

# Chapitre VI : Probabilités et combinatoire

## 1 Rappels

### 1.1 Probabilités d'événements

- En général, on note  $e_1, e_2, \dots, e_n$ , les résultats possibles d'une expérience aléatoire. L'ensemble de ces éventualités est appelé l'**univers**, on note  $\Omega = \{e_1, e_2, \dots, e_n\}$ . Un événement  $A$  est un sous-ensemble ou partie de  $\Omega$ .
- Un **événement élémentaire** est un événement qui ne contient qu'une seule issue, par exemple  $\{e_i\}$ . A chaque événement élémentaire  $\{e_i\}$ , on associe un nombre  $p_i$ , appelé **probabilité de  $\{e_i\}$**  tel que  $0 \leq p_i \leq 1$ . On note  $P(e_i) = p_i$ . En outre  $p_1 + p_2 + \dots + p_n = 1$ . On dit que l'on a défini une probabilité sur  $\Omega$ .
- La probabilité d'un événement  $A$ , notée  $P(A)$ , est la somme des probabilités des événements élémentaires inclus dans  $A$ . On pose  $P(\Omega) = 1$  et  $P(\emptyset) = 0$ .
- Dire que les événements élémentaires sont **équiprobables** signifie que  $P(e_i) = P(e_j)$ , pour tout  $i$  et  $j$ . Si leur nombre est  $n$ , alors  $P(e_i) = \frac{1}{n}$ , et la probabilité d'un événement  $A$  est donnée par :
 
$$P(A) = \frac{\text{nombre d'éléments de } A}{\text{nombre d'éléments de } \Omega}.$$
- L'événement  $\bar{A}$  est l'ensemble de toutes les issues qui ne sont pas dans  $A$ ,  $P(A) + P(\bar{A}) = 1$ .  $\bar{A}$  est l'événement contraire de  $A$ . Les probabilités de  $A, B, A \cap B$  et  $A \cup B$  sont liées par l'égalité :  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .

### 1.2 variable aléatoire

$\Omega = \{e_1, e_2, \dots, e_n\}$  est l'univers d'une expérience aléatoire sur lequel est définie une probabilité. Une **variable aléatoire**  $X$  est une fonction de  $\Omega$  dans  $\mathbb{R}$ . Notons  $x_1, x_2, \dots, x_q$  les valeurs de  $X$  ( $q \leq n$ ). On note  $p_i$  la probabilité que  $X$  soit égal à  $x_i$ . L'ensemble des couples  $(x_i, p_i)$  est la **loi de probabilité de  $X$** . On la présente en général sous forme d'un tableau.

#### Paramètres d'une variable aléatoire

**Espérance :**  $E(X) = x_1p_1 + x_2p_2 + \dots + x_qp_q = \sum_{k=1}^q x_kp_k$

**Variance :**  $V(X) = [x_1 - E(X)]^2p_1 + \dots + [x_q - E(X)]^2p_q = E(X^2) - [E(X)]^2$

**Ecart-type :**  $\sigma(x) = \sqrt{V(X)}$

## 2 Conditionnement par un événement de probabilité non nulle

### 2.1 Définition

Exemple :

Dans un lycée, les 250 élèves qui étudient une seconde langue se répartissent, suivant le choix de cette langue, selon le tableau ci-contre.

fréquences	allemand	italien	espagnol	total
garçons	0,20	0,044	0,156	0,40
filles	0,12	0,156	0,324	0,60
total	0,32	0,20	0,48	1

Une expérience aléatoire consiste à choisir un élève au hasard. On modélise cette expérience par la loi

équiprobable sur l'ensemble  $\Omega$  des 250 élèves. On considère les événements  $A$  : « l'élève étudie l'allemand » et  $F$  : « l'élève est une fille », on s'intéresse à la probabilité notée  $P_F(A)$ , que l'élève étudie l'allemand sachant que c'est une fille.

Le lycée compte  $0,6 \times 250 = 150$  filles dont  $0,12 \times 250 = 30$  étudient l'allemand, donc  $P_F(A) = \frac{30}{150} = \frac{1}{5}$ .

Cette probabilité est ainsi égale à  $\frac{0,12}{0,6}$ , c'est à dire  $\frac{P(A \cap F)}{P(F)}$ .

Définition : Une loi de probabilité est définie sur un ensemble  $\Omega$ .

$A$  et  $B$  sont deux événements avec  $P(A) \neq 0$ .

