

# Probabilités et Statistique

Jean-Marc Azaïs  
Laboratoire de Statistique et Probabilités

1<sup>er</sup> avril 1999



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Analyse combinatoire</b>	<b>5</b>
2.1	Échantillons ordonnés avec répétitions . . . . .	5
2.2	Échantillons ordonnés sans répétition . . . . .	5
2.3	Échantillons non ordonnés sans répétition . . . . .	6
2.4	Échantillons non ordonnés avec répétitions . . . . .	7
<b>3</b>	<b>Définition d'une probabilité sur un ensemble discret</b>	<b>8</b>
3.1	Généralités . . . . .	8
3.2	Cas $\Omega$ fini, probabilité uniforme . . . . .	8
3.3	Probabilité générale sur un ensemble fini . . . . .	9
3.4	Probabilité sur un ensemble dénombrable . . . . .	11
<b>4</b>	<b>Conditionnement et indépendance</b>	<b>13</b>
4.1	Conditionnement . . . . .	13
4.2	Indépendance . . . . .	15
4.3	Expériences indépendantes, produits d'espaces de probabilité . . . . .	15
4.4	Variables aléatoires . . . . .	17
4.5	Variables aléatoires indépendantes . . . . .	18
<b>5</b>	<b>Espérance, variance, moments</b>	<b>20</b>
5.1	Espérance . . . . .	20
5.2	Variance . . . . .	23
5.3	Covariance et corrélation . . . . .	25
<b>6</b>	<b>Variables aléatoires réelles admettant une densité</b>	<b>27</b>
6.1	Introduction . . . . .	27
6.2	Espérance d'une fonction d'une variable aléatoire . . . . .	28
6.3	Changement de variables . . . . .	29
<b>7</b>	<b>Observations vectorielles</b>	<b>31</b>
7.1	Introduction . . . . .	31
7.2	Indépendance . . . . .	33

7.3	Densités conditionnelles . . . . .	33
<b>8</b>	<b>Régression linéaire et théorème limites</b>	<b>34</b>
8.1	Théorèmes limites . . . . .	34
8.2	Régression linéaire . . . . .	35
<b>9</b>	<b>Statistique</b>	<b>37</b>
9.1	Généralités sur la statistique . . . . .	37
9.1.1	La population et l'échantillon . . . . .	37
9.1.2	Différent types de variables . . . . .	38
9.2	Étude d'une variable quantitative continue . . . . .	38
9.2.1	Indicateurs de position ou indicateurs centraux . . . . .	38
9.2.2	Indicateurs de dispersion . . . . .	39
9.2.3	Représentations graphique . . . . .	39
9.3	Échantillonnage quantitatif dans une population . . . . .	39
9.4	Échantillonnage pour une variable qualitative . . . . .	42

# Chapitre 1

## Introduction

Ce cours a pour but de vous familiariser avec le raisonnement probabiliste. Par rapport à un autre cours de mathématiques, il se distingue par l'ambition de modéliser certains phénomènes réels. Un modèle mathématiquement correct ne suffira donc pas toujours, il faudra encore que celui-ci corresponde aux observations. Ce document se divise en deux parties principales (ch 3,4,5 et ch 6,7,8) : la première concerne les probabilités discrètes et la seconde les probabilités continues.

A ses débuts, la théorie du “calcul des probabilités” a concerné principalement l'étude et la modélisation des jeux de hasard. Les premiers travaux sont attribués à Pascal et à Fermat (1654) sur des problèmes posés par le Chevalier de Méré, joueur professionnel et mathématicien amateur. La probabilité est définie comme le nombre de cas favorables sur le nombre de cas possibles et la solution fait souvent appel au dénombrement.

Imaginons maintenant que je joue une partie de tennis contre P. Sampras (Numéro 1 mondial). Si on veut bien avoir l'indulgence d'admettre qu'il y a deux solutions pour le vainqueur, il ne paraît pas du tout raisonnable de leur donner la même probabilité. Il faut donc généraliser la notion de probabilité, c'est ce qui est fait au Chapitre 3. On y définit dans sa plus grande généralité la notion de probabilité sur un ensemble fini et même dénombrable. C'est la partie “probabilités discrètes”.

Dans certaines situations, le cadre théorique précédent est insuffisant, c'est la cas en particulier quand on s'intéresse à une mesure physique ( poids tension électrique, etc.) qui prend ses valeurs sur  $\mathbb{R}$  qui n'est malheureusement pas dénombrable. Ce sont alors d'autres techniques qui sont employées aux chapitres 7 et 8 avec la notion de densité de probabilité. C'est la partie “probabilités continues”.

Afin de mettre l'accent au maximum sur les problèmes fondamentaux de conditionnement et de modélisation, nous avons volontairement réduit à sa plus simple expression le matériel théorique de ce cours. Nous avons en particulier toujours privilégié la démarche constructive plutôt que la démarche axiomatique.

Quelques ouvrages

Frugier G. (1992). “Exercice ordinaire de Probabilités” Ellipses.

Kriekeberg & Ziezold. “initiation aux probabilités” Belin.

Leboeuf Roque & Guegan. “probabilités” Ellipse.

Lepage Moore & Roy. “Introduction a la théorie des probabilités”. Presses de l’Université du Québec.

# Chapitre 2

## Analyse combinatoire

On se place sur un ensemble  $E$  **fini**.  $E$  est appelée la **population**. On note  $n$  son cardinal (nombre d'éléments). On va s'intéresser à des objets appelés **échantillons sur  $E$**  que l'on va définir cas par cas.

### 2.1 Échantillons ordonnés avec répétitions

**Définition 2.1.1** *Un échantillon ordonné avec répétitions de longueur  $p$  est une suite ordonnée de  $p$  éléments de  $E$  :  $(e_1, \dots, e_p)$ , c'est également un  $p$ -uplet à valeurs dans  $E$ . Il correspond à une **application** de  $\{1, \dots, p\}$  vers  $E$ .*

**Proposition 2.1.1** *Le nombre de d'échantillons ordonnés avec répétitions de longueur  $p$  sur  $E$  vaut :*

$$n^p.$$

*démonstration :*

Le résultat vient de la définition même de la multiplication : le cardinal du produit cartésien de deux ensembles est le produit des cardinaux des deux ensembles.  $\square$

*Exemple*

Combien de sigles de trois lettres peut on former avec les 26 lettres de l'alphabet?

Il y a **répétition** car  $AAA$ , par exemple, est un sigle.

Il y a **ordre** car  $ABC \neq BCA$ .

La taille de  $E$  est  $n = 26$ ,  $p$  vaut 3.

La réponse est donc :  $26^3 = 17576$ .

### 2.2 Échantillons ordonnés sans répétition

**Définition 2.2.1** *Un échantillon ordonné sans répétition de longueur  $p$  sur  $E$  est un  $p$ -uplet dont tous les éléments sont différents. Il faut donc  $p \leq n$ . Un tel*

échantillon est également appelé arrangement de longueur  $p$ . Il correspond à une **injection** de  $\{1, \dots, p\}$  vers  $E$ .

**Proposition 2.2.1** *Le nombre d'échantillons ordonnés sans répétition de longueur  $p$  sur  $E$  vaut*

- zéro, si  $p > n$
- $A_n^p = \frac{n!}{(n-p)!}$ , si  $p \leq n$ .

*Démonstration:* Il y a  $n$  possibilités pour choisir le premier élément,  $n - 1$  possibilités pour le second, par une récurrence immédiate on obtient le résultat.  $\square$

*Remarque:* Quand  $p = n$ , on obtient les **permutations de  $E$** , leur nombre vaut  $n!$ .

*Exemple: Nombre de tirages possibles du Loto.* Lors du tirage du Loto à la télévision, on tire 7 boules (6+ la complémentaire) en notant l'**ordre** (même si cet ordre n'intervient pas dans la suite du jeu). Il y a donc ordre et non-répétition car non-remise. Le nombre de tirages possibles vaut donc:  $A_{49}^7 = 49 \times 48 \times 47 \times 46 \times 45 \times 44 \times 43 = 432\ 938\ 943\ 360$ .

## 2.3 Échantillons non ordonnés sans répétition

**Définition 2.3.1** *Un échantillon non ordonné sans répétition de taille  $p$  sur  $E$  est tout simplement un sous ensemble de taille  $p$  de  $E$ . Un tel sous ensemble était appelé, par le passé, une combinaison.*

**Proposition 2.3.1** *Le nombre de sous ensembles de taille  $p$  de  $E$  vaut*

$$\binom{n}{p} = C_n^p = \frac{n!}{p!(n-p)!}, \text{ si } p \leq n ; 0 \text{ sinon.}$$

*Démonstration:*  $p!$  arrangements correspondent au même sous ensemble.  $\square$

*Exemple:* Combien de couleurs secondaires peut-on obtenir en mélangeant 2 couleurs primaires. Il y a trois couleurs primaires: rouge, bleu et jaune;  $n = 3$ . On considère des échantillons non ordonnés: bleu + rouge = rouge + bleu. Il n'y a pas de répétition: jaune + jaune ne convient pas. La réponse est donc  $C_3^2 = 3$ .

## 2.4 Échantillons non ordonnés avec répétitions

Nous allons introduire la définition par un exemple.

*Exemple:* On dispose d'un grand nombre de paquets de beurre, paire d'oeufs, sucre et farine pesant tous exactement 100g. Combien peut-on faire de gâteaux d'un kilo différents avec ces ingrédients? On admettra comme "gâteau" des solutions triviales comme: omelette pure, beurre pur, farine pure etc ... Un exemple de solution non triviale est donné par :

$$(Be, Be, Be, Oe, Oe, Oe, Su, Su, Fa, Fa)$$

Il est clair que l'ordre des ingrédients n'intervient pas.

