P(\Omega) = \sum_{i=1}^k P(\{\omega_i\}) = 1.$$

Si les résultats élémentaires sont équiprobables, on a donc  $P(\{\omega_i\}) = 1/k \forall k$  et on retrouve les résultats classiques:

si  $A = \{\omega_i, \dots, \omega_{i_r}\}$ ,  $P(A) = \frac{r}{k}$  ( $r = \#A, k = \#\Omega$ ).

**Définition 1.2** Si on s'est donné un espace fondamental  $\Omega$  et la classe des événements  $\mathcal{A} \subset \mathcal{P}(\Omega)$ , une mesure de probabilité est une application qui associe à tout événement  $A \in \mathcal{A}$  un nombre  $P(A)$ , appelé probabilité de  $A$

de manière telle que:

$$0 \leq P(A) \leq 1 \quad (1.1)$$

$$P(\Omega) = 1 \quad (1.2)$$

Si  $A_1$  et  $A_2$  sont deux événements disjoints

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) \quad (1.3)$$

Plus généralement, si  $A_1, A_2, \dots$  est une suite d'événements disjoints 2 à 2 et si leur union est un événement, alors

$$P(A_1 \cup A_2 \cup A_3 \dots) = P(A_1) + P(A_2) + P(A_3) + \dots \quad (1.3')$$

**Définition 1.3** Une famille  $\mathcal{A}$  de sous-ensemble de  $\Omega$  est appelée tribu ou  $\sigma$ -algèbre si:

- $\Omega \in \mathcal{A}$
- $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
- $A_1, A_2, \dots \in \mathcal{A} \Rightarrow A_1 \cup A_2 \cup \dots \in \mathcal{A}$

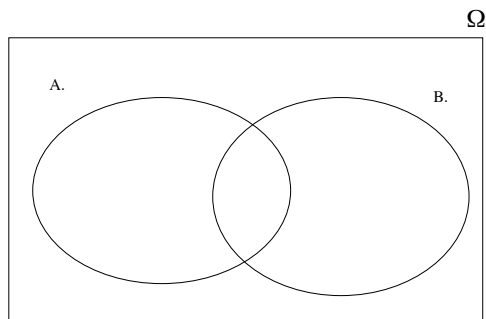
On peut montrer que toute algèbre de Boole de parties de  $\Omega$ ,  $\mathcal{A}$ , est contenue dans une  $\sigma$ -algèbre dite engendrée par  $\mathcal{A}$ ; si on a une mesure de probabilité sur  $\mathcal{A}$  elle s'étend en une mesure de probabilité sur cette  $\sigma$ -algèbre. La propriété (1.3') est appelée la  $\sigma$ -additivité de la mesure de probabilité.

**Dans la suite, nous supposons qu'on considère toujours une classe d'événements  $\mathcal{A}$  qui est une  $\sigma$ -algèbre; donc les unions et les intersections dénombrables d'événements sont des événements.**

**Définition 1.4** Le triple  $(\Omega, \mathcal{A}, P)$  est un espace probabilisé, quand  $\Omega$  est un ensemble (appelé l'ensemble des résultats possibles), quand  $\mathcal{A}$  est une tribu de parties de  $\Omega$  (ces parties sont appelées les événements) et quand  $P : \mathcal{A} \rightarrow \mathbb{R}$  est une mesure de probabilité sur  $\mathcal{A}$ .

**Propriétés:**

- $P(A^c) = 1 - P(A) \quad \forall \text{ événement } A$   
(Car  $\Omega$  est l'union disjointe de  $A$  et  $A^c$  donc  $P(A) + P(A^c) = P(\Omega) = 1$ )
- $P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \forall \text{ événements } A \text{ et } B$



En effet, on les unions disjointes suivantes:

$$\begin{cases} A = (A \setminus B) \cup (A \cap B) \\ B = (B \setminus A) \cup (A \cap B) \\ A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B) \end{cases}$$

- $P(\emptyset) = 0$

## 1.2 Exemples

1) Si  $\Omega$  est fini,  $\Omega = \{w_1, \dots, w_k\}$ , on prend  $\mathcal{A} = \mathcal{P}(\Omega)$  et on peut définir  $P(A) = \frac{1}{k} \#A \quad \forall A \in \mathcal{P}(\Omega)$   
(probabilité correspondant à des événements équiprobables).

2) Si  $\Omega$  est (fini ou) dénombrable  $\Omega = \{w_1, w_2, \dots\}$ , on prend  $\mathcal{A} = \mathcal{P}(\Omega)$  et on peut définir  $P(\{w_i\}) = p_i$  avec  $\sum_{i=1}^{\infty} p_i = 1$  alors

$$P(A) = \sum_{w_i \in A} p_i$$

Par exemple, si  $\Omega = \{P, FP, FFP, FFFP, \dots\}$  (lancer d'une pile jusqu'au 1<sup>er</sup> pile):

$$P(\{P\}) = 1/2, P(\{FP\}) = 1/4, P(\{FFP\}) = 1/8, \dots,$$

$$P(\{\overbrace{F \dots F}^{k \text{ fois}} P\}) = \frac{1}{2^{k+1}}, \dots$$

### 3) Problème du Chevalier de Mere

L'expérience 1 consiste à lancer 4 fois un dé. Donc  $\Omega = \{(a, b, c, d) \text{ avec } 1 \leq a, b, c, d \leq 6\}$  et tous les résultats sont également probables.

Soit  $A$ , l'événement: avoir au moins un 6.

# cas favorable =  $6^3 + 5 \times 6^2 + 25 \times 6 + 125 = 671$

# cas possibles =  $6^4 = 1296$  donc  $P(A_1) = \frac{671}{1296} \simeq 0,518$

Remarquons qu'il est plus facile de calculer  $P(A^c)$  !

$A_1^c$  = n'avoir aucun 6 donc # cas favorables à  $A_1^c = 5^4$

$P(A_1^c) = \frac{5^4}{6^4} \simeq 0,482$

Donc, on obtient également ainsi  $P(A_1) = 1 - P(A_1^c) = 0,518$ .

L'expérience 2 consiste à lancer 24 fois un double dé. L'événement  $A_2$  est : avoir au moins un double 6. Il est de nouveau plus aisé de calculer la probabilité de  $A_2^c$ .

$A_2^c$  = n'avoir aucun double 6

# cas favorables à  $A_2^c = (35)^{24}$

# cas possibles =  $(36)^{24}$

Donc  $P(A_2^c) = \frac{35^{24}}{36^{24}} \simeq 0,509$

et  $P(A_2) = 1 - P(A_2^c) = 0,491$

### 4) Expérience de l'aiguille de Buffon (1777)

On jette une aiguille de longueur 1 sur un plancher dont les planches sont de largeur 1. Quelle est la probabilité que l'aiguille touche une ligne entre deux planches? On regarde la planche où se trouve le milieu de l'aiguille. Si l'aiguille fait un angle  $\theta$  avec les lignes entre les planches, la probabilité que l'aiguille touche une ligne est  $P_\theta = \sin \theta$  car l'aiguille touchera une ligne ssi le centre de l'aiguille est à une distance au plus égale à  $\frac{1}{2} \sin \theta$  de la ligne. Quand on laisse tomber l'aiguille, la valeur de l'angle qu'elle forme avec la ligne des planches est un nombre pris de manière uniforme sur  $[0, \pi[$  donc

$$P = \frac{1}{\pi} \int_0^\pi P_\theta d\theta = \frac{1}{\pi} \int_0^\pi \sin \theta d\theta = \frac{-1}{\pi} [\cos \theta]_0^\pi = \frac{2}{\pi} = 0,6366....$$

Cette expérience donne une estimation de  $\pi$ !

### 1.2.1 Récréation: Le paradoxe de Simpson

On étudie deux méthodes de traitement des calculs rénaux, l'une par chirurgie, l'autre par néphrolithotomie. On prend un groupe de 700 patients. La moitié d'entre eux, 350, sont traités par chirurgie (C), l'autre moitié par néphrolithotomie (N).

Parmi ceux traités par (C), la proportion de guérison est de  $\frac{273}{350} \approx 0,78$ .  
Parmi ceux traités par (N), la proportion de guérison est de  $\frac{289}{350} \approx 0,83$ .

La néphrolithotomie semble plus efficace....mais si on considère la taille du calcul on a les résultats suivants:

Parmi ceux traités par (C)  $\left\{ \begin{array}{l} \text{la proportion de guérison de grand calcul est de } \approx 0,93 \\ \text{la proportion de guérison de petit calcul est de } \approx 0,73 \end{array} \right.$

Parmi ceux traités par (N)  $\left\{ \begin{array}{l} \text{la proportion de guérison de grand calcul est de } \approx 0,87 \\ \text{la proportion de guérison de petit calcul est de } \approx 0,69 \end{array} \right.$

Donc, aussi bien pour les petits calculs que pour les grands calculs, la chirurgie semble plus efficace !!

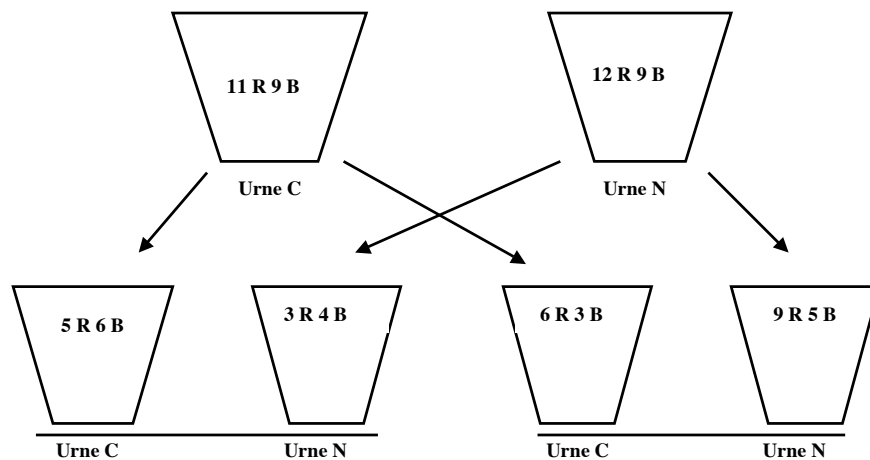
### COMMENT EST-CE POSSIBLE ?

Rien d'étonnant pourtant ! Un exemple où l'objectif est d'extraire une boule rouge prise au hasard dans une urne où il y a des boules rouges et des boules blanches. Dans l'urne  $C$  il y a 11 boules rouges et 9 boules blanches. Dans l'urne  $N$ , il y a 12 boules rouges et 9 boules blanches.

Dans l'urne  $C$  la probabilité de tirer une boule rouge est  $= \frac{11}{20}$

Dans l'urne  $N$  la probabilité de tirer une boule rouge est  $= \frac{12}{21}$

$\Rightarrow$  Il vaut mieux piocher dans l'urne  $N$ . Divisons maintenant l'urne  $C$  en deux urnes  $C_1$  contenant 5 boules rouges et 6 boules blanches et  $C_2$  contenant 6 boules rouges et 3 boules blanches. Divisons également l'urne  $N$  en une urne  $N_1$  contenant 3 boules rouges et 4 boules blanches et une urne  $N_2$  contenant 9 boules rouges et 5 blanches.



Dans  $C_1$  la proba est  $5/11$     Dans  $C_2$  proba =  $6/9 = 2/3$      $\Rightarrow$  Dans chaque cas, il vaut  
 Dans  $N_1$  la proba est  $3/7$     Dans  $N_2$  proba =  $9/14$     mieux piocher dans l'urne  $C$ !

$a_1 = \# \text{boules rouges dans la 1ere petite urne } C_1$     ( $= 5$ )

$b_1 = \# \text{boules (R ou B) dans la 1ere petite urne } C_1$     ( $= 11$ )

$c_1 = \# \text{boules rouges dans la 1ere petite urne } N_1$     ( $= 3$ )

$d_1 = \# \text{boules (R ou B) dans la 1ere petite urne } N_1$     ( $= 7$ )

$\frac{a_1}{b_1} = \text{proba de piocher R dans la 1re petite urne } C_1 = 5/11$

$\frac{c_1}{d_1} = \text{proba de piocher R dans la 1re petite urne } N_1 = 3/7$

$$\frac{a_1}{b_1} > \frac{c_1}{d_1}$$

de même

$$\frac{a_2}{b_2} > \frac{c_2}{d_2}$$

(Le 2 réfère aux deuxièmes petites urnes)( $a_2 = 6, b_2 = 9, c_2 = 9, d_2 = 14$ )

Mais

$a_1 + a_2 = \#$  boules rouges dans les urnes C réunies (11)

$b_1 + b_2 = \#$  boules R ou B dans les urnes C réunies (20)

$c_1 + c_2 = \#$  boules rouges dans les urnes N réunies (12)

$d_1 + d_2 = \#$  boules R ou B dans les urnes N réunies (21)

et on a

Proba de pioche R dans C =  $\frac{a_1+a_2}{b_1+b_2} < \frac{c_1+c_2}{d_1+d_2}$  ( $\frac{11}{20} < \frac{12}{21}$ )

C'est la même situation pour les guérisons (boules rouges) avec le traitement C ou avec le traitement N !!

# Chapitre 2

## Probabilités conditionnelles

### 2.1 Définition

Jetons deux dés (non truqués!). Quelle est la probabilité que la somme des dés soit 2 ? Réponse:  $\frac{1}{36}$ .

Supposons avoir déjà jeté un dé avec le résultat 1; lançons l'autre dé. Quelle est la probabilité que la somme des dés soit 2 ? Réponse:  $\frac{1}{6}$ .

On l'appelle la probabilité conditionnelle de l'événement  $A$  (obtenir une somme de 2 en jetant deux dés), étant donné qu'une condition  $C$  est réalisée (que le premier dé lancé ait donné 1). On la note  $P(A|B)$  et on dit: probabilité de  $A$ , étant donné  $C$ .

Autre exemple: prenons une carte au milieu d'un paquet de 52 cartes. Quelle est la probabilité de tirer un as ?  $\frac{4}{52} = \frac{1}{13}$ .

Quelle est la probabilité de tirer un as, sachant que la carte du bas du paquet est l'as de pique ?  $\frac{3}{51} = \frac{1}{17}$ .

#### Approche fréquentielle:

Considérons une expérience  $E$ , des événements  $A$  et  $B$ . Supposons  $P(B) \neq 0$ . Faisons  $n$  répétitions de l'expérience  $E$  et comptons la fréquence  $n(A)$  où l'événement  $A$  se réalise, et de même  $n(B), n(A \cap B)$ . Alors:

$\frac{n(A)}{n}$  converge vers  $P(A)$  quand  $n$  converge vers  $\infty$ ,

$\frac{n(B)}{n}$  converge vers  $P(B)$  quand  $n$  converge vers  $\infty$ ,

$\frac{n(A \cap B)}{n}$  converge vers  $P(A \cap B)$  quand  $n$  converge vers  $\infty$ .



Intéressons nous aux seules répétitions de  $E$  où l'événement  $B$  s'est réalisé (il y en a  $n(B)$ ), et parmi celles-là, regardons le nombre de fois que  $A$  s'est aussi réalisée: c'est  $n(A \cap B)$ . La fréquence relative de  $A$  sachant  $B$  est donc

$$\frac{n(A \cap B)}{n(B)} = \frac{n(A \cap B)/n}{n(B)/n} \text{ qui converge vers } \frac{P(A \cap B)}{P(B)} \text{ quand } n \rightarrow \infty$$

Ainsi la probabilité conditionnelle de  $A$  sachant  $B$ , qui est la limite de  $\frac{n(A \cap B)}{n(B)}$  quand  $n$  tend vers l'infini, est donnée par

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

**Définition 2.1** Soit  $B$  un événement tel que  $P(B) \neq 0$ .

La probabilité conditionnelle d'un événement  $A$  sachant  $B$  (ou sous la condition  $B$ ), notée  $P(A|B)$  est définie par

$$P(A|B) := P(A \cap B)/P(B) \quad (2.1)$$

**Remarque 2.1** Si  $P(A|B)$  est très différent de  $P(A)$ , cela veut dire qu'il y a un lien important entre  $A$  et  $B$ . Intuitivement, si  $P(A|B) = P(A)$  (avec  $P(B) \neq 0$ ) la probabilité de  $A$  sous la condition de  $B$  est la même que la probabilité de  $A$ , donc  $A$  est "indépendant" de  $B$ .

**Définition 2.2** Deux événements  $A$  et  $B$  sont dits indépendants si

$$P(A \cap B) = P(A).P(B) \quad (2.2)$$

Plus généralement, les événements  $A_1, A_2, A_3, \dots, A_n$  sont dits indépendants si

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1}).P(A_{i_2}) \dots P(A_{i_k}) \quad (2.3)$$

pour tout  $k$  avec  $2 \leq k \leq n$  et pour toute suite  $i_1 < i_2 < \dots < i_k$  d'entiers appartenant à  $\{1, \dots, n\}$ .

**Exemple 2.1** On jette une pièce de monnaie 3 fois (pièce non truquée!)

L'événement  $A$  est: obtenir pile au 3<sup>ème</sup> lancer  $\rightarrow P(A) = 1/2$

L'événement  $B$  est: obtenir pile aux 1<sup>er</sup> et 2<sup>ème</sup> lancers  $\rightarrow P(B) = 1/4$

La probabilité d'obtenir pile au 3<sup>ème</sup> lancer, sachant qu'on a eu pile aux 1<sup>er</sup> et 2<sup>ème</sup> lancer est  $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/8}{1/4} = 1/2 = P(A)$ . Le résultat au 3<sup>ème</sup> lancer est indépendant des résultats précédents!

## 2.2 Théorèmes

**Théorème 2.1 (Théorème de multiplication)** Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé. Soient  $A$  et  $B$  deux événements tels que  $P(B) \neq 0$ . Alors

$$P(A \cap B) = P(A|B)P(B). \quad (2.4)$$

Plus généralement, si  $A_1, \dots, A_n \in \mathcal{A}$  sont tels que  $P(A_2 \cap \dots \cap A_n) \neq 0$ , alors

$$P(A_1 \cap A_2 \cap A_3) = P(A_1|A_2 \cap A_3)P(A_2|A_3)P(A_3) \quad (2.5)$$

$$\begin{aligned} P(A_1 \cap A_2 \cap \dots \cap A_n) &= P(A_1|A_2 \cap \dots \cap A_n) \\ &\times P(A_2|A_3 \cap \dots \cap A_n) \\ &\times \dots \\ &\times P(A_{n-1}|A_n) \\ &\times P(A_n) \end{aligned} \quad (2.6)$$

**Théorème 2.2 (Théorème des probabilités totales)** Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé. Soient  $A, B$  deux événements, tels que  $0 < P(B) < 1$ . Alors,

$$P(A) = P(A|B).P(B) + P(A|B^c)P(B^c). \quad (2.7)$$

Plus généralement, soit  $\{B_1, B_2, \dots\} \subset \mathcal{A}$  un système exhaustif d'événements, c'est-à-dire tels que

$$\cup_i B_i = \Omega \quad (2.8)$$

$$B_i \cap B_j = \emptyset \quad \forall i \neq j \quad (2.9)$$

$$P(B_i) \neq 0 \quad \forall i \quad (2.10)$$

Alors,

$$P(A) = \sum_i P(A|B_i).P(B_i)$$

### Démonstration

$$\begin{aligned} P(A) &= P(A \cap (B \cup B^c)) \\ &= P((A \cap B) \cup (A \cap B^c)) \\ &= P(A \cap B) + P(A \cap B^c) \\ &\quad \text{car les événements } A \cap B \text{ et } A \cap B^c \text{ sont disjoints} \\ &= P(A|B).P(B) + P(A|B^c).P(B^c). \end{aligned}$$

### Application: (paradoxe de Lewis Carroll)

On remplit une urne avec 2 boules Blanches ou Noires en tirant au hasard, avec chaque fois  $P(B) = 1/2$ ,  $P(N) = 1/2$ . On tire ensuite une boule au hasard dans l'urne. Quelle est la probabilité d'avoir une boule noire?

$$\begin{aligned} P(\text{boule noire}) &= P(N|2B).P(2B) + P(N|1B, 1N) + P(N|2N).P(2N) \\ &= 0 + 1/2(1/4 + 1/4) + 1.1/4 = 1/2 \end{aligned}$$

Si on ajoute aux 2 boules de l'urne, 1 boule noire avant de tirer la boule au hasard, que devient la probabilité d'avoir une boule noire?

$$\begin{aligned} P(\text{boule noire}) &= P(N|2B1N)1/4 + P(N|1B2N)1/2 + P(N|3N)(1/4) \\ &= 1/3 \times 1/4 + 2/3 \times 1/2 + 1/4 = 2/3 \end{aligned}$$

Raisonnement erroné:  $P(\text{boule noire}) = 2/3 \rightarrow$  l'urne contient  $(2N, 1B) \rightarrow$  au départ elle contenait  $(1N, 1B)!!$

**Théorème 2.3 (Théorème de Bayes ou Probabilité des causes)** Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé. Soient  $A$  et  $B$  deux événements, avec  $P(B) \neq 0$ ,  $P(A) \neq 0$ ,  $P(A) \neq 1$ . Alors,

$$P(A|B) = \frac{P(B|A).P(A)}{P(B|A).P(A) + P(B|A^c).P(A^c)}. \quad (2.11)$$

Plus généralement, soit  $\{B_1, B_2, \dots\} \subset \mathcal{A}$  un système exhaustif d'événements, (donc tels que  $\cup_i B_i = \Omega$ ,  $B_i \cap B_j = \emptyset \forall i \neq j$ ,  $P(B_i) \neq 0 \forall i$ ) et soit  $A \in \mathcal{A}$  tel que  $P(A) \neq 0$ . Alors, pour tout  $j$  fixé

$$P(B_j|A) = \frac{P(A|B_j).P(B_j)}{\sum_i P(A|B_i).P(B_i)}. \quad (2.12)$$

**Démonstration** On a  $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A).P(A)}{P(B)}$  et, par le théorème des proba totales,  $P(B) = P(B|A).P(A) + P(B|A^c).P(A^c)$ .

### Application

Supposons que le taux d'incidence d'une maladie rare soit de  $\frac{1}{100.000}$ .  
Soit  $T$  un test fiable tel que:

$P(T^+ = \text{test positif} \mid S = \text{avoir la maladie}) = 0,99$

$P(T^+ = \text{test positif} \mid S^c = \text{ne pas avoir la maladie}) = 0,002$  (les faux positifs!)

Quelle est la probabilité  $P(S)$  d'avoir la maladie si le test est positif ?

$$\begin{aligned} P(S|T^+) &= \frac{P(T^+|S) \cdot P(S)}{P(T^+|S) \cdot P(S) + P(T^+|S^c) \cdot P(S^c)} \\ &= \frac{0,99 \times \frac{1}{100.000}}{0,99 \times \frac{1}{100.000} + \frac{0,002 \times 99.999}{100.000}} \\ &= \frac{0,99}{0,99 + 199,998} = \frac{0,99}{201,988} \approx 0,005 \end{aligned}$$

Pas de quoi s'affoler !

# Chapitre 3

## Variables aléatoires

L'espace fondamental  $\Omega$  des résultats possibles d'une expérience est un ensemble "amorphe"; il n'a ni ordre, ni notion de somme, produit,... On désire avoir des mesures quantitatives liées à des résultats d'expérience; on veut donc associer un (ou plusieurs) nombre(s) au résultat  $\omega$  de l'expérience. On regarde ainsi une application:

$$X : \Omega \rightarrow \mathbb{R} : \omega \rightarrow X(\omega).$$

Mais toute application  $X$  n'est pas adéquate; on veut pouvoir donner un sens à l'expression "quelle est la probabilité que la valeur de  $X$  soit dans l'intervalle  $(a, b)$ ?" Il faut donc que  $\{\omega \in \Omega | X(\omega) \in (a, b)\}$  appartienne à la classe  $\mathcal{A}$  des événements.

**Définition 3.1** Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé. On appelle variable aléatoire réelle toute application

$$X : \Omega \rightarrow \mathbb{R} \quad \omega \rightarrow X(\omega) \tag{3.1}$$

mesurable pour  $P$ , c'est-à-dire telle que  $\forall x \in \mathbb{R}$ ,

$$X^{-1}(]-\infty, x]) := \{\omega \in \Omega | X(\omega) \leq x\} \in \mathcal{A}. \tag{3.2}$$

**Remarque 3.1** Observons qu'alors  $X^{-1}(I) \in \mathcal{A}$  pour tout intervalle  $I \subset \mathbb{R}$  car tout intervalle est obtenu à partir des intervalles  $]-\infty, x]$   $x \in \mathbb{R}$  par des unions ou intersections dénombrables et des passages aux complémentaires.

**Exemple 3.1** Contenu d'un gobelet de café, sondage préélectoral, variable qui vaut 1 sur l'un des  $\omega$  et 0 ailleurs, total obtenu en lançant 3 dés.....

**Définition 3.2** On appelle loi de probabilité de la variable aléatoire  $X$  (ou distribution de probabilité de  $X$ ) la mesure de probabilité  $P^X$  induite par  $X$  sur  $\mathbb{R}$  définie par

$$P^X(I) := P\{\omega \in \Omega \mid X(\omega) \in I\} \quad (3.3)$$

pour tout intervalle  $I$  de  $\mathbb{R}$  (en se rappelant que l'ensemble des événements sur  $\mathbb{R}$  est l'ensemble des unions finies ou dénombrables d'intervalles de  $\mathbb{R}$ ).

**Remarque 3.2** La distribution de probabilité d'une variable aléatoire  $X$  est entièrement déterminée par les  $P(X^{-1}(]-\infty, x])) = P^X(]-\infty, x]) \forall x \in \mathbb{R}$ .

**Définition 3.3** Soit  $X$  une variable aléatoire. La fonction de répartition de la variable  $X$ , notée  $F^X$ , est la fonction sur  $\mathbb{R}$  à valeurs dans  $\mathbb{R}$  définie par:

$$F^X(x) := P[X \leq x] = P^X(]-\infty, x]) = P(\{\omega \in \Omega \mid X(\omega) \leq x\}) \quad \forall x \in \mathbb{R} \quad (3.4)$$

### Propriétés

- 1)  $\forall x \in \mathbb{R} \quad 0 \leq F^X(x) \leq 1$
- 2) La fonction  $F^X$  est monotone non décroissante, continue à droite
- 3)  $\lim_{x \rightarrow -\infty} F^X(x) = 0 \quad ; \lim_{x \rightarrow +\infty} F^X(x) = 1$ .

## 3.1 Variable aléatoire discrète

On regarde ici une variable aléatoire qui ne prend qu'un nombre fini ou dénombrable de valeurs:  $x_1, x_2, \dots, x_n, \dots$

**Définition 3.4** Une fonction  $X : \Omega \rightarrow \{x_1, x_2, \dots\} \subset \mathbb{R}$  qui associe à tout résultat possible  $\omega$  d'une expérience aléatoire, un nombre  $X(\omega)$  appartenant à une liste donnée, finie ou dénombrable, de nombres réels, et telle que  $X^{-1}\{x_i\} \in \mathcal{A} \forall i \in \mathbb{N}$ , est une variable aléatoire discrète sur  $\Omega$ .

**Remarque 3.3** Si  $\Omega$  est fini ou dénombrable et  $\mathcal{A} = \mathcal{P}(\Omega)$ , toute fonction  $X : \Omega \rightarrow \mathbb{R}$  est une variable aléatoire discrète sur  $\Omega$ .

**Remarque 3.4** La loi de probabilité  $P^X$  d'une valeur aléatoire discrète  $X : \Omega \rightarrow \{x_1, x_2, \dots\} \subset \mathbb{R}$  est entièrement déterminée par les nombres:

$$p^X(x_i) := P["X = x_i"] := P(\{\omega \in \Omega \mid X(\omega) = x_i\}) \text{ pour } i = 1, 2, \dots$$

**Définition 3.5** On appelle fonction de probabilité de la variable aléatoire discrète  $X$  la fonction

$$\{x_1, \dots, x_n\} \rightarrow \mathbb{R} \quad x_i \rightarrow p^X(x_i) = P["X = x_i"] = P(\{\omega \in \Omega \mid X(\omega) = x_i\}). \quad (3.5)$$

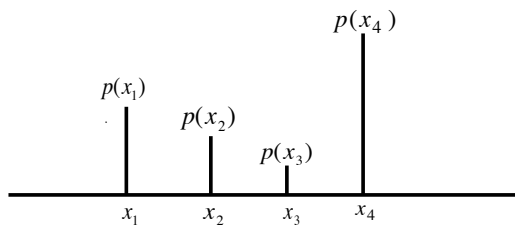
On a

$$p^X(x_1) + p^X(x_2) + \dots = 1 \quad (3.6)$$

car  $\Omega$  est l'union disjointe des  $X^{-1}\{x_i\}$ . De plus

$$P^X((a, b]) = P(\{\omega \in \Omega \mid a < X(\omega) \leq b\}) = \sum_{a < x_i \leq b} p^X(x_i). \quad (3.7)$$

On représente la fonction de probabilité par un diagramme en bâtonnets:



**Exemple 3.2** L'expérience consiste à jeter 3 pièces.

On a  $\Omega = \{ppp, ppf, pfp, fpp, pff, fpf, ffp, fff\}$  et on considère toute partie  $A$  de  $\Omega$  comme un événement. On considère

$X : \Omega \rightarrow \mathbb{R} \quad \omega \rightarrow X(\omega) :=$  le nombre de fois qu'on a eu "face". Donc  $X$  est à valeur dans  $\{0, 1, 2, 3\}$

$$p^X(0) = P(\{\omega \in \Omega \mid X(\omega) = 0\}) = P(\{ppp\}) = 1/8$$

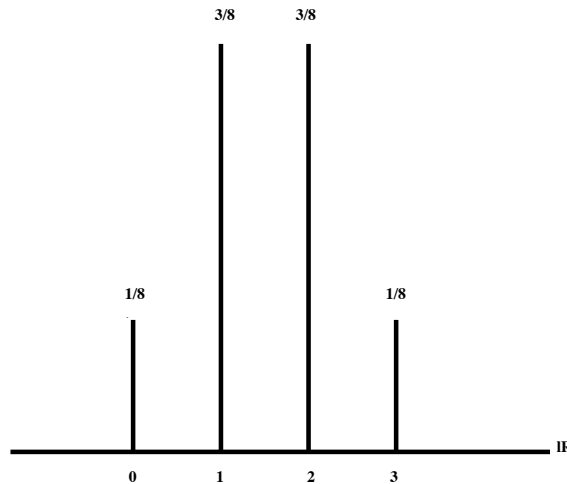
$$p^X(1) = P(\{\omega \in \Omega \mid X(\omega) = 1\}) = P(\{ppf, pfp, fpp\}) = 3/8$$

$$p^X(2) = P(\{\omega \in \Omega \mid X(\omega) = 2\}) = P(\{pff, fpf, ffp\}) = 3/8$$

$$p^X(3) = P(\{\omega \in \Omega \mid X(\omega) = 3\}) = P(\{fff\}) = 1/8$$

$$\begin{aligned} P^X((0, 2]) &= P\{\omega \in \Omega \mid 0 < X(\omega) \leq 2\} \\ &= P\{\omega \in \Omega \mid X(\omega) = 1\} + P\{\omega \in \Omega \mid X(\omega) = 2\} \\ &= p^X(1) + p^X(2) = \frac{3}{8} + \frac{3}{8} = \frac{3}{4} \end{aligned}$$

La fonction de probabilité de  $X$  associe à  $x \in \mathbb{R}$ , la probabilité que la variable  $X$  prenne la valeur  $x$ ; donc elle est nulle si  $x \notin \{0, 1, 2, 3\}$ , elle vaut  $p^X(0) = 1/8$  si  $x = 0$ ,  $p^X(1) = 3/8$  si  $x = 1$ ,  $p^X(2) = 3/8$  si  $x = 2$ ,  $p^X(3) = 1/8$  si  $x = 3$ . Son diagramme en bâtonnets est





**Exemple 3.3** On jette une pièce non truquée jusqu'à avoir un pile. On a  $\Omega = \{p, fp, ffp, fffp, fffffp, \dots\}$  et on considère  $\mathcal{A} = \mathcal{P}(\Omega)$ . Soit  $X : \Omega \rightarrow \mathbb{N}_0 = \{1, 2, 3, \dots\} \subset \mathbb{R}$  la variable aléatoire qui associe à un élément  $\omega \in \Omega$  le nombre  $X(\omega) :=$  le nombre de fois que l'on a du lancer la pièce. Ainsi,

$$X(p) = 1, \quad X(fp) = 2, \quad X(ffp) = 3, \dots$$

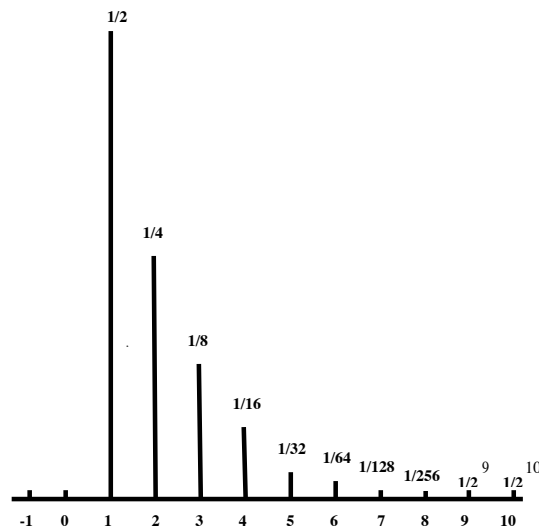
La fonction de probabilité associée à  $X$  est nulle en tout point  $x \in \mathbb{R} \setminus \mathbb{N}_0$  et vaut  $p^X(n)$  pour  $n \in \mathbb{N}_0$  avec

$$p^X(n) = P^X(\{n\}) = P(\{\omega \in \Omega \mid X(\omega) = n\}) = P(\underbrace{\{f \dots f\}}_{n-1} p) = \frac{1}{2^n}$$

Clairement,

$$\sum_{n \in \mathbb{N}_0} p^X(n) = p^X(0) + p^X(1) + \dots = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots = 1$$

Le diagramme en bâtonnets correspondant est



### 3.1.1 La variable binomiale

Supposons qu'on ait une expérience aléatoire avec seulement deux résultats possibles: notons les  $S$  (pour "succès") et  $F$  (pour "échec"); il peut s'agir de lancer une pièce de monnaie pour avoir pile ou face, de voir si l'équipe de football belge gagne ou perd un match, de tester si une pièce sortant d'une chaîne de fabrication est bonne ou mauvaise...