La probabilité de l'événement  $B$  sachant  $A$ , notée  $P_A(B)$ , est définie par :  $P_A(B) = \frac{P(A \cap B)}{P(A)}$ .

Remarque : La probabilité conditionnelle  $P_A(B)$  se note parfois  $P(B \setminus A)$ .

Propriété :  $A$  et  $B$  sont deux événements de  $\Omega$  de probabilité non nulle.

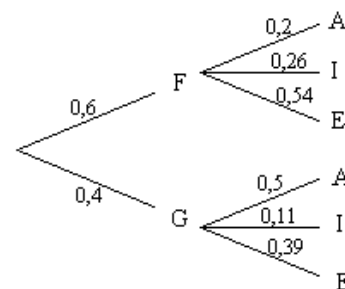
- $P(A \cap B) = P_B(A) \times P(B) = P_A(B) \times P(A)$
- $P_A(B) + P_A(\bar{B}) = 1$

## 2.2 Représentation à l'aide d'un arbre de probabilités

On peut représenter l'expérience aléatoire, décrite dans l'exemple précédent, par un arbre de probabilités. Considérons les événements :  $G$  : « l'élève est un garçon »,  $I$  : « l'élève étudie l'italien » et  $E$  : « l'élève étudie l'espagnol ».

Sur les deux premières branches, figurent les probabilités des événements  $F$  et  $G$  :  $P(F) = 0,6$  et  $P(G) = 0,4$ .

Puis sur les branches suivantes, on note les probabilités de chacun des événements  $A$ ,  $I$  ou  $E$  sachant  $F$  ou  $G$ .



## 2.3 Formule des probabilités totales

Définition : Dire que les événements  $B_1, B_2, \dots, B_n$  forment une partition de  $\Omega$  signifie que les ensembles  $B_i$  sont deux à deux disjoints et que leur réunion est  $\Omega$ .

Théorème : Les événements  $B_1, B_2, \dots, B_n$  forment une partition de  $\Omega$ . Alors, pour tout événement  $A$  :  $P(A) = P(A \cap B_1) + P(A \cap B_2) + \dots + P(A \cap B_n)$ , c'est à dire :  $P(A) = P_{B_1}(A) \times P(B_1) + P_{B_2}(A) \times P(B_2) + \dots + P_{B_n}(A) \times P(B_n)$ .

Démonstration : Les événements  $A \cap B_1, A \cap B_2, \dots, A \cap B_n$  sont deux à deux incompatibles et leur réunion est  $A$ , la formule en découle.

Remarque : On obtient en particulier, pour tous événements  $A$  et  $B$  de  $\Omega$ ,  $P(B) = P(B \cap A) + P(B \cap \bar{A})$ .

Exemple : On dispose de trois urnes contenant chacune cinq boules, rouges ou noires. La première contient 3 rouges et 2 noires, la deuxième 2 rouges et 3 noires et la troisième 1 rouge et 4 noires. Julien lance un dé bien équilibré.

S'il obtient « 1 », il extrait au hasard une boule de l'urne 1.

S'il obtient « 3 ou 5 », il extrait au hasard une boule de l'urne 2.

S'il obtient « 2, 4 ou 6 », il extrait au hasard une boule de l'urne 3.

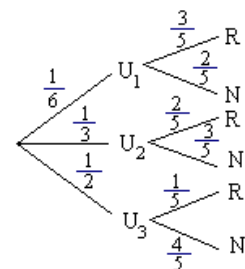
- Quelle est la probabilité que la boule tirée soit rouge et provienne de l'urne 1 ?
- Quelle est la probabilité que la boule tirée soit rouge ?

Solution :

- Illustrons cette situation par un arbre pondéré où  $U_k$  ( $k \in \{1, 2, 3\}$ ), et  $R$  désignent respectivement les événements : « la boule est extraite de l'urne  $k$  » et « la boule obtenue est rouge » :

La probabilité cherchée est  $P(R \cap U_1)$ .  $R \cap U_1$  est représenté par la branche supérieure de l'arbre.

On a :  $P(R \cap U_1) = P_{U_1}(R) \times P(U_1)$  et donc  $P(R \cap U_1) = \frac{3}{5} \times \frac{1}{6} = \frac{1}{10}$ .



- Les événements  $U_1, U_2$  et  $U_3$  représentent une partition de  $\Omega$ .