**Définition 2.4.1** *Un échantillon non ordonné avec répétitions de taille  $p$  sur  $E$  est caractérisé par la suite **ordonnée**  $j_1, \dots, j_n$  des indices de répétition des différents éléments de  $E$ . Les  $j_i$  sont  $\geq 0$  et vérifient*

$$j_1 + j_2 + \dots + j_n = p$$

*Un tel échantillon est également appelé multiensemble (dans la mesure où un élément peut lui appartenir plusieurs fois).*

Dans l'exemple des gâteaux on a  $n = 4$ ,  $p = 10$  et pour la solution particulière  $j_1 = j_2 = 3$ ,  $j_3 = j_4 = 2$  en supposant que l'ordre des éléments de  $E$  est (Be, Oe, Su, Fa).

**Proposition 2.4.1** *Le nombre d'échantillons non ordonnés avec répétitions de taille  $p$  vaut*

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

*Démonstration:* On considère  $n+p-1$  sites.

• • • • •

On place  $n-1$  barres en des sites quelconques

• • | • • | | • • | •

La suite des nombres de sites libres entre les barres définit  $n$  entiers dont la somme vaut  $p$ . Réciproquement à tout multiensemble on peut associer une configuration comme ci-dessus. Dans l'exemple de la figure:  $j_1 = 2$ ,  $j_2 = 2$ ,  $j_3 = 0$ ,  $j_4 = 2$ ,  $j_5 = 1$ ,  $n = 5$ ,  $p = 7$ . Pour finir la démonstration, il suffit de remarquer que

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

□

Dans l'exemple, il y a  $C_{13}^{10} = 286$  recettes.

# Chapitre 3

## Définition d'une probabilité sur un ensemble discret

### 3.1 Généralités

L'étude d'un phénomène aléatoire commence par la description de l'ensemble des résultats possibles de l'expérience.

- on note  $\Omega$  cet ensemble : "ensemble des réalisations".
- Les éléments  $\omega$  ou points de  $\Omega$  sont appelés des "réalisations".
- Les parties ou sous ensembles de  $\Omega$ , sont appelés des événements. Ils sont notés traditionnellement par des majuscules latines: A,B,etc...

### 3.2 Cas $\Omega$ fini, probabilité uniforme

Considérons l'expérience suivante :

*exemple 1* : D'un jeu de 32 cartes neuf et battu on extrait une carte. Pour des raisons de symétrie, chaque carte joue le même rôle, on dira qu'elle "a le même nombre de chances " d'être tirée et on traduit cela mathématiquement par

$$P\{\text{Roi de coeur}\} = 1/32$$

Cette notion notée P est appelée **probabilité**. La probabilité de tout événement est donnée par la formule classique

$$P(A) = \frac{\text{nb de cas favorables}}{\text{nb de cas possibles}} = \frac{|A|}{|\Omega|} \quad (3.1)$$

où  $|A|$  est le cardinal ou le nombre d'éléments de A.

*Exemple 1 (suite)* : la probabilité de "tirer un as" est

$$P(A) = \frac{\text{nb d'as}}{\text{nb de cartes}} = 4/32 = 1/8.$$

**Proposition 3.2.1** La fonction définie par la formule (3.1) apparaît comme une fonction  $P$

$$\mathcal{P}(\Omega) \rightarrow [0, 1]$$

vérifiant

1.  $P(\Omega) = 1$
2. si  $A \cap B = \emptyset$ ,  $P(A \cup B) = P(A) + P(B)$ .

La propriété (2) est appelée additivité.

démonstration: (1) évident

(2) si  $A \cap B = \emptyset$ , alors  $|A \cup B| = |A| + |B|$  ce qui donne le résultat.  $\square$

Cette probabilité est appelée probabilité uniforme car elle donne le même poids à chaque réalisation  $\omega$ . C'est celle que l'on sous-entend quand on dit "tirer au hasard" sans autre précision.

*Exemple 2:* Les tirages du Loto sont fait au hasard, c'est à dire suivant une probabilité uniforme.

### 3.3 Probabilité générale sur un ensemble fini

Il existe des expériences dont le résultat est aléatoire mais dont certaines réalisations sont plus probables que d'autres. Pour pouvoir modéliser ces situations, on doit étendre la notion de probabilité.

Soit  $\Omega$  un ensemble fini et soit  $K(\omega)$  une fonction de poids de  $\Omega \rightarrow [0, 1]$  vérifiant :

$$\sum_{\omega \in \Omega} K(\omega) = 1$$

(La somme des poids vaut 1)

On associe à la fonction  $K$  une probabilité par la relation

$$A \rightarrow P(A) = \sum_{\omega \in A} K(\omega)$$

*Exemple 3* On jette deux dés de couleurs différentes. On a donc

$$\Omega = \{(i, j), 1 \leq i \leq 6 ; 1 \leq j \leq 6\}.$$

On munit cet ensemble de la probabilité uniforme. On note  $(i, j)$  le résultat du premier dé puis du second et on s'intéresse à la somme  $i + j$  des deux dés pour obtenir par exemple

$$P(2) = P\{(1, 1)\} = 1/36$$

$$P(3) = P\{(1, 2), (2, 1)\} = 2/36$$

$$P(2) = P\{(1, 3), (3, 1), (2, 2)\} = 3/36$$

$$P(7) = P\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\} = 6/36$$

$$P(12) = P\{(6, 6)\} = 1/36$$

Cela définit une probabilité sur

$$\Omega' = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

associée aux poids

$$1/36, 2/36, 3/36, 4/36, 5/36, 6/36, 5/36, 4/36, 3/36, 2/36, 1/36$$

**Définition 3.3.1** Soit  $\Omega$  fini, on appelle probabilité sur  $\Omega$  toute fonction  $P: \mathcal{P}(\Omega) \rightarrow [0, 1]$  qui satisfait

$$P(A) = \sum_{\omega \in A} K(\omega)$$

où  $K$  est une fonction quelconque  $\Omega \rightarrow [0, 1]$  vérifiant

$$\sum_{\omega \in \Omega} K(\omega) = 1.$$

$K$  est appelée la fonction de poids associée à la probabilité.

**Proposition 3.3.1** 1.  $K$  est définie de manière unique connaissant  $P$  et réciproquement

2.  $P$  vérifie

- $P(A) \in [0, 1]$
- $P(\Omega) = 1$
- $P$  est additive.

*Démonstration*

$K(\omega) = P(\{\omega\})$  définit  $K$  de manière unique, le reste est très facile. □

*Remarques*

- $K(\omega) = 1/|\Omega|$  définit la probabilité uniforme
- la partie 2 de la proposition est une caractérisation; toute fonction vérifiant ces trois propriétés est une probabilité .

### 3.4 Probabilité sur un ensemble dénombrable

*Exemple 4:* L'auto-stoppeur. On s'intéresse au nombre de voitures qui vont passer devant un auto-stoppeur avant que la dernière ne s'arrête. Ce nombre n'est pas borné (sauf par des quantités très grandes sans intérêt). L'ensemble des possibles  $\Omega$  est donc égal à  $\{1, 2, \dots, n, \dots\}$ . C'est un ensemble infini dénombrable.

*Rappel sur les ensembles dénombrables*

- Un ensemble  $\Omega$  est dénombrable si il existe une bijection de  $\Omega$  vers une partie de  $\mathbb{N}$
- Les ensembles dénombrables sont soit finis, soit infinis. Dans ce dernier cas il existe un "énumération", c-à-d une bijection avec  $\mathbb{N}$ .
- Toute partie d'un ensemble dénombrable est dénombrable.
- Si  $\{A_i, i \in I\}$  est un famille de parties non vides 2 à 2 disjointes de  $\Omega$  dénombrable, alors  $I$  est dénombrable.
- Le produit cartésien de deux ensembles dénombrables est encore dénombrable.
- Deux ensembles infinis dénombrables sont en bijection.

**Définition 3.4.1** Soit  $\Omega$  infini dénombrable:  $\Omega = \{\omega_1, \omega_2, \dots, \omega_n, \dots\}$ , on appelle probabilité sur  $\Omega$  toute fonction  $P: \mathcal{P}(\Omega) \rightarrow [0, 1]$  qui satisfait

$$P(A) = \sum_{\omega \in A} K(\omega) \quad (3.2)$$

où  $K$  est un fonction quelconque  $\Omega \rightarrow [0, 1]$  telle que la série  $\sum_{n \in \mathbb{N}} K(\omega_n)$  converge et vaut 1.  $K$  est appelée la fonction de poids associée à la probabilité.

*Remarque fondamentale.* La série de terme général  $K(\omega_i)$  est positive donc absolument convergente donc commutativement et associativement convergente. Sa somme est donc indépendante de l'ordre choisi pour les élément de  $\Omega$  on peut donc noter

$$\sum_{\omega \in \Omega} K(\omega) = 1$$

Par ailleurs :

$$\sum_{\omega \in A} K(\omega)$$

est soit une somme finie, soit une série absolument convergente comme sous série de la série (3.2). Là encore l'ordre des éléments de  $A$  n'intervient pas.

**Proposition 3.4.1**

1.  $K$  est définie de manière unique connaissant  $P$  et réciproquement

2.  $P$  vérifie

- $P(A) \in [0, 1]$
- $P(\Omega) = 1$
- $P$  est  $\sigma$ -additive : si  $A_1, A_2, \dots, A_n, \dots$  est une suite infinie d'événements disjoints

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} P(A_n)$$

*Démonstration* : Seule la démonstration de la dernière propriété change par rapport au cas fini. Elle est une conséquence de la propriété de convergence associative des séries absolument convergentes intervenant.  $\square$

*Remarque* :

- Il ne peut exister de probabilité uniforme.
- La propriété (2) de la proposition caractérise une probabilité. Elle pourrait être prise comme définition.

**Définition 3.4.2 Espace de probabilité.**

Soit  $\Omega$  dénombrable et  $P$  une probabilité sur  $\Omega$ , le couple  $(\Omega, P)$  est appelé espace de probabilité.

