

# PROBABILITÉS ET STATISTIQUE

ECOLE CENTRALE DE LYON

PREMIÈRE ANNÉE

## Table des matières

<b>I</b>	<b>Théorie des probabilités</b>	<b>3</b>
<b>1</b>	<b>Modèle probabiliste</b>	<b>4</b>
1.1	Univers et événements. . . . .	4
1.2	Probabilités. . . . .	4
1.3	Probabilités conditionnées par un événement. . . . .	8
1.4	Indépendance deux à deux ou mutuelle d'événements. . . . .	9
<b>2</b>	<b>Variables aléatoires</b>	<b>10</b>
2.1	Définition. . . . .	10
2.2	Lois discrètes. . . . .	11
2.2.1	Définition, première propriété et exemple. . . . .	11
2.2.2	Une seule issue : la mesure de Dirac. . . . .	12
2.2.3	Deux issues : la loi de Bernoulli. . . . .	12
2.2.4	Issues équiprobables : la loi uniforme. . . . .	12
2.2.5	Tirages sans remise : la loi hypergéométrique. . . . .	12
2.2.6	Tirages avec remise : la loi binomiale. . . . .	13
2.2.7	Rang du premier succès : la loi géométrique. . . . .	14
2.2.8	Loi de Poisson . . . . .	15
2.3	Lois à densité. . . . .	15
2.3.1	Définition . . . . .	15
2.3.2	Loi uniforme. . . . .	16
2.3.3	Loi exponentielle. . . . .	16
2.3.4	Loi normale ou gaussienne. . . . .	16
2.4	Fonction de répartition d'une variable aléatoire . . . . .	16
<b>3</b>	<b>Moments de variables aléatoires</b>	<b>18</b>
3.1	Définitions . . . . .	18
3.2	Propriétés de l'espérance . . . . .	20
3.3	Changement de variable. . . . .	20
3.3.1	Changement de variable en discret. . . . .	21
3.3.2	Changement de variable par la fonction de répartition. . . . .	22
3.3.3	Changement de variable pour les v.a. à densité. . . . .	22
3.4	Variance et écart-type d'une variable aléatoire réelle. . . . .	23

<b>4</b>	<b>Vecteurs aléatoires</b>	<b>25</b>
4.1	Définition . . . . .	25
4.2	Lois discrètes. . . . .	25
4.2.1	Définition . . . . .	25
4.2.2	Tirages avec plus de deux couleurs : la loi multinomiale . . . . .	26
4.3	Lois à densité. . . . .	26
4.3.1	Définition. . . . .	26
4.3.2	Loi uniforme sur un ensemble de mesure bornée. . . . .	26
4.4	Lois marginales. . . . .	26
4.5	Moments de vecteurs aléatoires. . . . .	28
4.5.1	Formules de transfert . . . . .	28
4.5.2	Covariance de deux variables réelles. . . . .	28
4.5.3	Espérance d'un vecteur aléatoire. . . . .	30
4.5.4	Covariance de deux vecteurs aléatoires. . . . .	30
4.5.5	Variance (ou covariance) d'un vecteur aléatoire. . . . .	30
4.6	Variables aléatoires indépendantes. . . . .	31
4.7	Somme de v.a. indépendantes. . . . .	32
<b>5</b>	<b>Suites de variables aléatoires - Résultats asymptotiques</b>	<b>35</b>
5.1	Un exemple important. . . . .	35
5.2	Différents types de convergence. . . . .	35
5.3	Théorèmes de convergence. . . . .	36
5.4	Simulation de variables aléatoires . . . . .	40
5.4.1	Simulation d'une loi uniforme sur $[0, 1[$ . . . . .	40
5.4.2	Simulation d'une v.a. de Bernoulli . . . . .	41
5.4.3	Simulation d'une v.a. binomiale . . . . .	41
5.4.4	Simulation d'une v.a. admettant une densité continue. . . . .	41
5.4.5	Simulation d'une v.a. bornée à densité bornée. . . . .	41
5.4.6	Simulation d'une loi normale . . . . .	42
5.5	Méthode de Monte Carlo pour le calcul des intégrales et des moments. . . . .	42
5.5.1	Méthode . . . . .	43
5.5.2	Précision . . . . .	43
5.5.3	Intérêts de la méthode . . . . .	44
5.6	Méthode de "simulation Monte Carlo". . . . .	44
5.7	Rééchantillonnage (bootstrap) . . . . .	45
<b>II</b>	<b>Statistique inférentielle</b>	<b>46</b>
<b>6</b>	<b>Estimateurs, estimation par intervalle de confiance</b>	<b>46</b>
6.1	Echantillons . . . . .	46
6.1.1	Problématique type et vocabulaire de base . . . . .	46
6.1.2	Formalisation. . . . .	46
6.1.3	Définition. . . . .	46
6.2	Moments empiriques d'un échantillon. . . . .	46
6.2.1	Définition et premières propriétés. . . . .	46
6.2.2	Cas des échantillons gaussiens. . . . .	48
6.2.3	Loi du chi-deux. . . . .	48
6.2.4	Loi de Student. . . . .	49
6.3	Fonction de répartition empirique d'un échantillon . . . . .	50
6.4	Estimation par intervalle de confiance . . . . .	51
6.4.1	Introduction et définition. . . . .	51

6.4.2	Exemple de construction d'un intervalle de confiance. . . . .	51
6.4.3	Cas des variables gaussiennes. . . . .	52
6.4.4	Cas de variables non gaussiennes de carré intégrable. . . . .	53
<b>7</b>	<b>Théorie des tests</b>	<b>54</b>
7.1	Un exemple : les faiseurs de pluie. . . . .	54
7.2	Notions générales. . . . .	56
7.3	Tests paramétriques. . . . .	57
7.3.1	Test entre deux hypothèses simples. . . . .	57
7.3.2	Test d'une hypothèse simple contre une hypothèse composite : la fonction puissance. . . . .	58
7.4	Tests d'ajustement. . . . .	58
7.4.1	Test d'ajustement du Chi-deux. . . . .	59
7.4.2	Test d'ajustement du Chi-deux avec estimation de paramètres. . . . .	61
7.4.3	Test d'ajustement de Kolmogorov-Smirnov. . . . .	61
7.5	Tests de comparaison entre échantillons indépendants. . . . .	62
7.5.1	Tests paramétriques pour la comparaison de deux échantillons. . . . .	62
7.5.2	Test non paramétrique de comparaison de deux échantillons ou plus : le test du Chi-deux. . . . .	63
7.5.3	Test d'indépendance du chi-deux. . . . .	64
<b>A</b>	<b>Eléments de théorie des ensembles et de théorie de la mesure</b>	<b>65</b>
A.1	Ensembles dénombrables . . . . .	65
A.2	Ensembles négligeables. Intégrale des fonctions nulles presque partout. . . . .	65
A.3	Mesure de Lebesgue des ensembles de $\mathbb{R}^n$ . . . . .	65
<b>B</b>	<b>Lois usuelles</b>	<b>66</b>
B.1	Lois discrètes. . . . .	66
B.1.1	Lois, espérances, variances. . . . .	66
B.1.2	Propriétés. . . . .	69
B.2	Lois à densité . . . . .	70
B.2.1	Lois, espérances, variances. . . . .	70
B.2.2	Propriétés. . . . .	74
<b>C</b>	<b>Intervalles de Confiance</b>	<b>75</b>
C.1	IC sur la moyenne et la variance d'un échantillon gaussien. . . . .	75
C.2	IC pour le paramètre d'un échantillon d'une loi de Bernoulli. . . . .	76
C.3	IC pour la moyenne pour d'un échantillon non gaussien de carré intégrable. . . . .	76
<b>D</b>	<b>Tests statistiques</b>	<b>76</b>
D.1	Tests paramétriques pour un échantillon. . . . .	76
D.1.1	Tests sur la moyenne et la variance d'un échantillon gaussien. . . . .	77
D.1.2	Test sur le paramètre d'un échantillon d'une loi de Bernoulli. . . . .	77
D.1.3	Tests sur la moyenne d'un échantillon quelconque. . . . .	77
D.2	Tests paramétriques de comparaison d'échantillons indépendants. . . . .	78
D.2.1	Cas des échantillons gaussiens. . . . .	78
D.2.2	Cas des échantillons de Bernoulli. . . . .	78
D.2.3	Cas des échantillons quelconques. . . . .	79
D.3	Tests non paramétriques. . . . .	79
D.3.1	Tests d'ajustement d'un échantillon à une loi donnée. . . . .	79
D.3.2	Test de provenance d'échantillons d'une même population : le test du chi-deux. . . . .	80
D.3.3	Test d'indépendance : le test du chi-deux. . . . .	80

## Première partie

# Théorie des probabilités

Les parties vues en classes préparatoires (à peu près tout ce qui concerne le modèle discret) ne seront pas refaites en cours et sont considérées comme acquises.

## 1 Modèle probabiliste

### 1.1 Univers et événements.

Pour modéliser un *phénomène aléatoire* ou une *expérience aléatoire*, on considère qu'il existe un ensemble non vide  $\Omega$ , appelé l'*univers* ou l'*ensemble des possibles*, dont les éléments sont tous les états du monde possibles.

Un *événement* est un ensemble d'états du monde, c'est-à-dire un sous-ensemble de  $\Omega$ . En général, cet événement s'exprime par une assertion à laquelle on peut répondre par oui ou par non (proposition logique).

**Exemple :** « Je lance un dé non pipé et j'obtiens le résultat 3 » ; « Il fera 23 degrés demain à Lyon à 14h ».

Si l'on se trouve dans un état du monde  $\omega \in A$ , on dit que l'événement  $A$  est réalisé.

On note  $\mathcal{P}(\Omega)$  l'ensemble des parties de  $\Omega$ .

#### Opérations sur les événements

Soient  $A$  et  $B$  deux événements.

- $A^c$  est l'événement *contraire* de  $A$ .
- $A \cup B$  est l'événement «  $A$  ou  $B$  ».
- $A \cap B$  est l'événement «  $A$  et  $B$  ».
- On note  $A \setminus B = A \cap B^c$  l'événement «  $A$  et non  $B$  ».
- Lorsque  $A$  et  $B$  sont deux événements disjoints ( $A \cap B = \emptyset$ ), on dit que  $A$  et  $B$  sont *incompatibles*.
- $A \subset B$  signifie  $A \Rightarrow B$ .

**Exercice :** Exprimer, avec des opérations ensemblistes, les événements suivants :

1.  $A$  ou  $B$  sont réalisés mais pas  $C$ .
  2. L'un des  $A_n, n \geq 1$ , est réalisé.
  3. Tous les  $A_n, n \geq 1$ , sont réalisés.
1.  $(A \cup B) \cap C^c$ .
  2.  $\bigcup_{n \geq 1} A_n$ .
  3.  $\bigcap_{n \geq 1} A_n$ .

### 1.2 Probabilités.

#### Définition 1.1

On appelle *probabilité* sur  $\Omega$  toute application

$$\begin{aligned} \mathbf{P} : \mathcal{P}(\Omega) &\longrightarrow [0, 1] \\ A &\longmapsto \mathbf{P}[A] \end{aligned}$$

satisfaisant

- (i)  $\mathbf{P}[\Omega] = 1$  ;
- (ii) pour toute suite  $(A_n)_{n \geq 1}$  d'événements deux à deux incompatibles (disjoints) :

$$\mathbf{P} \left[ \bigcup_{n \geq 1} A_n \right] = \sum_{n \geq 1} \mathbf{P}[A_n].$$

Lorsque l'on modélise une expérience aléatoire, on doit préciser sa *loi*, c'est-à-dire la probabilité que n'importe quel événement se produise.

**Exemple :** [Expérience dont les résultats possibles sont en nombre fini et équiprobables] On note  $A_i$  l'événement « on a obtenu le résultat  $i$  », pour  $i = 1$  à  $n$ . Par exemple, on lance un dé à 6 faces non pipé et on s'intéresse au résultat obtenu. Puisque les résultats sont équiprobables, notons

$$p = \mathbf{P}[A_i] \text{ pour tout } i \in \{1, \dots, n\}.$$

Alors, comme les événements forment une partition,

$$\mathbf{P}[\Omega] = \mathbf{P}\left[\bigcup_{i=1}^n A_i\right] = \sum_{i=1}^n \mathbf{P}[A_i] = np$$

et comme  $\mathbf{P}[\Omega] = 1$ ,  $p = 1/n$ .

Si  $A$  est un événement du type « on a obtenu l'un des résultats  $i_1, \dots, i_k$  » où les  $i_k$  sont tous distincts, alors

$$\mathbf{P}[A] = \mathbf{P}\left[\bigcup_{j=1}^k A_{i_j}\right] = \sum_{j=1}^k \mathbf{P}[A_{i_j}] = kp = \frac{k}{n}.$$

On obtient donc la formule célèbre « nombre de cas favorables sur nombre de cas possibles ». Autrement dit, dans le cas d'expériences dont les issues sont équiprobables, le calcul des probabilités se réduit à un dénombrement.

**Remarque :**

- La remarque précédente est fausse lorsque les issues ne sont pas équiprobables.
- Il n'existe pas d'expérience aléatoire ayant un ensemble infini d'issues équiprobables.

**Exemple :** [Expérience dont les résultats sont en nombre fini] Cette fois-ci les  $\mathbf{P}[A_i]$  dépendent du résultat obtenu. Par exemple, dans une famille de deux enfants, quelles sont les probabilités que ces enfants soient deux garçons, deux filles, un garçon et une fille ?

$$\mathbf{P}[2 \text{ garçons}] = \mathbf{P}[2 \text{ filles}] = 1/4$$

$$\mathbf{P}[\text{une fille et un garçon}] = 1/2.$$

**Exemple :** [Expérience dont les résultats sont dénombrables] On peut alors toujours modéliser l'expérience par la donnée d'une partition dénombrable de  $\Omega : \{A_i, i \in \mathbb{N}\}$  ainsi que de la suite  $(p_i)_{i \in \mathbb{N}}$  des probabilités des événements  $A_i$ . Bien sûr, on a alors

$$\mathbf{P}[\Omega] = 1 = \sum_{i=1}^{+\infty} \mathbf{P}[A_i] = \sum_{i=1}^{+\infty} p_i \Rightarrow \sum_{i=1}^{+\infty} p_i = 1.$$

et, pour tout  $A = A_{i_1} \cup \dots \cup A_{i_j} \cup \dots$ , pour  $i_1, \dots, i_j, \dots \in \mathbb{N}$ ,

$$\mathbf{P}[A] = \sum_{j=1}^{\infty} \mathbf{P}[A_{i_j}] = \sum_{j=1}^{\infty} p_{i_j}.$$

Par exemple, on lance une infinité de fois une pièce de monnaie équilibrée et on note  $A_k$ , pour  $k \geq 1$ , l'événement « la première apparition de « pile » est le  $k$ -ième lancer ».

On a alors, pour tout  $k \geq 1$ ,  
 $p_k = \mathbf{P}[A_k] = 1/2^k$ .  
 Remarquons que l'on a bien

$$\sum_{k=1}^{+\infty} \frac{1}{2^k} = \frac{1}{2} \cdot 2 = 1.$$

### Proposition 1.1

Soit  $\mathbf{P}$  une probabilité sur  $\Omega$ . On a les propriétés :

1.  $\mathbf{P}[\emptyset] = 0$  ;
2. pour tout  $A \subset \Omega$ ,  $\mathbf{P}[A^c] = 1 - \mathbf{P}[A]$  ;
3. pour tous  $A, B \subset \Omega$ , si  $A \subset B$  alors

$$\mathbf{P}[A] \leq \mathbf{P}[B] \text{ et } \mathbf{P}[B \setminus A] = \mathbf{P}[B] - \mathbf{P}[A].$$

#### Démonstration :

1.  $\mathbf{P}[\emptyset] = \mathbf{P}[\emptyset \cup \emptyset] = 2\mathbf{P}[\emptyset]$  donc  $\mathbf{P}[\emptyset] = 0$ .
2.  $1 = \mathbf{P}[\Omega] = \mathbf{P}[A] + \mathbf{P}[A^c] \Rightarrow \mathbf{P}[A^c] = 1 - \mathbf{P}[A]$ .
3.  $\mathbf{P}[B] = \mathbf{P}[A] + \mathbf{P}[B \cap A^c] \Rightarrow \mathbf{P}[B] \geq \mathbf{P}[A]$  et  $\mathbf{P}[B \cap A^c] = \mathbf{P}[B] - \mathbf{P}[A]$ .

### Proposition 1.2

Pour tous  $A, B \subset \Omega$ ,  $\mathbf{P}[A \cup B] = \mathbf{P}[A] + \mathbf{P}[B] - \mathbf{P}[A \cap B]$  ; plus généralement, si  $A_1, \dots, A_n$  sont  $n$  événements, on a la formule du crible (ou formule de Poincaré) suivante :

$$\mathbf{P}\left[\bigcup_{i=1}^n A_i\right] = \sum_{k=1}^n (-1)^{k+1} \sum_{I \in \mathcal{P}_k(n)} \mathbf{P}\left[\bigcap_{i \in I} A_i\right]$$

où  $\mathcal{P}_k(n)$  est l'ensemble des parties à  $k$  éléments de l'ensemble  $\{1, \dots, n\}$ .

#### Démonstration :

Idée de la preuve (qui se fait par récurrence à partir de  $n = 2$ ). Dans le cas  $n = 2$ , on a  $\mathbf{P}[A \cup B] = \mathbf{P}[A] + \mathbf{P}[B] - \mathbf{P}[A \cap B]$ . Dans le cas  $n = 3$ ,  $\mathbf{P}[A \cup B \cup C] = \mathbf{P}[A] + \mathbf{P}[B] + \mathbf{P}[C] - \mathbf{P}[A \cap B] - \mathbf{P}[A \cap C] - \mathbf{P}[B \cap C] + \mathbf{P}[A \cap B \cap C]$ .

### Proposition 1.3

1. Pour toute suite croissante  $(A_n)_{n \geq 1}$  d'événements, c'est-à-dire telle que  $A_n \subset A_{n+1}$  pour tout  $n \geq 1$ , on a

$$\mathbf{P}\left[\bigcup_{n \geq 1} A_n\right] = \lim_{n \rightarrow +\infty} \mathbf{P}[A_n] = \sup_{n \geq 1} \mathbf{P}[A_n];$$

2. pour toute suite décroissante  $(A_n)_{n \geq 1}$  d'événements, c'est-à-dire telle que  $A_{n+1} \subset A_n$  pour tout  $n \geq 1$ , on a

$$\mathbf{P}\left[\bigcap_{n \geq 1} A_n\right] = \lim_{n \rightarrow +\infty} \mathbf{P}[A_n] = \inf_{n \geq 1} \mathbf{P}[A_n].$$

#### Démonstration :

1. Pour tout  $n$ , on écrit  $A_n$  comme l'union disjointe des anneaux  $B_k = A_k \setminus A_{k-1}$ . C'est-à-dire qu'avec  $B_1 = A_1$ ,

$$A_n = \bigcup_{k=1}^n B_k \text{ et } \bigcup_{n \geq 1} A_n = \bigcup_{n \geq 1} B_n \Rightarrow \mathbf{P} \left[ \bigcup_{n \geq 1} A_n \right] = \sum_{n \geq 1} \mathbf{P}[B_n].$$

Or, pour  $n \geq 2$ ,  $\mathbf{P}[B_n] = \mathbf{P}[A_n \setminus A_{n-1}] = \mathbf{P}[A_n] - \mathbf{P}[A_{n-1}]$ . On en déduit

$$\sum_{k=1}^n \mathbf{P}[B_k] = \mathbf{P}[A_n]$$

et on obtient le résultat en passant à la limite.

2. La suite  $(A_n)_{n \geq 1}$  est croissante. On peut donc utiliser le point précédent et on conclut en remarquant que

$$\begin{aligned} \mathbf{P} \left[ \bigcap_{n \geq 1} A_n \right] &= 1 - \mathbf{P} \left[ \bigcup_{n \geq 1} A_n^c \right] \\ &= 1 - \lim_{n \rightarrow +\infty} \mathbf{P}[A_n^c] = \lim_{n \rightarrow +\infty} \mathbf{P}[A_n]. \end{aligned}$$

#### **Proposition 1.4**

1. Pour toute suite  $(A_n)_{n \geq 1}$  d'événements,

$$\mathbf{P} \left[ \bigcup_{n \geq 1} A_n \right] \leq \sum_{n \geq 1} \mathbf{P}[A_n];$$

2. soit  $(A_n)_{n \geq 1}$  une partition dénombrable de  $\Omega$  alors, pour tout  $B \subset \Omega$ ,

$$\mathbf{P}[B] = \sum_{n \geq 1} \mathbf{P}[B \cap A_n].$$

#### Démonstration :

1. La suite  $\left( \bigcup_{k=1}^n A_k \right)_{n \geq 1}$  est croissante et d'après le point (1),

$$\mathbf{P} \left[ \bigcup_{k=1}^{+\infty} A_k \right] = \lim_{n \rightarrow +\infty} \mathbf{P} \left[ \bigcup_{k=1}^n A_k \right].$$

On pose,  $B_1 = A_1$  et, pour tout  $n \geq 2$ ,

$$B_n = A_n \setminus \bigcup_{k=1}^{n-1} A_k \Rightarrow \bigcup_{k=1}^n A_k = \bigcup_{k=1}^n B_k \text{ pour tout } n.$$

Comme les  $B_k$  sont disjoints, la probabilité de l'union est la somme des probabilités et, pour tout  $k$ ,  $\mathbf{P}[B_k] \leq \mathbf{P}[A_k]$ , ce qui permet de conclure.

2. Comme  $(A_n)_{n \geq 1}$  forme une partition de  $\Omega$ , on remarque que

$$B = \bigcup_{n \geq 1} [B \cap A_n]$$

et que cette union est disjointe. D'où le résultat.

### 1.3 Probabilités conditionnées par un événement.

Comment calculer la probabilité de la réalisation d'un événement  $A$  sachant qu'un événement  $B$  s'est réalisé ?

**Exemple :** On lance un dé. Quelle est la probabilité que le résultat soit inférieur ou égal à 3 sachant qu'il est pair ?

On tient le raisonnement suivant : les résultats pairs possibles sont 2, 4 et 6. Il y a donc 1 chance sur 3 que le dé tombe sur un résultat inférieur ou égal à 3 sachant qu'il est pair. Ce que l'on a fait, c'est que l'on a ramené le calcul de  $A = \ll \text{le résultat est inférieur ou égal à 3} \gg$  à l'événement  $B = \ll \text{le résultat est pair} \gg$  soit

$$\mathbf{P}[A \text{ sachant } B] = \frac{\mathbf{P}[A \cap B]}{\mathbf{P}[B]} = \frac{1/6}{1/2} = \frac{1}{3} \neq \mathbf{P}[A] = \frac{1}{2}$$

#### Définition 1.2

Soit  $B$  un événement de probabilité non nulle. Pour tout  $A \subset \Omega$ , on appelle probabilité de  $A$  sachant  $B$  la quantité définie par :

$$\mathbf{P}[A|B] = \frac{\mathbf{P}[A \cap B]}{\mathbf{P}[B]}.$$

#### Proposition 1.5 (Première formule de Bayes.)

Soient  $A$  et  $B$  deux événements de probabilité non nulle. Alors

$$\mathbf{P}[A|B] = \frac{\mathbf{P}[B|A]\mathbf{P}[A]}{\mathbf{P}[B]}.$$

Démonstration :

$$\mathbf{P}[A|B] = \frac{\mathbf{P}[A \cap B]}{\mathbf{P}[B]} = \frac{\mathbf{P}[B|A]\mathbf{P}[A]}{\mathbf{P}[B]}.$$

#### Proposition 1.6

Soit  $(A_n)_{n \geq 1}$  une partition de  $\Omega$  t.q.  $\mathbf{P}[A_n] \neq 0$  pour tout  $n \geq 1$ .

(i) pour tout  $B \subset \Omega$ , la formule des probabilités totales s'écrit :

$$\mathbf{P}[B] = \sum_{n \geq 1} \mathbf{P}[B|A_n]\mathbf{P}[A_n];$$

(ii) deuxième formule de Bayes : si  $B$  est un événement de probabilité non nulle, pour tout  $n \geq 1$ , on a :

$$\mathbf{P}[A_n|B] = \frac{\mathbf{P}[B|A_n]\mathbf{P}[A_n]}{\sum_{k \geq 1} \mathbf{P}[B|A_k]\mathbf{P}[A_k]}.$$

Démonstration :

(i) Puisque  $(A_n)_{n \geq 1}$  est une partition,

$$\Omega = \bigcup_{n \geq 1} A_n \Rightarrow B = B \cap \bigcup_{n \geq 1} A_n = \bigcup_{n \geq 1} (B \cap A_n)$$

et les  $A_n$  sont disjoints, donc

$$\mathbf{P}[B] = \sum_{n \geq 1} \mathbf{P}[B \cap A_n] = \sum_{n \geq 1} \mathbf{P}[B|A_n]\mathbf{P}[A_n].$$

(ii) On utilise la première formule de Bayes puis le (i).



## 1.4 Indépendance deux à deux ou mutuelle d'événements.

Comment exprimer l'idée que la probabilité de la réalisation d'un événement  $A$  ne dépend pas de la réalisation d'un événement  $B$  ?

**Exemple :** On lance deux fois une pièce de monnaie. On veut exprimer le fait que le second résultat obtenu ne dépend pas du premier.

Cela signifie que la probabilité que la seconde pièce tombe sur pile ou face est la même, que la première soit tombée sur pile ou sur face ; et c'est la probabilité des résultats du second lancer sans condition sur le premier. Autrement dit

$$\mathbf{P} [2^{\text{nd}} \text{ rés.} = \alpha | 1^{\text{er}} \text{ rés.} = \beta] = \mathbf{P} [2^{\text{nd}} \text{ rés.} = \alpha].$$

$$\Leftrightarrow \mathbf{P} [2^{\text{nd}} \text{ rés.} = \alpha \text{ et } 1^{\text{er}} \text{ rés.} = \beta] = \mathbf{P} [2^{\text{nd}} \text{ rés.} = \alpha] \mathbf{P} [1^{\text{er}} \text{ rés.} = \beta].$$

On en déduit la définition suivante.

### Définition 1.3

On dit que les événements  $A$  et  $B$  sont *indépendants* si  $\mathbf{P} [A \cap B] = \mathbf{P} [A] \mathbf{P} [B]$ .

### Proposition 1.7

Soient  $A$  et  $B$  deux événements. Si  $B$  est de probabilité non nulle,  $A$  et  $B$  sont indépendants si et seulement si

$$\mathbf{P} [A|B] = \mathbf{P} [A].$$

Démonstration : On suppose  $\mathbf{P} [B] \neq 0$ .  $A$  et  $B$  sont indépendants si et seulement si

$$\mathbf{P} [A \cap B] = \mathbf{P} [A] \mathbf{P} [B] \Leftrightarrow \frac{\mathbf{P} [A \cap B]}{\mathbf{P} [B]} = \mathbf{P} [A].$$

### Définition 1.4

Soit  $(A_n)_{n \geq 1}$  une suite d'événements. On dit que ces événements sont :

- (i) *indépendants deux à deux* si, pour tous  $n, m \geq 1, n \neq m$ ,  $A_n$  et  $A_m$  sont indépendants ;
- (ii) *mutuellement indépendants* (ou, plus simplement, *indépendants*) si, pour tout  $I \subset \mathbb{N}^*$  fini,

$$\mathbf{P} \left[ \bigcap_{i \in I} A_i \right] = \prod_{i \in I} \mathbf{P} [A_i].$$

Plus généralement, on définit les notions suivantes :

### Proposition 1.8

Soit  $(A_n)_{n \geq 1}$  une suite d'événements (mutuellement) indépendants. Alors :

1. les événements sont deux à deux indépendants ;
2. si l'on définit la suite  $(B_n)_{n \geq 1}$  par

$$B_n = A_n \text{ ou } B_n = A_n^c \text{ pour tout } n \geq 1,$$

alors les  $(B_n)_{n \geq 1}$  sont indépendants.

Démonstration :

1. est évident.

2. On se contente de la preuve pour deux événements : on suppose que  $A$  et  $B$  sont indépendants et on montre que  $A$  et  $B^c$  le sont ainsi bien sûr que  $A^c$  et  $B$  d'une part et  $A^c$  et  $B^c$  d'autre part. En effet, si  $A$  et  $B$  sont indépendants, on a :

$$\mathbf{P}[A] = \mathbf{P}[A \cap B] + \mathbf{P}[A \cap B^c]$$

puisque  $B$  et  $B^c$  forment une partition de  $\Omega$ . On en déduit

$$\begin{aligned}\mathbf{P}[A \cap B^c] &= \mathbf{P}[A] - \mathbf{P}[A] \mathbf{P}[B] \\ &= \mathbf{P}[A](1 - \mathbf{P}[B]) = \mathbf{P}[A] \mathbf{P}[B^c].\end{aligned}$$

Les autres cas se déduisent immédiatement du premier en intervertissant les rôles de  $A$  et  $B$  puis de  $A$  et  $A^c$ .

### Application importante :

Lorsque l'on dit que l'on réalise successivement  $n$  expériences identiques, on suppose en général que les résultats sont mutuellement indépendants. En particulier, si l'expérience de base a  $k$  issues équiprobables, la probabilité d'avoir un résultat donné à une série de  $n$  expériences réalisées indépendamment est  $1/k^n$  (jeu de pile ou face, jets de dés etc...).

**Remarque :** Des événements 2 à 2 indépendants ne sont pas en général mutuellement indépendants, comme l'illustre l'exemple suivant.

**Exemple :** Dans une famille de deux enfants, les événements  $A$  = « le premier est une fille »,  $B$  = « le second est une fille »,  $C$  = « les deux sont de même sexe » sont deux à deux indépendants mais pas mutuellement indépendants.

En effet, si  $D$  = « les deux sont des filles », alors  $A \cap B = A \cap C = B \cap C = D$ . Par ailleurs,  $\mathbf{P}[A] = \mathbf{P}[B] = \mathbf{P}[C] = 1/2$  et  $\mathbf{P}[D] = 1/4 = 1/2 \cdot 1/2$ , donc  $A$ ,  $B$  et  $C$  sont 2 à 2 indépendants. Par contre,  $A \cap B \cap C = D$  et donc  $\mathbf{P}[A \cap B \cap C] \neq \mathbf{P}[A] \mathbf{P}[B] \mathbf{P}[C]$ , et les événements ne sont pas mutuellement indépendants. En fait  $A \cap B \subset C$ , i.e.  $A$  et  $B$  implique  $C$ .