On a alors :  $P(R) = P(R \cap U_1) + P(R \cap U_2) + P(R \cap U_3)$

$P(R) = P_{U_1}(R) \times P(U_1) + P_{U_2}(R) \times P(U_2) + P_{U_3}(R) \times P(U_3)$

d'où  $P(R) = \frac{3}{5} \times \frac{1}{6} + \frac{2}{5} \times \frac{1}{3} + \frac{1}{5} \times \frac{1}{2} = \frac{1}{3}$ .

### 3 Événements indépendants

Une loi de probabilité est définie sur un ensemble d'éventualité  $\Omega$ .

#### 3.1 Indépendance de deux événements

Définition : Dire que deux événements  $A$  et  $B$  sont indépendants signifie que :  $P(A \cap B) = P(A) \times P(B)$ .

Remarque : Si  $A$  et  $B$  sont indépendants et de probabilités non nulles alors :

$$P_A(B) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B) \text{ et de même } P_B(A) = P(A).$$

Propriété : Si deux événements  $A$  et  $B$  sont indépendants, alors :

1. les événements  $\bar{A}$  et  $B$  sont indépendants
2. les événements  $A$  et  $\bar{B}$  sont indépendants
3. les événements  $\bar{A}$  et  $\bar{B}$  sont indépendants

Démonstration :  $A$  et  $B$  sont indépendants, donc  $P(A \cap B) = P(A)P(B)$ .

1.  $P(\bar{A} \cap B) = P(B) - P(A \cap B) = P(B) - P(A)P(B) = P(B)(1 - P(A)) = P(B)P(\bar{A})$   
donc  $\bar{A}$  et  $B$  sont indépendants.
2. C'est la même démonstration, il suffit de changer  $A$  et  $B$ .
3. d'après 1.,  $\bar{A}$  et  $B$  sont indépendants, et d'après 2.,  $\bar{A}$  et  $\bar{B}$  sont indépendants.

Exemple : On fait l'hypothèse que chacun des moteurs d'un avion bi-moteur tombe en panne avec une probabilité égale à 0,0001 et ceci d'une façon indépendante de l'autre moteur. Quelle est la probabilité que l'avion arrive à bon port sachant qu'il peut voler avec un seul moteur ?

Solution : On note  $A$  l'événement « le premier moteur tombe en panne » et  $B$  l'événement « le deuxième moteur tombe en panne ».

Les deux événements  $A$  et  $B$  étant indépendants, la probabilités que les deux moteurs tombent en panne pendant le même vol est :  $P(A \cap B) = P(A)P(B) = 0,0001^2 = 10^{-8}$ .

La probabilité que l'avion arrive à bon port est donc l'événement contraire  $\overline{A \cap B}$ , soit  $1 - 10^{-8} = 0,99999999$ .

#### 3.2 Indépendance de deux variables aléatoires

Définition :  $X$  et  $Y$  sont deux variables aléatoires sur  $\Omega$ .

On note  $x_1, x_2, \dots, x_n$  les valeurs prises par  $X$  et  $y_1, y_2, \dots, y_n$  les valeurs prises par  $Y$ . Dire que  $X$  et  $Y$  sont indépendantes signifie que pour tous  $i$  et  $j$  les événements  $(X = x_i)$  et  $(Y = y_j)$  sont indépendants.

Exemple : Une chaîne fabrique des boîtes rectangulaires.  $L$  et  $l$  sont les variables aléatoires égales respectivement à la longueur et la largeur (en cm) d'une boîte tirée au hasard en sortie de chaîne. Les lois de probabilités de  $L$  et  $l$  sont :

$x_i$	13,9	14	14,1
$P(L = x_i)$	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$

$y_i$	7,9	8	8,1
$P(l = y_i)$	$\frac{1}{8}$	$\frac{3}{4}$	$\frac{1}{8}$

$L$  et  $l$  sont indépendantes. On s'intéresse à la probabilité de  $E$  « le périmètre de la boîte est de 44 cm exactement ».

On a  $E = ((L = 13,9) \cap (l = 8,1)) \cup ((L = 14) \cap (l = 8)) \cup ((L = 14,1) \cap (l = 7,9))$ .