# Chapitre 4

## Conditionnement et indépendance

### 4.1 Conditionnement

*Exemple 5 :*

On tire au hasard une carte dans un paquet de 32 cartes. La probabilité de tirer un valet est :

$$P(\text{valet}) = 4/32 = 1/8$$

On renouvelle maintenant cette expérience mais une tierce personne regarde la carte sans nous la montrer et nous dit

*“c’est un honneur rouge”*

L’ensemble des réalisations s’est modifié est vaut

$$\Omega' = \{VK, VC, DK, DC, RK, RC, AK, AC\}$$

l’événement ”valet” devient maintenant  $VK, VC$  et la probabilité est modifiée : on la note  $P'$ . On a

$$P'\{\text{valet}\} = \frac{\#\{\text{valets qui sont des honneurs rouges}\}}{\#\{\text{honneurs rouges}\}} = 2/8 = 1/4$$

Cette nouvelle probabilité est notée

$$P(\text{valet/honneur rouge}) = \frac{P(\text{valet et honneur rouge})}{P(\text{honneur rouge})}$$

Par généralisation :

**Définition 4.1.1** Soit  $B$  un événement tel que  $P(B) \neq 0$  , la **probabilité conditionnelle** de  $A$  sachant  $B$  se note  $P(A/B)$  et vaut

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

On laisse à titre d'exercice de vérifier que c'est bien une probabilité sur  $\Omega$ . C'est une modélisation de la notion intuitive de probabilité de  $A$  quand on a l'information que  $B$  est bien réalisé.

*Formule*: Si  $P(B) \neq 0$  alors  $P(A \cap B) = P(A/B)P(B)$ .

**Formule de Bayes.** Exemple d'utilisation : on veut apprécier la fiabilité d'un test de dépistage d'une maladie. On recherche la probabilité d'être effectivement malade quand le test est positif. On la note  $P(M/T)$ . On peut mesurer facilement en laboratoire :

- la probabilité que le test soit positif sur un sujet sain :  $P(T/M^c)$
- la probabilité que le test soit positif sur un sujet malade :  $P(T/M)$ .

On suppose connaître par ailleurs la fréquence de la maladie dans la population de référence :  $P(M)$ . On recherche

$$P(M/T) = \frac{P(M \cap T)}{P(T)} = \frac{P(M)P(T/M)}{P(T \cap M) + P(T \cap M^c)} = \frac{P(M)P(T/M)}{P(M)P(T/M) + P(M^c)P(T/M^c)}$$

On a donc inversé les conditionnements. Le raisonnement ci dessus est parfaitement général dès que  $P(M)$ ,  $P(M^c)$  et  $P(T)$  sont non nulles. Nous avons donc prouvé

**Proposition 4.1.1 Formule de Bayes.** *Si  $P(M)$ ,  $P(M^c)$  et  $P(T)$  sont non nulles, alors :*

$$P(M/T) = \frac{P(M)P(T/M)}{P(M)P(T/M) + P(M^c)P(T/M^c)}$$

**Proposition 4.1.2 Formule des probabilités totales et formule de Bayes généralisée.** *Soient  $A_1, \dots, A_n$ , une partition de  $\Omega$  en ensembles de probabilités strictement positives alors*

- Pour tout événement  $B$  on a

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

- Soit  $B$  de probabilité strictement positive,

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

*Remarque* : Les deux formules ci-dessus s'étendent au cas d'un nombre dénombrable de  $A_i$ .

## 4.2 Indépendance

**Définition 4.2.1** Soit  $(\Omega, P)$  un espace de probabilité, deux événements  $A$  et  $B$  de cet espace sont dits indépendants ssi :

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

*Remarque :* On peut remplacer  $A$  et  $B$  par leurs complémentaires.

**Définition 4.2.2 Indépendance mutuelle de plusieurs événements** Soient  $A_1, \dots, A_n$ ,  $n$  événements d'un même espace de probabilité, il sont dits indépendants ssi pour toute sous suite  $i_1, \dots, i_k$  de  $1, \dots, n$  :

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

Pour trois événements  $A, B, C$  il faut

$$P(AB) = P(A)P(B)$$

$$P(BC) = P(B)P(C)$$

$$P(AC) = P(A)P(C)$$

$$P(ABC) = P(A)P(B)P(C)$$

(pour alléger les notations, on a oublié le symbole  $\cap$ ).

La définition implique clairement la proposition suivante

**Proposition 4.2.1** Soient  $A_1, \dots, A_n$ ,  $n$  événements indépendants d'un même espace de probabilité, tout sous ensemble :  $\{A_{i_1}, \dots, A_{i_k}, (k \leq n)\}$  est formé d'événements mutuellement indépendants.

## 4.3 Expériences indépendantes, produits d'espaces de probabilité

*Exemple 6 :* Considérons un exemple très simple qui avait mal été résolu en son temps par le mathématicien d'Alembert : *On jette deux pièces identiques, quelle est la probabilité d'obtenir **pile et face** ?*

A ce problème nous proposons deux solutions :

*Solution 1 (D'Alembert)*

Les deux pièces sont indiscernables, on s'intéresse donc aux paires non ordonnées.

$$\Omega = [\{p, f\}, \{f, f\}, \{p, p\}]$$

On le munit de sa probabilité uniforme pour en déduire :

$$P(\{p, f\}) = 1/3$$

*Solution 2*

On distingue les deux pièces, on s'intéresse donc aux couples ordonnés.

$$\Omega = [(p, f), (f, p), (f, f), (p, p)]$$

On le munit encore de sa probabilité uniforme :

$$P[\{p, f\}] = P[(f, p), (p, f)] = 2/4 = 1/2.$$

Mathématiquement, ces deux solutions sont parfaitement correctes. Cependant, il est clair que seule l'une d'entre elles peut refléter correctement la réalité. En fait, la première solution est expérimentalement fautive à cause du

**Principe d'indépendance** : Deux expériences conduites dans des conditions indépendantes au sens courant, donnent des résultats indépendants au sens probabiliste.

Cet énoncé est un principe expérimental au même titre que les principes de thermodynamique

**Application** : Les mouvements des deux pièces sont indépendants, donc les résultats des deux pièces sont indépendants au sens probabiliste. Par exemple

$$P((p, p)) = P(p) \times P(p) = 1/2 \times 1/2 = 1/4.$$

C'est donc la solution 2 qui reflète correctement la réalité. Avec deux pièces et un peu de temps on peut le vérifier.

Soient maintenant deux expériences indépendantes, à chaque expérience est associé un espace de probabilité. A l'aide de ces deux espaces, on va réaliser un nouvel espace qui va englober les deux espaces précédents de manière indépendante, on l'appelle "produit des espaces de probabilité"

**Théorème 4.3.1** Soient  $(\Omega_1, P_1)$  et  $(\Omega_2, P_2)$  deux espaces de probabilité, soient  $K_1$  et  $K_2$  les fonctions de poids associées, alors il existe une unique probabilité  $P$  sur  $\Omega = \Omega_1 \times \Omega_2$  telle que

$$\forall A_1 \in \Omega_1, \forall A_2 \in \Omega_2, P(A_1 \times A_2) = P(A_1) \times P(A_2) \quad (4.1)$$

Cette probabilité est déterminée par la fonction de poids :

$$K(\omega_1, \omega_2) = K_1(\omega_1) \times K_2(\omega_2) \quad (4.2)$$

*Démonstration* :

Unicité : Si  $P$  vérifie (4.1) alors

$$K(\omega_1, \omega_2) = P(\{\omega_1\} \times \{\omega_2\}) = P(\{\omega_1\}) \times P(\{\omega_2\}) = K_1(\omega_1) \times K_2(\omega_2).$$

La fonction de poids est donc déterminée de manière unique.

Existence : Définissons la fonction de poids  $K$  et donc la probabilité  $P$  par (4.2), il suffit de vérifier que

1.  $K$  est une fonction de poids

2. La probabilité  $P$  associée à  $K$  vérifie (4.1)

1.

$$\sum_{\omega \in \Omega} K(\omega) = \sum_{\omega_1 \in \Omega_1; \omega_2 \in \Omega_2} K(\omega_1, \omega_2) = \left( \sum_{\omega_1 \in \Omega_1} K(\omega_1) \right) \left( \sum_{\omega_2 \in \Omega_2} K(\omega_2) \right) = 1.$$

Ce résultat est dû, soit au produit de sommes finies si  $\Omega$  est fini, soit au produit de séries si  $\Omega$  est infini.  $P$  est donc une probabilité .

2. Soient  $A_1$  et  $A_2$  deux événements de  $\Omega_1$  et  $\Omega_2$ , le même raisonnement que ci-dessus montre que

$$P(A_1 \times A_2) = P(A_1) \times P(A_2).$$

□

*Notation*: L'espace de probabilité ainsi obtenu est parfois noté

$$(\Omega_1 \times \Omega_2, P_1 \otimes P_2).$$

La solution (2) du problème des deux pièces revient à créer le produit par lui-même de l'espace de probabilité associé à une pièce qui est

$$\Omega = \{p, f\}; P\{p\} = P\{f\} = 1/2.$$

On admettra la généralisation du théorème au produit de plusieurs espaces de probabilité .

## 4.4 Variables aléatoires

Pour résoudre certains problèmes de probabilité on a besoin de se placer dans un espace où on peut calculer, c'est à dire un groupe ou un corps. Ce sera presque toujours  $\mathbb{Z}$  ou  $\mathbb{R}$ , pour cela on définit une application de  $\Omega \rightarrow E \subset \mathbb{R}$ .

*Exemple*: Deux dés (suite).

Nous avons vu que l'espace  $\Omega$  était constitué des paires  $(i, j)$ . Cet ensemble était muni de la probabilité uniforme. Nous pouvons introduire sur cet exemple la fonction  $f$ : somme des deux dés

$$(i, j) \in \Omega \rightarrow f(i, j) = i + j \in \mathbb{R}.$$

C'est un exemple de variable aléatoire.

**Définition 4.4.1** *Soit  $(\Omega, P)$  un espace de probabilité, on appelle variable aléatoire (abrev. v.a.) à valeurs dans  $E$ , toute application de  $\Omega \rightarrow E$ . Dans cette définition  $E$  est un espace quelconque.*

**Proposition 4.4.1** –  $X(\Omega)$  est toujours dénombrable.

- Toute fonction d'une ou plusieurs variables aléatoires est une variable aléatoire.
- Si les  $X_i$  sont des v. a. réelles alors  $Y = \sum_{i=1,n} a_i X_i$  est encore une v.a. réelle.