Supposons que la probabilité d'obtenir le résultat  $S$  soit  $p$  (où  $p$  est un nombre réel appartenant à  $[0, 1]$ ), donc celle d'obtenir  $F$  est  $1 - p$ .

On répète cette expérience  $n$  fois.

**Définition 3.6** On parle de schéma de Bernoulli si on répète ainsi  $n$  fois une expérience et si

- chaque fois le résultat de l'expérience est  $S$  ou  $F$ ;
- la probabilité  $p$  d'obtenir le résultat  $S$  est la même à chacune des répétitions;
- le résultat de chaque expérience est indépendant des résultats précédents.

**Définition 3.7** La variable aléatoire binomiale  $X$  mesure le nombre de succès obtenus dans un schéma de Bernoulli où l'on répète  $n$  fois une expérience pour laquelle la probabilité d'un succès vaut  $p$ ; la loi binomiale de paramètre  $n$  et  $p$ ,  $\mathcal{B}(n, p)$ , est la loi de probabilité correspondante. On note  $X \sim \mathcal{B}(n, p)$ .

**Proposition 3.1** La variable binomiale  $X \sim \mathcal{B}(n, p)$  est une variable aléatoire discrète prenant les valeurs  $\{0, 1, 2, \dots, n\}$  et la fonction de probabilité de  $X$  est donnée par

$$\begin{aligned} p^X(k) &= P["X = k"] = \text{probabilité d'avoir } k \text{ succès dans les } n \text{ essais} \\ &= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \text{ pour } 0 \leq k \leq n, k \text{ entier (rappel: } 0! = 1) \end{aligned} \tag{3.8}$$

En effet, la probabilité d'une séquence d'expériences qui comporte  $k$  succès et  $(n - k)$  échecs est de  $p^k(1 - p)^{n-k}$ ; le nombre  $C_n^k = \frac{n!}{k!(n-k)!}$  est le nombre de séquences différentes de  $n$  expériences qui comportent  $k$  succès et  $(n - k)$  échecs.

**Exemple 3.4** Quelle est la probabilité d'avoir deux fois un 5 en lançant 4 fois un dé?

$$\begin{cases} p = \text{avoir un 5 en lançant un dé} = 1/6 \\ 1 - p = 5/6 \end{cases}$$

$X$  mesure le nombre de fois qu'on obtient 5 dans  $n = 4$  lancers de dé

$$P["X = 2"] = \frac{4!}{2!2!}(1/6)^2(5/6)^2 = \frac{2 \times 25}{12 \times 36} = \frac{25}{216}$$

**Exemple 3.5** Quelle est la probabilité d'avoir cinq fois pile en lançant 10 fois une pièce non truquée?

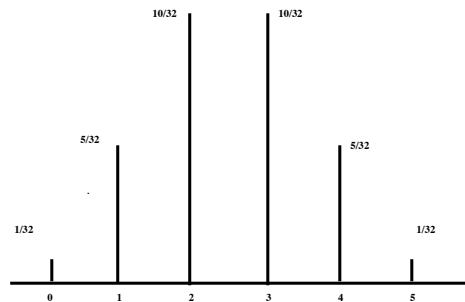
$p$ =avoir pile en lançant la pièce= $1/2$

$X$  mesure le nombre de fois qu'on obtient pile dans  $n = 10$  essais.

$$P["X = 5"] = \frac{10!}{5!5!}(1/2)^5(1/2)^5 = \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{5 \times 4 \times 3 \times 2 \times 1 \times 5 \times 4 \times 3 \times 2 \times 1} \times \frac{1}{2^{10}}$$

$$= \frac{9 \times 7}{2^8} = \frac{63}{256} \approx 1/4$$

**Diagramme en bâtonnets pour  $\mathcal{B}(5, 1/2)$**

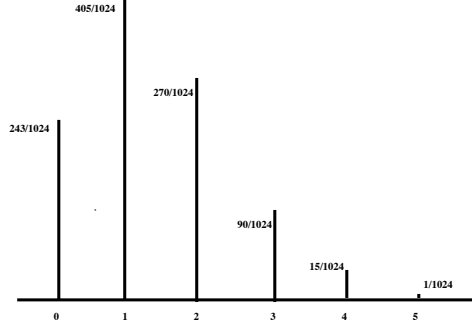


$$p^X(1) = P["X = 1"] = \frac{5!}{4!}(1/2)^5 = \frac{5}{32} = p^X(4)$$

$$p^X(2) = P["X = 2"] = \frac{5!}{3!2!}(1/2)^5 = \frac{10}{32} = p^X(3)$$

$$p^X(0) = P["X = 0"] = \frac{1}{32} = p^X(5)$$

### Diagramme en bâtonnets pour $\mathcal{B}(5, 1/4)$



$$\begin{aligned}
 p^X(0) &= \left(\frac{3}{4}\right)^5 = \frac{243}{1024} & p^X(1) &= 5 \cdot \frac{1}{4} \cdot \left(\frac{3}{4}\right)^4 = \frac{405}{1024} & p^X(2) &= 10 \cdot \frac{27}{4^5} = \frac{270}{1024} \\
 p^X(3) &= 10 \cdot \frac{9}{4^5} = \frac{90}{1024} & p^X(4) &= \frac{5 \cdot 3}{4^5} = \frac{15}{1024} & p^X(5) &= \frac{1}{4^5} = \frac{1}{1024}
 \end{aligned}$$

### 3.1.2 La loi de Poisson

**Définition 3.8** Une variable aléatoire  $X$  suit une loi de Poisson de paramètre  $\lambda$ , où  $\lambda$  est un nombre réel  $> 0$ , si  $X$  ne prend que des valeurs entières  $\geq 0$  et si

$$p^X(n) = e^{-\lambda} \frac{\lambda^n}{n!} \quad \text{pour tout entier } n \geq 0. \quad (3.9)$$

**Remarque 3.5** Considérons un événement qui se produit plusieurs fois à des instants aléatoires après l'instant  $t = 0$  (par exemple l'émission d'une particule  $\alpha$  par un corps radioactif). Soit  $N(t)$  le nombre d'événements dans l'intervalle  $[0, t]$ , et  $N(0) = 0$ . On dit que ces événements constituent un processus de Poisson lorsque:

- 1) les nombres d'événements se produisant dans des intervalles de temps disjoints sont indépendants;
- 2) la loi de probabilité de  $N(t + \Delta t) - N(t)$  ne dépend que de  $\Delta t$  (et pas de  $t$ );
- 3)  $\lim_{\Delta t \rightarrow 0} \frac{P[N(\Delta t)=1]}{\Delta t} = \lambda > 0$
- 4)  $\lim_{\Delta t \rightarrow 0} \frac{P[N(\Delta t)=2]}{\Delta t} = 0$ .

Alors on peut montrer (on ne le fera pas ici!), que pour un processus de Poisson, la loi de probabilité de la variable aléatoire  $N(t)$  est une loi de Poisson de paramètre  $\lambda t$ .

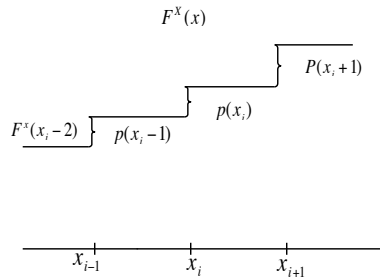
### 3.1.3 Fonction de répartition

Si  $X$  est une variable aléatoire discrète, prenant les valeurs  $x_1, \dots, x_n, \dots$  avec les probabilités  $p^X(x_1), \dots, p^X(x_n), \dots$ , alors sa fonction de répartition vaut,

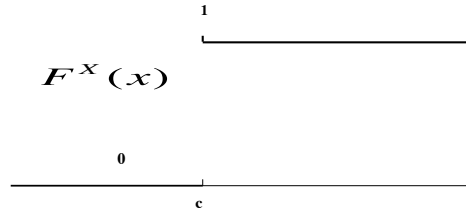
$$F^X(x) = P[“X \leq x”] = \sum_{x_i \leq x} p(x_i). \quad (3.10)$$

On obtient ainsi une fonction “en escalier”.

En rangeant les  $x_1 < x_2 < x_3 \dots x_n < x_{n+1} \dots$  de manière croissante, la fonction  $F^X$  est constante et vaut  $\sum_{j=1}^k p^X(x_j)$  pour  $x \in [x_k, x_{k+1}[$ .



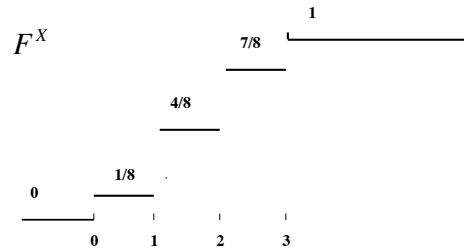
**Exemple 3.6** Considérons la loi de Dirac, c’est à dire celle d’une variable aléatoire  $X$  qui ne prend qu’une seule valeur  $c \in \mathbb{R}$ . On a donc  $p^X(c) = 1$  et  $p^X(x) = 0 \forall x \neq c$  donc,



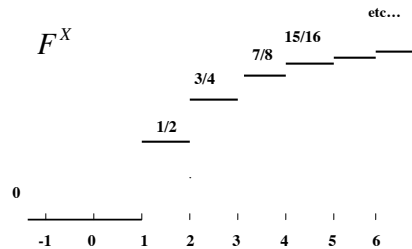
**Exemple 3.7** Si  $X$  est le nombre de faces obtenues en jetant trois pièces (non truquées!), on a,

$$p^X(0) = 1/8 = p^X(3) \quad p^X(1) = 3/8 = p^X(2)$$

donc



**Exemple 3.8** Si  $X$  est le nombre de fois qu'on doit lancer une pièce non truquée pour obtenir un pile pour la première fois, on a  $p^X(n) = \frac{1}{2^n}$  et donc,



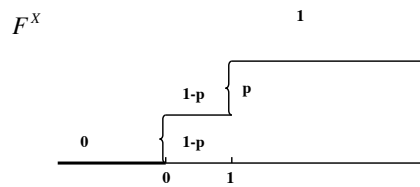
**Exemple 3.9** Pour une expérience de Bernoulli, ayant 2 résultats possibles ( $S$  avec la probabilité  $p$  et  $F$  avec la probabilité  $1 - p$ ), si  $X$  est le nombre de succès ( $S$ ) après

$$\left( \begin{array}{c} \\ \end{array} \right)$$

une expérience, on a

$$p^X(0) = 1 - p \quad p^X(1) = p \quad \text{et} \quad p^X(x) = 0 \quad \forall x \notin \{0, 1\}$$

On a,



### 3.1.4 Espérance mathématique-Variance

Soit  $X$  une variable aléatoire discrète prenant les valeurs  $\{x_1, x_2, \dots, x_n, \dots\}$  avec les probabilités  $p^X(x_1), p^X(x_2), \dots, p^X(x_n) \dots$

**Définition 3.9** L'espérance de  $X$ ,  $E[X]$  est donnée par

$$E[X] := \sum_i x_i p^X(x_i) \quad (\text{On la note aussi } \mu \text{ ou } \mu_X) \quad (3.11)$$

quand cette série est absolument convergente !

C'est donc la moyenne pondérée des valeurs prises par  $X$  avec les poids de pondération définis par les probabilités. On l'appelle parfois aussi la moyenne de la variable aléatoire  $X$ .

**Exemple 3.10** Considérons un jeu où l'on jette 2 dés (non truqués); il faut payer 1 euro pour jouer et l'on reçoit 4 euros si le total des dés vaut 5. Soit  $X$  la variable aléatoire définie par le gain. Alors  $X$  prend les valeurs  $-1$  (si le total des dés ne vaut pas 5) et  $3$  (si le total des dés vaut 5).

L'espérance de gain est donnée par

$$\begin{aligned} E[X] &= -1.p^X(-1) + 3p^X(3) \\ &= -1.P[\text{la somme des 2 dés ne vaut pas 5}] \\ &\quad + 3P[\text{la somme des 2 dés vaut 5}] \\ &= -1.\frac{8}{9} + 3\frac{1}{9} = \frac{-5}{9}!!! \end{aligned}$$

A partir de quelle somme reçue si le total des dés vaut 5 le jeu vaut-il la peine d'être joué?

**Définition 3.10** La variance  $Var[X]$ , également notée  $\sigma_X^2$  ou  $\sigma^2$  d'une variable aléatoire discrète  $X$  prenant les valeurs  $\{x_1, x_2, \dots, x_n, \dots\}$  avec les probabilités  $p^X(x_1), p^X(x_2), \dots, p^X(x_n) \dots$  est

$$Var[X] := \sum_i (x_i - \mu)^2 p^X(x_i) \quad (3.12)$$

et l'écart-type (ou déviations standard), noté  $\sigma_X$  ou  $\sigma$  est la racine carrée de  $\sigma_X^2$ .

**Exemple 3.11** Soit  $X$  une variable aléatoire binomiale de paramètre  $n$  et  $p$ ;  $X$  prend donc des valeurs entières  $0 \leq k \leq n$  avec les probabilités  $p^X(k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$ . Alors on a :



$$\begin{aligned}
E(X) &= \sum_{k=0}^n \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} k = \sum_{k=1}^n \frac{kn!p^k(1-p)^{n-k}}{k!(n-k)!} \\
&= np \sum_{k=1}^n \frac{(n-1)!p^{k-1}(1-p)^{(n-1)-(k-1)}}{(k-1)!((n-1)-(k-1))!} \\
&= np \sum_{k'=0}^{n-1} \frac{(n-1)!p^{k'}(1-p)^{(n-1-k')}}{k'!(n-1-k')!} = np(p + (1-p))^{n-1} \\
&= np
\end{aligned} \tag{3.13}$$

Donc l'espérance - ou la moyenne - d'une variable binomale de loi  $\mathcal{B}(n, p)$  vaut  $np$ .

$$\begin{aligned}
\sigma_X^2 &= \sum_{k=0}^n p^X(k)(k - np)^2 = \sum_{k=0}^n \frac{n!}{k!(n-k)!} (k^2 - 2knp + n^2p^2) p^k (1-p)^{n-k} \\
&= \sum_{k=0}^n \frac{k^2 n! p^k (1-p)^{n-k}}{k!(n-k)!} - 2np \sum_{k=0}^n \frac{kn!p^k(1-p)^{n-k}}{k!(n-k)!} + n^2p^2 \\
&= \sum_{k=1}^n \frac{k n! p^k (1-p)^{n-k}}{(k-1)!(n-k)!} - n^2 p^2 \\
&= \sum_{k=2}^n \frac{(k-1)n!p^k(1-p)^{n-k}}{(k-1)!(n-k)!} + \sum_{k=1}^n \frac{n!p^k(1-p)^{n-k}}{(k-1)!(n-k)!} - n^2p^2 \\
&= \sum_{k'=0}^{n-2} \frac{(n-2)!p^{k'}(1-p)^{(n-2-k')}}{k'!(n-2-k')!} n(n-1)p^2 \\
&\quad + pn \sum_{k'=0}^{n-1} \frac{(n-1)!p^{k'}(1-p)^{(n-1-k')}}{k'!(n-1-k')!} - n^2p^2 \\
&= p^2 n(n-1 - n + pn) = np(1-p).
\end{aligned} \tag{3.14}$$

Donc la variance d'une variable binomale de loi  $\mathcal{B}(n, p)$  vaut  $np(1-p)$  et son écart-type  $\sqrt{n}\sqrt{p(1-p)}$ . Par exemple, si on jette 46 fois une pièce non truquée, l'espérance (=la moyenne) du nombre de piles sera de 23 !!!

**Exemple 3.12** Si on considère une variable aléatoire  $X$  suivant une loi de Poisson de paramètre  $\lambda$ ; alors  $X$  prend des valeurs entières  $n \geq 0$  avec les

probabilités  $p^X(n) = e^{-\lambda} \frac{\lambda^n}{n!}$ . Dès lors,

$$E(X) = \sum_{n \geq 0} n e^{-\lambda} \frac{\lambda^n}{n!} = \lambda e^{-\lambda} \sum_{n \geq 1} \frac{\lambda^{n-1}}{(n-1)!} = \lambda \quad (3.15)$$

donc l'espérance de  $X$  vaut  $\lambda$  et on a

$$\begin{aligned} \sigma_X^2 &= \sum_{n \geq 0} \frac{e^{-\lambda} \lambda^n}{n!} (n - \lambda)^2 = \sum_{n \geq 1} \frac{e^{-\lambda} n \lambda^n}{(n-1)!} - \lambda^2 \\ &= \lambda^2 \sum_{n \geq 2} \frac{e^{-\lambda} \lambda^{n-2}}{(n-2)!} + \sum_{n \geq 1} \frac{e^{-\lambda} \lambda^n}{(n-1)!} - \lambda^2 \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda \end{aligned}$$

Donc la variance de  $X$  vaut  $\lambda$  et son écart-type  $\sigma_X = \sqrt{\lambda}$ .

### 3.1.5 Fonction d'une variable aléatoire

**Proposition 3.2** *Soit  $g : \mathbb{R} \rightarrow \mathbb{R}$  une fonction "mesurable" (par exemple une fonction continue) et considérons  $Y = g(X)$ . Donc,*

$$Y : \Omega \rightarrow \mathbb{R} \quad \omega \rightarrow Y(\omega) := g(X(\omega)) \quad \forall \omega \in \Omega. \quad (3.16)$$

*Si  $X$  est une variable aléatoire discrète prenant les valeurs  $\{x_1, x_2, \dots\}$  avec les probabilités  $\{p^X(x_1), p^X(x_2), \dots\}$  alors  $Y$  est aussi une variable aléatoire discrète prenant les valeurs (non nécessairement distinctes !)  $\{g(x_1), g(x_2), \dots\}$ . L'espérance de  $Y$  vaut*

$$E[y] = \sum_i g(x_i) p^X(x_i) \quad (3.17)$$

*quand cette série est absolument convergente.*

En effet  $Y$  prend les valeurs  $\{y_1, y_2, \dots\} = \{g(x_1), g(x_2), \dots\}$  et

$$p^Y(y_j) := P^Y(\{y_j\}) = \sum_{i \mid g(x_i) = y_j} P^X(\{x_i\})$$

; donc

$$y_j p^Y(y_j) = \sum_{i \mid g(x_i) = y_j} g(x_i) p^X(x_i)$$

et

$$E[Y] = \sum_j y^j p^Y(y_j) = \sum_i g(x_i) p^X(x_i).$$

En particulier la variance  $\sigma_X^2$  de  $X$  est égale à

$$\sigma_X^2 = E[(X - \mu)^2] \quad \text{où } \mu = E(X)$$

**Proposition 3.3** *La variance d'une variable aléatoire discrète  $X$  est égale à*

$$\sigma_X^2 = E[X^2] - \mu^2 \quad (3.18)$$

où  $\mu = E(X)$ .

$$\begin{aligned} \text{Dém: } \sigma_X^2 &= \sum_i (x_i - \mu)^2 p(x_i) = \sum_i (x_i^2 - 2\mu x_i + \mu^2) p(x_i) \\ &= \sum_i x_i^2 p(x_i) - 2\mu \sum_i x_i p(x_i) + \mu^2 \sum_i p(x_i) \\ &= E(X^2) - 2\mu \cdot \mu + \mu^2 = E(X^2) - \mu^2 \end{aligned}$$

**Proposition 3.4** *Soient  $X$  et  $Y$  des variables aléatoires discrètes sur un espace probabilisé. Quand les espérances  $E[\cdot]$  existent, on a*

$$E[a + bX + cY] = a + bE[X] + cE[Y] \quad a, b, c \in \mathbb{R} \quad (3.19)$$

Si  $P[X = c] = 1$  où  $c \in \mathbb{R}$  alors  $E[X] = c$ .

On a toujours  $(E[X])^2 \leq E[X^2]$  car  $\sigma_X^2 \geq 0$  et  $\sigma_X^2 = E[X^2] - (E[X])^2$ .

**Définition 3.11** On appelle moment d'ordre  $k$  d'une variable aléatoire  $X$ , le nombre  $\mu'_k := E[X^k]$  (quand cela existe) et moment centré d'ordre  $k$  de  $X$   $\mu_k := E[(X - \mu'_1)^k]$ ;  $\mu'_1 = \mu$  est la moyenne et  $\mu_2 = \sigma_X^2$  est la variance.

## 3.2 Variables aléatoires (absolument) continues

**Définition 3.12** Soit  $X$  une variable aléatoire définie sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ :

$$X : \Omega \rightarrow \mathbb{R}$$

et soit  $F^X$  la fonction de répartition associée

$$F^X : \mathbb{R} \rightarrow \mathbb{R} \quad x \rightarrow F^X(x) = P[X \leq x] = P^X([-\infty, x]).$$

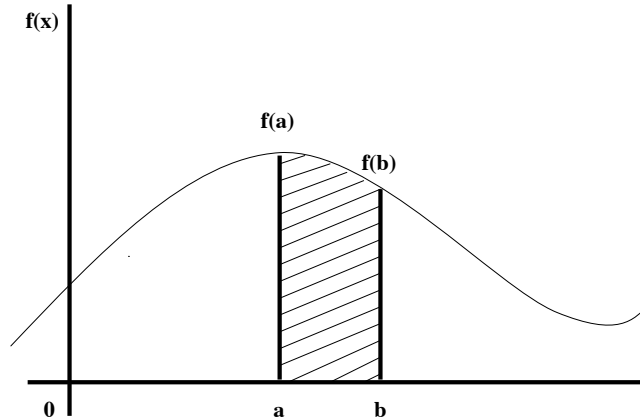
On dit que  $X$  est une variable aléatoire continue s'il existe une fonction  $f : \mathbb{R} \rightarrow \mathbb{R} \quad x \rightarrow f(x)$ , non négative, telle que

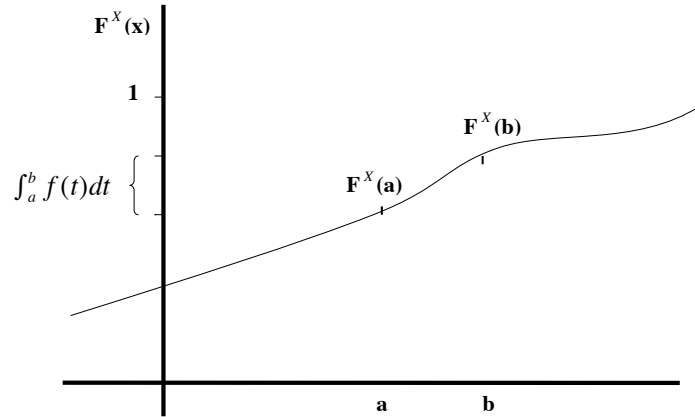
$$F^X(x) = \int_{-\infty}^x f(t)dt \quad \forall x \in \mathbb{R} \quad (3.20)$$

La fonction  $f$  est appelée densité de probabilité de  $X$ .

### Propriétés

- 1)  $P[a < X \leq b] = F^X(b) - F^X(a) = \int_a^b f(t)dt;$
- 2)  $\int_{-\infty}^{\infty} f(t)dt = 1;$





3) Presque partout  $F^X$  est dérivable et, presque partout

$$F'(x) := \frac{dF}{dx}(x) = f(x) = \lim_{\Delta x \rightarrow 0} \frac{P[x < X \leq x + \Delta x]}{\Delta x};$$

$$4) P[X = a] = \lim_{\varepsilon \rightarrow 0} \left( P[a - \varepsilon/2 < X \leq a + \varepsilon/2] = \int_{a-\varepsilon/2}^{a+\varepsilon/2} f(t)dt \approx \varepsilon f(a) \right) = 0.$$

**Définition 3.13** Soit  $X$  une variable aléatoire continue.

On définit l'espérance de  $X$ ,  $E[X]$ , appelée aussi moyenne et notée aussi  $\mu$  ou pour plus de précision  $\mu_X$ :

$$\mu := E[X] = \int_{-\infty}^{\infty} x f(x) dx \quad (3.21)$$

quand l'intégrale converge absolument!

On définit la variance de  $X$ , notée aussi  $\sigma^2$  ou  $\sigma_X^2$ :

$$\sigma^2 := Var[X] := E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2, \quad (3.22)$$

et l'écart-type (ou déviation standard) de  $X$ ,

$$\sigma := \sigma_X := \sqrt{\sigma^2} \quad (3.23)$$

**Proposition 3.5** *Soit  $X$  une variable aléatoire continue. Soit  $g : \mathbb{R} \rightarrow \mathbb{R}$  une fonction mesurable (par exemple continue). Considérons la variable aléatoire  $Y = g(X)$ . Son espérance est égale à:*

$$E[Y] = \int_{-\infty}^{\infty} g(x)f(x)dx. \quad (3.24)$$

La variance d'une variable aléatoire discrète  $X$  est égale à

$$\sigma_X^2 = E[X^2] - (E(X))^2. \quad (3.25)$$

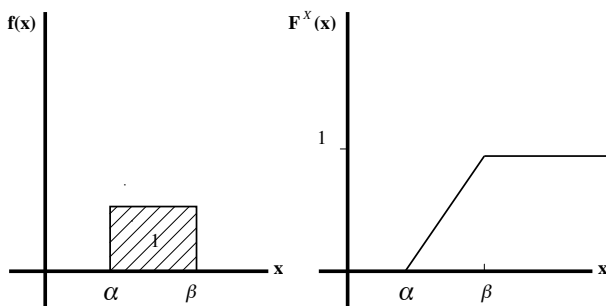
Soient  $X$  et  $Y$  des variables aléatoires continues sur un espace probabilisé. Quand les espérances  $E[\ ]$  existent, on a

$$E[a + bX + cY] = a + bE[X] + cE[Y] \quad a, b, c \in \mathbb{R} \quad (3.26)$$

**Exemple 3.13 (La loi uniforme)** On dit qu'une variable aléatoire  $X$  suit une loi de probabilité uniforme sur l'intervalle  $[\alpha, \beta]$  si c'est une variable aléatoire continue avec densité de probabilité donnée par

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{pour } \alpha \leq x \leq \beta \\ 0 & \text{ailleurs} \end{cases}$$

Le graphe de la fonction de répartition de  $X$  est donné par:

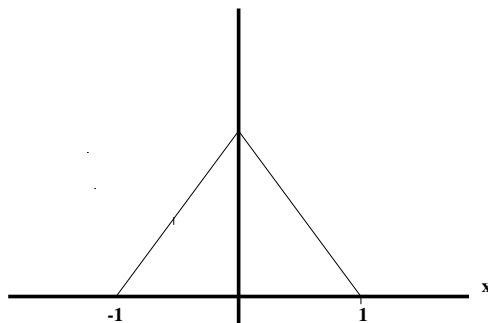


$$\mu = E[X] = \int_{-\infty}^{\infty} x f(x) dx = \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} x dx = \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} = 1/2(\alpha + \beta)$$

$$\begin{aligned} \sigma^2 = E[X^2] - \mu^2 &= \int_{-\infty}^{\infty} x^2 f(x) dx - 1/4(\alpha + \beta)^2 = \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} x^2 dx - 1/4(\alpha + \beta)^2 \\ &= \frac{1}{3} \frac{\beta^3 - \alpha^3}{\beta - \alpha} - \frac{1}{4}(\alpha^2 + 2\alpha\beta + \beta^2) = \frac{1}{12}(4\beta^2 + 4\beta\alpha + 4\alpha^2 - 3\alpha^2 - 6\alpha\beta - 3\beta^2) \\ &= \frac{\beta^2 - 2\alpha\beta + \alpha^2}{12} = \frac{(\beta - \alpha)^2}{12} \end{aligned}$$

**Exemple 3.14 ( La loi triangulaire )** La fonction densité de probabilité de la loi triangulaire est

$$f(x) = \begin{cases} x & \text{pour } 0 \leq x \leq 1 \\ 2 - x & \text{pour } 1 \leq x \leq 2 \\ 0 & \text{ailleurs} \end{cases}$$

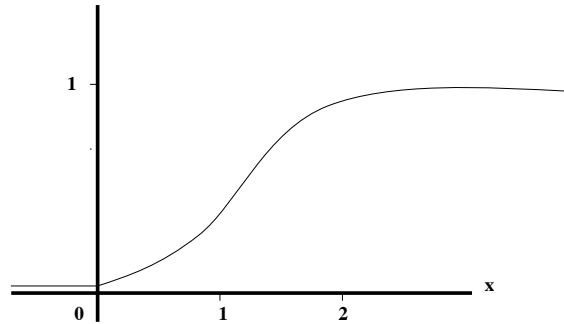


On a bien  $\int_{-\infty}^{\infty} f(t) dt = \int_0^1 t dt + \int_1^2 (2-t) dt = [\frac{t^2}{2}]_0^1 - [\frac{(2-t)^2}{2}]_1^2 = 1/2 + 1/2 = 1$

La fonction de répartition vaut:  $F(x) = \int_{-\infty}^x f(t) dt$  donc

$$F(x) = \begin{cases} 0 & \text{pour } x \leq 0 \\ \frac{x^2}{2} & \text{pour } 0 \leq x \leq 1 \\ -\frac{x^2}{2} + 2x - 1 & \text{pour } 1 \leq x \leq 2 \\ 1 & \text{pour } 2 \leq x \end{cases}$$

Graphique de F(x)



L'espérance ou moyenne est donnée par:

$$\begin{aligned}\mu &= \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 x^2 dx + \int_1^2 x(2-x) dx = \left[ \frac{x^3}{3} \right]_0^1 + \left[ x^2 - \frac{x^3}{3} \right]_1^2 \\ &= \frac{1}{3} + 4 - \frac{8}{3} - 1 + \frac{1}{3} = 1\end{aligned}$$

La variance est égale à

$$\begin{aligned}\sigma^2 &= E[X^2] - \mu^2 = \int_{-\infty}^{\infty} x^2 f(x) dx - 1 = \int_0^1 x^3 dx + \int_1^2 x^2(2-x) dx \\ &= \left[ \frac{x^4}{4} \right]_0^1 + \left[ \frac{2x^3}{3} - \frac{x^4}{4} \right]_1^2 \\ &= \frac{1}{4} + \frac{16}{3} - \frac{16}{4} - \frac{2}{3} + \frac{1}{4} - 1 = \frac{14}{3} - \frac{14}{4} - 1 = \frac{14}{12} - 1 = \frac{1}{6}\end{aligned}$$

**Exemple 3.15 (La loi de Cauchy)** Une variable aléatoire  $X$  suit une loi de Cauchy notée  $(\theta, c)$  si c'est une variable aléatoire continue avec densité de probabilité donnée par:

$$f(x) = \frac{c}{\pi} \frac{1}{c^2 + (x - \theta)^2} \quad \text{pour } c > 0 \text{ et } \theta \in \mathbb{R} \text{ fixés}$$



C'est bien une densité parce que

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)dx &= \frac{c}{\pi} \int_{-\infty}^{\infty} \frac{cdy}{c^2(1+y^2)} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{dy}{1+y^2} = \frac{2}{\pi} \int_0^{\infty} \frac{dy}{1+y^2} \text{ avec } y = \frac{x-\theta}{c} \\ &= \frac{2}{\pi} [\arctgy]_0^{\infty} = \frac{2}{\pi} (\pi/2 - 0) = 1\end{aligned}$$

! Il n'y a pas de moyenne. Pour  $c = 1$  et  $\theta = 0$ , on a

$$\int_{-\infty}^{\infty} xf(x)dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{xdx}{1+x^2} \text{ qui diverge !}$$

### 3.2.1 La loi normale standard

**Définition 3.14** Une variable aléatoire  $Z$  suit la loi normale standard (ou normale centrée réduite) notée  $\mathcal{N}(0, 1)$  si c'est une variable aléatoire continue avec densité de probabilité donnée par:

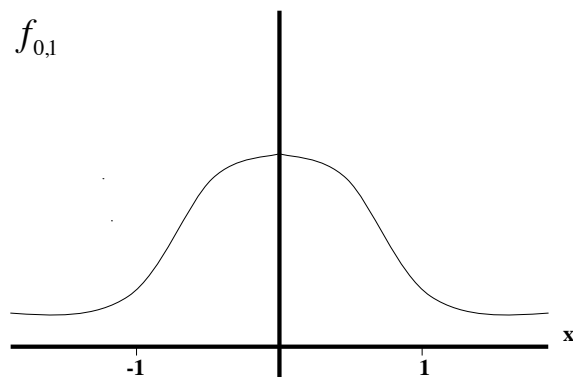
$$f_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-1/2 x^2}. \quad (3.27)$$

Vérifions que c'est une densité

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-1/2 x^2} dx > 0$$

$$\begin{aligned}\text{et } \left( \sqrt{2\pi} \int_{-\infty}^{\infty} e^{-1/2 x^2} dx \right)^2 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-1/2 x^2} dx \int_{-\infty}^{\infty} e^{-1/2 y^2} dy \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-1/2 (x^2+y^2)} dx dy = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-1/2 r^2} r dr d\theta \\ &= \frac{1}{2\pi} 2\pi [e^{-1/2 r^2}]_0^{\infty} = 1\end{aligned}$$

Le graphe de la densité de probabilité est



La moyenne est donnée par  $\mu = \int_{-\infty}^{\infty} x f_{0,1}(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-1/2 x^2} dx = 0$  par symétrie et car il y a convergence:

$$\frac{1}{\sqrt{2\pi}} \int_0^{\infty} x e^{-1/2 x^2} dx = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-u} du = \frac{1}{\sqrt{2\pi}} [-e^{-u}]_0^{\infty} = \frac{1}{\sqrt{2\pi}}$$

$$(u = \frac{1}{2}x^2)$$

La variance est donnée par

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{\infty} x^2 f_{0,1}(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-1/2 x^2} dx = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} x^2 e^{-1/2 x^2} dx \\ &= \frac{2}{\sqrt{2\pi}} ([-x \cdot e^{-1/2 x^2}]_0^{\infty} + \int_0^{\infty} e^{-1/2 x^2} dx) = 1 \end{aligned}$$

La fonction de répartition  $\phi(z) = \int_{-\infty}^z f_{0,1}(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-1/2 x^2} dx$  est donnée dans des tables.