$E$  est la réunion de trois événements disjoints et par l'hypothèse d'indépendance :

$$P(E) = P(L = 13,9) \times P(l = 8,1) + P(L = 14) \times P(l = 8) + P(L = 14,1) \times P(l = 7,9)$$

$$P(E) = \frac{1}{6} \times \frac{1}{8} + \frac{2}{3} \times \frac{3}{4} + \frac{1}{6} \times \frac{1}{8} = \frac{13}{24}.$$

### 4 Dénombrement

#### 4.1 Différents tirages dans une urne

De nombreuses expériences aléatoires peuvent s'assimiler à des tirages de boules dans une urne et se modélisent par une loi équirépartie.

Ce sont des situations de référence :

– **Tirages successifs avec remise**

On tire une boule de l'urne U, on note son numéro, puis on la remet dans U. On effectue de la sorte p tirages (dits *successifs avec remise*).

Le nombre de listes de longueur p formées de numéros de U est n^p.

– **Tirages successifs sans remise**

On tire une boule de l'urne U, on note son numéro et on ne la remet pas dans l'urne. On effectue de la sorte p tirages, avec p ≤ n (dits *successifs sans remise*).

Le nombre de listes formées de p numéros distincts pris dans U est égal à n(n - 1) ... (n - p + 1).

Cas particulier : Lorsque p = n, toutes les boules sont tirées une à une.

Le nombre de listes formées des n numéros de U est égal à n(n - 1) ... 2 × 1.

Définition : Pour n ∈ ℕ, n ≥ 2, le nombre n(n - 1) ... 2 × 1 est noté n! (il se lit « factorielle n »).

Par convention : 1! = 1 et 0! = 1.

– **Tirages simultanés**

On tire simultanément p boules de l'urne U (p ≤ n). On obtient ainsi un ensemble de p numéros pris parmi n, que l'on appelle combinaison.

Théorème : Le nombre de combinaisons de p objets pris parmi n (p ≤ n) est noté  $\binom{n}{p}$  et est égal à

$$\frac{n(n - 1) \dots (n - p + 1)}{p!} = \frac{n!}{p!(n - p)!}.$$

On le lit « p parmi n ».

Exercice : Au Loto national, un joueur coche 6 numéros sur une grille où figurent les nombres 1 à 49. De combien de façons le joueur peut-il remplir sa grille ?

Solution : L'ordre dans lequel les 6 numéros sont choisis n'intervient pas. Remplir une grille de Loto revient alors à choisir 6 nombres parmi 49, c'est à dire à constituer une combinaison de 6 éléments de {1, 2, 3, ..., 49}.

Il y a donc  $\binom{49}{6}$  possibilités pour remplir une grille de Loto.

$$\binom{49}{6} = \frac{49 \times 48 \times 47 \times 46 \times 45 \times 44}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = 13\,983\,816.$$

Il y a donc 13 983 816 grilles différentes au Loto.

## 4.2 Propriétés des combinaisons

Propriété 1 : Pour tout entier naturel n non nul :  $\binom{n}{0} = 1$  ;  $\binom{n}{n} = 1$  ;  $\binom{n}{1} = n$ .

Propriété 2 : Pour tous entiers naturels n et p tels que p ≤ n :  $\binom{n}{p} = \binom{n}{n - p}$

En effet choisir de prendre p objets parmi n revient au même que de prendre n - p objets parmi n.

Propriété 3 : Pour tous entiers naturels n et p tels que 1 ≤ p ≤ n :  $\binom{n - 1}{p - 1} + \binom{n - 1}{p} = \binom{n}{p}$ .

Démonstration :

1. E est un ensemble à n éléments, on note a un élément fixé de E. Pour dénombrer les parties à p éléments de E, on peut distinguer :

– celles qui ne contiennent pas a : ce sont les parties à p éléments choisis parmi les n - 1 éléments restants; elles sont au nombre de  $\binom{n - 1}{p}$  ;

– celles qui contiennent a : il reste à choisir p - 1 éléments parmi les n - 1 éléments restants; on en compte :  $\binom{n - 1}{p - 1}$ .

Or il y a  $\binom{n}{p}$  parties à p éléments de E, donc  $\binom{n}{p} = \binom{n - 1}{p - 1} + \binom{n - 1}{p}$ .

2. deuxième façon de démontrer la propriété :

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

$$\binom{n-1}{p-1} + \binom{n-1}{p} = \frac{n(n-1)!}{p!(n-p)!} = \frac{n!}{p!(n-p)!} = \binom{n}{p}.$$

Conséquence : Le triangle de Pascal

On peut calculer les  $\binom{n}{p}$  de proche en proche à l'aide du tableau ci-contre. Pour cela :

- on place des 1 sur la colonne «  $p = 0$  » et la diagonale
  - on obtient un autre nombre du tableau en additionnant le nombre juste au-dessus et celui situé à gauche sur la ligne précédente
- On utilise en fait la propriété 3.

