Ricco Rakotomalala

Probabilités et Statistique

Notes de cours

Université Lumière Lyon 2

Avant-propos

Ce document est un support de cours pour les enseignements des probabilités et de la statistique. Il couvre l'analyse combinatoire, le calcul des probabilités, les lois de probabilités d'usage courant et les tests d'adéquation à une loi.

Ce support correspond approximativement aux enseignements en L2 de la filière Sciences Économiques et Gestion, Administration Économique et Sociale (AES).

Un document ne vient jamais du néant. Pour élaborer ce support, je me suis appuyé sur différentes références, des ouvrages reconnus dans la discipline, mais aussi des ressources en ligne qui sont de plus en plus présents aujourd'hui dans la diffusion de la connaissance.

Les seuls bémols par rapport à ces documents en ligne sont le doute que l'on pourrait émettre sur l'exactitude des informations prodiguées, mais la plupart de leurs auteurs sont des enseignants-chercheurs qui font sérieusement leur travail ; une disponibilité plus ou moins aléatoire, au gré des migrations des serveurs et de la volonté de leurs auteurs, auquel il est très difficile de remédier ; les informations sont disparates, avec une absence d'organisation, à la différence des ouvrages qui suivent une ligne pédagogique très structurante.

Néanmoins, ces ressources en ligne renouvellent profondément le panorama des documents disponibles pour les enseignements. La gratuité n'est pas le moindre de leurs atouts.

Les références que nous avons utilisées pour la composition de ce support sont détaillées dans la bibliographie. Les excellents ouvrages de Grais (Chapitres 1 à 3, pages 1 à 168) et de Saporta (Chapitres 1 et 2, pages 1 à 65), en particulier, m'ont beaucoup aidé à structurer ce document.

Enfin, selon l'expression consacrée, ce texte n'engage que son auteur. Toutes suggestions ou commentaires qui peuvent l'améliorer sont le bienvenu.

Table des matières

Partie I Introduction aux méthodes de calcul des probabilités

Partie I

Introduction aux méthodes de calcul des probabilités

1

Éléments d'analyse combinatoire

1.1 Généralités

L'analyse combinatoire (A.C.) est le dénombrement des dispositions que l'on peut former à l'aide des éléments d'un ensemble fini.

1.1.1 L'ensemble étudié : éléments discernables et éléments indiscernables

Représentons les éléments d'un ensemble par $e_1, e_2, \ldots, e_n$. L'ensemble Ω comporte n éléments, c.-à-d. $card(\Omega) = n$.

Si e_i et e_j sont équivalents, on dit qu'ils sont indiscernables. Si e_i et e_j sont différents, on dit qu'ils sont discernables.

L'ensemble Ω de n éléments peut être constitué d'éléments discernables 2 à 2. Ex. $\{a, b, c, d, e\}$. Tous les éléments de Ω peuvent aussi être tous indiscernables. Ex. $\{a, a, a, a, a\}$.

Les éléments d'un ensemble peuvent être discernables ou indiscernables. Ex.$\Omega = \{a, b, a, a, c, d, d, c, a\}$, $card(\Omega) = 9$.

1.1.2 Les différentes dispositions

Une disposition est l'ensemble formé d'éléments choisis parmi les n éléments de l'ensemble étudié. Un élément figurant dans une disposition est caractérisé par :

- le nombre de fois où il figure dans l'ensemble ;
- sa place dans la disposition.

Exemple 1. Soit un ensemble de 4 cartes $\{9^p, 9^t, 7^p, 7^c\}$. La hauteur 9 se répète 2 fois, le 9^p se trouve en première position.

Quelques définitions

Définition 1. ***Disposition sans répétition.*** *C'est une disposition où un élément peut apparaître* 0 *ou* 1 *fois.*

Définition 2. ***Disposition avec répétition.*** *Un élément peut figurer plus d'une fois.*

Définition 3. ***Disposition ordonnée.*** *L'ordre d'obtention d'un élément est important. Ex. les éléments constituant la plaque minéralogique d'un véhicule.*

Définition 4. ***Disposition non-ordonnée.*** *L'ordre d'obtention d'un élément n'est pas important, on n'en tient pas compte dans la caractérisation de la disposition. Ex. Les numéros issu d'un tirage du loto.*

Exemple 2. Prenons un jeu de dé à 6 faces (éléments discernables) $\Omega = \{1, 2, 3, 4, 5, 6\}$. Après 3 jets, nous obtenons $A = (2, 5, 1)$; nous réitérons les jets et nous obtenons $B = (5, 1, 2)$. A et B sont équivalents si nous considérons que les dispositions sont non-ordonnées. En revanche, ils ne sont pas équivalents si nous sommes dans le cadre d'une disposition non-ordonnée.

1.2 Formules classiques d'analyse combinatoire

1.2.1 Les paires

Soient 2 ensembles constitués respectivement de α éléments discernables et β éléments discernables : $A = \{a_1, a_2, a_3, \ldots, a_\alpha\}$ et $B = \{b_1, b_2, b_3, \ldots, b_\beta\}$.

On appelle une paire une disposition de 2 éléments dont le 1^{er} appartient à A, et le 2^{nd} appartient à B. La paire est une disposition ordonnée, mais les 2 éléments sont de nature différentes.

Pour décrire l'ensemble des paires possibles, il suffit de composer chaque élément de A avec chaque élément de B. Il y a donc $M = \alpha \times \beta$ paires possibles.

Exercice 1. Combien y a-t-il de mots de 2 lettres ? Il y a 26 lettres dans l'alphabet. On peut donc former $26 \times 26 = 676$ mots de 2 lettres.

Exercice 2. Combien y a-t-il de mots de 2 lettres formés d'une consonne et d'une voyelle (sans tenir compte de l'ordre) ? Il y a 20 consonnes et 6 voyelles. Nous pouvons former $20 \times 6 = 120$ couples "‘consonne $\times$ voyelle"’ et, $6 \times 20 = 120$ couples "‘voyelle $\times$ consonne"’. Il y a donc 240 mots possibles formés par une consonne et une voyelle.

1.2.2 Les multiplets

Soient λ ensembles distincts formés d'éléments complètement discernables.

$$
\begin{aligned}
A &= (a_1, a_2, \ldots, a_\alpha) \\
B &= (b_1, b_2, \ldots, b_\beta) \\
C &= (c_1, c_2, \ldots, b_\rho) \\
&\cdots \\
S &= (s_1, s_2, \ldots, s_\lambda)
\end{aligned}
$$

On appelle multiplets un disposition ordonnée de λ éléments dont le 1er est un élément de A, le second un élément de B,..., et le λème un élément de S.

Les multiplets sont de la forme $(a_{i_1}, b_{i_2}, c_{i_\rho}, \ldots, s_{i_\lambda})$.

Un multiplet de λ éléments peut être considéré comme une paire où : le 1er élément est constitué des $(\lambda - 1)$ éléments, et le second élément, un élément de S.

$$(a, b, c, \ldots, s) = [(a, b, c, \ldots), s]$$

Pour obtenir le nombre de multiplets, nous pouvons raisonner par récurrence.

$$\begin{aligned} M_\lambda &= M_{\lambda-1} \times \lambda \\ &= M_{\lambda-2} \times \rho \times \lambda \\ &= \cdots \\ &= \alpha \times \beta \times \cdots \times \lambda \end{aligned}$$

Exercice 3. Quelle est la capacité d'un code constitué de mots où les 2 premiers symboles sont des lettres de l'alphabet, et les 2 derniers symboles sont des chiffres ? Réponse : $M = 26 \times 26 \times 10 \times 10 = 67600$

1.2.3 Arrangement avec répétition

Soit Ω l'ensemble fondamental (référentiel, ensemble de référence, population mère) composé de n éléments : $card(\Omega) = n$.

Nous constituons une échantillon ω de taille p : $card(\omega) = p$.

Si nous avons à choisir p éléments parmi n, **la disposition étant ordonnée et avec répétition**, on dit qu'on a un arrangement de p éléments parmi n.

$$\mathcal{A}_n^p = n^p \tag{1.1}$$

N.B. Il est possible que $p > n$.

Exemple 3. Dans un pays imaginaire, un numéro de téléphone comporte 5 chiffres. Il doit commencer par 0, le second chiffre est compris entre 1 et 5, il indique la région. Les autres chiffres sont libres. Combien de numéros de téléphones différents peut-on former dans ce pays ?

$\Omega = \{0, 1, \ldots, 9\}$; $\omega = \{a, b, c, d, e\}$. Pour les 3 derniers chiffres, le nombre d'arrangements possible est $\mathcal{A}_{10}^3 = 10^3 = 1000$. La capacité totale est $1 \times 5 \times \mathcal{A}_{10}^3 = 1 \times 5 \times 1000 = 5000$.

1.2.4 Arrangement sans répétition

Soit Ω une population mère, avec $card(\Omega) = n$. On constitue un échantillon de taille p, la disposition est **ordonnée et sans répétition**. On dit qu'on a un arrangement sans répétition de p éléments parmi n.

$$A_n^p = \frac{n!}{(n-p)!} \tag{1.2}$$

Définition 5 (Factorielle n). *Factorielle n, notée $n!$ est définie par $n! = \prod_{i=1}^{n} i = 1 \times 2 \times \ldots \times n$. Par convention $0! = 1$. Nous pouvons également utiliser une définition[1] récursive $n! = n \times (n-1)!$.*

[1] Pour plus d'informations, voir `http://fr.wikipedia.org/wiki/Factorielle`

Exemple 4. Dans notre pays imaginaire ci-dessus, combien y a-t-il de numéros comportant des chiffres tous différents ?

$$01abc \rightarrow A_8^3 = 336$$
$$02abc \rightarrow A_8^3 = 336$$
$$03abc \rightarrow A_8^3 = 336$$
$$04abc \rightarrow A_8^3 = 336$$
$$05abc \rightarrow A_8^3 = 336$$

$$\text{Total} = 1680$$

N.B. : Attention, dans $01abc$ par exemple, les chiffres (a, b et c) ne peuvent provenir que de $\Omega - \{0,1\}$, d'où la formule de l'arrangement A_8^3.

1.2.5 Permutation sans répétition

C'est un arrangement sans répétitions de n éléments parmi n.

$$P_n = A_n^p = \frac{n!}{(n-n)!} = n! \tag{1.3}$$

1.2.6 Permutation avec répétition

On appelle permutation avec répétition de p éléments où n sont distincts ($n \leq p$), une disposition ordonnée de l'ensemble de ces p éléments où le 1er figure p_1 fois, le second p_2 fois, etc., tel que $p_1 + p_2 + \cdots + p_n = p$.

$$\mathcal{P}_p^{(p_1, p_2, \cdots)} = \frac{p!}{p_1! p_2! \cdots} \tag{1.4}$$

Exemple 5. On jette successivement 12 dés. On appelle "‘résultat"’ une suite ordonnée de 12 points emmenés. Combien y a-t-il de résultats possibles ?

$\Omega = \{1,2,3,4,5,6\}$, $n = 6$, $p = 12$ éléments, ordonnés, avec répétition $\rightarrow \mathcal{A}_6^{12} = 6^{12}$

Exemple 6. Combien y a-t-il de résultats où chaque face est emmenée 2 fois ?

$n = 6$ éléments distincts, $p = 12$ éléments $\rightarrow \mathcal{P}_{12}^{(2,2,2,2,2,2)} = \frac{12!}{2!2!2!2!2!2!}$

Exemple 7. Combien y a-t-il de résultats où la face "‘1"’ se retrouve 5 fois, "‘2"’ 3 fois, "‘3"’ 3 fois, et "‘4"’ 1 fois ?

$\rightarrow \mathcal{P}_{12}^{(5,3,3,1,0,0)} = \frac{12!}{5!3!3!1!0!0!}$

1.2.7 Combinaison sans répétition

On considère une population mère Ω constitué de n éléments tous discernables. On forme un échantillon de taille p. Si la disposition est **non-ordonnée et sans répétition**, on dit que l'on a une combinaison sans répétition de p éléments parmi n.

$$C_n^p = \frac{n!}{p!(n-p)!} \tag{1.5}$$

Exemple 8. On tire 5 cartes dans un jeu de 32 cartes. Combien y a-t-il de résultats possibles ?
$C_{32}^5 = \frac{32!}{5!(32-5)!}$

1.2.8 Combinaison avec répétition

C'est une disposition non-ordonnée de p éléments, à choisir parmi n éléments discernables, avec répétition.

$$K_n^p = C_{n+p-1}^p \tag{1.6}$$

Exemple 9. Dans un jeu de domino, il y a 7 valeurs possibles $\{\text{blanc}, 1, 2, 3, 4, 5, 6\}$. Sur un domino, 2 valeurs sont inscrites, qu'importe l'ordre d'apparition des valeurs (retourner le domino n'a aucune influence sur ce qu'elle représente). Le nombre total de dominos que nous pouvons former est de $p = 2$ éléments choisis parmi $n = 7$, c.-à-d. $K_7^2 = C_{7+2-1}^2 = 28$.