### 3.2.2 La loi normale

**Définition 3.15** Une variable aléatoire  $X$  suit une loi normale  $\mathcal{N}(\mu, \sigma^2)$ , où  $\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_0^+$ , si c'est une variable aléatoire continue avec densité de

probabilité donnée par:

$$\begin{aligned} f_{\mu, \sigma^2}(x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\ &= \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} e^{-1/2(\frac{x-\mu}{\sigma})^2} \end{aligned} \quad (3.28)$$

On a bien  $\int_{-\infty}^{\infty} f_{\mu, \sigma^2}(x) dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-1/2 y^2} dy = 1$  avec  $y = \frac{x-\mu}{\sigma}$

La moyenne vaut

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{x}{\sigma} e^{-1/2(\frac{x-\mu}{\sigma})^2} dx &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{x-\mu}{\sigma} e^{-1/2(\frac{x-\mu}{\sigma})^2} dx + \mu \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \int_{-\infty}^{\infty} e^{-1/2(\frac{x-\mu}{\sigma})^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma y e^{-1/2 y^2} dy + \mu \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-1/2 y^2} dy = 0 + \mu = \mu \end{aligned}$$

La variance vaut

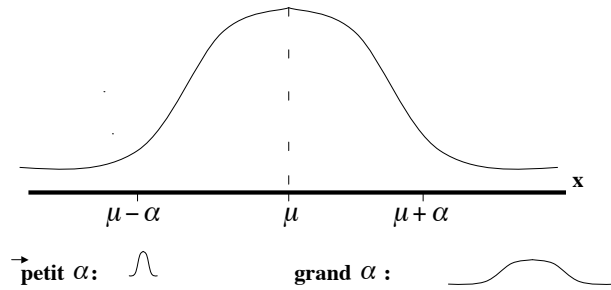
$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{\sigma} e^{-1/2(\frac{x-\mu}{\sigma})^2} dx &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(\frac{x-\mu}{\sigma}\right)^2 e^{-1/2(\frac{x-\mu}{\sigma})^2} d\left(\frac{x-\mu}{\sigma}\right) \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-1/2 y^2} dy = \sigma^2 \end{aligned}$$

et l'écart type vaut  $\sigma$ .

### Graphes de la densité de probabilité

$$\begin{aligned} F_{\mu, \sigma^2}(x) &= \int_{-\infty}^x f_{\mu, \sigma^2}(t) dt \\ &= \frac{1}{\sigma} \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-1/2(\frac{t-\mu}{\sigma})^2} dt \\ &= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-1/2 y^2} dy = F_{0,1}\left(\frac{x-\mu}{\sigma}\right) \end{aligned}$$

$f_{\mu,\alpha}$



**Théorème 3.1** Si  $X$  est une variable aléatoire de loi normale  $\mathcal{N}(\mu, \sigma^2)$  alors  $Z = \frac{X - \mu}{\sigma}$  est une variable aléatoire de loi normale standard  $\mathcal{N}(0, 1)$ .

Cette loi de probabilité est très souvent utilisée, par exemple pour décrire des erreurs de mesure, la taille d'un être humain, les résultats de tests,...

**Exemple 3.16** Le taux de respiration des tissus dans les diaphragmes de rats (mesuré en microlitre par mg de tissu sec et par heure) sous des conditions normales de température, est distribué suivant une loi normale et moyenne  $\mu = 2,03$  et écart-type  $\alpha = 0,44$ . Quelle est la probabilité d'obtenir chez un rat choisi au hasard un taux d'au moins 2,5 ?

$$P[X \geq 2,5] = P\left[\frac{X - 2,03}{0,44} \geq \frac{2,5 - 2,03}{0,44}\right] = P[Z \geq 1,07]$$

où  $Z$  est une variable aléatoire de loi normale standard; donc c'est donné par  $1 - \phi(1,07) = 0,1423$

La fonction de répartition de la loi normale standard  $\phi(z) = P[Z \leq z]$  est donné dans des tables!

## Propriété

Environ 68% des valeurs d'une variable aléatoire normale se trouvent dans l'intervalle  $(\mu - \sigma, \mu + \sigma)$ , 95% des valeurs dans l'intervalle  $(\mu - 2\sigma, \mu + 2\sigma)$  et 99,7% des valeurs dans  $(\mu - 3\sigma, \mu + 3\sigma)$ .

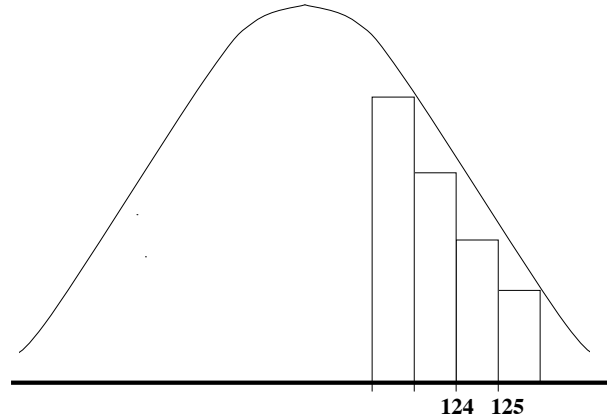
**Exemple 3.17** Supposons que le pH d'échantillons de sol prélevés dans une région donnée soit distribué suivant une loi normale de moyenne pH=6,00 et la déviation standard 0,1. Quelle est la probabilité que le pH d'un échantillon pris au hasard dans cette région soit compris entre 5,9 et 6,15 ?

$$\begin{aligned} P[5,9 \leq pH \leq 6,15] &= P\left[\frac{5,9 - 6}{0,1} \leq Z \leq \frac{6,15 - 6}{0,1}\right] \\ &= P[-1 \leq Z \leq 1,5] = \phi(1,5) - \phi(-1) \\ &= 0,9332 - 0,1587 = 0,7755 \end{aligned}$$

Quelle valeur  $A$  de pH ne sera dépassée que par 5% de toutes les mesures ? On cherche  $A$  tel que  $P(pH \leq A) = 0,95$  donc  $\frac{A-6}{0,1} = z$  avec  $\phi(z) = 0,95$  ce qui donne  $z = 1,645$  donc  $A = 6,1645$ .

La loi normale est souvent utilisée comme approximation d'une loi d'une variable discrète.

**Exemple 3.18** Les tests de QI d'une population européenne sont distribués de manière approximative suivant une loi normale de moyenne 100 et d'écart-type 15. Quelle est la probabilité qu'un individu sélectionné au hasard ait un QI  $\geq 125$ ? Attention: ici les tests de QI ne prennent que des valeurs entières; donc un QI  $\geq 125$  correspond à un QI  $\geq 124,5$ ; on approxime l'histogramme de la fonction de probabilité de QI avec des rectangles centrés sur les entiers par la courbe de la densité de la loi normale.



$$\begin{aligned} \text{Donc } P[QI \geq 125] &= P\left[\frac{QI - 100}{15} \geq \frac{124,5 - 100}{15}\right] = P[Z \geq 1,63] \\ &= 1 - \phi(1,63) = 1 - 0,9484 = 0,0516 \end{aligned}$$

Un grand nombre de méthodes de statistiques (inférentielle) sont basées sur l'hypothèse que la variable que l'on "échantillonne" suit une loi normale. Quand on a un grand échantillon, comment savoir s'il est raisonnable ou non de supposer avoir une loi normale ? On calcule la moyenne  $\bar{x}$  et l'écart-type  $S$  de l'échantillon constitué de  $n$  éléments ( $n$  grand!).

Si soit (le nombre d'observations dans  $(\bar{x} - S, \bar{x} + S) - 0,68n \geq 1,41\sqrt{n}$   
 soit (le nombre d'observations dans  $(\bar{x} - 2S, \bar{x} + 2S) - 0,95n \geq 0,654\sqrt{n}$   
 soit (le nombre d'observations dans  $(\bar{x} - 3S, \bar{x} + 3S) - 0,997n \geq 0,164\sqrt{n}$

alors l'hypothèse d'avoir une loi normale est douteuse.

**Exemple 3.19 ( Les lois gamma )** (larges familles de densités de probabilité non symétriques)

Pour  $\alpha > 0$  la fonction  $\Gamma(\alpha)$  est définie par

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \quad (3.29)$$

Elle a les propriétés suivantes:

- $\forall \alpha > 1, \Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$   
 $[car \int_0^\infty x^{\alpha-1} e^{-x} dx = [-e^{-x} x^{\alpha-1}]_0^\infty + (\alpha - 1) \int_0^\infty x^{\alpha-2} e^{-x} dx]$
- pour tout entier positif  $n, \Gamma(n) = (n - 1)!$   
 $[car \Gamma(1) = \int_0^\infty e^{-x} dx = [-e^{-x}]_0^\infty = 1]$
- $\Gamma(1/2) = \sqrt{\pi}$   
 $[car \Gamma(1/2) = \int_0^\infty x^{-1/2} e^{-x} dx = 2 \int_0^\infty e^{-y^2} dy = \frac{2}{\sqrt{2}} \int_0^\infty e^{-\frac{t^2}{2}} dt = \sqrt{2} \cdot \frac{\sqrt{2\pi}}{2}.$

**Définition 3.16** Une variable aléatoire continue  $X$  a une loi de probabilité gamma si la densité de probabilité de  $X$  est

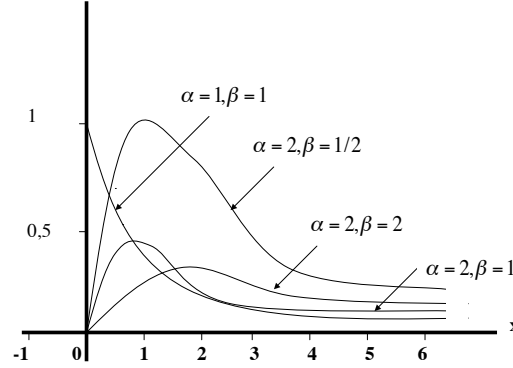
$$f(x; \alpha; \beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & \text{pour } x \geq 0 \\ 0 & \text{pour } x < 0 \end{cases}$$

où les paramètres  $\alpha$  et  $\beta$  sont des réels strictement positifs. Pour  $\beta = 1$ , on parle de loi de probabilité gamma standard!

On a bien

$$\int_{-\infty}^{\infty} f(t; \alpha, \beta) dt = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty t^{\alpha-1} e^{-t/\beta} dt = \frac{1}{\Gamma(\alpha)} \int_0^\infty (t')^{\alpha-1} e^{-t'} dt' = 1$$

## Quelques graphes de densité de probabilité



La moyenne et la variance sont données par:

$$E[X] = \mu = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty x^\alpha e^{-x/\beta} dx = \frac{\beta \Gamma(\alpha + 1)}{\Gamma(\alpha)} \int_0^\infty f(x; \alpha + 1, \beta) dx = \alpha \beta$$

$$\alpha^2 = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty x^{\alpha+1} e^{-x/\beta} dx - (\alpha \beta)^2 = \frac{\beta^2 \Gamma(\alpha + 2)}{\Gamma(\alpha)} \int_0^\infty f(x; \alpha + 2, \beta) dx - \alpha^2 \beta^2$$

$$= \beta^2 \alpha (\alpha + 1 - \alpha) = \beta^2 \alpha$$

### **Exemple 3.20 (La loi de Chi-Carré)**

**Définition 3.17** Une variable aléatoire continue  $X$  a une loi de probabilité Chi-Carré de paramètre  $\nu$  (où  $\nu$  est un entier positif) si la densité de probabilité est une densité gamma  $f(x; \alpha; \beta)$  pour  $\alpha = \nu/2$  et  $\beta = 2$  donc si la densité de probabilité est de la forme:

$$f(x, \nu) = \begin{cases} \frac{1}{2^{\nu/2} \Gamma(\nu/2)} x^{\nu/2-1} e^{-\frac{x}{2}} & \text{pour } x \geq 0 \\ 0 & \text{pour } x < 0 \end{cases}$$

Le paramètre  $\nu$  est appelé le nombre de degrés de liberté de  $X$ . On note cette loi: la loi  $\mathcal{X}_\nu^2$ .



**Remarque 3.6** Nous verrons que cette loi est importante car elle est à la base de nombreuses méthodes en statistiques. La moyenne d'une variable  $X$  de loi  $\mathcal{X}_\nu^2$  de paramètre  $\nu$  est égale à  $\nu$  et sa variance est égale à  $2\nu$ .

Lien avec la distribution normale

Si  $X$  a une loi de distribution normale standard, alors  $X^2$  a une loi  $\mathcal{X}_1^2$  à  $\nu = 1$  degré de liberté. En effet,

$$P[X^2 \leq y] = P[-\sqrt{y} \leq X \leq \sqrt{y}] \quad \forall y \geq 0.$$

Si  $f$  est la densité de probabilité de  $X$  et  $g$  celle de  $X^2$ , on a donc:

$$g(x) = 0 \text{ si } x < 0; \text{ si } y > 0 \int_{-\infty}^y g(t)dt = \int_{-\sqrt{y}}^{\sqrt{y}} f(t)dt$$

donc en dérivant

$$\begin{aligned} g(y) &= \frac{d}{dy} \left( \int_{-\sqrt{y}}^{\sqrt{y}} f(t)dt \right) = f(\sqrt{y}) \frac{1}{2\sqrt{y}} - f(-\sqrt{y}) \left( \frac{-1}{2\sqrt{y}} \right) \\ &= \left( \frac{2}{2\sqrt{\pi}} e^{-\frac{1}{2}y} \right) \frac{1}{2\sqrt{y}} = \frac{y^{-1/2} e^{-y/2}}{\sqrt{2}\sqrt{\pi}} = \frac{1}{2^{1/2}\Gamma(1/2)} y^{1/2-1} e^{-y/2} \end{aligned}$$

**Proposition 3.6** Si  $X_1, \dots, X_n$  sont  $n$  variables aléatoires indépendantes identiquement distribuées selon une loi normale standard, alors la variable aléatoire  $Y$  définie par la somme des carrés des  $X_i$  a une loi de probabilité Chi-Carré de paramètre  $n$  :

$$Y = \sum_{i=1}^n X_1^2 + \dots X_N^2 \quad \sim \quad \chi_n^2.$$

# Chapitre 4

## Distribution jointe et corrélation de deux variables aléatoires

Dans le chapitre précédent, nous avons regardé des modèles de loi de probabilité d'une variable aléatoire. De nombreux problèmes en probabilités et statistiques requièrent la considération simultanée de plusieurs variables aléatoires.

**Définition 4.1** Un vecteur aléatoire  $(X_1, \dots, X_k)$  défini sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  est la donnée de  $k$  variables aléatoires définies sur  $(\Omega, \mathcal{A}, P)$ .

Nous allons étudier ici le cas ( $k = 2$ ) de deux variables aléatoires.

**Définition 4.2** Soient  $X$  et  $Y$  deux variables aléatoires définies sur l'espace fondamental  $\Omega$  d'une expérience. On appelle fonction de répartition jointe du vecteur aléatoire  $(X, Y)$  la fonction  $F^{XY}$  définie par

$$F^{XY}(x, y) := P["X \leq x \text{ et } Y \leq y"] \quad (4.1)$$

**Définition 4.3** Si  $X$  et  $Y$  sont deux variables discrètes,  $X$  prenant des valeurs  $x_1, x_2, \dots, x_i, \dots$  et  $Y$  prenant des valeurs  $y_1, y_2, \dots, y_j, \dots$  on définit la fonction de probabilité jointe:

$$p^{XY}(x_i, y_j) := P["X = x_i \text{ et } Y = y_j"] \quad (4.2)$$

Pour tout sous-ensemble “mesurable”  $A$  de  $\mathbb{R} \times \mathbb{R}$ , (par exemple un rectangle) la probabilité que le couple  $(X, Y)$  prenne une valeur dans  $A$  est obtenue en sommant les valeurs de la fonction de probabilité jointe sur les paires  $(x_i, y_j)$  dans  $A$ :

$$P[(X, Y) \in A] = \sum_{(x_i, y_j) \in A} p^{XY}(x_i, y_j)$$

Les lois marginales de  $X$  et  $Y$  sont définies par:

$$P[“X = x_i”] = p^X(x_i) = \sum_{y_j} p^{XY}(x_i, y_j) \quad (4.3)$$

$$P[“Y = y_j”] = p^Y(y_j) = \sum_{x_i} p^{XY}(x_i, y_j) \quad (4.4)$$

Remarquons que  $\sum_{i,j} p^{XY}(x_i, y_j) = 1$ .

**Définition 4.4**  $X$  et  $Y$  sont deux variables aléatoires conjointement continues, s’il existe une fonction  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  non négative telle que

$$F^{X,Y}(x, y) (= P[“X \leq x \text{ et } Y \leq y”]) = \int_{-\infty}^x \left( \int_{-\infty}^y f(u, v) dv \right) du. \quad (4.5)$$

La fonction  $f(x, y)$  est appelée densité de probabilité jointe. On a

$$\frac{\partial^2 F}{\partial x \partial y} = f(x, y),$$

et si  $D$  est un domaine du plan, la probabilité que  $(X, Y)$  prenne des valeurs dans  $D$  est:

$$P[(X, Y) \in D] = \int \int_D f(x, y) dx dy$$

La loi marginale de  $X$  est déterminée par sa fonction de répartition

$$F^X(x) = P[“X \leq x”] = \int_{-\infty}^x \left( \int_{-\infty}^{\infty} f(u, v) dv \right) du$$

donc la densité de probabilité pour la loi marginale de la variable  $X$  est donnée par

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy. \quad (4.6)$$

De même, la loi marginale de  $Y$  est déterminée par la fonction de répartition

$$F^Y(y) = P[Y \leq y] = \int_{-\infty}^{\infty} \left( \int_{-\infty}^y f(u, v) dv \right) du = \int_{-\infty}^y \left( \int_{-\infty}^{\infty} f(u, v) du \right) dv$$

donc la densité de probabilité pour la loi marginale de la variable  $Y$  est donnée par

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx \quad (4.7)$$

**Exemple 4.1** Une compagnie d'assurances a un certain nombre de clients qui possèdent chez elle une assurance-auto et une assurance-maison (incendie). La franchise doit être spécifiée pour chacune de ces assurances avec un choix de 200 euros ou 500 euros pour l'assurance-auto et 0 euro, 250 euros ou 500 euros pour l'assurance-maison. Soit  $X$  la franchise sur l'assurance-auto et  $Y$  celle sur l'assurance-maison. Les couples de valeurs possibles pour  $(X, Y)$  sont (200,0) (200,250) (200,500) (500,0) (500,250) (500,500).

Supposons que la fonction de probabilité jointe soit donnée par la table:

|       |                | $y_j$ |      |     |
|-------|----------------|-------|------|-----|
|       |                | 0     | 250  | 500 |
| $x_i$ | $p^{XY}(x, y)$ |       |      |     |
|       |                | 0     | 250  | 500 |
|       | 200            | 0,2   | 0,1  | 0,2 |
|       | 500            | 0,05  | 0,15 | 0,3 |

(On a bien  $\sum_{i,j} p^{XY}(x_i, y_j) = 1!$ )

(Ceci veut dire que  $P[X = 200 \text{ et } Y = 500] = 0,2$  etc...)

Si l'on veut connaître

$$\begin{aligned} P[Y \geq 250] &= p^{XY}(200, 250) + p^{XY}(200, 500) + p^{XY}(500, 250) + p^{XY}(500, 500) \\ &= 0,1 + 0,2 + 0,15 + 0,3 = 0,75 \end{aligned}$$

La fonction de probabilité de  $X$  est donnée par

$$\begin{aligned} p_X(x) &= P[X = x] \\ p_X(200) &= 0,2 + 0,1 + 0,2 = 0,5 \\ p_X(500) &= 0,05 + 0,15 + 0,3 = 0,5 \end{aligned}$$

La fonction de probabilité de  $Y$  est donnée par

$$\begin{aligned}p_Y(0) &= 0,2 + 0,05 = 0,25 \\p_Y(250) &= 0,1 + 0,15 = 0,25 \\p_Y(500) &= 0,2 + 0,3 = 0,5\end{aligned}$$

**Exemple 4.2** Dans un restaurant fast-food il y a un guichet à l'intérieur et un guichet à l'extérieur (drive-in). Un jour pris au hasard, soit  $X$  la proportion du temps où le guichet extérieur est utilisé (au moins un client y est servi ou attend d'être servi) et soit  $Y$  la proportion du temps où le guichet intérieur est utilisé. Les valeurs possibles pour  $(X, Y)$  sont tous les points du rectangle  $D = \{(x, y) \mid 0 \leq x \leq 1, 0 \leq y \leq 1\}$ . Supposons que la densité de probabilité jointe soit donnée par

$$f(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & 0 \leq x \leq 1 \quad 0 \leq y \leq 1 \\ 0 & \text{ailleurs} \end{cases}$$

(On a bien  $\int_{-\infty}^{\infty} (\int_{-\infty}^{\infty} f(u, v) dv) du = \int_0^1 (\int_0^1 \frac{6}{5}(u + v^2) dv) du = \int_0^1 [\frac{6}{5}(uv + \frac{v^3}{3})]_0^1 du = \int_0^1 (\frac{6}{5}u + \frac{2}{5}) du = [\frac{3u^2}{5} + \frac{2u}{5}]_0^1 = 1$ )

La probabilité qu'aucun des deux guichets ne soit utilisé plus d'un quart d'heure de temps est donnée par:

$$\begin{aligned}P[0 \leq X < 1/4, 0 < Y \leq 1/4] &= \int_0^{1/4} \int_0^{1/4} \frac{6}{5}(x + y^2) dy dx \\&= \frac{6}{5} \int_0^{1/4} (\frac{x}{4} + \frac{1}{3 \times 64}) dx = \frac{6}{20} [\frac{x^2}{2}]_0^{1/4} + \frac{6}{20} \times \frac{1}{3 \times 64} \\&= \frac{6}{640} + \frac{1}{640} = \frac{7}{640} = 0,0109\end{aligned}$$

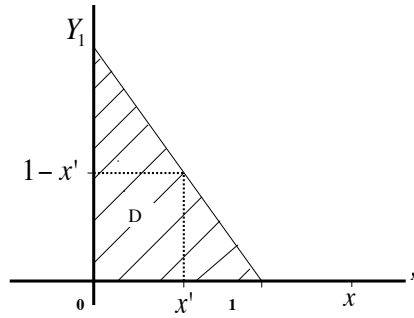
**Exemple 4.3** Quand on récolte des échantillons de roches dans une région, il y a 3 types de roches. Soit  $X$  la proportion de la première roche et  $Y$  la proportion de la deuxième roche; la connaissance de  $X$  et  $Y$  détermine bien entendu la proportion de la troisième roche! Les valeurs que peuvent prendre

$(X, Y)$  sont dans

$$D = \{(x, y) \mid 0 \leq x \leq 1, 0 \leq y \leq 1, x + y \leq 1\}$$

Supposons que la densité de probabilité jointe soit

$$f(x, y) = \begin{cases} 24xy & \text{si } (x, y) \in D \\ 0 & \text{ailleurs} \end{cases}$$



$$\begin{aligned} (\text{On a bien } \int_{-\infty}^{\infty} (\int_{-\infty}^{\infty} f(x, y) dy) dx &= \int_0^1 (\int_0^{1-x} 24(xy) dy) dx = \int_0^1 24 \left[ \frac{xy^2}{2} \right]_{y=0}^{y=1-x} dx \\ &= \int_0^1 24 \frac{x(1-x)^2}{2} dx = 12 \int_0^1 (x - 2x^2 + x^3) dx = 12 \left( \frac{1}{2} - \frac{2}{3} + \frac{1}{4} \right) = 1) \end{aligned}$$

La probabilité que les deux premiers types de roches fassent plus de 50% du volume est

$$\begin{aligned} P[X + Y > 0,5] &= 1 - P[X + Y \leq 0,5] \\ &= 1 - \int_0^{0,5} 24x \left( \int_0^{0,5-x} y dy \right) dx \\ &= 1 - 12 \int_0^{0,5} x \left( \frac{1}{2} - x \right)^2 dx \\ &= 1 - 12 \left[ \frac{x^2}{8} - \frac{x^3}{3} + \frac{x^4}{4} \right]_0^{0,5} \\ &= 1 + 3 \left( \frac{1}{8} - \frac{1}{6} + \frac{1}{16} \right) = \frac{15}{16} \end{aligned}$$

## 4.1 Espérance d'une fonction de deux variables aléatoires

Soient  $X$  et  $Y$  deux variables aléatoires (définies sur le même espace de probabilisé  $\Omega$ ). Soit  $Z = g(X, Y)$  où  $g$  est une fonction (mesurable) de  $\mathbb{R}^2$  dans  $\mathbb{R}$ . Alors  $Z$  est une variable aléatoire.

**Proposition 4.1** *Si  $X$  et  $Y$  sont deux variables aléatoires discrètes,  $X$  prenant des valeurs  $x_1, x_2, \dots$ ,  $Y$  des valeurs  $y_1, y_2, \dots$  avec la fonction de probabilité jointe  $p^{XY}(x_i, y_j) := P[“X = x_i \text{ et } Y = y_j”]$*

$$E(Z) = E(g(X, Y)) = \sum_{x_i} \sum_{y_j} g(x_i, y_j) p^{XY}(x_i, y_j). \quad (4.8)$$

*Quand  $X$  et  $Y$  sont deux variables aléatoires conjointement continues de densité de probabilité jointe  $f^{XY}$ , donc  $(P[“X \leq x \text{ et } Y \leq y”] = \int_{-\infty}^x (\int_{-\infty}^y f^{XY}(u, v) dv) du)$ , on a:*

$$E(Z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f^{XY}(x, y) dx dy. \quad (4.9)$$

## 4.2 Indépendance de deux variables aléatoires

**Définition 4.5** Deux variables aléatoires  $X$  et  $Y$  sont dites indépendantes ssi  $F^{XY}(x, y) = F^X(x) F^Y(y) \quad \forall x, y \in \mathbb{R}$ .

Si ce n'est pas le cas, on dit que  $X$  et  $Y$  sont dépendantes.

**Proposition 4.2** *Deux variables aléatoires  $X$  et  $Y$  sont indépendantes ssi*  
1) *quand  $X$  et  $Y$  sont discrètes,*

$$p^{XY}(x, y) = p^X(x) \cdot p^Y(y) \quad (4.10)$$

*pour toutes les paires  $(x, y)$  de valeurs de  $(X, Y)$ ;*

2) *quand  $X$  et  $Y$  sont conjointement continues,*

$$f^{XY}(x, y) = f_X(x) \cdot f_Y(y) \quad \forall (x, y) \quad (4.11)$$

**Exemple 4.4** Dans l'exemple plus haut de la compagnie d'assurance

$$p^{XY}(200, 0) = 0, 2 \quad p^X(200) = 0, 5 \quad p^Y(0) = 0, 25$$

donc  $X$  et  $Y$  sont dépendantes.

**Exemple 4.5** Dans l'exemple des échantillons de roches, soit  $(x, y) \in D$

$$f^{XY}(x, y) = 24 \, xy \quad f_X(x) = \int_0^{1-x} 24 \, xy \, dy = 12 \, x(1-x)^2$$

$$f_Y(y) = \int_0^{1-y} 24 \, xy \, dx = 12 \, y(1-y)^2$$

donc  $X$  et  $Y$  sont dépendantes.

### 4.3 Corrélation de deux variables aléatoires

Quand deux variables aléatoires ne sont pas indépendantes, il est intéressant de mesurer “combien elles sont liées”.

**Définition 4.6** La covariance entre deux variables aléatoires  $X$  et  $Y$ , notée  $cov(X, Y)$  est définie par

$$cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] \quad (4.12)$$

où  $\mu_X$  est la moyenne de la loi marginale de  $X$  et  $\mu_Y$  celle de  $Y$ .

**Proposition 4.3**

$$cov(X, Y) = E[XY] - \mu_X \mu_Y$$

**Définition 4.7** Le coefficient de corrélation de deux variables aléatoires  $X$  et  $Y$ , noté  $corr(X, Y)$  ou  $\rho_{X,Y}$  est défini par

$$\rho_{X,Y} = \frac{cov(X, Y)}{\sigma_X \sigma_Y} \quad (4.13)$$

où  $\sigma_X$  est l'écart-type de la loi marginale de  $X$  et  $\sigma_Y$  celui de  $Y$ .

**Proposition 4.4** Si  $X$  et  $Y$  sont indépendantes alors  $\rho_{X,Y} = 0$ . Mais attention, si  $\rho_{X,Y} = 0$  cela ne veut pas nécessairement dire que  $X$  et  $Y$  sont indépendantes.

**Dém:** ( $X, Y$  continues)

$$E[XY] = \int \int f^{XY}(x, y) xy \, dx \, dy = \int f_X(x) x \, dx \int f_Y(y) y \, dy = \mu_X \mu_Y.$$

**Proposition 4.5** Si  $X$  et  $Y$  sont indépendantes alors la variance de  $X + Y$  est donnée par  $Var(X + Y) = E[(X + Y)^2] - (\mu_X + \mu_Y)^2 = Var(X) + Var(Y)$ .

Dém:  $E[XY] = \mu_X \mu_Y$  puisque  $cov(X, Y) = 0$ .