$n \setminus p$	0	1	2	3	4	5
0	1					
1	1	1				
2	1	2	1			
3	1	3	3	1		
4	1	4	6	4	1	
5	1	5	10	10	5	1

### 4.3 Formule du binôme

Propriété : Pour tous nombres complexes  $a$  et  $b$ , et pour tout entier  $n \geq 1$ ,

$$(a+b)^n = a^n + \binom{n}{1} a^{n-1}b + \binom{n}{2} a^{n-2}b^2 + \dots + \binom{n}{n-1} ab^{n-1} + b^n$$

$$= \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}.$$

Exemple : Calculer  $(5+i)^5$ .

$$(5+i)^5 = 5^5 + \binom{5}{1} 5^4 i + \binom{5}{2} 5^3 i^2 + \binom{5}{3} 5^2 i^3 + \binom{5}{4} 5 i^4 + i^5 = 3125 + 3125i - 1250 - 250i + 25 + i$$

$$(5+i)^5 = 1900 + 2876i.$$

Exercice :

1. Démontrer que pour tout entier  $n \geq 1$ ,  $\sum_{k=0}^n \binom{n}{k} = 2^n$ .
2. En déduire le nombre de toutes les parties d'un ensemble à  $n$  éléments.

Solution :

1. D'après la formule du binôme :  $(1+1)^n = \sum_{k=0}^n \binom{n}{k} \times 1^k \times 1^{n-k}$  donc  $2^n = \sum_{k=0}^n \binom{n}{k}$ .
2. Le nombre de toutes les parties d'un ensemble à  $n$  éléments est la somme du nombre de parties à 0 élément, à 1 élément, à 2 éléments, ..., à  $n$  éléments.

Le nombre total est donc  $\sum_{k=0}^n \binom{n}{k}$ , c'est à dire  $2^n$  d'après la question 1.

## 5 Lois de probabilité discrètes

### 5.1 loi de Bernoulli

Définition : Une épreuve de Bernoulli est une expérience aléatoire qui ne comporte que deux issues appelées succès (noté  $S$ ) et échec (noté  $\bar{S}$ ), de probabilités respectives  $p$  et  $1-p$ .

La loi de probabilité ci-contre est appelée loi de Bernoulli de paramètre  $p$ .

issue	$S$	$\bar{S}$
probabilité	$p$	$1-p$

Remarque : Si  $X$  est une variable aléatoire prenant les valeurs 0 ou 1, telle que  $P(X=1) = p$  et  $P(X=0) = 1-p$ , alors on dit que  $X$  est une variable de Bernoulli de paramètre  $p$  (ou encore que  $X$  suit une loi de Bernoulli de paramètre  $p$ ). On note  $X$  suit  $\mathcal{B}(p)$ .

### 5.2 Loi binomiale

Définition : Un schéma de Bernoulli est la répétition d'épreuves de Bernoulli identiques dans des conditions d'indépendance.

Propriété : Un schéma de Bernoulli est constitué de  $n$  épreuves indépendantes.  $X$  est la variable aléatoire qui à chaque liste de  $n$  résultats associe le nombre de succès.

Alors pour tout entier  $k$ , avec  $0 \leq k \leq n$ ,  $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ .

La loi de probabilité de la variable aléatoire  $X$  est appelée loi binomiale de paramètres  $n$  et  $p$ . Cette loi est notée  $\mathcal{B}(n; p)$ .

Démonstration : Chaque liste formée de  $k$  succès et donc de  $n - k$  échecs, a pour probabilité :  $p^k (1-p)^{n-k}$ . Le nombre de telles listes est égal au nombre de façons différentes de choisir la position des  $k$  succès parmi les  $n$  résultats. Il y a  $\binom{n}{k}$  telles listes. Donc  $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ .

Remarque : d'après la formule du binôme,  $\sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1$ .

Propriété : La loi binomiale de paramètres  $n$  et  $p$  a pour :

- espérance mathématique :  $E(X) = np$
- variance :  $V(X) = np(1-p)$
- écart-type :  $\sigma(X) = \sqrt{np(1-p)}$ .