1.2.9 Partition généralisée

Exemple introductif

On partage 52 cartes entre 4 joueurs (A, B, C, D). Combien y a-t-il de façons de distribuer ces cartes de manière à ce que :
- le joueur A reçoive 5 trèfles, 2 piques, 4 coeurs, 2 carreaux ;
- B reçoive $(4, 0, 7, 2)$;
- C reçoive $(3, 10, 0, 0)$;
- D reçoive $(1, 1, 2, 9)$.

Essayons de représenter cela dans un tableau.

-	A	B	C	D	Total
trèfle	5	4	3	1	13
carreau	2	2	0	9	13
coeur	4	7	0	0	12
pique	2	0	10	1	13
Total	13	13	13	13	52

- Considérons tout d'abord la répartition des trèfles. Le nombre de distribution des 13 trèfles suivant les joueurs est $\mathcal{P}_{13}^{(5,4,3,1)}$;
- Pour les carreaux, ce sera $\mathcal{P}_{13}^{(2,2,0,9)}$;
- Pour les coeurs, ce sera $\mathcal{P}_{13}^{(4,7,0,2)}$;
- Pour les piques, ce sera $\mathcal{P}_{13}^{(2,0,10,1)}$;

Cette répartition est effectuée simultanément, nous effectuons le produit pour obtenir le nombre total de possibilités :

$$M = \mathcal{P}_{13}^{(5,4,3,1)} \times \mathcal{P}_{13}^{(2,2,0,9)} \times \mathcal{P}_{13}^{(4,7,0,2)} \times \mathcal{P}_{13}^{(2,0,10,1)}$$

Remarque 1. Il est possible d'effectuer l'analyse duale ç.-à-d. détecter la façon pour A de recevoir 5 trèfles, 2 carreaux, 4 coeurs, 2 piques ; et ainsi de suite pour chaque joueur. Nous obtenons alors le nombre total de possibilité en effectuant également le produit. Nous obtiendrons les mêmes résultats.

Généralisation

Nous pouvons généraliser la solution ci-dessus en considérant la répartition de n objets classés en s catégories entre p joueurs. Le tableau de travail est le suivant :

-	A_1	$\cdots$	A_j	$\cdots$	A_p	Total
1	n_{11}	.	n_{1j}	.	n_{1p}	$n_{1.}$
$\ldots$	.	.	.	.	.	$\vdots$
i	n_{i1}	.	n_{ij}	.	n_{ip}	$n_{i.}$
$\ldots$	.	.	.	.	.	$\vdots$
s	n_{s1}	.	n_{sj}	.	n_{sp}	$n_{s.}$
Total	$n_{.1}$	.	$n_{.j}$	.	$n_{.p}$	$n_{..} = n$

Pour la ligne i, le nombre de dispositions possibles est $\frac{n_{i.}!}{n_{i1}!\cdots n_{ij}!\cdots n_{ip}!}$.

Ainsi, le nombre total de dispositions est obtenu avec :

$$\frac{\prod_{i=1}^{s} n_{i.}!}{\prod_{i=1}^{s}\prod_{j=1}^{p} n_{ij}!} \tag{1.7}$$

2

Définition de la probabilité

2.1 Notion de probabilité

Une épreuve est une expérience dont l'issue est incertaine. Les résultats éventuels d'une épreuve font généralement appel au hasard. L'ensemble des résultats éventuels (les résultats possibles, les éventualités) s'appelle **ensemble fondamental** (référentiel, ensemble de référence, population mère).

Exemple 10. Une pièce de monnaie possède deux figures (éventualités) : *pile* et *face*. Si la pièce n'est pas trafiquée et lancée loyalement, *pile* a autant de chances d'apparaître que *face*. On dit alors que les deux éventualités sont équiprobables.

A chaque élément de l'ensemble des éventualités, on peut associer un nombre, la probabilité d'obtenir l'éventualité.

2.2 Extension de la définition

Nous appelons évènement un sous-ensemble quelconque A_i de Ω, constitué de N_A éventualités équiprobables ç.-à-d. $A_i \subseteq \Omega$, et $card(A_i) = N_A$.

Une éventualité est un évènement élémentaire tel que la probabilité de réalisation de l'évènement A_i est $P(A_i \text{ se réalise}) = \frac{N_A}{N}$, où $card(\Omega) = N$.

Exemple 11. Ω est constitué d'un jeu de 32 cartes. Il comporte 8 hauteurs $\{7, 8, 9, 10, J, D, K, A\}$. Dans chaque hauteur, il y a 4 couleurs $\{pique, coeur, carreau, trefle\}$.

– $\rightarrow P\{\text{tirer la hauteur } 7\} = \frac{4}{32}$

– $\rightarrow P\{\text{tirer la couleur pique}\} = \frac{7}{32}$

2.3 Axiomes

Axiome d'uniformité. $P\{\text{éventualité équiprobable}\} = \frac{1}{N}$.
Domaine de définition. $0 \leq P\{\} \leq 1$.
Évènement certain. Si $A \equiv \Omega$, alors $P\{A\} = 1$.
Évènement impossible. Si $A \equiv \emptyset$, alors $P\{A\} = 0$.

2.4 Définition relative à des évènements : les diagrammes de Venn

2.4.1 Inclusion

Soit $A \subset \Omega$ un évènement (un ensemble). A est inclus dans Ω si chaque élément e de A appartient également à Ω.

$$\forall e \in A \Rightarrow e \in \Omega, \text{alors} A \subset \Omega$$

Propriétés de l'inclusion

Caractéristique 1 (Transitivité).

$$\left.\begin{matrix} A \subset B \\ B \subset C \end{matrix}\right\} \Rightarrow A \subset C$$

Caractéristique 2 (Réflexivité). $A \subset A$ est toujours vraie.

Caractéristique 3 (Antisymétrie).

$$\left.\begin{matrix} A \subset B \\ B \subset A \end{matrix}\right\} \Rightarrow A \equiv C$$

L'inclusion est donc une relation d'ordre parmi les sous-ensembles de l'ensemble fondamental.

2.4.2 La complémentarité

Soit un sous-ensemble A de Ω. On appelle complémentaire de A par rapport à Ω, noté $\bar{A}$ ou $\neg A$, le sous-ensemble de Ω constitué de tous les éléments qui n'appartiennent pas à A.

$$\bar{A} = \complement_{\Omega}^{A}$$

2.4.3 Ensembles disjoints

Deux ensembles A et B sont dits disjoints s'ils n'ont pas d'éléments en commun. On termes d'évènements, on dit que A et B sont incompatibles.

2.5 Opérations logiques sur les ensembles

2.5.1 Réunion ou addition logique

On appelle réunion de A et B, l'ensemble dont les éléments appartiennent soit à A, soit à B, soit simultanément à A et B. On note généralement $A \cup B$ ou encore $A + B$ [1].

Caractéristique 4.

$$\left.\begin{matrix} A \subset C \\ B \subset C \end{matrix}\right\} \Rightarrow A \cup B \subset C$$

[1] Attention, ne pas confondre avec l'addition arithmétique

Caractéristique 5. $A \subset B \Rightarrow A \cup B \equiv B$

Caractéristique 6 (Commutativité). $A \cup B \equiv B \cup A$

Caractéristique 7 (Associativité). $(A \cup B) \cup C \equiv A \cup (B \cup C) \equiv A \cup B \cup C$

2.5.2 Intersection ou produit logique

Soient 2 ensembles A et B. On appelle intersection de A et B l'ensemble des éléments qui appartiennent à la fois à A et à B. On note $A \cap B = A.B$

Caractéristique 8. Si A et B sont disjoints, alors $A \cap B \equiv \emptyset$

Caractéristique 9. $A \cap B \subset A \cup B$

Caractéristique 10.

$$\left.\begin{array}{r} A \subset C \\ B \subset C \end{array}\right\} \Rightarrow A \cap B \subset A \cup B \subset C$$

Caractéristique 11. $B \subset A \Rightarrow A \cap B \equiv B$

Caractéristique 12 (Commutativité). $A \cap B \equiv B \cap A$

Caractéristique 13 (Associativité). $(A \cap B) \cap C \equiv A \cap (B \cap C) \equiv A \cap B \cap C$

Caractéristique 14 (Distributivité à droite par rapport à la réunion).

$$\begin{aligned} A \cap (B \cup C) &\equiv (A \cap B) \cup (A \cap C) \\ (A \cup B) \cap (C \cup D) &\equiv ((A \cup B) \cap C) \cup ((A \cup B) \cap D) \\ &\equiv (A \cap C) \cup (B \cap C) \cup (A \cap D) \cup (B \cap D) \end{aligned}$$

La distributivité à gauche est valable également.

3

Axiomes du calcul des probabilités

3.1 Notion d'évènement

Soit un évènement fondamental Ω constitué de plusieurs éventualités équiprobables. L'évènement fondamental peut être subdivisé en plusieurs sous-parties. Dans chaque sous-partie, il y a plusieurs éventualités équiprobables.

Un évènement est défini comme un sous-ensemble des parties de Ω. Il est constitué d'une ou plusieurs éventualités équiprobables.

Une sous-partie E est dite élémentaire si elle n'est constituée que d'une seule éventualité.

Remarque 2. On peut appliquer le diagramme de Venn dans la théorie des évènements. Les opérations logiques sur les ensembles peuvent y être transposées, permettant ainsi l'élaboration d'une algèbre sur les évènements.

Caractéristique 15 (Complémentarité). $\complement_{\Omega}^{A} = \bar{A}$, c'est la non-réalisation de l'évènement A c.-à-d. si A est réalisé, $\bar{A}$ ne l'est pas.

Caractéristique 16 (Réunion). On peut créer un evènement C qui soit la réunion de deux évènements A et B. La réalisation de C est soit la réalisation de A, soit de B, soit de A et B simultanément.

Caractéristique 17 (Intersection). La réalisation de C entraîne la réalisation de A et de B.

3.2 Mesure de la probabilité

Si nous lançons un dé à 6 faces, les résultats possibles sont $\{1, 2, 3, 4, 5, 6\}$.

L'obtention de la face "'1"' constitue un évènement. La "'chance"' d'avoir la face "'1"' est de $\frac{1}{6}$. C'est la probabilité d'avoir une face.

Il en est de même pour chacune des faces du dé.

Soit $\mathcal{P}(\Omega)$ l'ensemble des sous-parties de Ω. La probabilité est une application de chaque évènement des parties de Ω vers $\mathbb{R}$. Cette application noté P ou Pr s'appelle "'mesure de la probabilité"'

$$E_i \in \mathcal{P}(\Omega) \xrightarrow{P} P\{E_i\} \in \mathbb{R} \tag{3.1}$$

Le nombre réel $P\{E_i\}$ s'appelle probabilité de réalisation de l'évènement E_i. Elle quantifie la crédibilité de (ou la chance d'obtenir) l'évènement E_i.

3.3 Axiomes de Kolmogorov

L'application P est restreinte par 3 axiomes, dites "axiomes de Kolmogorov".

Axiome de positivité Toute probabilité de réalisation d'un évènement est toujours positif.

$$P\{A\} \geq 0, \forall A \in \Omega$$

Axiome de certitude $P\{\Omega\} = 1$, on dit que Ω est un évènement certain.

Remarque 3. La probabilité est normée, c.-à-d. $0 \leq P \leq 1$.

Axiome d'additivité Si A et B sont incompatibles ($A \cap B \equiv \emptyset$), alors $P\{A \cup B\} = P\{A\} + P\{B\}$.

3.4 Quelques propriétés

Caractéristique 18 (Propriété de l'évènement impossible). La probabilité de l'évènement impossible est nulle. En d'autres termes $P(\emptyset) = 0$.

Preuve. Soit un évènement quelconque A. On sait que $A \cap \emptyset \equiv \emptyset$ et $A \cup \emptyset \equiv A$ (tout évènement est incompatible avec l'évènement impossible).

A partir de l'axiome d'additivité de Kolmogorov,

$$\begin{aligned} P\{A \cup \emptyset\} &= P\{A\} + P\{\emptyset\} \\ &= P\{A\} \end{aligned}$$

Par conséquent, $P(\emptyset) = 0$.

□

Caractéristique 19 (Probabilité et inclusion). Si $B \subset A$, alors $P\{B\} \leq P\{A\}$

Caractéristique 20 (Probabilité d'un évènement complémentaire). Soit, $\bar{A} = \complement_{\Omega}^{A}, P\{\bar{A}\} = 1 - P\{A\}$

Preuve. $A \cup \bar{A} = \Omega$, et $A \cap \bar{A} = \emptyset$.
$P\{A \cup \bar{A}\} = P\{A\} + P\{\bar{A}\} = 1 \Rightarrow P\{\bar{A}\} = 1 - P\{A\}$

3.5 Théorème des probabilités totales

Proposition 1. *Considérons* 2 *évènements* A *et* B *tels que* $A \cap B \neq \emptyset$.

$$P\{A \cup B\} = P\{A\} + \{B\} - P\{A \cap B\} \tag{3.2}$$

Preuve. On définit C tel que $C = \bar{A} \cap B$. On en déduit que $A \cap C = \emptyset$. En effet, $A \cap (\bar{A} \cap B) = \emptyset$. A et $\bar{A} \cap B$ sont incompatibles.