## 2 Variables aléatoires

### 2.1 Définition.

Pour modéliser une expérience aléatoire, on a souvent recours à une *variable aléatoire* (v.a.). C'est une application définie sur  $\Omega$  qui à chaque état du monde associe le résultat d'une expérience donnée dans cet état. Lorsque la valeur prise par la variable aléatoire est réelle, on parle de variable aléatoire réelle (v.a.r.). En général, on ne sait pas décrire explicitement une variable aléatoire associée à une expérience aléatoire mais on connaît sa *loi*, c'est-à-dire la probabilité qu'elle prenne telle ou telles valeurs.

**Exemple :** [Modélisation du lancer d'un dé à six faces non pipé.] On considère que  $\Omega$  contient tous les états du monde possibles. Si dans l'état du monde  $\omega \in \Omega$  le jet du dé donne la face  $i$  alors on note  $X(\omega) = i$ . La variable aléatoire  $X$  prend ses valeurs dans  $\{1, \dots, 6\}$ . La loi de  $X$  est la donnée des probabilités que  $X$  prenne comme valeur chacun des entiers 1,...,6 (ici, 1/6 dans chaque cas).

**Exemple :** [Modélisation du tirage au hasard d'un point entre 0 et 1.] La variable aléatoire modélisant l'expérience prend ses valeurs dans l'intervalle  $[0, 1]$ . Pour modéliser l'uniformité du tirage, on comprend que la probabilité qu'une valeur particulière soit prise doit être nulle (comme dans le cas infini dénombrable). En revanche, la probabilité de tomber dans un intervalle particulier doit être proportionnelle à la longueur de l'intervalle et comme la longueur de  $[0, 1]$  est 1, cette probabilité est égale à la longueur de l'intervalle.

#### Définition 2.1

Une variable aléatoire réelle est une application

$$\begin{aligned}X : \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto X(\omega)\end{aligned}$$

A partir de cette variable aléatoire, on peut définir les événements du type  $[X \in B]$ , pour  $B \subset \mathbb{R}$ , c'est-à-dire l'ensemble des états du monde  $\omega$  dans lesquels le résultat de l'expérience aléatoire est dans l'ensemble  $B$ . Du point de vue ensembliste,

$$[X \in B] = \{\omega \in \Omega / X(\omega) \in B\} = X^{-1}(B).$$

Plus concrètement, pour le jet d'un dé à six faces, par exemple, «  $X$  paire » s'écrit aussi «  $X \in \{2, 4, 6\}$  » ; pour l'exemple du choix d'un nombre dans  $[0, 1]$ , « obtenir un nombre supérieur à  $1/2$  » s'écrit «  $X \in [1/2, 1]$  ».

Connaître la « loi » d'une variable aléatoire  $X$  c'est connaître les probabilités de tous ces événements  $[X \in B]$ . Dans la suite, on va s'intéresser à deux cas particuliers de lois, les lois discrètes et les lois à densité, que l'on décrira complètement.

### Notation

On note parfois  $\mathcal{L}(X)$  la loi de  $X$ . Si  $X$  a pour loi  $\mu$ , on dit que  $X$  suit la loi  $\mu$  et on note  $X \rightsquigarrow \mu$  ou  $\mathcal{L}(X) = \mu$ .

## 2.2 Lois discrètes.

Ces lois servent à modéliser les expériences aléatoires ayant un ensemble (au plus) dénombrable d'issues possibles. Les variables aléatoires dont les lois sont discrètes s'appellent des v.a. *discrètes* (elles prennent un ensemble dénombrable de valeurs).

### 2.2.1 Définition, première propriété et exemple.

#### Définition 2.2

Une variable aléatoire réelle  $X$  est dite *discrète* s'il existe une suite  $\{x_1, \dots, x_k, \dots\}$  finie ou non de réels et une suite  $\{p_1, \dots, p_k, \dots\}$  de  $[0, 1]$  telles que,

$$\text{pour tout } k \geq 1, \mathbf{P}[X = x_k] = p_k \text{ et } \sum_{k \geq 1} p_k = 1$$

La donnée de ces deux suites détermine la loi de la variable aléatoire discrète  $X$ .

Une expérience à issues dénombrables modélisée par une partition d'événements  $(A_k)_{k \in \mathbb{N}}$  peut être également modélisée par une variable aléatoire à valeurs entières satisfaisant

$$\mathbf{P}[X = k] = \mathbf{P}[A_k], \text{ pour tout } k \in \mathbb{N}.$$

Réciproquement, si  $X$  est une variable aléatoire à valeurs dans  $\{x_k\}_{k \geq 1}$ , alors les événements  $([X = x_k])_{k \geq 1}$  forment une partition de  $\Omega$ . Les propriétés des probabilités entraînent alors que la suite  $(p_k)_{k \geq 1}$  de la définition ci-dessus doit satisfaire

$$p_k \geq 0 \text{ pour tout } k \text{ et } \sum_{k \geq 1} p_k = 1.$$

#### Proposition 2.1

Soit  $X$  une v.a. discrète prenant ses valeurs dans  $\{x_k, k \geq 1\}$  avec les probabilités  $(p_k)_{k \geq 1}$ . Alors, pour tout  $B \subset \mathbb{R}$ ,

$$\mathbf{P}[X \in B] = \sum_{x_k \in B} p_k.$$

**Exemple :** On lance deux dés équilibrés simultanément. Quelle est la loi de la somme des points indiqués sur les faces visibles des dés ?

On note  $S$  la somme des faces visibles des deux dés.  $S$  prend ses valeurs dans  $\{2, \dots, 12\}$ . Sans complément d'information, on peut supposer que les résultats des dés sont indépendants et donc que la probabilité d'obtenir un couple quelconque est  $1/36$ . On en déduit la loi de  $S$  :

$k$	2	3	...	7	...	11	12
$\mathbf{P}[S = k]$	1/36	2/36	...	6/36	...	2/36	1/36.

### 2.2.2 Une seule issue : la mesure de Dirac.

On suppose que l'expérience aléatoire considérée n'a qu'une issue possible,  $x_0 \in \mathbb{R}$ . Cette expérience peut être modélisée par la variable aléatoire constante  $X = x_0$ . Autrement dit,  $X$  sera l'application constante :  $X(\omega) = x_0$  pour tout  $\omega \in \Omega$ . La loi de  $X$  est alors, par définition, donnée par :

$$\mathbf{P}[X = x_0] = 1.$$

### 2.2.3 Deux issues : la loi de Bernoulli.

On suppose que l'expérience aléatoire considérée n'a que deux issues possibles  $x_1$  et  $x_2 \in \mathbb{R}$ . Cette expérience peut être modélisée par une variable aléatoire  $X$  définie sur un univers  $\Omega$  et prenant uniquement les valeurs  $x_1$  ou  $x_2$ . La loi de  $X$  est alors entièrement déterminée par les données de  $p = \mathbf{P}[X = x_1]$  (puisque alors  $\mathbf{P}[X = x_2] = 1 - p$ ).

La loi d'une telle variable s'appelle une *loi de Bernoulli de paramètre  $p$*  et se note  $\mathcal{B}(p)$ . Remarquer que l'on peut modéliser avec une telle loi toutes les expériences à deux issues, même lorsqu'elles ne sont pas numériques. Par exemple, on modélise les expériences consistant en un succès et un échec par une v.a. prenant les valeurs 1 (succès) et 0 (échec),  $p$  étant alors par convention la probabilité du succès.

**Exemple :** Jeu de Pile (1) ou Face (0) ; sexe d'un enfant à naître, tirage d'une boule blanche ou noire.

### 2.2.4 Issues équiprobables : la loi uniforme.

Cette loi est celle des v.a. servant à modéliser les expériences aléatoires ayant un nombre fini d'issues possibles ayant la même probabilité d'apparition. Une v.a. suivant la loi uniforme prend ses valeurs dans un ensemble  $\{x_1, \dots, x_n\}$ ,  $n \geq 1$ , et, pour tout  $k \in \{1, \dots, n\}$ ,

$$\mathbf{P}[X = x_k] = \frac{1}{n}.$$

**Exemple :** Tirage d'une boule « au hasard » parmi  $n$ . De façon arbitraire, on numérote les boules ( $x_k = k$  pour tout  $k$ ). Jet d'un dé.

### 2.2.5 Tirages sans remise : la loi hypergéométrique.

**Exemple :** Dans une urne, on dispose  $N_1 + N_2$  boules numérotées dont  $N_1$  blanches et  $N_2$  noires. On tire au hasard  $n$  ( $n \leq \min(N_1, N_2)$ ) boules et on s'intéresse au nombre de boules blanches obtenues lors de ce tirage lorsqu'on le fait sans remise.

Le résultat obtenu, que l'on note  $X$  est un entier entre 0 et  $n$  car  $n \leq \min(N_1, N_2)$  (on peut donc n'avoir que des boules blanches ou que des boules noires).  $X$  est une variable aléatoire définie sur un univers  $\Omega$ . Le résultat d'un tirage sans remise est un sous-ensemble de l'ensemble des boules, de cardinal  $n$ , que les tirages soient simultanés ou successifs. Chaque sous-ensemble a la même probabilité d'être tiré.

On est donc dans une configuration équiprobable. La probabilité que  $X$  prenne la valeur  $k \in \{0, \dots, n\}$  est donc le nombre de tirages donnant exactement  $k$  boules blanches sur le nombre total de tirages. Le nombre de tirages total possible est le nombre d'ensembles à  $n$  éléments pris parmi les  $N_1 + N_2$  boules, soit  $\binom{N_1 + N_2}{n}$ . Le nombre de tirages donnant  $k$  boules blanches est le nombre de sous-ensembles à  $n$  éléments dont  $k$  sont tirés parmi les  $N_1$  blanches et  $n - k$  parmi les  $N_2$  noires, soit  $\binom{N_1}{k} \binom{N_2}{n - k}$ . Finalement, pour tout  $k \in \{0, \dots, n\}$ ,

$$\mathbf{P}[X = k] = \frac{\binom{N_1}{k} \binom{N_2}{n - k}}{\binom{N_1 + N_2}{n}}.$$

**Remarque :** Comme les ensembles  $[X = k]$ , pour  $k \in \{0, \dots, n\}$ , forment une partition de  $\Omega$ , on en déduit que

$$\sum_{k=0}^n \frac{\binom{N_1}{k} \binom{N_2}{n-k}}{\binom{N_1+N_2}{n}} = 1$$

soit

$$\sum_{k=0}^n \binom{N_1}{k} \binom{N_2}{n-k} = \binom{N_1+N_2}{n}.$$

La loi de la v.a.  $X$  décrite ci-dessus s'appelle la loi hypergéométrique. On la note

$$\mathcal{H}(n, N_1, N_2) \text{ ou } \mathcal{H}(n, N_1, N_1 + N_2) \text{ ou encore } \mathcal{H}(n, p, N)$$

où  $N$  est la taille de la population totale et  $p$  la proportion des individus de la population possédant le caractère que l'on veut compter, i.e.

$$N = N_1 + N_2 \text{ et } p = \frac{N_1}{N_1 + N_2}.$$

### 2.2.6 Tirages avec remise : la loi binomiale.

**Exemple :** On reprend l'exemple précédent mais on suppose cette fois que le tirage s'effectue avec remise, c'est-à-dire que chaque boule tirée est remise dans l'urne avant le tirage suivant. Remarque que cela revient à tirer chacune des boules dans  $n$  urnes identiques mais distinctes (comme lorsqu'on lance plusieurs dés). On ne suppose plus que  $n \leq \min(N_1, N_2)$ .

Le résultat d'un tirage est une suite ordonnée de boules, non nécessairement distinctes. Ces tirages sont équiprobables. A chaque tirage, tous les choix sont envisageables, donc le nombre de tirages possibles est  $(N_1 + N_2)^n$ .

La variable  $X$  prend les mêmes valeurs que précédemment. Soit  $k \in \{0, \dots, n\}$ . On cherche le nombre de configurations avec  $k$  boules blanches. Il faut choisir la place de ces  $k$  boules blanches :  $C_n^k$ . Ensuite, on doit choisir les boules blanches : on a  $N_1^k$  choix puisqu'il s'agit, pour chacune des  $k$  places, de choisir une boule parmi  $N_1$ . De même, on a  $N_2^{n-k}$  choix pour les boules noires. Finalement,

$$\mathbf{P}[X = k] = \binom{n}{k} \frac{N_1^k N_2^{n-k}}{(N_1 + N_2)^n}.$$

On peut récrire cette probabilité

$$\mathbf{P}[X = k] = \binom{n}{k} \left( \frac{N_1}{N_1 + N_2} \right)^k \left( \frac{N_2}{N_1 + N_2} \right)^{n-k}.$$

On dit alors que  $X$  suit une loi binomiale de paramètres  $n$  et  $p = N_1/(N_1 + N_2)$ . Dans ce cas,  $n$  est le nombre de tirages (avec remise) et  $p$  est la proportion de caractères qui nous intéresse (ici « blanc »). On note  $X \sim \mathcal{B}(n, p)$  et, pour tout  $k \in \{0, \dots, n\}$ ,

$$\mathbf{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}.$$

**Application importante : nombre de succès dans une série d'expériences aléatoires identiques et indépendantes.**

On lance  $n$  fois une pièce de monnaie ayant pour probabilité  $p$  de tomber sur Pile et on compte le nombre  $X$  de fois où la pièce tombe sur Pile.

Alors  $X$  suit une loi binomiale de paramètres  $n$  et  $p$ . En effet, dans cet exemple classique, les lancers sont toujours supposés indépendants. Jouer à pile ou face revient donc à tirer au hasard et avec remise une boule dans une urne comprenant une proportion  $p$  de boules blanches et  $1-p$  de boules noires, comme on va le voir maintenant : on note  $X_1, \dots, X_n$  les résultats des  $n$  lancers. Alors le nombre de Pile est  $X = X_1 + \dots + X_n$  et, pour tout  $k \in \{0, \dots, n\}$ ,

$[X = k]$  est l'union des événements disjoints de type  $A_1 \cap A_2 \dots \cap A_n$  où  $k$  des  $A_i$  sont des  $[X_i = 1]$  et les  $n - k$  restants sont des  $[X_i = 0]$ . La probabilité d'un tel événement, comme intersection d'événements indépendants, est  $p^k(1 - p)^{n-k}$ . Reste à compter ces événements : c'est le nombre de façons de choisir  $k$  "1" dans une suite ordonnée de  $n$  "0" ou "1", soit  $C_n^k$ . Finalement  $\mathbf{P}[X = k] = \binom{n}{k} p^k(1 - p)^{n-k}$ .

### 2.2.7 Rang du premier succès : la loi géométrique.

**Exemple :** On lance une pièce de monnaie qui a la probabilité  $p$  de tomber sur Pile et on note  $X$  le premier rang d'apparition d'un Pile.

La variable  $X$  prend ses valeurs dans  $\mathbb{N}^*$ . On a

$$\mathbf{P}[X = 1] = p$$

et, pour tout  $k \geq 2$ ,

$$\mathbf{P}[X = k] = \mathbf{P}[1^{\text{er}} \text{ l.} = \text{Face}, \dots, (k - 1)\text{-ième l.} = \text{Face}, k\text{-ième l.} = \text{Pile}]$$

et du fait de l'indépendance mutuelle des lancers,

$$\mathbf{P}[X = k] = (1 - p)^{k-1}p,$$

formule qui est aussi valable pour  $k = 1$ .

La loi du premier succès (ici Pile) dans une succession d'expériences identiques et indépendantes s'appelle la loi Géométrique de paramètre  $p$ , noté  $\mathcal{G}(p)$ . C'est une loi dite « sans mémoire ». Plus précisément, on a le résultat suivant :

#### Proposition 2.2

Une v.a.  $X$  à valeurs dans  $\mathbb{N}^*$  suit la loi géométrique si et seulement si, pour tout  $k \geq 1$ ,

$$\mathbf{P}[X = k + 1 | X > k] = \mathbf{P}[X = 1].$$

Démonstration : On remarque tout d'abord que

$$\mathbf{P}[X = k + 1 | X > k] = \frac{\mathbf{P}[X = k + 1]}{1 - \mathbf{P}[X \leq k]}.$$

L'égalité de la proposition s'écrit donc

$$\mathbf{P}[X = k + 1] = \mathbf{P}[X = 1] \left( 1 - \sum_{j=1}^k \mathbf{P}[X = j] \right).$$

On suppose que  $X$  suit une loi géométrique de paramètre  $p = \mathbf{P}[X = 1]$ . Alors

$$\sum_{j=1}^k \mathbf{P}[X = j] = p \sum_{j=0}^{k-1} (1 - p)^j = p \frac{1 - (1 - p)^k}{1 - (1 - p)} = 1 - (1 - p)^k,$$

d'où le résultat.

Réciproquement, on a, pour  $k = 1$

$$\mathbf{P}[X = 2] = \mathbf{P}[X = 1] (1 - \mathbf{P}[X = 1])$$

et pour  $k \geq 2$

$$\mathbf{P}[X = k + 1] = \mathbf{P}[X = 1] \left( 1 - \sum_{j=1}^{k-1} \mathbf{P}[X = j] - \mathbf{P}[X = k] \right).$$

Avec l'hypothèse, on en déduit

$$\begin{aligned} \mathbf{P}[X = k + 1] &= \mathbf{P}[X = k] - \mathbf{P}[X = 1]\mathbf{P}[X = k] \\ &= (1 - \mathbf{P}[X = 1])\mathbf{P}[X = k]. \end{aligned}$$

On en déduit, pour  $k \geq 1$ ,

$$\mathbf{P}[X = k + 1] = (1 - \mathbf{P}[X = 1])^k \mathbf{P}[X = 1]$$

qui est le résultat souhaité.

### 2.2.8 Loi de Poisson

Cette loi, à valeur dans  $\mathbb{N}$  sert à modéliser les arrivées indépendantes de clients à un guichet, de véhicules à un péage, etc. Elle dépend d'un paramètre  $\lambda$  qui est le nombre moyen d'arrivées par unité de temps.

Plus précisément, on dit que  $X$  suit une loi de Poisson de paramètre  $\lambda > 0$  et on note  $X \rightsquigarrow \mathcal{P}(\lambda)$  si pour tout  $k \in \mathbb{N}$ ,

$$\mathbf{P}[X = k] = \frac{e^{-\lambda} \lambda^k}{k!}.$$

## 2.3 Lois à densité.

Les v.a. à densité modélisent des expériences dont le résultat est quantitatif et dont les valeurs prises sont réparties « suffisamment régulièrement » sur  $\mathbb{R}$ . Par exemple : taille d'un composant produit par une machine, durée de vie d'une particule, tirage d'un point au hasard dans  $[0, 1]$ .

### 2.3.1 Définition

#### Définition 2.3

Soit  $f : \mathbb{R} \rightarrow \mathbb{R}$  une fonction positive d'intégrale 1. On dit qu'une variable aléatoire  $X$  a pour densité  $f$  si sa loi est définie pour tout  $B \subset \mathbb{R}$  par

$$\mathbf{P}[X \in B] = \int_B f(x) dx.$$

**Remarque :** Si  $B$  est un ensemble négligeable et si  $X$  est une v.a. à densité,

$$\mathbf{P}[X \in B] = \int_B f(x) dx = 0.$$

Comme corollaire, on montre que les v.a. discrètes n'admettent pas de densité puisqu'elles sont « portées » par des ensembles dénombrables donc négligeables, i.e. si  $X$  est une v.a. à valeurs dans  $D$  dénombrable alors  $\mathbf{P}[X \in D] = 1$ .

Les lois à densité sont moins faciles à appréhender que les lois discrètes, sauf la loi uniforme. Elles apparaissent naturellement par passage à la limite dans des lois discrètes, par changement de variable à partir d'une loi uniforme, ou « par expérience » dans les séries statistiques. Dans la suite, nous proposons quelques exemples usuels. L'ensemble des lois classiques à connaître se trouve en annexe.

### 2.3.2 Loi uniforme.

Une variable aléatoire  $X$  est dite uniforme si la probabilité qu'elle prenne ses valeurs dans un ensemble donné est proportionnelle à la « longueur » de cet ensemble, c'est donc l'intégrale d'une constante sur l'ensemble.

Pour un intervalle  $[a, b]$  la longueur  $b - a$  est aussi l'intégrale sur l'intervalle. Cela implique que la densité d'une variable aléatoire uniforme doit être constante.

Plus précisément : on dit que  $X$  suit la loi uniforme sur l'intervalle borné  $I$  non réduit à un point si elle admet pour densité la fonction

$$f(x) = \begin{cases} 1/\lambda(I) & \text{si } x \in I \\ 0 & \text{sinon} \end{cases}$$

où  $\lambda(I)$  est la longueur de  $I$ .

Cette définition peut-être étendue à des ensembles  $B$  autres que des intervalles en prenant comme définition de la longueur la quantité

$$\lambda(B) = \int_B dx$$

à condition que cette quantité soit bien définie, strictement positive et finie.

Si  $X$  suit une loi uniforme sur  $B$ , on note  $X \rightsquigarrow \mathcal{U}(B)$ .

### 2.3.3 Loi exponentielle.

La loi exponentielle est le pendant continu de la loi géométrique : c'est la seule loi à densité continue sans mémoire, c'est-à-dire satisfaisant

$$\mathbf{P}[X > s + t | X > t] = \mathbf{P}[X > s] \text{ pour tous } t, s > 0. \quad (1)$$

On peut montrer que la densité d'une loi sans mémoire à densité continue est nécessairement définie par

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases}$$

où  $\lambda > 0$  est un paramètre. On dit qu'une v.a. admettant cette fonction pour densité est une loi exponentielle de paramètre  $\lambda$  que l'on note  $\mathcal{E}(\lambda)$ . Une telle variable aléatoire satisfait la propriété (1).

### 2.3.4 Loi normale ou gaussienne.

Cette loi apparaît comme loi limite lorsque l'on répète une infinité de fois des expériences identiques et indépendantes. Elle est également utilisée très fréquemment en statistique. Sa densité est strictement positive sur  $\mathbb{R}$ , symétrique, et présente une forme caractéristique dite « en cloche ».

On dit que  $X$  suit une loi normale de paramètres  $m \in \mathbb{R}$  et  $\sigma^2 > 0$  et on note  $X \rightsquigarrow \mathcal{N}(m, \sigma^2)$  si  $X$  admet pour densité la fonction définie par

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/(2\sigma^2)} \text{ pour tout } x \in \mathbb{R}.$$

## 2.4 Fonction de répartition d'une variable aléatoire

La fonction de répartition d'une variable aléatoire est une caractérisation de tous les types de variables aléatoires réelles.

### Définition 2.4

Soit  $X$  une v.a. réelle sur  $\Omega$ . On appelle fonction de répartition de  $X$  et on note  $F_X$  la fonction, définie pour tout



$t \in \mathbb{R}$  par

$$F_X(t) = \mathbf{P}[X \leq t] = \mathbf{P}[\{\omega \in \Omega / X(\omega) \leq t\}].$$

### Proposition 2.3

Soit  $F_X$  la fonction de répartition d'une v.a.r.  $X$  définie sur  $\Omega$ . Alors, pour tous  $a, b \in \mathbb{R}$  tels que  $a < b$ , on a :

$$\mathbf{P}[a < X \leq b] = F_X(b) - F_X(a).$$

Démonstration :

### Proposition 2.4

Soit  $F_X$  la fonction de répartition d'une v.a.r.  $X$  définie sur  $\Omega$ . Alors  $F_X$  est croissante, continue à droite et satisfait

$$\lim_{t \rightarrow -\infty} F_X(t) = 0, \quad \lim_{t \rightarrow +\infty} F_X(t) = 1.$$

**Remarque :** La loi d'une v.a.r.  $X$  est entièrement déterminée par la fonction de répartition  $F_X$  de  $X$ . Plus précisément, pour tout  $t \in \mathbb{R}$ ,

$$F_X(t) = \mathbf{P}[X \in ]-\infty, t]]$$

et la donnée de ces probabilités suffit pour caractériser la loi de  $X$ .

**Exemple :** Soit  $X$  une v.a. réelle.

1. Si  $X$  est une v.a. discrète prenant ses valeurs dans  $\{x_1, \dots, x_k, \dots\}$ , les valeurs étant distinctes, alors

$$F_X(t) = \sum_{k | x_k \leq t} \mathbf{P}[X = x_k] \text{ pour tout } t \in \mathbb{R}.$$

En reprenant le cas de l'exemple de la somme  $S$  de deux dés :

2. Si  $X$  est une v.a.r. admettant une densité  $f$  alors

$$F_X(t) = \int_{-\infty}^t f(x) dx \text{ pour tout } t \in \mathbb{R}.$$

En effet,

Par exemple, si  $X$  suit une loi uniforme sur  $[0, a]$ , où  $a > 0$  :

En particulier, dans le cas à densité, on a le résultat suivant :

**Proposition 2.5**

Soit  $X$  une v.a.r. de densité  $f$  et de fonction de répartition  $F$ . Alors  $F$  est continue sur  $\mathbb{R}$  et si  $f$  est continue en  $x \in \mathbb{R}$ , alors  $F$  est dérivable en  $x$  et

$$F'(x) = f(x).$$

Réciproquement, si  $X$  admet une fonction  $F$  continue et dérivable par morceaux, alors  $X$  admet pour densité  $f = F'$ .

**Définition 2.5**

Soit  $X$  une v.a. réelle de fonction de répartition  $F_X$ . On appelle quantile d'ordre  $\alpha$  le réel  $q_\alpha$  défini par

$$q_\alpha = F_X^{-1}(\alpha)$$

### 3 Moments de variables aléatoires

#### 3.1 Définitions

Il existe des données plus synthétiques que la loi pour décrire une expérience aléatoire. La moyenne ou l'écart-type des résultats possibles par exemple. Voyons sur un exemple comment on calcule la moyenne en pratique.

**Exemple :** On lance une pièce de monnaie équilibrée. On considère la variable aléatoire qui vaut 1 lorsque la pièce tombe sur pile et -1 sinon. Quelle est la valeur moyenne attendue ?

$$0$$

On suppose maintenant que l'on gagne deux points si on obtient pile et que l'on perd 1 point si on obtient face. Quelle est la moyenne attendue de nos points ?

$$\frac{-1 + 2}{2} = \frac{1}{2}.$$

Enfin, on suppose à nouveau que l'on gagne ou perd un point mais que la pièce tombe deux fois plus souvent sur pile que sur face. Quelle est la valeur moyenne attendue ?

$$\frac{2x1 - 1}{3}$$

Ce que l'on fait, c'est donc une moyenne des points que l'on peut obtenir, pondérée par les probabilités d'obtenir ces points. Cela conduit à la définition suivante dans le cas des variables aléatoires discrètes :

### Définition 3.1

Soit  $X$  est une v.a. prenant ses valeurs dans  $\{x_k, k \in \mathbb{N}\}$ , alors si

$$\sum_{k \in \mathbb{N}} |x_k| \mathbf{P}[X = x_k] < +\infty$$

on dit que  $X$  est intégrable et on définit l'espérance de  $X$  par

$$\mathbf{E}[X] = \sum_{k \in \mathbb{N}} x_k \mathbf{P}[X = x_k].$$

Pour les variables à densité, l'idée est la même : la valeur moyenne de la variable aléatoire, doit être la somme des valeurs prises par la variable pondérées non plus par des probabilités discrètes mais par la densité de probabilité de la variable. Cela conduit à la définition suivante :

### Définition 3.2

Soit  $X$  est une variable aléatoire admettant pour densité  $f$ . Alors si

$$\int_{\mathbb{R}} |x| f(x) dx < +\infty$$

on dit que  $X$  est intégrable et on définit l'espérance de  $X$  par

$$\mathbf{E}[X] = \int_{\mathbb{R}} x f(x) dx.$$

L'espérance (ou la *moyenne*) d'une variable aléatoire est une caractéristique de tendance centrale de la variable. On l'appelle également le *moment d'ordre 1* de la variable aléatoire.

**Exemple :** [Calculs d'espérances.]

1. Soit  $X$  une v.a. suivant une loi binomiale de paramètres  $n \in \mathbb{N}^*$  et  $p \in ]0, 1[$ . On rappelle que  $X$  prend ses valeurs dans  $\{0, \dots, n\}$  et  $\mathbf{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$  pour tout  $k \in \{0, \dots, n\}$ .

Comme  $X$  prend un nombre fini de valeurs, elle est intégrable et

$$\mathbf{E}[X] = \sum_{k=1}^n k \binom{n}{k} p^k (1-p)^{n-k}.$$

$$\mathbf{E}[X] = \sum_{k=1}^n \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k}$$

$$\mathbf{E}[X] = \sum_{k=1}^n \frac{n(n-1)!}{(k-1)!(n-1-(k-1))!} p p^{k-1} (1-p)^{n-1-(k-1)}$$

$$= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!(n-1-k)!} p^k (1-p)^{n-1-k}$$

$$= np(p + 1 - p)^{n-1} = np.$$

Donc  $\mathbf{E}[X] = np$ .

2. Soit  $X$  une v.a. suivant la loi exponentielle de paramètre  $\lambda > 0$ . On rappelle qu'une telle variable admet pour densité la fonction

$$x \mapsto f(x) = \lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x).$$

Comme  $X$  est une v.a. positive, elle est intégrable si son intégrale est finie. On a

Cette intégrale est finie et se calcule par intégration par parties. Donc  $X$  est intégrable et

$$\mathbf{E}[X] =$$

Finalement,  $X$  est intégrable et

$$\mathbf{E}[X] =$$

### 3.2 Propriétés de l'espérance

L'espérance est d'une variable aléatoire  $X$  est en réalité une intégrale en un sens que l'on ne veut pas préciser ici. A ce titre, elle vérifie un certain nombre de propriétés que nous énonçons sans démonstration.

#### Proposition 3.1

Soient  $X$  et  $Y$  des v.a.r. intégrables définie sur  $\Omega$ . Alors,

1. pour tous  $a, b \in \mathbb{R}$ ,  $aX + b$  est intégrable et  $\mathbf{E}[aX + b] = a\mathbf{E}[X] + b$  et  $\mathbf{E}[b] = b$  ;
2.  $X + Y$  est intégrable et  $\mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y]$  ;
3. si  $X$  est positive sur  $\Omega$  alors  $\mathbf{E}[X] \geq 0$ .