Exercice : Un élève répond au hasard aux 10 questions d'un Q.C.M. Pour chaque question, 5 réponses sont proposées dont une seule est exacte.  $X$  est la variable aléatoire égale au nombre de bonnes réponses.

1. Montrer que la loi de probabilité de  $X$  est une loi binomiale.
2. Calculer la probabilité d'avoir au moins 5 bonnes réponses.
3. Calculer l'espérance mathématique du nombre de bonnes réponses.

Solution :

1. Chaque question est une épreuve de Bernoulli où le succès est « la réponse est exacte » ; alors  $p = \frac{1}{5}$ . Le Q.C.M. est la répétition de 10 épreuves identiques et indépendantes ; il correspond alors à un schéma de Bernoulli. Le nombre  $X$  de bonnes réponses au Q.C.M. est une variable aléatoire de loi binomiale de paramètres 10 et  $\frac{1}{5}$ .
2.  $P(X \geq 5) = P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10)$   
 $= 252 \left(\frac{1}{5}\right)^5 \left(\frac{4}{5}\right)^5 + 210 \left(\frac{1}{5}\right)^6 \left(\frac{4}{5}\right)^4 + 120 \left(\frac{1}{5}\right)^7 \left(\frac{4}{5}\right)^3 + 45 \left(\frac{1}{5}\right)^8 \left(\frac{4}{5}\right)^2 + 10 \left(\frac{1}{5}\right)^9 \left(\frac{4}{5}\right)^1 + \left(\frac{1}{5}\right)^{10}$ .  
D'où  $P(X \geq 5) \simeq 0,03$ .
3.  $E(X) = 10 \times \frac{1}{5} = 2$ .  
Ainsi, en moyenne, un candidat peut espérer avoir 2 réponses exactes à ce Q.C.M.

## 6 Equirépartition

On considère une épreuve aléatoire ayant  $k$  issues équiprobables  $\omega_1, \omega_2, \dots, \omega_k$ .

On peut la représenter par un tirage dans une urne de Bernoulli contenant  $k$  boules numérotées de 1 à  $k$ .

**On répète  $n$  fois cette épreuve dans les mêmes conditions.**

On note  $f_1, f_2, \dots, f_k$  les fréquences d'apparition de chaque boule dans l'échantillon.

On peut s'attendre à ce que les valeurs  $f_1, f_2, \dots, f_k$  diffèrent peu de  $\frac{1}{k}$  puisque les  $k$  boules ont par définition autant de chances d'apparaître les unes que les autres.

On va alors mesurer l'écart entre les valeurs  $f_1, f_2, \dots, f_k$  et  $\frac{1}{k}$ , on pose :

$$d^2 = \left(f_1 - \frac{1}{k}\right)^2 + \left(f_2 - \frac{1}{k}\right)^2 + \dots + \left(f_k - \frac{1}{k}\right)^2.$$

Cette quantité est toujours positive et n'est nulle que si toutes les fréquences sont égales à  $\frac{1}{k}$ . Elle est d'autant plus petite que les fréquences sont proches de  $\frac{1}{k}$ .

Théorème : Quand on répète  $n$  fois ( $n \geq 100$ ) une même épreuve ayant  $k$  issues équiprobables et qu'on note  $f_1, f_2, \dots, f_k$  les fréquences d'apparition de ces issues, alors il est **très rare** (moins de 5% des échantillons possibles) que :  $\left(f_1 - \frac{1}{k}\right)^2 + \left(f_2 - \frac{1}{k}\right)^2 + \dots + \left(f_k - \frac{1}{k}\right)^2$  dépasse  $\frac{2}{n}$

Remarque : Notons  $n_1, n_2, \dots, n_k$  le nombre d'apparitions, respectivement de  $\omega_1, \omega_2, \dots, \omega_k$ .

Alors :  $n_i = n f_i$ ; donc  $n_i - \frac{n}{k} = n \left( f_i - \frac{1}{k} \right)$ .

D'après le théorème, il est très rare que :  $\left( n_1 - \frac{n}{k} \right)^2 + \left( n_2 - \frac{n}{k} \right)^2 + \dots + \left( n_k - \frac{n}{k} \right)^2$  dépasse  $2n$ .

Conséquence : Test d'équipartition

Considérons une épreuve aléatoire admettant  $k$  issues. On voudrait savoir si ces issues sont équiprobables.