*Exemple* : Variable indicatrice. Soit  $A$  un événement sur  $(\Omega, P)$ , on lui associe la variable aléatoire  $\mathbb{1}_A$  à valeurs dans  $\{0,1\}$  qui vaut

- 1 si  $\omega \in A$
- 0 sinon.

**Définition 4.4.2** La loi de probabilité de  $X$  v.a. à valeurs dans  $E$ , est la probabilité  $P_X$  sur  $X(\Omega)$  définie par :

$$P_X(A) = P(X \in A) = P(\{\omega / X(\omega) \in A\})$$

Remarquez que l'événement à droite est défini sur  $\Omega$ . Dans l'exemple des deux dés on "transporte" ainsi la probabilité de  $\Omega$  sur  $\{2, 12\}$ .

**Définition 4.4.3** Fonction de répartition d'une variable aléatoire à valeurs dans un sous ensemble de  $\mathbf{R}$ . C'est la fonction  $F(x)$  de  $\mathbb{R} \rightarrow [0, 1]$  définie par :

$$F(x) = P(X \leq x).$$

## 4.5 Variables aléatoires indépendantes

**Théorème-Définition 4.5.1** Soient  $X_1, \dots, X_n$ ,  $n$  variables aléatoires à valeurs dans  $E_1, \dots, E_n$ , les quatre propriétés ci-dessous sont équivalentes.

1. Pour tout  $A_i$  inclus dans  $E_i$  les événements  $\{X_i \in A_i\}$  sont indépendants ( $i = 1, \dots, n$ ).
2. Pour tout  $A_i$  inclus dans  $E_i$ ,  $P[X_1 \in A_1, \dots, X_n \in A_n] = P[X_1 \in A_1] \times \dots \times P[X_n \in A_n]$ .
3. La loi de  $X = (X_1, \dots, X_n)$  est le produit des lois des  $X_i$ .
4.  $\forall (x_1, \dots, x_n) \in (E_1, \dots, E_n); P[X_1 = x_1, \dots, X_n = x_n] = P[X_1 = x_1] \times \dots \times P[X_n = x_n]$ .

On dira dans ce cas que les v.a. sont indépendantes.

*Démonstration:* (1)  $\Rightarrow$  (2) de manière évidente.

(2)  $\Rightarrow$  (1) : La définition d'événements indépendants concerne une suite extraite  $A_{i_1}, \dots, A_{i_k}, (k \leq n)$  Dans les espaces "oubliés" on pose  $A_i = \Omega_i$  pour obtenir le résultat.

(3)  $\Leftrightarrow$  (2) par le théorème sur les produits d'espaces de probabilité.

(4)  $\Leftrightarrow$  "la fonction de poids est le produit des fonctions de poids"  $\Leftrightarrow$  (3).

□

**Proposition 4.5.1** Soient  $X_1, \dots, X_n$ ,  $n$  variables aléatoires indépendantes.

- Toute suite extraite  $X_{i_1}, \dots, X_{i_k} (k \leq n)$  est une suite indépendante
- Soient  $f_1, \dots, f_n$ ,  $n$  fonctions, on définit:  $Y_i = f_i(X_i), i = 1, \dots, n$ , alors les variables  $Y_i$  sont indépendantes.

*Démonstration:* Nous laissons le premier résultat en exercice, pour le second, on utilise la caractérisation (2).

$$\begin{aligned} P\{(Y_1 \in B_1), \dots, (Y_n \in B_n)\} &= P\{(X_1 \in f_1^{-1}(B_1)), \dots, (X_n \in f_n^{-1}(B_n))\} \\ &= P\{(X_1 \in f_1^{-1}(B_1))\} \times \dots \times P\{(X_n \in f_n^{-1}(B_n))\} = P\{Y_1 \in B_1\} \dots P\{Y_n \in B_n\} \end{aligned}$$

□

**La loi de Bernoulli et la loi binomiale.** Nous définissons d'abord la loi de Bernoulli de paramètre  $p$ .  $X_1$  suit une telle loi ssi

- $X_1 = 0$  ou  $1$
- $P\{X_1 = 1\} = p$  et  $P\{X_1 = 0\} = 1 - p$ .

Il s'agit d'une expérience en tout ou rien : succès ou échec. Nous définissons maintenant la loi binomiale de paramètres  $n$  et  $p$ , notée  $\mathcal{B}(n, p)$ , on considère maintenant  $n$  copies indépendantes  $X_1, \dots, X_n$  de la loi précédente et la variable

$$S = X_1 + \dots + X_n$$

$S$  est le nombre de succès en  $n$  tentatives indépendantes. Cherchons la fonction de poids de cette probabilité :

$$P(S = k) ?$$

- Si  $k > n$  il est clair qu'il n'y a pas de solution et que la probabilité est nulle.
- Si  $k \leq n$  On recherche d'abord tous les  $n$ -uplets comportant  $k$  "un" et  $n - k$  "zéros". Il y en a  $C_n^k$ . A cause de l'indépendance, chacune de ces configurations a la probabilité:  $p^k(1 - p)^{n-k}$ . au total nous obtenons donc :

**Proposition 4.5.2** La loi binomiale de paramètres  $n$  et  $p$ , notée  $\mathcal{B}(n, p)$  est définie par la fonction de poids

$$P(S = k) = C_n^k p^k (1 - p)^{n-k}.$$

# Chapitre 5

## Espérance, variance, moments

### 5.1 Espérance

*Exemple 8: Jeu de dé à la sauvette.* Dans les jeux de dés à la sauvette dans la rue, le joueur mise 1 Fr sur un numéro. Si ce numéro sort, la banque ajoute 4 fois la mise et donne le total soit 5 Frs au joueur. Sinon elle ne verse rien. Intuitivement le jeu est défavorable au joueur car "une fois sur 6" il gagne 4 Frs et "5 fois sur 6" il perd 1Fr. Le "gain moyen" du joueur est donc

$$4/6 - 5/6 = -1/6.$$

Le jeu est favorable à la banque.

On formalise la notion précédente de la façon suivante.

**Définition 5.1.1** Soit  $X$  une v.a. réelle qui prend les valeurs  $x_1, x_2, \dots, x_k, \dots$ . Si la série

$$\sum_{k=0}^{\infty} x_k P(X = x_k)$$

est absolument convergente, on dit que l'espérance de  $X$  existe et on note

$$E(X) = \sum_{k=0}^{\infty} x_k P(X = x_k) = \sum_{x \in X(\Omega)} x P(X = x).$$

*Remarques :*

- Par abus de notation, on appelle série absolument convergente toute somme finie.

–

$$E(X) \text{ existe} \Leftrightarrow E(|X|) \text{ existe.}$$

*Exemple* : Un dé

$$E(X) = 1/6(1 + 2 + 3 + 4 + 5 + 6) = 3,5.$$

*Contre exemple* : Jeu de Saint-Petersbourg.

Un joueur joue contre la banque de la manière suivante : On lance une pièce jusqu'à la première occurrence de "face" . Si face apparaît au nième lancer le joueur reçoit  $2^n$  francs. Combien le joueur doit il reverser à la banque pour que le jeu soit équitable?

On montre :

$$P\{\text{face au nième coup}\} = P\{\overbrace{ppp\dots p}^{n-1} f\} = 2^{-n}.$$

de sorte que

$$E(X) = \sum_{n=1}^{\infty} 2^{-n} 2^n = 1 + 1 + 1 + \dots = +\infty.$$

L'espérance n'existe pas, il n'y a pas de façon de rendre le jeu équitable.

**Proposition 5.1.1** *Espérance d'une fonction d'une variable aléatoire. Soit  $f$  une fonction réelle, on pose  $Y=f(X)$ . Alors*

$$E(Y) = \sum_{x \in X(\Omega)} f(x)P(X = x)$$

*si cette somme est absolument convergente.*

*Démonstration* : En exercice.

**Proposition 5.1.2 Propriétés élémentaires de l'espérance.** *Soient  $X$  et  $Y$ , deux variables aléatoires réelles,*

(i) *Bornitude* : *si  $E(X)$  existe et si  $|Y| \leq |X|$  alors  $E(Y)$  existe.*

*Nous supposons maintenant que  $E(X)$  et  $E(Y)$  existent.*

(ii) *Linéarité.*  *$E(aX + bY)$  existe et vaut  $aE(X) + bE(Y)$ .*

(iii) *Monotonie.* *Si  $X \geq Y$ , alors  $E(X) \geq E(Y)$ .*

(iv) *Inégalité de Markov.* *Si  $X \geq 0$  et  $a \geq 0$ ,*

$$P(X \geq a) \leq (1/a)E(X).$$

*Démonstration* :

(i)  $E(Y)$  existe  $\Leftrightarrow \sum |Y(\omega)| K(\omega) < \infty$  mais le terme général de cette série positive est majoré par le terme général de la série  $\sum |X(\omega)| K(\omega)$  qui est convergente par hypothèse.

(ii)  $|aX + bY| \leq |a||X| + |b||Y|$  ce qui montre que l'espérance existe bien.

$$E(aX + bY) = \sum_{x,y} (ax + by)P[(X, Y) = (x, y)] = a \sum_{x,y} xP[(X, Y) = (x, y)] + b \sum_{x,y} yP[(X, Y) = (x, y)] = a \sum_x xP[X = x] + b \sum_y yP[Y = y] = aE(X) + bE(Y)$$

(iii)  $Z = X - Y \geq 0$ . La définition de l'espérance implique que  $E(Z) \geq 0$ , on applique alors (ii).

(iv) Comme  $X$  est positive,

$$X \geq a \cdot \mathbb{1}_{\{X \geq a\}}$$

et on applique (iii).