**Exemple 4.6** Dans l'exemple de la compagnie d'assurance, la variable aléatoire  $X$  prend les valeurs 200 et 500 avec les probabilités  $p^X(200) = 0,5$  et  $p^X(500) = 0,5$  donc  $\mu_X = 200 \times 0,5 + 500 \times 0,5 = 350$  et l'écart-type  $\sigma_X = \sqrt{(150)^2 0,5 + (150)^2 0,5} = 150$ .

La variable aléatoire  $Y$  prend les valeurs 0,250 et 500 avec les probabilités  $p^Y(0) = 0,25$ ,  $p^Y(250) = 0,25$  et  $p^Y(500) = 0,5$ , donc  $\mu_Y = 250 \times 0,25 + 500 \times 0,5 = 312,5$  et

$\sigma_Y = \sqrt{(312,5)^2 \times \frac{1}{4} + (62,5)^2 \times \frac{1}{4} + (187,5)^2 \times \frac{1}{2}} \simeq 207$ . Dès lors,

$$\begin{aligned} Cov(X, Y) &= E[XY] - \mu_X \mu_Y \\ &= \sum_{ij} p(x_i, y_j) x_i y_j - \mu_X \mu_Y \\ &= 200 \times 250 \times 0,1 + 200 \times 500 \times 0,2 + 500 \times 250 \times 0,15 \\ &\quad + 500 \times 500 \times 0,3 - 350 \times 312,5 \\ &= 9.375 \end{aligned}$$

$$\text{donc } \rho_{XY} = \frac{9.375}{150 \times 207} \simeq 0,3$$

# Chapitre 5

## Quelques théorèmes importants

**Théorème 5.1** Si  $X$  et  $Y$  sont des variables aléatoires indépendantes de lois normales,  $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$  et  $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$  alors  $X + Y$  est normale de loi  $\mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$ .

(Sans démonstration ici)

**Théorème 5.2 (Inégalité de Bienaymé-Tchebychev)** Soit  $X$  une variable aléatoire d'espérance (moyenne)  $\mu$  et d'écart-type  $\sigma$ . alors pour tout  $k > 0$

$$P[|X - \mu| \geq k \sigma] \leq \frac{1}{k^2} \quad (5.1)$$

Démonstration: (dans le cas d'une variable aléatoire continue  $X$ , de densité  $f$ ):

$$\begin{aligned} \sigma^2 &= E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \\ &\geq \int_{-\infty}^{\mu - k\sigma} (x - \mu)^2 f(x) dx + \int_{\mu + k\sigma}^{\infty} (x - \mu)^2 f(x) dx \\ &\geq k^2 \sigma^2 \left[ \int_{-\infty}^{\mu - k\sigma} f(x) dx + \int_{\mu + k\sigma}^{\infty} f(x) dx \right] \\ &= k^2 \sigma^2 \{P[X \leq \mu - k\sigma] + P[X \geq \mu + k\sigma]\} \\ &= k^2 \sigma^2 (P[(X - \mu) \leq -k\sigma] + P[(X - \mu) \geq k\sigma]) \\ &= k^2 \sigma^2 P[|X - \mu| \geq k\sigma] \end{aligned}$$

donc  $P[|X - \mu| \geq k\sigma] \leq \frac{\sigma^2}{k^2 \sigma^2} = \frac{1}{k^2}$ .

**Théorème 5.3 (Loi faible des grands nombres)** (*Situation typique: on répète une même expérience aléatoire*)

Soient  $X_1, X_2, \dots$  une suite de variables aléatoires indépendantes, de même loi de moyenne  $\mu$  et d'écart-type  $\sigma$ . Alors,  $\forall \epsilon > 0$

$$P\left\{\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| \geq \epsilon\right\} \rightarrow 0 \text{ pour } n \rightarrow \infty. \quad (5.2)$$

Démonstration:

$$E\left[\frac{X_1 + \dots + X_n}{n}\right] = \frac{n\mu}{n} = \mu$$

$$Var\left(\frac{X_1 + \dots + X_n}{n}\right) = Var\left(\frac{X_1}{n}\right) + \dots + Var\left(\frac{X_n}{n}\right) = n\left(\frac{\sigma^2}{n^2}\right) = \frac{\sigma^2}{n}$$

Le théorème de Bienaymé-Tchebychev pour la variable aléatoire  $X = \frac{X_1 + \dots + X_n}{n}$  implique que

$$P\left[\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| \geq k \frac{\sigma}{\sqrt{n}}\right] \leq \frac{1}{k^2}$$

Donc

$$P\left[\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| \geq \epsilon\right] \leq \frac{\sigma^2}{n\epsilon^2} \rightarrow 0 \text{ pour } n \rightarrow \infty (\epsilon \text{ fixé})$$

En posant  $\epsilon = \frac{k\sigma}{\sqrt{n}}$  donc  $k = \frac{\epsilon\sqrt{n}}{\sigma}$  et  $\frac{1}{k^2} = \frac{\sigma^2}{\epsilon^2 n}$ .

**Théorème 5.4 (Théorème central limite)** (*sans démonstration*)

Si  $X_1, X_2, \dots$  est une suite de variables aléatoires indépendantes, de même loi de moyenne  $\mu$  et d'écart-type  $\sigma$  alors la loi de la variable aléatoire

$$Z_n = \frac{\left(\frac{X_1 + \dots + X_n}{n}\right) - \mu}{(\sigma/\sqrt{n})}$$

tend vers la loi normale standard pour  $n \rightarrow \infty$ , c'est à dire

$$P[Z_n \leq z] \rightarrow_{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{x^2}{2}} dx = \phi(z) \quad (5.3)$$

**Remarque 5.1**  $X_1, X_2, \dots, X_n$  sont des variables aléatoires indépendantes

- si la fonction de probabilité jointe est donnée par

$$P[X_1 = x_{i_1}, X_2 = x_{i_2}, \dots, X_n = x_{i_n}] = P[X_1 = x_{i_1}] \cdot P[X_2 = x_{i_2}] \cdot \dots \cdot P[X_n = x_{i_n}]$$

dans le cas de variables aléatoires discrètes;

- si la densité de probabilité jointe  $f^{X_1, X_2, \dots, X_n}(x_1, \dots, x_n)$  est donnée par

$$f^{X_1, X_2, \dots, X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdot \dots \cdot f_{X_n}(x_n)$$

dans le cas de variables aléatoires conjointement continues.

# **Partie II**

# **STATISTIQUE**

# Introduction

Considérons un certain phénomène (expérience) aléatoire et l'observation  $x$  d'une variable aléatoire réelle  $X$ .

La théorie des probabilités donne une description du comportement de  $x$  en spécifiant la loi de  $X$ , c'est à dire en donnant l'expression de la fonction de répartition  $F^X$  ou, ce qui revient au même, en donnant la fonction de probabilité  $x_i \rightarrow p^X(x_i)$  dans le cas d'une variable aléatoire discrète et la densité de probabilité  $f^X(x)$  dans le cas d'une variable aléatoire continue.

Dans la pratique, le modèle probabiliste décrivant le comportement de l'observation  $x$  n'est souvent que partiellement connu. Il faut alors faire appel, non plus à un modèle probabiliste particulier, mais à une famille de modèles probabilistes. Généralement on a une famille de lois permettant de décrire le comportement aléatoire de l'observation  $x$  qui est indexée par une famille de paramètres. (On parle de modèle statistique paramétrique).

**Exemples 5.1** 1. On injecte un sérum à un patient atteint d'une certaine maladie (ceci est l'expérience aléatoire). On s'intéresse à l'efficacité du sérum sur le patient (survie ou décès du malade). On décrit cette situation par l'observation  $x$  de la variable aléatoire

$$X = \begin{cases} 0 & \text{si le patient décède au bout du mois} \\ 1 & \text{si le patient survit} \end{cases}$$

La loi de  $X$  choisie est la loi binomiale  $\text{bin}(1, p)$  où  $p$  est la probabilité que le patient survive mais on ne connaît pas à priori  $p$  qui est donc un paramètre.

2. On étudie les punaises en laiton produites par la firme  $A$  et on s'intéresse à la proportion de punaises défectueuses. On décrit cette situation en

imaginant une "machine de Bernoulli" : chaque punaise produite est le résultat d'une expérience dans un schéma de Bernoulli avec une probabilité  $(1 - p)$  d'être bonne.

3. On fait un sondage avant un référendum pour savoir si le "oui" va l'emporter. L'expérience aléatoire consiste à "prélever" au hasard un électeur dans la population et à l'interroger sur ses intentions de vote (intention de voter OUI ou, au contraire, intention de voter NON ou de s'abstenir). On décrit cette situation par le modèle suivant de variable aléatoire  $X$ :

$$X = \begin{cases} 1 & \text{si intention de voter OUI} \\ 0 & \text{si intention de voter NON ou de s'abstenir} \end{cases}$$

où  $X$  a la distribution d'une loi binomiale  $Bin(1, p)$ ,  $p$  étant la proportion d'intentions de vote "OUI" parmi tous les électeurs.

4. On regarde l'expérience aléatoire (de Mendel) consistant à semer une graine provenant du croisement à la deuxième génération d'un pois jaune et ridé avec un pois vert et lisse. On observe la couleur des pois issus de cette graine. On décrit ceci par l'observation d'une variable aléatoire:

$$X = \begin{cases} 1 & \text{si les pois sont jaunes} \\ 0 & \text{si les pois sont verts} \end{cases}$$

avec la loi de  $X$  donnée par  $P["X = 1"] = \frac{3}{4}$   $P["X = 0"] = \frac{1}{4}$  donc  $X$  a pour loi la loi binomiale  $Bin(1, \frac{3}{4})$ .

5. On étudie le temps de vie (en heures d'écriture) d'une marque de stylos-billes. Le modèle est le suivant:
  - \*  $X$  mesure la durée de vie d'un stylo bille
  - \* On suppose que la distribution de  $X$  est donnée par une loi normale de moyenne  $\mu$  et d'écart-type  $\sigma$ .

Dans les exemples donnés ci-dessus, le modèle n'est souvent que partiellement connu car  $p$  n'est pas connu dans les modèles 1, 2 et 3 et  $\mu$  et  $\sigma$  ne sont pas connus dans le modèle 5. Ce sont donc des paramètres.

Cette seconde partie du cours va étudier différentes démarches permettant de

récolter de l'information sur la valeur du (ou des) paramètre(s)  $\theta$  correspondant à la loi  $P_\theta^X$  de la variable  $X$  correspondant à l'observation  $x$ . L'ensemble de ces techniques s'appelle l'inférence statistique. On va s'intéresser:

1. au problème de l'estimation de la valeur du (ou des) paramètre(s)  $\theta$ ;
2. au problème de la détermination d'un intervalle de confiance pour cette valeur (= une "fourchette" de valeur);
3. aux tests d'hypothèses permettant de décider si le paramètre  $\theta$  prend ou ne prend pas certaines valeurs.

### Exemples

Dans le modèle d'injection du sérum,  $p$  n'est pas connu. Des problèmes d'inférence statistique que l'on peut envisager sont :

- quelle est la probabilité  $p$  qu'un patient ayant reçu le sérum survive au bout d'un mois ?
- peut-on donner une "fourchette" pour  $p$  ?
- $p$  a-t-il une valeur supérieure à  $\frac{1}{2}$ ?

Le même type de questions peut se poser pour les exemples 2 et 3. Dans l'exemple 5,  $\mu$  et  $\sigma^2$  ne sont pas connus et on peut se demander quelles sont les valeurs de  $\mu$  et  $\sigma^2$ , peut-on réfuter les affirmations du fabricant selon lesquelles  $\mu \geq 15, \dots$ ?



# Chapitre 6

## Echantillons-Estimeurs

Comme on ne peut généralement pas connaître l'information sur tous les éléments d'une population (quel pourcentage de la population va voter OUI, quelle proportion de punaises produites sont défectueuses, quelle est la durée de vie de chaque stylo-bille...) On va étudier l'information donnée par un échantillon pris au hasard dans la population.

### 6.1 Echantillons aléatoires simples

Prendre un échantillon aléatoire simple dans une population est une procédure pour sélectionner un nombre  $n$  d'individus dans la population de manière telle que tous les choix de  $n$  individus soient également probables; que tout individu ait la même probabilité d'être choisi et que le choix d'un individu n'influence pas le choix des individus suivants...

Prendre un échantillon consiste ainsi (le plus souvent) à répéter une expérience aléatoire  $E$   $n$  fois, chaque répétition fournissant une observation.

1<sup>ère</sup> expérience  $\rightarrow$  observation  $x_1$  de la variable  $X_1$ ;

$n^{\text{ème}}$  expérience  $\rightarrow$  observation  $x_n$  de la variable  $X_n$ .

Les expériences étant identiques et indépendantes, les variables aléatoires  $X_i$  ( $i = 1, \dots, n$ ) sont indépendantes et identiquement distribuées; elles ont toutes la même fonction de probabilités  $y_j \rightarrow p_\theta^{X_i}(y_j)$  (cas discret) ou la même densité de probabilité  $x \rightarrow f_\theta^{X_i}(x)$  dépendant d'un (ou plusieurs) paramètre(s)  $\theta$  inconnu(s).

Ce type de procédure est appelé modèle d'échantillonnage simple.

\* Le n-uple de variables indépendantes et identiquement distribuées

$$(X_1, \dots, X_n)$$

est appelé échantillon aléatoire simple de taille  $n$ .

\* Le n-uple de valeurs observées  $(x_1, \dots, x_n)$  est appelé valeurs de l'échantillon de taille  $n$ .

\*  $\theta$  est le paramètre inconnu apparaissant dans la loi commune des  $X_i$ .

\* La loi commune des  $X_i$  est appelée la loi-population.

**Définition 6.1** On appelle statistique toute fonction  $S$  des variables aléatoires  $(X_1, \dots, X_n)$  constituant l'échantillon, qui ne dépend pas explicitement du paramètre inconnu; en d'autres termes la valeur  $s$  de cette fonction peut être calculée sur base des observations  $(x_1, \dots, x_n)$  uniquement.

Remarquons qu'une statistique est une variable aléatoire dont la distribution est déterminée par la distribution des  $X_i$  donc par celle de  $X$ ; ainsi la distribution de la statistique dépend du paramètre  $\theta$ .

**Exemples 6.1** La moyenne échantillon (ou moyenne observée), lorsque l'on a des valeurs de l'échantillon, est le nombre donné par

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

Ce nombre est la valeur de la statistique définie par

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

La variance-échantillon est le nombre défini par

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

il correspond à la statistique

$$S^2 = \frac{1}{n-1} \sum_{c=i}^n (X_i - \bar{X})^2.$$

La proportion-échantillon d'une grandeur suivant une loi de Bernoulli est

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i;$$

c'est la valeur de la statistique

$$\hat{P} = \frac{1}{n} \sum_{i=1}^n X_i.$$

## 6.2 Estimateurs

Une estimation d'un paramètre  $\theta$  est un nombre, obtenu à partir de l'information (c'est-à-dire des valeurs  $(x_1, \dots, x_n)$ ) d'un échantillon, qui servira comme approximation de la vraie valeur de  $\theta$ ; on demande bien entendu que ce nombre fasse partie des valeurs que peut prendre le paramètre  $\theta$ . La statistique dont la valeur est la fonction des observations qui est utilisée pour obtenir cette estimation est appelée l'estimateur du paramètre  $\theta$ ; ici aussi on demande que les valeurs que peut prendre cette statistique appartiennent à l'ensemble des valeurs que peut prendre le paramètre  $\theta$ . Donc on a

**Définition 6.2** Un estimateur d'un paramètre  $\theta$  est une statistique prenant ses valeurs dans l'ensemble  $\Theta$  des valeurs que peut prendre  $\theta$ .

Dans les exemples repris ci-dessus, si l'on veut estimer la moyenne d'une certaine grandeur dans une population, on prend un échantillon et on mesure la grandeur en question chez chacun de ces individus. La moyenne-échantillon  $\bar{x}$  est une estimation de la moyenne de la population dont on a extrait l'échantillon, et la variable  $\bar{X}$  est l'estimateur correspondant de la moyenne.

### 6.2.1 Estimation d'une probabilité ou d'une proportion

On veut savoir dans une population quelle est la proportion d'éléments (ou individus) [donc de résultats fondamentaux de notre espace fondamental  $\Omega$ ] qui ont une certaine propriété. On considère donc la variable aléatoire  $X$  définie sur  $\Omega$  et pouvant prendre deux valeurs:  $X(\omega) = 1$  si  $\omega$  a la propriété étudiée et  $X(\omega) = 0$  si  $\omega$  n'a pas la propriété étudiée. La variable  $X$  suit une

loi de Bernouilli  $Bin(1, p)$  et ce que l'on recherche est la proportion inconnue  $p$ .

Un problème où il s'agit de savoir avec quelle probabilité  $p$  on a un certain phénomène se traite de la même manière en considérant la variable aléatoire  $X$  définie sur  $\Omega$  telle que  $X(\omega) = 1$  si le phénomène est vrai pour  $\omega$  et  $X(\omega) = 0$  sinon.

Le paramètre inconnu est  $p$  qui prend ses valeurs dans  $[0, 1]$ . L'objectif est d'avoir de l'information sur  $p$ .

Le modèle est le suivant: l'expérience consiste à mesurer la valeur de la variable aléatoire

$$X = \begin{cases} = 1 & \text{avec probabilité } p \\ = 0 & \text{avec probabilité } (1 - p) \end{cases}$$

sur un individu. Prendre un échantillon aléatoire simple consiste à faire  $n$  répétitions identiques et indépendantes de cette expérience aléatoire. On obtient ainsi des mesures  $(x_1, \dots, x_n)$  de variables aléatoires  $(X_1, \dots, X_n)$  toutes indépendantes et chacune distribuée suivant une loi  $Bin(1, p)$ .

### Estimateurs de $p$

On peut construire différentes estimations de  $p$ . Par exemple:

$$\hat{p}_1 = \frac{1}{n}(x_1 + \dots + x_n)$$

$$\hat{p}_2 = x_1 \quad (\text{estimation très grossière! elle vaut 0 ou 1})$$

$$\hat{p}_3 = \frac{x_1 + 2x_2 + x_3}{4}$$

.....

Ces estimations correspondent aux estimateurs

$$\hat{P}_1 = \frac{1}{n}(X_1 + \dots + X_n)$$

$$\hat{P}_2 = X_1$$

$$\hat{P}_3 = \frac{X_1 + 2X_2 + X_3}{4}$$

....

La question se pose: quel estimateur choisir ?

L'estimateur usuel est donné par

$$\hat{P} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

donc, dans un échantillon, l'estimation de  $p$  est donnée généralement par la proportion  $\hat{p}$  de succès dans l'échantillon.

Question 1 : Cet estimateur est-il bon ?

Question 2 : Que veut dire bon ?

On ne peut pas dire ce que vaut  $\hat{p} - p$  puisqu'on ne connaît pas  $p$  ! La vraie question est celle-ci : si on prend beaucoup d'échantillons de  $n$  individus, comment les valeurs des estimations  $\hat{p}$  se placeront-elles par rapport à la vraie valeur (inconnue)  $p$  ? On va donc étudier la loi de probabilité de la variable aléatoire  $\hat{P}$  (l'estimateur) dont la valeur dans un échantillon donné est  $\hat{p}$  (l'estimation).

Remarquons que pour l'estimateur usuel,  $n\hat{P} = \sum_{i=1}^n X_i$  compte le nombre de succès dans  $n$  répétitions de l'expérience (i.e. dans l'échantillon) donc la distribution de l'estimateur  $n\hat{P}$  est celle d'une binomiale  $Bin(n, p)$

$$n\hat{P} \sim Bin(n, p)$$

Ainsi la statistique  $\hat{P}$  prend les valeurs  $0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1$  avec les probabilités (qui dépendent du paramètre  $p$ ):

$$P_p\left[\hat{P} = \frac{r}{n}\right] = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r}.$$

Le comportement de  $\hat{P}$  est fixé si on connaît  $p$ . Inversement, dans le comportement de  $\hat{P}$  se cache de l'information sur  $p$  et c'est cette information qu'on veut extraire.

L'espérance (moyenne) de  $\hat{P}$ , pour une valeur fixée quelconque de  $p \in [0, 1]$ , est :

$$E_p[\hat{p}] = \frac{1}{n} E(n\hat{p}) = \frac{1}{n} np = p$$

La variance de  $\hat{P}$  pour une valeur fixée quelconque de  $p \in [0, 1]$ , est :

$$Var_p[\hat{p}] = \frac{1}{n^2} Var_p(n\hat{P}) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$$

donc l'écart type  $\sigma$  de  $\hat{P}$ , pour une valeur fixée quelconque de  $p \in [0, 1]$ , vaut  $\frac{\sqrt{p(1-p)}}{\sqrt{n}}$ .

Rappelons que pour  $n$  grand

$$\frac{Bin(n, p) - np}{\sqrt{np(1-p)}} \rightarrow \mathcal{N}(0, 1)$$

Donc, pour  $n$  grand et pour un  $p \in [0, 1]$  fixé quelconque:

$$\hat{P} \approx \mathcal{N}(\mu = p, \sigma^2 = \frac{p(1-p)}{n}).$$

## 6.2.2 estimateurs non biaisés, estimateurs de la moyenne et de la variance

**Définition 6.3** Un estimateur  $\hat{\theta}$  de  $\theta$  est dit non biaisé si

$$E_{\theta}[\hat{\theta}] = \theta$$

pour toute valeur possible de  $\theta$ .

Remarquons que  $E_{\theta}$  indique que l'on prend l'espérance en considérant la distribution de probabilité donnée pour la valeur  $\theta$  du paramètre.

### Principe d'estimation non biaisée:

Quand on doit choisir entre différents estimateurs, on en prend un non biaisé.

**Proposition 6.1** • *Quand les  $X_i$  sont des variables aléatoires binomiales  $B(1, p)$ , la proportion échantillon  $\hat{P} = \frac{1}{n}(X_1 + \dots + X_n)$  est un estimateur non biaisé de  $p$ .*

- *Soient  $X_1, X_2, \dots, X_n$  des variables aléatoires où chaque  $X_i$  est distribué suivant la même loi de moyenne  $\mu$ . Alors,*

$$\hat{\mu} := \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

*est un estimateur non biaisé de  $\mu$ .*

- *Soient  $X_1, X_2, \dots, X_n$  des variables aléatoires indépendantes où chaque  $X_i$  est distribué suivant la même loi de moyenne  $\mu$  et de variance  $\sigma^2$ . Alors*

$$\hat{\sigma}^2 = S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

est un estimateur non biaisé de  $\sigma^2$ .

PREUVE Le point 1 a été vu page précédente. Pour les points 2 et 3, on a:

$$\begin{aligned} E_\mu[\hat{\mu}] &= E_\mu\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E_\mu[X_i] = \frac{n\mu}{n} = \mu \\ E_{\mu,\sigma^2}[S^2] &= E_{\mu,\sigma^2}\left[\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}\right] = \frac{1}{n-1} \left( \sum_{i=1}^n E_{\mu,\sigma^2}[X_i^2] - \frac{1}{n} E_{\mu,\sigma^2}\left[\left(\sum_{i=1}^n X_i\right)^2\right] \right) \\ &= \frac{1}{(n-1)} \left( \sum_{i=1}^n (\sigma^2 + \mu^2) - \frac{1}{n} (Var_{\mu,\sigma^2}\left[\sum_{i=1}^n X_i\right] + (E_{\mu,\sigma^2}\left[\sum_{i=1}^n X_i\right])^2 \right) \end{aligned}$$

(car pour toute variable aléatoire  $Y$ , on a  $Var(y) = E[Y^2] - (E[Y])^2$ ). Donc

$$\begin{aligned} E_{\mu,\sigma^2}[S^2] &= \frac{1}{n-1} (n\sigma^2 + n\mu^2 - \frac{1}{n}(n\sigma^2) - \frac{1}{n}(n\mu)^2) \\ &= \frac{1}{n-1} ((n-1)\sigma^2) + \frac{1}{n-1} (n\mu^2 - n\mu^2) = \sigma^2 \end{aligned}$$

■

### **Principe d'estimation non biaisée à variante minimum:**

Parmi tous les estimateurs de  $\theta$  qui sont non biaisés, choisissons celui dont la variance est minimum.

**Théorème 6.1 ( Sans démonstration ici!)** Si  $X_1, \dots, X_n$  est un échantillon aléatoire, avec tous les  $X_i$  indépendants et distribués suivant une loi normale de moyenne  $\mu$  et de variance  $\sigma^2$ , alors l'estimateur

$$\hat{\mu} = \bar{X}$$

est l'estimation non biaisée de variance minimum pour la moyenne  $\mu$ .

**Attention:** Il existe des distributions pour lesquelles  $\bar{X}$  n'est pas un bon estimateur pour la moyenne. Dans de nombreux cas, si l'on ne connaît pas la nature de la distribution identique de  $X_i$ , un bon estimateur de la moyenne est donné par la moyenne tronquée à 10 ou 20 % d'un échantillon !

### **6.2.3 méthode du maximum de vraisemblance**

Une méthode générale pour trouver une bonne estimation d'un paramètre lorsqu'on a un échantillon est la méthode du maximum de vraisemblance.

**Définition 6.4** Supposons que les variables aléatoires  $(X_1, \dots, X_n)$  ont une fonction (ou densité de) probabilité jointe donnée par

$$f(x_1, \dots, x_n, \theta_1, \dots, \theta_r) \quad (*)$$

où les  $\theta_1, \dots, \theta_r$  sont des paramètres inconnus. Prenons les valeurs observées  $(y_1, \dots, y_n)$  de ces variables pour un échantillon et regardons la fonction  $(*)$  calculée pour ces valeurs, donc  $f(y_1, \dots, y_n, \theta_1, \dots, \theta_r)$ , comme une fonction de  $\theta_1, \dots, \theta_r$ . On l'appelle la fonction de vraisemblance. La méthode du maximum de vraisemblance donne pour estimation  $\hat{\theta}_1, \dots, \hat{\theta}_r$  de  $\theta_1, \dots, \theta_r$  les valeurs des  $\theta_i$  qui maximisent cette fonction.

**Exemple 6.1** Un échantillon de 10 nouveaux compacts disques de marque A est prélevé. On constate que les 1<sup>er</sup>, 3<sup>ième</sup> et 10<sup>ième</sup> disques sont défectueux.

$$\text{soit } X_i = \begin{pmatrix} = 1 \text{ si le } i^{\text{ième}} \text{ disque est défectueux} \\ = 0 \text{ si le } i^{\text{ième}} \text{ disque est bon} \end{pmatrix}$$

Les  $X_i (i = 1, \dots, 10)$  sont des variables aléatoires indépendantes et identiquement distribuées suivant une loi binomiale  $B(1, p)$  et  $p$  est le paramètre inconnu. Les valeurs observées sont  $(1, 0, 1, 0, 0, 0, 0, 0, 0, 1)$  et la fonction de vraisemblance est

$$f(1, 0, 1, 0, 0, 0, 0, 0, 0, 1, p) = p(1-p)p(1-p)\dots(1-p)p = p^3(1-p)^7$$

Pour quelle valeur  $\hat{p}$  de  $p$ , cette fonction est-elle maximale?

$\hat{p}$  maximise  $p^3(1-p)^7$  ssi il maximise  $\log(p^3(1-p)^7) = 3\log p + 7\log(1-p)$ .

La dérivée de  $3\log p + 7\log(1-p)$  s'annule en  $\hat{p}$  donc

$$\frac{3}{\hat{p}} - \frac{7}{1-\hat{p}} = 0 \rightarrow 3(1-\hat{p}) - 7(\hat{p}) = 0 \rightarrow 3 - 10\hat{p} = 0 \rightarrow \hat{p} = \frac{3}{10}.$$

Remarquons que l'estimation  $\hat{p}$  de  $p$  obtenue par la méthode de maximum de vraisemblance coïncide avec l'estimation  $\frac{3}{10}$  obtenue par la proportion échantillon.



# Chapitre 7

## Intervalles de confiance

**Définition 7.1** On choisit un nombre  $\alpha$  dans  $[0, 1]$ , généralement très petit (les valeurs usuelles de  $\alpha$  sont 10%, 5%, 1% et 1/1000).

Un intervalle de confiance, au niveau de confiance  $1 - \alpha$ , pour un paramètre  $\theta$  est un intervalle de la droite réelle (contenu dans l'ensemble  $\Theta$  des valeurs que peut prendre le paramètre  $\theta$ ), défini à partir des observations d'un échantillon, et qui recouvre la "vraie" valeur (inconnue) de  $\theta$  avec une probabilité au moins égale à  $1 - \alpha$ .

En d'autres termes, on veut un intervalle  $[l^-, l^+] \subset \Theta$  dont les bornes ( $l^-$  et  $l^+$ ) sont des fonctions des observations  $(x_1, \dots, x_n)$  d'un échantillon, donc sont les valeurs de statistiques  $L^-$  et  $L^+$ , prenant des valeurs dans  $\Theta$ , telles que

$$P_\theta[L^- \leq \theta \leq L^+] \geq 1 - \alpha \quad \forall \theta \in \Theta.$$

Rappelons que des statistiques sont des fonctions de  $X_1, \dots, X_n$  ne dépendant pas explicitement des paramètres inconnus; la distribution d'une statistique dépend de la distribution de  $X$  et donc de la valeur du paramètre  $\theta$ . Quand on note  $P_\theta$ , on indique qu'on calcule la probabilité pour une valeur fixée  $\theta$  du paramètre et donc en considérant la distribution des statistiques pour cette valeur fixée du paramètre.

## 7.1 intervalle de confiance pour une proportion ou une probabilité

Prenons comme précédemment un échantillon aléatoire simple  $(X_1, \dots, X_n)$  avec  $X_i \sim \text{Bin}(1, p)$  et considérons l'estimateur  $\hat{P} = \frac{1}{n} \sum_{i=1}^n X_i$ . Si l'échantillon est grand,

$$\hat{P} \approx \mathcal{N}(\mu = p, \sigma^2 = \frac{p(1-p)}{n})$$

Fait: Lorsque  $np(1-p) > 9$ , on admet que l'approximation normale est une approximation raisonnable de la loi de  $\hat{P}$ .

Nous supposons dans la suite que c'est le cas.

Trouvons dans notre table de la fonction  $\phi$  de répartition de la loi normale standard le nombre réel  $z_{\alpha/2}$  tel que

$$\phi(z_{\alpha/2}) = P[\mathcal{N}(0, 1) \leq z_{\alpha/2}] = 1 - \alpha/2.$$

Alors,

$$P[-z_{\alpha/2} \leq \mathcal{N}(0, 1) \leq z_{\alpha/2}] = 1 - \alpha,$$

donc,

$$P_p \left[ -z_{\alpha/2} \leq \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{\alpha/2} \right] \approx 1 - \alpha$$

et l'on en déduit que

$$P_p \left[ \hat{P} - z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{P} + z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \right] \approx 1 - \alpha.$$

Si l'on connaît les valeurs  $(x_1, \dots, x_n)$  obtenues dans un échantillon, on calcule l'estimation  $\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i$  et l'on sait que l'intervalle déterminé par  $[\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$  recouvrira  $p$  avec une probabilité  $\approx 1 - \alpha$ . Attention: les bornes de cet intervalle dépendent de  $p$  ce qui n'aurait aucun sens puisqu'on ne connaît pas  $p$ ; ce ne sont pas les valeurs de statistiques. On va donc prendre comme intervalle de confiance pour  $p$  au niveau de confiance  $1 - \alpha$

$$\left[ \hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right].$$

Les bornes de cet intervalle  $l_- = \hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$  et  $l_+ = \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$  sont les valeurs dans l'échantillon des statistiques  $L^- = \hat{P} - z_{\alpha/2} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$  et  $L^+ = \hat{P} + z_{\alpha/2} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$  et on a bien, car  $n$  est grand:

$$P_p \left[ \hat{P} - z_{\alpha/2} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \leq p \leq \hat{P} + z_{\alpha/2} \sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \right] \approx 1 - \alpha.$$

**Exemple 7.1** Une nouvelle méthode "graphie stress téléthermometry" a permis de détecter 23 parmi 29 cancers du sein en phase initiale.

Pour avoir une idée de l'efficacité de cette méthode, on aimerait connaître une intervalle de confiance (au niveau de confiance 90%) pour la proportion  $p$  de cancers détectés par cette méthode.