#### Proposition 3.2

Si  $X$  est une variable aléatoire bornée alors  $X$  est intégrable et

$$\mathbf{E}[|X|] \leq \sup_{\omega \in \Omega} |X(\omega)|.$$

#### Proposition 3.3 (Inégalité de Markov.)

Soit  $X$  une v.a.r. intégrable. On a,

$$\mathbf{P}[|X| > \varepsilon] \leq \frac{\mathbf{E}[|X|]}{\varepsilon} \text{ pour tout } \varepsilon > 0.$$

### 3.3 Changement de variable.

Il arrive qu'un modèle aléatoire utilise une variable aléatoire  $Y$  dont on ne connaît pas la loi mais qui s'exprime fonctionnellement à travers une autre variable  $X$  mieux connue. On verra dans cette section que l'on peut non seulement calculer l'espérance de  $Y$  en utilisant la loi de  $X$  mais aussi, dans certains cas, déterminer la loi de  $Y$ .

Les formules suivantes permettent de calculer l'espérance de fonctions de variables aléatoires.

**Proposition 3.4 (Formules de transfert.)**

Soit  $X$  une v.a. sur  $\Omega$  et à valeurs dans  $\mathbb{R}$ . Soit  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ .

1. Si  $X$  est une v.a. discrète prenant ses valeurs dans  $\{x_k, k \geq 1\}$  et si  $\varphi(X)$  est intégrable, c'est-à-dire si

$$\sum_{k \geq 1} |\varphi(x_k)| \mathbf{P}[X = x_k] < +\infty \text{ alors } \mathbf{E}[\varphi(X)] = \sum_{k \geq 1} \varphi(x_k) \mathbf{P}[X = x_k].$$

2. Si  $X$  est une v.a. admettant pour densité  $f$  et si  $\varphi(X)$  est intégrable, c'est-à-dire si

$$\int_{\mathbb{R}} |\varphi(x)| f(x) dx < +\infty \text{ alors } \mathbf{E}[\varphi(X)] = \int_{\mathbb{R}} \varphi(x) f(x) dx.$$

**3.3.1 Changement de variable en discret.**

**Exemple :** On considère un insecte se déplaçant sur une ligne matérialisée par des morceaux de sucres. Son point de départ, supposé connu, est indicé par 0 et les sucres sont numérotés suivant  $\mathbb{Z}$ . On suppose que l'insecte ne fait que des déplacements de + ou - 1. On suppose que sur 100 déplacements, le nombre de déplacements à droite suit une loi binomiale de paramètres  $p$ , 100. Quelle est la loi de la position de l'insecte après 100 déplacements ?

De façon générale, lorsque la nouvelle variable discrète s'exprime de façon bijective par rapport à l'ancienne, on a le résultat suivant.

**Proposition 3.5**

Soit  $X$  une variable aléatoire discrète prenant ses valeurs dans l'ensemble  $\{x_i, i \in \mathbb{N}\}$ . Soit  $\psi$  une application bijective de  $\mathbb{R}$  dans  $\mathbb{R}$  et soit  $Y$  la variable aléatoire définie par  $Y = \psi(X) = \psi \circ X$ . Alors  $Y$  prend ses valeurs dans  $\{\psi(x_i), i \in \mathbb{N}\}$  et pour tout  $i \in \mathbb{N}$ , si  $y_i = \psi(x_i)$

$$\mathbf{P}[Y = y_i] = \mathbf{P}[X = \psi^{-1}(y_i)].$$

On en déduit aisément le cas général (non nécessairement bijectif).

**Proposition 3.6**

Soit  $X$  une variable aléatoire discrète prenant ses valeurs dans l'ensemble  $\{x_i, i \in \mathbb{N}\}$ . Soit  $\psi$  une application de  $\mathbb{R}$  dans  $\mathbb{R}$  et soit  $Y$  la variable aléatoire définie par  $Y = \psi(X) = \psi \circ X$ . Alors  $Y$  prend ses valeurs dans  $\{\psi(x_i), i \in \mathbb{N}\}$  et pour tout  $y \in \{\psi(x_i), i \in \mathbb{N}\}$ ,

$$\mathbf{P}[Y = y] = \sum_{x \in \psi^{-1}(y)} \mathbf{P}[X = x].$$

### 3.3.2 Changement de variable par la fonction de répartition.

**Exemple :** Soit  $X$  une variable aléatoire de fonction de répartition  $F_X$ . Soit  $\psi$  une application bijective croissante de  $\mathbb{R}$  dans  $\mathbb{R}$  et soit  $Y = \psi(X)$ . Calculer la fonction de répartition  $F_Y$  de  $Y$ .

Ce résultat n'est pas aisément généralisable alors que les calculs peuvent manifestement être poussés plus loin. Cela est fait ci-dessous à travers les densités.

### 3.3.3 Changement de variable pour les v.a. à densité.

Il existe une sorte de réciproque à la formule de transfert qui sera utile pour déterminer des lois inconnues dans le cas des variables à densité.

#### Proposition 3.7

Soit  $X$  une v.a. définie sur  $\Omega$  et à valeurs dans  $\mathbb{R}$ . La loi de  $X$  est entièrement caractérisée par la donnée des

$$\mathbf{E} [\varphi(X)], \text{ pour tout } \varphi : \mathbb{R} \rightarrow \mathbb{R} \text{ bornée.}$$

Plus précisément, s'il existe une fonction  $f$  telle que pour toute fonction  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  bornée

$$\mathbf{E} [\varphi(X)] = \int_{\mathbb{R}} \varphi(x) f(x) dx,$$

alors  $f$  est une fonction positive d'intégrale 1 et  $X$  a pour densité  $f$ .

**Remarque :** Dans le résultat précédent, les applications  $\varphi$  sont bornées, ce qui garantit l'existence de l'espérance et rend inutile sa vérification.

**Application :** Utilisation de la formule de transfert pour un changement de variable. On connaît la loi d'une v.a. à densité et une seconde v.a. s'exprime, par l'intermédiaire d'une application, en fonction de la première. Comment calculer la loi de la seconde v.a. ? On donne ci-dessous un exemple pour présenter la méthode préconisée.

**Exemple :** On suppose que  $\Theta$  suit une loi uniforme sur  $] -\pi/2, \pi/2[$ . On pose  $X = \tan \Theta$ . Quelle est la loi de  $X$  ?

Soit  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  une application bornée. Alors  $\varphi(X)$  est intégrable et

$$\mathbf{E} [\varphi(X)] =$$

On applique la formule de transfert à  $\varphi \circ \tan$ . Il vient :

Puisque cette égalité est satisfaite pour toute application bornée, on en déduit que la v.a.  $X$  admet pour densité

Cette méthode nous permet de montrer le résultat suivant, qui reste assez limité.

**Proposition 3.8**

Soit  $X$  une variable aléatoire de densité  $f$ . On pose  $Y = \psi(X)$  où  $\psi$  est un difféomorphisme de  $\mathbb{R}$  sur  $\mathbb{R}$  (respectivement d'un intervalle ouvert  $I$  tel que  $X(\Omega) \subset I$  sur  $\psi(I)$ ). Alors  $Y$  admet pour densité la fonction  $g$  définie, pour tout  $y \in \mathbb{R}$ , par

$$g(y) = \frac{f(\psi^{-1}(y))}{|\psi'(\psi^{-1}(y))|}.$$

$$\left( \text{repectivement } g(y) = \frac{f(\psi^{-1}(y))}{|\psi'(\psi^{-1}(y))|} \mathbf{1}_{\psi(I)}(y). \right)$$

**Proposition 3.9**

Application Soit  $X$  une v.a. suivant une loi normale de paramètre  $m$  et  $\sigma^2$ . Soient  $a$  et  $b$  des réels. Alors  $Y = aX + b$  suit une loi normale de paramètres  $am + b$  et  $a^2\sigma^2$ .

Démonstration : Si  $a = 0$ ,  $Y = b$  est bien une v.a. normale de moyenne  $b$  et de variance 0. Sinon,  $Y$  est obtenue par changement de variable bijectif sur  $\mathbb{R}$  et en appliquant la formule ci-dessus, sa densité  $g$  est définie, pour tout  $y \in \mathbb{R}$ , par

$$g(y) =$$

Avec

$$f(x) = \quad \text{pour tout } x \in \mathbb{R},$$

on obtient bien

$$g(y) = \quad \text{pour tout } y \in \mathbb{R}.$$

### 3.4 Variance et écart-type d'une variable aléatoire réelle.

**Définition 3.3**

Soit  $X$  une v.a.r. définie sur  $\Omega$  telle que  $X^2$  soit intégrable. On appelle alors la variance de  $X$  le réel positif

$$\text{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2].$$

La variance est une caractéristique de dispersion (on utilise aussi la racine de la variance, appelée l'écart-type de la variable). On appelle également la variance le *moment centré d'ordre 2* de la variable aléatoire réelle.

**Remarque :** Si  $X$  une v.a.r. définie sur  $\Omega$  telle que  $X^2$  soit intégrable, alors  $X$  est intégrable.

Grâce aux formules de transfert, on peut montrer les résultats suivants :

**Proposition 3.10**

Si  $X$  est une v.a.r. discrète prenant ses valeurs dans  $\{x_k, k \geq 1\}$  et si

$$\sum_{k \geq 1} x_k^2 \mathbf{P}[X = x_k] < +\infty$$

alors la variance de  $X$  est la quantité

$$\mathbf{Var} [X] = \mathbf{E} [(X - \mathbf{E} [X])^2] = \sum_{k \geq 1} (x_k - \mathbf{E} [X])^2 \mathbf{P} [X = x_k].$$

### Proposition 3.11

Si  $X$  est une v.a.r. admettant une densité  $f$  et si

$$\int_{\mathbb{R}} x^2 f(x) dx < +\infty,$$

alors la variance de  $X$  est la quantité

$$\mathbf{Var} [X] = \mathbf{E} [(X - \mathbf{E} [X])^2] = \int_{\mathbb{R}} (x - \mathbf{E} [X])^2 f(x) dx.$$

### Proposition 3.12

Soit  $X$  une v.a.r. de carré intégrable.

1.  $\mathbf{Var} [X] = \mathbf{E} [X^2] - \mathbf{E} [X]^2$ .
2. Pour tous  $a, b \in \mathbb{R}$ ,  $aX + b$  est de carré intégrable et

$$\mathbf{Var} [aX + b] = a^2 \mathbf{Var} [X].$$

Démonstration :

1.

2.

**Exemple :** [Calculs de variances.] On reprend les exemples de calculs d'espérances.

1.  $X \rightsquigarrow \mathcal{B}(n, p)$ . Comme  $X$  prend un nombre fini de valeurs, elle est de carré intégrable. On commence par calculer  $\mathbf{E} [X^2]$ .

Il vient, d'après la proposition précédente,

$$\mathbf{Var} [X] =$$

2.  $X \rightsquigarrow \mathcal{E}(\lambda)$ .

### Proposition 3.13 (Inégalité de Bienaymé-Tchebychev.)

Soit  $X$  une v.a.r. de carré intégrable. Alors,

$$\mathbf{P} [|X - \mathbf{E} [X]| > \varepsilon] \leq \frac{\mathbf{Var} [X]}{\varepsilon^2} \text{ pour tout } \varepsilon > 0.$$



Démonstration :

## 4 Vecteurs aléatoires

### 4.1 Définition

Dans cette partie, il s'agit de modéliser des phénomènes aléatoires prenant leurs valeurs dans  $\mathbb{R}^n$ ,  $n \geq 2$ .

- Exemple :**
1. On lance deux dés à 6 faces. Le résultat observé est un élément de  $\{1, \dots, 6\}^2$ , chaque élément du couple correspondant au résultat de l'un des dés.
  2. On lance une flèche sur une cible circulaire. On peut repérer l'impact de la flèche par un couple de réels, l'abscisse et l'ordonnée de l'impact dans un repère orthonormé par exemple.

Dans les deux exemples ci-dessus, le résultat peut être considéré comme un couple de variables aléatoires réelles ou encore un vecteur aléatoire de dimension deux  $(X, Y)$ . Le but de ce chapitre sera de décrire la loi des vecteurs aléatoires et de comprendre les liens entre loi du vecteur et loi des composantes du vecteur.

#### Définition 4.1

Un vecteur aléatoire  $X = (X_1, \dots, X_n)$  est une application

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R}^n \\ \omega &\longmapsto X(\omega) = (X_1(\omega), \dots, X_n(\omega)). \end{aligned}$$

### 4.2 Lois discrètes.

#### 4.2.1 Définition

**Exemple :** La loi du résultat du lancer des deux dés, si on ne donne pas plus de précision, peut être donnée comme suit : le couple  $(X, Y)$  prend ses valeurs dans  $\{1, \dots, 6\}^2$  et, pour tout  $(i, j) \in \{1, \dots, 6\}^2$ ,

$$\mathbf{P}[(X, Y) = (i, j)] = \mathbf{P}[[X = i] \cap [Y = j]] = \frac{1}{36}.$$

**Notation.** On note également, la probabilité que les dés donnent  $i$  et  $j$  par  $\mathbf{P}[X = i, Y = j]$ .

#### Définition 4.2

Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire prenant ses valeurs dans  $\{(x_{k_1}^1, \dots, x_{k_n}^n) / (k_1, \dots, k_n) \in \mathbb{N}^n\}$ . La loi du vecteur  $(X_1, \dots, X_n)$  est dite discrète et c'est la donnée, pour tous  $k_1, \dots, k_n$ , de

$$\mathbf{P}[X_1 = x_{k_1}^1, \dots, X_n = x_{k_n}^n] = \mathbf{P}[[X_1 = x_{k_1}^1] \cap \dots \cap [X_n = x_{k_n}^n]].$$

#### 4.2.2 Tirages avec plus de deux couleurs : la loi multinomiale

**Exemple :** On procède à  $n$  tirages avec remise dans une urne qui contient des boules de  $I$  couleurs distinctes,  $I \geq 2$ . Pour  $i = 1..I$ ,  $p_i$  est la proportion de boules de couleur  $i$ . On note  $X = (X_1, \dots, X_I)$  le vecteur aléatoire dont les composantes sont les nombres de boules de la couleurs  $i$  tirées. On cherche la loi de ce vecteur.

Le vecteur  $X$  prend ses valeurs dans l'ensemble  $\{0, \dots, n\}^I$ . Le résultat d'un tirage est un  $n$ -uplet dont les coordonnées sont les couleurs des boules tirées successivement. Le tirage est équiprobable.

La probabilité d'avoir un tirage donnant exactement  $x_1$  boules de couleur 1,  $x_2$  boules de couleur 2 etc. est  $\frac{n!}{x_1! \dots x_I!} p_1^{x_1} \dots p_I^{x_I}$ . Pour avoir tous les tirages donnant des nombres par couleur égaux, il faut faire toutes les permutations possibles, soit multiplier par  $I!$ . Ce faisant, on compte aussi les permutations des boules d'une même couleur, dont il ne faut pas tenir compte; il faut donc diviser le résultat par le nombre de ces permutations. Finalement, pour tout  $x = (x_1, \dots, x_I) \in \{0, \dots, n\}^I$  tel que  $x_1 + \dots + x_I = n$ ,

$$\mathbf{P}[X = x] = \mathbf{P}[X_1 = x_1, \dots, X_I = x_I] = \frac{n!}{x_1! \dots x_I!} p_1^{x_1} \dots p_I^{x_I}.$$

### 4.3 Lois à densité.

#### 4.3.1 Définition.

**Exemple :** Si l'on considère que l'impact de la flèche est réparti uniformément sur la cible et que la cible peut être assimilée à un disque  $D$  de rayon 1 du plan, alors on dira que le couple  $(X, Y)$  admet comme densité la fonction

$$f(x, y) = \frac{1}{\pi} \mathbf{1}_D(x, y),$$

où  $\lambda(D)$  est l'aire du disque  $D$  ( $\lambda(D) = \pi$ ).

#### Définition 4.3

Soit  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  une fonction positive d'intégrale 1. On dit qu'un vecteur aléatoire  $X = (X_1, \dots, X_n)$  admet pour densité  $f$  si sa loi est définie, pour tout  $B \subset \mathbb{R}^n$  par

$$\mathbf{P}[X \in B] = \int_B f(x) dx = \int \dots \int_B f(x_1, \dots, x_n) dx_1 \dots dx_n.$$

#### 4.3.2 Loi uniforme sur un ensemble de mesure bornée.

Soit  $B \subset \mathbb{R}^n$  un ensemble de mesure non nulle et bornée. On dit que  $X$  suit la loi uniforme sur  $B$  si elle admet pour densité la fonction  $f$  définie, pour tout  $x \in \mathbb{R}^n$ , par

$$f(x) = \frac{1}{\lambda_n(B)} \mathbf{1}_B(x),$$

où  $\lambda_n(B)$  est la mesure de  $B$ .

### 4.4 Lois marginales.

Les lois marginales sont les lois des composantes d'un vecteur aléatoire. Connaissant la loi d'un vecteur aléatoire, on peut déduire les lois marginales mais l'inverse est faux en général.

**Exemple :** On dispose de 10 jetons numérotés de 1 à 10 dont 5 sont blancs et 5 sont noirs. Si on tire un jeton au hasard, la probabilité que son numéro soit pair est 1/2, la probabilité que sa couleur soit blanc est 1/2 mais pour autant, on ne connaît pas la probabilité que le jeton soit pair et blanc. Pour cela on a besoin de savoir la proportion (ou le nombre) de jetons blancs parmi les jetons pairs (par exemple).

La loi marginale de la première composante d'un vecteur  $(X_1, \dots, X_n) \in \mathbb{R}^n$  s'obtient en remarquant que si  $B$  est un sous-ensemble de  $\mathbb{R}$ ,

$$\mathbf{P}[X_1 \in B] = \mathbf{P}[(X_1, X_2, \dots, X_n) \in B \times \mathbb{R}^{n-1}].$$

On déduit de cette remarque la loi marginale de n'importe quel vecteur extrait de  $(X_1, \dots, X_n)$ . Concrètement, calculer une loi marginale revient à "sommer" les probabilités sur toutes les valeurs possibles des autres composantes.

**Exemple :** [Cas discret] On reprend l'exemple de la loi multinomiale. La loi de  $X_1$  est donnée par les probabilités

$$\begin{aligned} \mathbf{P}[X = x_1] = \\ = \end{aligned}$$

Donc  $X_1$  suit une loi binomiale de paramètres  $(n, p_1)$ .

**Exemple :** [Cas à densité] On reprend l'exemple de la flèche lancée sur le disque  $D$  de centre 0 et de rayon 1. On suppose toujours que le point d'impact de la flèche est uniformément réparti sur la cible et donc que la loi de ce couple est  $f = \mathbf{1}_D/\pi$ . On cherche les lois de  $X$  et  $Y$ . Par symétrie, ces v.a. ont même loi. Pour avoir la loi de  $X$ , à nouveau l'idée est de « sommer » sur les valeurs de  $Y$ . Plus précisément, la densité de  $X$  (et donc de  $Y$ ) sera définie par

Généralisons ces résultats.

#### Proposition 4.1

Soit  $(X, Y)$  un couple de variables aléatoires réelles définies sur  $\Omega$ .

Si  $(X, Y)$  prend ses valeurs dans  $\{(x_i, y_j) / (i, j) \in \mathbb{N}^2\}$ ,

$$\mathbf{P}[X = x_i] = \sum_{j \in \mathbb{N}} \mathbf{P}[X = x_i, Y = y_j] \text{ pour tout } i \in \mathbb{N}.$$

Si  $(X, Y)$  admet pour densité  $f_{(X,Y)} : \mathbb{R}^2 \rightarrow \mathbb{R}$ , la densité de  $X$  est la fonction  $f_X$  définie par

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) dy \text{ pour tout } x \in \mathbb{R}.$$

**Remarque :** Par symétrie, on obtient des résultats analogues pour la v.a.r.  $Y$ . Les lois de  $X$  et de  $Y$ , quand elles sont calculées à partir de la loi du couple  $(X, Y)$ , sont appelées des *lois marginales*.

Nous présentons ici le cas de la dimension  $n$ .

#### Proposition 4.2

Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire défini sur  $\Omega$ .

Si  $X$  prend ses valeurs dans  $\{(x_{k_1}^1, \dots, x_{k_n}^n) / (k_1, \dots, k_n) \in \mathbb{N}^n\}$ , alors pour tout  $i \in \{1, \dots, n\}$ , pour tout  $k \in \mathbb{N}$ ,

$$\mathbf{P}[X_i = x_k^i] = \sum_{(k_1, \dots, k_{i-1}, k_{i+1}, \dots, k_n) \in \mathbb{N}^{n-1}} \mathbf{P}[X_1 = x_{k_1}^1, \dots, X_{i-1} = x_{k_{i-1}}^{i-1}, X_i = x_k^i, X_{i+1} = x_{k_{i+1}}^{i+1}, \dots, X_n = x_{k_n}^n].$$

Si  $X$  admet pour densité  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , la densité de  $X$  est la fonction  $f_X$  définie par

$$f_X(x) = \int_{\mathbb{R}^{n-1}} f(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n \text{ pour tout } x \in \mathbb{R}.$$

## 4.5 Moments de vecteurs aléatoires.

### 4.5.1 Formules de transfert

Pour pouvoir faire le calcul des moment de vecteurs aléatoires, il faudra savoir calculer l'espérance de fonctions réelles de vecteurs aléatoires. Ce résultat est donné par les formules de transfert.

#### Définition 4.4

Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire de dimension  $n$  prenant ses valeurs dans  $\{x_k, k \in \mathbb{N}\} \subset \mathbb{R}^n$  et soit  $\varphi$  une application de  $\mathbb{R}^n$  dans  $\mathbb{R}$ . On dit que la v.a.r.  $\varphi(X)$  est intégrable si

$$\sum_{k \in \mathbb{N}} |\varphi(x_k)| \mathbf{P}[X = x_k] < +\infty.$$

#### Définition 4.5

Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire de dimension  $n$  et de densité  $f$ , et soit  $\varphi$  une application de  $\mathbb{R}^n$  dans  $\mathbb{R}$ . On dit que la v.a.r.  $\varphi(X)$  est intégrable si

$$\int_{\mathbb{R}^n} |\varphi(x)| f(x) dx < +\infty.$$

#### Proposition 4.3 (Formules de transfert)

Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire de dimension  $n$  et soit  $\varphi$  une application de  $\mathbb{R}^n$  dans  $\mathbb{R}$ .

1. Si  $X$  prend ses valeurs dans  $\{x_k, k \in \mathbb{N}\} \subset \mathbb{R}^n$  et si  $\varphi(X)$  est intégrable, alors

$$\mathbf{E}[\varphi(X)] = \sum_{k \in \mathbb{N}} \varphi(x_k) \mathbf{P}[X = x_k].$$

2. Si  $X$  admet une densité  $f$  et si  $\varphi(X)$  est intégrable, alors

$$\mathbf{E}[\varphi(X)] = \int_{\mathbb{R}^n} \varphi(x) f(x) dx.$$

### 4.5.2 Covariance de deux variables réelles.

#### Définition 4.6

Soient  $X$  et  $Y$  deux v.a.r. de carré intégrable. Alors on peut définir la covariance de  $X$  et  $Y$  par

$$\text{cov}[X, Y] = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])].$$

On dit que  $X$  et  $Y$  sont décorrélées si  $\text{cov}[X, Y] = 0$ .

**Remarque :** Pour calculer la covariance de deux variables aléatoires, on a donc besoin de l'espérance du produit des variables  $X - \mathbf{E}[X]$  et  $Y - \mathbf{E}[Y]$ . Ce calcul, comme on le voit dans les formules de transfert ci-dessus, nécessite la loi jointe des variables aléatoires.

**Proposition 4.4**

1. La covariance est bilinéaire, i.e. les applications  $X \rightarrow \text{cov}[X, Y]$  et  $Y \rightarrow \text{cov}[X, Y]$  sont linéaires. La covariance est symétrique.
2. Pour tout  $X$  de carré intégrable et pour réel  $a$ ,  $\text{cov}[a, X] = 0$ .

Démonstration :

1.

2.

**Proposition 4.5**

Soient  $X$  et  $Y$  deux v.a.r. de carré intégrable. Alors

1.  $\text{cov}[X, Y] = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$ .
2.  $X + Y$  est de carré intégrable et  $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{cov}[X, Y]$ . En particulier, si  $X$  et  $Y$  sont décorrélées, alors  $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$ .

Démonstration :

1.

2.

**Proposition 4.6**

Si  $X_1, \dots, X_n$  sont  $n$  v.a.r. alors

$$\text{Var}\left[\sum_{k=1}^n X_k\right] = \sum_{k=1}^n \text{Var}[X_k] + \sum_{i \neq j} \text{cov}[X_i, X_j]$$

et si les v.a.r.  $X_1, \dots, X_n$  sont deux à deux décorrélées alors

$$\text{Var}\left[\sum_{k=1}^n X_k\right] = \sum_{k=1}^n \text{Var}[X_k].$$

Démonstration :

#### 4.5.3 Espérance d'un vecteur aléatoire.

(pas faite en cours)

##### Définition 4.7

Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire défini sur  $\Omega$  et à valeurs dans  $\mathbb{R}^n$ . On suppose que toutes les composantes du vecteur sont intégrables. Alors  $X$  est dit intégrable et son espérance est le vecteur de  $\mathbb{R}^n$

$$\mathbf{E}[X] = (\mathbf{E}[X_1], \dots, \mathbf{E}[X_n]).$$

On déduit aisément des propriétés de l'espérance des variables réelles, celles de l'espérance d'un vecteur aléatoire :

##### Proposition 4.7

Soient  $X$  et  $Y$  deux vecteurs aléatoires intégrables à valeur dans  $\mathbb{R}^n$  et soit  $a \in \mathbb{R}$  et  $b \in \mathbb{R}^n$  des constantes. Alors

1.  $aX + b$  est intégrable et  $\mathbf{E}[aX + b] = a\mathbf{E}[X] + b$ .
2.  $X + Y$  est intégrable et  $\mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y]$ .

#### 4.5.4 Covariance de deux vecteurs aléatoires.

(pas faite en cours)

On peut prendre la covariance de vecteurs aléatoires de taille différente du moment qu'ils sont définis sur le même univers  $\Omega$ .

##### Définition 4.8

Soient  $X = (X_1, \dots, X_n)$  et  $Y = (Y_1, \dots, Y_m)$  deux vecteurs aléatoires définis sur  $\Omega$  et à valeurs dans  $\mathbb{R}^n$  et  $\mathbb{R}^m$  respectivement. On suppose que les composantes des deux vecteurs sont de carré intégrable. Alors, la covariance des vecteurs  $X$  et  $Y$  est la matrice réelle  $n \times m$

$$\mathbf{cov}[X, Y] = (\mathbf{cov}[X_i, Y_j])_{i=1..n, j=1..m}.$$

Si  $\mathbf{cov}[X, Y] = 0$ ,  $X$  et  $Y$  sont dits décorrélés.

**Remarque :** Remarquer que  $X$  et  $Y$  sont décorrélés si et seulement si les composantes de  $X$  et les composantes de  $Y$  sont deux à deux décorrélées.

##### Proposition 4.8

La covariance est bilinéaire.

#### 4.5.5 Variance (ou covariance) d'un vecteur aléatoire.

(pas faite en cours)

##### Définition 4.9

Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire défini sur  $\Omega$  et à valeurs dans  $\mathbb{R}^n$ . On suppose que toutes les composantes du vecteur sont de carré intégrable. On définit alors la variance ou la covariance de  $X$  par la matrice

$n \times n$

$$\mathbf{Var} [X] = \mathbf{cov} [X, X] = (\mathbf{cov} [X_i, X_j])_{i=1..n, j=1..n}.$$

De la bilinéarité de la covariance de vecteurs, on déduit la

#### Proposition 4.9

1. Soient  $X$  un vecteur aléatoire de  $\mathbb{R}^n$  de composantes de carré intégrable,  $a$  un réel et  $b$  un vecteur de  $\mathbb{R}^n$ . Alors

$$\mathbf{Var} [aX + b] = a^2 \mathbf{Var} [X].$$

2. Soient  $X$  et  $Y$  deux vecteurs aléatoires de  $\mathbb{R}^n$  de composantes de carré intégrable. Alors

$$\mathbf{Var} [X + Y] = \mathbf{Var} [X] + \mathbf{cov} [X, Y] + \mathbf{cov} [Y, X] + \mathbf{Var} [Y].$$

## 4.6 Variables aléatoires indépendantes.

#### Définition 4.10

Soient  $X$  et  $Y$  deux v.a.r. sur  $\Omega$ . On dit que  $X$  et  $Y$  sont indépendantes si, pour tous  $A, B \subset \mathbb{R}$ ,

$$\mathbf{P} [X \in A, Y \in B] = \mathbf{P} [X \in A] \mathbf{P} [Y \in B].$$

#### Proposition 4.10

Si  $(X, Y)$  est un couple de variables aléatoires à valeurs dans  $\{(x_i, y_j) / (i, j) \in \mathbb{N}^2\}$ ,  $X$  et  $Y$  sont indépendantes si et seulement si

$$\mathbf{P} [X = x_i, Y = y_j] = \mathbf{P} [X = x_i] \mathbf{P} [Y = y_j] \text{ pour tout } (i, j) \in \mathbb{N}^2.$$

#### Proposition 4.11

Si  $(X, Y)$  est un couple de variables aléatoires qui admet pour densité  $f_{(X,Y)} : \mathbb{R}^2 \rightarrow \mathbb{R}$ , et si  $f_X, f_Y$  sont les densités marginales de  $X$  et  $Y$  respectivement,  $X$  et  $Y$  sont indépendantes si et seulement si

$$f_{(X,Y)}(x, y) = f_X(x) f_Y(y) \text{ pour presque tout } (x, y) \in \mathbb{R}^2.$$

#### Définition 4.11

Plus généralement, si  $(X_1, \dots, X_n)$  est un vecteur aléatoire, on dit que les v.a.r.  $X_1, \dots, X_n$  sont indépendantes si, pour tous  $B_1, \dots, B_n \subset \mathbb{R}$ ,

$$\mathbf{P} [X_1 \in B_1, \dots, X_n \in B_n] = \mathbf{P} [X_1 \in B_1] \dots \mathbf{P} [X_n \in B_n].$$

L'expression de cette indépendance dans les cas discrets et à densité est une extension évidente du cas  $n = 2$  ci-dessus et est laissée au lecteur.