□

**Proposition 5.1.3** Si  $X$  et  $Y$  sont indépendantes et admettent une espérance, alors  $E(XY)$  existe et vaut

$$E(XY) = E(X)E(Y)$$

*Démonstration:*

$$E(XY) = \sum_{x,y} xyP[(X, Y) = (x, y)] = \sum_{x,y} xyP[X = x]P[Y = y] = \left[ \sum_x xP(X = x) \right] \left[ \sum_y yP(Y = y) \right].$$

□

**Corollaire** Si  $X_1, X_2, \dots, X_n$  sont  $n$  variables aléatoires indépendantes et si  $f_1, \dots, f_n$  sont  $n$  fonctions réelles telles que  $E(f_i(X_i))$  existe alors

$$E(f_1(X_1) \dots f_n(X_n)) = (E(f_1(X_1)))(E(f_2(X_2))) \dots (E(f_n(X_n))).$$

### Exemples de lois et calculs d'espérances

Loi de Bernoulli de paramètre  $p$ : Si  $X$  suit une loi de Bernoulli de paramètre  $p$

$$E(X) = p.$$

Loi binomiale: Si  $X$  suit la loi  $B(n, p)$

$$E(X) = np.$$

Loi géométrique: (Loi de l'auto-stoppeur). On considère une suite infinie d'épreuves de Bernoulli indépendantes  $X_1 \dots X_n \dots$  de même paramètre  $p$  (chaque voiture qui passe prend ou ne prend pas indépendamment l'auto-stoppeur avec

la même proba). On définit la variable aléatoire  $Y$  : numéro d'ordre du premier succès.

$$\{Y = k\} = \{(k - 1) \text{ échecs puis un succès sur les } k \text{ premiers essais}\}$$

L'indépendance des tentatives implique

$$P(Y = k) = (1 - p)^{k-1}p \quad , \quad k = 1, 2, \dots$$

En utilisant les résultats sur les séries géométriques, on vérifie que la somme des poids vaut bien 1. On montre également :

$$E(Y) = 1/p.$$

## 5.2 Variance

**Définition 5.2.1** Une variable aléatoire  $Y$  est dite centrée ssi son espérance est nulle :  $E(Y) = 0$ . Si  $X$  admet une espérance alors  $\tilde{X} = X - E(X)$  est centrée.

**Définition 5.2.2** On appelle moment d'ordre 2 de la variable  $X$  la quantité  $E(X^2)$  si elle existe. Sinon on dit que la variable n'a pas de moment d'ordre 2.

**Définition 5.2.3** Variance d'une variable aléatoire ayant un moment d'ordre 2. C'est la quantité

$$\text{Var}(X) = E(\tilde{X}^2) = E(X - E(X))^2.$$

C'est le moment d'ordre 2 de la variable centrée. La variance mesure l'écart quadratique de la variable à son espérance.

*Remarques :*

- $|X| \leq X^2 + 1$ ; Donc si le moment d'ordre 2 existe, l'espérance existe aussi et la définition ci-dessus a un sens.
- $\text{Var}(X) = \sum (x - m)^2 P(X = x)$  avec  $m = E(X)$ .
- Nous dirons indifféremment que la variable  $X$  admet une variance ou un moment d'ordre 2 si  $E(X^2) \leq \infty$ .

### Proposition 5.2.1

- $\text{Var}(X + a) = \text{Var}(X)$
- $\text{Var}(aX) = a^2 \text{Var}(X)$

- $Var(X) = E(X^2) - (E(X))^2$
- $E((X - a)^2) = Var(X) + (E(X) - a)^2$  est donc minimum quand  $a = E(X)$  et le minimum vaut  $Var(X)$ .

*Démonstration:*

- Les deux premiers résultats sont faciles.
- $Var(X) = E((X - E(X))^2) = E(X^2) - 2E(XE(X)) + E(X)^2 = E(X^2) - 2E(X)^2 + E(X)^2$
- par le même développement :

$$E((X - a)^2) = E(X^2) - (E(X))^2 + (E(X))^2 - 2aE(X) + a^2 = Var(X) + (E(X) - a)^2$$

□

**Proposition 5.2.2** *Inégalité de Bienaymé-Chebychev. Si  $Var(X)$  existe, alors : pour tout  $a > 0$   $P(|X - E(X)| \geq a) \leq 1/a^2 Var(X)$ .*

*Démonstration:*  $Z = (X - E(X))^2$  est une variable positive et  $E(Z) = Var(X)$   
On applique l'inégalité de Markov à  $Z$ .

$$P(|X - E(X)| \geq a) = P(Z \geq a^2) \leq 1/a^2 E(Z) = 1/a^2 Var(X)$$

□

**Proposition 5.2.3** *Si  $X$  et  $Y$  sont indépendantes et ont toutes deux une variance*

$$Var(X + Y) = Var(X) + Var(Y).$$

*Démonstration:* On considère d'abord les variables centrées associées  $\tilde{X}$  et  $\tilde{Y}$ .  
 $E(\tilde{X} + \tilde{Y}) = 0$  donc

$$\begin{aligned} Var(\tilde{X} + \tilde{Y}) &= E(\tilde{X} + \tilde{Y})^2 = E(\tilde{X})^2 + E(\tilde{Y})^2 + 2E(\tilde{X})E(\tilde{Y}) \\ &= E(\tilde{X})^2 + E(\tilde{Y})^2 = Var(\tilde{X}) + Var(\tilde{Y}). \end{aligned}$$

Dans le cas général

$$Var(X + Y) = Var(\tilde{X} + \tilde{Y}) = Var(\tilde{X}) + Var(\tilde{Y}) = Var(X) + Var(Y)$$

□

*Exemples :* Pour la loi de Bernoulli le calcul montre que :  $Var(X) = p(1 - p)$ .  
On en déduit que si  $S$  suit la loi  $B(n,p)$  :

$$Var(S) = np(1 - p).$$

Si  $Y$  suit la loi géométrique de paramètre  $p$ , le calcul montre

$$Var(Y) = \frac{1 - p}{p^2}.$$

**Loi de Poisson** Cette loi est utilisée pour modéliser les occurrences d'événements rares, par exemple : nombre de cas d'une maladie rare dans une grande population. C'est la loi sur  $\mathbb{N}$  donnée par la fonction de poids :

$$P(X = k) = e^{-\lambda} \lambda^k / k!$$

on montre que

$$E(X) = Var(X) = \lambda.$$

### 5.3 Covariance et corrélation

Soient  $X$  et  $Y$  deux variables qui ont une loi jointe  $P(X = x, Y = y)$  et qui admettent une variance (un moment d'ordre 2), alors l'espérance de  $XY$  existe car

$$|XY| \leq 1/2(X^2 + Y^2).$$

**Proposition 5.3.1** (*Inégalité de Schwarz*). Soient  $X$  et  $Y$  admettant une variance, alors

$$|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$$

*Démonstration :* Soit  $\lambda$  un réel, on a

$$E((\lambda X + Y)^2) = \lambda^2 E(X^2) + 2\lambda E(XY) + E(Y^2) \geq 0.$$

Pour que le trinôme soit toujours positif il faut que le discriminant  $b^2 - 4ac$  soit négatif, soit

$$(E(XY))^2 - E(X^2)E(Y^2) \leq 0$$

en prenant la racine, on obtient le résultat.

□

*Remarques :*

– en appliquant l'inégalité de Schwarz à  $|X|$   $|Y|$  on obtient

$$E(|X| |Y|) \leq \sqrt{E(X^2)E(Y^2)}.$$

– En l’appliquant à  $|X|$  et 1, on obtient

$$E(|X|) \leq \sqrt{E(X^2)}.$$

**Définition 5.3.1** Soient  $X$  et  $Y$  deux variables définies sur le même espace de probabilité et admettant des moments d’ordre 2. On appelle covariance de  $X$  et  $Y$  la quantité

$$\text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))).$$

Remarques

- $\text{Cov}(X, X) = \text{Var}(X)$ ,
- $\text{Cov}(X, Y) = E(XY) - E(X).E(Y)$ ,
- si  $X$  et  $Y$  sont indépendantes  $\text{Cov}(X, Y) = 0$ ,
- si  $a$  et  $b$  sont des réels
  - $\text{Cov}(X - a, Y - b) = \text{Cov}(X, Y)$
  - $\text{Cov}(aX + b, Y) = a \text{Cov}(X, Y)$
  - $\text{Cov}(aX, bY) = ab \text{Cov}(X, Y)$ .

**Proposition 5.3.2** Soient  $n$  variables  $(X_1, \dots, X_n)$  pourvues de moments d’ordre 2, alors

$$\text{Var}(X_1 + \dots + X_n) = \sum_{k=1}^n \text{Var}(X_k) + 2 \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j).$$

*Démonstration:* Comme les variances et covariances ne sont pas modifiées par addition d’une constante, on peut remplacer les  $X_k$  par les  $Y_k = X_k - E(X_k)$ , la formule à démontrer revient alors à

$$E((Y_1 + \dots + Y_n)^2) = \sum_{k=1, n} E(Y_k^2) + 2 \sum_{1 \leq i < j \leq n} E(Y_i Y_j),$$

qui résulte directement du développement du carré de la somme.

□

# Chapitre 6

## Variables aléatoires réelles admettant une densité

### 6.1 Introduction

La définition d'une probabilité sur  $\mathbb{R}$  dans le sens le plus général est basée sur la théorie de la mesure et sera traitée dans les années ultérieures.

Nous allons nous intéresser ici au cas particulier beaucoup plus simple des probabilités qui admettent une densité.

**Définition 6.1.1 Densité de probabilité sur  $\mathbb{R}$ .** *C'est toute fonction positive Riemann intégrable telle que*

$$\int_{\mathbb{R}} f(x) dx = 1$$

*Remarque :* Dans la suite de ce cours nous ne considérerons que des fonctions  $f$  continues ou continues par morceaux. Ces fonctions sont toutes Riemann intégrable.

**Définition 6.1.2** *On dit qu'une variable aléatoire réelle  $X$  a pour densité  $f(x)$  si pour tout intervalle  $]a, b[$  réel ( $a$  et  $b$  étant éventuellement infini), on a*

$$P(a < X < b) = \int_a^b f(x) dx \tag{6.1}$$

*Remarque* La formule 6.1 définit une fonction que nous appellerons probabilité qui va de l'ensemble des intervalles réels dans  $[0,1]$ . On veut que cette fonction satisfasse la propriété de  $\sigma$ -additivité introduite dans le chapitre 2. Ce faisant, on peut prolonger  $P$  à une classe d'ensembles qui contient les unions dénombrables d'intervalles. Nous admettrons qu'une telle probabilité est  $\sigma$ -additive.