On a  $1 - \alpha = 90\% \rightarrow \alpha = 0,1$ ; dès lors  $z_{\alpha/2} = 1,645$ . D'autre part  $n = 29$  et l'estimation de  $p$  dans notre échantillon est  $\hat{p} = \frac{23}{29}$ . L'intervalle de confiance recherché est

$$\left[ \frac{23}{29} - 1,645 \sqrt{\frac{\frac{23}{29}(1 - \frac{23}{29})}{29}}, \quad \frac{23}{29} + 1,645 \sqrt{\frac{\frac{23}{29}(1 - \frac{23}{29})}{29}} \right] \approx [0,669; 0,917].$$

Il y a donc 9 chances sur 10 pour que la vraie probabilité de détection soit comprise entre 0,669 et 0,917.

**Remarque 7.1** On peut "agir" sur l'intervalle de confiance soit à travers  $\alpha$  soit à travers  $n$ .

- Pour un  $n$  fixé, si  $\alpha$  diminue (donc si le niveau de confiance augmente) alors la longueur de l'intervalle de confiance (I.C.) augmente.
- Pour un  $\alpha$  fixé, si  $n$  augmente alors la longueur de l'I.C. diminue.

### Question

Après l'observation ci-dessus sur les cancers du sein, combien de patients faut-il observer pour obtenir une intervalle de confiance (au niveau de confiance

90%) dont la longueur est inférieure à 0,10?

Remarquons que la longueur de l'intervalle de confiance est égale à

$$2z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}};$$

on veut que cette quantité soit  $\leq 0,10$ , donc on a

$$\frac{(2.1,645)^2 \frac{23}{29} (1 - \frac{23}{29})}{0,01} \leq n$$

ce qui implique  $n \geq 177,68$ . Il faut donc faire au moins 178 observations.

**Remarque 7.2** Ce calcul s'est fait avec une idée préalable de l'ordre de grandeur de  $p$  (donné pour le  $\hat{p}$  d'un échantillon; ici  $\frac{23}{29}$ ). Sinon on utilise le fait que  $\hat{p}(1-\hat{p}) \leq 1/4$  quelle que soit la valeur de  $\hat{p} \in [0,1]$ .

## 7.2 Moyenne d'une grandeur, grand échantillon

On considère l'observation d'une grandeur  $X$  dans une population et on aimerait estimer la moyenne  $\mu$  de cette grandeur. On prend un échantillon aléatoire simple de  $n$  individus et on considère la moyenne  $\bar{x}$  des grandeurs  $x_i$  mesurées sur les  $n$  individus. Comme précédemment, l'estimation de la moyenne est donnée par la moyenne-échantillon  $\bar{x} = \frac{1}{n}(x_1 + \dots + x_n)$  et l'estimateur correspondant est

$$\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$$

où les  $X_i$  sont des variables aléatoires indépendantes et identiquement distribuées (leur distribution est celle de  $X$ ).

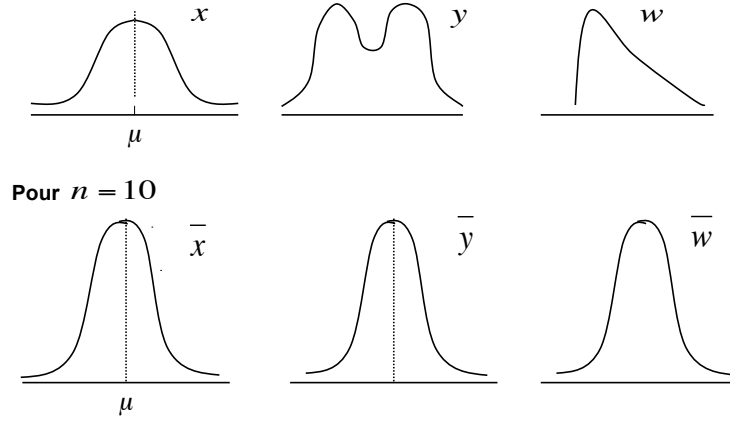
Quelle est la distribution (= loi de probabilité) de  $\bar{X}$  ?

On sait que  $E[X] = \mu$ . De plus, le théorème central limite nous dit que, si  $n$  tend vers l'infini,

$$\frac{\bar{X} - \mu}{\sigma\sqrt{n}} \text{ tend vers la loi normale standard } \mathcal{N}(0,1),$$

si  $\mu$  est la moyenne et  $\sigma$  l'écart-type de la loi de chacun des  $X_i$ .

Ceci veut dire que quelle que soit la forme de la distribution originale  $X$ , le fait de prendre des moyennes  $\frac{X_1 + \dots + X_n}{n}$  résulte en une distribution normale.



Pour notre estimateur  $\bar{X}$  de la moyenne (dont la distribution dépend des paramètres  $\mu$  et  $\sigma$ ), on a, pour  $n$  grand:

$$P_{\mu, \sigma} \left[ -z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2} \right] = 1 - \alpha$$

quelles que soient les valeurs  $\mu$  et  $\sigma$  fixées donc

$$P_{\mu, \sigma} \left[ \bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right] = 1 - \alpha \quad \forall \mu, \sigma.$$

Ainsi, on a

**Proposition 7.1 (I.C. pour la moyenne, grand échantillon,  $\sigma$  connu)**

*Si la variance  $\sigma$  de la variable  $X$  est connue, on obtient un intervalle de confiance pour la moyenne  $\mu$  de cette grandeur  $X$  au niveau de confiance  $1 - \alpha$  en prenant un grand échantillon, en calculant la moyenne-échantillon  $\bar{x}$ , et en définissant l'intervalle:*

$$\left[ \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right].$$

Si la variance n'est pas connue, en remplaçant  $\sigma$  par son estimation  $S$ , on a, pour un grand échantillon

$$P_{\mu, \sigma} \left[ \bar{X} - z_{\alpha/2} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} \frac{S}{\sqrt{n}} \right] \approx 1 - \alpha \quad \forall \mu, \sigma..$$

Donc

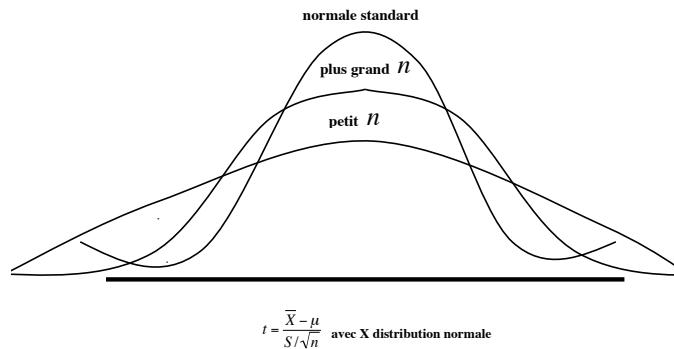
**Proposition 7.2 (I.C. pour la moyenne, grand échantillon,  $\sigma$  inconnu)**

*On prend un grand échantillon, on calcule la valeur  $\bar{x}$  pour la moyenne échantillon et la valeur  $s^2$  pour la variance échantillon. Un intervalle de confiance pour la moyenne au niveau de confiance  $1 - \alpha$  est donné par*

$$\left[ \bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}} \right].$$

**Attention**

La variable  $\frac{\bar{X} - \mu}{s/\sqrt{n}}$  est approximativement distribuée suivant la loi normale quand  $n$  est grand! Si  $n$  est petit, la distribution de  $\frac{\bar{X} - \mu}{s/\sqrt{n}}$  sera plus dispersée que la normale standard.



### 7.3 Moyenne d'une grandeur distribuée suivant une normale, avec un petit échantillon (variable de Student)

Supposons avoir une grandeur  $X$  distribuée suivant une loi normale de moyenne  $\mu$  et de variance  $\sigma^2$ . Prenons un échantillon  $X_1, X_2, \dots, X_n$ ; donc les  $X_i$  sont

des variables aléatoires indépendantes toutes distribuées suivant une loi normale  $\mathcal{N}(\mu, \sigma^2)$ . L'estimateur pour la moyenne donné par  $\bar{X} = \frac{X_1 + \dots + X_n}{n}$  a alors la distribution d'une loi normale de moyenne  $\mu$  et de variance  $\frac{\sigma^2}{n}$ . Donc,

**Proposition 7.3 (I.C. pour la moyenne, petit échantillon,  $\sigma$  connu)**

*Si une grandeur est distribuée suivant une loi normale de moyenne  $\mu$  inconnue et de variance  $\sigma^2$  connue, un intervalle de confiance pour la moyenne au niveau de confiance  $1 - \alpha$  est donné en calculant la moyenne-échantillon  $\bar{x}$  des valeurs mesurées dans un échantillon de taille  $n$  (quelconque) et en considérant l'intervalle*

$$\left[ \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right].$$

Si la variance n'est pas connue, on doit la remplacer par une estimation.

Fait: La variable aléatoire

$$T_n = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

a une loi de probabilité appelée loi de student à  $n - 1$  degrés de liberté et notée  $T_{n-1}$  (ou  $t_{n-1}$ ) .

Il existe des tables donnant les valeurs de la fonction de répartition d'une loi de Student  $t_k$  à  $k$  degrés de liberté.

En pratique, on peut utiliser la loi de Student pour estimer des intervalles de confiance à partir d'un petit échantillon même quand la loi  $X$  de la population n'est pas normale, si la densité de cette loi a approximativement une forme de cloche:



**Proposition 7.4** *Pour avoir un intervalle de confiance pour la moyenne  $\mu$  au niveau de confiance  $1 - \alpha$  à partir d'un échantillon aléatoire simple*

de  $n$  éléments, d'une variable  $X$  normale (ou ayant une forme approximativement "en cloche"), on regarde dans les tables les valeurs  $t_{\alpha/2, n-1}$  telles que

$$P[-t_{n-1; \alpha/2} \leq T_n \leq t_{n-1; \alpha/2}] = 1 - \alpha,$$

on calcule la moyenne échantillon  $\bar{x}$  et l'écart-type échantillon  $s$ , et l'intervalle de confiance pour  $\mu$  au niveau de confiance  $1 - \alpha$  est:

$$\left[ \bar{x} - t_{n-1; \alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1; \alpha/2} \frac{s}{\sqrt{n}} \right]$$

**Remarque 7.3** Le nombre réel noté  $t_{m; \beta}$  est tel que la fonction de répartition de la loi de Student calculé en ce nombre est égale à  $1 - \beta$ .

Attention : cet élément est souvent noté dans les livres  $t_{m; 1-\beta}$ ; nous avons adopté dans ce cours une notation uniforme basée sur la notation traditionnelle pour la loi normale standard.



# Chapitre 8

## Tests d'hypothèses

Dans un test d'hypothèses, on a deux hypothèses à propos de la vraie valeur d'un paramètre dont l'une et seulement une est vraie. En se basant sur un échantillon de la population, on veut décider laquelle de ces deux hypothèses est correcte.

Si  $\Theta$  est l'ensemble des valeurs possibles du paramètre  $\theta$ , on a donc une partition de  $\Theta$  en deux sous-ensemble  $\mathcal{H}_o$  et  $\mathcal{H}_1$ :

$$\Theta = \mathcal{H}_o \cup \mathcal{H}_1 \quad \mathcal{H}_o \cap \mathcal{H}_1 = \emptyset$$

On désire tester l'hypothèse nulle  $H_o$  selon laquelle la vraie valeur de  $\theta$  est dans  $\mathcal{H}_o$ , contre l'hypothèse alternative  $H_1$ , aussi appelée contre hypothèse, selon laquelle la vraie valeur de  $\theta$  appartient à  $\mathcal{H}_1$ .

Face à un problème de test, seules deux décisions sont possibles:

- 1- soit on décide de rejeter l'hypothèse  $H_o$  :  $RH_o$ ;
- 2- soit on décide de ne pas rejeter l'hypothèse  $H_o$  :  $\bar{R}H_o$ .

**Exemples 8.1** 1. On jette une pièce de monnaie  $n$  fois et on s'intéresse à la probabilité  $p$  d'obtenir face avec cette pièce. Trois problèmes de test sont envisageables:

[a.] L'hypothèse  $H_o$  correspond à l'hypothèse selon laquelle la pièce est correctement équilibrée ( $H_o$  correspond à  $p = 1/2$ ); l'hypothèse

alternative  $H_1$  correspond à une pièce déséquilibrée ( $H_1$  correspond à  $p \neq 1/2$ ).

[b.] L'hypothèse  $H_o$ : la pièce est correctement équilibrée ou déséquilibrée en faveur de face ( $H_o : p \geq 1/2$ ) est opposée à l'hypothèse alternative  $H_1$  selon laquelle la pièce est déséquilibrée en faveur de pile ( $H_1 : p < 1/2$ ).

[c.] L'hypothèse  $H_o$  : la pièce est correctement équilibrée ou déséquilibrée en faveur de pile ( $H_o : p \leq 1/2$ ) est opposée à l'hypothèse alternative selon laquelle la pièce est déséquilibrée en faveur de face ( $H_i : p > 1/2$ ).

2. Dans un sondage pré-référendum, on s'intéresse à la proportion  $p$  d'intentions de vote "oui" dans le corps électoral. Dans le cas où le sondage est commandé par les partisans du oui, on s'intéresse au test :  
 $H_o : p \leq 1/2$  (le non l'emporte ou on a un ex-aequo)  
 $H_1 : p > 1/2$  (le oui l'emporte).

3. Dans un traitement classique contre l'obésité, seuls 25 % des patients ont perdu 9 kg ou plus à la fin du traitement. On imagine une nouvelle procédure de modification de comportement pour introduire une perte de kilos chez les obèses. Ceux qui ont conçu cette nouvelle méthode pensent que plus de 25 % des patients traités perdront au moins 9 kg. Soit  $p$  la proportion d'individus obèses qui perdraient plus de 9 kg avec cette méthode. On s'intéresse au problème de test:  
 $H_o$  : la nouvelle méthode n'est pas plus efficace que la méthode classique:  $p \leq 0,25$   
 $H_i$  : la nouvelle méthode est plus efficace :  $p > 0,25$ .

4. Un fabricant vend des boîtes de céréales dont le contenu annoncé est de 250 gr. On considère le poids moyen  $\mu$  contenu dans une telle boîte. On s'intéresse au problème de test.  
 $H_o$ : le contenu des boîtes pèse effectivement en moyenne un poids  $\geq 250$  gr, donc  $\mu \geq 250$  gr  
 $H_1$  : le fabricant "triche" systématiquement sur le poids du contenu de ses boîtes de céréales, soit  $\mu < 250$  gr.

Dans notre étude  $H_o$  est l'hypothèse "favorisée" par la méthode; on la fera correspondre souvent à une situation de statu-quo. On continuera à croire

l'hypothèse  $H_0$  sauf si l'évidence expérimentale la contredit fortement.

Un test d'hypothèse est une règle de comportement (ou règle de décision) basé sur les valeurs observées  $(x_i, \dots, x_n)$  dans un échantillon; le test est caractérisée par un sous-ensemble  $\tilde{R}$  de l'ensemble des valeurs possibles des observations dans  $\mathbb{R}^n$  et la règle s'exprime de la manière suivante:

si  $(x_1, \dots, x_n) \in \tilde{R}$  alors on rejette l'hypothèse  $H_0$ .

**Exemples 8.2** 1. Reprenons la pièce de monnaie avec le test d'hypothèse  $H_0$  = la pièce n'est pas déséquilibrée, ( $p = 1/2$ ), donc l'hypothèse  $H_1$  = la pièce est déséquilibrée ( $p \neq 1/2$ ). Jetons la pièce 100 fois et considérons la proportion  $\hat{p}$  de "faces" observées. On pourrait décider de rejeter  $H_0$  si les observations sont telles que  $\hat{p} > 0,75$  ou  $\hat{p} < 0,25$  Mais on pourrait aussi décider de rejeter  $H_0$  si  $\hat{p} > 0,8$  ou  $\hat{p} < 0,2$ .

2. Reprenons le traitement des obèses par la nouvelle méthode; faisons un test avec pour hypothèse  $H_0$  que la nouvelle méthode n'est pas plus efficace que la méthode classique.. Appliquons ce traitement à 20 patients. Soit  $x$  nombre des individus qui ont perdu plus de 9 kilos après ce traitement. On pourrait décider  $RH_0$  (de rejeter  $H_0$  donc de considérer la nouvelle méthode de traitement comme plus efficace) si la valeur observée de  $x$  appartient à  $\{8, 9, 10, \dots, 20\}$ .

3. Pour les boîtes de céréales, on peut décider de  $RH_0$  si  $\bar{x} \leq 220gr$  où  $\bar{x}$  est le poids moyen du contenu mesuré de 30 boîtes prises au hasard.

**Définition 8.1** En prenant une décision du type décrit plus haut (rejeter ou ne pas rejeter l'hypothèse  $H_0$  opposée à l'hypothèse  $H_1$ ) on risque de commettre deux types d'erreur:

-rejeter l'hypothèse  $H_0$  alors qu'elle est vraie; on appelle cette erreur une erreur de première espèce;

- ne pas rejeter l'hypothèse  $H_0$  alors qu'elle est fausse; on appelle cette erreur une erreur de deuxième espèce.

|              | $H_0$ vraie ( $\theta \in \mathcal{H}_0$ ) | $H_0$ fausse ( $\theta \in \mathcal{H}_1$ ) |
|--------------|--|---|
| $RH_0$       | erreur de 1 <sup>ère</sup> espèce          | bonne décision                              |
| $\bar{R}H_0$ | bonne décision                             | erreur de 2 <sup>ème</sup> espèce           |

Un test d'hypothèse ne pourra jamais être infallible ! On voudrait minimiser les risques de 1<sup>ère</sup> et de 2<sup>ème</sup> espèce. Cependant, minimiser le risque de 1<sup>ère</sup> espèce (qui est de rejeter l'hypothèse  $H_o$  alors qu'elle est vraie) conduit à adopter un comportement prudent quant au rejet de l'hypothèse  $H_o$ , donc à considérer une région de rejet petite, ce qui conduit à augmenter le risque de 2<sup>ème</sup> espèce (qui est ne pas rejeter l'hypothèse  $H_o$  alors qu'elle est fausse). On doit donc faire un compromis appelé le principe de Neyman

**Définition 8.2 (principe de Neyman)**

[1.] On ne minimise pas le risque de 1<sup>ère</sup> espèce mais on lui impose une contrainte dite condition de niveau  $\alpha$ . On se donne un petit nombre  $\alpha \in [0, 1]$  (les valeurs usuelles de  $\alpha$  sont 10%, 5%, 1%,...) et on demande que la probabilité de rejeter  $H_o$ , alors que  $\theta \in \mathcal{H}_o$  (donc que  $H_o$  est vraie), soit inférieure à  $\alpha$ ; donc on demande

$$P_\theta[RH_o] \leq \alpha \quad \forall \theta \in \mathcal{H}_o$$

où  $P_\theta$  vaut dire qu'on calcule la probabilité en supposant que la valeur du (ou des) paramètre(s) est le(s) nombre(s)  $\theta$ .

[2.] Parmi les procédures de tests satisfaisant à la condition de niveau, on choisit celle qui minimise le risque de 2<sup>ème</sup> espèce, c'est à dire qui maximise la puissance du test où l'on définit la puissance du test  $\Pi_\theta$  par

$$\begin{aligned} \Pi_\theta &= 1 - \text{le risque de 2<sup>ème</sup> espèce} \\ &= 1 - P_\theta[\neg RH_o \mid H_o \text{ est fausse}] \\ &= P_\theta[RH_o \mid \theta \in \mathcal{H}_1] \end{aligned}$$

$$(\text{= probabilité de rejeter } H_o \text{ quand } H_o \text{ est fausse})$$

Comment sélectionner la région  $\tilde{R}$ ?

**Définition 8.3** De manière pratique, pour faire un test, on choisit une statistique de test  $T$  qui est une variable aléatoire, fonction des variables définissant l'échantillon  $(X_1, \dots, X_n)$  ne dépendant pas explicitement du(des) paramètre(s) inconnu(s). On définit une région  $R$  de la droite réelle appelée région de rejet et on définit la règle de comportement suivante, en termes de la valeur  $t$  de la statistique  $T$  dans l'échantillon:

- si  $t \in R$  alors  $RH_o$  (on rejette l'hypothèse  $H_o$ )
- si  $t \notin R$  alors  $\neg RH_o$  (on ne rejette pas l'hypothèse  $H_o$ ).

La région  $R$  induit une région  $\tilde{R}$  dans l'espace des observations:

$$(x_1, \dots, x_n) \in \tilde{R} \iff t \in R.$$

Par exemple dans l'exemple de la pièce de monnaie  $T = \hat{P}$ , dans l'exemple des obèses  $T = \sum_{i=1}^{20} X_i$  si  $X_i$  est la variable binomale qui vaut 1 quand la perte de poids est  $\geq 9$  kgs, et dans l'exemple des boîtes de céréales  $T = \bar{X}$ .

**Proposition 8.1** *La statistique  $T$  est une fonction des variables aléatoires indépendantes  $(X_1, \dots, X_n)$ , donc sa distribution dépend de la valeur du(des) paramètre(s) inconnu(s)  $\theta$ . Quand la décision de rejet est définie à partir d'une statistique  $T$  et d'une région  $R$ , la condition de niveau au niveau  $\alpha$  s'écrit*

$$P_\theta[T \in R] \leq \alpha \quad \forall \theta \in \mathcal{H}_o.$$

**Exemple 8.1** Dans l'exemple de la pièce de monnaie, si le nombre  $n$  de jets de la pièce est grand, la distribution de  $\hat{P}$  sous la condition que la probabilité d'avoir face est  $p_o$  est une normale

$$\mathcal{N}(p_o, \frac{p_o(1-p_o)}{n}).$$

On va définir une région de rejet en termes de valeurs prises par  $\hat{p}$  : si la valeur observée de  $\hat{p} \in R$  alors on rejette l'hypothèse  $H_o$ .

Si l'hypothèse  $H_o$  pour la pièce est que  $p_o = 1/2$  (i.e. la pièce est équilibrée); on décide de rejeter cette hypothèse si le  $\hat{p}$  observé est tel que  $|\hat{p} - 1/2| > B$ . On veut déterminer  $B$  pour avoir une condition de niveau de 10% donc pour que

$$P_{p_o=1/2} \left[ |\hat{P} - 1/2| > B \right] \leq \alpha = 10\%.$$

Remarquons que

$$P_{p_o} \left[ -B < \hat{P} - p_o < B \right] = P_{p_o} \left[ \frac{-B}{\sqrt{\frac{p_o(1-p_o)}{n}}} < \frac{\hat{P} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}} < \frac{B}{\sqrt{\frac{p_o(1-p_o)}{n}}} \right].$$

Rappelons que  $\frac{\hat{P} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}}$  a la distribution d'une normale  $\mathcal{N}(0, 1)$  si la valeur de  $p$  est  $p_o$ ; de plus on a défini  $z_\alpha \in \mathbb{R}$  par  $P[\mathcal{N}(0, 1) < z_\alpha] = 1 - \alpha$ . Donc

$$P[-z_{\alpha/2} < \mathcal{N}(0, 1) < z_{\alpha/2}] = 1 - \alpha$$

et donc

$$\alpha = 1 - P_{p_0}[-z_{\alpha/2} < \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} < z_{\alpha/2}].$$

Il faut donc choisir un  $B$  tel que  $z_{\alpha/2} \leq \frac{B}{\sqrt{\frac{p_0(1-p_0)}{n}}}$  pour avoir un risque de 1<sup>ère</sup> espèce  $\leq \alpha$ . Maintenant on veut minimiser le risque de 2<sup>ème</sup> espèce tout en respectant cette condition de niveau; on veut donc une région de rejet la plus grande possible ce qui revient à choisir un  $B$  le plus petit possible puisqu'on rejette  $H_0$  si  $|\hat{p} - 1/2| > B$ . Donc on prend  $\frac{B}{\sqrt{\frac{p_0(1-p_0)}{n}}} = z_{\alpha/2}$ .

Ainsi on est conduit à rejeter l'hypothèse que la pièce est équilibrée si la proportion de faces  $\hat{p}$  observée dans l'échantillon est telle que

$$|\hat{p} - 1/2| > z_{\alpha/2} \sqrt{\frac{1/2(1/2)}{n}}.$$

L'intervalle d'acceptation de l'hypothèse "la pièce n'est pas truquée" pour un test au niveau  $\alpha$  est

$$\left(\frac{1}{2} - \frac{1}{2\sqrt{n}}z_{\alpha/2}, \quad \frac{1}{2} + \frac{1}{2\sqrt{n}}z_{\alpha/2}\right)$$

## 8.1 Tests sur une probabilité ou une proportion

On veut faire des tests sur la valeur  $p$  d'une proportion ou d'une probabilité à partir des résultats d'un échantillon.

On considère trois types de tests:

$$\left. \begin{array}{ll} \text{Test 1} & H_0 : p = p_0 \quad (H_1 : p \neq p_0) \\ \text{Test 2} & H_0 : p \geq p_0 \quad (H_1 : p < p_0) \\ \text{Test 3} & H_0 : p \leq p_0 \quad (H_1 : p > p_0) \end{array} \right\} \text{ où } p_0 \text{ est choisi dans } [0, 1].$$

On considère donc comme précédemment un modèle échantillon aléatoire simple  $(X_1, \dots, X_n)$  avec  $X_i \sim \text{Bin}(1, p)$  et l'estimateur  $\hat{P} = \frac{1}{n} \sum_{i=1}^n X_i$ . Les valeurs échantillon sont  $(x_1, \dots, x_n)$  qui sont chacune égale à 0 ou 1 et la proportion échantillon est le nombre  $\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i$ . Si l'échantillon est grand, la distribution de  $\hat{P}$ , pour une valeur  $p$  du paramètre, est:

$$\hat{P} \approx \mathcal{N}(\mu = p, \sigma^2 = \frac{p(1-p)}{n});$$

et l'approximation normale est une approximation raisonnable de la loi de  $\hat{P}$  lorsque  $np(1-p) > 9$ , . Nous supposons dans la suite que c'est le cas. (Si l'échantillon est petit, il faut utiliser le fait que  $n\hat{P}$  a la distribution d'une loi binomiale  $Bin(n, p)$ ).

### 8.1.1 Test 1 au niveau $\alpha$ ; $H_0 : p = p_0$ ( $H_1 : p \neq p_0$ )

On va rejeter l'hypothèse  $H_0$  si la valeur de la proportion échantillon  $\hat{p}$  n'est pas proche de  $p_0$ , donc si  $|\hat{p} - p_0| > C$ . La condition de niveau,  $P_p[RH_0] \leq \alpha \forall p \in \mathcal{H}_0$  s'écrit:

$$P_{p=p_0}[|\hat{P} - p_0| > C] \leq \alpha.$$

On sait que

$$P_{p_0} \left[ -z_{\alpha/2} \leq \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \leq z_{\alpha/2} \right] \approx 1 - \alpha$$

donc

$$P_{p_0} \left[ |\hat{P} - p_0| > z_{\alpha/2} \sqrt{\frac{p_0(1-p_0)}{n}} \right] \approx \alpha.$$

On en déduit que pour satisfaire la condition de niveau au niveau  $\alpha$ , il faut que

$$C \geq z_{\alpha/2} \sqrt{\frac{p_0(1-p_0)}{n}}.$$

Maintenant, on veut minimiser la probabilité d'erreur de seconde espèce (qui consiste à ne pas rejeter  $H_0$  alors que l'hypothèse  $H_0$  est fausse); on veut donc une région de rejet la plus grande possible (tout en satisfaisant la condition de niveau); on veut donc le  $C$  le plus petit possible. Ainsi, on a

$$C = z_{\alpha/2} \sqrt{\frac{p_0(1-p_0)}{n}}.$$

La règle de comportement sera donc la suivante: on rejette  $H_0$  si

$$|\hat{p} - p_0| > z_{\alpha/2} \sqrt{\frac{p_0(1-p_0)}{n}}.$$

### 8.1.2 Test 2 au niveau $\alpha$ ; $H_0 : p \geq p_0$ ( $H_1 : p < p_0$ )

On va rejeter l'hypothèse  $H_0$  si la valeur de la proportion échantillon  $\hat{p}$  est telle que  $\hat{p} - p_0 < -D$  où  $D > 0$ .

La condition de niveau,  $P_p[RH_0] \leq \alpha \ \forall p \in \mathcal{H}_0$  s'écrit:

$$P_p[\hat{P} - p_0 < -D] \leq \alpha \quad \forall p \geq p_0.$$

Si le paramètre vaut  $p$ ,  $\hat{P}$  a une distribution normale centrée en  $p$  donc  $\hat{P} - p_0$  est une distribution normale centrée en  $p - p_0$ . Donc la probabilité  $P_p[\hat{P} - p_0 < -D]$  est inférieure à  $P_{p_0}[\hat{P} - p_0 < -D]$  pour  $p > p_0$ . La condition de niveau s'écrit donc

$$P_{p_0}[\hat{P} - p_0 < -D] \leq \alpha.$$

Comme  $\frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim \mathcal{N}(0, 1)$  si  $p = p_0$ , on sait que

$$P_{p_0} \left[ \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \leq -z_\alpha \right] \approx \alpha$$

donc

$$P_{p_0} \left[ \hat{P} - p_0 < -z_\alpha \sqrt{\frac{p_0(1-p_0)}{n}} \right] \approx \alpha.$$

On en déduit que pour satisfaire la condition de niveau au niveau  $\alpha$ , il faut que

$$-D \leq -z_\alpha \sqrt{\frac{p_0(1-p_0)}{n}}.$$

Maintenant, on veut minimiser la probabilité d'erreur de seconde espèce, donc une région de rejet la plus grande possible (tout en satisfaisant la condition de niveau); on veut donc le  $-D$  le plus proche de zéro possible. Ainsi, on a

$$D = z_\alpha \sqrt{\frac{p_0(1-p_0)}{n}}.$$

La règle de comportement sera donc la suivante: on rejette  $H_0$  si

$$\hat{p} - p_0 < -z_\alpha \sqrt{\frac{p_0(1-p_0)}{n}}.$$



### 8.1.3 Test 3 au niveau $\alpha$ ; $H_0 : p \leq p_0$ ( $H_1 : p > p_0$ )

On va rejeter l'hypothèse  $H_0$  si la valeur de la proportion échantillon  $\hat{p}$  est telle que  $\hat{p} - p_0 > E$  où  $E > 0$ . La condition de niveau,  $P_p[RH_0] \leq \alpha \forall p \in \mathcal{H}_0$  s'écrit:

$$P_p[\hat{P} - p_0 > E] \leq \alpha \quad \forall p \leq p_0.$$

Sous la condition que le paramètre vaut  $p$ ,  $\hat{P}$  a une distribution normale centrée en  $p$  donc  $\hat{P} - p_0$  est une distribution normale centrée en  $p - p_0$ . Donc la probabilité  $P_p[\hat{P} - p_0 > E]$  est inférieure à  $P_{p_0}[\hat{P} - p_0 > E]$  pour  $p < p_0$ . La condition de niveau s'écrit donc

$$P_{p_0}[\hat{P} - p_0 > E] \leq \alpha.$$

Comme  $\frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim \mathcal{N}(0, 1)$  si  $p = p_0$ , on sait que

$$P_{p_0} \left[ \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \geq z_\alpha \right] \approx \alpha$$

donc

$$P_{p_0} \left[ \hat{P} - p_0 > z_\alpha \sqrt{\frac{p_0(1-p_0)}{n}} \right] \approx \alpha.$$

On en déduit que pour satisfaire la condition de niveau au niveau  $\alpha$ , il faut que

$$E \geq z_\alpha \sqrt{\frac{p_0(1-p_0)}{n}}.$$

Maintenant, on veut minimiser la probabilité d'erreur de seconde espèce, donc une région de rejet la plus grande possible (tout en satisfaisant la condition de niveau); on veut donc le  $E$  le plus petit possible. Ainsi, on a

$$E = z_\alpha \sqrt{\frac{p_0(1-p_0)}{n}}.$$

La règle de comportement sera donc la suivante: on rejette  $H_0$  si

$$\hat{p} - p_0 > z_\alpha \sqrt{\frac{p_0(1-p_0)}{n}}.$$

## 8.2 Tests sur la moyenne d'une grandeur

### 8.2.1 exemple de test sur la moyenne d'une population normale

Un fabricant de pneus utilise une machine pour tester la vie des pneus. Il obtient après un grand nombre d'expériences une vie moyenne de 25.000 km avec un écart type de 2.500 km. On envisage un changement dans le processus de fabrication; on imagine qu'il ne va pas modifier l'écart type mais on espère qu'il augmentera la moyenne de vie des pneus. Un échantillon de 25 pneus est fabriqué par le nouveau processus; leur durée de vie moyenne est de 27.200 km. Mettre en place ce nouveau processus serait coûteux et ne serait envisagé que si la vraie moyenne de durée de vie de ces nouveaux pneus était supérieure à 26.000 km. La moyenne échantillon de 27.200 km justifie-t-elle ce changement?