Plus précisément, on se demande si le modèle attribuant à chaque issue la même probabilité  $\frac{1}{k}$  est adapté à cette épreuve.

Pour cela, on observe les résultats de  $n$  répétitions de cette épreuve, on mesure les fréquences d'apparition  $f_1, f_2, \dots, f_k$  de chacune des issues et on calcule la valeur de  $d^2$ .

– Si  $d^2$  est inférieur à  $\frac{2}{n}$ , on dira que l'échantillon est compatible avec l'hypothèse d'équiprobabilité; en effet, sous cette hypothèse, plus de 95% des échantillons possibles vérifient cette condition.

On retiendra donc comme plausible l'hypothèse d'équiprobabilité.

– Si  $d^2$  est supérieur à  $\frac{2}{n}$ , on dira que cet échantillon n'est pas compatible avec cette hypothèse.

On considérera donc que l'hypothèse d'équiprobabilité est peu vraisemblable.

Bien entendu, dans les deux cas, on n'a pas de garantie absolue et il subsiste toujours un risque de se tromper.

On dira que la conclusion a un **niveau de confiance de 95%**.

Remarque : On pourrait prendre un autre niveau de confiance. La valeur  $\frac{2}{n}$  du théorème devrait alors être modifiée :

- pour un niveau de confiance de 90%, il faudrait la remplacer par  $\frac{1,6}{n}$ ;
- pour un niveau de confiance de 99%, il faudrait la remplacer par  $\frac{3,4}{n}$ .

Exemple : Un joueur lance un dé cubique 1000 fois et on obtient la distribution de fréquences suivante :

valeur $i$	1	2	3	4	5	6
fréquence $f_i$	0,186	0,165	0,162	0,149	0,190	0,148

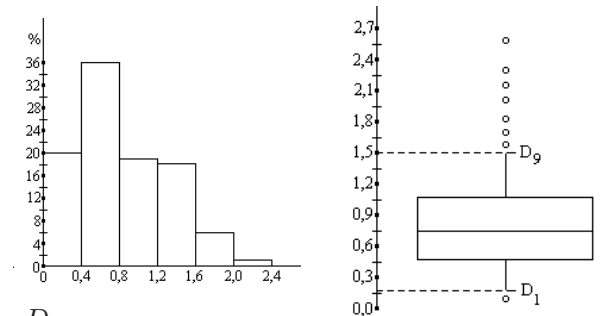
Ce joueur veut savoir si le dé est équilibré, c'est à dire si les résultats obtenus sont compatibles avec le modèle d'équiprobabilité sur  $\{1,2,3,4,5,6\}$ .

On note  $d_{obs}^2 = \left( f_1 - \frac{1}{6} \right)^2 + \left( f_2 - \frac{1}{6} \right)^2 + \left( f_3 - \frac{1}{6} \right)^2 + \left( f_4 - \frac{1}{6} \right)^2 + \left( f_5 - \frac{1}{6} \right)^2 + \left( f_6 - \frac{1}{6} \right)^2$ .

$d_{obs}^2 \simeq 0,001603 > \frac{1,6}{1000}$ .

On rejette alors l'hypothèse que le dé est bien équilibré avec un niveau de confiance de 90%.

Deuxième test possible : On simule un grand nombre de fois (500 fois, par exemple) une série de 1000 lancers d'un dé bien équilibré et on regarde comment se situe la valeur ( $\simeq 1,603$ ) de  $1000d_{obs}^2$  par rapport aux valeurs de  $1000d^2$  obtenues dans la simulation. L'histogramme et le graphique (diagramme) en boîte ci-contre représente la série des 500 valeurs de  $1000d^2$  obtenues par simulation.



L'histogramme ne permet pas de trouver la valeur du 9<sup>ème</sup> décile  $D_9$ .

On peut seulement dire que le neuvième décile est dans l'intervalle

$]1,2; 1,6[$ . Avec le diagramme en boîte, on obtient  $D_9 = 1,5$ .

Règle de décision

- si  $1000d_{obs}^2 < D_9$ , on accepte le modèle d'équipartition, c'est à dire que le considère que les résultats observés sont compatibles avec l'hypothèse que le dé est bien équilibré;
- si  $1000d_{obs}^2 > D_9$ , on rejette l'hypothèse que le dé est bien équilibré au risque d'erreur de 10%.

Dans l'exemple, on rejette donc l'hypothèse que le dé est bien équilibré avec un risque d'erreur de 10%.