## Conséquences

– Si  $a$  est un réel,

$$P(X = a) = \lim_{n \rightarrow \infty} P(a - 1/n < X < a + 1/n) = \lim_{n \rightarrow \infty} \int_{a-1/n}^{a+1/n} f(x) dx = 0.$$

On a donc pour  $a, b$  finis :

$$P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b) = \int_a^b f(x) dx.$$

– La fonction de répartition de la variable  $X$  est définie par

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(x) dx.$$

La fonction  $F$  a alors les propriétés suivantes :

- Elle est continue et dérivable en tout point où  $f$  est continue.
- Si  $f(a^-)$  et  $f(a^+)$  existent, les dérivées à gauche et à droite de  $F$  :  $F'(a^-)$  et  $F'(a^+)$  existent et

$$F'(a^-) = f(a^-) ; F'(a^+) = f(a^+).$$

La fonction de répartition est donc continue alors que dans le cas discret elle admet une discontinuité en tout point de  $X(\Omega)$ .

*Exemple :* Soient  $-\infty < a < b < \infty$ , on définit la fonction

$$f(x) = \frac{1}{b-a} \text{ si } x \in [a, b] ; 0 \text{ sinon}$$

on note cette fonction

$$f(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x).$$

On vérifie facilement que cette fonction est bien une densité de probabilité. La probabilité ou loi associée est appelée loi uniforme continue sur l'intervalle  $[a, b]$ .

## 6.2 Espérance d'une fonction d'une variable aléatoire

Soit  $I$  un intervalle réel.

$$E(\mathbb{1}_I(X)) = P(X \in I) = \int_I f(x) dx$$

comme nous l'avons vu dans la partie concernant les probabilités discrètes.

Soit maintenant  $g$  une fonction en escalier,

$$g(x) = \sum_{i=1}^n \alpha_i \mathbb{I}_{[a_i, b_i[}(x)$$

où les intervalles  $[a_i, b_i[$  sont disjoints. Alors, si  $g(X)$  ne prend qu'un nombre fini de valeurs, l'espérance est linéaire

$$\begin{aligned} E(g(X)) &\stackrel{\text{lin}}{=} \sum_{i=1}^n \alpha_i E(\mathbb{I}_{[a_i, b_i[}(X)) = \sum_{i=1}^n \alpha_i \int_{[a_i, b_i[} f(x) dx \\ &= \int_{\mathbb{R}} \sum_{i=1}^n \alpha_i \mathbb{I}_{[a_i, b_i[}(x) f(x) dx = \int g(x) f(x) dx. \end{aligned}$$

On généralise ce résultat en admettant le

**Théorème 6.2.1** *Soit  $X$  de densité  $f$ , pour toute fonction  $g$  continue ,*

$$E(g(X)) = \int_{-\infty}^{\infty} f(x)g(x)dx, \quad (6.2)$$

*à condition que l'intégrale généralisée soit absolument convergente. Réciproquement, si la relation (6.2) est vraie pour toute fonction  $g$  continue bornée, alors  $f$  est la densité de  $X$ .*

## 6.3 Changement de variables

Considérons le problème suivant :  $X$  est une variable aléatoire réelle de densité  $f(x)$ , on considère  $Y = h(X)$  où  $h$  est une certaine fonction.  $Y$  a t'elle une densité et que vaut elle?

On suppose qu'il existe  $-\infty \leq a < b \leq \infty$  tels que

- $f(x)$  est nulle en dehors de  $]a, b[$
- $h$  est dérivable strictement croissante de  $]a, b[$  vers  $]c, d[$ , ( $h(a) = c$  ;  $h(b) = d$ ).

Nous appliquons le théorème 6.2 sur les densités. Soit  $g$  continue bornée,

$$E(g(Y)) = E(g(h(X))) = E(g \bullet h(X)) = \int_a^b g \bullet h(x) f(x) dx.$$

Dans l'intégrale on fait le changement de variable

$$y = h(x) \ ; \ x = h^{-1}(y) \ ; \ dx = (h^{-1})'(y) dy.$$

On obtient donc

$$E(g(Y)) = \int_{h(a)}^{h(b)} g(y) f(h^{-1}(y)) (h^{-1})'(y) dy = \int_c^d g(y) (f(h^{-1}(y)) (h^{-1})'(y)) dy$$

Par le théorème des densités (6.2), la densité de Y est donc

$$u(y) = f(h^{-1}(y)) (h^{-1})'(y).$$

Quand h est décroissante, on obtient le même résultat avec un changement de signe et  $h(a) = d$  ;  $h(b) = c$ . Au total on a bien montré

**Proposition 6.3.1** *Soit X une variable aléatoire de densité  $f(x)$  nulle en dehors de  $]a, b[$ ,  $-\infty \leq a < b \leq \infty$  et soit h une application dérivable, strictement monotone de  $]a, b[$  vers  $]c, d[$ , alors  $Y = h(X)$  est une variable aléatoire réelle de densité*

$$u(y) = f(h^{-1}(y)) |(h^{-1})'(y)| \mathbb{I}_{]c, d[}(y).$$

*Exemple :* Si X est de densité  $f(x)$  quelle est la densité de  $Y = \exp(X)$ ?

- $h = \exp$  est bijective de  $] - \infty, +\infty[$  vers  $]0, +\infty[$
- $h^{-1} = \log$  ;  $(h^{-1})'(y) = 1/y$ .
- La densité de Y est donc  $u(y) = f(\log(y)) 1/y \mathbb{I}_{]0, +\infty[}(x)$ .

# Chapitre 7

## Observations vectorielles

### 7.1 Introduction

Il existe beaucoup de situations où les mesures physiques ne sont pas isolées. Par exemple on peut mesurer

- un rayonnement dans plusieurs longueurs d'ondes,
- Une longueur et une largeur de feuilles,
- Des coordonnées cartésiennes ou polaires d'un point.

On modélisera cela en disant que l'on observe un variable vectorielle  $X$  à valeurs dans  $\mathbb{R}^n$ . On se limitera la plupart du temps à  $n = 2$ .

**Définition 7.1.1 Intégrale multiple.** *Soit  $f$  une fonction de deux variables, continue par morceaux. Si  $f$  est positive on admet que*

$$\int_{\mathbb{R}} \left[ \int_{\mathbb{R}} f(x, y) dy \right] dx = \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} f(x, y) dx \right] dy$$

la valeur pouvant être infinie. Elle est notée

$$\iint_{\mathbb{R}^2} f(x, y) dx dy.$$

Si  $f$  change de signe et si

$$\iint_{\mathbb{R}^2} |f(x, y)| dx dy < +\infty,$$

on dit que  $f$  est intégrable et on a encore :

$$\int_{\mathbb{R}} \left[ \int_{\mathbb{R}} f(x, y) dy \right] dx = \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} f(x, y) dx \right] dy$$

cette valeur est notée

$$\iint_{\mathbb{R}^2} f(x, y) dx dy.$$

Si

$$\iint_{\mathbb{R}^2} |f(x, y)| dx dy = +\infty,$$

on dit que l'intégrale

$$\iint_{\mathbb{R}^2} f(x, y) dx dy$$

n'est pas définie.

**Définition 7.1.2** La variable aléatoire vectorielle  $X$  a pour densité  $f(x, y)$  ssi pour tout couple  $]a, b[ \times ]c, d[$  d'intervalles éventuellement infinis, on a

$$P(X \in ]a, b[ \times ]c, d[) = \iint_{]a, b[ \times ]c, d[} f(x, y) dx dy$$

où  $f$  est une fonction positive, continue par morceaux dont l'intégrale sur  $\mathbb{R}^2$  vaut 1.

De la même façon que dans le cas univariable on montrerait le théorème :

**Théorème 7.1.1** Si le vecteur  $X$  a pour densité  $f(x, y)$  et si  $g$  est une fonction continue,

$$E(g(X)) = \iint_{\mathbb{R}^2} g(x, y) f(x, y) dx dy$$

à condition que l'intégrale double existe.

Ce théorème sera utilisé pour le calculs de moments.

**Lois marginales** Si  $X = (X_1, X_2)$  a pour densité  $f(x, y)$  alors

–  $X_1$  a pour densité

$$f_1(x) = \int_{\mathbb{R}} f(x, y) dy$$

–  $X_2$  a pour densité

$$f_2(y) = \int_{\mathbb{R}} f(x, y) dx.$$

Il suffit pour le vérifier de se ramener à la définition de la densité jointe en prenant un des intervalles égaux à  $\mathbb{R}$ .

**Covariance** La covariance se définit comme dans le cas discret. Si  $(X_1, X_2)$  est de densité  $f(x_1, x_2)$  et

$$m_i = E(X_i) = \iint_{\mathbb{R}^2} x_i f(x_1, x_2) dx_1 dx_2 \quad i = 1, 2$$

on a alors

$$Cov(X_1, X_2) = E[(X_1 - m_1)(X_2 - m_2)] = \iint_{\mathbb{R}^2} (x_1 - m_1)(x_2 - m_2) f(x_1, x_2) dx_1 dx_2.$$

On obtient exactement les mêmes formules de calcul de variance que dans le cas discret.

## 7.2 Indépendance

**Définition 7.2.1** Soit  $X = (X_1, X_2)$  un vecteur aléatoire, on dit que ses composantes sont indépendantes ssi pour tout couple d'intervalles  $I_1$  et  $I_2$  on a

$$P(X_1 \in I_1; X_2 \in I_2) = P(X_1 \in I_1)P(X_2 \in I_2)$$

On admet le

**Théorème 7.2.1** Soit  $X = (X_1, X_2)$  un vecteur de densité  $f(x, y)$ , de densités marginales  $f_1(X_1)$  et  $f_2(X_2)$ , les variables  $X$  et  $Y$  sont indépendantes ssi

$$f(x, y) = f_1(x)f_2(y)$$

*Remarque :* Comme dans le cas discret on a  $X, Y$  indépendants  $\Rightarrow f_1(X), f_2(Y)$  indépendants  $\Rightarrow E(f_1(X)f_2(Y)) = E(f_1(X))E(f_2(Y))$ . La dernière égalité se vérifie en remarquant que l'intégrale se factorise, la première implication sera par contre admise.