On peut également déduire que $A \cup C = A \cup B$ (cf. les propriétés des probabilités).

$P\{A \cup C\} = P\{A\} + P\{C\} = P\{A \cup B\}$

$$P\{C\} = P\{A \cup B\} - P\{A\} \tag{3.3}$$

Considérons maintenant les évènements C et $A \cap B$. Ils sont incompatibles c.-à-d. $C \cap (A \cap B) = \emptyset$. On constate aisément que $C \cup (A \cap B) = B$, $P\{C \cup (A \cap B) = B\} = P\{B\} = P\{C\} + P\{A \cap B\}$

$$P\{C\} = P\{B\} - P\{A \cap B\} \tag{3.4}$$

Avec les équations (3.3) et (3.4), nous en déduisons :

$P\{A \cup B\} - P\{A\} = P\{B\} - P\{A \cap B\}$, d'où le résultat (Équation 3.2).
□

3.6 Inégalité de Boole

D'après les axiomes des probabilités totales,

$$\begin{aligned} P\{A + B\} &= P\{A\} + P\{B\} - P\{A.B\} \\ P\{A + B\} &\leq P\{A\} + P\{B\} \end{aligned}$$

On peut généraliser à un nombre quelconque d'évènements :

$$P\{\sum_i A_i\} \leq \sum_i P\{A_i\} \tag{3.5}$$

Caractéristique 21 (Évènements incompatibles). Si les évènements A_i sont incompatibles, $P\{\sum_i A_i\} = \sum_i P\{A_i\}$

3.7 Théorème des probabilités composées

3.7.1 Les probabilités conditionnelles ou liées

Soient 2 évènements A et B tels que $A \cap B \neq \emptyset$. La probabilité de réalisation de l'évènement B lorsque l'évènement A est réalisé s'appelle "probabilité conditionnelle de B par rapport à A", que l'on note $P\{B/A\}$. On dit également "probabilité de B si A" ou encore "probabilité de B sachant A".

On calcule cette probabilité de la manière suivante :

$$P\{B/A\} = \frac{P\{B \cap A\}}{P\{A\}} \tag{3.6}$$

Exemple 12. Soit un jeu de 52 cartes. Les cartes ont été regroupées selon leur couleur. Quelle est la probabilité d'extraire la hauteur "roi" sachant que l'on a puisé dans le paquet des "coeur" ?

$P\{coeur\} = \frac{13}{52}$, et $P\{roi \cap coeur\} = \frac{1}{52}$. On calcule alors $P\{roi/coeur\} = \frac{1/52}{13/52} = \frac{1}{13}$.

3.7.2 Les probabilités composées

A partir du résultat précédent, nous pouvons écrire

$$P\{A \cap B\} = P\{B/A\} \times P\{A\} \quad (3.7)$$
$$= P\{A/B\} \times P\{B\} \quad (3.8)$$

3.8 Notion d'indépendance en probabilité

Soient 2 évènements A et B. Ils sont indépendants si la réalisation de A n'affecte pas la réalisation de B, et inversement.

On peut alors écrire

$$P\{A/B\} = P\{A\}$$
$$P\{A/A\} = P\{B\}$$

On dit encore que A et B sont indépendants si et seulement la probabilité de réalisation simultanée de ces évènements est égal au produit des probabilités individuelles.

$$P\{A \cap B\} = P\{A\} \times P\{A\}$$

4

Les schémas de tirages probabilistes

4.1 Exposé du problème

Nous disposons d'une urne avec N boules. Dans l'urne, il y a N_1 boules de couleur c1, N_2 boules de couleurs c2, …, N_k boules de couleur ck.

On extrait n boules de l'urne avec, respectivement pour chaque couleur, $n_1, n_2, \ldots, n_k$. Quelle est la probabilité d'obtenir cette configuration $(n_1, n_2, \ldots, n_k)$?

L'ensemble fondamental est $\Omega = \{c1, c2, \ldots, ck\}$.

Nous pouvons adopter le mode de représentation suivant :

Élément	Effectif	Échantillon ω
c1	N_1	n_1
…	…	…
ck	N_k	n_k
Total	N	n

Question 1. Pour calculer les probabilités, il faut se poser et répondre correctement aux questions suivantes :

– Le tirage est-il exhaustif ou avec remise ?
– La configuration est-elle ordonnée ?

4.2 Tirage exhaustif

Un tirage est dit exhaustif s'il est effectué sans remise (sans répétition). Il importe alors de savoir si avec un tirage sans remise :

– on s'intéresse à l'ordre d'apparition des boules c.-à-d. la disposition est-elle ordonnée ?
– on s'intéresse uniquement au nombre d'apparition des boules c.-à-d. obtenir n_i boules de couleur ci.

4.2.1 Tirage exhaustif ordonné

Considérons que nous avons N boules de 2 couleurs br (rouges) et bn (noires). On extrait un échantillon ω de taille n, et on obtient n_1 br et n_2 bn.

On cherche à calculer $P\{n_1(br)\,\text{et}\,n_2(bn)\}$.

Revenons à notre tableau.

Couleur	Effectif	Échantillon ω
br	N_1	n_1
bn	N_2	n_2
Total	N	n

- Nombre d'éventualités possibles : A_N^n
- Nombre d'éventualités favorables : $A_{N_1}^{n_1} \times A_{N_2}^{n_2}$

On en déduit la probabilité :

$$P\{n_1(br)\,\text{et}\,n_2(bn)\} = \frac{A_{N_1}^{n_1} \times A_{N_2}^{n_2}}{A_N^n} \tag{4.1}$$

4.2.2 Tirage exhaustif non-ordonné

Qu'importe l'ordre d'apparition des boules cette fois-ci. Nous calculons donc :

- Nombre d'éventualités possibles : C_N^n
- Nombre d'éventualités favorables : $C_{N_1}^{n_1} \times C_{N_2}^{n_2}$

On en déduit la probabilité :

$$P\{n_1(br)\,\text{et}\,n_2(bn)\} = \frac{C_{N_1}^{n_1} \times C_{N_2}^{n_2}}{C_N^n} \tag{4.2}$$

C'est le **schéma hypergéométrique**.

Remarque 4. Il est possible d'exprimer cette seconde probabilité (Équation 4.2 en fonction de la précédente (Équation 4.1). En effet, $C_N^n = \frac{N!}{n!(N-n)!} = \frac{A_N^n}{n!}$. Il arrive alors $P\{n_1(br)\,\text{et}\,n_2(bn)\} = \frac{A_{N_1}^{n_1}/n_1! \times A_{N_2}^{n_2}/n_2!}{A_N^n/n!}$, ainsi

$$P\{n_1(br)\,\text{et}\,n_2(bn)\} = \frac{A_{N_1}^{n_1} A_{N_2}^{n_2}}{A_N^n} \times \frac{n!}{n_1!n_2!} \tag{4.3}$$

Dans la première partie de l'équation 4.3, nous retrouvons la disposition ordonnée sans remise (Équation 4.1) ; dans la seconde partie, les différentes permutations possibles des n_1 et n_2 éléments parmi n c.-à-d. $\mathcal{P}_n^{(n_1,n_2)}$.

4.3 Tirage de Bernoulli

Nous nous plaçons maintenant dans le cadre du **tirage avec remise**.

4.3.1 Tirage de Bernoulli ordonné

Tirage avec remise et ordonné.

Couleur	Effectif	Échantillon ω
br	N_1	n_1
bn	N_2	n_2
Total	N	n

– Nombre d'éventualités possibles : $\mathcal{A}_N^n$
– Nombre d'éventualités favorables : $\mathcal{A}_{N_1}^{n^1} \times \mathcal{A}_{N_2}^{n^2}$

On en déduit la probabilité :

$$P\{n_1(br)\,\mathrm{et}\,n_2(bn)\} = \frac{\mathcal{A}_{N_1}^{n_1} \times \mathcal{A}_{N_2}^{n_2}}{\mathcal{A}_N^n} \tag{4.4}$$

Posons $p = \frac{N_1}{N}$ et $q = \frac{N_2}{N} = 1 - p$. Nous pouvons écrire l'équation (4.4) de la manière suivante :

$$\begin{aligned} P\{n_1(br)\,\mathrm{et}\,n_2(bn)\} &= \frac{N_1^{n_1} N_2^{n_2}}{N^n} \\ &= \frac{N_1^{n_1} N_2^{n_2}}{N^{n_1} N^{n_2}} \\ &= (\frac{N_1}{N})^{n_1} (\frac{N_2}{N})^{n_2} \\ &= p^{n_1} q^{n_2} \end{aligned}$$

4.3.2 Tirage de Bernoulli non-ordonné

Tirage avec remise, non-ordonné.

A la lumière de ce que nous avons pu voir précédemment, du passage du tirage ordonné vers le non-ordonné pour le schéma exhaustif (Équation 4.3) et le tirage avec remise ordonné (Équation 4.4), nous pouvons déduire la formule :

$$P\{n_1(br)\,\mathrm{et}\,n_2(bn)\} = \mathcal{P}_n^{(n_1,n_2)} p^{n_1} q^{n_2} \tag{4.5}$$

Nous pouvons également écrire :

$$P\{n_1(br)\,\mathrm{et}\,n_2(bn)\} = C_n^{n_1} p^{n_1} (1-p)^{n-n_1} \tag{4.6}$$

C'est le **schéma binomial**.

4.4 Comparaison des 2 modes de tirage, exhaustif et avec remise

Intéressons-nous au tirage non-ordonné avec k couleurs.

Dans un tirage exhaustif, $P_E = \frac{C_{N_1}^{n_1} C_{N_2}^{n_2} \cdots C_{N_k}^{n_k}}{C_N^n}$

Dans un tirage de Bernoulli, $P_B = \frac{n!}{n_1! n_2! \cdots n_k!} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$

Remarque 5. Si $n = 1$, alors $P_E = P_B$

Remarque 6. Dans un tirage sans remise (exhaustif), la structure de l'ensemble fondamental est modifié à chaque tirage. Ce n'est pas le cas du tirage de Bernoulli, la composition de l'urne n'est jamais modifiée puisque chaque objet extrait y est remise avant le tirage suivant : de fait, il y a indépendance totale des tirages. C'est la raison pour laquelle, lors du tirage avec remise, l'expression de la probabilité ne dépend pas de N, nombre total de boules, mais uniquement de la structure de l'ensemble fondamental (p_i).

Remarque 7. Lorsque les N_k sont très grands devant les n_k (c.-à-d. $\frac{N!}{(N-n)!} \approx N^n$, cf. les développement limités), nous avons l'approximation $P_E \approx P_B$. En effet, les tirages successifs modifient très peu la structure de l'ensemble fondamental, elle reste proche de la situation initiale.

4.5 Généralisation du tirage exhaustif ordonné

Nous voulons distribuer n objets, répartis en m groupes, entre p personnes. On cherche à savoir quelle est la probabilité d'une configuration donnée.

Pour mieux appréhender le problème, utilisons le tableau suivant :

Obj vs. Pers	1	$\cdots$	j	$\cdots$	p	Total
1	n_{11}	$\cdots$	n_{1j}	$\cdots$	n_{1p}	$n_{1.}$
$\vdots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
i	n_{i1}	$\cdots$	n_{ij}	$\cdots$	n_{ip}	$n_{i.}$
$\vdots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
m	n_{m1}	$\cdots$	n_{mj}	$\cdots$	n_{mp}	$n_{m.}$
Total	$n_{.1}$	$\cdots$	$n_{.j}$	$\cdots$	$n_{.p}$	$n_{..} = n$

La probabilité s'écrit :

$$P\{\} = \frac{\frac{n_{1.}!}{n_{11}!n_{12}!\cdots n_{1p}!} \frac{n_{2.}!}{n_{21}!n_{22}!\cdots n_{2p}!} \cdots \frac{n_{m.}!}{n_{m1}!n_{m2}!\cdots n_{mp}!}}{\frac{n!}{n_{.1}!n_{.2}!\cdots n_{.p}!}}$$

En réorganisant un peu tout cela.

$$P\{\} = \frac{\prod_{i=1}^{m} n_{i.}! \prod_{j=1}^{p} n_{.j}!}{n! \prod_{i=1}^{m} \prod_{j=1}^{p} n_{ij}!} \tag{4.7}$$

Une règle mnémonique pour s'en rappeler est de considérer qu'il s'agit du rapport entre :
- le produit des factoriels des effectifs marginaux ;
- $n!$ fois le produit des factoriels des effectifs conjoints.

5

Probabilité de Bayes

5.1 Énoncé du problème

Un ensemble de boules rouges (br) et boules noires (bn) sont réparties dans 2 urnes U_1 et U_2.

- p_1 est la proportion de br dans U_1, p_2 dans U_2 ;
- $q_1 = 1 - p_1$ est la proportion de bn dans U_1, q_2 dans U_2 ;

On demande à un opérateur d'effectuer un tirage en deux temps :

1. d'abord, il choisit une urne, π_1 est la probabilité de choisir l'urne U_1, π_2 pour U_2 ;
2. puis il extrait une boule de l'urne.