On est donc dans la situation suivante: On a une population (conceptuelle) constituée par tous les pneus qui seraient fabriqués par la nouvelle méthode.

**Condition 1** On s'intéresse à la durée de vie de ces pneus et on suppose que cette grandeur a une distribution normale de moyenne  $\mu$  (inconnue) et d'écart type  $\sigma$  supposé connu, égal à celui du procédé précédent (donc  $\sigma = 25.000$  km).

**Condition 2** L'échantillon des vies des 25 pneus donne les valeurs  $x_1, \dots, x_{25}$  d'un échantillon aléatoire  $X_1, \dots, X_{25}$  de variables aléatoires indépendantes et identiquement distribuées selon  $\mathcal{N}(\mu, \sigma^2 = (2.500km)^2)$ .

Les deux hypothèses de notre problème de test sont :

$H_o : \mu \leq 26.500$  km (on ne change pas de mode de fabrication)

$H_1 : \mu > 26.500$  km

$H_o$  ne sera rejeté en faveur de  $H_1$  que si l'expérience est très en faveur de  $H_1$ .

On fait un test au niveau  $\alpha = 0,05$ .

**Etape 1: choix d'une statistique** On choisit une statistique de test ; sa valeur dans l'échantillon sera une fonction des observations d'échantillon (une fonction de  $x_1, \dots, x_{25}$ ) qui sera utilisée pour prendre une décision. Le fait que  $\bar{X}$  est un estimateur non biaisé à variance minimum de la moyenne  $\mu$  suggère d'utiliser une procédure de test avec règle de décision basée sur  $\bar{x}$ . On introduit la statistique de test

$$Z = \frac{\bar{X} - 26.500}{\sigma/\sqrt{n}} \simeq \frac{\bar{X} - 26.500}{500}.$$

**Etape 2: établissement d'une région de rejet** On veut définir la région de rejet; on dira qu'on rejette  $H_0$  si  $\frac{\bar{x}-26.500}{500} \geq C (> 0)$ . On détermine  $C$  pour avoir une condition de niveau  $\alpha = 0,05$  (i.e. une probabilité d'erreur de type 1  $\leq 0,05$ ). On veut donc que

$$P_\mu[Z \geq C] \leq 0,05 \quad \forall \mu \leq 26.500 km.$$

Remarquons que  $\bar{X} \sim \mathcal{N}(\mu, \sigma^2)$  donc pour  $\mu_0 = 26.500$ , on a:

$$P_{\mu_0}[Z \geq C] = P_{\mu_0}\left[\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \geq C\right] = P[\mathcal{N}(0, 1) \geq C].$$

D'autre part, si  $\mu < 26.500$  on a

$$P_\mu\left[\frac{\bar{X} - 26.500}{\sigma/\sqrt{n}} \geq C\right] \leq P_{26.500}\left[\frac{\bar{X} - 26.500}{\sigma/\sqrt{n}} \geq C\right].$$

En outre

$$P[\mathcal{N}(0, 1) \geq C] \leq 0,05 \text{ ssi } C \geq z_{0,05} = 1,645.$$

On en conclut que pour satisfaire une condition de niveau  $\alpha = 0,05$ , il faut et il suffit que  $C \geq 1,645$ . Maintenant, on veut minimiser la probabilité d'erreur de seconde espèce, donc prendre une région de rejet la plus grande possible (tout en satisfaisant la condition de niveau); on prend donc  $C$  le plus petit possible et on a  $C = z_\alpha = 1,645$

La règle de comportement sera donc la suivante:

$$\text{on rejette } H_0 \text{ ssi } \frac{\bar{x}-26.500}{500} \geq 1,645.$$

**Etape 3: décision avec l'échantillon donné.**

La valeur expérimentale de la moyenne-échantillon est  $\bar{x} = 27.200$  donc la valeur de notre statistique de test est

$$z = \frac{27.200 - 26.500}{500} = \frac{700}{500} = 1,4.$$

Ainsi  $z$  n'est pas dans la région de rejet. La conclusion est que le nouveau procédé ne fournit pas une amélioration de la durée de vie du pneu de la grandeur souhaitée.

La valeur de 1,645 a été choisie pour contrôler la probabilité d'une erreur de type I; le résultat du test dépend de la valeur choisie.

### Calcul de la puissance du test

La puissance du test est égale à  $1 - P[\mathcal{R}H_0 | H_0 \text{ est fausse}]$ . C'est donc une fonction  $\Pi_\mu$  définie pour toute valeur  $\mu > 26.500$  km par:

$$\begin{aligned}\Pi_\mu &= 1 - P_\mu[Z \leq 1,645] = P_\mu[Z > 1,645] \\ &= P_\mu[\bar{X} > 26.500 + 822] \\ &= P\left[\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{26.500 - \mu}{500} + 1,645\right] \\ &= P[\mathcal{N}(0,1) > \frac{26.500 - \mu}{500} + 1,645].\end{aligned}$$

Pour  $\mu = 26.500$  la puissance vaut  $\Pi(\mu) = 0,05$ ;  
pour  $\mu = 27.000$  la puissance vaut  $\Pi(\mu) = P[\mathcal{N}(0,1) > 0,645]$ ;  
pour  $\mu = 27.500$  la puissance vaut  $\Pi(\mu) = P[\mathcal{N}(0,1) > -0,355]$ ;  
pour  $\mu = 28.000$  la puissance vaut  $\Pi(\mu) = P[\mathcal{N}(0,1) > -1,355]$ .

### 8.2.2 Tests sur la moyenne d'une grandeur

On considère une grandeur  $X : \Omega \rightarrow \mathbb{R}$  telle que  $E(X) = \mu$  et  $Var(X) = \sigma^2$ . On ne connaît pas a priori la loi de  $X$ . On ne connaît pas la valeur du paramètre  $\mu$  et on ne connaît pas nécessairement la valeur de  $\sigma^2$ . On veut faire des test d'hypothèse sur le paramètre  $\mu$  à partir de l'information donnée par les valeurs  $(x_1, \dots, x_n)$  d'un échantillon aléatoire de taille  $n$  prélevé dans la population. Ainsi les  $x_i$  sont les valeurs de variables aléatoires  $(X_1, \dots, X_n)$  indépendantes et ayant toutes la même loi de probabilité que  $X$ . Rappelons que les estimateurs usuels de  $\mu$  et  $\sigma^2$  sont la moyenne échantillon

$$\hat{\mu} = \bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$$

et la variance-échantillon

$$\hat{\sigma}^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{X})^2.$$

On fera trois types de test:

$$\left. \begin{array}{ll} \text{Test 1} & H_0 : \mu = \mu_0 \quad (H_1 : \mu \neq \mu_0) \\ \text{Test 2} & H_0 : \mu \geq \mu_0 \quad (H_1 : \mu < \mu_0) \\ \text{Test 3} & H_0 : \mu \leq \mu_0 \quad (H_1 : \mu > \mu_0) \end{array} \right\} \text{ où } \mu_0 \text{ est choisi dans } \mathbb{R}.$$

Le théorème central limite nous dit que pour un grand échantillon ( $n > 30$ ), on a

$$\frac{\hat{X} - \mu}{\sigma/\sqrt{n}} \approx \mathcal{N}(0, 1)$$

et même (ce qui est important si on ne connaît pas la variance)

$$\frac{\hat{X} - \mu}{S/\sqrt{n}} \approx \mathcal{N}(0, 1).$$

D'autre part, si l'on sait que la distribution de  $X$  est une normale, donc si

$$X \sim \mathcal{N}(\mu, \sigma^2),$$

alors pour un  $n$  quelconque, on a

$$\frac{\hat{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1) \quad \text{et} \quad \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

où  $t_{n-1}$  désigne une loi de Student à  $n - 1$  degrés de liberté (car  $\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^n (\frac{X_i - \bar{X}}{\sigma})^2 \sim \chi_{n-1}^2$  est une loi chi-carré).

### Test avec de grands échantillons, quand la variance est connue

On choisit comme statistique de test

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}.$$

Si la valeur de  $\mu$  est  $\mu_0$ , on sait que  $Z \approx \mathcal{N}(0, 1)$ . On raisonne sur la région de rejet comme dans les cas précédents et on a

| Tests d'hypothèse                                | Règle de comportement<br>au niveau $\alpha$        |
|--|--|
| $H_0 : \mu \leq \mu_0 \quad (H_1 : \mu > \mu_0)$ | $RH_0$ si $z > z_\alpha$                           |
| $H_0 : \mu \geq \mu_0 \quad (H_1 : \mu < \mu_0)$ | $RH_0$ si $z < -z_\alpha$                          |
| $H_0 : \mu = \mu_0 \quad (H_1 : \mu \neq \mu_0)$ | $RH_0$ si $z \notin [-z_{\alpha/2}, z_{\alpha/2}]$ |

### Test avec de grands échantillons, quand la variance est inconnue

On choisit comme statistique de test

$$Z' = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}.$$

Si la valeur de  $\mu$  est  $\mu_0$ , on sait que  $Z' \approx \mathcal{N}(0, 1)$ . Alors

| Tests d'hypothèse                                | Règle de comportement<br>au niveau $\alpha$         |
|--|---|
| $H_0 : \mu \leq \mu_0 \quad (H_1 : \mu > \mu_0)$ | $RH_0$ si $z' > z_\alpha$                           |
| $H_0 : \mu \geq \mu_0 \quad (H_1 : \mu < \mu_0)$ | $RH_0$ si $z' < -z_\alpha$                          |
| $H_0 : \mu = \mu_0 \quad (H_1 : \mu \neq \mu_0)$ | $RH_0$ si $z' \notin [-z_{\alpha/2}, z_{\alpha/2}]$ |

### Test avec de petits échantillons, quand la distribution est normale de variance connue

On choisit comme statistique de test

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}.$$

Si la valeur de  $\mu$  est  $\mu_0$ , on sait que  $Z \sim \mathcal{N}(0, 1)$ . Donc

| Tests d'hypothèse                                | Règle de comportement<br>au niveau $\alpha$        |
|--|--|
| $H_0 : \mu \leq \mu_0 \quad (H_1 : \mu > \mu_0)$ | $RH_0$ si $z > z_\alpha$                           |
| $H_0 : \mu \geq \mu_0 \quad (H_1 : \mu < \mu_0)$ | $RH_0$ si $z < -z_\alpha$                          |
| $H_0 : \mu = \mu_0 \quad (H_1 : \mu \neq \mu_0)$ | $RH_0$ si $z \notin [-z_{\alpha/2}, z_{\alpha/2}]$ |

### Test avec de petits échantillons, quand la distribution est normale de variance inconnue

On choisit comme statistique de test

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}.$$

Si la valeur de  $\mu$  est  $\mu_0$ , on sait que  $T \sim t_{n-1}$ . Donc

| Tests d'hypothèse                                | Règle de comportement<br>au niveau $\alpha$                |
|--|--|
| $H_0 : \mu \leq \mu_0 \quad (H_1 : \mu > \mu_0)$ | $RH_0$ si $t > t_{n-1;\alpha}$                             |
| $H_0 : \mu \geq \mu_0 \quad (H_1 : \mu < \mu_0)$ | $RH_0$ si $t < -t_{n-1;\alpha}$                            |
| $H_0 : \mu = \mu_0 \quad (H_1 : \mu \neq \mu_0)$ | $RH_0$ si $t \notin [-t_{n-1;\alpha/2}, t_{n-1;\alpha/2}]$ |

où l'on a défini comme précédemment  $t_{m;\alpha}$  (avec  $m$  entier  $> 0$  et  $\alpha \in [0, 1]$ ) comme le nombre réel tel que la fonction de répartition d'une variable aléatoire de Student à  $m$  degrés de libertés soit égale à  $1 - \alpha$  en ce point, donc

$$P[t_m > t_{m;\alpha}] = \alpha.$$

Faites attention, la convention consiste souvent à noter ce nombre réel  $t_{m;1-\alpha}$ ; nous avons choisi ici une notation uniforme pour toutes les lois de probabilité en nous alignant sur les conventions de notations adoptées pour la loi normale standard.

**Pour de petits échantillons d'une population quelconque, on utilise des tests de rang (qui ne seront pas vus ici!).**

# Chapitre 9

## Comparaisons de deux populations

### 9.1 Comparaison de deux proportions

Une expérience faite à Harvard avait pour but de décider si la prise régulière d'aspirine était efficace dans la prévention des crises cardiaques. Bien sûr, comme dans la plupart de tels essais cliniques, la probabilité qu'un individu déterminé ait une crise cardiaque dans une année déterminée est (heureusement!) très faible. Mais on souhaite une réponse rapide. Que faire?

La solution simple (mais chère!) est de tester un grand nombre d'individus. Dans cette étude, 22.071 individus furent testés (tous des docteurs); ils furent répartis aléatoirement en deux groupes. Le premier groupe reçut un placebo (une pillule identique à l'aspirine mais sans aspirine), le deuxième groupe reçut une aspirine par jour. Après une période d'environ 5 ans (l'étude fut interrompue avant le délai initialement prévu vu son résultat très positif; il eût été injuste de ne pas faire profiter des résultats le groupe prenant le placebo), on enregistre les résultats suivants concernant le nombre d'individus ayant une crise cardiaque (fatale ou non) durant cette période.

|          | Crise card. | Pas de crise card. | n      | Proportion d'attaque                     |
|----------|-------------|--------------------|--------|--|
| Placebo  | 239         | 10.795             | 11.034 | $\hat{p}_1 = \frac{239}{11.034} = 0,021$ |
| Aspirine | 139         | 10.898             | 11.037 | $\hat{p}_2 = \frac{139}{11.037} = 0,012$ |

La différence de proportion est petite  $\hat{p}_1 - \hat{p}_2 = 0,0091$  mais le risque relatif  $\hat{p}_1/\hat{p}_2 = 1,72$  donc les gens prenant le placebo ont 1,72 fois plus de "chance"



d'avoir une crise cardiaque.

### 9.1.1 Le modèle lors d'une comparaison de deux proportions

On a deux échantillons aléatoires indépendants de deux populations binomiales. Dans notre exemple, la population 1 est celle des individus ne prenant pas d'aspirine. L'échantillon est constitué des personnes prenant le placebo. La population 2 est celle des individus prenant de l'aspirine; l'échantillon est constitué des personnes auxquelles on en donne.

La chance de "succès" dans la première population est  $p_1$ , la chance de "succès" dans la deuxième population est  $p_2$ .

On a donc des variables aléatoires  $(X_1, \dots, X_{n_1})$  et des variables aléatoires  $(Y_1, \dots, Y_{n_2})$  toutes indépendantes, avec tous les  $X_i \sim \text{Bin}(1, p_1)$  and tous les  $Y_j \sim \text{Bin}(1, p_2)$ .

L'estimateur de  $p_1$  est donné par  $\hat{P}_1 = \frac{X}{n_1}$  où  $X = X_1 + \dots + X_{n_1} \sim \text{Bin}(n_1, p_1)$  mesure le nombre de succès dans le premier échantillon avec  $n_1$  le nombre d'individus dans le premier échantillon.

L'estimateur de  $p_2$  est donné par  $\hat{P}_2 = \frac{Y}{n_2}$  où  $Y = Y_1 + \dots + Y_{n_2} \sim \text{Bin}(n_2, p_2)$  mesure le nombre de succès dans le deuxième échantillon avec  $n_2$  le nombre d'individus dans le deuxième échantillon.

Un estimateur pour la différence  $p_1 - p_2$  est donné par  $\hat{P}_1 - \hat{P}_2$

**Proposition 9.1** *Soit  $X \sim \text{Bin}(n_1, p_1)$  et  $Y \sim \text{Bin}(n_2, p_2)$  deux variables aléatoires indépendantes. Alors*

$$E\left(\frac{X}{n_1} - \frac{Y}{n_2}\right) = E(\hat{P}_1 - \hat{P}_2) = p_1 - p_2$$

$$\text{Var}(\hat{P}_1 - \hat{P}_2) = \text{Var}\left(\frac{X}{n_1} - \frac{Y}{n_2}\right) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}.$$

Pour de grands échantillon  $\hat{P}_1 - \hat{P}_2$  est approximativement normalement distribué et on a

$$Z = \frac{\hat{P}_1 - \hat{P}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}} \sim \mathcal{N}(0, 1).$$

### 9.1.2 Intervalle de confiance pour $p_1 - p_2$

Un intervalle de confiance au niveau  $1 - \alpha$  pour  $p_1 - p_2$ , est donné par

$$[ \hat{p}_1 - \hat{p}_2 - D_\alpha, \hat{p}_1 - \hat{p}_2 + D_\alpha ]$$

avec

$$D_\alpha = z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

où  $\hat{p}_1 - \hat{p}_2$  sont les valeurs observées de  $\hat{p}_1$  et  $\hat{p}_2$  dans les échantillons, et où comme précédemment  $z_{\alpha/2}$  est le nombre réel (obtenu dans les tables de la fonction de répartition d'une loi normale standard) tel que

$$\phi(z_{\alpha/2}) = P[\mathcal{N}(0, 1) \leq z_{\alpha/2}] = 1 - \alpha.$$

La probabilité pour que la vraie valeur de  $p_1 - p_2$  soit dans l'intervalle est supérieure ou égale à  $1 - \alpha$

Dans l'étude des aspirines  $\sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} = 0,00175$

Donc un intervalle de confiance, au niveau 95%, de  $p_1 - p_2$  est donné par  $[0,0091 - 0,0034, 0,0091 + 0,0034]$  car ( $z_{0,5\%} = 1,96$ )

On peut conclure qu'on est sûr à 95% que la différence entre "chance" d'avoir une crise cardiaque est entre 0,0057 et 0,0125 ce qui est un nombre positif. On est donc sûr à 95% que l'aspirine diminue réellement le risque de crise cardiaque!

### 9.1.3 Tests d'hypothèse pour $p_1 - p_2$

La question formelle est la suivante: si l'aspirine n'avait pas d'effet, quelle est la probabilité que les résultats mesurés soient dûs au simple hasard?

On considèrera pour cette question, un test dans lequel  $H_0$ , l'hypothèse nulle, est que  $p_1 \leq p_2$  (dans notre cas : l'aspirine n'a pas d'effet) et  $H_1$ , l'hypothèse alternative, est que  $p_1 > p_2$ .

De manière plus générale, on considère trois types de test: ceux pour lesquels  $H_0$  consiste à dire que  $p_1 = p_2$ , ceux pour lesquels  $H_0$  consiste à dire que  $p_1 \leq p_2$ , et ceux pour lesquels  $H_0$  consiste à dire que  $p_1 \geq p_2$ .

On impose une condition de niveau à la probabilité d'erreur de première espèce; on va donc imposer que la probabilité de rejeter l'hypothèse  $H_0$  soit inférieure à  $\alpha$  dès que les paramètres  $p_1$  et  $p_2$  sont tels que  $H_0$  est vraie. Comme précédemment, cette condition de niveau sera satisfaite si et seulement si elle est satisfaite pour la distribution de la statistique correspondant au cas limite des valeurs  $(p_1, p_2)$  satisfaisant  $H_0$  donc, pour les trois types de test, quand  $p_1 = p_2$ .

Sous l'hypothèse  $p_1 = p_2 = p$  on a

$$E(\hat{P}_1 - \hat{P}_2) = 0 \quad \text{et} \quad \text{Var}(\hat{P}_1 - \hat{P}_2) = p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$

Si  $n_1$  et  $n_2$  sont tous les deux grands,  $\hat{P}_1$  et  $\hat{P}_2$  ont approximativement une distribution normale, donc  $\hat{P}_1 - \hat{P}_2$  est également approximativement normale.

$$\frac{\hat{P}_1 - \hat{P}_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim \mathcal{N}(0, 1)$$

Attention, le dénominateur dépend de  $p$ ; pour avoir une statistique de test, on remplace  $p$  par son estimateur  $\hat{P}$ ; ici on a deux échantillons indépendants de deux populations où la proportion  $p$  est la même; on le regarde comme un grand échantillon d'une population et:

$$\hat{P} = \frac{X + Y}{n_1 + n_2} = \frac{n_1 \hat{P}_1 + n_2 \hat{P}_2}{n_1 + n_2}$$

La **statistique de test** est

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \approx \mathcal{N}(0, 1)$$

et la **règle de comportement** est déterminée par les **régions de rejet** suivantes:

| $H_0$          | Hypothèse alternative | Région de rejet de $H_0$ avec condition de niveau $x$ |
|----------------|-----------------------|---|
| $p_1 = p_2$    | $p_1 \neq p_2$        | $z \geq z_{x/2}$ ou $z \leq -z_{x/2}$                 |
| $p_1 \leq p_2$ | $p_1 > p_2$           | $z \geq z_x$  |
| $p_1 \geq p_2$ | $p_1 < p_2$           | $z \leq -z_x$   |

**Attention:** Les deux échantillons ont été supposés indépendants. Nous verrons plus loin comment traiter des observations (binomiales) non indépendantes.

## 9.2 Comparaison de deux moyennes

On considère une grandeur  $X$  de moyenne  $\mu_1$  et de variance  $\sigma_1^2$  et une grandeur  $Y$  de moyenne  $\mu_2$  et de variance  $\sigma_2^2$ . Les moyennes sont inconnues et les variances sont connues ou inconnues. On veut de l'information sur la différence des moyennes à partir des données d'échantillons.

Soient  $(x_1, \dots, x_{n_1})$  les valeurs mesurées de la grandeur  $X$  dans un échantillon de taille  $n_1$ , donc les valeurs de variables aléatoires  $X_1, \dots, X_{n_1}$ , indépendantes, identiquement distribuées suivant la loi de  $X$  donc de moyenne  $\mu_1$  et de variance  $\sigma_1^2$ . Soient  $(y_1, \dots, y_{n_2})$  les valeurs mesurées de la grandeur  $Y$  dans un échantillon de taille  $n_2$ , donc les valeurs de variables aléatoires  $Y_1, \dots, Y_{n_2}$ , indépendantes, identiquement distribuées suivant la loi de  $Y$  donc de moyenne  $\mu_2$  et de variance  $\sigma_2^2$ . On suppose les deux échantillons indépendants donc toutes les variables  $X_i$  et  $Y_j$  sont indépendantes.

L'estimateur "naturel" de  $\mu_1 - \mu_2$  est

$$\bar{X} - \bar{Y} \quad \text{avec} \quad \bar{X} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i \quad \text{et} \quad \bar{Y} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i.$$

On a  $E(\bar{X} - \bar{Y}) = \mu_1 - \mu_2$  et  $Var(\bar{X} - \bar{Y}) = Var(\bar{X}) + Var(\bar{Y}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$ .

**Cas des grands échantillons** ( $n_1 > 30, n_2 > 30$ )

Si les variances sont connues, on a, par le théorème central-limite

$$Z := \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \approx \mathcal{N}(0, 1).$$

Si les variances ne sont pas connues, en notant  $S_1^2$  et  $S_2^2$  les estimateurs variance-échantillon – donc  $S_1^2 = \frac{1}{n_1-1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2$  et similairement pour  $S_2^2$  –, on a encore, comme les échantillons sont grands:

$$Z := \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \approx \mathcal{N}(0, 1.)$$

### Cas des petits échantillons pour des distributions $X$ et $Y$ normales

Si les variances sont connues, comme la somme de distributions normales indépendantes est une distribution normale, on a

$$Z := \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0, 1).$$

Si les variances ne sont pas connues, on SUPPOSE EN PLUS QUE CES VARIANCES SONT EGALES  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ ; on a alors  $Var(\bar{X} - \bar{Y}) = \sigma^2(\frac{1}{n_1} + \frac{1}{n_2})$ . On définit un estimateur non biaisé de cette variance par

$$S^2 := \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

On montre que la variable

$$T := \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{S^2(\frac{1}{n_1} + \frac{1}{n_2})}} \sim t_{n_1 + n_2 - 2}$$

a la distribution d'une variable de Student à  $n_1 + n_2 - 2$  degrés de liberté.

#### 9.2.1 Intervalle de confiance au niveau $(1 - \alpha)$ pour $\mu_1 - \mu_2$

Si les échantillons sont grands et si les variances sont connues, un tel intervalle est donné par

$$\left[ \bar{x} - \bar{y} - z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \bar{x} - \bar{y} + z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right].$$

Il en est de même si les échantillons sont petits mais les distributions sont normales (ou approximativement normales, en tout cas en forme de "cloche") et si les variances sont connues.

Si les échantillons sont grands et si les variances sont inconnues, un intervalle de confiance ou niveau  $(1 - \alpha)$  pour  $\mu_1 - \mu_2$  est donné par

$$\left[ \bar{x} - \bar{y} - z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \bar{x} - \bar{y} + z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right].$$

Si les échantillons sont petits et si les variances sont inconnues, lorsque les distributions sont normales et lorsque les variances sont égales, un intervalle de confiance ou niveau  $(1 - \alpha)$  pour  $\mu_1 - \mu_2$  est donné par

$$\left[ \bar{x} - \bar{y} - t_{n_1+n_2-2; \alpha/2} \sqrt{s^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}, \bar{x} - \bar{y} + t_{n_1+n_2-2; \alpha/2} \sqrt{s^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \right]$$

où, comme précédemment,  $t_{m; \beta}$  est le nombre réel tel que

$$P[t_m > t_{m; \beta}] = \beta.$$

[Attention, les conventions changent d'un livre à l'autre. Ce nombre est souvent noté  $t_{m; 1-\beta}$ .]

### 9.2.2 Test au niveau $\alpha$ pour $\mu_1 - \mu_2$

On fera trois types de test:

$$\left. \begin{array}{l} \text{Test 1} \quad H_0 : \mu_1 \leq \mu_2 \quad (H_1 : \mu_1 > \mu_2) \\ \text{Test 2} \quad H_0 : \mu_1 \geq \mu_2 \quad (H_1 : \mu_1 < \mu_2) \\ \text{Test 1} \quad H_0 : \mu_1 = \mu_2 \quad (H_1 : \mu_1 \neq \mu_2) \end{array} \right\}$$

**Cas des variances connues, soit pour grands échantillons, soit pour échantillons quelconques de grandeurs normales**

| Tests de comparaisons de moyennes                             | Statistique de test   | Règle de comportement au niveau $\alpha$           |
|---|---|--|
| $H_0 : \mu_1 - \mu_2 \leq 0$<br>( $H_1 : \mu_1 - \mu_2 > 0$ ) | $Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$<br><br>( si $\mu_1 = \mu_2$ , $Z \sim \mathcal{N}(0, 1)$ ) | $RH_0$ si $z > z_\alpha$                           |
| $H_0 : \mu_1 - \mu_2 \geq 0$<br>( $H_1 : \mu_1 - \mu_2 < 0$ ) |   | $RH_0$ si $z < -z_\alpha$                          |
| $H_0 : \mu - \mu_2 = 0$<br>( $H_1 : \mu_1 \neq \mu_2$ )       |   | $RH_0$ si $z \notin [-z_{\alpha/2}, z_{\alpha/2}]$ |

### Cas des variances inconnues pour grands échantillons

| Tests de comparaisons de moyennes                           | Statistique de test   | Règle de comportement au niveau $\alpha$           |
|---|---|--|
| $H_0 : \mu_1 - \mu_2 \leq 0$<br>$(H_1 : \mu_1 - \mu_2 > 0)$ | $Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$<br>( si $\mu_1 = \mu_2$ , $Z \sim \mathcal{N}(0, 1)$ ) | $RH_0$ si $z > z_\alpha$                           |
| $H_0 : \mu_1 - \mu_2 \geq 0$<br>$(H_1 : \mu_1 - \mu_2 < 0)$ |   | $RH_0$ si $z < -z_\alpha$                          |
| $H_0 : \mu_1 - \mu_2 = 0$<br>$(H_1 : \mu_1 - \mu_2 \neq 0)$ |   | $RH_0$ si $z \notin [-z_{\alpha/2}, z_{\alpha/2}]$ |

### Cas des variances inconnues mais égales, pour petits échantillons de grandeurs normales

| Tests de comparaisons de moyennes                           | Statistique de test  | Règle de comportement au niveau $\alpha$  |
|---|--|---|
| $H_0 : \mu_1 - \mu_2 \leq 0$<br>$(H_1 : \mu_1 - \mu_2 > 0)$ | $T = \frac{\bar{X} - \bar{Y}}{\sqrt{S^2(\frac{1}{n_1} + \frac{1}{n_2})}}$<br>( si $\mu_1 = \mu_2$ , $T \sim t_{n_1+n_2-2}$ ) | $RH_0$ si $t > t_{n_1+n_2-2; \alpha}$     |
| $H_0 : \mu_1 - \mu_2 \geq 0$<br>$(H_1 : \mu_1 - \mu_2 < 0)$ |  | $RH_0$ si $t < -t_{n_1+n_2-2; \alpha}$    |
| $H_0 : \mu_1 - \mu_2 = 0$<br>$(H_1 : \mu_1 - \mu_2 \neq 0)$ |  | $RH_0$ si $ t  > t_{n_1+n_2-2; \alpha/2}$ |

**Exemple 9.1** La proximité d'une route à trafic élevé a-t-elle une influence sur la teneur en plomb du sang? (Arch. Environmental health, 1967).

On considère les mesures effectuées sur un premier groupe de  $n_1 = 35$  femmes vivant en bordure d'autoroute; on a

$$\bar{x} = 16,7 \quad s_1^2 = 49,0,$$

et les mesures effectuées sur un deuxième groupe de  $n_2 = 30$  femmes vivant à la campagne loin d'une autoroute où

$$\bar{y} = 9,9 \quad s_2^2 = 24,01.$$

On veut tester:  $H_0 : \mu_1 - \mu_2 \leq 0$ , i.e. la teneur n'est pas plus élevée si on vit près d'une autoroute, contre  $H_1 : \mu_1 - \mu_2 > 0$ . On a donc deux grands échantillons, des variances inconnues, et on calcule la valeur de la statistique

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{s_1^2/35 + s_2^2/30}} = \frac{6,80}{1,483} = 4,58$$

Puisque  $z_{0,01} = 2,33$  on rejette  $H_0$  car  $(4,58 > 2,33!)$  au niveau 1%.

**Exemple 9.2** On veut tester les effets d'une exposition à l'ozone chez les rats. Un échantillon de 20 rats est plongé dans une atmosphère d'ozone à 2% pendant 30 jours. Leur volume pulmonaire mesuré après ces 30 jours est en moyenne de 9,28 ml avec un écart-type de  $\sqrt{0,37}$  ml. Un groupe contrôle de 17 rats de mêmes caractéristiques initiales fournit un volume pulmonaire moyen de 7,97 ml et un écart-type de  $\sqrt{0,41}$  ml. Ces données mettent-elles en évidence une action positive de l'ozone sur le volume pulmonaire? on a donc

$$\bar{x} = 9,28ml \quad \bar{y} = 7,97ml \quad s_1^2 = 0,37ml^2 \quad s_2^2 = 0,41ml^2$$

et on veut tester l'hypothèse

$$H_0 : \mu_1 - \mu_2 \leq 0 \text{ contre } H_1 : \mu_1 - \mu_2 > 0.$$

On suppose que les variances sont égales et que les distributions sont normales; l'estimateur sans biais de la variance commune  $\sigma^2$  vaut  $S^2 = 0,151ml^2$  donc  $S = 0,39ml$

La statistique de test vaut

$$t = \frac{9,28 - 7,97}{0,39\sqrt{1/20 + 1/17}} = 10,18.$$

Il y a  $n_1 + n_2 - 2 = 35$  degrés de liberté. Le quantile d'ordre 0,99 de la variable de Student à 35 degrés de liberté est

$$t_{35;0,01} = 2,44$$

Donc puisque  $10,18 > 2,44$  on rejette  $H_0$ !

Le problème de la comparaison de deux moyennes dans le cas où les variances ne sont pas égales soulève des questions théoriques fondamentales qui ne peuvent être résolues de manière totalement satisfaisante. On trouve certaines procédures suggérées dans les livres de statistiques, en particulier si l'on veut juste tester si les moyennes sont égales.