**Exemple :** [Cas discret] On reprend l'exemple des 10 jetons. On suppose qu'il y a  $n$  jetons blancs pairs et  $n'$  jetons blancs impairs (sans supposer  $n + n' = 5$  mais seulement  $n + n' \leq 10$ ).

<i>couleur \ parite</i>	pair	impair	couleur
blanc	n/10	n'/10	
noir			
parite	1/2	1/2	1

**Exemple :** Dans le cas de la flèche lancée sur la cible, les variables  $X$  et  $Y$  sont-elles indépendantes ?

Plus généralement, on peut montrer qu'une condition nécessaire pour que les v.a.  $X$  et  $Y$  d'un couple de v.a.  $(X, Y)$  admettant une densité  $f$  soient indépendantes est que le support de  $f$  (l'adhérence de l'ensemble où  $f$  ne s'annule pas) soit du type  $A \times B$  où  $A, B \subset \mathbb{R}$ .

**Proposition 4.12**

*Si  $X$  et  $Y$  sont deux v.a.r. indépendantes et de carré intégrable alors elles sont décorrélées, c'est-à-dire que  $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$ .*

**Remarque :** La réciproque est fautive en général. En effet, toujours avec notre cible, il est clair que

$$\mathbf{E}[X] = \mathbf{E}[Y] =$$

D'autre part,  $XY$  prend ses valeurs dans  $[-1, 1]$  et, grâce à la formule de transfert, on a :

$$\mathbf{E}[XY] = \int_{\mathbb{R}^2} xy \frac{1}{\pi} \mathbf{1}_D(x, y) dx dy =$$

Par conséquent,  $\text{cov}[(X, Y)] = 0$ , les variables sont décorrélées mais non indépendantes (vu plus haut).

**Démonstration :** On démontre la proposition dans le cas d'un couple de variables à densité (pour des variables discrètes, la preuve est similaire). On suppose  $X$  et  $Y$  indépendantes, de carré intégrable, et que le couple  $(X, Y)$  admet une densité  $f$ . On note  $f_X, f_Y$  les densités marginales de  $X$  et  $Y$  respectivement. Par la formule de transfert,

$$\mathbf{E}[XY] =$$

Comme  $X$  et  $Y$  sont indépendantes, on a :  
théorème de Fubini,

. Et donc, par le

**Proposition 4.13**

*Soient  $X_1, \dots, X_n$   $n$  v.a.r. indépendantes.*

1. *Elles sont indépendantes deux à deux mais la réciproque est fautive.*
2. *On connaît la loi de  $(X_1, \dots, X_n)$  dès que l'on connaît les lois de chacune des variables  $X_i, i \in \{1, \dots, n\}$ .*
3. *Soit  $p \in \{1, \dots, n-1\}$ . Alors, pour toutes fonctions  $\phi : \mathbb{R}^p \rightarrow \mathbb{R}$  et  $\psi : \mathbb{R}^{n-p} \rightarrow \mathbb{R}$ , les variables  $\phi(X_1, \dots, X_p)$  et  $\psi(X_{p+1}, \dots, X_n)$  sont indépendantes.*

## 4.7 Somme de v.a. indépendantes.

Un cas où l'indépendance de variables aléatoires est particulièrement utile est celui du calcul de la loi de la somme de v.a. En effet, en général, on ne sait rien de la somme de v.a. si l'on ne fait pas d'hypothèse supplémentaire.



En toute généralité, la loi de  $X + Y$  est connue dès que l'on connaît la loi du couple. En effet, si par exemple  $(X, Y)$  admet une densité  $f$ , alors pour toute fonction bornée  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ ,

$$\mathbf{E}[\varphi(X + Y)] = \int_{\mathbb{R}^2} \varphi(x + y) f(x, y) dx dy.$$

Cela dit, lorsque les v.a. sont indépendantes, on peut pousser les calculs plus loin.

**Proposition 4.14 (Cas discret)**

Soient  $X$  et  $Y$  deux v.a.r. indépendantes. On pose  $Z = X + Y$ . On suppose que  $X$  et  $Y$  prennent leurs valeurs dans  $\{x_i, i \in \mathbb{N}\}$  et  $\{y_j, j \in \mathbb{N}\}$  respectivement. On note  $Z(\Omega)$  l'ensemble (dénombrable) de toutes les valeurs possibles de  $x_i + y_j$  quand  $(i, j) \in \mathbb{N}^2$ . Alors, pour tout  $z \in Z(\Omega)$ ,

$$\mathbf{P}[Z = z] = \sum_{i \in \mathbb{N}} \mathbf{P}[X = x_i] \mathbf{P}[Y = z - x_i]$$

Démonstration : En effet, soit  $z \in Z(\Omega)$ ,

$$\mathbf{P}[Z = z] =$$

$$= \sum_{j \in \mathbb{N}} \mathbf{P}[X = z - y_j] \mathbf{P}[Y = y_j].$$

**Proposition 4.15**

Soient  $X, Y$  deux v.a. indépendantes et soit  $Z = X + Y$ . On suppose que  $X$  et  $Y$  admettent des densités  $f$  et  $g$  respectivement. Alors  $Z$  admet pour densité  $h = f * g$ .

Démonstration : Les v.a. étant indépendantes, le couple  $(X, Y)$  admet pour densité  $(x, y) \mapsto f(x)g(y)$ .

Pour trouver la loi de  $Z = X + Y$ , on utilise la méthode vue dans le chapitre sur les moments : soit  $\varphi$  une application bornée de  $\mathbb{R}$  dans  $\mathbb{R}$ . On a, en utilisant la formule de transfert,

$$\mathbf{E}[\varphi(Z)] =$$

$$\Rightarrow \mathbf{E}[\varphi(Z)] =$$

Puis, en faisant le changement de variable  $z = x + y$ , on obtient

$$\mathbf{E}[\varphi(Z)] =$$

$$\mathbf{E}[\varphi(Z)] =$$

Donc  $Z$  a pour densité la fonction  $h = f * g$  (la convolution de  $f$  et  $g$ ) définie, pour presque tout  $z \in \mathbb{R}$ , par

$$h(z) = \int_{\mathbb{R}} f(x)g(z - x)dx = \int_{\mathbb{R}} f(z - y)g(y)dy$$

Les exemples qui suivent et des résultats du même type sont dans l'annexe « lois usuelles ». Il faut connaître ces résultats.

**Exemple :** On suppose que  $X \rightsquigarrow \mathcal{B}(n, p)$  et  $Y \rightsquigarrow \mathcal{B}(m, p)$  et que  $X$  et  $Y$  sont indépendantes. On cherche la loi de  $X + Y$ .

La v.a.  $Z$  prend ses valeurs dans \_\_\_\_\_ et, pour tout  $k \in$  \_\_\_\_\_, en appliquant la proposition ci-dessus :

$$\mathbf{P}[Z = k] =$$

$$\mathbf{P}[Z = k] =$$

Donc  $Z \rightsquigarrow \mathcal{B}(n + m, p)$ .

Bien sûr, si  $X$  et  $Y$  ne sont pas indépendantes, le résultat est faux : par exemple, si  $Y = X$ , on a  $Z = 2X$  qui est une variable paire (et donc pas une v.a. binomiale).

**Exemple :** On suppose que  $X$  et  $Y$  sont indépendantes et suivent des lois normales  $\mathcal{N}(0, 1)$ . Alors  $X + Y \rightsquigarrow \mathcal{N}(0, 2)$ .

Alors, d'après la proposition précédente, si on note respectivement  $f$  et  $g$  les densités de  $X$  et  $Y$  (ici,  $f = g$ ),  $Z$  a pour densité la fonction  $h$  définie sur  $\mathbb{R}$  par

$$h(z) = \int_{\mathbb{R}} f(x)g(z - x)dx.$$

Rappelons que la densité d'une v.a.  $\mathcal{N}(m, \sigma^2)$  est

$$x \mapsto \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m)^2/(2\sigma^2)} \text{ pour tout } x \in \mathbb{R}.$$

Donc on a ( $\sigma^2 = 1$  et  $m = 0$ )

$$f(x) = g(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ pour tout } x \in \mathbb{R}$$

et on veut ( $\sigma^2 = 2$ ,  $m = 0$ )

$$h(z) = \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-z^2/4} \text{ pour tout } z \in \mathbb{R}.$$

Or,

$$h(z) =$$

On remarque que

$$(*) \quad x^2 + (x - z)^2 =$$

d'où le résultat.

Remarquer que dans « lois usuelles » le résultat est plus général : si  $X_1$  et  $X_2$  sont deux v.a. indépendantes suivant respectivement des lois  $\mathcal{N}(m_1, \sigma_1^2)$  et  $\mathcal{N}(m_2, \sigma_2^2)$ , alors  $X_1 + X_2 \rightsquigarrow \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$ . Pour montrer cela, on commence par montrer que la propriété est satisfaite lorsque  $m_1 = m_2 = 0$  (comme ci-dessus mais la factorisation (\*) est plus difficile) puis on remarque que, si  $X_i \rightsquigarrow \mathcal{N}(m_i, \sigma_i^2)$  alors  $X_i - m_i \rightsquigarrow \mathcal{N}(0, \sigma_i^2)$ .

La proposition établit en outre que la somme de deux v.a. indépendantes discrètes est une v.a. discrète et que la somme de deux v.a. indépendantes à densité est une v.a. à densité. Cette remarque suggère qu'il existe des variables aléatoires ni discrètes ni à densité (en sommant deux v.a. indépendantes de nature différente).

Voici un autre exemple dans  $\mathbb{R}^2$  de v.a. ni discrète ni à densité. On suppose que  $X$  suit la loi uniforme sur  $[0, 1]$ . Alors  $(X, X)$  prend ses valeurs uniformément dans  $\Delta = \{(x, x) / x \in [0, 1]\}$  qui n'est pas un ensemble discret de  $\mathbb{R}^2$  et qui est négligeable dans  $\mathbb{R}^2$  (et donc  $(X, X)$  ne peut admettre une densité).

Tous ces cas « particuliers » qui constituent en fait le cas général justifient l'utilité d'une théorie des probabilités unifiée entièrement fondée sur la théorie de la mesure.

## 5 Suites de variables aléatoires - Résultats asymptotiques

### 5.1 Un exemple important.

Une des grandes questions de la théorie des probabilités a été la suivante : si je lance un très grand nombre de fois une pièce de monnaie et que je calcule la *fréquence observée* de « pile », est-ce que j'obtiens, à la limite quand le nombre de tirages tend vers l'infini, la *fréquence théorique* de « pile », à savoir, si la pièce est équilibrée,  $1/2$  ? De plus, puis-je connaître la distribution de cette fréquence observée quand  $n$  devient grand ?

Modélisons le problème. On modélise chaque tirage par une v.a. de Bernoulli de paramètre  $1/2$ .

Comme le nombre de tirages est destiné à tendre vers l'infini, on se donne donc une suite  $(X_n)_{n \geq 1}$  de variables suivant la loi  $\mathcal{B}(1/2)$ . Il est assez naturel de supposer que les tirages sont indépendants, ce qui se traduit par le fait que les variables sont indépendantes, c'est-à-dire que tout vecteur aléatoire formé par un nombre fini quelconque de v.a. de cette suite forme une famille de v.a. indépendantes.

Si on compte 1 lorsque la pièce tombe sur « pile » et 0 sinon, la fréquence de « pile » observée après  $n$  tirages sera

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k.$$

La première question à laquelle il faut répondre est : a-t-on

$$\lim_{n \rightarrow +\infty} \bar{X}_n = \frac{1}{2} ?$$

Et si oui, en que sens ?

Les théorèmes qui répondent à ce type de questions sont connus sous le nom de « lois des grands nombres ». On en verra un en particulier.

La seconde question, à savoir : « quelle est la distribution, quand  $n$  devient grand, de la fréquence observée de « pile »  $\bar{X}_n$  ? » est plus délicate. La réponse est contenue dans un résultat célèbre appelé le « théorème limite central » qui dit que pour  $n$  assez grand,  $\bar{X}_n$  suit approximativement une loi normale.

Ces résultats dépassent largement le cadre de l'exemple proposé puisqu'ils ne dépendent pas de la loi de la variable dont on évalue la moyenne (ici la loi de Bernoulli).

Avant de voir ces résultats, définissons deux notions de convergence.

### 5.2 Différents types de convergence.

#### Définition 5.1

Soit  $(X_n)_{n \geq 1}$  une suite de v.a.r. définies sur  $\Omega$  et soit  $X$  une v.a.r. définie sur le même univers.

1.  $(X_n)_{n \geq 1}$  converge presque sûrement vers  $X$  s'il existe un négligeable  $N \subset \Omega$  (i.e. un élément de  $\mathcal{P}(\Omega)$  satisfaisant  $\mathbf{P}[N] = 0$ ) tel que,  $X_n(\omega) \rightarrow X(\omega)$  pour tout  $\omega \in N^c$ .
2. On dit que  $(X_n)_{n \geq 1}$  converge en loi vers  $X$  si, pour tout  $t \in \mathbb{R}$  auquel  $F_X$  est continue,  $\lim_{n \rightarrow +\infty} F_{X_n}(t) = F_X(t)$ .

**Remarque :** La notion de convergence presque sûre correspond à celle de convergence presque partout pour les suites de fonctions réelles. En revanche, la notion de convergence en loi est totalement neuve. On utilise couramment d'autres types de convergence "en moyenne", "en moyenne quadratique", "en probabilité" que nous ne définissons pas ici.

**Proposition 5.1 (admise)**

Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires réelles définies sur  $\Omega$  et soit  $X$  une v.a.r. définie sur le même univers. Si  $(X_n)_{n \geq 1}$  converge vers  $X$  presque sûrement alors  $(X_n)_{n \geq 1}$  converge en loi vers  $X$ .

**Remarque :** La réciproque de l'implication de la proposition précédente est fautive en général. Deux variables de même loi peuvent être différentes en chaque point  $\omega \in \Omega$  et  $(X_n)_{n \geq 1}$  converge en loi vers  $X$  signifie seulement que la loi de  $(X_n)_{n \geq 1}$  converge vers la loi de  $X$ . On notera cette convergence :  $\mathcal{L}(X_n) \rightarrow \mathcal{L}(X)$  quand  $n \rightarrow +\infty$ .

**Exemple :** Soit  $Y$  définie sur  $\Omega$  une variable aléatoire réelle de loi  $\mathcal{B}(1/2)$ . Soit  $X = 1 - Y$ . Alors  $X$  a la même loi que  $Y$ . Pour tout  $n$ , on pose  $X_n = Y$  et  $X = 1 - Y$ .

Alors  $(X_n)_{n \geq 1}$  converge trivialement vers  $X$  en loi. Et pourtant  $|X_n - X| = 1$  sur  $\Omega$  et donc on n'a pas convergence presque sûre.

**Remarque :** Lorsque les v.a.r. sont entières, la convergence en loi peut s'exprimer plus simplement. Plus précisément, on suppose que, pour tout  $n \geq 1$ ,  $X_n$  prend ses valeurs dans  $\mathbb{N}$ . Alors  $(X_n)_{n \geq 1}$  converge en loi vers  $X$  si et seulement si, pour tout  $k \in \mathbb{N}$ ,

$$\mathbf{P}[X = k] = \lim_{n \rightarrow +\infty} \mathbf{P}[X_n = k].$$

### 5.3 Théorèmes de convergence.

**Proposition 5.2 (Convergence de la loi hypergéométrique vers la loi binomiale)**

Soit  $(X_N)_{N \geq N}$  une suite de v.a. de loi  $\mathcal{H}(n, N, p)$ . Alors  $(X_N)_{N \geq 1}$  converge en loi, lorsque  $N$  tend vers  $+\infty$  vers la loi  $\mathcal{B}(n, p)$ .

L'idée est la suivante : on souhaite faire un sondage avant des élections mettant en compétition deux candidats. Pour cela, on tire un échantillon de  $n$  individus dans une population de taille  $N$  que l'on suppose comporter une proportion  $p$  d'individus préférant le candidat 1 au candidat 2.

Le résultat que nous allons montrer dit en substance que, lorsque  $N$  est suffisamment grand, la loi du nombre d'individus tirés sans remise préférant le candidat 1, soit

est proche de la loi de ce nombre lorsque les individus sont tirés avec remise, soit

Remarque qu'en pratique, on procède à ce genre de tirage sans remise, mais le résultat est utile car les calculs pour la loi binomiale sont plus simples.

**Démonstration :** On note donc  $X_N$  le nombre d'individus préférant le candidat 1 sur les  $n$  tirés dans une population de  $N \gg n$  individus. La v.a.  $X_N$  prend ses valeurs dans  $\{0, \dots, n\}$  et, pour tout  $k \in \{0, \dots, n\}$ ,

$$\mathbf{P}[X_N = k] =$$

=

On remarque que, pour  $N$  au voisinage de  $+\infty$ ,

$$\frac{(N-1)!}{N!} = \frac{(N-2)!}{N!} = \dots, \frac{(N-n)!}{N!} \simeq$$

Par conséquent

$$\lim_{N \rightarrow +\infty} \mathbf{P}[X_N = k] =$$

pour tout  $k \in \{0, \dots, n\}$  et donc  $(X_N)_N$  converge bien en loi vers une  $\mathcal{B}(n, p)$ .

**Proposition 5.3 (Convergence de la loi binomiale vers la loi de Poisson)**

Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires définies sur un même  $\Omega$ . On suppose que, pour tout  $n \geq 1$ ,  $X_n$  suit une loi  $\mathcal{B}(n, p_n)$ . On suppose en outre que  $np_n \rightarrow \lambda > 0$  quand  $n \rightarrow +\infty$ . Alors  $(X_n)_{n \geq 1}$  converge en loi vers la loi de Poisson  $\mathcal{P}(\lambda)$ .

On verra la démonstration de ce résultat en TD.

**Théorème 5.1 (Une loi des grands nombres pour des v.a.r. décorrélées.)**

Soit  $(X_n)_{n \geq 1}$  une suite de v.a.r. de carré intégrable, décorrélées, d'espérance  $m$  et dont la suite des variances est bornée par une constante  $C > 0$ . Pour tout  $n$ , on note

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}.$$

Alors, pour tout  $\varepsilon > 0$ , on a les estimations

$$\mathbf{P}[|\bar{X}_n - m| > \varepsilon] \leq \frac{C}{n\varepsilon^2} \quad \text{et} \quad \mathbf{E}[(\bar{X}_n - m)^2] \leq \frac{C}{n}.$$

De plus, la suite  $(\bar{X}_n)_{n \geq 1}$  converge presque sûrement vers  $m$ .

Démonstration : On calcule (cf. TD)

$$\mathbf{E}[\bar{X}_n] =$$

et, puisque les v.a. sont décorrélées,

. On en déduit que

On remarque que cette quantité tend vers 0 quand  $n \rightarrow +\infty$  : cette convergence s'appelle la convergence en moyenne quadratique.

En appliquant l'inégalité de

On remarque que cette quantité tend vers 0 à  $\varepsilon$  fixé quand  $n$  tend vers l'infini : ce type de convergence s'appelle la convergence en probabilité.

On admet la convergence presque sûre.

Le dernier théorème que nous allons voir dans ce cours est principalement utilisé pour faire des approximations de probabilités. Avant de l'énoncer, on va introduire la problématique sur un exemple (réel) :

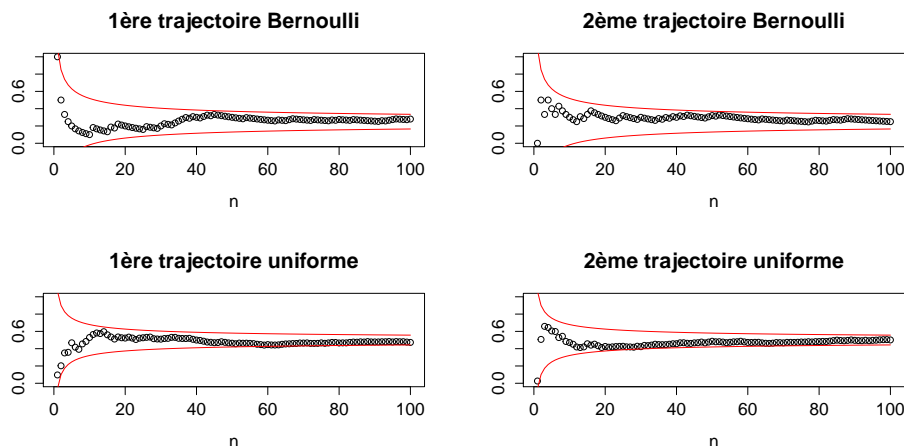


FIGURE 1 – Loi des Grands Nombres (Bernoulli ( $p = 0.25$ ) et uniforme sur  $[0, 1]$ )

**Exemple :** 1.359.670 ( $N_1$ ) garçons et 1.285.086 ( $N_2$ ) filles sont nés en Suisse entre 1871 et 1900. Ces données sont-elles compatibles avec l’hypothèse que les sexes des nouveaux nés sont déterminés indépendamment les uns des autres et également distribués ?

On pose  $N = N_1 + N_2 = 2.644.756$ . Pour tout  $n \in \{1, \dots, N\}$ , on note  $X_n$  le sexe du  $n$ -ième né sur la période concernée. Plus précisément, on suppose que  $X_n$  vaut 1 lorsque l’enfant est un garçon et  $-1$  sinon. Si l’on admet l’hypothèse ci-dessus alors les  $(X_n)_{n=0..N}$  forment une suite de v.a. indépendantes et toutes d’espérance et de variance .

On pose  $S_N = X_1 + \dots + X_N$ . Cette variable est l’écart entre le nombre de garçons et le nombre de filles. Pour vérifier si notre hypothèse est juste, on va vérifier que l’événement  $[S_N \geq N_1 - N_2]$  a une probabilité suffisante sous cette hypothèse. Il faut donc calculer

$$\mathbf{P}[S_N \geq N_1 - N_2].$$

On devrait pouvoir le faire car  $S_N$  suit (à peu de chose près) une loi binomiale. Mais les calculs sont très gros (des factorielles avec de très grands nombres). On cherche donc une bonne approximation de cette probabilité. Le théorème limite central va nous la fournir.

Ce théorème repose non plus sur la décorrélation des variables de la suite mais sur leur indépendance. Précisons la définition.

### Définition 5.2

Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires. On dit que les v.a. de cette suite sont indépendantes si, pour tout  $N \geq 2$ , les v.a.  $X_1, \dots, X_N$  sont indépendantes, au sens vu au chapitre précédent.

### Théorème 5.2 (Théorème limite central.)

Soit  $(X_n)_{n \geq 1}$  une suite de v.a.r. de carré intégrable, indépendantes et identiquement distribuées (c’est-à-dire de même loi — i.i.d. en abrégé). On pose  $m = \mathbf{E}[X_1]$ ,  $\sigma^2 = \mathbf{Var}[X_1]$ . Alors, quand  $n \rightarrow +\infty$ ,

$$\mathcal{L}\left(\frac{X_1 + \dots + X_n - nm}{\sqrt{n}\sigma}\right) \rightarrow \mathcal{N}(0, 1).$$

Ce théorème, qui sera très largement utilisé en Statistique, est admis.

Remarquer que

$$\frac{X_1 + \dots + X_n - nm}{\sqrt{n}\sigma} = \frac{\bar{X}_n - m}{\sigma/\sqrt{n}} = \frac{\bar{X}_n - \mathbf{E}[\bar{X}_n]}{\sqrt{\mathbf{Var}[\bar{X}_n]}}$$

qui est de moyenne 0 et de variance 1.

On peut utiliser ce théorème pour faire l'approximation suivante :

$$\bar{X}_n \rightsquigarrow \mathcal{N}\left(m, \frac{\sigma^2}{n}\right) \text{ approximativement.}$$

Appliqué à notre problème, il donne

$$\mathbf{P}[S_N \geq N_1 - N_2]$$

Par le théorème limite central,

Conclusion :

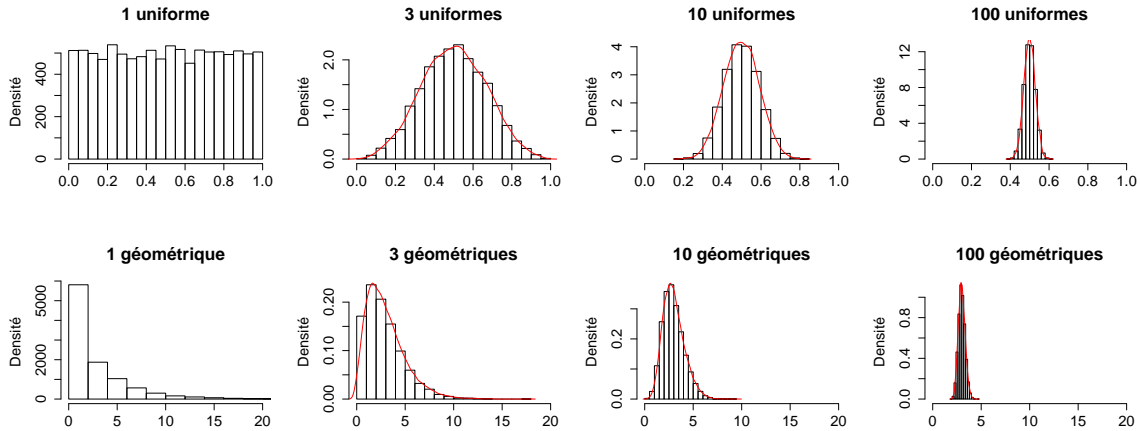


FIGURE 2 – TCL-Illustration

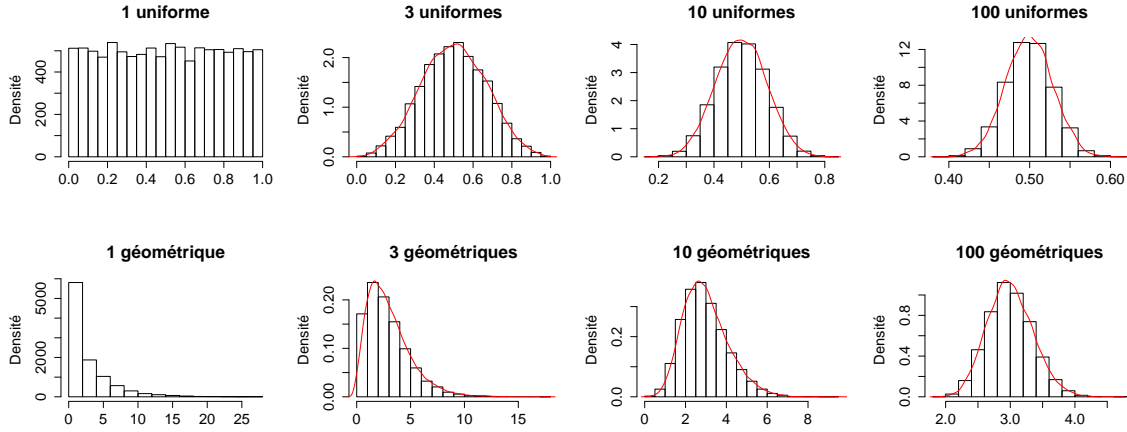


FIGURE 3 – TCL-Illustration - zoom

## 5.4 Simulation de variables aléatoires

La plupart des langages informatiques mettent à disposition de l'utilisateur un générateur de nombre pseudo-aléatoires. Il s'agit en fait de suites récurrentes entières avec une très grande périodicité qui sont initialisées via des paramètres liés à l'ordinateur (par exemple l'horloge interne). On considère qu'approximativement, via sa fonction "random", un langage fournit une suite de v.a. entières indépendantes de loi uniforme sur l'ensemble  $\{0, \dots, N_0\}$ ,  $N_0$  étant l'entier maximal de l'ordinateur.

A partir de cette suite de v.a. indépendantes, on doit être capable de construire une suite de v.a. indépendantes suivant n'importe quelle loi.

### 5.4.1 Simulation d'une loi uniforme sur $[0, 1[$ .

#### Proposition 5.4

On définit  $(U_N)_{N \geq 1}$  une suite de v.a. telle que, pour  $N$  fixé,  $U_N$  suit une loi uniforme sur

$$\left\{0, \frac{1}{N}, \frac{2}{N}, \dots, \frac{N-1}{N}\right\}.$$

Alors  $(U_N)_{N \geq 1}$  converge en loi vers la loi uniforme sur  $[0, 1[$ .

Démonstration : Pour tout  $N \geq 1$ , on définit  $F_N$  la fonction de répartition de  $U_N$ , soit

La fonction de répartition d'une loi uniforme est



Cette fonction est continue et par conséquent, la convergence en loi est la convergence ponctuelle de la suite  $(F_N)_{N \geq 1}$ . Pour  $t < 0$  et  $t \geq 1$ ,  $\lim_{N \rightarrow +\infty} F_N(t) = 0$  et  $1$  respectivement. Pour  $0 \leq t < 1$ ,  $\lim_{N \rightarrow +\infty} F_N(t) = t$ , où  $[x]$  désigne la partie entière d'un réel  $x$ . Or, pour tout  $t \in \mathbb{R}$ ,  $\lim_{N \rightarrow +\infty} [Nt]/N = t$ , d'où le résultat.

Cette convergence en loi nous permet de considérer,  $N_0$  (l'entier maximal de l'ordinateur) étant très grand, que la fonction "random" de l'ordinateur divisée par  $N_0 + 1$  fournit la réalisation d'une v.a. suivant une loi uniforme sur  $[0, 1[$  (ou  $]0, 1[$  ou  $]0, 1]$  ou  $[0, 1]$ ). Des réalisations successives fournies par l'ordinateur seront considérées comme mutuellement indépendantes par construction.

#### 5.4.2 Simulation d'une v.a. de Bernoulli

Pour simuler une v.a. de Bernoulli de paramètre  $p \in ]0, 1[$ , il suffit de simuler une v.a.  $U$  uniforme sur  $]0, 1[$  et de poser  $X = 1_{[U \leq p]}$  (i.e.  $X = 1$  si  $U \leq p$  — donc avec probabilité  $p$  — et 0 sinon).