## 7.3 Densités conditionnelles

On fait un raisonnement en utilisant des infiniment petits  $dx$  et  $dy$ . Soient  $X, Y$  de densité jointe  $f(x, y)$  et soit  $f_1(x)$  la densité marginale de  $X$ . Soient  $dx$  et  $dy$  deux "infiniment petits"

$$\begin{aligned} P(y \leq Y \leq y + dy | x \leq X \leq x + dx) &= \frac{P(y \leq Y \leq y + dy \text{ et } x \leq X \leq x + dx)}{P(x \leq X \leq x + dx)} \\ &= \frac{f(x, y)dx dy}{f_1(x)dx} = \frac{f(x, y)}{f_1(x)} dy \end{aligned}$$

Pour prendre en compte les problèmes de division par zéro aux points où  $f_1$  s'annule, on pose la définition suivante.

**Définition 7.3.1** Soit  $(X, Y)$  un couple de variables de densité  $f(x, y)$  et de densité marginale en  $X : f_1(x)$ . Soit  $x$  fixé, on appelle densité conditionnelle de  $Y$  sachant  $X = x$  la fonction de  $y$  définie par

$$f(y|X = x) = \frac{f(x, y)}{f_1(x)} \quad \text{si } f_1(x) \neq 0 \quad ; \quad 0 \text{ sinon}$$

La valeur que l'on donne à la fonction quand  $f_1(x) = 0$  est sans importance car  $X$  ne prend "jamais" ces valeurs.

# Chapitre 8

## Régression linéaire et théorème limites

### 8.1 Théorèmes limites

**Définition 8.1.1 Convergence en probabilité** . Une suite  $A_n$  de variables aléatoires converge en probabilité vers  $l \in \mathbb{R}$  ssi

$$\text{pour tout } \epsilon > 0 : P(|A_n - l| > \epsilon) \rightarrow 0, n \rightarrow +\infty$$

Il faut penser que  $\epsilon$  est tout petit pour comprendre le sens de cette définition.

**Définition 8.1.2 Convergence en moyenne quadratique**. Une suite  $A_n$  de variables aléatoires converge en moyenne quadratique vers  $l \in \mathbb{R}$  ssi

$$E((A_n - l)^2) \rightarrow 0, n \rightarrow +\infty$$

**Proposition 8.1.1** La convergence en moyenne quadratique implique la convergence en probabilité .

*Démonstration*: On applique l'inégalité de Markov à  $(A_n - l)^2$

$$P(|A_n - l| > \epsilon) = p((A_n - l)^2 > \epsilon^2) \leq 1/\epsilon^2 V_n \rightarrow 0.$$

□

La convergence en moyenne quadratique est souvent un moyen commode de démontrer la convergence en probabilité. C'est le cas pour la démonstration de la loi des grands nombres. Cependant il existe des exemples où on a convergence en probabilité sans avoir convergence en moyenne quadratique.

**Proposition 8.1.2 Loi des Grands Nombres**. Soient  $X_1, \dots, X_n$  une suite de variables aléatoires réelles indépendantes et de même loi admettant un moment d'ordre 2, alors  $\bar{X}_n = 1/n(X_1 + \dots + X_n)$  converge en moyenne quadratique et donc en probabilité vers  $E(X_1)$ .

*Démonstration:*

$$E(\overline{X_n}) = E(X_1) \quad ; \quad \text{var}(\overline{X_n}) = \text{var}(X_1)/n$$

ce qui implique la convergence en moyenne quadratique.

□

**Définition 8.1.3 Convergence en loi.** Une suite  $A_n$  de variables aléatoires converge en loi vers un variable aléatoire  $B$  ssi

$$P(A_n \leq x) \rightarrow F_B(x) \quad , \quad n \rightarrow +\infty$$

en tout point où la fonction de répartition  $F_B$  de  $B$  est continue.

Rappelons que si  $B$  admet une densité sa fonction de répartition  $F_B$  est continue en tout point.

Nous sommes donc maintenant en mesure d'énoncer le théorème central limite qui donne la vitesse de convergence de  $\overline{X_n}$  vers  $E(X_1)$

**Théorème 8.1.1** Soient  $X_1, \dots, X_n$  une suite de variables aléatoires réelles indépendantes et de même loi admettant un moment d'ordre 2, On pose

$$S_n = X_1 + \dots + X_n$$

alors

$$\frac{S_n - nE(X_1)}{\sqrt{n\text{Var}(X_1)}} \xrightarrow{\text{loi}} N(0, 1).$$

Dans le cas particulier où les variables suivent des lois de Bernoulli, on obtient le théorème de Moivre-Laplace qui montre la convergence de la loi binomiale, convenablement normalisée, vers la loi normale.

## 8.2 Régression linéaire

On considère deux variables aléatoires  $X$  et  $Y$  qui admettent un moment d'ordre 2 et une loi jointe. On désire approcher  $Y$  par une combinaison affine de  $X$  :  $aX + b$  au sens de l'erreur quadratique minimum. Plus précisément on va chercher les valeurs  $\hat{a}$  et  $\hat{b}$  telles que

$$r(a, b) = E[Y - (aX + b)]^2$$

soit minimum au point  $a = \hat{a}$ ,  $b = \hat{b}$ .

Remarquons que la quantité que l'on minimise va être une variance puisqu'au minimum, la variable dont on prend le moment d'ordre 2 sera centrée.

### Solution du problème.

On suppose que la variance de  $X$  est strictement positive car sinon le problème est sans intérêt. On reparamétrise en posant :

$$aX + b = a(X - E(X)) + b + aE(X) = aZ + \beta.$$

La variable  $Z$  est maintenant centrée. En développant :

$$R(a, \beta) = E(Y^2) + a^2 E(Z^2) + \beta^2 - 2aE(ZY) - 2\beta E(Y).$$

On cherche les solutions en dérivant :

$$\frac{\partial R(a, \beta)}{\partial \beta} = 2\beta - 2E(Y) = 0$$

$$\frac{\partial R(a, \beta)}{\partial a} = 2aE(Z^2) - 2E(ZY) = 0$$

on trouve donc comme solution

$$\hat{a} = \frac{E(ZY)}{E(Z^2)} = \frac{Cov(X, Y)}{var(X^2)}$$

$$\hat{\beta} = E(Y) \Rightarrow \hat{b} = E(Y) - aE(X).$$

La fonction de régression de  $Y$  sur  $X$  est donc

$$\frac{Cov(X, Y)}{Var(X^2)}(X - E(X)) + E(Y)$$

Le minimum  $R(\hat{a}, \hat{b})$  vaut

$$Var(Y) - \frac{Cov^2(X, Y)}{Var(X)}.$$

On suppose maintenant que  $Var(Y) > 0$  pour définir le coefficient de corrélation linéaire entre  $X$  et  $Y$

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{var(X)Var(Y)}}$$

on a alors

$$R(\hat{a}, \hat{b}) = var(Y)[1 - \rho^2]$$

*Remarques :*

Le coefficient  $\rho$  est à valeurs dans  $[-1, 1]$

Si on essaye d'approcher  $Y$  par une seule constante  $b$ , nous avons déjà vu que le minimum de  $R$  vaut :  $Var(X)$ , la valeur  $\rho^2$  apparait donc comme le pourcentage de variance expliquée par  $Y$ .

# Chapitre 9

## Statistique

### 9.1 Généralités sur la statistique

La statistique poursuit deux buts:

- résumer de manière synthétique un corpus de données, on parle alors de statistique descriptive ou exploratoire,
- confronter ces données à un modèle probabiliste, on parle alors de statistique inférentielle.

Nous allons décrire très brièvement ces deux aspects. A part dans les cas de recensement, la statistique travaille sur deux notions qui convient de bien distinguer.

#### 9.1.1 La population et l'échantillon

- L'échantillon est un ensemble explicitement connu d'individus ou de données ayant le même statut. Schématiquement il correspond à un fichier informatique. A titre d'exemple, on peut citer
  - un ensemble de trente arbres mesurés au hasard dans un forêt,
  - un ensemble de 1000 sondés interrogés dans le cas d'un sondage d'opinion.

Cet échantillon est supposé être issu et représenter un ensemble plus grand et inaccessible :

- La population ( $\Omega$ ) qui est dans les cas ci dessus
  - tous les arbres de la forêt,
  - les 34 millions d'électeurs français.

Remarquez que la notation n'est pas anodine puisque la population est la plus souvent l'ensemble des réalisations d'un espace de probabilité.

### 9.1.2 Différent types de variables

Après avoir bien identifié l'échantillon et la population, il faut classer les types d'observations ou variables (aléatoires). Cette classification sera en dernier lieu subjective. On appellera

- Variable quantitative toute variable qui exprime une quantité en unité physique ou de comptage. C'est le cas systématiquement de tout ce qui s'exprime en Frs, Kg, Volts, mètres, années etc...

Pour une variable quantitative, 4 est 4 fois plus que 1.

Une variable quantitative peut prendre un nombre infini de valeurs dans un intervalle réel : on parle alors de variable quantitative continue. Elle peut également prendre qu'un petit nombre de valeurs, par exemple le nombre d'enfants d'une famille, on parle alors de variable quantitative discrète. Elle peut enfin être observée seulement à travers son appartenance à diverses classes on parle alors de variable continue discrétisée.

- Variable qualitative ordinale. Il s'agit d'une variable qui prend un nombre fini de valeurs qui sont ordonnées mais qui n'expriment pas une quantité. Il s'agit par exemple de rangs, de degrés d'évolution de maladies. Pour une variable qualitative ordinale, 4 n'est pas 4 fois 1 mais est plus grand que 1.
- Variable qualitative nominale. Il s'agit d'un variable prenant un nombre fini de valeurs sans hiérarchie. Par exemple une catégorie socio professionnelle, une couleur, un numéro de lot. Pour un variable qualitative nominale, 4 n'est ni plus ni moins que 1.