# Chapitre 10

## Tests chi-carré ( $\chi^2$ )

### 10.1 Variables normales, chi-carré et de Student

**Définition 10.1** Une variable aléatoire continue  $X$  a une loi de probabilité chi-carré de paramètre  $n$  (où  $n$  est un entier positif), et on note  $X \sim \chi_n^2$ , si c'est une variable aléatoire continue dont la densité de probabilité est :

$$f(x, n) = \frac{1}{2^{n/2} \Gamma(n/2)} x^{n/2-1} e^{-\frac{x}{2}} \quad \text{pour } x \geq 0$$
$$= 0 \quad \text{pour } x < 0.$$

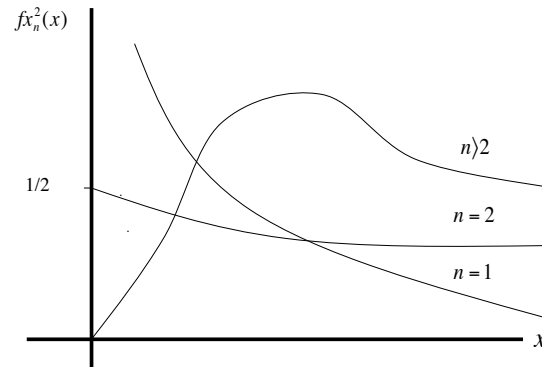
avec  $\Gamma(1/2) = \sqrt{\pi}$ ,  $\Gamma(1) = 1$  and  $\Gamma(\beta) = (\beta - 1)\Gamma(\beta - 1) \forall \beta$ .

Le paramètre  $n$  est appelé le nombre de degrés de liberté de  $X$ .

La moyenne et la variance d'une loi chi-carré sont:

$$E(\chi_n) = n \quad \text{Var}(\chi_n) = 2n$$

Les densités de probabilité des variables chi-carré sont de la forme:



**Théorème 10.1** -Si  $Y$  a une distribution normale standard, alors  $Y^2 \sim \chi_1^2$ .  
 -Plus généralement, si  $(X_1, \dots, X_n)$  sont  $n$  variables aléatoires indépendantes suivant toutes la loi normale standard  $\mathcal{N}(0, 1)$ , alors la somme de leur carré suit une loi de probabilité chi-carré à  $n$  degrés de liberté:

$$(X_1)^2 + \dots + (X_n)^2 \sim \chi_n^2$$

-Si  $X$  et  $Y$  sont deux variables aléatoires chi-carré indépendantes  $X \sim \chi_n^2$  et  $Y \sim \chi_m^2$  alors leur somme est une chi-carré :

$$X + Y \sim \chi_{n+m}^2.$$

- Si  $X$  et  $Y$  sont deux variables aléatoires indépendantes, telles que  $X \sim \mathcal{N}(0, 1)$  et  $Y \sim \chi_n^2$ , la loi suivie par  $T := \frac{X}{\sqrt{Y/n}}$  est la loi de Student à  $n$  degrés de liberté, notée  $t_n$ :

$$\frac{X}{\sqrt{Y/n}} \sim t_n$$

La densité d' une loi de Student  $t_n$  à  $n$  degrés de liberté est :

$f(x) = c_{(n)}^{ste} (1 + \frac{x^2}{n})^{-\frac{n+1}{2}}$ ; on a  $E(t_n) = 0$  et  $Var(T_n) = \frac{n}{n-2}$  quand  $(n > 2)$ .  
 D'autre part, quand  $n$  est grand, la loi de  $t_n$  tend vers une normale standard:  
 $t_n \rightarrow \mathcal{N}(0, 1)$  quand  $n \rightarrow \infty$ .

Nous allons présenter trois types de tests utilisant les lois chi-carré: des tests d'ajustement, des tests d'homogénéité et des tests d'indépendance.

## 10.2 Les tests $\chi^2$ d'ajustement

Supposons, lors d'une expérience aléatoire  $E$ , mesurer une grandeur prenant un nombre fini  $I$  de résultats possibles  $(r_1, \dots, r_I)$  avec les probabilités  $(p_1, \dots, p_I)$  (bien-sûr  $p_1 + \dots + p_I = 1$ ).

On répète cette expérience  $n$  fois dans des conditions uniformes et de manière indépendante et on mesure les nombres  $(n_1, \dots, n_I)$  de fois que l'on a obtenu les différents résultats :  $n_j$  = nombre de fois qu'on a eu le résultat  $r_j$  ; (bien sûr  $n_1 + \dots + n_I = n$ ).

Chacun des  $n_i$  est la mesure d'une variable aléatoire  $N_j$  qui a la distribution d'une binomiale  $\text{Bin}(n, p_i)$ ; attention, les variables  $N_1, N_2, \dots, N_j$  ne sont **pas** indépendantes.

Remarquons que  $E(N_i) = np_i$  et  $\text{Var}(N_i) = np_i(1 - p_i)$ .

La variable  $\vec{N} = (N_1, \dots, N_I)$ , à  $I$  dimensions, est appelée variable multinomiale.

**Théorème 10.2** (non prouvé ici!) La variable aléatoire  $Q = \sum_{i=1}^I \frac{(N_i - np_i)^2}{np_i}$  a une loi approximative de distribution  $\chi^2_{I-1}$ :

$$Q = \sum_{i=1}^I \frac{(N_i - np_i)^2}{np_i} \approx \chi^2_{I-1}$$

L'approximation considérée est valide si

$$\begin{aligned} n &\geq 0 \\ np_i &> 0 & \forall i = 1, \dots, I \\ np_i &\geq 5 & \text{dans } 80\% \text{ des cas au moins.} \end{aligned}$$

**Remarque 10.1** La valeur  $q$  de la variable  $Q$  dans un échantillon donné est égale à

$$q = \sum_{i=1}^I \frac{(\text{effectifs observés}_i - \text{effectifs théoriques}_i)^2}{\text{effectifs théoriques}_i}$$

Le **test  $\chi^2$  d'ajustement** consiste à tester si les valeurs des probabilités  $p_1, p_2, \dots, p_I$  sont des quantités prédéterminées  $p_1^\circ, p_2^\circ, \dots, p_I^\circ$ . Donc

$$H_0 : p_1 = p_1^\circ, p_2 = p_2^\circ, \dots, p_I = p_I^\circ$$

$$H_1 : \exists i \text{ tel que } p_i \neq p_i^\circ$$

La Statistique de test est

$$Q = \sum_{i=1}^I \frac{(N_i - np_i^\circ)^2}{np_i^\circ}.$$

On sait par le théorème mentionné que, sous l'hypothèse  $H_0$ , on a  $Q \sim \chi_{I-1}^2$ . Dès lors, la **règle de comportement**, pour ce test d'ajustement au niveau  $\alpha$  est:

$$RH_0 \text{ si } q > \chi_{I-1;\alpha}^2.$$

Comme pour les autres lois de probabilité, on a noté  $\chi_{m;\alpha}^2$  le nombre réel tel que

$$P[\chi_m^2 > \chi_{m;\alpha}^2] = \alpha.$$

Faites attention que ceci n'est pas une notation standard!

### Exemple 1 : expérience de Mendel

Un échantillon de 1611 graines appartenant à la seconde génération d'un croisement entre pois verts et lisses et pois jaunes et ridés donne les résultats suivants:

|       | lisse       | ridé        |
|-------|-------------|-------------|
| vert  | $n_1 = 926$ | $n_2 = 288$ |
| jaune | $n_3 = 293$ | $n_4 = 104$ |

Ces résultats sont-ils compatibles avec les thèses de Mendel?

Les thèses de Mendel assignent à chacun des 4 cas ci-dessous les probabilités  $p_i^\circ$  suivantes:

|       | lisse                      | ridé                       |
|-------|----------------------------|----------------------------|
| vert  | $p_1^\circ = \frac{9}{16}$ | $p_2^\circ = \frac{3}{16}$ |
| jaune | $p_3^\circ = \frac{3}{16}$ | $p_4^\circ = \frac{1}{16}$ |

Les “effectifs théoriques”,  $np_i^\circ$ , sont:

|       | lisse   | ridé  |
|-------|---|---|
| vert  | $np_1^\circ = 1611 \times \frac{9}{16} = 906,2$ | $np_2^\circ = 1611 \times \frac{3}{16} = 302,1$ |
| jaune | $np_3^\circ = 1611 \times \frac{3}{16} = 302,1$ | $np_4^\circ = 1611 \times \frac{1}{16} = 100,7$ |

On obtient

$$q = \frac{(926 - 906,2)^2}{906,2} + \frac{(288 - 302,1)^2}{302,1} + \frac{(293 - 302,1)^2}{302,1} + \frac{(104 - 100,7)^2}{100,7} = 1,473$$

Dans ce problème, il y a  $4 - 1 = 3$  degrés de liberté. Comme  $\chi_{3;0,1}^2 = 6,251$ , on ne rejette pas l’hypothèse que les lois de Mendel sont vérifiées, même au niveau  $\alpha = 10\%$ .

### 10.2.1 Tests d’ajustement avec paramètres

On veut étudier le cas où les  $p_i^\circ$  ne sont pas complètement spécifiées, mais sont des fonctions données de paramètres inconnus  $p_i^\circ = p_i^\circ(\theta_1, \dots, \theta_k)$ .

**Proposition 10.1** *Soit  $\vec{N} = (N_1, \dots, N_I)$  une variable multinomiale à  $I$  dimensions telle que  $N_i \sim \text{Bin}(n, p_i^\circ(\theta_1, \dots, \theta_k))$ . Alors*

$$Q = \sum_{i=1}^I \frac{(N_i - np_i^\circ(\hat{\Theta}_1, \dots, \hat{\Theta}_k))^2}{np_i^\circ(\hat{\Theta}_1, \dots, \hat{\Theta}_k)} \approx \chi_{I-k-1}^2$$

si les  $\hat{\Theta}_i$  sont les estimateurs des paramètres  $\theta_i$  donnés par le maximum de vraisemblance dans la loi multinomiale qui gouverne l’observation de la variable  $\vec{N} = (N_1, \dots, N_I)$ . On notera  $\hat{P}_i := p_i^\circ(\hat{\Theta}_1, \dots, \hat{\Theta}_k)$ . L’approximation considérée est valide si, comme précédemment, on regarde un échantillon avec  $n \geq 0, n\hat{p}_i > 0 \forall i = 1, \dots, I$  et  $n\hat{p}_i \geq 5$  dans 80% des cas au moins.

En bref, chaque paramètre estimé  $\theta_i$  “coûte” un degré de liberté.

Le **test  $\chi^2$  d’ajustement en présence de paramètres** consiste à tester si les valeurs des probabilités  $p_1, p_2, \dots, p_I$  sont de la forme  $p_i^\circ = p_i^\circ(\theta_1, \dots, \theta_k)$  avec des valeurs de  $\theta_1 \dots, \theta_k$  inconnues.

La Statistique de test est

$$Q = \sum_{i=1}^I \frac{(N_i - np_i^\circ(\hat{\Theta}_1, \dots, \hat{\Theta}_k))^2}{np_i^\circ(\hat{\Theta}_1, \dots, \hat{\Theta}_k)}$$

où les  $\hat{\theta}_i$  sont les estimateurs des paramètres  $\theta_i$  donnés par le maximum de vraisemblance. Donc, sous l'hypothèse  $H_0$ , on a  $Q \approx \chi^2_{I-k-1}$ .

**Remarque 10.2** La valeur  $q$  de la variable  $Q$  dans un échantillon donné est égale à

$$q = \sum_{i=1}^I \frac{(\text{effectifs observés}_i - \text{effectifs théoriques estimés}_i)^2}{\text{effectifs théoriques estimés}_i}$$

La règle de comportement, pour ce test d'ajustement au niveau  $\alpha$  est:

$$RH_0 \text{ si } q > \chi^2_{I-k-1; \alpha}.$$

### Exemple 2 Principe de Hardy-Weinberg

Le groupe sanguin  $MN$  se compose de trois phénotypes distincts:  $M$ ,  $MN$  et  $N$ . Une population d'individus appartenant à ce groupe est stable dans sa composition si les proportions relatives à ces trois phénotypes sont de la forme:

$$\begin{aligned} p(M) &= p_1(\theta) = \theta^2 \\ p(MN) &= p_2(\theta) = 2\theta(1 - \theta) \quad \text{où } \theta \in (0, 1) \\ p(N) &= p_3(\theta) = (1 - \theta)^2 \end{aligned}$$

Un échantillon de taille  $n = 500$  est prélevé dans une population. Les valeurs de l'échantillon se répartissent ainsi

$$\begin{array}{ccc} M & MN & N \\ n_1 = 125 & n_2 = 225 & n_3 = 150 \end{array}$$

Ces observations sont-elles compatibles avec l'hypothèse que la population est stable dans sa composition?

Ici on doit estimer un paramètre inconnu  $\theta$ . L'estimation maximum de vraisemblance  $\hat{\theta}$  de  $\theta$  est la valeur  $\hat{\theta}$  de  $\theta$  qui maximise la probabilité, sous la condition  $\theta$ , que l'on ait observé  $n_1$  individus  $M$ ,  $n_2$   $MN$  et  $n_3$  donc

$$P_{\theta}(n_1, n_2, n_3) = c^{ste} (p_1(\theta))^{n_1} (p_2(\theta))^{n_2} (p_3(\theta))^{n_3} = c^{ste} \theta^{2n_1} (2\theta(1-\theta))^{n_2} (1-\theta)^{2n_3}$$

ou de façon équivalente, la valeur de  $\theta$  qui maximise

$$\log P_\theta(n_1, n_2, n_3) = (2n_1 + n_2)\log \theta + (n_2 + 2n_3)\log (1 - \theta) + c^{ste}$$

La dérivée de  $\log P_\theta(n_1, n_2, n_3)$  par rapport à  $\theta$  s'annule en un maximum, ce qui donne:

$$\frac{2n_1 + n_2}{\hat{\theta}} - \frac{n_2 + 2n_3}{1 - \hat{\theta}} = 0$$

donc

$$\hat{\theta} = \frac{2n_1 + n_2}{2n_1 + 2n_2 + 2n_3} = \frac{2n_1 + n_2}{2n}$$

qui prend ici la valeur

$$\hat{\theta} = (2X125 + 225)/2X500 = \frac{475}{1000}$$

donc

$$\begin{aligned}\hat{p}_1^\circ &= p_1^\circ(\hat{\theta}) = (\hat{\theta})^2 = \left(\frac{475}{1000}\right)^2 = 0,225625 \\ \hat{p}_2^\circ &= p_2^\circ(\hat{\theta}) = 2\theta(1 - \hat{\theta}) = 2X0,475X0,525 = 0,498750 \\ \hat{p}_3^\circ &= p_3^\circ(\hat{\theta}) = (1 - \hat{\theta})^2 = (0,525)^2 = 0,275625\end{aligned}$$

Ce qui donne les “effectifs théorique estimés”  $n\hat{p}_i$ :

$$n\hat{p}_1 = 112,81 \quad n\hat{p}_2 = 249,38 \quad n\hat{p}_3 = 137,81$$

et

$$q = \frac{(125 - 112,81)^2}{112,81} + \frac{(225 - 249,38)^2}{249,38} + \frac{(150 - 137,81)^2}{137,81} = 4,89$$

Le nombre de degrés de liberté est  $3 - 1 - 1 = 1$  (cf  $I = 3$ ,  $k = 1$ ).

Comme  $Z^2 \sim \mathcal{X}_1^2$  si  $Z \sim \mathcal{N}(0, 1)$  on a:

$$\begin{aligned}P[\mathcal{X}_1^2 \geq 4,89] &= 1 - P[(\mathcal{N}(0, 1))^2 \leq 4,89] \\ &= 1 - P[-\sqrt{4,89} \leq \mathcal{N}(0, 1) \leq \sqrt{4,89}] \\ &= 1 - P[-2,211 \leq \mathcal{N}(0, 1) \leq 2,211]\end{aligned}$$

$$\begin{aligned}P[\mathcal{N}(0, 1) \leq 2,211] &= 0,9826 \\ P[X_1^2 \geq 4,89] &= 2.(1 - 0,9826) \\ &= 2.0,0174 \\ &= 0,0348\end{aligned}$$

On rejette donc l'hypothèse de stabilité au niveau 5% ( mais on ne la rejetterait pas au niveau 1%!).

### Test d'une distribution donnée pour une variable

Les tests chi-carré d'ajustement sont utilisés pour tester si une variable aléatoire suit une distribution donnée (par exemple une loi de Poisson de paramètre  $\lambda$  ou une loi normale de moyenne  $\mu$  et de variance  $\sigma^2$ ).

**Exemple 10.1** On mesure le taux de cholestérol en mg/l, de 49 individus indiens vivant en zone rurale en Amérique centrale et ayant des revenus bas. Les mesures obtenues sont les nombres à 3 chiffres suivants:

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 204 | 108 | 140 | 152 | 158 | 129 | 175 | 146 | 157 | 174 | 192 | 194 | 144 | 152 |
| 7   | 1   | 3   | 4   | 5   | 2   | 6   | 4   | 5   | 6   | 7   | 7   | 3   | 4   |
| 135 | 223 | 145 | 231 | 115 | 131 | 129 | 142 | 114 | 173 | 226 | 155 | 166 | 220 |
| 3   | 7   | 4   | 7   | 1   | 2   | 2   | 3   | 1   | 6   | 7   | 4   | 5   | 7   |
| 180 | 172 | 143 | 148 | 171 | 143 | 124 | 158 | 144 | 108 | 189 | 136 | 136 | 197 |
| 6   | 6   | 3   | 4   | 6   | 3   | 2   | 5   | 3   | 1   | 7   | 3   | 3   | 7   |
| 131 | 95  | 139 | 181 | 165 | 142 | 162 |     |     |     |     |     |     |     |
| 2   | 1   | 3   | 6   | 5   | 3   | 5   |     |     |     |     |     |     |     |

Ces données peuvent-elles être considérées comme provenant d'une population normale? Les 49 observations sont réparties en 7 classes construites sur la base d'une division en 7 intervalles équiprobables sous une loi à priori  $\mathcal{N}(150, (30)^2)$ . On choisit des valeurs pour les paramètres (ici 150 pour la moyenne et  $30^2$  pour la variance) qui paraissent grosso modo compatibles avec les données. (La classe à laquelle appartient une mesure est le chiffre écrit en-dessous de cette mesure dans le tableau ci-dessus.)

Pour calculer ces 7 classes, on procède comme suit:

Rappelons que si  $\phi$  est la fonction de répartition d'une loi normale standard, on désigne par  $z_\beta$ , pour un  $\beta \in [0, 1]$ , le nombre réel tel que  $\phi(z_\beta) = 1 - \beta$ : donc si  $\beta < \beta'$ , on a

$$P[z_{\beta'} \leq \mathcal{N}(0, 1) \leq z_\beta] = \phi(z_\beta) - \phi(z_{\beta'}) = \beta' - \beta.$$

Rappelons et que,  $\phi(1,07) = 6/7 = 0,8577$ , donc  $z_{1/7} = 1,07$  et  $z_{6/7} = -1,07$ . On trouve de même dans les tables les valeurs de  $z_{2/7} = -z_{5/7}$  et



$z_{3/7} = -z_{4/7}$ . Maintenant

$$\begin{aligned}
P[\mathcal{N}(150; (30)^2) \leq 150 - 30 \times 1,07 = 117,9] &= P[\mathcal{N}(0,1) \leq -1,07] = 1/7, \\
P[\mathcal{N}(150; (30)^2) \geq 150 + 30 \times 1,07 = 182,1] &= P[\mathcal{N}(0,1) \geq 1,07] = 1/7, \\
P[150 + 30 \times z_{2/7} \leq \mathcal{N}(150; (30)^2) \leq 150 + 30 \times 1,07 = 182,1] \\
&= P[z_{2/7} \leq \mathcal{N}(0,1) \leq z_{1/7}] = 1/7,
\end{aligned}$$

etc... donc la division de la droite en 7 classes équiprobables pour  $\mathcal{N}(150, (30)^2)$  donne les intervalles  $] -\infty, 117,9[$   $[117,9, 132,9[$   $[132,9, 144,6[$   $[144,9, 155,4[$   $[155,4, 167,1[$   $[167,1, 182,1[$   $[182,1, \infty[$ .

La répartition des observations dans ces classe est la suivante:

| Classes             | $i$ | $n_i$ = effectifs<br>observé | $n\hat{p}_i$ =effectifs théoriques<br>estimés |
|---------------------|-----|------------------------------|---|
| $] -\infty, 117,9[$ | 1   | 5                            | 5,17  |
| $[117,9, 132,9[$    | 2   | 5                            | 5,64  |
| $[132,9, 144,6[$    | 3   | 11                           | 6,08  |
| $[144,9, 155,4[$    | 4   | 6                            | 6,64  |
| $[155,4, 167,1[$    | 5   | 6                            | 7,12  |
| $[167,1, 182,1[$    | 6   | 7                            | 7,98  |
| $[182,1, \infty[$   | 7   | 9                            | 10,38   |

On a calculé les “effectifs théoriques estimés”,  $n\hat{p}_i$ , en prenant pour estimation de  $\mu$  le nombre  $\hat{\mu} = \bar{x} = 157,02$ , et pour estimation de  $\sigma^2$  le nombre  $\hat{\sigma}^2 = S^2 = (31,4)^2$ . Alors,

$$\hat{p}_1 = P[\mathcal{N}(\hat{\mu}, \hat{\sigma}^2) \in \text{classe 1}] = P\left[-\infty < \mathcal{N}(0,1) < \frac{117,9 - 157,02}{31,4}\right] = 0,1055$$

$$\begin{aligned}
\hat{p}_2 &= P[\mathcal{N}(\hat{\mu}, \hat{\sigma}^2) \in \text{classe 2}] = P\left[\frac{117,9 - 157,02}{31,4} < \mathcal{N}(0,1) < \frac{132,9 - 157,02}{31,4}\right] \\
&= 0,1151
\end{aligned}$$

etc...

En toute rigueur, l'estimation de  $\hat{\mu}$  et  $\hat{\sigma}^2$  aurait dû se faire en utilisant le maximum de vraisemblance, donc en maximisant  $p_1(\mu, \sigma^2)^{n_1} p_2(\mu, \sigma^2)^{n_2} \dots p_7(\mu, \sigma^2)^{n_7}$ .

On ne le fera pas ici.

Nous obtenons comme valeur pour notre statistique de test:

$$q = \frac{(5 - 5, 17)^2}{5, 17} + \dots + \frac{(9 - 10, 38)^2}{10, 38} = 4, 60$$

Il y a  $7-1-2=4$  degrés. Puisque  $\chi_{4,0,05}^2 = 9, 488 > 4, 60$ , on ne rejette pas l'hypothèse. La conclusion est donc que ces données peuvent être considérées comme provenant d'une population normale  $\mathcal{N}(157, 02, \sigma^2 = (31, 4)^2)$ .

### 10.3 Les tests $\chi^2$ d'homogénéité

Supposons regarder  $J$  expériences  $E_1, \dots, E_j$ , multinomiales, chacune donnant lieu à  $I$  résultats possibles. On note  $p_{1j}, \dots, p_{ij}, \dots, p_{Ij}$  les probabilités associées à l'expérience  $j$ , donc  $p_{ij}$  est la probabilité d'obtenir le  $i^{eme}$  résultat lorsqu'on effectue la  $j^{eme}$  expérience. Pour l'ensemble des  $J$  expériences, on a un tableau de probabilités, tableau  $I \times J$  (à  $I$  lignes et  $J$  colonnes):

| $E_1$          | $E_j$          | $E_J$    |
|----------------|----------------|----------|
| $p_{11} \dots$ | $p_{1j} \dots$ | $p_{1J}$ |
| $\cdot$        | $\cdot$        | $\cdot$  |
| $\cdot$        | $\cdot$        | $\cdot$  |
| $\cdot$        | $\cdot$        | $\cdot$  |
| $p_{i1} \dots$ | $p_{ij} \dots$ | $p_{iJ}$ |
| $\cdot$        | $\cdot$        | $\cdot$  |
| $\cdot$        | $\cdot$        | $\cdot$  |
| $p_{I1} \dots$ | $p_{Ij} \dots$ | $p_{IJ}$ |

L'hypothèse d'**homogénéité** entre les  $J$  expériences consiste à dire que, pour chaque  $i$  entre 1 et  $I$ , la probabilité d'obtenir le  $i^{eme}$  résultat est la même pour les  $J$  expériences. On veut tester cette homogénéité. L'hypothèse  $H_0$  correspond au cas d'homogénéité donc

$$H_0 : \begin{aligned} p_{11} = p_{12} = \dots = p_{1j} = \dots = p_{1J} \\ p_{21} = p_{22} = \dots = p_{2j} = \dots = p_{2J} \quad \text{c'est à dire } p_{ij} = p'_{ij} \quad \forall i, j, j'. \\ p_{I1} = p_{I2} = \dots = p_{Ij} = \dots = p_{IJ} \end{aligned}$$

Ou encore : sous l'hypothèse  $H_0$ , toutes les colonnes du tableau sont identiques.

L'expérience  $E_j$  est répétée  $n_j$  fois (sous des conditions uniformes en veillant à l'indépendance), indépendamment des répétitions des autres expériences. Ces  $n_j$  répétitions fournissent les effectifs observés: on note  $n_{ij}$  le nombre de fois qu'on a obtenu le  $i^{eme}$  résultat au cours de ces  $n_j$  répétitions de l'expérience  $E_j$ . Ces effectifs observés sont présentés dans un tableau  $I \times J$  appelé **table de contingence**. On ajoute une colonne et une ligne avec la somme des éléments dans chaque ligne et chaque colonne; on définit donc

$$n_{i.} := n_{i1} + \dots + n_{iJ} \quad \forall 1 \leq i \leq I$$

Remarquons que  $n_{1j} + \dots + n_{ij} + \dots + n_{Ij} = n_j$  est le nombre de fois qu'on effectue la  $j^{eme}$  expérience

| $E_1$                               | $E_j$          | $E_J$    |                                    |
|-------------------------------------|----------------|----------|------------------------------------|
| $n_{11} \dots$                      | $n_{1j} \dots$ | $n_{1J}$ | $n_{1.} = n_{11} + \dots + n_{1J}$ |
| $\cdot$                             |                | $\cdot$  |                                    |
| $\cdot$                             |                | $\cdot$  |                                    |
| $n_{i1} \dots$                      | $n_{ij} \dots$ | $n_{iJ}$ | $n_{i.} = n_{i1} + \dots + n_{iJ}$ |
| $\cdot$                             |                | $\cdot$  |                                    |
| $\dots$                             |                | $\cdot$  |                                    |
| $\cdot$                             |                | $\cdot$  |                                    |
| $n_{I1} \dots$                      | $n_{Ij} \dots$ | $n_{IJ}$ | $n_{I.} = n_{I1} + \dots + n_{IJ}$ |
| $n_{1.} := n_{11} + \dots + n_{I1}$ | $n_j$          | $n_J$    | $n = n_1 + \dots + n_J$            |

Les  $n_j$  sont fixés par les conditions d'expérience; l'effectif observé  $n_{ij}$  est aléatoire; c'est la valeur observée d'une variable binomiale  $N_{ij} \sim Bin(n_j, p_{ij})$ . Les  $n_{i.}$  sont aléatoires et sont les valeurs des variables  $N_{i.} := N_{i1} + \dots + N_{iJ}$ . "L'effectif "théorique" est donné par  $E(N_{ij}) = n_j p_{ij}$ .

La procédure de test pour l'homogénéité est la suivante:

$$\begin{array}{ll} \text{Problème de test} & H_0 \quad p_{ij} = p_{ij'} \quad \forall i, j, j' \\ & H_1 \quad \exists i, j, j' \text{ tels que } p_{ij} \neq p_{ij'} \end{array}$$

La statistique de test est donnée par

$$Q = \sum_{I=1}^I \sum_{j=1}^J \frac{(N_{ij} - N_{i.} \frac{n_j}{n})^2}{N_{i.} \frac{n_j}{n}}.$$

Sous l'hypothèse  $H_0$ , la distribution de  $Q$  est approximativement une loi  $\chi^2$  à  $(I-1)(J-1)$  degrés de liberté (et l'approximation est valide sous les mêmes conditions que précédemment) :

$$Q \sim X_{(I-1)(J-1)}^2$$

La règle de comportement pour un test d'homogénéité au niveau  $\alpha$  est donc la suivante:

$$RH_0 \text{ si } q = \sum_{I=1}^I \sum_{j=1}^J \frac{(n_{ij} - n_{i.} \frac{n_{.j}}{n})^2}{n_{i.} \frac{n_{.j}}{n}} > \chi_{(I-1)(J-1); \alpha}^2$$

**Remarque 10.3** Sous l'hypothèse  $H_0$ , une estimation des  $p_{ij}$  est en effet donnée par  $\hat{p}_{11} = \hat{p}_{12} = \dots = \hat{p}_{1J} = \frac{n_{11}+n_{12}+\dots+n_{1J}}{n} = \frac{n_{1.}}{n}$  et de même  $\hat{p}_{ij} = \frac{n_{i.}}{n}$  et donc les effectifs théoriques correspondants au résultats  $i$  dans la  $j^{\text{ème}}$  expérience faite  $n_j$  fois sont:

$$n_j \hat{p}_{ij} = \frac{n_{i.} n_j}{n}.$$

**Exemple 10.2** Trois antidouleurs ont été administrés chacun à 75 patients souffrant de névralgie chronique. Les résultats obtenus sont les suivants:

|                      | antidouleur     |                 |                 |                |
|----------------------|-----------------|-----------------|-----------------|----------------|
| degré de soulagement | 1               | 2               | 3               |                |
| faible               | 15              | 19              | 13              | $n_{1.} = 47$  |
| moyen                | 42              | 32              | 40              | $n_{2.} = 114$ |
| complet              | 18              | 24              | 22              | $n_{3.} = 64$  |
|                      | $\overline{75}$ | $\overline{75}$ | $\overline{75}$ |                |

Ces données mettent-elles en évidence une action différenciée des trois antidouleurs? Le problème posé est celui de l'homogénéité des résultats obtenus pour les trois anti-douleurs. Les effectifs théoriques estimés  $\frac{n_{i.} n_j}{n}$  sont

$$\begin{array}{ccc} 15,67 & 15,67 & 15,67 \\ 38 & 38 & 38 \\ 21,33 & 21,33 & 21,33 \end{array}$$

La statistique de test vaut:

$$q = \frac{(15 - 15,67)^2}{15,67} + \frac{(42 - 38)^2}{38} + \frac{(18 - 21,33)^2}{21,33} + \frac{(19 - 15,67)^2}{15,67} + \dots = 3,54$$

Comme  $X_{4,10\%}^2 = 7,779 > 3,54$ , on ne rejette pas l'hypothèse que les trois antidouleurs sont équivalents.

## 10.4 Les tests $\chi^2$ d'indépendance

Soient  $X$  et  $Y$  deux variables aléatoires discrètes. On veut décrire une procédure de test pour voir si l'on peut considérer ces deux variables comme indépendantes.

Désignons par  $A_1, \dots, A_I$  les résultats possibles pour  $X$  et par  $B_1, \dots, B_J$  les résultats possibles pour  $Y$ . La loi bivariée de  $(X, Y)$  est caractérisée par les probabilités jointes

$$p_{ij} = P[X = A_i \text{ et } Y = B_j] \quad i = 1, \dots, I, j = 1, \dots, J$$

On range ces probabilités dans un tableau  $I \times J$

|          | $B_1 \dots$                            | $B_j \dots$                            | $B_J$         |   |
|----------|--|--|---------------|---|
| $A_1$    | $p_{11} \dots$                         | $p_{1j} \dots$                         | $p_{1J}$      | $p_{1\cdot} := p_{11} + \dots + p_{1J}$ |
| $\vdots$ | $\vdots$                               | $\vdots$                               | $\vdots$      |   |
| $A_i$    | $p_{i1} \dots$                         | $p_{ij} \dots$                         | $p_{iJ}$      | $p_{i\cdot} := p_{i1} + \dots + p_{iJ}$ |
| $\vdots$ | $\vdots$                               | $\vdots$                               | $\vdots$      |   |
| $A_I$    | $p_{I1} \dots$                         | $p_{IJ} \dots$                         | $p_{IJ}$      | $p_{I\cdot} := p_{I1} + \dots + p_{IJ}$ |
|          | $p_{\cdot 1} := p_{11} \dots + p_{I1}$ | $p_{\cdot j} := p_{1j} \dots + p_{Ij}$ | $p_{\cdot J}$ | 1                                       |

L'expérience aléatoire  $E$  à laquelle est associée une mesure de  $(X, Y)$  peut donc être vue comme une expérience relevant d'un schéma multinomial:

Résultats    Probabilités

|              |          |
|--------------|----------|
| $(A_1, B_1)$ | $p_{11}$ |
| $(A_1, B_2)$ | $p_{12}$ |
| $\vdots$     |          |
| $(A_I, B_J)$ | $p_{IJ}$ |

Il y a  $(I \times J)$  résultats possibles. Rappelons que l'indépendance de  $X$  et  $Y$  s'énonce :

$$p_{ij} = p_{i\cdot} p_{\cdot j} \quad \forall i, j.$$

En effet  $p_{i\cdot} = \sum_j p_{ij} = P[X = A_i]$  définit la fonction de probabilité marginale de  $X$  et  $p_{\cdot j} = \sum_i p_{ij} = P[Y = B_j]$  définit la fonction de probabilité marginale de  $Y$ .