#### 5.4.3 Simulation d'une v.a. binomiale

Pour simuler une v.a. binomiale de paramètres  $n \geq 1$  et  $p \in ]0, 1[$ , il suffit de sommer  $n$  v.a. de Bernoulli de paramètre  $p$  indépendantes.

#### 5.4.4 Simulation d'une v.a. admettant une densité continue.

##### Proposition 5.5

Soit  $X$  une v.a. de densité  $f$  continue et de fonction de répartition  $F$  inversible. On pose  $Y = F(X)$ . Alors  $Y$  suit une loi uniforme sur  $[0, 1]$ .

Démonstration : D'après la proposition 3.8 sur les changements de variables bijectifs pour les v.a. à densité,  $Y$  admet pour densité la fonction  $g$  définie, pour  $y \in ]0, 1[$  par

Par ailleurs, pour  $y < 0$ ,  $\mathbf{P}[F(X) \leq y] = 0$  et pour  $y \geq 1$ ,  $\mathbf{P}[F(X) \leq y] = 1$  et le théorème est démontré.

Pour simuler la v.a.  $X$ , si  $F^{-1}$  est connue analytiquement, il suffit donc de simuler une loi uniforme sur  $]0, 1[$  et de calculer  $F^{-1}$  du résultat obtenu.

#### 5.4.5 Simulation d'une v.a. bornée à densité bornée.

##### Proposition 5.6

Soit  $X$  une v.a. presque sûrement bornée de densité  $f$  presque partout bornée. On suppose pour simplifier que  $X \in [0, 1]$  et on choisit  $m$  tel que  $f \leq m$  p.p. Soient  $U$  et  $V$  deux v.a. indépendantes de lois uniformes sur  $[0, 1]$  et  $[0, m]$  respectivement. Alors, conditionnellement à  $[V < f(U)]$ ,  $U$  a même loi que  $X$ .

Démonstration : Soit  $x \in \mathbb{R}$ . On calcule  $\mathbf{P}[U \leq x | V < f(U)]$ .

Par la formule de transfert appliquée à une fonction du couple  $(U, V)$  de densité égale au produit de ses marginales, on a, si  $x \in [0, 1]$

$$\begin{aligned} \mathbf{P}[U \leq x, V < f(U)] &= \int_0^x \int_0^{f(u)} \frac{dv}{m} du \\ &= \int_0^x \frac{f(u)}{m} du = \int_{-\infty}^x \frac{f(u)}{m} du \end{aligned}$$

puisque  $X$  prend ses valeurs dans  $[0, 1]$ . De plus,

$$\mathbf{P}[V < f(U)] = \int_0^1 \int_0^{f(u)} \frac{1}{m} dv du = \int_0^1 \frac{f(u)}{m} du = \frac{1}{m}.$$

Finalement,  $\mathbf{P}[U \leq x|V < f(U)] = \int_{-\infty}^x f(u)du$  pour  $x \in [0, 1[$  et on peut vérifier facilement que cela reste vrai pour  $x < 0$  (probabilité nulle) et  $x \geq 1$  (probabilité égale à 1).

A partir de ce résultat, on construit la procédure de simulation suivante : étant donnée une v.a.  $X$  comme dans la proposition, on procède à la simulation de  $U$  et  $V$  indépendantes. Si les réalisations  $u$  et  $v$  de ces v.a. satisfont  $v < f(u)$  la réalisation de la simulation de  $X$  est  $u$ , sinon on procède à un autre tirage des v.a.  $U$  et  $V$ .

Cette méthode est utilisée pour simuler des lois bêta.

### Définition 5.3 (Lois Beta)

1. On dit que la v.a.  $X$  suit une loi bêta de type I de paramètres réels  $n$  et  $p > 0$  et on note  $X \rightsquigarrow \text{bêta I}(n, p)$  si elle est comprise entre 0 et 1 et si sa densité est

$$f(x) = \frac{1}{B(n, p)} x^{n-1} (1-x)^{p-1} \mathbf{1}_{]0,1[}(x)$$

où la fonction  $B$  est donnée par

$$B(n, p) = \int_0^1 x^{n-1} (1-x)^{p-1} dx.$$

2. On dit que la v.a.  $Y \geq 0$  suit une loi bêta de type II de paramètres  $n, p$  et on note  $Y \rightsquigarrow \text{bêta II}(n, p)$  si elle admet pour densité la fonction

$$f(y) = \frac{1}{B(n, p)} \frac{y^{n-1}}{(1+y)^{n+p}} \mathbf{1}_{]0,+\infty[}(y).$$

### 5.4.6 Simulation d'une loi normale

Parfois, on ne connaît les v.a. qu'à travers un changement de variable et si une v.a. s'exprime comme une fonction de v.a. indépendantes il peut être plus simple de simuler ces v.a. et de reconstituer la v.a. de départ à travers une fonction que de chercher à la simuler directement. La loi normale est un exemple de ce principe.

### Proposition 5.7

Soient  $U$  et  $V$  deux v.a. indépendantes de loi uniforme sur  $]0, 1[$ . On pose

$$X = \sqrt{-2 \ln U} \cos 2\pi V, \quad Y = \sqrt{-2 \ln U} \sin 2\pi V.$$

Alors  $X$  et  $Y$  sont des v.a. indépendantes de loi  $\mathcal{N}(0, 1)$ .

Générer une v.a.  $\mathcal{N}(0, 1)$  nécessite donc de générer deux v.a. indépendantes de loi uniforme sur  $]0, 1[$  puis d'utiliser l'une des deux formules proposées.

## 5.5 Méthode de Monte Carlo pour le calcul des intégrales et des moments.

La simulation aléatoire permet de faire des calculs d'intégrales. Cette méthode, appelée la méthode de Monte Carlo, repose sur la loi des grands nombres : sous de bonnes hypothèses, on peut approcher presque sûrement l'espérance d'une variable aléatoire (une intégrale dans le cas à densité) par la moyenne arithmétique de variables aléatoires de même espérance. L'idée de la méthode est alors de simuler ces v.a. et de considérer la réalisation de leur moyenne arithmétique comme une approximation numérique de leur espérance commune.

### 5.5.1 Méthode

On suppose que l'on cherche à calculer numériquement

$$\int_{\mathbb{R}} f(x)dx \text{ pour une fonction } f \text{ intégrable.}$$

On remarque que, par des changements de variable successifs, on peut se ramener au calcul de

$$\int_0^1 g(u)du, \text{ pour une certaine fonction } g.$$

Mais, étant donné une v.a.  $U$  de loi uniforme sur  $]0, 1[$ , par la formule de transfert,

$$\mathbf{E}[g(U)] = \int_0^1 g(u)du.$$

Plus précisément, si  $f$  est intégrable sur  $\mathbb{R}$ , on remarque par exemple que

$$\int_{\mathbb{R}} f(x)dx =$$

puis que

$$\begin{aligned} \int_{\mathbb{R}} f(x)dx &= \\ &= \int_0^1 g(u)du \end{aligned}$$

en posant et

$$g(u) = f\left(\frac{u}{1-u}\right) + f\left(-\frac{u}{1-u}\right) \text{ pour } u \in [0, 1]$$

Soit alors  $(U_n)_{n \geq 1}$  une suite de v.a. indépendantes suivant une loi uniforme sur  $]0, 1[$ . D'après la loi des grands nombres,

$$\frac{1}{n} \sum_{i=1}^n g(U_i) \rightarrow \int_0^1 g(u)du \text{ p.s. quand } n \rightarrow +\infty.$$

Une réalisation de cette somme donnera donc une approximation numérique de l'intégrale.

### 5.5.2 Précision

Par ailleurs, on a les estimations suivantes :

$$\begin{aligned} \mathbf{E}\left[\frac{1}{n} \sum_{i=1}^n g(U_i)\right] &= \mathbf{E}[g(U)] = \int_0^1 g(u)du, \\ \mathbf{Var}\left[\frac{1}{n} \sum_{i=1}^n g(U_i)\right] &= \frac{1}{n} \mathbf{Var}[g(U)]. \end{aligned}$$

Il est important de pouvoir donner une précision sur l'approximation numérique de l'intégrale. C'est ce que l'on fait grâce aux intervalles de confiance qui seront introduits dans la partie "Statistique" du cours. Cela permet de dire que l'approximation est dans un intervalle donné avec une certaine probabilité (confiance). Cet intervalle dépend du nombre de simulations  $n$  disponibles. Plus  $n$  est grand, plus précis sera l'intervalle à confiance fixée.

Nous donnons ici le résultat qui pourra être utilisé dès ce chapitre et compris par la suite :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n g(U_i), \quad S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (g(U_i) - \bar{X}_n)^2$$

et  $u_{1-\alpha/2}$  le quantile d'ordre  $1 - \alpha/2$  de la loi normale centrée réduite. Alors l'intervalle

$$\left[ \bar{X}_n - u_{1-\alpha/2} \frac{S_n}{\sqrt{n}}, \bar{X}_n + u_{1-\alpha/2} \frac{S_n}{\sqrt{n}} \right]$$

est un intervalle de confiance au seuil  $1 - \alpha$ .

### 5.5.3 Intérêts de la méthode

Cette méthode converge assez lentement et repose sur la qualité du générateur de nombres aléatoires dont on dispose mais elle est très robuste. En particulier, en grandes dimensions (dès la dimension 3), elle est plus rapide et plus simple d'utilisation que les intégrateurs numériques.

Une autre utilisation évidente de cette méthode est l'approximation des moments des v.a. En effet, toujours par la loi des grands nombres, l'espérance  $\mathbf{E}[\varphi(X)]$  d'une v.a.  $X$  pour une fonction à valeurs réelles  $\varphi$  telle que  $\varphi(X)$  est intégrable, peut être approchée p.s. quand  $n$  grand par

$$\mathbf{E}[\varphi(X)] \approx \frac{1}{n} \sum_{i=1}^n \varphi(X_i)$$

pour toute suite  $(X_i)_{i \geq 1}$  iid à  $X$ . Même lorsque  $X$  admet une densité  $f$ , cette approximation peut-être beaucoup plus précise à  $n$  petit que le calcul direct de la valeur

$$\int_{\mathbb{R}} \varphi(x) f(x) dx$$

car les points alors choisis par la méthode de Monte Carlo se font selon la densité  $f$  et non uniformément.

## 5.6 Méthode de “simulation Monte Carlo”.

La simulation permet également de trouver la distribution approchée de phénomènes aléatoires complexes. Dans le cas d'une variable aléatoire à densité, par exemple, il s'agit d'obtenir une densité approchée, par exemple par une fonction constante par morceaux. Plus précisément, si  $Y$  admet une densité  $f$ , pour tout intervalle  $I$  de  $\mathbb{R}$ ,

$$\mathbf{P}[Y \in I] = \int_I f(x) dx = \int_{\mathbb{R}} \mathbf{1}_I(x) f(x) dx = \mathbf{E}[\mathbf{1}_I(Y)].$$

Ceci, d'après la section précédente, peut s'approcher en simulant un grand nombre de fois la variable  $Y$  et en faisant la moyenne arithmétique des valeurs des réalisations obtenues, soit en calculant la proportion empirique des réalisations tombées dans l'intervalle  $I$ .

Pour avoir la distribution approchée, on discrétise l'ensemble des valeurs prises par  $Y$  en intervalles  $I$  « assez petits », puis on approche  $\mathbf{P}[Y \in I]$  par la méthode de Monte Carlo pour chaque intervalle  $I$ . On obtient ainsi une approximation discrète de la densité  $f$  de  $Y$  (cf Fig. 4).

Cette méthode est utilisée quand la loi de  $Y$  est connue comme une fonction de variables aléatoires indépendantes, i.e. quand  $Y = \psi(X_1, \dots, X_n)$  où  $X_1, \dots, X_n$  sont des v.a. indépendantes de lois connues et  $\psi$  une fonction connue. Pour simuler  $Y$ , on simule alors chacune des variables  $X_1, \dots, X_n$  puis on applique la fonction  $\psi$  au vecteur trouvé.

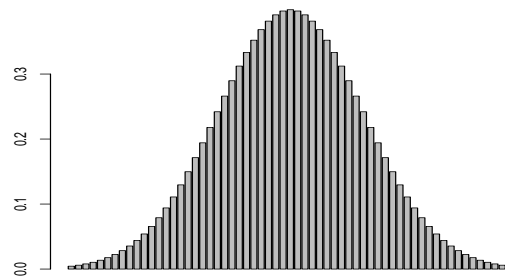


FIGURE 4 – Distribution empirique d'une loi normale

### 5.7 Rééchantillonnage (bootstrap)

Le bootstrap est une méthode qui permet de faire des calculs (de moments, de fonction de répartition d'une v.a.  $X$ ) à partir d'une v.a. dont on ne connaît des réalisations du phénomène aléatoire. Elle consiste à utiliser, comme distribution (inconnue) de la v.a.  $X$  sa distribution empirique (l'histogramme des données). Pour bien comprendre cette méthode, il faut déjà avoir vu la notion d'échantillon qui est traitée dans la partie suivante.

## Deuxième partie

# Statistique inférentielle

## 6 Estimateurs, estimation par intervalle de confiance

### 6.1 Echantillons

#### 6.1.1 Problématique type et vocabulaire de base

On veut estimer les chances d'être élu de deux candidats au second tour d'une élection présidentielle.

On fait un *sondage*, c'est-à-dire que l'on constitue un *échantillon* de  $n$  personnes et on note  $x_1, \dots, x_n$  les intentions de vote des  $n$  individus issus de la *population* des inscrits.

Pour tout  $i \in \{1, \dots, n\}$ , on note  $x_i = 1$  si l'individu  $i$  déclare qu'il votera pour le candidat 1 et  $x_i = 0$  si son vote ira au candidat 2 (pas de vote nul ou blanc). On peut estimer le pourcentage qui sera obtenu par le candidat 1 par

#### 6.1.2 Formalisation.

Si les individus sont choisis selon un *échantillonnage aléatoire simple* (c'est-à-dire uniformément et avec remise) dans la population, on peut considérer chaque  $x_i$  comme la *réalisation* d'une variable aléatoire  $X_i$  suivant une loi de Bernoulli  $\mathcal{B}(p)$ , où  $p$  est la proportion d'individus qui voteront pour le candidat 1 dans la population totale, et que les  $(X_i)_{i=1..n}$  sont indépendantes.

#### 6.1.3 Définition.

##### Définition 6.1

On dit que le vecteur aléatoire  $(X_1, \dots, X_n)$  est un échantillon de la v.a.r.  $X$  s'il est constitué de variables i.i.d. à  $X$  (i.e. indépendantes et de même loi que  $X$ , appelée la variable parente). On appelle réalisation de l'échantillon  $(X_1, \dots, X_n)$  toute observation  $(x_1, \dots, x_n)$ , autrement dit tout  $n$ -uple  $(X_1(\omega), \dots, X_n(\omega))$ , pour un  $\omega \in \Omega$ .

Exemple :

## 6.2 Moments empiriques d'un échantillon.

### 6.2.1 Définition et premières propriétés.

#### Définition 6.2

Soit  $(X_1, \dots, X_n)$  un échantillon d'une v.a.  $X$ .

1. On appelle *moyenne empirique* de l'échantillon la statistique (ou fonction de l'échantillon)

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

2. On appelle variance empirique de l'échantillon la statistique (ou fonction de l'échantillon)

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

Les statistiques  $\bar{X}$  et  $\Sigma^2$  sont ce que l'on appelle des *estimateurs* des moments de la variable parente. La moyenne empirique  $\bar{X}$  est un estimateur de l'espérance de  $X$  et  $\Sigma^2$  est un estimateur de sa variance.

La qualité des estimateurs est mesurée par leur écart avec le paramètre qu'ils estiment :

- le biais est l'écart moyen entre l'estimateur et le paramètre ;
- la précision est l'écart quadratique moyen entre l'estimateur et le paramètre. C'est aussi la somme du carré du biais et de la variance de l'estimateur.

### Proposition 6.1

Soit  $(X_1, \dots, X_n)$  un échantillon d'une v.a.  $X$ .

1. Si  $X$  est intégrable,  $\bar{X}$  est intégrable et  $\mathbf{E}[\bar{X}] = \mathbf{E}[X]$ , et si  $X$  est de carré intégrable,  $\bar{X}$  est de carré intégrable et

$$\mathbf{Var}[\bar{X}] = \frac{\mathbf{Var}[X]}{n}.$$

2.  $\Sigma^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mathbf{E}[X])^2 - (\bar{X} - \mathbf{E}[X])^2$  et, si  $X$  est de carré intégrable,

$$\mathbf{E}[\Sigma^2] = \frac{n-1}{n} \mathbf{Var}[X].$$

Si  $\mu_4 = \mathbf{E}[(X - \mathbf{E}[X])^4]$  est défini, on a

$$\mathbf{Var}[\Sigma^2] = \frac{n-1}{n^3} \left[ (n-1)\mu_4 - (n-3)\mathbf{Var}[X]^2 \right].$$

**Démonstration :** Le (1) et la première partie du (2) ont été vus en TD. La seconde partie du (2) est un calcul fastidieux et est admise.

Cette proposition montre que la moyenne empirique est un estimateur sans biais de l'espérance de  $X$  et que sa variance (et donc sa précision) converge vers 0 quand la taille de l'échantillon tend vers  $+\infty$ . Par ailleurs la variance empirique a un biais qui disparaît quand  $n$  tend vers  $+\infty$  (on dit que cet estimateur est asymptotiquement sans biais) et sa précision converge également vers 0. Plutôt que  $\Sigma^2$ , on utilisera dans la suite l'estimateur sans biais de la variance

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

### Proposition 6.2

(admise). Si  $|X|^3$  est intégrable, on a :

$$\mathbf{cov}[\bar{X}, \Sigma^2] = \frac{n-1}{n^2} \mathbf{E}[(X - \mathbf{E}[X])^3].$$

En particulier,  $\bar{X}$  et  $\Sigma^2$  sont asymptotiquement décorréliées et si la distribution de  $X$  est symétrique alors elles sont décorréliées.

Pour pouvoir utiliser ces estimateurs dans la théorie de l'estimation, il nous faut connaître leur loi ou, tout au moins, leur loi approchée ou asymptotique. L'ensemble des résultats de convergence utilisés dans la suite est résumé dans la proposition suivante.

**Proposition 6.3**

Soit  $(X_i)_{i \geq 1}$  une suite i.i.d. à  $X$ , une v.a.r. de carré intégrable d'espérance  $m$  et de variance  $\sigma^2$ . Pour tout  $n \geq 1$ , on note  $\bar{X}_n$  la moyenne empirique de l'échantillon  $(X_1, \dots, X_n)$  et  $\Sigma_n^2$  sa variance empirique. Alors

1.  $(\bar{X}_n)_{n \geq 1}$  converge presque sûrement et en moyenne quadratique vers  $m$  et

$$\mathcal{L} \left( \frac{\bar{X}_n - m}{\sigma/\sqrt{n}} \right) \rightarrow \mathcal{N}(0, 1) \text{ quand } n \rightarrow +\infty;$$

2. si  $X^2$  est de carré intégrable alors  $(\Sigma_n^2)_{n \geq 1}$  converge presque sûrement et en moyenne quadratique vers  $\sigma^2$  et

$$\mathcal{L} \left( \frac{\Sigma_n^2 - \mathbf{E}[\Sigma_n^2]}{\sqrt{\mathbf{Var}[\Sigma_n^2]}} \right) \rightarrow \mathcal{N}(0, 1) \text{ quand } n \rightarrow +\infty.$$

Démonstration : Applications directe de la loi des grands nombres et du théorème limite central.

**6.2.2 Cas des échantillons gaussiens.**

On suppose que  $X \rightsquigarrow \mathcal{N}(m, \sigma^2)$ .

**Proposition 6.4**

Soit  $(X_1, \dots, X_n)$  un échantillon de taille  $n$  de la v.a.  $X$ . On a les propriétés suivantes :

$$\bar{X} \rightsquigarrow \mathcal{N} \left( m, \frac{\sigma^2}{n} \right), \quad \frac{n\Sigma^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2 \quad \left( \text{ou } \frac{(n-1)S^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2 \right),$$

$\bar{X}$  et  $\Sigma^2$  (ou  $S^2$ ) sont indépendantes et

$$\frac{\bar{X} - m}{\Sigma/\sqrt{n-1}} \rightsquigarrow T_{n-1} \quad \left( \text{ou } \frac{\bar{X} - m}{S/\sqrt{n}} \rightsquigarrow T_{n-1} \right).$$

Dans cette proposition apparaissent deux nouvelles lois que nous allons maintenant étudier.

**6.2.3 Loi du chi-deux.**

Soient  $X_1, \dots, X_n$   $n$  v.a. indépendantes de loi  $\mathcal{N}(0, 1)$ .

**Lemme 6.1**

Soit  $X$  une v.a. suivant une loi  $\mathcal{N}(0, 1)$ . Alors

$$X^2 \rightsquigarrow \text{Gamma} \left( \frac{1}{2}, \frac{1}{2} \right).$$



**Définition 6.3**

La loi de  $X_1^2 + \dots + X_n^2$  est appelée une loi du chi-deux à  $n$  degrés de liberté, notée  $\chi_n^2$ .

En consultant dans « lois usuelle » la partie concernant la somme de v.a. indépendantes suivant une loi Gamma, on se convainc facilement des résultats suivants :

**Proposition 6.5**

$\chi_n^2 = \text{Gamma}(1/2, n/2)$ . Par ailleurs, si  $X$  et  $Y$  sont indépendantes et suivent respectivement des lois du chi-deux à  $n$  et  $p$  degrés de liberté alors  $X + Y$  suit une loi du chi-deux à  $n + p$  degrés de liberté.

**Application : démonstration de la proposition 6.4.**  $X \rightsquigarrow \mathcal{N}(m, \sigma^2)$ . Or

$$\Sigma^2 = \frac{1}{n} \sum_{i=1}^n (X_i - m + m - \bar{X})^2$$

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

En multipliant par  $n/\sigma^2$ , on obtient :

$$\sum_{i=1}^n \left( \frac{X_i - m}{\sigma} \right)^2 =$$

soit

$$\sum_{i=1}^n \left( \frac{X_i - m}{\sigma} \right)^2 =$$

Le théorème dit « de Cochran » (admis — une sorte de réciproque à la construction de la loi du chi-deux) nous permet de conclure ici par : puisque

$$\sum_{i=1}^n \left( \frac{X_i - m}{\sigma} \right)^2 \rightsquigarrow \text{et} \left( \frac{\bar{X} - m}{\sigma/\sqrt{n}} \right)^2 \rightsquigarrow ,$$

$$n \frac{\Sigma^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2 \quad \text{et} \quad \Sigma^2 \text{ et } \bar{X} \text{ indépendants.}$$

**6.2.4 Loi de Student.**

Soient  $X$  et  $Y$  des v.a. indépendantes suivant respectivement des lois  $\mathcal{N}(0, 1)$  et  $\chi_n^2$ .

**Définition 6.4**

La loi de

$$T = \frac{X}{\sqrt{Y/n}}$$

s'appelle la loi de Student à  $n$  degrés de liberté, notée  $T_n$ . Sa densité est la fonction

$$f(x) = \frac{(1 + x^2/n)^{-(n+1)/2}}{\sqrt{n} B(1/2, n/2)} \text{ pour } x \in \mathbb{R}.$$

**Proposition 6.6**

La loi de Student à 1 degré de liberté est la loi de Cauchy de paramètre 1 (qui n'admet aucun moment).

Si  $n > 1$ ,  $\mathbf{E}[T_n] = 0$  et, si  $n > 2$ ,

$$\mathbf{Var}[T_n] = \frac{n}{n-2}.$$

**Application : preuve de la prop. 6.4.** On suppose que  $X \rightsquigarrow \mathcal{N}(m, \sigma^2)$ . Alors

$$\frac{\bar{X} - m}{\sigma/\sqrt{n}} \rightsquigarrow \quad \text{et } n \frac{\Sigma^2}{\sigma^2} \rightsquigarrow$$

et ces variables sont indépendantes. Donc, d'après la définition de la loi de Student,

soit

Remarquer que cette statistique ne dépend pas de  $\sigma$ .

**6.3 Fonction de répartition empirique d'un échantillon**

D'une façon générale, une *statistique d'échantillon* est une fonction d'un échantillon. Nous n'étudions dans ce cours que les plus usuelles.

On peut définir comme statistique d'échantillon pertinente, la fonction de répartition empirique d'un échantillon.

**Définition 6.5**

Soit  $(X_1, \dots, X_n)$  un échantillon d'une v.a.  $X$ . On appelle fonction de répartition empirique de l'échantillon la famille de statistiques définie par

$$F_n^*(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{]-\infty, x]}(X_i) \text{ pour tout } x \in \mathbb{R}.$$

**Théorème 6.1**

Soit  $(X_i)_{i \geq 1}$  une suite de v.a. i.i.d. à une v.a.  $X$ . Alors, pour tout  $x \in \mathbb{R}$ ,

$$\lim_{n \rightarrow +\infty} F_n^*(x) = F_X(x) \text{ presque sûrement.}$$

Démonstration : Soit  $x \in \mathbb{R}$  fixé. Pour tout  $i \in \{1, \dots, n\}$ , on pose

$$Y_i = \mathbf{1}_{]-\infty, x]}(X_i).$$

La loi de  $Y_i$  est : puisque  $Y_i$  prend les valeurs 0 ou 1,  $Y_i \rightsquigarrow \mathcal{B}(p)$  avec  $p = \mathbf{P}[Y_i = 1] = \mathbf{P}[X_i \leq x] = F_{X_i}(x) = F_X(x)$ .

On réécrit alors

$$F_n^*(x) = \frac{Y_1 + \dots + Y_n}{n}$$

et, puisque les  $Y_i$  sont indépendants, on peut appliquer la loi des grands nombres et il vient, quand  $n \rightarrow +\infty$ ,

$$\text{p.s. } F_n^*(x) \rightarrow \mathbf{E}[Y_1] = p = F_X(x).$$

## 6.4 Estimation par intervalle de confiance

### 6.4.1 Introduction et définition.

Jusqu'ici, nous avons proposé une valeur empirique de l'espérance ou de la variance comme réalisation d'un estimateur, choisi « au mieux ». Ceci n'est pas suffisant. En effet, il faut savoir avec certitude l'erreur commise lorsqu'on fait une telle approximation. Si les « fourchettes » d'erreur ne sont pas données, l'estimation ne dit rien. Le problème est de répondre aux questions suivantes : soit  $\theta$  le paramètre à estimer et soit  $t$  la réalisation observée de l'estimateur.

- Etant donné  $\varepsilon > 0$ , avec quelle probabilité a-t-on  $\theta \in [t - \varepsilon, t + \varepsilon]$  ?
- Dans quel intervalle se situe mon paramètre avec probabilité 0,9 ?
- Quelle taille doit avoir mon échantillon si je veux avoir un intervalle de taille  $2\varepsilon$  pour mon paramètre avec probabilité 0,9 ?

Dans la définition qui suit,  $\theta$  est le paramètre à estimer (la moyenne ou la variance d'une variable aléatoire  $X$  dont on dispose d'un échantillon  $X_1, \dots, X_n$ ).

#### Définition 6.6

Soit  $0 < \alpha < 1$ . On appelle intervalle de confiance pour le paramètre  $\theta$  de niveau de confiance  $1 - \alpha$  (ou au seuil de risque  $\alpha$ ) un intervalle  $[L_1; L_2]$  où  $L_1$  et  $L_2$  sont des variables aléatoires telles que  $P(L_1 \leq \theta \leq L_2) = 1 - \alpha$ .

Par construction  $L_1$  et  $L_2$  sont des variables aléatoires car elles dépendent des observations donc des  $X_1, \dots, X_n$ . On note  $l_1$  et  $l_2$  la réalisation de ces v.a. sur l'échantillon. On appelle alors aussi intervalle de confiance au seuil de risque  $\alpha$ , l'intervalle  $[l_1, l_2]$ .

Il est difficile de proposer une méthode systématique pour déterminer les intervalles de confiance. On va ici traiter un exemple puis passer en revue les cas classiques et essayer de comprendre la méthode sans la formaliser.

### 6.4.2 Exemple de construction d'un intervalle de confiance.

**Exemple :** On cherche à estimer la durée de vie moyenne  $m$  d'un pneu jusqu'à l'usure. Plus précisément, on cherche un intervalle de confiance au seuil de risque  $\alpha = 0,05$  pour  $m$ , c'est-à-dire un intervalle  $I \subset \mathbb{R}$  tel que  $m \in I$  à 95%.

Pour cela, on dispose de mesures (i.e. on a réalisé des expériences aléatoires) faites sur  $n = 10$  pneus. On a noté, pour  $i = 1, \dots, 10$ ,  $x_i$  la durée de vie du pneu  $i$  et on a calculé

$$\bar{x} = \frac{x_1 + \dots + x_n}{n} = 20000 \text{ km.}$$

La modélisation de l'expérience conduit, si elle a été réalisée dans les « bonnes » conditions (tirages uniformes indépendants), à admettre que chaque  $x_i$  est la réalisation d'une v.a.  $X_i$  et que  $(X_1, \dots, X_n)$  est un échantillon d'une v.a.  $X$  de moyenne  $m$ . Dans cet exemple précis, on suppose de plus que l'on sait que  $X$  suit une loi normale d'écart-type  $\sigma = 2000 \text{ km}$ .

On cherche des variables aléatoires  $L_1$  et  $L_2$  telles que

$$\mathbf{P}[L_1 \leq m \leq L_2] = 0,95.$$

La v.a.  $\bar{X}$  étant l'estimateur de  $m$ , et la loi de

$$\frac{\bar{X} - m}{\sigma/\sqrt{n}}$$

étant complètement connue et symétrique — c'est une  $\mathcal{N}(0, 1)$  — nous allons construire l'intervalle de confiance avec cet estimateur, c'est-à-dire centré autour de cet estimateur, en choisissant

$$L_1 = \bar{X} - \frac{\sigma}{\sqrt{n}}u_0, \quad L_2 = \bar{X} + \frac{\sigma}{\sqrt{n}}u_0$$

avec  $u_0 > 0$  tel que

$$\mathbf{P}\left[-u_0 \leq \frac{\bar{X} - m}{\sigma/\sqrt{n}} \leq u_0\right] = 0,95.$$

On note

$$X^* = \frac{\bar{X} - m}{\sigma/\sqrt{n}}.$$

Soit  $\phi$  la fonction de répartition de  $X^*$ . On a :

$$\mathbf{P}[-u_0 < X^* < u_0] = \quad .$$

et, par symétrie de  $\phi$ ,  $\phi(-u_0) = \quad$  . Donc, finalement, on cherche  $u_0 > 0$  tel que

$$\Leftrightarrow$$

$$\Leftrightarrow$$

$u_0$  est donc le quantile d'ordre 0.975 de la loi normale centrée réduite.