## 9.2 Étude d'une variable quantitative continue

Nous aborderons dans cette partie seulement le cas d'une variable quantitative continue c'est à dire observée avec suffisamment de précision pour que l'on puisse considérer qu'elle est observée de manière exacte.

### 9.2.1 Indicateurs de position ou indicateurs centraux

Pour situer en premier lieu la position d'une variable, on se sert principalement de deux indicateurs:

- La moyenne  $\bar{X} = 1/n \sum_i X_i$
- La médiane empirique  $X_{1/2}$ . C'est la donnée qui sépare en deux l'échantillon,

$$X_{1/2} = X_{(m+1)}, \text{ si } n \text{ est impair } = 2m + 1 \quad ; \quad \frac{X_{(m)} + X_{(m+1)}}{2} \text{ si } n = 2m$$

où  $X_{(i)}$  est la donnée qui a le rang  $i$ .

## 9.2.2 Indicateurs de dispersion

- La variance empirique  $s^2 = 1/(n - 1) \sum (X_i - \bar{X})^2$
- l'intervalle interquartile. On définit les quartiles  $X_{1/4}$  et  $X_{3/4}$  qui coupent l'échantillon en 1/4, 3/4 (il y a plusieurs méthodes de gestion des arrondis qui sont sans importance). L'intervalle interquartile vaut  $X_{1/4} - X_{3/4}$ .
- l'étendue  $X_{max} - X_{min}$  (dépend beaucoup de  $n$ ),
- l'écart moyen à la médiane

$$\sum_i |X_i - X_{1/2}|$$

- l'écart type empirique  $s = \sqrt{s^2}$ .

## 9.2.3 Représentations graphique

- La boîte à moustache qui représente dans un même graphique la médiane les quartiles, le min et le max.
- Le "stem and leaf" (feuille et branche) qui crée une ligne par valeur entière possible de la variable et qui les remplit avec les premières décimales.
- L'histogramme. Soit  $A_1, \dots, A_n$  une partition de  $\mathbb{R}$  ou d'une sous partie, en  $n$  sous intervalles. Soit  $n_i$  le nombre d'observations dans l'intervalle  $A_i$ . L'histogramme consiste en la représentation par un diagramme en bâtons de la *densité d'effectif* définie par

$$\text{à } A_i \text{ on associe } \frac{n_i}{|A_i|}$$

$|A_i|$  est la taille de l'intervalle.

## 9.3 Échantillonnage quantitatif dans une population

On considère une population  $\Omega$  (les arbres de la forêt) et une variable aléatoire  $X(\omega)$  (taille de l'arbre  $\omega$  tiré au hasard). On suppose que  $E(X^2) = \sigma^2$  existe. On veut savoir si la vraie espérance (la hauteur moyenne de tous les arbres de la forêt)  $\mu$  vaut une certaine valeur donnée  $\mu_0$ . Pour cela on va construire une procédure de prise de décision appelée **test** basée sur deux hypothèses

- l'hypothèse générale :  $\mu \neq \mu_0$

– l’hypothèse particulière ou nulle  $H_0 : \mu = \mu_0$ .

Pour décrire cette procédure nous avons besoin de trois lemmes.

**Lemme 9.3.1** *Si  $A_n$  est une suite de variables aléatoires qui converge en probabilité vers  $l$  et si  $g$  est une fonction continue en  $l$ , alors  $g(A_n)$  converge en probabilité vers  $g(l)$ .*

*Démonstration* :Ecrivons que  $g$  est continue,

$$\forall \epsilon > 0 \exists \eta / |A_n - l| < \eta \Rightarrow |g(A_n) - g(l)| < \epsilon$$

C’est à dire

$$\{|A_n - l| < \eta\} \subset \{|g(A_n) - g(l)| < \epsilon\}$$

où encore

$$\{|A_n - l| \geq \eta\} \supset \{|g(A_n) - g(l)| \geq \epsilon\}$$

donc

$$P\{|g(A_n) - g(l)| \geq \epsilon\} \leq P\{|A_n - l| \geq \eta\} \rightarrow 0.$$

□

**Lemme 9.3.2** *Si  $A_n$  est une suite de variables aléatoires qui tend en probabilité vers 1 et si  $B_n$  est une suite de variables aléatoires qui tend en loi vers  $B$ , alors  $A_n B_n$  tend en loi vers  $B$ .*

*Démonstration* : On se limite au cas où les fonctions de répartition de  $B : F_B$  et  $B_n : F_{B_n}$  sont continues. Soit  $x$  quelconque, il suffit de montrer que

$$P\{A_n B_n \leq x\} \rightarrow F_B(x).$$

Pour ce faire on pose

$$E_n = \{A_n \in [1 - \epsilon, 1 + \epsilon]\}$$

On note  $E_n^c$  son complémentaire, on a alors

$$\{B_n \leq \frac{x}{1 + \epsilon}\} \subset \{A_n B_n \leq x\} \cup E_n^c$$

donc

$$P\{A_n B_n \leq x\} \geq F_{B_n}(\frac{x}{1 + \epsilon}) - P(E_n^c)$$

Comme  $P(E_n^c)$  tend vers zéro cela donne la majoration d’un coté. Pour l’autre

$$\{A_n B_n \leq x\} \subset \{B_n \leq \frac{x}{1 - \epsilon}\} \cup E_n^c$$

et on raisonne comme précédemment. □

**Lemme 9.3.3** *Soient  $X_1 \dots X_n$ ,  $n$  observations indépendantes et de même loi admettant un moment d’ordre deux, l’estimateur de la variance  $s^2$  converge en probabilité vers la variance  $\sigma^2$ .*

*Démonstration :*

$$s^2 = 1/(n-1) \left[ \sum_i X_i^2 - 1/n(\sum_i X_i)^2 \right].$$

Par le lemme (9.3.2) on peut remplacer  $1/(n-1)$  par  $1/n$  sans rien changer. Par la loi des grands nombres appliquée à  $X^2$  le premier terme tend vers  $E(X^2)$  et par la loi des grands nombres appliquée à  $X$  le second terme tend vers  $(E(X))^2$  ce qui démontre le résultat. Nous pouvons maintenant énoncer le principal résultat.

**Théorème 9.3.1** *Définissons la statistique*

$$T_n^h = \frac{\bar{X}_n - h}{\sqrt{s^2/n}}$$

alors,

- sous l'hypothèse  $H_0 : \mu = \mu_0$ ,  $T_n^{\mu_0}$  converge en loi vers une loi normale  $N(0,1)$ .
- sous l'hypothèse générale:  $\mu \neq \mu_0$   $T_n^{\mu_0}$  tend vers l'infini en probabilité.

*Démonstration :* Plaçons nous sous l'hypothèse nulle, alors

$$T_n^{\mu_0} = \frac{\bar{X}_n - \mu_0}{\sqrt{\sigma^2/n}} \frac{\sigma}{s}$$

Le théorème central limite implique la convergence en loi du premier terme vers une  $N(0,1)$ , les lemmes (9.3.1 et 9.3.3) impliquent la convergence de  $\frac{\sigma}{s}$  vers 1 et donc maintenant le lemme (9.3.2) donne le résultat.

Sous l'hypothèse générale :

$$T_n^{\mu_0} = T_n^\mu + \sqrt{n}(\mu - \mu_0).$$

□

Nous pouvons construire maintenant la procédure de test de  $H_0$ . On se donne une valeur numérique  $\alpha$  que l'on considère comme faible pour une probabilité. On définit une région de plausibilité pour une variable aléatoire de loi normale  $(0,1)$ . Soit  $Z_\alpha$  le fractile  $1 - \alpha/2$  de la loi  $N(0,1)$  défini par

$$P(X < Z_\alpha) = 1 - \alpha/2 \quad \text{avec } X \text{ de loi } N(0,1).$$

On alors

$$P\{|X| > Z_\alpha\} = \alpha.$$

La règle de décision est alors la suivante, si

$$|T_n^{\mu_0}| > Z_\alpha$$

on considère que cette valeur est "trop grande" pour être une réalisation de loi normale et donc on rejette  $H_0$ .

Dans le cas contraire  $H_0$  est acceptable.

## 9.4 Échantillonnage pour une variable qualitative

Dans le cas d'une variable qualitative, on calcule les fréquences sur l'échantillon ( fréquences empiriques) qui sont des quantités qui estiment les vraies probabilités sur la population entière. On peut représenter ces fréquences par des diagrammes en "camemberts" en bâtons ou barres.

**Test de comparaison à une probabilité théorique** Considérons le problème suivant : sur 1000 sondés, 280 personnes connaissent la lessive "Y". La vraie fréquence dans la population peut-elle être de 0.25? On construit le modèle suivant :  $X_1, \dots, X_n$  sont  $n$  (1000) observations indépendantes de variable aléatoires de Bernoulli qui prennent la valeur 1 si la personne connaît le produit et 0 sinon. On note  $S$  le nombre de 1 sur les  $n$  sondés.  $S$  suit une loi binomiale de paramètres  $n$  et  $p$ .

On considère

- l'hypothèse nulle  $H_0 : p = p_0$
- l'hypothèse générale  $H_1 : p \neq p_0$ .

où  $p_0$  est une certaine valeur (0.25 dans notre exemple). Si on note  $f$  la fréquence empirique, cette fréquence est la moyenne des  $X_i$ . On a de plus

$$E(f) = p \quad Var(f) = \frac{p(1-p)}{n}.$$

Par les mêmes argument que pour le théorème (9.3.1) on montre que sous  $H_0 : p = p_0$

$$Z^{p_0} = \frac{f - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \rightarrow N(0, 1) \text{ en loi.}$$

Cette même expression tend vers l'infini sous l'hypothèse générale.

De même que précédemment on définit la règle de test

- $|Z^{p_0}| > Z_\alpha$  on rejette  $H_0$ ; le test est dit significatif.
- $|Z^{p_0}| \leq Z_\alpha$   $H_0$  est acceptable; Le test est dit non significatif.

Dans l'exemple numérique on trouve avec  $\alpha = 0.05$ ,  $p_0 = 0.25$ ,  $Z^{0.25} = 2.19$ ,  $Z_\alpha = 1.96$  on conclut au rejet de  $H_0$ . La probabilité diffère de 0.25.