Constatant la couleur de la boule (br pour fixer les idées), nous devons calculer la probabilité que la boule provienne de l'urne U_1 (ou U_2). Cette probabilité a posteriori est dite **Probabilité de Bayes**.

5.2 Le théorème de Bayes

Posons B l'évènement "tirer une boule rouge". On cherche à calculer la probabilité $P\{U_1/B\}$.

$$\begin{aligned} P\{U_1/B\} &= \frac{P\{U_1.B\}}{P\{B\}} \\ &= \frac{P\{B/U_1\}P\{U_1\}}{P\{B\}} \end{aligned}$$

Or $U_1 + U_2 = \Omega$ (ensemble fondamental) ; $B.(U_1 + U_2) = B.\Omega = B$, on en déduit :

$$\begin{aligned} P\{U_1/B\} &= \frac{P\{B/U_1\}P\{U_1\}}{P\{B.(U_1 + U_2)\}} \\ &= \frac{P\{B/U_1\}P\{U_1\}}{P\{B.U_1 + B.U_2\}} \end{aligned}$$

Si U_1 et U_2 sont incompatibles, BU_1 et BU_2 le sont également. Nous pouvons appliquer l'axiome d'additivité de Kolmogorov.

$$\begin{aligned} P\{U_1/B\} &= \frac{P\{B/U_1\}P\{U_1\}}{P\{B.U_1\} + P\{B.U_2)\}} \\ &= \frac{P\{B/U_1\}P\{U_1\}}{P\{B/U_1\}P\{U_1\} + P\{B/U_2\}P\{U_2\}} \end{aligned}$$

5.3 Généralisation du théorème de Bayes

Considérons le cas général où nous avons m urnes et k couleurs de boules : U_i désigne l'urne i, et B_j indique l'apparition d'une boule de couleur j.

La probabilité a posteriori que l'urne U_i ait été utilisée sachant qu'une boule j est apparue (évènement B_j) s'écrit :

$$P\{U_i/B_j\} = \frac{p_{j/i} \times \pi_i}{\sum_{i=1}^{k} p_{j/i} \times \pi_i} \tag{5.1}$$

où $p_{j/i}$ est la proportion de boules de couleur j dans l'urne i ; π_i est la probabilité d'obtenir l'urne i.

Exemple 13. Un appareil peut être monté avec des pièces de haute qualité (40% des cas) et avec des pièces ordinaires (60% des cas). Dans le premier cas, sa fiabilité (probabilité de fonctionnement) sur une durée t est égale à 0.95 ; dans le second, elle est de 0.7. L'appareil a été soumis à un essai et s'est avéré fiable (fonctionnement sans défaillance sur la durée de référence). Déterminer la probabilité que l'appareil soit composé de pièces de haute qualité.

Solution : Notons B l'évènement "l'appareil a fonctionné sans défaillance" ; U_1 (resp. U_2) la probabilité qu'un appareil soit composé de pièces de haute qualité (resp. de qualité ordinaire), $P(U_1) = 0.4$ (resp. $P(U_2) = 0.6$).

On nous indique que $P(B/U_1) = 0.95$ et $P(B/U_2) = 0.7$.

En appliquant le théorème de Bayes, nous obtenons :

$$P(U_1/B) = \frac{0.4 \times 0.95}{0.4 \times 0.95 + 0.6 \times 0.7} = \frac{0.38}{0.8} = 0.475$$

6

Les variables aléatoires

6.1 Définition d'une variable aléatoire

Exemple introductif. On lance 3 fois une pièce de monnaie (qui prend deux valeurs possibles "pile" ou "face", est-il besoin de le préciser). L'ensemble fondamental est $\Omega = \{(P, F), (P, F), (P, F)\}$.

Notons X le nombre de "face" apparue au bout des trois jets. Les valeurs possibles de X sont $D_X = \{0, 1, 2, 3\}$. X est une application de Ω dans $\mathcal{D}_X$.

De manière générale, X est une application :

$$X : \Omega \mapsto \mathbb{R} \tag{6.1}$$

X s'appelle une **variable aléatoire**, on s'intéressera aux valeurs prises par X ($X = x$), et plus particulièrement à la probabilité d'obtenir ces valeurs $P(X = x)$ [1].

Définition 6. *Soient une épreuve e_i et Ω l'ensemble fondamental des éventualités. L'application $X : \Omega \mapsto \mathbb{R}$ fait correspondre à tout élément de Ω un nombre réel*

$$\forall e_i \in \Omega, e_i \to X(e_i) \in \mathbb{R}$$

X désigne la variable aléatoire ; x désigne la valeur possible de la variable aléatoire ; le domaine de définition de X est $\mathcal{D}_X = \{x_1, \ldots, x_n\}$.

6.2 Probabilité d'une variable aléatoire

Soit X une variable aléatoire définie par une application de Ω dans $\mathbb{R}$, pouvant prendre des valeurs dans $\mathcal{D}_X = \{x_1, \ldots, x_n\}$. Par définition, la probabilité pour que la variable X soit égale à x est la probabilité des éléments de Ω ayant pour image la valeur x dans l'application. Cette probabilité mesure le degré de crédibilité de réalisation de l'évènement.

$P(X = x_i)$ est régie par les 3 axiomes de Kolmogorov.

1. $P(X = x_i) \geq 0$
2. $P(X = \mathcal{D}_X) = 1$
3. Si les x_i sont incompatibles entre eux, $P[(X = x_i)\, ou\, (X = x_j)] = P(X = x_i) + P(X = x_j)$

[1] A partir de maintenant, nous utiliserons les parenthèses, et non les accolades, pour désigner un évènement "valeur(s) prise(s) par une variable aléatoire".

6.3 Les différents types de variables aléatoire

6.3.1 Variable aléatoire discrète

Une variable aléatoire est dite discrète lorsque les valeurs x_i qu'elle est susceptible de prendre sont en nombre fini, ou encore formés exclusivement de nombres entiers.

Quelques exemples :
- nombre de "face" apparaissant après 10 jets d'une pièce ;
- nombre de véhicules passant à un carrefour dans une journée ;
- nombre de clients entrant dans un magasin le samedi.

6.3.2 Variable aléatoire continue

Une variable aléatoire est dite continue si elle peut prendre *toutes* les valeurs d'un *intervalle*.

Quelques exemples :
- intervalle de temps entre 2 passages de train ;
- longueur de cheveux ;
- durée de vie en secondes d'une pièce mécanique.

Remarque 8 (Distinction variable discrète et continue). Parfois, la distinction est purement formelle. Le type de la variable aléatoire dépend du degré de précision que l'on désire appréhender. Par exemple, on s'intéresse à l'espérance de vie d'un chat, si on travaille sur le nombre d'années, le nombre de valeurs possible est fini ; si on travaille en secondes, ce sera différent. La situation sera un peu plus claire lorsque l'on abordera la question de la probabilité ponctuelle. Dans le cas d'une variable continue, elle est par définition nulle.

6.4 La loi de probabilité

Une variable aléatoire est totalement définie par sa **loi de probabilité**. Cette dernière est caractérisée par :

1. l'ensemble des valeurs qu'elle peut prendre (son domaine de définition $\mathcal{D}_X$) ;
2. les probabilités attribuées à chacune de ses valeurs $P(X = x)$.

Exemple 14. On lance une pièce de monnaie une fois, l'ensemble fondamental est $\Omega = \{P, F\}$. Soit X le nombre d'apparition de "pile", $\mathcal{D}_X = \{0, 1\}$ et on peut calculer facilement $P(X = x)$, $P(X = 0) = \frac{1}{2}$ et $P(X = 1) = \frac{1}{2}$.

Exemple 15. On lance une pièce de monnaie 2 fois, l'ensemble fondamental est $\Omega = \{P, F\}$. Soit X le nombre d'apparition de "pile", $\mathcal{D}_X = \{0, 1, 2\}$. On veut calculer $P(X = 1)$, pour cela on s'intéresse aux deux séquences possibles qui lui sont associées (Pile d'abord, Face ensuite), ou inversement (F, P).

$$
\begin{aligned}
P(X=1) &= P\{(P,F)\,ou\,(F,P)\} \\
&= P\{(P,F)\} + P\{(F,P)\}\,,\ \text{\scriptsize évènements incompatibles} \\
&= P\{P\} \times P\{F\} + P\{F\} \times P\{P\}\,,\ \text{\scriptsize jets indépendants} \\
&= \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} \\
&= \frac{1}{2}
\end{aligned}
$$

De la même manière, nous pouvons calculer $P(X=0)$ et $P(X=2)$.

Exemple 16. Sur le même exemple du jet de 2 pièces, nous pouvons adopter l'approche du tirage de Bernoulli non-ordonné. Nous utilisions le tableau suivant pour caractériser l'expérience :

Pièce	Ω	ω
Face	$q=0.5$	$2-x$
Pile	$p=0.5$	x
Total	1	2

La probabilité d'obtenir x "pile" est :

$$P(X=x) = \frac{2!}{x!(2-x)!} p^x q^{2-x}$$

De manière générale, pour n jets de pièces, si l'on se réfère au schéma binomial, la probabilité d'obtenir x-fois le côté "pile" est égal à :

Pièce	Ω	ω
Face	$q=0.5$	$n-x$
Pile	$p=0.5$	x
Total	1	n

et

$$P(X=x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

Il est d'usage de représenter graphiquement la distribution de probabilité d'une variable discrète à l'aide d'un diagramme en bâtons. A chaque valeur de X est associée un trait (un bâton) dont la hauteur est proportionnelle à $P(X=x)$ (Figure 6.1). Certains parlent également d'histogramme de fréquences, mais dans ce cas, la largeur de chaque barre est obligatoirement constante.

Remarque 9 (Probabilité cumulée). Dans certains cas, on peut être intéressé par l'obtention d'au moins "x" valeurs. Dans l'exemple des jets de pièces, on peut vouloir poser la question : "quelle est la probabilité que l'on obtienne *au moins* x-fois le côté pile de la pièce ?". Dans ce cas, on parle de probabilité cumulée ou encore de **fonction de répartition**.

Il s'agit de la somme des probabilités ponctuelles :

$$
\begin{aligned}
P(X \leq x) &= P(X=0\,\text{ou}\,X=1\,\text{ou}\,\ldots\,\text{ou}\,X=x) \\
&= P(X=0) + P(X=1) + \ldots + P(X=x)\ \text{\scriptsize évènements incompatibles} \\
&= \sum_{j=0}^{x} P(X=j)
\end{aligned}
$$

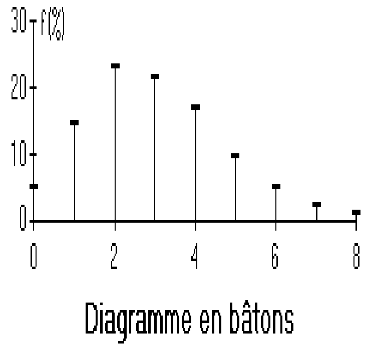

Fig. 6.1. Diagramme en bâtons (variable discrète)

Essayons de préciser cela de manière plus formelle.

6.5 Fonction de répartition

6.5.1 Pour une variable discrète

La fonction de répartition n'est autre que le cumul des probabilités individuelles. La probabilité pour que la variable aléatoire X prenne une valeur inférieure à x est une fonction $F(x)$ que l'on appelle *fonction de répartition*, $P(X < x) = F(x)$.

Si $\mathcal{D}_X = \{0, 1, 2, \ldots, n\}$,

$$F(x) = P(X < x) = \sum_{j=0}^{x-1} P(X = j) \tag{6.2}$$

On représente graphiquement cette fonction de répartition à l'aide d'une courbe cumulative en escalier (Figure 6.2)[2].

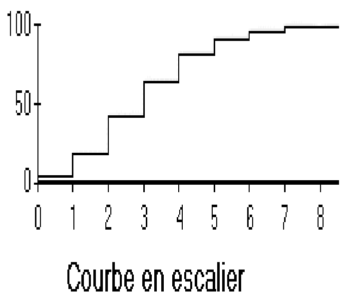

Fig. 6.2. Courbe cumulative en escalier (variable discrète)

[2] Source : `http://nte-serveur.univ-lyon1.fr/nte/immediato/math2002/Mass11/cours/chapitr2a.htm`

6.5.2 Pour une variable continue

Dans ce cas, la plupart du temps, $\mathcal{D}_X \equiv \mathbb{R}$. La probabilité ponctuelle $P(X = x) = f(x)$ est la fonction de densité. La fonction de répartition $F(x) = P(X < x)$ est définie par :

$$F(x) = \int_{-\infty}^{x} f(t)dt \tag{6.3}$$

La fonction de répartition $F(x)$ est une fonction continue, monotone et croissante dans le domaine $[0, 1]$. On note deux cas extrêmes :

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0$$
$$F(+\infty) = \lim_{x \to +\infty} F(x) = 1$$

On retrouve bien là les axiomes de positivité et de certitude.

La fonction de répartition est généralement représentée graphiquement par une courbe cumulative (Figure 6.3).