On inverse la fonction de répartition et, finalement, on trouve

$$\mathbf{P} \left[ \quad < \frac{\bar{X} - m}{\sigma/\sqrt{n}} < \quad \right] = 0,95.$$

Avec les données  $\sigma = 2000$  et  $n = 10$ , on trouve  $\sigma/\sqrt{n} \approx 632$  et  $\sigma u_0/\sqrt{n} \approx 1240$ .

Comme  $\bar{x} = 20000$  est une réalisation de  $\bar{X}$ , on en déduit que  $m \in ]18760, 21240[$  avec probabilité 0,95, ou encore que  $]18760, 21240[$  est un intervalle de confiance de  $m$  au seuil de risque 0,05.

**Remarque :**

1. Etant donnée la situation (recherche d'une norme manifestement minimale), on aurait pu chercher un intervalle monolatère  $[L, +\infty[$ .
2. Dans cet exemple,  $m$  était inconnu et  $\sigma$  connu et on a eu besoin de  $\sigma$  dans notre calcul. Comment faire si on cherche à estimer  $m$  alors que  $\sigma$  est inconnu ? Ou alors pour estimer  $\sigma$  que  $m$  soit connu ou non ? Tout le problème ici est de faire intervenir dans le calcul une statistique dont on connaît la loi, indépendamment des paramètres inconnus. Nous allons pouvoir passer en revue le cas des variables gaussiennes qui a été entièrement explicité.
3. Que se passe-t-il alors dans le cas non gaussien ? C'est le théorème limite central qui nous permettra de conclure dans certains cas.

### 6.4.3 Cas des variables gaussiennes.

On suppose que la variable parente suit une loi gaussienne de paramètres  $m$  et  $\sigma^2$  dont on cherche des intervalles de confiance au seuil de risque donné  $\alpha$ .

**Estimation de la moyenne lorsque  $\sigma$  est connu (cas de l'exemple).** On utilise l'estimateur  $\bar{X}$ . On sait que

$$X^* = \frac{\bar{X} - m}{\sigma/\sqrt{n}} \rightsquigarrow \mathcal{N}(0, 1).$$

Pour tout  $\beta \in ]0, 1[$ , on note  $u_\beta$  le quantile d'ordre  $\beta$  de la loi normale centrée réduite. On a alors

$$\mathbf{P} \left[ u_{\alpha/2} < \frac{\bar{X} - m}{\sigma/\sqrt{n}} < u_{1-\alpha/2} \right] = \phi(u_{1-\alpha/2}) - \phi(u_{\alpha/2}) = 1 - \alpha,$$

on choisira l'intervalle de confiance

$$\left] \bar{X} - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2} \right[$$

puisque  $u_{1-\alpha/2} = -u_{\alpha/2}$ .

**Estimation de la moyenne lorsque  $\sigma$  est inconnu.** On sait que la variable

$$\frac{\bar{X} - m}{S/\sqrt{n}} \rightsquigarrow T_{n-1}.$$

On note à nouveau, pour tout  $\beta \in ]0, 1[$ ,  $t_\beta$  le quantile d'ordre  $\beta$  de la loi de Student à  $n - 1$  degrés de liberté. Alors, comme la densité d'une loi de Student est symétrique,  $t_{1-\alpha/2} = -t_{\alpha/2}$  et donc

$$\mathbf{P} \left[ -t_{1-\alpha/2} < \frac{\bar{X} - m}{S/\sqrt{n}} < t_{1-\alpha/2} \right] = 1 - \alpha.$$

L'intervalle

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

est un intervalle de confiance pour  $m$  au seuil de risque  $\alpha$ .

**Estimation de la variance lorsque  $m$  est connu.** On utilise l'estimateur

$$T = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2.$$

On sait que  $nT/\sigma^2 \rightsquigarrow \chi_n^2$ . Si on note  $x_\beta$  le quantile d'ordre  $\beta$  de la loi du chi-deux à  $n$  degrés de liberté pour tout  $\beta \in ]0, 1[$ , alors

$$\mathbf{P} \left[ x_{\alpha/2} < n \frac{T}{\sigma^2} < x_{1-\alpha/2} \right] = 1 - \alpha$$

et donc un intervalle de confiance pour  $\sigma^2$  au seuil de risque  $\alpha$  est

$$\left[ \frac{nT}{x_{1-\alpha/2}}, \frac{nT}{x_{\alpha/2}} \right].$$

**Estimation de la variance lorsque  $m$  est inconnu.** On utilise l'estimateur  $S^2$ . On a

$$(n-1) \frac{S^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2$$

et on procède comme ci-dessus : on obtient un intervalle de confiance au seuil de risque  $\alpha$

$$\left[ \frac{(n-1)S^2}{x_{1-\alpha/2}}, \frac{(n-1)S^2}{x_{\alpha/2}} \right].$$

#### 6.4.4 Cas de variables non gaussiennes de carré intégrable.

Par le théorème limite central, un intervalle de confiance pour la moyenne peut-être obtenu si  $n$  est assez grand. L'intervalle de confiance au seuil de risque  $\alpha$  a la forme suivante :

$$\left[ \bar{X} - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2} \right]$$

où  $u_{1-\alpha}$  est le quantile d'ordre  $1 - \alpha/2$  de la loi normale centrée réduite.

**Remarque :**

1. Quand la loi est connue et que la variance est une fonction  $g$  de l'espérance, on remplace  $\sigma$  par  $\sqrt{g(\bar{X})}$ . C'est le cas pour la loi de Bernoulli (cf. annexe C.2), de Poisson, exponentielle etc.
2. Quand la loi n'est connue qu'au travers de réalisations de la variable aléatoire, on remplace  $\sigma$  par l'estimateur habituel  $S$ .

## 7 Théorie des tests

### 7.1 Un exemple : les faiseurs de pluie.

**Situation :**

Le niveau naturel des précipitations dans la Beauce en  $mm$  par an suit une loi normale  $\mathcal{N}(600, 100^2)$ .

Les « faiseurs de pluie » prétendent augmenter le niveau moyen annuel des précipitations par insémination des nuages avec de l'iodure d'argent.

Les agriculteurs souhaitent que le niveau augmente d'au moins 50mm par an en moyenne avant de financer ce projet.

Après insémination, on obtient les mesures suivantes :

année	1951	1952	1953	1954	1955
mm	510	614	780	512	501

année	1956	1957	1958	1959
mm	534	603	788	650

On a le choix entre deux hypothèses :

$H_0$  : l'insémination est sans effet,

$H_1$  : l'insémination augmente  
de 50  $mm$  le niveau moyen de pluie.

1. *Le point de vue des agriculteurs* : ils adoptent  $H_0$  et n'acceptent de l'abandonner que si la probabilité de le faire à tort est très faible, mettons  $\alpha \ll 1$ .

On cherche un événement  $A$  qui se produit avec probabilité  $\alpha$  sous l'hypothèse  $H_0$  :

$$\mathbf{P}[A|H_0] = \alpha \text{ et tel que}$$

- (i) connaissant les résultats des essais, on puisse déterminer s'il a été réalisé ou non ;
- (ii) l'événement  $A$  se réalise avec une forte probabilité sous l'hypothèse  $H_1$ .

**Remarque :**

Le test consiste ensuite à vérifier si, lors des essais, cet événement s'est réalisé ou non. Deux cas se produisent :

- Si  $A$  s'est réalisé, alors qu'il avait une probabilité  $\alpha$  très faible de se réaliser sous  $H_0$ , les agriculteurs décideront de rejeter l'hypothèse  $H_0$ , et donc d'accepter  $H_1$  ; ce faisant, ils ont une probabilité  $\alpha$  de se tromper.
- Si  $A$  ne s'est pas réalisé, les agriculteurs décideront alors de conserver l'hypothèse  $H_0$  faute de raisons suffisantes de la rejeter.

**Remarque :**

2. *Le point de vue des faiseurs de pluie*. Leur risque est mesuré différemment : ils redoutent que l'hypothèse  $H_1$  soit rejetée alors qu'elle est bonne. On vient de voir que cette hypothèse sera rejetée par les agriculteurs si  $A$  n'est pas réalisé. Ils calculent donc

$$\beta = \mathbf{P}[A^c|H_1].$$

- Si  $\beta$  est petit, cela signifiera donc que, sous l'hypothèse  $H_1$ , l'événement qui a conduit à rejeter  $H_1$  a très peu de chances de se réaliser, et l'idée que l'hypothèse  $H_1$  n'est pas la bonne est confirmée.
- Si  $\beta$  est grand, cela signifiera que l'événement qui a conduit à rejeter  $H_1$  se réalise avec une forte probabilité sous  $H_1$  et donc que le test que l'on a mis en place n'est pas significatif.

**Formalisation mathématique.** On suppose que,

- $X$  suit une loi  $\mathcal{N}(m, 100^2)$ .
- Les mesures effectuées sont des réalisations d'un échantillon de taille  $n = 9$  de la v.a.  $X$ .

On doit choisir entre

$$H_0 : [m = 600], \quad H_1 : [m \geq 650].$$

#### Etape 1

On fixe le seuil de risque accepté par les agriculteurs.

Par exemple  $\alpha = 0,05$ .

#### Etape 2

On détermine la région de rejet de  $H_0$ ,  $A$ .

On la choisit de la forme

$$A = [(X_1, \dots, X_n) \in W]$$

$$W = \{(x_1, \dots, x_n) \in \mathbb{R}^n / \bar{x} > k\}.$$

Reste à déterminer  $k$ . On a  $A = [\bar{X} > k]$  donc  $k$  doit être tel que  $\mathbf{P}[\bar{X} > k | H_0] = \alpha$ . Or, sous  $H_0$ ,

$$\bar{X} \rightsquigarrow \mathcal{N}\left(600, \frac{100^2}{n}\right) \text{ et donc } \frac{\bar{X} - 600}{100/3} \rightsquigarrow \mathcal{N}(0, 1).$$

Il suffit alors d'inverser la fonction de répartition de la loi  $\mathcal{N}(0, 1)$  : on trouve que

et donc

$$\mathbf{P}\left[\frac{\bar{X} - 600}{100/3} > 1,645 | H_0\right] = 0,05.$$

On choisit donc  $k$  tel que

$$\text{soit } k = 600 + \frac{100}{3} \times 1,645 = 655 \text{ environ.}$$

Finalement  $A = [\bar{X} > 655]$ .

#### Etape 3

On procède au test :  $(x_1, \dots, x_n) \in W$  ?

On trouve  $\bar{x} \approx 610,2$ . Donc  $\bar{x} \leq k$  et  $(x_1, \dots, x_n) \notin W$ ,  $A$  n'est pas réalisé.

Sans complément d'information, les agriculteurs sont donc conduits à conserver  $H_0$ .

#### Etape 4

Calcul de la probabilité  $\beta = \mathbf{P}[A^c | H_1]$  d'avoir conservé  $H_0$  à tort.

Sous  $H_1$ ,  $m \geq 650$  et donc la loi de  $\bar{X}$  dépend de  $m$ . On calcule le pire des cas,

$$\sup_{m \geq 650} \mathbf{P} [\bar{X} \leq k | \mathbf{E} [\bar{X}] = m].$$

Ce sup est obtenu pour  $m = 650$ .

Sous l'hypothèse  $m = 650$ ,  $\frac{\bar{X}-650}{100/3} \rightsquigarrow \mathcal{N}(0, 1)$  et on trouve

$$\begin{aligned} \beta &= \mathbf{P} \left[ \frac{\bar{X} - 650}{100/3} \leq \frac{k - 650}{100/3} | \mathbf{E} [\bar{X}] = 650 \right] \\ &= \mathbf{P} \left[ \frac{\bar{X} - 650}{100/3} \leq 0,15 | \mathbf{E} [\bar{X}] = 650 \right] \approx 0,56. \end{aligned}$$

Le risque de conserver  $H_0$  à tort est donc considérable et les agriculteurs ont peut-être eu tort de le faire. On peut aussi interpréter ce résultat avant même de faire le test sur les données en disant que le test n'est pas très bon puisque l'événement  $A$  dont la réalisation conduit à rejeter  $H_0$  et accepter  $H_1$ , a une probabilité relativement petite  $1 - \beta = 0,44$  de se réaliser sous  $H_1$ .

## 7.2 Notions générales.

On dispose d'observations  $(x_1, \dots, x_n)$  qui sont la réalisation d'un échantillon  $(X_1, \dots, X_n)$  d'une variable aléatoire  $X$  (réelle ou vectorielle). Un test statistique définit une règle de décision pour choisir entre deux hypothèses  $H_0$  et  $H_1$  faites sur la loi de  $X$  au vu des données recueillies.

Les hypothèses  $H_0$  et  $H_1$  ne jouent pas le même rôle, l'hypothèse  $H_0$  est celle à laquelle on tient le plus, qu'on ne veut rejeter qu'avec une faible probabilité de le faire à tort. De plus, pour pouvoir procéder à un test il faut impérativement être capable de faire des calculs sous l'hypothèse  $H_0$ , elle doit donc être suffisamment précise alors que l'hypothèse  $H_1$  peut être relativement vague (la négation de  $H_0$  par exemple). Bien sûr les hypothèses  $H_0$  et  $H_1$  doivent s'exclure mutuellement.

### Construction et utilisation du test :

1. On fixe  $\alpha > 0$  petit, le *risque de première espèce* qui est la probabilité de rejeter  $H_0$  à tort.
2. On détermine une *région de rejet* de  $H_0$ ,  $W \subset \mathbb{R}^n$ , telle que

$$\mathbf{P} [(X_1, \dots, X_n) \in W | H_0] = \alpha.$$

Cette région dépend fortement des hypothèses que l'on considère. En particulier, elle dépend de  $H_1$  en ce sens que l'on souhaite que la probabilité

$$1 - \beta = \mathbf{P} [(X_1, \dots, X_n) \in W | H_1]$$

soit la plus grande possible. En effet,  $1 - \beta$ , qui s'appelle la *puissance du test*, mesure la probabilité que les données soient dans la région de rejet de  $H_0$  lorsque  $H_1$  est vraie.

Cette région de rejet est en règle générale construite à partir d'une *variable de décision* qui dépend de la nature du test à effectuer. Dans le cas de l'exemple des faiseurs de pluie, le test portait sur l'espérance de la variable parente de l'échantillon et la variable de décision était donc celle que l'on utilise pour construire un intervalle de confiance pour l'espérance (ici dans le cas gaussien à variance connue)

$$\frac{\bar{X} - m}{\sigma/\sqrt{n}}.$$



A partir de cette variable de décision, on a construit une zone de rejet à tort comme le complémentaire d'un intervalle de confiance pour le paramètre (sous l'hypothèse  $H_0$ , le paramètre a seulement une probabilité  $\alpha$  de ne pas se trouver dans l'intervalle de confiance). Se donner une zone de rejet consistera donc à se donner une variable de décision  $D$  et le complémentaire d'un intervalle de confiance  $I$  pour cette variable au seuil de risque  $\alpha$ .

3. Règle de décision : si la réalisation  $(x_1, \dots, x_n)$  de notre échantillon est dans  $W$ , on rejette  $H_0$  ; sinon, on conserve  $H_0$ . On fait ce test sur la réalisation  $d$  de la variable de décision  $D$  : si  $d$  est dans l'intervalle de confiance  $I$  on ne rejette pas  $H_0$  ; si  $d \notin I$ , on rejette  $H_0$ .

Finalement, construire un test, c'est se donner les hypothèses  $H_0$  et  $H_1$ , le seuil de risque  $\alpha$  petit, la variable de décision  $D$  et l'intervalle de confiance  $I$  et, si on peut la calculer, la puissance du test  $1 - \beta$ .

**Remarque :** Le paramètre

$$\beta = \mathbf{P}[(X_1, \dots, X_n) \in W^c | H_1]$$

s'appelle le *risque de seconde espèce* ; c'est la probabilité de conserver  $H_0$  alors que  $H_1$  est vraie. Ce risque doit être aussi petit que possible à  $\alpha$  fixé.

**Remarque :** Heuristiquement, il est assez facile de se convaincre que, lorsqu'on diminue  $\alpha$ , on diminue la taille de la région de rejet  $W$  et donc on diminue également la puissance du test (ou on augmente le risque de seconde espèce). Par conséquent, on ne peut choisir  $\alpha$  trop petit (les valeurs usuelles de  $\alpha$  sont , 0.1 ou 0.05 ou 0.01).

**Qualité d'un test :** On a vu que la qualité d'un test est mesurée par sa puissance  $1 - \beta$ . D'autre part :

- si  $1 - \beta > \alpha$ , on dit que le test est sans biais.
- si  $1 - \beta \rightarrow 1$  lorsque la taille de l'échantillon  $n$  tend vers l'infini, on dit que le test est convergent.

**Classification des tests.** On dit que le test est *paramétrique* lorsque les hypothèses portent sur la valeur d'un ou plusieurs paramètres de la loi de  $X$ . Si un même test convient pour différentes lois, on dit que le test est *robuste* (comme les tests de moyenne, par exemple). Parmi les tests *non paramétriques* (qui sont robustes), on trouve les *tests d'ajustement* à une loi donnée. Il existe également des tests de *comparaison de plusieurs échantillons* qui permettent de déterminer si des échantillons sont issus d'une même population. Enfin, on verra comment tester si deux variables aléatoires sont indépendantes.

### 7.3 Tests paramétriques.

On fait une hypothèse sur le paramètre  $\theta$  (l'espérance ou la variance) de la loi d'une v.a.  $X$ . On dispose de la réalisation d'un échantillon  $(X_1, \dots, X_n)$  de la v.a.  $X$ .

Les hypothèses que l'on peut formuler sont de deux types :

- hypothèse simple :  $[\theta = \theta_0]$  où  $\theta_0 \in \mathbb{R}^d$  est une valeur fixée du paramètre ;
- hypothèse composite :  $[\theta \in B]$  où  $B$  est une partie de  $\mathbb{R}^d$  non réduite à un point.

Noter que, lorsque le paramètre est réel, une hypothèse composite a souvent la forme  $[\theta < \theta_0]$ ,  $[\theta > \theta_0]$  ou  $[\theta \neq \theta_0]$  pour une valeur fixée  $\theta_0$  du paramètre.

**Remarque :** Pour nous, l'hypothèse  $H_0$  sera toujours une hypothèse simple, pour pouvoir faire tous les calculs.

#### 7.3.1 Test entre deux hypothèses simples.

On suppose  $\theta_0 \neq \theta_1$  et

$$H_0 : [\theta = \theta_0], \quad H_1 : [\theta = \theta_1].$$

La variable de décision est la variable qui sert à construire un intervalle de confiance pour le paramètre  $\theta$ , comme cela a été fait chapitre 6. Ces variables de décision sont rappelées en annexe D.

On suppose en toute généralité que  $D$  est un estimateur de  $\theta$ . Alors

- si  $\theta_1 > \theta_0$ , la zone de rejet a la forme  $[D > d]$  ;
- si  $\theta_1 < \theta_0$ , la zone de rejet a la forme  $[D < d]$ .

**Remarque :** On peut montrer que les tests ainsi construits sont les meilleurs possibles au sens que la zone de rejet construite est celle qui, parmi toutes celles de probabilité  $\alpha$  sous  $H_0$ , a la plus forte probabilité sous  $H_1$ .

### 7.3.2 Test d'une hypothèse simple contre une hypothèse composite : la fonction puissance.

On suppose

$$H_0 : [\theta = \theta_0], \quad H_1 : [\theta \in B]$$

où  $\theta_0$  est une valeur fixée du paramètre et  $B$  une partie de  $\mathbb{R}$  ne contenant pas  $\theta_0$ .

Même si l'on connaît la loi de la variable parente  $X$ , on ne peut calculer la puissance d'un test car  $H_1$  n'est pas assez précise. Par contre, pour tout  $\theta_1 \in B$ , on peut calculer la puissance d'un test pour les hypothèses

$$H_0 : [\theta = \theta_0], \quad H_1 : [\theta = \theta_1].$$

On appelle alors *fonction puissance du test* la fonction, définie sur  $B$  par  $\theta_1 \in B \mapsto 1 - \beta(\theta_1)$ . On recherche alors le test *uniformément le plus puissant* (UPP en abrégé), c'est-à-dire, s'il existe, celui tel que, pour tout  $\theta_1 \in B$ , sa puissance en  $\theta_1$  est supérieure à celle de tout autre test.

Lorsque  $H_1$  a la forme  $[\theta > \theta_1]$  avec  $\theta_1 \geq \theta_0$  ou  $[\theta < \theta_1]$  avec  $\theta_1 \leq \theta_0$ , on peut démontrer que les tests utilisés dans la section précédente sont UPP. Ce sont donc ceux-là que l'on utilisera.

Lorsque  $H_1$  est de la forme  $[\theta \neq \theta_0]$ , on utilise encore ces tests de la façon suivante : on construit la région de rejet de la forme  $[D < d_1] \cup [D > d_2]$  de sorte que

$$\mathbf{P}[D < d_1] = \mathbf{P}[D > d_2] = \frac{\alpha}{2}.$$

**Exemple :** *Test bilatéral :* Soit  $(X_1, \dots, X_n)$  un échantillon de taille  $n$  d'une v.a.  $X$  suivant une loi normale de paramètres  $m$  et  $\sigma^2$  inconnus. On veut tester les hypothèses :

$$H_0 : [m = m_0], \quad H_1 : [m \neq m_0]$$

avec un risque de première espèce  $\alpha$ .

**Sous l'hypothèse  $H_0$ ,**

On choisit la région critique de la forme

$$A = [|T| > k]$$

où  $k$  est choisi de sorte que

En utilisant la symétrie de la loi de Student et en inversant sa fonction de répartition, on trouve, pour  $n = 15$  et  $\alpha = 0,1$  (par exemple),  $k =$  .

Le test est donc le suivant : on calcule  $t$  avec les données observées pour  $m_0$  donné.

Si , on rejette  $H_0$  (avec probabilité  $\alpha$  de se tromper) ; sinon, on accepte  $H_0$ .

## 7.4 Tests d'ajustement.

On dispose de la réalisation d'un échantillon  $(X_1, \dots, X_n)$  d'une v.a.r.  $X$  et on souhaite déterminer la loi de  $X$ . La première étape consiste à « deviner » une loi possible pour  $X$ , en regardant l'histogramme des fréquences constitué par la réalisation de notre échantillon par exemple. On construit alors un test pour savoir si  $X$  suit ou non la loi que l'on a devinée, mettons  $\mathcal{L}$  ; autrement dit, on pose

$$H_0 : [X \text{ suit la loi } \mathcal{L}],$$

$$H_1 : [X \text{ ne suit pas la loi } \mathcal{L}].$$

### 7.4.1 Test d'ajustement du Chi-deux.

**Exemple :** Test d'ajustement du Chi-deux à une loi  $\mathcal{G}(1/2)$ .

On veut vérifier l'ajustement de la « loi du premier succès », lorsque l'on répète indéfiniment des tirages d'une pièce de monnaie, à la loi  $\mathcal{G}(1/2)$ .

On réalise 50 fois l'expérience, c'est-à-dire que 50 fois de suite, on lance la pièce de monnaie jusqu'à obtenir « pile » et on note le rang d'arrivée de ce « pile ».

Les 50 résultats sont : 1 1 3 2 3 1 1 1 4 3 2 7 1 2 1 1 2 4 2 1 1 1 2 2 5 2 1 1 1 1 3 2 2 1 1 1 1 1 6 1 3 1 1 3 2 1 1 2 1 4.  
On représente ces résultats par

On suppose que ces résultats sont les réalisations d'un échantillon de taille 50 d'une v.a.  $X$  et on teste les hypothèses

$$H_0 : [X \rightsquigarrow \mathcal{G}(1/2)] \quad H_1 : \overline{H_0}.$$

Pour cela, on commence par faire une partition en classe des valeurs prises par la v.a.  $X$ . Ici,  $X$  prend ses valeurs dans  $\mathbb{N}^*$ . On choisit ces classes de sorte que l'*effectif empirique des classes*, c'est-à-dire le nombre d'observations de chaque classe ne soit pas trop petit. Par exemple, ici, on choisit les classes  $C_1 =$ ,  $C_2 =$ ,  $C_3 =$ ,  $C_4 =$ .

Les effectifs empiriques de ces classes sont donc les  $n_j$  dans le tableau suivant :

$j$	$n_j$		
1	26		
2	12		
3	6		
4	6		
totaux	n=50		

Le but du test est d'étudier la différence entre ces effectifs empiriques et les *effectifs théoriques* des classes. En effet, sous l'hypothèse  $H_0$ ,  $X$  suit une loi  $\mathcal{G}(1/2)$  et, sous cette hypothèse, l'effectif que devrait avoir une classe est le nombre total d'observations (la taille de l'échantillon) multiplié par la probabilité qu'une observation soit dans la classe.

Calculons, pour chaque classe, la probabilité d'être dans chacune des classes sous  $H_0$  : pour tout  $k \in \mathbb{N}^*$ ,  $\mathbf{P}[X = k] = (1 - 1/2)^{k-1} 1/2 = (1/2)^k$  donc, si on note  $p_j = \mathbf{P}[X \in C_j]$  pour  $j = 1$  à 4, on a

Finalement, les effectifs théoriques des classes sont donc les  $np_j$ , pour  $j = 1$  à 4, que l'on peut mettre dans le tableau.

Avant de passer au test proprement dit, formalisons ce que nous venons de faire.

Soit  $X(\Omega)$  l'ensemble des valeurs prises par  $X$  sous l'hypothèse  $H_0$ . On choisit une partition  $C_1, \dots, C_J$  de  $X(\Omega)$ , chaque  $C_j$ , pour  $j \in \{1, \dots, J\}$ , étant appelé une *classe* (voir plus loin les commentaires sur le choix des classes). On définit alors les variables aléatoires  $N_1, \dots, N_J$ , *effectifs empiriques des classes*, comme les nombres de v.a. de l'échantillon appartenant aux classes  $C_1, \dots, C_J$  respectivement.

On peut donc calculer ces effectifs par les formules : pour tout  $j \in \{1, \dots, J\}$

$$N_j = \sum_{i=1}^n \mathbf{1}_{C_j}(X_i)$$

ou bien de façon équivalente

$$N_j = \text{Card} \{i \in \{1, \dots, n\} / X_i \in C_j\},$$

et les  $n_j$  du tableau sont les réalisations des  $N_j$ .

On note  $p_j = \mathbf{P}[X \in C_j]$  pour tout  $j \in \{1, \dots, J\}$ . Alors  $p_j$  est la proportion *théorique* de résultat que l'on doit trouver dans la classe  $j$ . On appelle alors *effectif théorique de la classe  $C_j$*  la quantité  $np_j$ .

**Remarque :** Le  $J$ -uplet  $(N_1, \dots, N_J)$  suit une loi multinomiale de paramètres  $(n, p_1, \dots, p_J)$ . En particulier, chaque  $N_j$  suit une loi binomiale de paramètres  $(n, p_j)$ . L'effectif théorique de la classe  $j$  est l'effectif théorique **moyen** de la classe, soit  $\mathbf{E}[N_j] = np_j$ .

$j$	$n_j$		.
1	26		
2	12		
3	6		
4	6		
totaux	n=50		

### Théorème 7.1

Soit  $D^2$  la variable aléatoire définie par

$$D^2 = \sum_{j=1}^J \frac{(N_j - np_j)^2}{np_j}.$$

Alors, sous l'hypothèse  $H_0$ , lorsque la taille de l'échantillon  $n$  tend vers l'infini,  $D^2$  converge en loi vers une variable du chi-deux à  $J - 1$  degrés de liberté.

### Test d'ajustement du chi-deux

- On admet que  $D^2$  suit approximativement une loi du Chi-deux à  $J - 1$  degrés de liberté.
- On fixe le risque de première espèce  $\alpha$  petit.
- La région de rejet de  $H_0$  est choisie de la forme  $[D^2 > c]$  où  $c$  est à déterminer en fonction de  $\alpha$  dans en inversant la fonction de répartition de la loi du Chi-deux à  $J - 1$  degrés de liberté.
- Pour appliquer le test, il suffit alors de calculer la réalisation de la variable  $D^2$  que l'on obtient avec nos données et de constater si elle se trouve ou non dans la région de rejet de  $H_0$ .

**Exemple :** (suite).  $J = 4$ .  $D^2$  suit approximativement une loi du chi-deux à 3 degrés de liberté. On se fixe un seuil de risque (par exemple  $\alpha = 0.01$ ).

En inversant la fonction de répartition d'une loi du Chi-deux à 3 degrés de liberté, on trouve que, sous  $H_0$ ,  $\mathbf{P}[D^2 > c] = 0.01$ . On rejettera  $H_0$  si notre réalisation de  $D^2$  est supérieure à ce seuil. On calcule donc la réalisation de  $D^2$ ,

$$d^2 = \sum_{j=1}^4 \frac{(n_j - np_j)^2}{np_j}.$$

$j$	$n_j$	$np_j$	
1	26	25	
2	12	12,5	
3	6	6,25	
4	6	6,25	
totaux	n=50	n=50	

On fait le calcul à partir du tableau, et on trouve  $d^2 = 1/12,5 < 11,345$ , donc on conserve  $H_0$  (i.e. la pièce utilisée est équilibrée et les lancers ont été faits indépendamment).

**Remarque :** L'approximation que l'on fait en supposant que  $D^2$  suit une loi du Chi-deux à  $J - 1$  degrés de liberté n'est admise que si les effectifs empiriques et théoriques des classes "ne sont pas trop petits". Le seuil fixé dépend largement des auteurs de traités de statistique. Nous demanderons que les réalisations des  $N_j$  et que les  $np_j$  soient supérieurs à 5. Dans le cas contraire, on modifiera les classes en les regroupant pour obtenir ces conditions.