L'expérience  $E$  est répétée  $n$  fois (sous des conditions uniformes et de façon à assurer l'indépendance entre les  $n$  répétitions). Ces  $n$  répétitions donnent lieu à des effectifs observés  $n_{ij}$  (qui sont les valeurs observées de variables aléatoires  $N_{ij} \sim \text{Bin}(n, p_{ij})$ ) présentés sous la forme d'une table de contingence  $I \times J$ .

|       | $B_1 \dots$                            | $B_j \dots$    | $B_J$         |   |
|-------|--|----------------|---------------|---|
| $A_1$ | $n_{11} \dots$                         | $n_{1j} \dots$ | $n_{1J}$      | $n_{1\cdot} := n_{11} + \dots + n_{1J}$ |
|       | $\vdots$                               | $\vdots$       | $\vdots$      |   |
| $A_i$ | $n_{i1} \dots$                         | $n_{ij} \dots$ | $n_{iJ}$      | $n_{i\cdot} := n_{i1} + \dots + n_{iJ}$ |
|       | $\vdots$                               | $\vdots$       | $\vdots$      |   |
| $A_I$ | $n_{I1} \dots$                         | $n_{Ij} \dots$ | $n_{IJ}$      | $n_{I\cdot}$                            |
|       | $n_{\cdot 1} := n_{11} \dots + n_{I1}$ | $n_{\cdot j}$  | $n_{\cdot J}$ |   |

**Remarque 10.4** Les effectifs théoriques sont donnés par  $E(N_{ij}) = np_{ij} = np_{i\cdot}p_{\cdot j}$  sous l'hypothèse que les variables sont indépendantes.

#### Test d'indépendance:

$$\begin{aligned} H_0 &: p_{ij} = p_{i\cdot} p_{\cdot j} \quad \forall i, j \\ H_1 &: \exists i, j \text{ tels que } p_{ij} \neq p_{i\cdot} p_{\cdot j}. \end{aligned}$$

#### Statistique de test :

$$Q = \sum_{I=1}^I \sum_{j=1}^J \frac{(N_{ij} - \frac{N_{i\cdot} N_{\cdot j}}{n})^2}{\frac{N_{i\cdot} N_{\cdot j}}{n}}.$$

où  $N_{i\cdot} = \sum_j N_{ij}$  et  $N_{\cdot j} = \sum_i N_{ij}$ .

La loi de  $Q$  sous l'hypothèse  $H_0$  est approximativement une loi chi-carré

$$Q \approx \chi_{(I-1)(J-1)}^2$$

et l'approximation est valable sous les conditions énoncées précédemment.

### Règle de comportement au niveau $\alpha$ :

$$RH_0 \text{ si } q = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - \frac{n_{i.} n_{.j}}{n})^2}{\frac{n_{i.} n_{.j}}{n}} > \chi_{(I-1)(J-1); \alpha}$$

**Remarque 10.5** Les quantités  $\frac{n_{i.} n_{.j}}{n}$  sont les effectifs théoriques estimés sous la condition  $H_0$ , soit  $n\hat{p}_{ij} = n\hat{p}_{i.}\hat{p}_{.j}$  avec les estimations  $\hat{p}_{i.} = \frac{n_{i.}}{n}$  et  $\hat{p}_{.j} = \frac{n_{.j}}{n}$ .

**Exemple 10.3** Des certificats de décès, au nombre de 2.100, ont été prélevés au hasard dans l'ensemble des certificats de décès des hopitaux publics. On les a classé selon deux critères :

- la cause de décès d'une part (qu'on notera  $(X)$ ) et
- les habitudes tabagiques du défunt  $(Y)$

Les résultats sont les suivants:

| <u>Cause du décès</u> | <u>Non fumeur</u> | <u>Fumeur modéré</u> | <u>Grand fumeur</u> | <u>Total</u> |
|-----------------------|-------------------|----------------------|---------------------|--------------|
| Maladie respiratoire  | 55                | 120                  | 162                 | 337          |
| Maladie cardiaque     | 49                | 388                  | 315                 | 752          |
| Autre                 | 61                | 300                  | 650                 | 1011         |
| Total                 | 165               | 808                  | 1127                | 2100         |

Le tableau des effectifs théoriques estimés  $n\hat{p}_{i.}\hat{p}_{.j} = \frac{n_{i.} n_{.j}}{n}$  sous la condition d'indépendance  $H_0$ , est

| <u>Cause du décès</u> | <u>Non fumeur</u>                     | <u>Fumeur modéré</u> | <u>Grand fumeur</u> | <u>Total</u> |
|-----------------------|---------------------------------------|----------------------|---------------------|--------------|
| Maladie respiratoire  | $26,6 = \frac{337 \times 165}{2100}$  | 129,7                | 180,8               | 337          |
| Maladie cardiaque     | $59,1 = \frac{752 \times 165}{2100}$  | 289,3                | 403,6               | 752          |
| Autre                 | $79,4 = \frac{1011 \times 165}{2100}$ | 389                  | 542,6               | 1011         |
| Total                 | 165                                   | 808                  | 1127                |              |

La valeur  $q$  de la statistique de test est:

$$q = \frac{(55 - 26,5)^2}{26,5} + \frac{(120 - 129,7)^2}{129,7} + \frac{(162 - 180,8)^2}{180,8} + \frac{(49 - 59,1)^2}{59,1} + \frac{(388 - 289,3)^2}{289,3} \\ + \frac{(315 - 403,6)^2}{403,6} + \frac{(61 - 79,4)^2}{79,4} = 134,1.$$

Le nombre de degrés de liberté est  $(3 - 1)(3 - 1) = 4$ . Donc on rejette  $H_0$ ; les deux variables ne sont pas indépendantes.



# Chapitre 11

## Le modèle de régression linéaire simple

L'objectif de l'analyse de régression est d'exploiter le lien entre deux (ou plusieurs) variables pour obtenir de l'information sur l'une d'entre elles à travers la connaissance des valeurs des autres. Deux variables  $x$  et  $y$  sont liées de manière déterministe si, quand on connaît la valeur de  $x$ , celle de  $y$  est complètement déterminée. Par exemple, si une particule se déplace sur une droite avec une vitesse initiale  $v_0$  et une accélération constante égale à  $a_0$ , la distance parcourue ( $y$ ) est connue si on connaît le temps ( $x$ ) depuis le moment initial:

$$y = v_0 x + \frac{1}{2} a_0 x^2.$$

Il arrive souvent que deux variables soient liées, mais pas de manière déterministe. Par exemple si  $x$  est l'âge d'un enfant et  $y$  la taille de cet enfant ou encore si  $x$  est la moyenne obtenue en Bac1 et  $y$  la moyenne obtenue en Bac3 ...

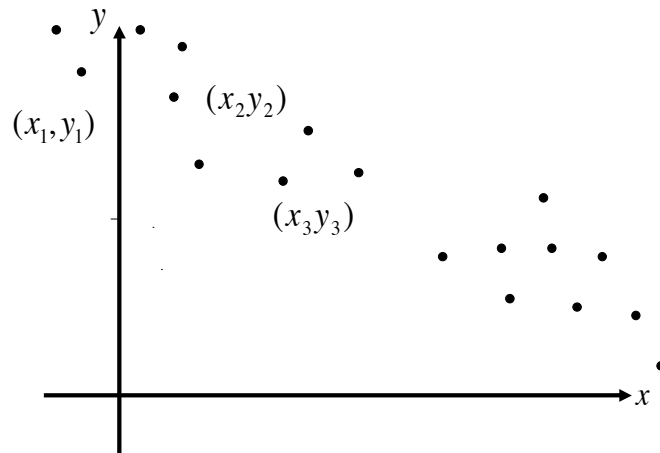
L'analyse de la régression est le domaine des statistiques qui étudie les relations entre deux (ou plusieurs) variables liées de manière non déterministe. La plus simple relation déterministe entre deux variables  $x$  et  $y$  est une relation linéaire  $y = \beta_0 + \beta_1 x$ . Si les deux variables ne sont pas liées de manière déterministe, alors pour une valeur fixée de la variable  $x$ , la valeur de la seconde variable  $y$  est aléatoire.

Généralement la variable dont la valeur est fixée par l'expérimentateur est notée  $x$  et est appelée variable indépendante. Pour une valeur fixée  $x$ , la

seconde variable est aléatoire; on la note  $Y$  et on note  $y$  la valeur observée. On appelle  $Y$  la variable dépendante.

En général, des observations sont faites pour un certain nombre de valeurs de la variable indépendante  $x$ , soit les valeurs  $x_1, \dots, x_N$ . notons  $Y_i$  la variable aléatoire - et  $y_i$  sa valeur observée - associée à  $x_i$ . Les données consistent en  $n$  paires  $(x_1, y_1) \dots (x_N, y_N)$ .

Une première étape dans l'analyse de régression de deux variables est de dessiner le "scatter plot" ( graphique de dispersion) des données.



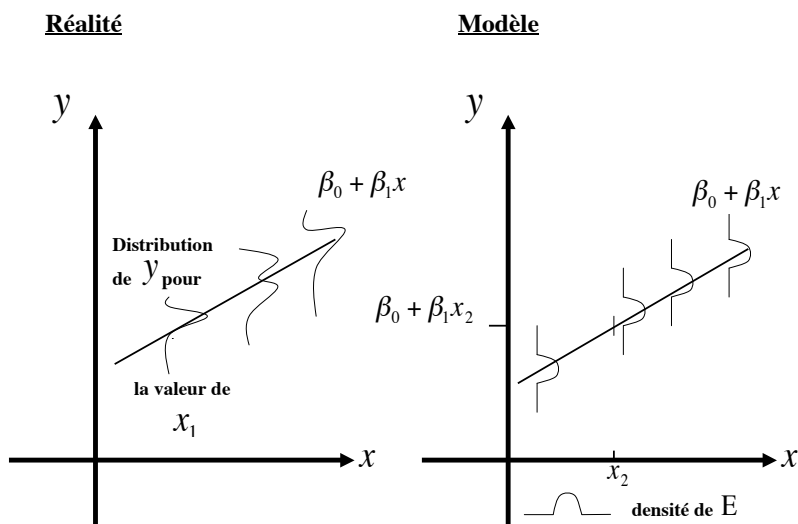
Il n'y a aucune courbe simple passant par ces points; cependant, il y a une nette tendance observée: la valeur de  $y$  diminue quand celle de  $x$  augmente.

La généralisation du modèle déterministe linéaire  $y = \beta_0 + \beta_1 x$  est de supposer que l'espérance de  $Y$  pour une valeur fixée de  $x$  est une fonction linéaire de  $x$  et, pour une valeur fixée de  $x$ , la variable  $Y$  diffère de son espérance par une variable aléatoire  $\varepsilon$ .

Pour un  $x$  fixé, si  $\mu_{Y|x} := E[Y|x]$  on a  $\mu_{Y|x} = \beta_0 + \beta_1 x$  et si  $\sigma_{Y|x}^2 := Var[Y|x]$  on a  $\sigma_{Y|x} = \sigma^2 = Var[\varepsilon]$ .

En réalité, la distribution de  $Y|x$  autour de  $\mu_{y|x}$ , donc la distribution de  $\varepsilon$  est différente suivant la valeur de  $x$  mais on fait dans ce modèle une hypothèse simplificatrice.

(Par exemple si  $x$  désigne la taille d'un homme et  $y$  son poids, les variations autour d'un poids moyen pour une taille donnée augmentent avec  $x$ , donc on devrait avoir  $Var(Y|x) = \sigma_x^2$  avec  $\sigma_x^2$  croissant avec  $x$  pour être plus réaliste mais on suppose ici  $Var(Y|x) = \sigma^2$  est constante).



La droite d'équation  $y = \beta_0 + \beta_1 x$  est appelée droite de régression;  $\beta_0$  et  $\beta_1$  sont appelés paramètres de la droite de régression.

## 11.1 Estimation des paramètres

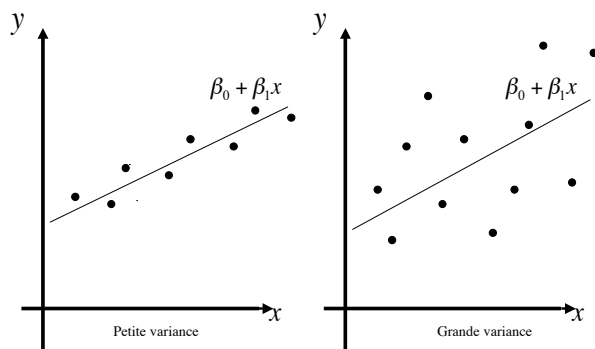
Avant de faire des inférences sur les paramètres  $\beta_0$  et  $\beta_1$ , on doit préciser comment les paires  $(x_i, y_i)$  observées sont obtenues à partir du modèle linéaire probabiliste exposé plus haut.

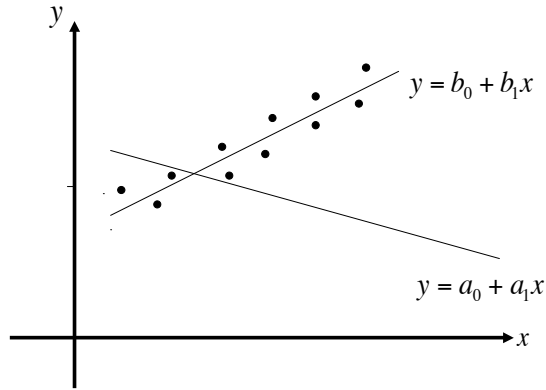
**Hypothèse :** Pour les valeurs fixées  $x_1, x_2, \dots, x_n$  de la variable  $x$ , les observations  $y_1, \dots, y_n$  sont les valeurs observées de variables aléatoires  $Y_1, \dots, Y_n$  telles que

$$\begin{cases} Y_1 = \beta_0 + \beta_1 x_1 + \varepsilon_1 \\ \vdots \\ Y_n = \beta_0 + \beta_1 x_n + \varepsilon_n \end{cases}$$

où les  $\varepsilon_i$  sont des variables aléatoires indépendantes identiquement distribuées, telles que  $E(\varepsilon_i) = 0$   $Var(\varepsilon_i) = \sigma^2$ .

En pratique, les valeurs des paramètres  $\beta_0$  et  $\beta_1$  ne sont pas connues; les paires  $(x_1, y_1)$  sont distribuées autour de la droite de régression; si la variance  $\sigma^2$  est grande, de nombreuses observations seront loin de la droite mais si  $\sigma^2$  est petite, les observations s'agglomèrent autour de la droite de régression.





Notre estimation de  $y = \beta_0 + \beta_1 x$  doit être une droite qui s'adapte le mieux possible aux données observées.

### 11.1.1 Méthode des moindres carrés (Gauss 1777–1855)

**Principe** : Parmi toutes les droites  $y = b_0 + b_1 x$ , la droite de régression estimée  $y = \hat{\beta}_0 + \hat{\beta}_1 x$  est celle qui minimise la somme des carrés des résidus

$$SC(b_0, b_1) = \sum_{i=1}^n (y_i - (b_0 + b_1 x_i))^2$$

donc

$$SC(\hat{\beta}_0, \hat{\beta}_1) \leq SC(b_0, b_1) \quad \forall \beta_0, \beta_1$$

Pour chercher le minimum on doit avoir  $\frac{\partial SC}{\partial b_0}(\hat{\beta}_0, \hat{\beta}_1) = 0$  et  $\frac{\partial SC}{\partial b_1} = 0$  donc

$$\sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)) = 0$$

$$\sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)) x_i = 0.$$

Ceci donne:

$$\begin{aligned} n\hat{\beta}_0 + \left(\sum_{i=1}^n x_i\right)\hat{\beta}_1 &= \sum_{i=1}^n y_i \\ \left(\sum_{i=1}^n x_i\right)\hat{\beta}_0 + \left(\sum_{i=1}^n (x_i)^2\right)\hat{\beta}_1 &= \sum_{i=1}^n x_i y_i \end{aligned}$$

et donc

$$\begin{aligned} \hat{\beta}_1 &= \frac{n \sum_i x_i y_i - (\sum_i x_i)(\sum_i y_i)}{n \sum_i x_i^2 - (\sum_i x_i)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \hat{\beta}_0 &= \frac{\sum_{i=1}^n y_i - \hat{\beta}_1 \sum_{i=1}^n x_i}{n} = \bar{y} - \hat{\beta}_1 \bar{x}. \end{aligned}$$

La valeur du minimum de la somme des carrés des résidus

$$SC_{\text{rés}} = SC(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^n (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$$

est appelée la somme des carrés résiduelle.

### 11.1.2 Estimateurs pour $\beta_0$ , $\beta_1$ et $\sigma^2$

On suppose que les valeurs des  $x_i$  sont choisies avant de faire l'expérience, donc seules les  $y_i$  sont les valeurs de variables aléatoires  $Y_i$ . Les estimateurs de  $\beta_0$  et  $\beta_1$  sont notés  $\hat{\beta}_0$  et  $\hat{\beta}_1$  sont définis par:

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{i=1}^n (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} (Y_i - \frac{1}{n} \sum_{j=1}^n Y_j) \\ \hat{\beta}_0 &= \frac{\sum Y_i - \hat{\beta}_1 \sum x_i}{n}. \end{aligned}$$

On a  $E[\hat{\beta}_0] = \beta_0$  et  $E[\hat{\beta}_1] = \beta_1$ ; ce sont donc des estimateurs non biaisés. D'autre part

$$\begin{aligned} Var[\hat{\beta}_0] &= \frac{\sigma^2}{n} \left(1 + \frac{\bar{x}^2}{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}\right), \\ Var[\hat{\beta}_1] &= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}. \end{aligned}$$

Un estimateur pour  $\sigma^2$  est donné par

$$\hat{S}^2 = \frac{1}{n-2} SC_{res} = \frac{1}{n-2} \sum_{i=1}^n (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2.$$

On divise par  $n-2$  car les 2 paramètres  $\beta_0$  et  $\beta_1$  ont été estimés; on peut montrer que  $E(\hat{S}^2) = \sigma^2$ .

### 11.1.3 Distribution des estimateurs dans le cas gaussien

Pour tester des hypothèses et construire des intervalles de confiance, on doit faire des hypothèses supplémentaires sur la distribution des  $Y_i$ .

**Hypothèse:** On suppose que les  $Y_i$  sont non seulement indépendants mais sont normalement distribués (i.e. on suppose les  $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ ).

Alors

$$\begin{aligned}\hat{\beta}_0 &\sim \mathcal{N}\left(\beta_0, \frac{\sigma^2}{n} \left(1 + \frac{\bar{x}^2}{\frac{1}{n} \sum_i (x_i - \bar{x})^2}\right)\right) \\ \hat{\beta}_1 &\sim \mathcal{N}\left(\beta_1, \sigma^2 / \sum_i (x_i - \bar{x})^2\right) \\ \frac{(n-2)S^2}{\sigma^2} &= \frac{1}{\sigma^2} SC_{res} \sim \chi_{n-2}^2.\end{aligned}$$

De plus, les deux variables  $\hat{\beta}_1$  et  $\frac{(n-2)S^2}{\sigma^2}$  sont indépendantes donc

$$\frac{\hat{\beta}_1 - \beta_1}{S / \sqrt{\sum_i (x_i - \bar{x})^2}} \sim t_{n-2}.$$

## 11.2 Intervalle de confiance et Tests d'hypothèse pour $\beta_1$ dans le cas gaussien

### 11.2.1 Intervalle de confiance pour $\beta_1$

Un intervalle de confiance au niveau  $(1 - \alpha)$  pour  $\beta_1$  est donné par

$$\left[ \hat{\beta}_1 - \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}} t_{n-2; \alpha/2}; \hat{\beta}_1 + \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}} t_{n-2; \alpha/2} \right]$$

où  $\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$  et où  $s = \sqrt{s^2}$  avec  
 $s^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$  pour  $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ .

### 11.2.2 Test d'hypothèse pour $\beta_1$

Pour faire des tests sur la pente  $\beta_1$  de la droite de régression on utilise la statistique de test

$$T = \frac{\hat{\beta}_1 - \beta_0}{S / \sqrt{\sum (x_i - \bar{x})^2}}.$$

Les tests au niveau  $\alpha$  sont les suivants:

| Hypothèse à tester              | Règle de décision  |
|---------------------------------|--|
| $H_0 ; \beta_1 \leq \beta_{10}$ | $RH_0$ si $T \geq t_{n-2;\alpha}$                            |
| $H_0 ; \beta_1 \geq \beta_{10}$ | $RH_0$ si $T \leq -t_{n-2;\alpha}$                           |
| $H_0 ; \beta_1 = \beta_{10}$    | $RH_0$ si $T \notin [-t_{n-2;\alpha/2}, t_{n-2;\alpha/2}]$ . |

## 11.3 Corrélation

On a défini la méthode des moindres carrés pour trouver les coefficients d'une droite qui passe à travers le nuage de points dans le plan défini par des mesures  $(x_i, y_i)$  de deux variables  $X$  et  $Y$ . Une question importante se pose cependant: comment mesurer la force d'une relation linéaire entre les deux variables  $X$  et  $Y$ ?

**Définition 11.1** Le coefficient de corrélation de l'échantillon  $(x_1, y_1) \dots (x_n, y_n)$  est défini par

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}.$$

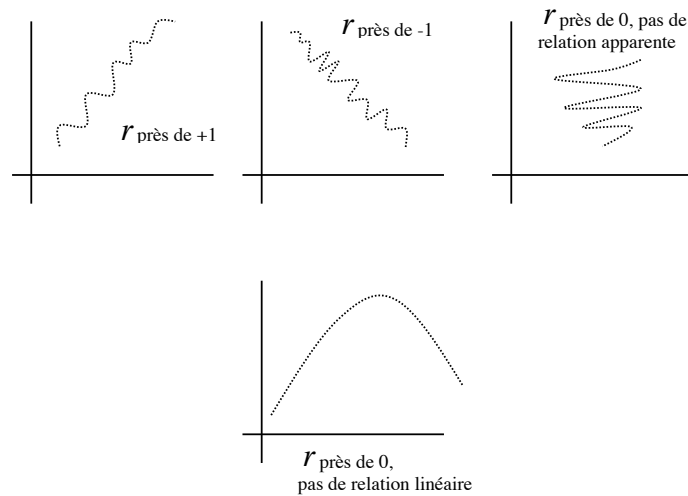
Remarquons que:

- $r$  est indépendant des unités dans lesquelles  $x$  et  $y$  sont mesurés.
- $-1 \leq r \leq 1$ .
- $r = 1$  ssi toutes les paires  $(x_i, y_i)$  sont sur une droite de pente positive et  $r = -1$  ssi toutes les paires  $(x_i, y_i)$  sont sur une droite de pente négative.

En règle usuelle, on dit que la corrélation est faible si  $0 \leq |r| \leq 1/2$  et que la corrélation est forte si  $|r| \geq 0,8$ .



### Exemples:



On peut penser aux paires  $(x_i, y_i)$  comme étant les valeurs prises par des paires de variables aléatoires  $(X_i, Y_i)$  ayant toutes la même distribution de probabilité jointe, définie par une fonction de probabilité jointe  $p^{(X,Y)}(x, y) = P["X = x \text{ et } Y = y"]$  dans le cas discret ou par une densité de probabilité jointe  $f^{(X,Y)}(x, y)$  dans le cas continu.

On a défini le coefficient de corrélation  $\rho(X, Y)$  de deux variables aléatoires  $X$  et  $Y$  par

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

où

$$\begin{aligned} \text{cov}(X, Y) &= \sum_x \sum_y (x - \mu_x)(y - \mu_y) p^{(X,Y)}(x, y) \text{ pour } (X, Y) \text{ discrètes} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f(x, y) dx dy \\ &\text{pour } (X, Y) \text{ conjointement continues.} \end{aligned} \quad (11.1)$$

Le coefficient de corrélation  $\rho := \rho(X, Y)$  est un paramètre de la population (comme  $\mu_X, \mu_Y, \sigma_X, \sigma_Y \dots$ ). Une estimation de ce paramètre  $\rho$  dans un échantillon  $\{(x_i, y_i) \mid 1 \leq i \leq n\}$  est donnée par  $r$ . L'estimateur correspondant est

$$R = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i (X_i - \bar{X})^2} \sqrt{\sum_i (Y_i - \bar{Y})^2}}.$$

Si la densité de probabilité jointe de  $(X, Y)$  est donnée par

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{\left(-\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]/2(1-\rho^2)\right)}$$

alors

$$T = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}} \sim t_{n-2} \text{ si } \rho = 0$$

Ceci permet de faire les tests d'hypothèse  $H_0 : \rho \geq 0$ , ou  $H_0 : \rho = 0$ , ou  $H_0 : \rho \leq 0$  sur le coefficient de corrélation.

Pour tester l'hypothèse  $H_0 : \rho = \rho_0$ , ou  $H_0 : \rho \geq \rho_0$ , ou  $H_0 : \rho \leq \rho_0$ , avec  $\rho_0 \neq 0$ , on utilise, pour une distribution jointe des variables  $X$  et  $Y$  comme ci-dessus, la statistique

$$V := \frac{1}{2} \ln \left( \frac{1+R}{1-R} \right)$$

qui a approximativement une distribution normale de moyenne  $\mu_V = \frac{1}{2} \ln \left( \frac{1+\rho}{1-\rho} \right)$  et de variance  $\frac{1}{n-3}$ .

# Chapitre 12

## ANOVA à un facteur

En étudiant les méthodes d'analyse de données, nous nous sommes d'abord intéressés à des problèmes concernant un seul échantillon, puis nous avons abordé l'analyse comparative de deux échantillons. Pour un seul échantillon, les données étaient des observations obtenues en choisissant au hasard des “individus” dans une population. Quand on comparait deux échantillons, ils étaient soit pris dans deux populations distinctes, soit ils concernaient deux “traitements” distincts appliqués à des individus issus d'une même population.

L'analyse de la variance (plus brièvement ANOVA) réfère à de nombreuses situations expérimentales, définissant des procédures statistiques pour analyser des données provenant d'expériences. Le problème le plus simple, appelé ANOVA à un facteur, concerne l'analyse de données

- soit issues de plus de deux populations,
- soit venant d'expériences sur une population où plus de deux traitements ont été utilisés.

La caractéristique qui différencie les populations ou les traitements est appelée le **facteur** et les différents traitements (ou populations) sont appelés les **niveaux** du facteur.

Par exemple, on peut faire une expérience étudiant l'effet du nombre de plantes par pot sur le poids moyen d'une plante. Le facteur est ici le nombre de plantes par pot. Il y a autant de niveaux de ce facteur que de nombres différents de plantes par pot dans les expériences faites (par exemple 3 niveaux: 4 plantes par pot, 12 plantes par pot et 28 plantes par pot). La grandeur étudiée est le poids d'une plante.

L'ANOVA à un facteur va étudier la comparaison des moyennes de la grandeur

étudiée dans ces différentes populations (ou lors de ces différents traitements).

Notons  $I$  pour le nombre de niveaux du facteur (donc le nombre de populations ou de traitements qu'on veut comparer) et notons  
 $\mu_1$  pour la moyenne de la grandeur étudiée dans le premier niveau (donc dans la première population ou lorsque le premier traitement est appliqué),  
 $\mu_2$  pour la moyenne de la grandeur étudiée dans le deuxième niveau,  
 $\vdots$   
 $\mu_I$  pour la moyenne de la grandeur étudiée dans le dernier ( $I^{eme}$ ) niveau.

Le test d'hypothèse qui nous intéresse ici est:

$H_0$  :  $\mu_1 = \mu_2 = \dots \mu_I$   
contre  
 $H_1$  : au moins deux de ces moyennes sont différentes.

Nous prenons un échantillon pour chacun des niveaux du facteur et notons :  
-  $X_{ij}$  la variable aléatoire qui dénote la  $j^{eme}$  mesure de la grandeur observée dans le  $i^{eme}$  niveau;  
-  $x_{ij}$  la valeur observée de  $X_{ij}$  quand on fait l'expérience.

**Hypothèse de travail:** on suppose que chacun des  $X_{ij}$  (qui sont bien sûr tous supposés indépendants et identiquement distribués pour chaque valeur de  $i$  fixée) a une distribution normale de moyenne  $\mu_i$  et de variance  $\sigma^2$ . On suppose donc que la variance est la même dans tous les niveaux!

Supposons avoir un échantillon de taille  $J_k$  pour le  $k^{eme}$  niveau; on a donc des observations

$$x_{k1}, x_{k2}, \dots x_{kJ_k}$$

qui sont les valeurs de variables aléatoires indépendantes

$$X_{k1}, X_{k2}, \dots X_{kJ_k} \sim \mathcal{N}(\mu_k, \sigma^2).$$

On désigne par  $n$  le nombre total d'observations, donc

$$n = J_1 + J_2 + \dots + J_I.$$

**Théorème 12.1** *Définissons:*

- La somme totale des carrés

$$SST := \sum_{i=1}^I \sum_{j=1}^{J_i} (X_{ij} - \bar{X}_{..})^2 = \sum_{i=1}^I \sum_{j=1}^{J_i} X_{ij}^2 - \frac{1}{n} X_{..}^2 ;$$

- La somme des carrés des traitements

$$SST_r := \sum_{i=1}^I \sum_{j=1}^{J_i} (\bar{X}_{i.} - \bar{X}_{..})^2 = \sum_{i=1}^I \frac{1}{J_i} X_{i.}^2 - \frac{1}{n} X_{..}^2 ;$$

- La somme des carrés des erreurs

$$SSE := \sum_{i=1}^I \sum_{j=1}^{J_i} (X_{ij} - \bar{X}_{i.})^2 = SST - SST_r ;$$

où on a défini

$$\begin{aligned} X_{..} &:= \sum_{i=1}^I \sum_{j=1}^{J_i} X_{ij} & \bar{X}_{..} &:= \frac{1}{n} X_{..} \\ X_{i.} &:= \sum_{j=1}^{J_i} X_{ij} & \bar{X}_{i.} &:= \frac{1}{J_i} X_{i.} \end{aligned}$$

donc  $\bar{X}_{i.}$  est la moyenne-échantillon de l'échantillon dans le niveau  $i$ ;

- $MST_r := \frac{1}{I-1} SST_r$      $MSE := \frac{1}{n-I} SSE$ .

Alors, sous l'hypothèse  $H_0$  que les moyennes sont égales,  $F := \frac{MST_r}{MSE}$  a la distribution d'une variable aléatoire  $\mathcal{F}$  de paramètres  $I-1, n-I$ :

$$F = \frac{MST_r}{MSE} \sim \mathcal{F}_{I-1, n-I}.$$

**Remarque 12.1** Les variables aléatoires  $\mathcal{F}_{n_1, n_2}$  sont définies pour des entiers positifs  $n_1$  et  $n_2$ ; leurs valeurs sont données dans des tables, en particulier les nombres réels que nous noterons  $\mathcal{F}_{n_1, n_2; \alpha}$  tels que

$$P[ \mathcal{F}_{n_1, n_2} > \mathcal{F}_{n_1, n_2; \alpha} ] = \alpha.$$

Si on a deux variables chi-carré, leur quotient définit une variable  $\mathcal{F}$ ; de manière plus précise, si  $Y_1 \sim \mathcal{X}_{n_1}$  et  $Y_2 \sim \mathcal{X}_{n_2}$  alors  $\frac{Y_1/n_1}{Y_2/n_2} \sim \mathcal{F}_{n_1, n_2}$ .

**La règle de décision** pour notre test au niveau  $\alpha$  est la suivante: on rejette l'homogénéité (donc on rejette l'hypothèse  $H_0$  selon laquelle les moyennes dans les différents niveaux sont égales :  $\mu_1 = \mu_2 = \dots \mu_I$ ) si la valeur  $f$  de la statistique  $F$  est supérieure à  $\mathcal{F}_{I-1, n-I; \alpha}$ .

Si on rejette l'homogénéité, il existe de nombreuses méthodes statistiques permettant des comparaisons multiples entre les différents niveaux. Ceci sort du cadre de ce cours d'introduction aux méthodes statistiques.

De même, les ANOVA à deux facteurs ou plus étudient des situations où plusieurs facteurs sont étudiés, chacun avec un certain nombre de niveaux.

Remarquons encore que notre étude de l'ANOVA à un facteur peut se décrire sous la forme

$$X_{ij} = \mu_i \varepsilon_{ij}$$

où les  $\varepsilon_{ij}$  sont des variables indépendantes et identiquement distribuées suivant une normale  $\mathcal{N}(0, \sigma^2)$ .