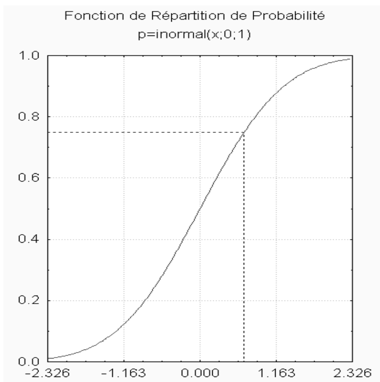

Fig. 6.3. Courbe cumulative (variable continue)

Remarque 10. Souvenons-nous que la fonction de répartition est une intégrale de la fonction de densité. Il est également possible d'utiliser cette dernière pour représenter graphiquement la fonction de répartition, il s'agit d'une surface dans ce cas (Figure 6.4).

6.6 Densité d'une variable aléatoire continue

La densité de probabilité d'une variable aléatoire continue est la dérivée première par rapport à x de la fonction de répartition. Cette dérivée prend le nom de fonction de densité, notée $f(x) = \frac{dF(x)}{dx}$. Elle est équivalente à $P(X = x)$ dans le cas des variables discrètes.

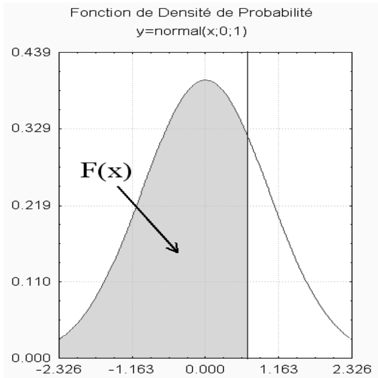

Fig. 6.4. Valeur de la fonction de répartition à partir de la fonction de densité

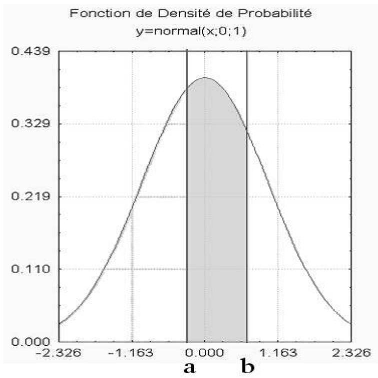

Fig. 6.5. Probabilité d'un intervalle à partir de la fonction de densité

6.7 Probabilité d'un intervalle

On cherche à calculer la probabilité $P(a < X < b)$. Graphiquement, cela se traduit par la surface comprise entre a et b (Figure 6.5).

Analytiquement, il s'agit de $P(a < X < b) = \int_a^b f(x)dx$. Développons cette expression :

$$\begin{aligned} P(a < X < b) &= P(X < b) - P(X < a) \\ &= \int_{-\infty}^{a} f(x)dx - \int_{-\infty}^{b} f(x)dx \\ &= F(b) - F(a) \end{aligned}$$

Remarque 11. Probabilité ponctuelle pour une variable continue. La vraie distinction entre variables continues et discrètes tient dans le calcul de la probabilité ponctuelle. La probabilité d'un point c situé entre a et b serait $\lim_{b \to a} P(a < X < b) = 0$. Ainsi, la probabilité d'une valeur est par définition nulle pour les variables continues. En réalité, il s'agit bien souvent d'un problème de point de vue, voire d'échelle ou de précision de mesure. La probabilité que la durée de vie d'un véhicule soit égal à 4 an est loin d'être nulle (beaucoup de véhicules partent à la casse au bout de 4 an). En revanche, la probabilité que cette durée de vie soit exactement de 126144000 secondes (très approximativement 4 an) est quasi-nulle (on peut difficilement dire à la seconde près la perte d'un véhicule).

7

Caractéristiques d'une variable aléatoire

7.1 Les caractéristiques de tendance centrale

7.1.1 Les quantiles

On appelle **quantile** ou **fractile** d'ordre α $(0 \leq \alpha \leq 1)$ d'une variable aléatoire X dont la fonction de répartition est $F(x)$, la valeur x_α telle que $F(x_\alpha) = \alpha$.

x_α s'appelle quantile d'ordre α.

Remarque 12. Dans le cas où X est une variable discrète, $F(x_\alpha) = \alpha$ s'entend $P(X < x_\alpha) = \alpha$.

Nous énumérons ici quelques quantiles particuliers.

La médiane

La médiane est le quantile d'ordre $\alpha = \frac{1}{2}$, en d'autres termes la médiane Me est définie par $\int_{-\infty}^{Me} f(x)dx = 0.5$. La médiane partage en deux parties égales la population, c'est une caractéristique de tendance centrale.

Les quartiles

Les quartiles, notés Q_i (respectivement $i = 1, 2, 3$) correspondent aux quantiles d'ordre ($\alpha = 0.25, 0.5, 0.75$).

Notons que $Q_2 = Me$.

Les déciles

Le k-ème décile ($k = 1, \ldots, 9$) est le quantile d'ordre $\frac{k}{10}$. En particulier, le 5-ème décile correspond à la médiane.

7.1.2 Le mode

On appelle mode (valeur dominante, valeur la plus probable) d'une variable aléatoire, la valeur Mo pour laquelle l'histogramme de fréquence présente son maximum.

Lorsque la variable aléatoire X est continue, avec une fonction de densité pourvue d'une dérivée première et d'une dérivée seconde, le mode Mo satisfait à $f'(Mo) = 0$ et $f''(Mo) < 0$) (concavité vers le bas).

Dans le cas des variables discrètes, Mo est la valeur de X associée à la plus grande probabilité, d'où l'appellation *valeur la plus probable*.

7.1.3 Espérance mathématique

Soit $\varphi(X)$ une fonction définie pour tout X appartenant au domaine de définition $\mathcal{D}_X$. On appelle **espérance mathématique** de $\varphi(X)$, que l'on note $E[\varphi(X)]$ l'expression :

$$E[\varphi(X)] = \int_{\mathcal{D}_X} \varphi(x)f(x)dx = \int_{\mathcal{D}_X} \varphi(x)dF(x) \tag{7.1}$$

Remarque 13. En particulier, si $\varphi(X) = X$, alors $E[X] = \int_{\mathcal{D}_X} xf(x)dx$.

Remarque 14. L'espérance mathématique existe si l'intégrale est convergente. Elle est indépendante de x.

Remarque 15. Lorsque X est une variable aléatoire discrète, $E[\varphi(X)] = \sum_{\mathcal{D}_X} \varphi(x)P(X = x)$.

Si $\mathcal{D}_X \equiv \mathbb{N}, E[\varphi(X)] = \sum_{x=}^{+\infty} \varphi(x)P(X = x)$

Caractéristique 22 (Espérance d'une constante). $E[a] = a$.

Preuve.

$$\begin{aligned} E[a] &= \int_{\mathcal{D}_X} af(x)dx \\ &= a \times \int_{\mathcal{D}_X} f(x)dx \\ &= a \times 1 \\ &= a \end{aligned}$$

□

On peut en déduire que $E[E[X]] = E[X]$, puisque $E[X]$ n'est pas une variable aléatoire.

Caractéristique 23 (L'espérance mathématique est un opérateur linéaire). $E[aX + bY] = aE[X] + bE[Y]$.

De manière générale,

$$E[\sum_i a_iX_i + b] = \sum_i a_iE[X_i] + b$$

Caractéristique 24 (Espérance du produit de 2 v.a.). $E[XY] = E[X]E[Y] + COV(X,Y)$, $COV()$ est la covariance, nous la détaillerons plus loin.

En particulier, lorsque X et Y sont indépendantes, $E[XY] = E[X]E[Y]$.

7.2 Les caractéristiques de dispersion

7.2.1 Les moments non-centrés d'ordre r

Un moment non-centré d'ordre r est définie de la manière suivante :

$$m_r(X) = E[X^r] \tag{7.2}$$

- Pour une v.a. (variable aléatoire) discrète, $m_r(X) = \sum_{x=0}^{+\infty} x^r p_x$, où $p_x = P(X = x)$;
- Pour une v.a. continue, $m_r(X) = \int_{\mathcal{D}_X} x^r f(x)dx$

Remarque 16 (Moments non-centrés empiriques, statistique descriptive). Rappelons qu'en statistique descriptive, ce moment non-centré, pour une v.a. discrète par exemple, est obtenue avec la formule $m_r = \sum_{i=0}^{n} f_i x_i$, où f_i est la fréquence observée de la valeur x_i.

Remarque 17 (Cas particuliers).
- $r = 0 \Rightarrow m_0(X) = 1$;
- $r = 1 \Rightarrow m_1(X) = E[X]$, espérance mathématique ;
- $r = 2 \Rightarrow m_2(X) = E[X^2]$.

7.2.2 Les moments centrés d'ordre r

Un moment centré d'ordre r est défini de la manière suivante :

$$\mu_r(X) = E[(E - E[X])^r] \tag{7.3}$$

- pour une v.a. discrète : $\mu_r(X) = \sum_{x=0}^{+\infty}(x - E[X])^r p_x$
- pour une v.a. continue : $\mu_r(X) = \int_{\mathcal{D}_X}(x - E[X])^r f(x)dx$

Remarque 18 (Statistique descriptive). En statistique descriptive, pour une variable discrète, le moment centré d'ordre r est obtenu avec $m_r = \sum_{i=0}^{n} f_i(x_i - \bar{x})^r$

Remarque 19 (Cas particuliers).
- $r = 0 \Rightarrow \mu_0(X) = E[(X - E[X])^0] = E[1] = 1$
- $r = 1 \Rightarrow \mu_1(X) = E[(X - E[X])] = E[X] - E[E[X]] = E[X] - E[X] = 0$
- $r = 2 \Rightarrow \mu_2(X) = E[(X - E[X])^2] = V(X)$, la variance de X que nous détaillerons plus loin.

7.2.3 Les moments factoriels

On appelle moment factoriel d'ordre r la quantité :

$$\mu_{[r]}(X) = E[\frac{X!}{(X-r)!}] = E[X(X-1)\cdots(X-r+1)] \tag{7.4}$$

Les moments factoriels se calculent essentiellement sur les v.a. discrètes, il vient alors $\mu_{[r]}(X) = \sum_{x=0}^{n} \frac{x!}{(x-r)!} p_x$.

Remarque 20. Il est possible d'exprimer les moments non-centrés à l'aide des moments factoriels :

- $m_1 = \mu_{[1]}$
- $m_2 = \mu_{[2]} + \mu_{[1]}$
- $m_3 = \mu_{[3]} + 2\mu_{[2]} + \mu_{[1]}$
- $m_4 = \mu_{[4]} + 6\mu_{[3]} + 7\mu_{[2]} + \mu_{[1]}$

7.2.4 La variance

On appelle variance de X, noté $V(X)$, le moment centré d'ordre 2 de X.

$$V(X) = E[(X - E[X])^2] \tag{7.5}$$

La variance mesure la dispersion autour de la moyenne.

L'écart-type est la racine carrée de la variance :

$$\sigma(X) = \sqrt{V(X)} \tag{7.6}$$

Remarque 21 (Formule de Koenig). Il est possible d'exprimer la variance à partir des moments non-centrés.

$$\begin{aligned} V(X) &= E[(X - E[X])^2] = E[X^2 - 2XE[X] + E[X]^2] \\ &= E[X^2] + E[X]^2 - 2E[X.E[X]] \\ &= E[X^2] + E[X]^2 - E[X].E[X] \\ &= E[X^2] - E[X]^2 \\ &= m_2(X) - m_1(X)^2 \text{ , connue sous le nom de } \textbf{formule de Koenig} \end{aligned}$$

Caractéristique 25 (Variance d'une constante). Si a est une constante (non aléatoire), $V(a) = 0$.

Caractéristique 26 (Mise en facteur d'un coefficient non-aléatoire). Si a est une constante (non aléatoire), $V(aX) = a^2 V(X)$.

Caractéristique 27 (Variance d'une somme de v.a.). Pour $Z = aX + bY$, $V(Z) = a^2V(X) + b^2V(Y) + 2ab \times COV(X,Y)$

En particulier, si X et Y sont indépendants, $V(Z) = a^2V(X) + b^2V(Y)$

Caractéristique 28 (Variance du produit de 2 v.a.). $V(XY) = V(X)V(Y) + E[X]^2V(Y) + E[Y]^2V(X)$.

En particulier, si les v.a. sont centrés, c.-à-d. $E[X] = E[Y] = 0$, alors $V(XY) = V(X)V(Y)$

7.2.5 La covariance

Soient 2 v.a. X et Y, on appelle covariance de X et Y l'expression :

$$COV(X,Y) = E[(X - E[X])(Y - E[Y])] = E(XY) - E[X].E[Y] \tag{7.7}$$

- Pour une v.a. discrète, $COV(X,Y) = \sum_x \sum_y (x - E[X])(y - E[Y]).p_{xy}$, où $p_{xy} = P((X = x) \text{ et } (Y = y))$;

– Pour une v.a. continue, $COV(X,Y) = \int_{\mathcal{D}_X}\int_{\mathcal{D}_Y}(x - E[X])(y - E[Y])f(x,y)dxdy$, où $f(x,y)$ est la fonction de densité conjointe.

Caractéristique 29 (Covariance de 2 v.a. indépendantes). Soient X et Y, deux v.a. indépendantes, $COV(X,Y) = 0$.