#### 7.4.2 Test d'ajustement du Chi-deux avec estimation de paramètres.

Très souvent, on veut pouvoir ne spécifier, dans l'hypothèse  $H_0$ , que la loi de  $X$  et non les paramètres de cette loi, que l'on ignore a priori et que l'on ne peut qu'estimer. Le test vise donc à choisir entre les hypothèses

$$H_0 : [X \rightsquigarrow \mathcal{L}(t)], \quad H_1 : \overline{H}_0,$$

où  $t$  est la réalisation observée d'un estimateur  $T$  du paramètre  $\theta \in \mathbb{R}^p$  de la loi  $\mathcal{L}$ .

On procède exactement comme ci-dessus, en utilisant cette fois-ci le théorème :

##### **Théorème 7.2**

*Sous l'hypothèse  $H_0$ , lorsque la taille de l'échantillon  $n$  tend vers l'infini,  $D^2$  converge en loi vers une variable du chi-deux à  $J - 1 - p$  degrés de liberté ( $p$  est le nombre de paramètres estimés).*

#### 7.4.3 Test d'ajustement de Kolmogorov-Smirnov.

Lorsque la loi est entièrement spécifiée, on peut faire des tests d'ajustement qui reposent sur la convergence de la fonction de répartition empirique d'un échantillon vers la fonction de répartition de la variable parente de l'échantillon. Nous décrivons ici le test de Kolmogorov-Smirnov qui, comme le test de Cramer-von Mises, s'applique dans le cas de lois dont les fonctions de répartition sont continues.

##### **Théorème 7.3**

*Sous les hypothèses du théorème 6.1, la vitesse de convergence de  $F_n^*$  vers  $F_X$  est précisée par, pour tout  $y > 0$ ,*

$$\begin{aligned} \lim_{n \rightarrow +\infty} \mathbf{P} \left[ \sqrt{n} \sup_{x \in \mathbb{R}} |F_n^*(x) - F_X(x)| \leq y \right] \\ = K(y) = \sum_{k \in \mathbb{Z}} (-1)^k e^{-2k^2 y^2}. \end{aligned}$$

Pour le test d'ajustement de Kolmogorov-Smirnov, on procède comme suit : on veut tester l'hypothèse selon laquelle  $X$  suit une loi  $\mathcal{L}$  de fonction de répartition  $F$ . On teste toujours

$$H_0 : [X \text{ suit la loi } \mathcal{L}], \quad H_1 : \overline{H}_0.$$

On fixe un seuil de risque  $\alpha$  petit. Sous l'hypothèse  $H_0$ , la variable aléatoire

$$\sqrt{n}D_n = \sqrt{n} \sup_{x \in \mathbb{R}} |F_n^*(x) - F(x)|$$

admet approximativement comme fonction de répartition la fonction  $K$ . On cherche une région de rejet de la forme

$$[\sqrt{n}D_n > y]$$

en déterminant  $y$  en fonction de  $\alpha$  et  $n$  (la fonction de répartition inverse de  $D_n$  n'est pas donnée dans les logiciels de statistique en général, on utilise alors des tables comme celle donnée à la fin du Saporta). On procède ensuite au test en calculant la réalisation de  $\sqrt{n}D_n$  obtenue avec nos données.

## 7.5 Tests de comparaison entre échantillons indépendants.

On dispose de  $m$  échantillons indépendants d'une certaine variable  $X$  et on désire savoir si ces échantillons proviennent d'une même population. On peut reformuler le problème de la façon suivante : chaque échantillon est une suite finie i.i.d. d'une variable parente  $X^k$  et le problème est de savoir si les  $X^k$  ont même loi.

### 7.5.1 Tests paramétriques pour la comparaison de deux échantillons.

Lorsque l'on a deux échantillons indépendants ( $m = 2$ ), on peut chercher à savoir si les variables parentes ont même espérance et même variance.

On suppose que les variables parentes des deux échantillons,  $X^1$  et  $X^2$ , admettent des moments d'ordre 2 et on note

$$m_k = \mathbf{E} [X^k] \text{ et } \sigma_k^2 = \mathbf{Var} [X^k] \text{ pour } k = 1, 2.$$

On note  $(X_1^1, \dots, X_{n_1}^1)$  l'échantillon de la v.a.  $X^1$  et  $(X_1^2, \dots, X_{n_2}^2)$  celui de la v.a.  $X^2$ .

**Cas des échantillons gaussiens.** On suppose que  $X^1$  et  $X^2$  suivent des lois gaussiennes. On commence par comparer les variances et, si elles ne sont pas significativement différentes, on comparera les moyennes sous l'hypothèse que les variances sont égales.

Pour comparer les variances, on utilise le résultat suivant :

#### Théorème 7.4

Soient  $X$  et  $Y$  deux v.a. indépendantes suivant respectivement des loi du Chi-deux à  $n$  et  $p$  degrés de liberté. Alors

$$F = \frac{X/n}{Y/p}$$

suit une loi de Fisher-Snedecor à  $(n, p)$  degrés de liberté notée  $F_{n,p}$ .

On applique ce résultat en procédant au *test de Fisher-Snedecor* suivant : on pose

$$H_0 : [\sigma_1 = \sigma_2], \quad H_1 : [\sigma_1 \neq \sigma_2];$$

on choisit un seuil de risque  $\alpha$  petit. Si on note  $S_1^2$  et  $S_2^2$  les variances empiriques (sans biais) des échantillons, alors,

$$(n_k - 1) \frac{S_k^2}{\sigma_k^2} \rightsquigarrow \quad \text{pour } k = 1, 2.$$

et donc, sous l'hypothèse  $H_0$ ,

$$= S_1^2 / S_2^2 \rightsquigarrow$$

et de même

$$S_2^2 / S_1^2 \rightsquigarrow .$$

On note alors  $F$  la variable

$$F = S_1^2 / S_2^2 \text{ si } s_1^2 / s_2^2 \geq 1 \text{ et } F = S_2^2 / S_1^2 \text{ sinon,}$$

où  $s_k^2$  est la réalisation de la variable  $S_k^2$  pour  $k = 1, 2$ . On cherche alors la région de rejet sous la forme

$$[F > f_\alpha]$$

où  $f_\alpha$  est à déterminer en fonction de  $\alpha$  avec la fonction de répartition de la loi de Fisher-Snedecor (cf. exemple en TD).

Si le test précédent n'a pas conduit à rejeter l'hypothèse  $\sigma_1 = \sigma_2$ , on procède au test de Student sur les moyennes comme suit : on suppose  $\sigma = \sigma_1 = \sigma_2$  (inconnue) et on teste les hypothèses :

$$H_0 : [m_1 = m_2], \quad H_1 : [m_1 \neq m_2]$$

en choisissant un seuil de risque  $\alpha$  petit.

On note  $\bar{X}_k$  la moyenne empirique de l'échantillon  $k$ . Sous l'hypothèse  $H_0$ ,

$$\bar{X}_1 - \bar{X}_2 \rightsquigarrow$$

et donc

soit

$$T = \frac{(\bar{X}_1 - \bar{X}_2)\sqrt{n_1 + n_2 - 2}}{\sqrt{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}\sqrt{1/n_1 + 1/n_2}} \rightsquigarrow T_{n_1 + n_2 - 2}.$$

On choisit lors la région de rejet de la forme

$$[|T| > t_{\alpha/2}].$$

**Remarque :** Lorsque  $\sigma_1 \neq \sigma_2$  et que les échantillons sont suffisamment grands (i.e. quelques dizaines d'observations), on peut encore appliquer le test de Student.

#### Cas des échantillons non gaussiens.

Dans ce cas, le test de Fisher-Snedecor ne peut plus s'appliquer, mais on peut encore appliquer le test de comparaison des moyennes si les échantillons sont assez grands, en remplaçant la loi de Student par la loi normale.

#### 7.5.2 Test non paramétrique de comparaison de deux échantillons ou plus : le test du Chi-deux.

On dispose de  $m$  échantillons de v.a.  $X^1, \dots, X^m$ . Comme pour le test d'ajustement du Chi-deux, on partage en  $J$  classes l'ensemble des valeurs prises par ces variables aléatoires. Pour tout  $k \in \{1, \dots, m\}$  et pour tout  $j \in \{1, \dots, J\}$ , on note  $N_{kj}$  le nombre de réalisations de l'échantillon  $k$  qui sont dans la classe  $C_j$ .

On pose

$$N_{.j} = \sum_{k=1}^m N_{kj} \text{ l'effectif empirique de la classe } j \in \{1, \dots, J\},$$

$$N_{k.} = \sum_{j=1}^J N_{kj} = n_k \text{ la taille de l'échantillon } k \in \{1, \dots, m\} \text{ et}$$

$$N = \sum_{k=1}^m \sum_{j=1}^J N_{kj} = n \text{ le nombre total d'observations.}$$

On pose enfin

$$D_0^2 = \sum_{k=1}^m \sum_{j=1}^J \frac{(N_{kj} - N_{k.}N_{.j}/N)^2}{N_{k.}N_{.j}/N}.$$

On peut montrer que, sous l'hypothèse ( $H_0$ ) que les échantillons proviennent d'une même population et sont indépendants,  $D_0^2$  suit approximativement une loi du Chi-deux à

$$(m-1)(J-1)$$

degrés de liberté. On procède alors comme dans le test d'ajustement du Chi-deux avec les hypothèses  $H_0$  et  $H_1 = \bar{H}_0$ .

**Exemple :** On veut comparer 4 générateurs de nombres aléatoires entre 1 et 13. Pour cela, on fait 100 tirages aléatoires. Les résultats obtenus sont :

**gén. 1 :** 1 4 2 5 11 4 5 2 4 4 3 13 5 13 4 10 12 12 4 1 3 12 12 9 5 6 12 13 8,

**gén. 2 :** 7 11 13 5 2 2 4 4 1 12 13 11 6 1 9 13 12 12 12 7 5,

**gén. 3 :** 13 13 11 5 10 13 13 10 8 12 3 4 5 3 4 4 6 6 10 2 3 5 2 3 13 13 4 9 1,

**gén. 4 :** 2 1 3 3 11 10 12 8 4 7 2 11 13 1 10 1 12 12 4 2 12.

On choisit de partitionner  $\{1, \dots, 13\}$  en les classes

$$C_1 = \{1, 2, 3, 4, 5\}, C_2 = \{6, 7, 8, 9\}, C_3 = \{10, 11, 12, 13\}.$$

On dispose les résultats dans un tableau :

	$C_1$	$C_2$	$C_3$	Total
$G_1$	16 = $n_{1j}$	3	10	<u>29</u> = $n_{k.}$
$G_2$	8	4	9	<u>21</u>
$G_3$	14	4	11	<u>29</u>
$G_4$	10	2	9	<u>21</u>
Total	<u>48</u> = $n_{.j}$	<u>13</u>	<u>39</u>	

Sous l'hypothèse  $H_0$  selon laquelle les générateurs sont identiques et indépendants, on sait que  $D_0^2$  suit une loi du chi-deux à  $3 \times 2 = 6$  degrés de liberté et on applique le test du chi-deux habituel avec un risque de première espèce  $\alpha$  à choisir.

### 7.5.3 Test d'indépendance du chi-deux.

**Remarque :** Le test non paramétrique de comparaison d'échantillons indépendants s'applique également pour établir l'indépendance de deux variables aléatoires.

En effet, supposons que l'on observe un échantillon

$$((X_1, Y_1), \dots, (X_n, Y_n))$$

d'un couple de v.a.  $(X, Y)$ . L'ensemble des valeurs prises par  $X$  est partitionné en les classes  $C_i, i = 1$  à  $I$  et celui des valeurs prises par  $Y$  en les classes  $C^j, j = 1$  à  $J$ .

On note  $C_{ij} = C_i \times C^j$  alors

$$\{C_{ij}, i = 1..I, j = 1..J\}$$

forme une partition de l'ensemble des valeurs prises par  $(X, Y)$ . On note alors  $N_{ij}$  l'effectif de la classe  $C_{ij}$ .

On pose

$$H_0 : [X \text{ et } Y \text{ sont indépendantes}], \quad H_1 : \overline{H}_0.$$

Sous l'hypothèse  $H_0$ , la variable

$$D_0^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(N_{ij} - N_{i.}N_{.j}/n)^2}{N_{i.}N_{.j}/n} \rightsquigarrow \chi_{(I-1)(J-1)}^2$$

et on procède à un test du chi-deux.

#### Justification



## Annexes

### A Éléments de théorie des ensembles et de théorie de la mesure

#### A.1 Ensembles dénombrables

##### Définition A.1

Un ensemble  $D$  est dit dénombrable si l'on peut compter ses éléments, c'est-à-dire s'il existe une application  $\varphi : D \rightarrow \mathbb{N}$  injective (un même numéro ne peut correspondre à deux éléments distincts).

##### Proposition A.1

- Les ensembles  $\mathbb{N}$ ,  $\mathbb{Z}$  sont dénombrables.
- Tout sous-ensemble d'un ensemble dénombrable est dénombrable.
- Le produit cartésien fini d'ensembles dénombrables est dénombrable.
- $\mathbb{Q}$  est dénombrable.
- L'union dénombrable d'ensembles dénombrables est dénombrable.

#### A.2 Ensembles négligeables. Intégrale des fonctions nulles presque partout.

##### Définition A.2

Un ensemble  $N \subset \mathbb{R}$  est dit négligeable de  $\mathbb{R}$  si, pour tout  $\varepsilon > 0$ , il existe une famille dénombrable d'intervalles de  $\mathbb{R}$  recouvrant  $N$  et de somme des longueurs inférieure à  $\varepsilon$ , c'est-à-dire s'il existe une suite  $(I_n^\varepsilon)_{n \geq 1}$  d'intervalles de  $\mathbb{R}$  satisfaisant :

$$N \subset \bigcup_{n \geq 1} I_n^\varepsilon \quad \text{et} \quad \sum_{n \geq 1} |I_n^\varepsilon| < \varepsilon,$$

où  $|I_n^\varepsilon|$  désigne la longueur de l'intervalle  $I_n^\varepsilon$ .

##### Remarque :

- Un ensemble ouvert ne peut donc pas être négligeable puisqu'il contient un intervalle de longueur strictement positive.
- Toute union dénombrable de négligeables est négligeable.

##### Définition A.3

Un ensemble  $N \subset \mathbb{R}^n$  ( $n \geq 1$ ) est dit négligeable de  $\mathbb{R}^n$  si, pour tout  $\varepsilon > 0$ , il existe une famille dénombrable de pavés de  $\mathbb{R}^n$  recouvrant  $N$  et de somme des mesures inférieure à  $\varepsilon$ , avec la mesure d'un pavé prise comme le produit des longueurs de ses côtés.

##### Remarque :

- Les remarques ci-dessus s'appliquent à la dimension  $n$ .
- Les droites dans  $\mathbb{R}^2$ , les plans dans  $\mathbb{R}^3$  etc. sont négligeables.

##### Proposition A.2

Soit  $f$  une fonction définie sur  $\mathbb{R}^n$ , nulle en dehors d'un négligeable (on dit aussi nulle presque partout). Alors l'intégrale de Lebesgue de  $f$  est nulle. Réciproquement, si  $f$ , positive (respectivement négative) presque partout, est d'intégrale nulle, alors  $f$  est nulle presque partout.

#### A.3 Mesure de Lebesgue des ensembles de $\mathbb{R}^n$ .

On peut montrer que les notions de longueur (dans  $\mathbb{R}$ ) d'aire (dans  $\mathbb{R}^2$ ) et de volume (dans  $\mathbb{R}^3$ ) peuvent se généraliser à toutes dimensions. La mesure ainsi obtenue s'appelle la mesure de Lebesgue. Elle vérifie certaines propriétés que nous rappelons ici.

La mesure d'un pavé de  $\mathbb{R}^n$  (produit de  $n$  intervalles de  $\mathbb{R}$ ) est le produit des longueurs des intervalles.  
Un ensemble est négligeable si et seulement si sa mesure de Lebesgue est nulle.

## B Lois usuelles

### B.1 Lois discrètes.

#### B.1.1 Lois, espérances, variances.

**Mesure de Dirac en  $x_0 \in \mathbb{R}$ .** On dit que  $X$  suit une mesure de Dirac en  $x_0 \in \mathbb{R}$  et on note  $X \rightsquigarrow \delta_{x_0}$  si  $X$  ne prend que la valeur  $x_0$ . On a alors

$$\mathbf{P}[X = x_0] = 1, \quad \mathbf{E}[X] = x_0, \quad \mathbf{Var}[X] = 0.$$

**Loi de Bernoulli.** On dit que  $X$  suit une loi de Bernoulli de paramètre  $p \in [0, 1]$  et on note  $X \rightsquigarrow \mathcal{B}(p)$  si  $X$  prend les valeurs 0 et 1 et

$$\mathbf{P}[X = 1] = p.$$

On a

$$\mathbf{E}[X] = p \text{ et } \mathbf{Var}[X] = p(1 - p).$$

Remarque, si  $p = 0$ ,  $X \rightsquigarrow \delta_0$  et si  $p = 1$ ,  $X \rightsquigarrow \delta_1$ .

**Loi uniforme sur  $\{1, \dots, n\}$ .** On dit que  $X$  suit une loi uniforme sur  $\{1, \dots, n\}$  et on note  $X \rightsquigarrow \mathcal{U}(\{1, \dots, n\})$  si  $X$  prend ses valeurs dans  $\{1, \dots, n\}$  et

$$\mathbf{P}[X = k] = \frac{1}{n} \text{ pour tout } k \in \{1, \dots, n\}.$$

On a

$$\mathbf{E}[X] = \frac{n+1}{2} \text{ et } \mathbf{Var}[X] = \frac{n^2 - 1}{12}.$$

**Loi binomiale.** On dit que  $X$  suit une loi binomiale de paramètre  $n \geq 1$  et  $p \in [0, 1]$  et on note  $X \rightsquigarrow \mathcal{B}(n, p)$  si  $X$  prend ses valeurs dans  $\{0, \dots, n\}$  et

$$\mathbf{P}[X = k] = \binom{n}{k} p^k (1 - p)^{n-k} \text{ pour tout } k \in \{0, \dots, n\}.$$

On a

$$\mathbf{E}[X] = np \text{ et } \mathbf{Var}[X] = np(1 - p).$$

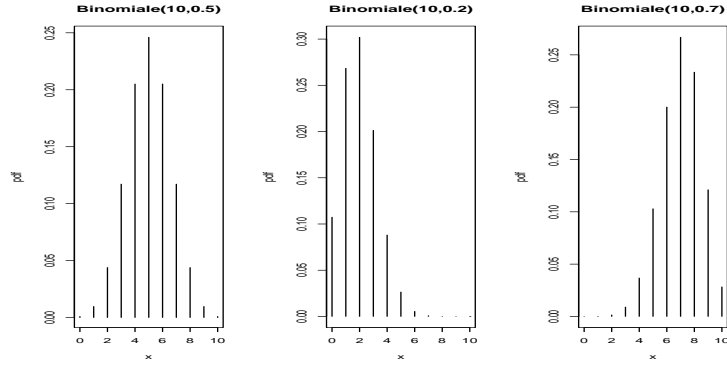


FIGURE 5 – Loi Binomiale

**Loi hypergéométrique.** On dit que  $X$  suit une loi hypergéométrique de paramètres  $N \geq 1$ ,  $n \geq 1$ ,  $p \in [0, 1]$  tels que  $pN \in \mathbb{N}$  et  $n \leq \min(pN, (1-p)N)$ , et on note

$$X \rightsquigarrow \mathcal{H}(N, n, p)$$

si  $X$  prend ses valeurs dans  $\{0, \dots, n\}$  et

$$\mathbf{P}[X = k] = \frac{\binom{pN}{k} \binom{(1-p)N}{n-k}}{\binom{N}{n}}.$$

On a

$$\mathbf{E}[X] = np \text{ et } \mathbf{Var}[X] = \frac{N-n}{N-1} np(1-p).$$

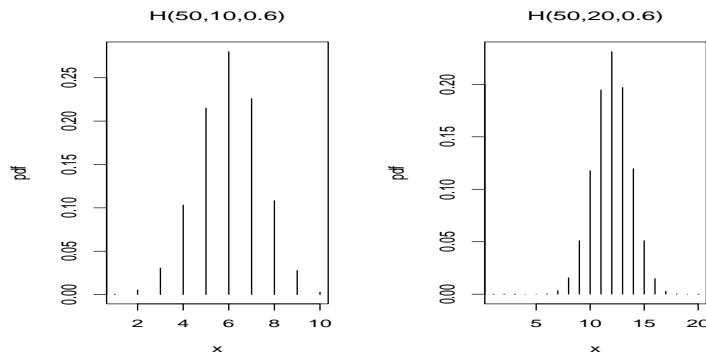


FIGURE 6 – Loi Hypergéométrique

**Loi de Poisson.** On dit que  $X$  suit une loi de Poisson de paramètre  $\lambda > 0$  et on note

$$X \rightsquigarrow \mathcal{P}(\lambda)$$

si  $X$  prend ses valeurs dans  $\mathbb{N}$  et

$$\mathbf{P}[X = k] = e^{-\lambda} \frac{\lambda^k}{k!} \text{ pour tout } k \in \mathbb{N}.$$

On a

$$\mathbf{E}[X] = \lambda \text{ et } \mathbf{Var}[X] = \lambda.$$

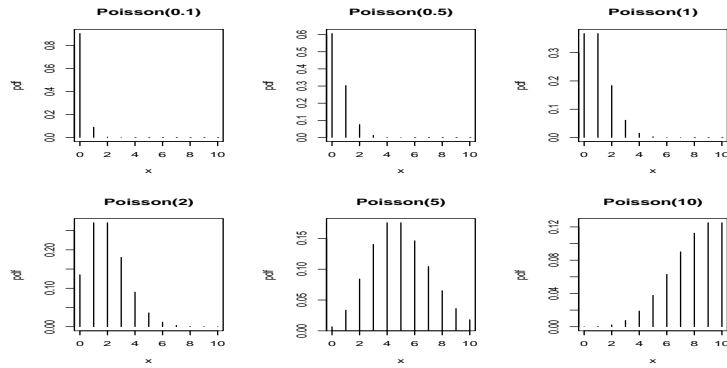


FIGURE 7 – Loi de Poisson

**Loi géométrique.** On dit que  $X$  suit une loi géométrique de paramètre  $p \in [0, 1]$  et on note

$$X \rightsquigarrow \mathcal{G}(p)$$

si  $X$  prend ses valeurs dans  $\mathbb{N}^*$  et

$$\mathbf{P}[X = k] = p(1 - p)^{k-1} \text{ pour tout } k \in \mathbb{N}^*.$$

On a

$$\mathbf{E}[X] = \frac{1}{p} \text{ et } \mathbf{Var}[X] = \frac{1 - p}{p^2}.$$

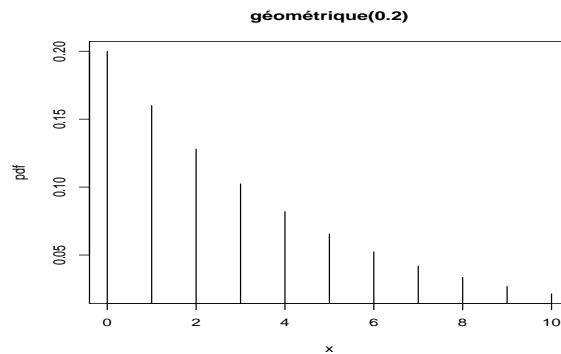


FIGURE 8 – Loi Géométrique

**Loi de Pascal.** On dit que  $X$  suit une loi de Pascal de paramètres  $n \in \mathbb{N}^*$  et  $p \in [0, 1]$  et on note

$$X \rightsquigarrow Pa(n, p)$$

si  $X$  prend ses valeurs dans  $\{k \in \mathbb{N} / k \geq n\}$  et

$$\mathbf{P}[X = k] = \binom{n-1}{k-1} p^n (1-p)^{k-n} \text{ pour tout } k \geq n.$$

On a

$$\mathbf{E}[X] = \frac{n}{p} \text{ et } \mathbf{Var}[X] = \frac{n(1-p)}{p^2}.$$

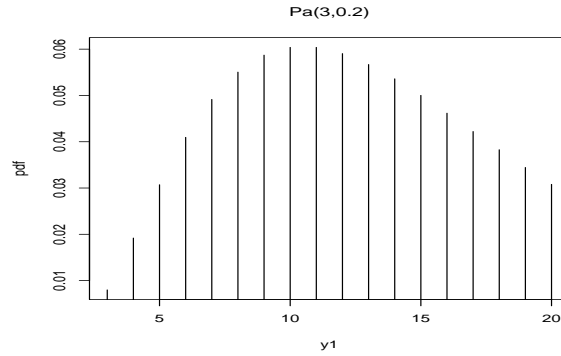


FIGURE 9 – Loi de Pascal

### B.1.2 Propriétés.

Une loi de Bernoulli de paramètre  $p$  est une loi binomiale de paramètres 1 et  $p$ .

La somme de 2 v.a. indépendantes suivant des lois binomiales de paramètres  $(n_1, p)$  et  $(n_2, p)$  respectivement suit une loi binomiale de paramètres  $n_1 + n_2$  et  $p$ .

La somme de deux v.a. indépendantes suivant respectivement des lois de Poisson de paramètre  $\lambda_1$  et  $\lambda_2$  suit une loi de Poisson de paramètre  $\lambda_1 + \lambda_2$ .

Une loi géométrique de paramètre  $p$  est une loi de Pascal de paramètres 1 et  $p$ .

La somme de 2 v.a. indépendantes suivant des lois de Pascal de paramètres  $(n_1, p)$  et  $(n_2, p)$  respectivement suit une loi de Pascal de paramètres  $n_1 + n_2$  et  $p$ .

### Proposition B.1

Convergence de la loi hypergéométrique vers la loi binomiale. On suppose que, pour tout  $k \geq 1$ ,  $X_k$  suit une loi hypergéométrique de paramètres  $N_k, n, p_k$  (avec  $p_k N_k \in \mathbb{N}^*$ ) et que

$$\lim_{k \rightarrow +\infty} N_k = +\infty, \quad \lim_{k \rightarrow +\infty} p_k = p \in ]0, 1[.$$

Alors

$$\mathcal{L}(X_k) \rightarrow \mathcal{B}(n, p) \text{ quand } k \rightarrow +\infty.$$

**Application.** *Approximation de la loi hypergéométrique par la loi binomiale.* Dans la pratique, si  $X \rightsquigarrow \mathcal{H}(N, n, p)$ , on peut approcher sa loi par une  $\mathcal{B}(n, p)$  dès que  $10n < N$ .

### Proposition B.2

*Convergence de la loi binomiale vers la loi de Poisson.* On suppose que, pour tout  $n \geq 1$ ,  $X_n$  suit une loi binomiale de paramètres  $n, p_n$  et que

$$\lim_{n \rightarrow +\infty} np_n = \lambda > 0.$$

Alors

$$\mathcal{L}(X_n) \rightarrow \mathcal{P}(\lambda) \text{ quand } n \rightarrow +\infty.$$

**Application.** *Approximation de la loi binomiale par la loi de Poisson.* Dans la pratique, si  $X \rightsquigarrow \mathcal{B}(n, p)$ , on peut approcher sa loi par une  $\mathcal{P}(np)$  dès que  $n \geq 30$  et  $p \leq 0,1$ .

## B.2 Lois à densité

### B.2.1 Lois, espérances, variances.

**Loi uniforme sur  $[a, b]$ .** On dit que  $X$  suit une loi uniforme sur  $[a, b]$ , où  $a < b$ , et on note

$$X \rightsquigarrow \mathcal{U}([a, b])$$

si  $X$  admet pour densité la fonction

$$f(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x) \text{ pour tout } x \in \mathbb{R}.$$

On a

$$\mathbf{E}[X] = \frac{a+b}{2} \text{ et } \mathbf{Var}[X] = \frac{(b-a)^2}{12}.$$

**Loi exponentielle.** On dit que  $X$  suit une loi exponentielle de paramètre  $\lambda > 0$  et on note

$$X \rightsquigarrow \mathcal{E}(\lambda)$$

si  $X$  admet pour densité la fonction

$$f(x) = \lambda e^{-\lambda x} \mathbf{1}_{]0,+\infty[}(x) \text{ pour tout } x \in \mathbb{R}.$$

On a

$$\mathbf{E}[X] = \frac{1}{\lambda} \text{ et } \mathbf{Var}[X] = \frac{1}{\lambda^2}.$$

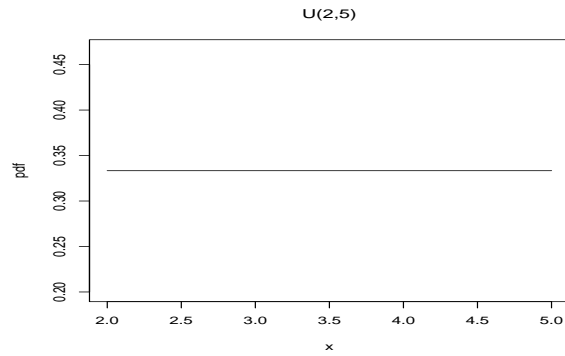


FIGURE 10 – Loi Uniforme

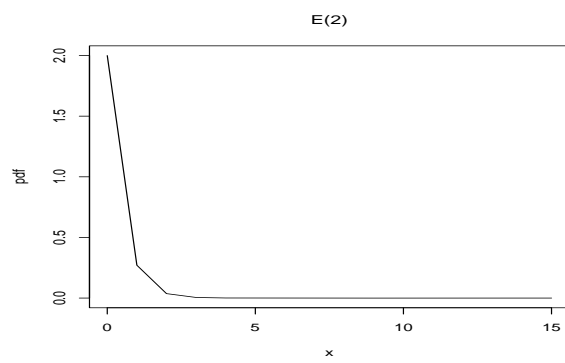


FIGURE 11 – Loi Exponentielle

**Loi Gamma.** On dit que  $X$  suit une loi Gamma de paramètres  $\lambda > 0$  et  $\alpha > 0$  et on note

$$X \rightsquigarrow \text{Gamma}(\lambda, \alpha)$$

si  $X$  admet pour densité la fonction

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \mathbf{1}_{]0, +\infty[}(x) \text{ pour tout } x \in \mathbb{R}$$

avec

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

On a :

$$\Gamma(\alpha + 1) = \alpha \Gamma(\alpha) \text{ pour tout } \alpha > 0 ;$$

$$\Gamma(n) = (n - 1)! \text{ pour tout } n \in \mathbb{N}^*.$$

On en déduit

$$\mathbf{E}[X] = \frac{\alpha}{\lambda} \text{ et } \mathbf{Var}[X] = \frac{\alpha}{\lambda^2}.$$

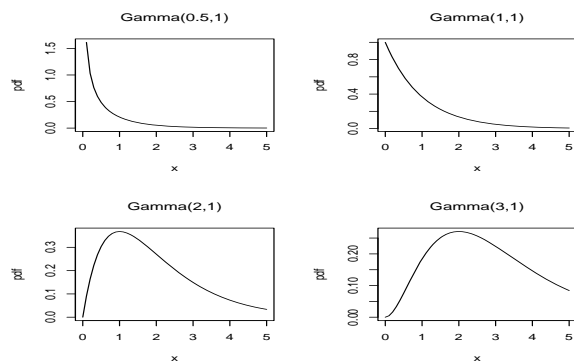


FIGURE 12 – Loi Gamma

**Lois Bêta.** On dit que  $X$  suit une loi Bêta de type I de paramètres  $\alpha > 0$  et  $\beta > 0$  et on note

$$X \rightsquigarrow \text{bêta } I(\alpha, \beta)$$

si  $X$  admet pour densité la fonction

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} \mathbf{1}_{[0,1]}(x) \text{ pour tout } x \in \mathbb{R}$$

avec

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx.$$

On a :

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \text{ pour tous } \alpha, \beta > 0.$$

On en déduit

$$\mathbf{E}[X] = \frac{\alpha}{\alpha + \beta} \text{ et } \mathbf{Var}[X] = \frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2}.$$

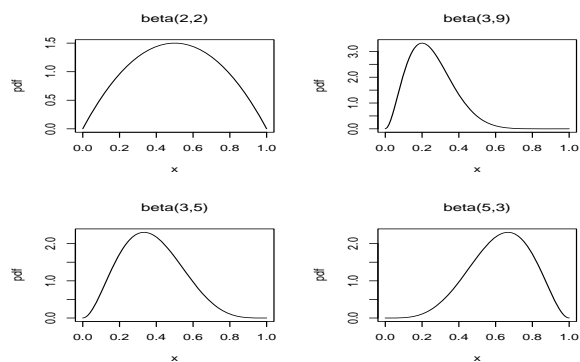


FIGURE 13 – Loi Bêta I



Si  $X$  suit une loi Bêta de type  $I$  de paramètres  $\alpha > 0$  et  $\beta > 0$ , alors la v.a.  $Y = X/(1 - X)$  suit une loi Bêta de type  $II$  notée  $\text{bêta } II(\alpha, \beta)$  et dont la densité s'obtient par changement de variable :

$$g(y) = \frac{1}{B(\alpha, \beta)} \frac{y^{\alpha-1}}{(1+y)^{\alpha+\beta}} \mathbf{1}_{]0, +\infty[}(y).$$

On a :

$$\mathbf{E}[Y] = \frac{\alpha}{\beta-1} \text{ et } \mathbf{Var}[Y] = \frac{\alpha(\alpha+\beta-1)}{(\beta-1)^2(\beta-2)}.$$

**Loi de Cauchy.** On dit que  $X$  suit une loi de Cauchy de paramètre  $c > 0$  et on note

$$X \rightsquigarrow \mathcal{C}(c)$$

si  $X$  admet pour densité la fonction

$$f(x) = \frac{c}{\pi(c^2 + x^2)} \text{ pour tout } x \in \mathbb{R}.$$

Un telle v.a. n'est pas intégrable.

**Loi normale (ou de Gauss ou de Laplace-Gauss).** On dit que  $X$  suit une loi normale de paramètres  $m \in \mathbb{R}$  et  $\sigma^2 > 0$  et on note

$$X \rightsquigarrow \mathcal{N}(m, \sigma^2)$$

si  $X$  admet pour densité la fonction

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \text{ pour tout } x \in \mathbb{R}.$$

On a :

$$\mathbf{E}[X] = m \text{ et } \mathbf{Var}[X] = \sigma^2.$$

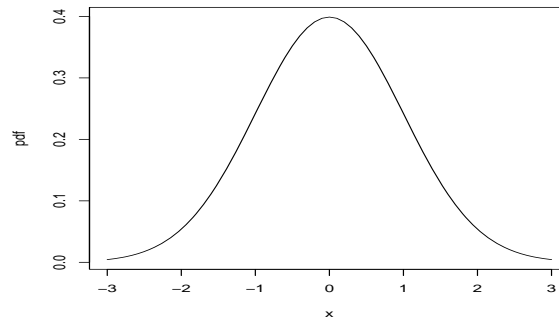


FIGURE 14 – Loi Normale

### B.2.2 Propriétés.

Une loi exponentielle de paramètre  $\lambda$  est une loi Gamma de paramètres  $\lambda$  et 1.

La somme de deux v.a. indépendantes suivant des lois Gamma de paramètres  $(\lambda, \alpha)$  et  $(\lambda, \beta)$  suit une loi Gamma de paramètres  $(\lambda, \alpha + \beta)$ .

Si  $X$  suit une loi normale de paramètres  $m$  et  $\sigma^2$  alors la v.a.

$$X^* = \frac{X - m}{\sigma}$$

suit une loi normale de paramètres 0 et 1.

La somme de deux v.a. indépendantes suivant des lois normales de paramètres  $(m_1, \sigma_1^2)$  et  $(m_2, \sigma_2^2)$  suit une loi normale de paramètres  $m_1 + m_2$  et  $\sigma_1^2 + \sigma_2^2$ .

#### Proposition B.3

Convergence de la loi binomiale vers la loi normale. On suppose que, pour tout  $n \geq 1$ ,  $X_n$  suit une loi binomiale de paramètre  $p \in ]0, 1[$ . Alors

$$\mathcal{L} \left( \frac{X_n - np}{\sqrt{np(1-p)}} \right) \rightarrow \mathcal{N}(0, 1) \text{ quand } n \rightarrow +\infty.$$

**Application.** Approximation de la loi binomiale par la loi normale. Dans la pratique, si  $X$  suit une loi  $\mathcal{B}(n, p)$ , on peut approcher sa loi par une  $\mathcal{N}(np, np(1-p))$  dès que  $np > 5$  et  $n(1-p) > 5$ .

#### Proposition B.4

Convergence de la loi de Poisson vers la loi normale. On suppose que, pour tout  $n \geq 1$   $X_n$  suit une loi normale de paramètre  $\lambda_n > 0$  et que

$$\lim_{n \rightarrow +\infty} \lambda_n = +\infty.$$

Alors

$$\mathcal{L} \left( \frac{X_n - \lambda_n}{\sqrt{\lambda_n}} \right) \rightarrow \mathcal{N}(0, 1) \text{ quand } n \rightarrow +\infty.$$

**Application.** Approximation de la loi de Poisson par la loi normale. Dans la pratique, si  $X \sim \mathcal{P}(\lambda)$  on peut approcher sa loi par une  $\mathcal{N}(\lambda, \lambda)$  dès que  $\lambda > 18$ .

**Remarque :** Dans les deux dernières approximations, il convient d'effectuer une correction de continuité, c'est-à-dire d'utiliser les égalités

$$\mathbf{P}[X = k] = \mathbf{P}[k - 1/2 < X \leq k + 1/2]$$

ou

$$\mathbf{P}[X \leq k] = \mathbf{P}[X \leq k + 1/2]$$

avant de faire l'approximation.

## C Intervalles de Confiance

Nous donnons ici les intervalles de confiance pour la moyenne et la variance.  
On suppose que  $(X_1, \dots, X_n)$  est un échantillon d'une v.a.  $X$ .

### C.1 IC sur la moyenne et la variance d'un échantillon gaussien.

On suppose que  $X$  suit une loi normale  $\mathcal{N}(m, \sigma)$ .

**IC sur la moyenne avec variance connue.** On utilise l'estimateur  $\bar{X}$ . On sait que

$$X^* = \frac{\bar{X} - m}{\sigma/\sqrt{n}} \rightsquigarrow \mathcal{N}(0, 1).$$

Pour tout  $\beta \in ]0, 1[$ , on note  $u_\beta$  le quantile d'ordre  $\beta$  de la loi normale centrée réduite. On a alors

$$\mathbf{P} \left[ u_{\alpha/2} < \frac{\bar{X} - m}{\sigma/\sqrt{n}} < u_{1-\alpha/2} \right] = \phi(u_{1-\alpha/2}) - \phi(u_{\alpha/2}) = 1 - \alpha,$$

on choisira l'intervalle de confiance

$$\left] \bar{X} - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2} \right[$$

puisque  $u_{1-\alpha/2} = -u_{\alpha/2}$ .

**IC sur la moyenne avec variance inconnue.** On sait que la variable

$$\frac{\bar{X} - m}{S/\sqrt{n}} \rightsquigarrow T_{n-1}.$$

On note à nouveau, pour tout  $\beta \in ]0, 1[$ ,  $t_\beta$  le quantile d'ordre  $\beta$  de la loi de Student à  $n - 1$  degrés de liberté. Alors, comme la densité d'une loi de Student est symétrique,  $t_{1-\alpha/2} = -t_{\alpha/2}$  et donc

$$\mathbf{P} \left[ -t_{1-\alpha/2} < \frac{\bar{X} - m}{S/\sqrt{n}} < t_{1-\alpha/2} \right] = 1 - \alpha.$$

L'intervalle

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

est un intervalle de confiance pour  $m$  au seuil de risque  $\alpha$ .

**IC sur la variance avec moyenne connue.** On utilise l'estimateur

$$T = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2.$$

On sait que  $nT/\sigma^2 \rightsquigarrow \chi_n^2$ . Si on note  $x_\beta$  le quantile d'ordre  $\beta$  de la loi du chi-deux à  $n$  degrés de liberté pour tout  $\beta \in ]0, 1[$ , alors

$$\mathbf{P} \left[ x_{\alpha/2} < n \frac{T}{\sigma^2} < x_{1-\alpha/2} \right] = 1 - \alpha$$

et donc un intervalle de confiance pour  $\sigma^2$  au seuil de risque  $\alpha$  est

$$\left] \frac{nT}{x_{1-\alpha/2}}, \frac{nT}{x_{\alpha/2}} \right[.$$

**IC sur la variance avec moyenne inconnue.** On utilise l'estimateur  $S^2$ . On a

$$(n-1) \frac{S^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2$$

et on procède comme ci-dessus : on obtient un intervalle de confiance au seuil de risque  $\alpha$

$$\left] \frac{(n-1)S^2}{x_{1-\alpha/2}}, \frac{(n-1)S^2}{x_{\alpha/2}} \right[.$$

## C.2 IC pour le paramètre d'un échantillon d'une loi de Bernoulli.

On suppose que  $X$  suit une loi de Bernoulli  $\mathcal{B}(p)$  et que l'échantillon est grand.

On sait que ,  $\frac{\bar{X} - p}{\sqrt{p(1-p)/n}} \rightsquigarrow \mathcal{N}(0, 1)$  approximativement.

Pour tout  $\beta \in ]0, 1[$ , on note  $u_\beta$  le quantile d'ordre  $\beta$  pour la loi normale centrée réduite. Un intervalle de confiance au seuil de risque  $\alpha$  pour le paramètre  $p$  est

$$\left] \bar{X} - \frac{u_{1-\alpha/2}}{\sqrt{n}} \sqrt{\bar{X}(1-\bar{X})}, \bar{X} + \frac{u_{1-\alpha/2}}{\sqrt{n}} \sqrt{\bar{X}(1-\bar{X})} \right[.$$

## C.3 IC pour la moyenne pour d'un échantillon non gaussien de carré intégrable.

Par le théorème limite central, un intervalle de confiance pour la moyenne peut-être obtenu si  $n$  est assez grand. L'intervalle de confiance au seuil de risque  $\alpha$  a la forme suivante :

$$\left] \bar{X} - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2} \right[$$

où  $u_{1-\alpha}$  est le quantile d'ordre  $1 - \alpha/2$  de la loi normale centrée réduite.

On n'en utilisera pas d'intervalle de confiance pour la variance dans le cadre non gaussien.

## D Tests statistiques

Nous donnons ici les *variables de décision* utilisées dans les tests statistiques usuels. Si la variable de décision est  $Z$  et  $z$  sa réalisation, la région de rejet de  $H_0$  est à choisir sous la forme  $W = \{z > c\}$ ,  $W = \{z \leq c\}$  ou  $W = \{|z| > c\}$  en fonction de la forme de  $H_1$ , la constante  $c$  étant déterminée par le risque de première espèce  $\alpha$ .

### D.1 Tests paramétriques pour un échantillon.

On suppose que  $(X_1, \dots, X_n)$  est un échantillon d'une v.a.  $X$ .

### D.1.1 Tests sur la moyenne et la variance d'un échantillon gaussien.

On suppose que  $X$  suit une loi normale  $\mathcal{N}(m, \sigma)$ .

#### Tests sur la moyenne avec variance connue.

$$H_0 : [m = m_0]$$

$$H_1 : [m = m_1] \text{ ou } [m < m_0] \text{ ou } [m > m_0] \text{ ou } [m \neq m_0].$$

$$\text{Sous } H_0, \frac{\bar{X} - m_0}{\sigma/\sqrt{n}} \rightsquigarrow \mathcal{N}(0, 1).$$

**Tests sur la moyenne avec variance inconnue.** Sous les mêmes hypothèses que ci-dessus,

$$\text{Sous } H_0, \frac{\bar{X} - m_0}{S/\sqrt{n}} \rightsquigarrow T_{n-1}.$$

#### Tests sur la variance avec moyenne connue.

$$H_0 : [\sigma = \sigma_0]$$

$$H_1 : [\sigma = \sigma_1] \text{ ou } [\sigma < \sigma_0] \text{ ou } [\sigma > \sigma_0] \text{ ou } [\sigma \neq \sigma_0].$$

$$\text{Sous } H_0, \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - m)^2 \rightsquigarrow \chi_n^2.$$

**Tests sur la variance avec moyenne inconnue.** Sous les mêmes hypothèses que ci-dessus,

$$\text{Sous } H_0, (n-1) \frac{S^2}{\sigma_0^2} \rightsquigarrow \chi_{n-1}^2.$$

### D.1.2 Test sur le paramètre d'un échantillon d'une loi de Bernoulli.

On suppose que  $X$  suit une loi de Bernoulli  $\mathcal{B}(p)$  et que l'échantillon est grand.

$$H_0 : [p = p_0]$$

$$H_1 : [p = p_1] \text{ ou } [p < p_0] \text{ ou } [p > p_0] \text{ ou } [p \neq p_0].$$

$$\text{Sous } H_0, \frac{\bar{X} - p_0}{\sqrt{p_0(1-p_0)}/\sqrt{n}} \rightsquigarrow \mathcal{N}(0, 1) \text{ approximativement.}$$

### D.1.3 Tests sur la moyenne d'un échantillon quelconque.

On suppose que  $X$  est une variable de carré intégrable, que  $n$  est grand. Alors on peut utiliser les tests de moyenne, en utilisant la loi normale à la place de la loi de Student.

## D.2 Tests paramétriques de comparaison d'échantillons indépendants.

On suppose que  $(X_1^1, \dots, X_{n_1}^1)$  et  $(X_1^2, \dots, X_{n_2}^2)$  sont des échantillons indépendants de v.a.  $X^1$  et  $X^2$  respectivement.

### D.2.1 Cas des échantillons gaussiens.

On suppose que  $X^1 \rightsquigarrow \mathcal{N}(m_1, \sigma_1^2)$  et  $X^2 \rightsquigarrow \mathcal{N}(m_2, \sigma_2^2)$ .

#### Test d'égalité des variances.

$$H_0 : [\sigma_1 = \sigma_2], \quad H_1 : [\sigma_1 \neq \sigma_2].$$

Si  $s_1^2 > s_2^2$ ,

$$\text{Sous } H_0, \quad \frac{S_1^2}{S_2^2} \rightsquigarrow F_{n_1-1, n_2-1}.$$

Si  $s_2^2 > s_1^2$ ,

$$\text{Sous } H_0, \quad \frac{S_2^2}{S_1^2} \rightsquigarrow F_{n_2-1, n_1-1}.$$

Nota Bene : Dans les deux cas, on choisit une région de rejet unilatérale.

#### Test d'égalité des moyennes si les variances sont connues.

$$H_0 : [m_1 = m_2], \quad H_1 : [m_1 \neq m_2].$$

$$\text{Sous } H_0, \quad \frac{\bar{X}^1 - \bar{X}^2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \rightsquigarrow \mathcal{N}(0, 1).$$

**Test d'égalité des moyennes si les variances sont égales.** Sous les mêmes hypothèses que ci-dessus,

$$\text{Sous } H_0, \quad T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{(n_1-1)S_1^2 + (n_2-1)S_2^2}} \frac{\sqrt{n_1+n_2-2}}{\sqrt{1/n_1 + 1/n_2}} \rightsquigarrow T_{n_1+n_2-2}.$$

Nota Bene : On choisit dans ces deux derniers tests une région de rejet bilatérale.

#### Test d'égalité des moyennes si les variances sont inconnues.

Si on se sait pas que les variances sont égales, mais que l'échantillon est grand, on peut encore utiliser le test 2.1.2,  $T$  suivant approximativement un  $T_{n_1+n_2-2}$ .

### D.2.2 Cas des échantillons de Bernoulli.

On suppose que  $X^1 \rightsquigarrow \mathcal{B}(p_1)$  et  $X^2 \rightsquigarrow \mathcal{B}(p_2)$ . On suppose également que les échantillons sont de grande taille et on pose

$$\hat{p} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}.$$

$$H_0 : [p_1 = p_2], \quad H_1 : [p_1 \neq p_2].$$

$$\text{Sous } H_0, \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\hat{p}(1-\hat{p})\sqrt{1/n_1 + 1/n_2}}} \rightsquigarrow \mathcal{N}(0, 1) \text{ approximativement.}$$

### D.2.3 Cas des échantillons quelconques.

On suppose que  $X^1$  et  $X^2$  sont des variables de carré intégrable. Si les échantillons sont assez grands, on peut encore utiliser le test 2.1.2, en remplaçant la loi de Student par la loi normale.

## D.3 Tests non paramétriques.

### D.3.1 Tests d'ajustement d'un échantillon à une loi donnée.

On suppose que  $(X_1, \dots, X_n)$  est un échantillon d'une v.a.  $X$ .

#### Loi entièrement déterminée.

$$H_0 : [X \rightsquigarrow \mathcal{L}], \quad H_1 : \overline{H_0}.$$

*Test d'ajustement du Chi-deux.* Soit  $Y$  une v.a. de loi  $\mathcal{L}$ . On choisit une partition finie  $C_1, \dots, C_J$  de l'ensemble des valeurs prises par  $Y$ . Pour tout  $j \in \{1, \dots, J\}$ , on note

$$p_j = \mathbf{P}[Y \in C_j] \text{ et } N_j = \text{Card} \{i \in \{1, \dots, n\} / X_i \in C_j\}$$

$$\text{Sous } H_0, D^2 = \sum_{j=1}^J \frac{(N_j - np_j)^2}{np_j} \rightsquigarrow \chi_{J-1}^2 \text{ approximativement.}$$

*Test de Kolmogorov-Smirnov.* Soit  $F$  la fonction de répartition d'une v.a.  $Y$  suivant la loi  $\mathcal{L}$ . On suppose que  $F$  est continue. Pour tout  $x \in \mathbb{R}$ , on note

$$F_n^*(x) = \frac{1}{n} \text{Card} \{i \in \{1, \dots, n\} / X_i \leq x\}$$

la fonction de répartition empirique de l'échantillon. On pose

$$D_n = \sup_{x \in \mathbb{R}} |F_n^*(x) - F(x)|.$$

$$\text{Sous } H_0, \mathbf{P}[\sqrt{n}D_n \leq y] \approx K(y) \text{ pour tout } y > 0$$

où la fonction  $K$  est tabulée. On choisit la région de rejet sous la forme  $W = \{d_n > c\}$  où  $d_n$  réalisation de  $D_n$ .

**Ajustement avec paramètres à estimer : test d'ajustement du chi-deux avec paramètres.** On suppose que  $\mathcal{L}$  dépend de paramètres  $\theta_1, \dots, \theta_p$ . On note  $T_1, \dots, T_p$  des estimateurs de ces paramètres et  $t_1, \dots, t_p$  les réalisations observées de ces paramètres sur l'échantillon.

$$H_0 : [X \rightsquigarrow \mathcal{L}(t_1, \dots, t_p)], \quad H_1 : \overline{H_0}.$$

Avec les mêmes notations que pour le test d'ajustement du chi-deux sans paramètre à estimer,

$$\text{Sous } H_0, D^2 \rightsquigarrow \chi_{J-1-p}^2 \text{ approximativement.}$$

### D.3.2 Test de provenance d'échantillons d'une même population : le test du chi-deux.

Soient  $(X_1^1, \dots, X_{n_1}^1), \dots, (X_1^m, \dots, X_{n_m}^m)$   $m$  échantillons indépendants de v.a.  $X^1, \dots, X^m$ .

$H_0$  : [les échantillons proviennent de la même population],  $H_1$  :  $\overline{H_0}$ .

On note

$$N_{kj} = \text{Card}\{i \in \{1, \dots, n_k\} / X_i^k \in C_j\},$$
$$N_{k\cdot} = \sum_{j=1}^J N_{kj}, \quad N_{\cdot j} = \sum_{k=1}^m N_{kj}, \quad N = \sum_{k=1}^m \sum_{j=1}^J N_{kj}.$$
$$\text{Sous } H_0, \quad D_0^2 = \sum_{k=1}^m \sum_{j=1}^J \frac{(N_{kj} - N_{k\cdot} N_{\cdot j} / N)^2}{N_{k\cdot} N_{\cdot j} / N} \rightsquigarrow \chi_{(J-1)(m-1)}^2.$$

### D.3.3 Test d'indépendance : le test du chi-deux.

On applique le test précédent pour tester l'indépendance de deux v.a.  $X$  et  $Y$  à partir d'un échantillon  $((X_1, Y_1), \dots, (X_n, Y_n))$  de  $(X, Y)$ . On note alors  $\{C_i, i \in \{1, \dots, I\}\}$  et  $\{C^j, j \in \{1, \dots, J\}\}$  des partitions des valeurs prises par  $X$  et  $Y$  respectivement, et

$$N_{ij} = \text{Card}\{l \in \{1, \dots, n\} / X_l \in C_i \text{ et } Y_l \in C^j\}.$$

$$H_0 : [X \text{ et } Y \text{ indépendantes}], \quad H_1 : \overline{H_0}.$$

On procède comme ci-dessus.



# Fonctions EXCEL

Voici les principales fonctions Excel utiles la simulation et la réalisation de tests d'hypothèses.

## Version Excel 2007 ou OpenOffice

**ALEA()** : génère un nombre pseudo aléatoire entre 0 et 1.

**MOYENNE(plage)** : fait la moyenne des données de la plage de données plage.

**VAR.P(plage)** : fait la variance des données de la plage de données plage (calcule l'estimateur  $\Sigma^2$ ).

**VAR(plage)** : calcule la valeur de l'estimateur  $S^2$  sur la plage de données plage.

A partir de la version Excel 2010, les fonctions suivantes existent aussi **VAR.P.N(plage)** : fait la variance des données de la plage de données plage (calcule l'estimateur  $\Sigma^2$ ).

**VAR.S(plage)** : calcule la valeur de l'estimateur  $S^2$  sur la plage de données plage.

**LOI.NORMALE.STANDARD(x)** =  $P[X < x]$  pour  $x$  réel et  $X$  loi normale de paramètres 0 et 1.

La fonction inverse est **LOI.NORMALE.STANDARD.INVERSE(p)** pour une proba  $p$ .

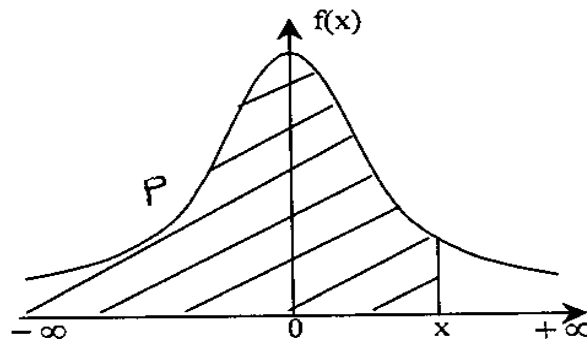


FIGURE 15 – LOI.NORMALE.STANDARD.INVERSE

**LOI.STUDENT(t;n;1)** =  $P[T > t]$  pour  $t$  réel,  $n$  entier et  $T$  loi de Student à  $n$  degrés de liberté (unilatérale).

**LOI.STUDENT(t;n;2)** =  $P[|T| > t] = G_n(t)$  pour  $t$  réel positif,  $n$  entier et  $T$  loi de Student à  $n$  degrés de liberté (bilatérale).

**LOI.STUDENT.INVERSE(p;n)** =  $G_n^{-1}(p)$  pour  $p$  proba est l'inverse de la précédente (bilatérale).

Par conséquent, si  $F_n$  est la fonction de répartition loi de Student à  $n$  degrés de liberté,  $F_n(t) = 1 - G_n(t)/2$  et  $F_n^{-1}(1 - \alpha/2) = G_n^{-1}(\alpha)$ .

L'inverse de la loi unilatérale en  $p$  et  $n$  est donc **LOI.STUDENT.INVERSE(2p;n)**.

A partir de la version 2010 |

**LOI.KHIDEUX(x;n)** =  $P[X > x] = H_n(x)$  pour  $x$  réel positif et  $X$  loi du Chi-deux à  $n$  degrés de liberté.

La fonction inverse est **KHIDEUX.INVERSE(p;n)** =  $H_n^{-1}(p)$  pour une proba  $p$  et un entier  $n$ .

Par conséquent, si  $F_n$  est la fonction de répartition loi du Chi-deux à  $n$  degrés de liberté,  $F_n(t) = 1 - G_n(t)$  et

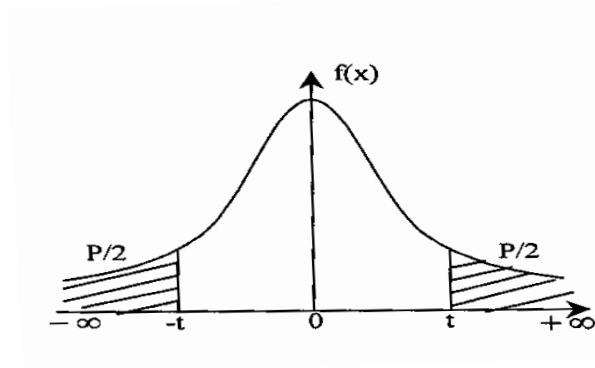


FIGURE 16 – LOI.STUDENT.INVERSE

$$F_n^{-1}(1 - \alpha/2) = H_n^{-1}(\alpha/2)$$

Dans OpenOffice, la fonction **LOI.KHIDEUX.INVERSE(p;n)** pour une proba  $p$  et un entier  $n$  est l'inverse de la fonction de répartition d'une loi Chi-deux à  $n$  degrés de liberté

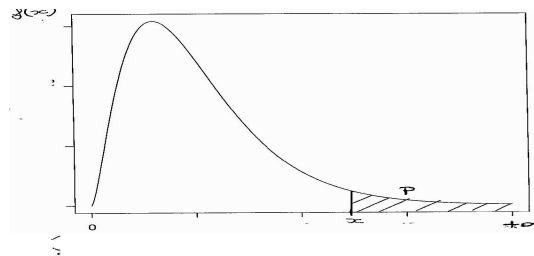


FIGURE 17 – KHIDEUX.INVERSE

**LOI.F(f;n1;n2)** =  $P[F > f] = I(f)$  pour  $f$  réel positif,  $n1$  et  $n2$  entiers et  $F$  loi de Fisher à  $n1, n2$  degrés de liberté.

La fonction inverse est **INVERSE.LOI.F(p;n1;n2)** =  $I^{-1}(p)$  pour une proba  $p$  et des entiers  $n1$  et  $n2$ .

Par conséquent, si  $F$  est la fonction de répartition loi de Fisher à  $n1, n2$  degrés de liberté,  $F(t) = 1 - I(t)$  et  $F^{-1}(1 - \alpha/2) = I^{-1}(\alpha/2)$

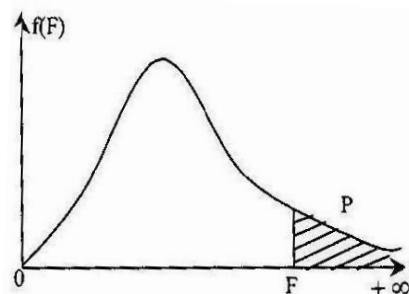


FIGURE 18 – INVERSE.LOIF

### Version Excel 2010 |

**ALEA()** : génère un nombre pseudo aléatoire entre 0 et 1.

**MOYENNE(plage)** : fait la moyenne des données de la plage de données plage.

**VAR.P.N(plage)** : fait la variance des données de la plage de données plage (calcule l'estimateur  $\Sigma^2$ ).

**VAR.S(plage)** : calcule la valeur de l'estimateur  $S^2$  sur la plage de données plage.

**LOI.NORMALE.STANDARD.N(x;Vrai)** =  $P[X < x]$  pour  $x$  réel et  $X$  loi normale de paramètres 0 et 1.

La fonction inverse est **LOI.NORMALE.STANDARD.INVERSE.N(p)** pour une proba  $p$ .

**LOI.STUDENT.N(t;n;Vrai)** =  $P[T \leq t]$  pour  $t$  réel,  $n$  entier et  $T$  loi de Student à  $n$  degrés de liberté (unilatérale).

**LOI.STUDENT.INVERSE.N(p;n)** pour  $p$  probabilité est l'inverse de la précédente.

**LOI.KHIDEUX.N(x;n;Vrai)** =  $P[X \leq x]$  pour  $x$  réel positif et  $X$  loi du Chi-deux à  $n$  degrés de liberté.

La fonction inverse est **LOI.KHIDEUX.INVERSE(p;n)** pour une proba  $p$  et un entier  $n$ .

**LOI.F.N(f;n1;n2.VRAI)** =  $P[F \leq f]$  pour  $f$  réel positif,  $n1$  et  $n2$  entiers et  $F$  loi de Fisher à  $n1, n2$  degrés de liberté.

La fonction inverse est **INVERSE.LOIF.N(p;n1;n2)** pour une proba  $p$  et des entiers  $n1$  et  $n2$ .