7.2.6 le coefficient de corrélation

Le coefficient de corrélation est défini par :

$$\rho_{XY} = \rho(X,Y) = \frac{cov(X,Y)}{\sigma(X)\sigma(Y)} \tag{7.8}$$

7.3 Les caractéristiques de forme ou coefficients de Fisher

7.3.1 Le coefficient d'asymétrie

Le coefficient d'asymétrie γ_1 est défini par rapport à l'espérance mathématique :

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{E[(X - E[X])^3]}{\sigma^3} \tag{7.9}$$

C'est une valeur sans dimension qui n'est pas affecté par un changement d'origine et d'échelle .

Selon la valeur du coefficient d'asymétrie, la fonction de densité (ou le diagramme en bâtons pour les v.a. discrètes) prend une forme différente : étalée à droite, symétrique ou étalée à gauche (Figure 7.1).

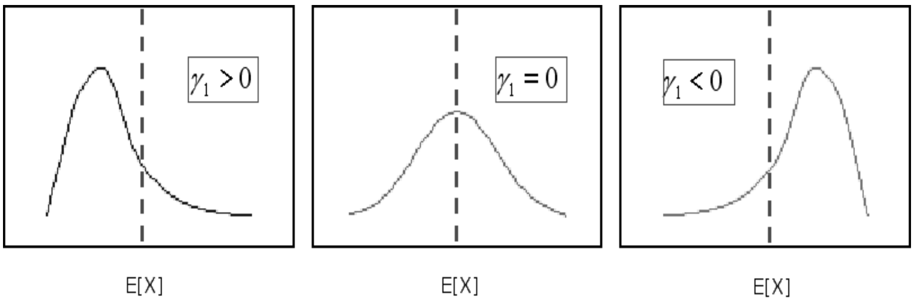

Fig. 7.1. Forme de la distribution selon le coefficient d'asymétrie

7.3.2 Le coefficient d'aplatissement

Le coefficient d'aplatissement vise à situer la hauteur de la courbe de desnité d'une loi par rapport à la référence qu'est la loi normale (Loi de Laplace-Gauss). Noté γ_2, sa formule est la suivante :

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 \tag{7.10}$$

C'est un coefficient sans dimension, invariant par changement d'échelle et de dimension. La constante 3 a été choisie de manière à ce que le coefficient d'aplatissement de la loi normale soit égale à $\gamma_2 = 0$.

Dans la figure 7.2, nous observons les aspects de la fonction de distribution correspond à différentes valeurs de γ_2.

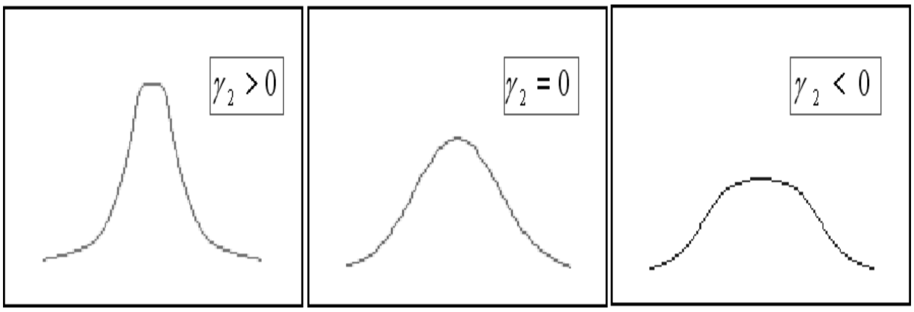

Fig. 7.2. Forme de la distribution selon le coefficient d'aplatissement

7.4 Fonctions génératrices

7.4.1 Fonction génératrice des moments

Définition 7. *Soit une variable réelle t, on appelle fonction génératrice des moments de X correspondant à t la quantité :*

$$\Psi_X(t) = E[e^{tX}] \tag{7.11}$$

Elle génère les moments non-centrés.

Exemple 17. Pour $t = 0$, nous avons :
- $\Psi_X(0) = 1$;
- $\Psi_X^{'}(0) = E[X]$;
- $\Psi_X^{''}(0) = E[X^2]$;
- $\cdots$;
- $\Psi_X^{(k)}(0) = E[X^k] = m_k$, moment non-centré d'ordre k.

7.4.2 Fonction génératrice des probabilités

Définition 8. *Soit une quantité certaine $u \in [0, 1]$. On appelle fonction génératrice des probabilités, notée $g_X(u)$, l'expression :*

$$g_X(u) = E[u^X] \tag{7.12}$$

La fonction génératrice des probabilités génère les moments factoriels.

Exemple 18. Soit X une v.a. discrète, $g_X(u) = E[u^X] = \sum_{x \in \mathcal{D}_X} u^x p_x$, nous observons alors :
- $\frac{dg_X(u)}{du} = \sum_{x=0}^{+\infty} x u^{x-1} p_x$; pour le cas particulier de $u = 1$, $\frac{dg_X(1)}{du} = \sum_{x=0}^{+\infty} x p_x = \mu_{[1]}(X)$
- $\frac{d^2 g_X(u)}{du^2} = \sum_{x=0}^{+\infty} x(x-1) u^{x-2} p_x$; pour e $u = 1$, $\frac{d^2 g_X(1)}{du} = \sum_{x=0}^{+\infty} x(x-1) p_x = \mu_{[2]}(X)$;
- $\cdots$;
- $\frac{d^k g_X(1)}{du} = \sum_{x=0}^{+\infty} x(x-1)\ldots(x-k+1) p_x = \mu_{[k]}(X)$

7.4.3 Fonctions caractéristiques

Définition 9 (Fonction caractéristique). *La fonction caractéristique d'une v.a. X est définie par*

$$\phi_X(t) = E[e^{itX}] \tag{7.13}$$

où i est l'unité imaginaire, avec $i^2 = -1$.

Son intérêt est qu'elle définit de manière univoque la loi d'une v.a. Si 2 v.a. présentent la même fonction caractéristique, on peut affirmer sans ambiguïté qu'elles suivent la même loi.

Selon le type de la v.a., la fonction caractéristique s'exprime différemment :
- pour une v.a. discrète, $\phi_X(t) = \sum_x e^{itx} p_x$
- pour une v.a. continue, $\phi_X(t) = \int e^{itX} f(x) dx$

Propriétés

Grâce à deux propriétés très importantes, les fonctions caractéristiques permettent de résoudre un grand nombre de problèmes de probabilités.

Caractéristique 30. Si $Y = aX$, où a est un scalaire non-aléatoire, alors $g_Y(t) = g_X(at)$

Caractéristique 31 (Fonction caractéristique d'une somme de v.a. indépendantes). La fonction caractéristique de $Z = Y + X$ est égale à $g_Z(t) = g_X(t) \times g_Y(t)$

Exemple 19 (Loi de la somme de 2 v.a. normales). Soit X (resp. Y) une v.a. distribuée selon une loi normale[1] de paramètres $\mathcal{N}(0, \sigma_x)$ (resp. $\mathcal{N}(0, \sigma_y)$) , sa fonction caractéristique est[2] égale à $g_X(t) = e^{-\frac{(\sigma_x t)^2}{2}}$ (resp. $g_Y(t) = e^{-\frac{(\sigma_y t)^2}{2}}$).

On veut déterminer la loi de distribution de $Z = X + Y$. Formons sa fonction caractéristique :

$$\begin{aligned} g_Z(t) &= g_X(t) \times g_Y(t) \\ &= e^{-\frac{(\sigma_x t)^2}{2}} \times e^{-\frac{(\sigma_y t)^2}{2}} \\ &= e^{-\frac{(\sigma_x t)^2 + (\sigma_y t)^2}{2}} \\ &= e^{-\frac{(\sigma_x^2 + \sigma_y^2) t^2}{2}} \end{aligned}$$

On en déduit aisément que $Z \sim \mathcal{N}(0, \sqrt{\sigma_x^2 + \sigma_y^2})$

7.5 Inégalité de Bienaymé-Chebychev

Définition 10 (Inégalité de Bienaymé-Chebychev). *Soit X une v.a. d'espérance mathématique μ et de variance σ^2. L'inégalité de Bienaymé-Chebychev indique que pour tout nombre réel positif t, la*

[1] Nous verrons plus loin dans ce support la définition de la loi normale. Ses deux paramètres sont la moyenne et l'écart-type.

[2] Pour plus de précisions, voir `http://www.proba.jussieu.fr/cours/processus-html/node18.html`

probabilité que X s'écarte de son espérance mathématique d'une grandeur inférieure à t, a pour limite supérieure $\frac{\sigma^2}{t^2}$:

$$P\{|X-\mu| \geq t\} \leq \frac{\sigma^2}{t^2} \tag{7.14}$$

Remarque 22 (Loi des grands nombres). Cette inégalité est importante car elle permet de démontrer la loi dite des "grands nombres". Elle stipule qu'il suffit d'extraire un *échantillon d'un effectif suffisant* dans une population pour estimer l'espérance mathématique à l'aide de la moyenne arithmétique, ou encore pour estimer une probabilité à l'aide d'une fréquence. On parle alors de *convergence en probabilité.*

Remarque 23 (Une expression plus générale de l'inégalité de Bienaymé-Chebychev). La portée de l'inégalité de Bienaymé-Chebychev est bien plus large que celle couramment présentée dans les ouvrages. Elle peut être définie pour une fonction $g(X)$ telle que

$$P\{g(X) \geq t\} \leq \frac{E[g(X)^k]}{t^k} \tag{7.15}$$

La présentation usuelle correspond à $k = 2$ et $g(x) = |X - \mu|$.

Définition 11 (Convergence en probabilité). *On considère une suite (X_n) d'une v.a. définie sur Ω, X une autre v.a. définie sur Ω. On dit que la suite (X_n) converge en probabilité vers X si*

$$\forall \epsilon > 0\,,\ \lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0 \tag{7.16}$$

Conséquence : Pour que (X_n) converge en probabilité vers X, il faut et il suffit que $E(X_n - X) \to 0$ et $V(X_n - X) \to 0$ lorsque $n \to \infty$ (la démonstration passe par l'inégalité de Bienaymé-Chebychev).

Partie II

Lois de probabilités d'usage courant

A priori, les lois de distribution des phénomènes physiques, économiques, etc. sont innombrables. Chaque cas semble particulier. En effet, quelle rapprochement pourrait-on faire entre la durée de vie d'une paire de chaussures et le temps d'attente à une caisse d'un grand magasin ?

En réalité, on se rend compte que la grande majorité des phénomènes statistiques peuvent être décrits par un nombre réduit de modèles probabilistes. Il importe dans un premier temps de pouvoir **décrire de façon adéquate le mécanisme du processus réel étudié** (temps d'attente, nombre de passages dans un intervalle de temps, nombre d'essais avant d'obtenir tel résultat, etc.).

Dans un second temps, une fois cette caractérisation réalisée, nous pouvons **choisir la loi théorique** qui paraît le mieux convenir pour modéliser le phénomène observé, l'étape suivante consistant à **estimer les paramètres de la loi**.

Enfin, dans un troisième et dernier temps, nous devons **nous assurer que le rapprochement** entre la loi théorique proposée et les données observées **est** statistiquement **crédible**.

Une fois la loi théorique identifiée et validée, il sera possible d'expliquer et d'interpréter le phénomène observé, voire de prédire les résultats futurs de l'expérience.

Dans cette partie, nous présentons les lois de probabilités les plus souvent utilisées dans les études. Elles permettent de modéliser une grande variété de problèmes. Notre démarche sera toujours la même : nous présentons d'abord le schéma de formation de la v.a., en le précisant le cas échéant à l'aide de quelques exemples ; la loi de probabilité est ensuite définie, nous mettons en exergue les paramètres de la loi ; les principales caractéristiques sont calculées (espérance, variance, coefficients d'asymétrie et d'aplatissement) ; et enfin, les estimateurs usuels des paramètres de la loi à partir d'un échantillon d'observations sont décrits.

Nous noterons X la v.a. étudiée. Une expérience correspond à une observation de la v.a. X, elle est notée x_i. Nous disposons d'une série d'observations $x_i, i = (1, 2, \ldots, n)$, donc d'un échantillon de taille n, pour estimer les paramètres de la loi à partir des données.

8

Les lois discrètes

8.1 Loi binomiale : suite d'épreuves de Bernoulli

8.1.1 Description

La loi binomiale décrit le nombre d'apparitions d'un évènement (disons "succès") dans une suite de m épreuves indépendantes. La probabilité d'apparition de l'évènement à une épreuve est p (a contrario la probabilité de non-apparition est $q = 1 - p$), elle est constante tout le long de l'expérience.

Exemple 20. On s'intéresse à l'apparition du côté "pile" si l'on lance m fois une pièce. A chaque jet, la probabilité d'apparition de "pile" est $p = \frac{1}{2}$.

8.1.2 Définition analytique

Le schéma d'échantillonnage peut être résumé dans le tableau suivant :

État	Probabilité	Tirage
"succès"	p	x
"échec"	$1 - p$	$m - x$
Total	1	m

La probabilité s'écrit :

$$P(X = x) = C_m^x p^x (1 - p)^{m-x} \tag{8.1}$$

La loi binomiale $\mathcal{B}(m, p)$ est définie par le paramètre p, qui indique la probabilité d'apparition de l'évènement pour une épreuve ; m est le nombre d'épreuves réalisées, c'est une donnée du problème, nous n'avons pas à l'estimer.

8.1.3 Caractéristiques de la loi

$$E(X) = mp$$
$$V(X) = mp(1-p)$$
$$\gamma_1 = \frac{1-2p}{\sqrt{mp(1-p)}}$$
$$\gamma_2 = \frac{1-6p(1-p)}{mp(1-p)}$$

Remarque 24. Il est possible d'obtenir ces caractéristiques à partir des fonctions génératrices des moments.

8.1.4 Estimation des paramètres. Loi de Bernoulli

Une expérience comprend m épreuves, le domaine de définition de x_i est $\{0, 1, \ldots, m\}$.

L'expérience est réitérée n fois. Nous disposons de n observations x_i suivant une loi $\mathcal{B}(m, p)$: m est une donnée, le seul paramètre à estimer est p

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{mn} \tag{8.2}$$

8.1.5 Cas particulier : la loi de Bernoulli

Bien souvent, l'expérience est réduite à une seule épreuve c.-à-d. $m = 1$. Dans ce cas, X suit la loi de Bernoulli $\mathcal{B}(1, p)$. La distribution de X est définie par $P(X = x) = p^x(1-p)^{1-x}$.

Dans ce contexte, x_i est une indicatrice égale à 1 si l'évènement survient à l'expérience (épreuve dans ce cas) numéro i, 0 sinon ; n étant le nombre d'épreuves réalisées. L'estimateur du paramètre p devient $\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}$

8.2 Loi de Pascal

8.2.1 Description

La loi de Pascal décrit le nombre d'épreuves indépendantes nécessaires pour obtenir k fois l'évènement étudié. La probabilité d'apparition de l'évènement à une épreuve est p, elle est constante tout le long de l' expérience.

Exemple 21. On s'intéresse au nombre de jets nécessaires pour le côté "pile" de la pièce apparaisse k fois. A chaque jet, la probabilité d'apparition de "pile" est $p = \frac{1}{2}$.

Exemple 22. Pour les grands joueurs, on peut également s'intéresser au nombre d'essais nécessaires pour qu'une combinaison apparaisse (enfin) au loto.

8.2.2 Définition analytique

$$P(X = x) = C_{x-1}^{k-1} p^k (1-p)^{x-k} \tag{8.3}$$

Le domaine de définition de X est $\mathcal{D}_X = \{k, k+1, \ldots\}$; k est le nombre de réalisation de l'évènement souhaité ; p est la probabilité de survenue de l'évènement à une épreuve.

8.2.3 Caractéristiques de la loi

$$
\begin{aligned}
E(X) &= \frac{k}{p} \\
V(X) &= \frac{k(1-p)}{p^2} \\
\gamma_1 &= \frac{2-p}{\sqrt{k(1-p)}} \\
\gamma_2 &= \frac{1+4(1-p)+(1-p)^2}{k(1-p)}
\end{aligned}
$$

8.2.4 Estimation des paramètres

Le seul paramètre à estimer est la probabilité p de survenue de l'évènement pour une épreuve. Nous disposons de n observations c.-à-d. l'expérience a été effectuée n fois. A chaque expérience, le nombre d'essais x_i nécessaires pour obtenir k succès (apparition de l'évènement) a été relevé.

$$
\hat{p} = \frac{k}{n}\sum_{i=1}^{n}\frac{1}{x_i} \tag{8.4}
$$

8.2.5 Cas particulier : la loi géométrique

La loi géométrique est la loi du nombre d'essais nécessaires pour faire apparaître l'évènement. Il correspond exactement à l'exemple du loto ci-dessus. La probabilité associée est $P(X = x) = p(1-p)^{x-1}$

8.2.6 Cas particulier : la loi binomiale négative

Dans certains ouvrages, il peut y avoir confusion entre la loi de Pascal et la loi Binomiale négative. La distinction est la suivante : la première indique *le nombre d'essais nécessaires pour obtenir le k-ème succès* ; la loi binomiale négative correspond au *nombre d'échecs précédent le k-ème succès.*

Si X suit une loi de Pascal, $Y = X - k$ suit une loi binomiale négative. Sa distribution est $P(Y = y) = C_{k+y-1}^{k-1}p^k(1-p)^y$.

8.3 Loi hypergéométrique

8.3.1 Description

Pour préciser les idées, nous dirons que nous sommes en présence d'une urne contenant M boules, B sont de couleur blanche qui nous intéresse, $M - B$ sont de couleur noire.

La loi hypergéométrique décrit le nombre X de boules blanches extraites de l'urne si nous effectuons un tirage exhaustif (sans remise) de m boules. La présentation sous forme de tableau du schéma d'échantillonnage permet d'appréhender le problème.

Couleur	Urne	Tirage
bb	B	x
bn	$M-B$	$m-x$
Total	M	m

Forcément, $m \leq M$, $x \leq B$ et $x \leq m$. Notons également que la composition de l'urne varie au cours de l'expérience.

8.3.2 Définition analytique

La distribution de la loi hypergéométrique s'écrit :

$$P(X=x) = \frac{C_B^x C_{M-B}^{m-x}}{C_M^m} \tag{8.5}$$

La loi est essentiellement décrite par son paramètre B ; M est une donnée de l'expérimentation.

Exemple 23. Une machine outil fournit M pièces par jour, M est une donnée du problème, nous nous intéressons au nombre B de pièces défectueuses produites par jour.

8.3.3 Caractéristiques de la loi

$$\begin{aligned}
E(X) &= m\frac{B}{M} \\
V(X) &= m\frac{B}{M-1}(1-\frac{B}{M})(1-\frac{m}{M}) \\
\gamma_1 &= \frac{(1-2\frac{B}{M})}{\sqrt{m\frac{B}{M}(1-\frac{B}{M})}\frac{(M-2m)\sqrt{M-1}}{(M-2)\sqrt{M-m}}} \\
\gamma_2 &= \ldots \text{expression très complexe et de peu d'utilité}
\end{aligned}$$

8.3.4 Estimation des paramètres

Si l'on reprend notre exemple de la machine outil ci-dessus, l'échantillonnage consiste à effectuer un prélèvement pendant n jours. Nous obtenons ainsi un échantillon de n observations, l'observation x_i indique le nombre de pièces défectueuses lors de chaque prélèvement.

B, le nombre de pièces défectueuses parmi les M pièces produites par jour, est le paramètre à estimer :

$$\hat{B} = \frac{M}{m}\frac{\sum_{i=1}^{n} x_i}{n} \tag{8.6}$$

8.3.5 Passage aux proportions et approximation par la loi binomiale

Travailler sur des proportions

Bien souvent, la loi hypergéométrique est présentée différemment dans la littérature. Nous avons bien une population de M individus, mais on s'intéresse plutôt à la proportion p des individus possédant la caractéristique étudiée en effectuant un tirage de m observations sans remise. Par rapport au schéma binomial, c'est le caractère exhaustif du tirage qui fait la différence.

Exemple 24. Dans un parking de supermarché, on s'intéresse à la proportion de véhicules ayant des jantes alliages. Nous n'allons faire un échantillonnage avec remise, au risque de reconsidérer 2 ou même plusieurs fois un même véhicule.

Par rapport à nos notations ci-dessus, on pose $B = Mp$, les caractéristiques de la loi deviennent :

$$E(X) = mp$$
$$V(X) = \frac{M-m}{M-1} mp(1-p)$$

Généralement $m > 1$, la variance de la loi hypergéométrique est inférieure à la variance de la loi binomiale, ce qui fait tout son intérêt.

Approximation

On peut montrer que lorsque la taille de la population M tend vers l'infini, les distributions des lois hypergéométriques et binomiales sont quasi-équivalentes.

On constate également que lorsque M est très grand par rapport à m, le rapport $\frac{M-m}{M-1} \rightarrow 1$ (Ex. *On interroge* 1000 *individus prélevés au hasard parmi les quelques* 44.5 *millions d'électeurs inscrits en France en* 2007), il est possible d'approximer la loi hypergéométrique à l'aide de la loi binomiale. Cela facilite le calcul des probabilités dans la pratique.

En général, on considère que l'approximation est valable dès que le rapport $\frac{m}{M} < 0.1$. Ce rapport est le *taux de sondage*.

8.4 Loi de Poisson

8.4.1 Description

La loi de Poisson, dite *loi des évènements rares*, est utilisée pour décrire les phénomènes où la probabilité de survenue de l'évènement qui nous intéresse est faible. Cette probabilité doit être constante tout le long de l'expérience. Dans une certaine mesure, on peut considérer la loi de Poisson comme un cas particulier de la loi binomiale.

Exemple 25. Nombre de pièces défectueuses produites par une machine outil.

La loi de Poisson peut être également utilisée spécifiquement pour décrire la survenue d'un évènement dans un intervalle de temps. On parle alors de *processus de Poisson*. Dans ce cas, la probabilité ne doit être influencée sur les expériences passées, elle ne doit pas influencer les expériences futures. Le nombre d'apparition de l'évènement ne dépend que de la durée d'observation.

Exemple 26. Nombre d'accidents sur une portion de route.

8.4.2 Définition analytique

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \tag{8.7}$$

avec $\lambda \geq 0$

8.4.3 Caractéristiques de la loi

La loi de Poisson $\mathcal{P}(\lambda)$ est entièrement définie par le paramètre λ, elle tient un rôle particulier. En effet :

$$\begin{aligned} E(X) &= \lambda \\ V(X) &= \lambda \\ \gamma_1 &= \frac{1}{\sqrt{\lambda}} \\ \gamma_2 &= \frac{1}{\lambda} \end{aligned}$$

8.4.4 Estimation des paramètres

Le seul paramètre à estimer est λ. Pour chaque expérience, nous disposons de x_i, le nombre d'apparition de l'évènement. L'estimation est :

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{8.8}$$

8.4.5 Approximation de la loi binomiale

Il y a des similitudes dans la définition de la loi binomiale et la loi de Poisson. La question est de savoir s'il est possible de procéder à des approximations et dans quel sens.

Comme son nom l'indique, la *loi des évènements rares*, la loi de Poisson propose une bonne approximation de la loi binomiale dès lors que la probabilité p de survenue de l'évènement élémentaire est suffisamment faible, $p < 0.1$ dans la pratique.

Exemple 27. Si m est le nombre de pièces sélectionnées au hasard dans la production d'une machine, p la proportion de pièces défectueuses – on peut espérer qu'elle est très petite – la loi du nombre de pièces défectueuses qui suit une loi binomiale $X \sim \mathcal{B}(m,p)$ peut être raisonnablement approchée à l'aide d'une loi de Poisson $\mathcal{P}(mp)$.

8.5 Loi multinomiale

8.5.1 Description

La loi multinomiale, ou *loi polynomiale*, est une généralisation de la loi binomiale, où l ($l > 2$) issues peuvent survenir lors d'une épreuve.

8.5.2 Définition analytique

Nous avons l états possibles E_k, $k = 1, \ldots, l$. La probabilité d'apparition de l'état k est p_k. Le schéma d'échantillonnage peut être résumé dans le tableau suivant :

État	Probabilité	Tirage
A_1	p_1	x_1
$\ldots$	$\ldots$	$\ldots$
A_k	p_k	x_k
$\ldots$	$\ldots$	$\ldots$
A_l	p_l	x_l
Total	1	m

La distribution est définie par :

$$P(X = \vec{x}) = P[X = (x_1, \ldots, x_l)] = \frac{m!}{x_1! \cdots x_l!} p_1^{x_1} \cdots p_l^{x_l} \tag{8.9}$$

Contrairement aux autres situations étudiées dans ce support, **la variable aléatoire est un vecteur**.

8.5.3 Caractéristiques de la loi

Nous proposons toujours l'espérance et la variance, mais pour chaque composante X_k du vecteur. De plus, comme elles ne sont pas indépendantes, nous devons également produire leur covariance.

$$\begin{aligned} E(X_k) &= mp_k \\ V(X_k) &= mp_k(1 - p_k) \\ COV(X_k, X_j) &= -mp_k p_j \,, k \neq j \end{aligned}$$

8.5.4 Estimation des paramètres

Chaque expérience permet de produire une observation, qui est un vecteur, $\vec{x_i} = (x_{i1}, \cdots, x_{ik}, \cdots, x_{il})$. La valeur x_{ik} correspond au nombre d'apparition de l'état E_k pour l'expérience numéro i. Au cours de cette expérience, m objets ont été tirés au hasard dans la population.

Nous disposons de n observations (n expériences). L'objectif est de proposer une estimation de la probabilité p_k de chaque état E_k.

$$\hat{p}_k = \frac{\sum_{i=1}^{n} x_{ik}}{mn} \tag{8.10}$$

9

Les lois continues

Les lois continues sont caractérisées à l'aide des fonctions de densité. Théoriquement, une v.a. continue doit pouvoir prendre n'importe quelle valeur dans un intervalle donné. C'est pour cela d'ailleurs que la probabilité ponctuelle est par définition nulle.

En réalité, dans certains cas, la distinction loi discrète - loi continue est essentiellement formelle. Il est possible d'approximer certaines lois discrètes à l'aide de lois continues.

9.1 Loi uniforme

9.1.1 Description

La principale caractéristique de la loi uniforme continue est que la probabilité d'être dans un intervalle dépend uniquement de la largeur de l'intervalle et non de la position de l'intervalle dans le domaine de définition.

La fonction de densité prend une forme rectangulaire. La loi est symétrique.

Remarque 25 (Loi discrète uniforme). On peut définir le pendant de la loi uniforme pour les v.a. discrètes. Il s'agit dans ce cas d'un cas particulier de la loi multinomiale où les états E_k sont équiprobables.

9.1.2 Définition analytique

La loi repose sur deux paramètres a et b qui sont respectivement la borne basse et la borne haute du domaine de définition de la v.a. X. La fonction de densité est alors :

$$f(x) = \begin{cases} \frac{1}{b-a}, \text{ si } \ a \leq x \leq b \\ 0, \text{ sinon} \end{cases}$$

Dans le cas de la loi uniforme, il est aisé de calculer la fonction de répartition :

$$F(x) = \frac{x-a}{b-a} \tag{9.1}$$

9.1.3 Caractéristiques de la loi

$$E(X) = \frac{a+b}{2}$$
$$V(X) = \frac{(b-a)^2}{12}$$
$$\gamma_1 = 0$$
$$\gamma_2 = -1.2$$

Comme la loi est par définition symétrique, il est naturel que $\gamma_1 = 0$.

9.1.4 Estimation des paramètres

Soient x_{min} (resp. x_{max}) la plus petite (resp. grande) valeur observée sur un échantillon de taille n. Les estimateurs des paramètres a et b de la loi sont :

$$\hat{a} = x_{min} - \frac{x_{max} - x_{min}}{n-1} \tag{9.2}$$
$$\hat{b} = x_{max} + \frac{x_{max} - x_{min}}{n-1} \tag{9.3}$$

On remarquera que l'on n'utilise pas directement les valeurs triviales x_{min} et x_{max} pour estimer les paramètres.

9.2 Loi normale

9.2.1 Description

Une v.a. X suit la loi normale (Loi gaussienne, Loi de Laplace-Gauss), lorsqu'elle est formée à partir d'un *grand nombre de facteurs s'additionnant les uns aux autres, aucune ne jouant un rôle prédominant, et agissant de manière indépendante.*

Considérée pendant longtemps comme la loi universelle, ses caractéristiques se retrouvant dans de nombreux domaines, dans les sciences dures (ex. erreurs de mesure en physique) mais également dans les sciences humaines et sociales (ex. fluctuations économiques), elle a été depuis recadrée.

Néanmoins, la loi normale garde une place particulière de par sa simplicité. Elle est à la base de nombreux résultats dans la statistique paramétrique, en théorie de l'estimation et théorie des tests.

Exemple 28 (Régression Linéaire Multiple). Dans la régression linéaire multiple, les lois utilisées pour l'évaluation globale de la régression et l'évaluation individuelle des coefficients reposent sur la normalité des résidus.

9.2.2 Définition analytique

La loi normale est paramétrée par la moyenne μ et l'écart-type σ, on note $X \sim \mathcal{N}(m, \sigma)$. Sa fonction de densité est :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \tag{9.4}$$

C'est la fameuse "courbe en cloche" (Figure 9.1). Elle est symétrique. Son aplatissement sert de référence pour situer les autres lois.

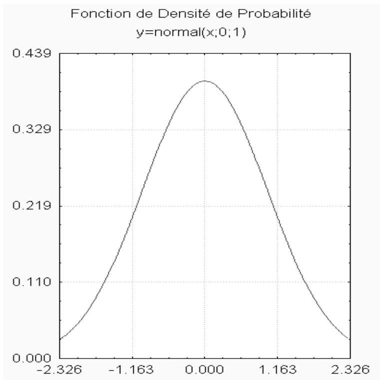

Fig. 9.1. Courbe de densité de la loi normale

Caractéristique 32 (Loi normale centrée réduite). Si $Z \sim \mathcal{N}(0, 1)$, alors $X = \sigma Z + \mu$ suit une loi $\mathcal{N}(\mu, \sigma)$. Grâce à cette propriété, seule Z est tabulée. Il est aisé d'extrapoler sur X, quelles que soient les valeurs de μ et σ.

A contrario, il est possible de ramener toute distribution normale à la loi normale centrée réduite en formant $Z = \frac{X-\mu}{\sigma}$. On montre facilement que $E[Z] = 0$ et $V(Z) = 1$.

Caractéristique 33 (Somme de v.a. normales indépendantes). Autre caractéristique très intéressante de la loi normale, la somme de lois normales indépendantes suit une loi normale.

Prenons le cas de 2 v.a. X_1 et X_2, si $X_1 \sim \mathcal{N}(\mu_1, \sigma_1)$ et $X_2 \sim \mathcal{N}(\mu_2, \sigma_2)$, alors $S \sim \mathcal{N}(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$.

9.2.3 Caractéristiques de la loi

$$E(X) = \mu$$
$$V(X) = \sigma^2$$
$$\gamma_1 = 0$$
$$\gamma_2 = 0$$

$\gamma_1 = 0$ et $\gamma_2 = 0$ sont des caractéristiques marquantes de la loi normale. Elles peuvent être utilisées pour vérifier la compatibilité d'une distribution empirique avec la loi normale (voir test de Bera-Jarque, test de d'Agostino).

9.2.4 Estimation des paramètres

Les deux paramètres à estimer sont μ et σ. Nous disposons de n réalisations x_i de la v.a. X.

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{9.5}$$

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{9.6}$$

9.2.5 Théorème central limite

La popularité de la loi normale repose, entre autres, sur une propriété remarquable que l'on appelle le *théorème limite central* (ou *théorème central limite*).

Théorème 1 (Théorème central limite). *Soient $X_1, X_2, \ldots, X_n$, des v.a. indépendantes de même loi,* ***quelle que soit cette loi,*** *de moyenne μ et d'écart type σ.*

La v.a. $S = \frac{1}{n}\sum_{i=1}^{n} X_i$ tend vers une loi normale $\mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$

Nous notons que ces éléments définissent les conditions de formation de la loi normale : les facteurs sont nombreux, indépendants, et agissent additivement.

Remarque 26. Ce théorème peut s'exprimer de manière plus générale encore. L'égalité des moyennes et des écarts type n'est pas requise pour que la somme tende vers la loi normale. L'expression des paramètres de S est simplement un peu plus compliquée.

9.2.6 Loi lognormale

La loi log-normale est un cas particulier de la loi normale. On dit que X suit une loi log-normale, si son logarithme, $Y = \ln(X)$, suit une loi normale.

Les conditions de formation de la loi log-normale est identique à la loi normale, à la différence que les facteurs agissent de manière multiplicative.

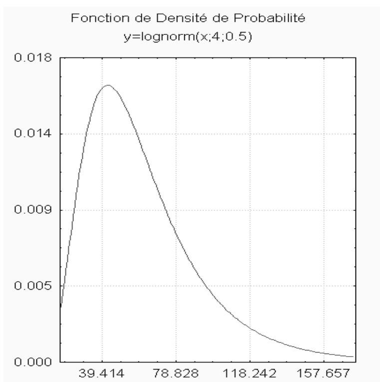

Fig. 9.2. Courbe de densité de la loi log-normale

C'est une loi définie dans $\mathbb{R}^+$, asymétrique, étalée à droite (Figure 9.2). Sa densité de probabilité est :

$$f(x) = \frac{1}{\sigma x\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\ln x-\mu}{\sigma})^2} \tag{9.7}$$

où $\mu = E[\ln(X)]$ et $\sigma^2 = V[\ln(X)]$.

On rencontre fréquemment cette loi dans la distribution des salaires ou des revenus. Un grand nombre de personnes disposent d'un salaire modeste (proche du SMIC par exemple), et à mesure que la valeur s'élève, la proportion des personnes concernées diminue rapidement. La médiane est très inférieure à la moyenne pour ce type de loi.

9.2.7 Approximation de la loi binomiale

L'approximation de la loi binomiale par la loi normale est une autre utilisation très commode de la loi normale. La v.a. $X \sim \mathcal{B}(m, p)$ peut être approximée par une loi normale de paramètres $\mu = mp$ et $\sigma = \sqrt{mp(1-p)}$ dès lors que m est *grand* (quelques dizaines d'unités) et que p n'est *pas trop proche* de 0 ou 1.

En effet, lorsque $m \to \infty$, X peut prendre une infinité de valeurs. En pratique, $mp(1-p) > 9$ suffit à assurer une approximation satisfaisante.

9.2.8 Approximation de la loi de Poisson

Dans le même ordre d'idée, lorsque $\lambda \to \infty$, il est possible d'approximer une loi de Poisson $\mathcal{P}(\lambda)$ par une loi normale $\mathcal{N}(\lambda, \sqrt{\lambda})$.

En pratique, dès $\lambda \geq 18$, nous obtenons une approximation satisfaisante (Figure 9.3).

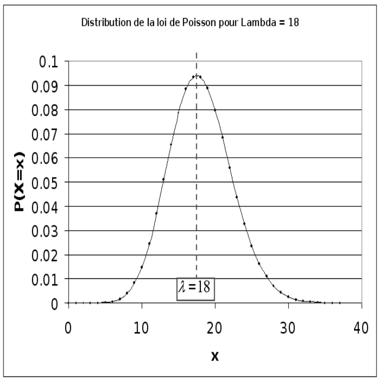

Fig. 9.3. Distribution de la loi de Poisson pour $\lambda = 18$

9.2.9 Transformation de variable

La normalité des variables, tout du moins leur symétrie, est une condition souvent exigée dans les tests paramétriques. Or les données observées peuvent ne pas être compatible avec cette hypothèse. Il est alors commode d'introduire des transformations de variables permettant de se ramener à une distribution proche de la loi normale. Le logarithme, vu plus haut, en est une ; la racine carrée peut être utilisée également.

Il existe plusieurs classes de fonctions paramétrables permettant d'accéder à des transformations que l'on peut ajuster selon notre convenance. Par ces fonctions, une des plus connue est la transformation de Box-Cox[1]. La fonction dépend de 2 paramètres k_1 et k_2, elle s'écrit :

$$z = \begin{cases} \frac{(x+k_2)^{k_1}-1}{k_1}\,, & (k_1 \neq 0) \\ \ln(x+k_2)\,, & (k_1 = 0) \end{cases}$$

Les paramètres k_1 et k_2 peuvent être déterminés par tâtonnement, ou encore par maximisation de la vraisemblance (*N.A. : mais c'est quand-même un peu le canon pour tuer la mouche...*).

9.3 Loi exponentielle

9.3.1 Description

Si la loi de Poisson peut être considérée comme le nombre de survenue d'un évènement dans un intervalle de temps, la loi exponentielle est l'intervalle de temps séparant deux évènements poissonniens.

Cela peut être la durée de vie d'une pièce, l'intervalle de temps séparant deux pannes consécutives (*N.A. : si on a une voiture comme la mienne, le temps d'attente ne sera pas excessif*), etc.

Cette loi est très utilisée en étude de fiabilité.

9.3.2 Définition analytique

La v.a. X est définie dans $\mathbb{R}^+$, elle est notée $\mathcal{E}(\lambda)$, sa fonction de densité est :

$$f(x) = \lambda e^{-\lambda x} \tag{9.8}$$

λ est le seul paramètre de la loi. La fonction de répartition est facilement obtenu :

$$F(x) = 1 - e^{-\lambda x} \tag{9.9}$$

Caractéristique 34 (Somme de 2 v.a. exponentielles). Soient $X_1 \sim \mathcal{E}(\lambda_1)$ et $X_2 \sim \mathcal{E}(\lambda_2)$, alors la v.a. $S = X_1 + X_2$ suit une loi exponentielle, de paramètre $S \sim \mathcal{E}(\lambda_1 + \lambda_2)$.

9.3.3 Caractéristiques de la loi

$$\begin{aligned} E(X) &= \frac{1}{\lambda} \\ V(X) &= \frac{1}{\lambda^2} \\ \gamma_1 &= 2 \\ \gamma_2 &= 6 \end{aligned}$$

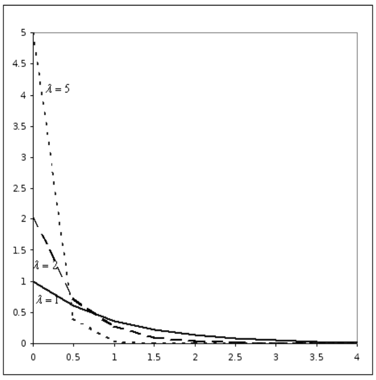

Fig. 9.4. Distribution de la loi exponentiel selon quelques valeurs de λ

Dans la figure 9.4, nous observons l'influence de λ sur la forme de la fonction de densité ; $\gamma_1 > 0$, la loi est effectivement étalée à droite.

9.3.4 Estimation des paramètres

Nous disposons que n réalisations x_i de la v.a. X, l'estimation du paramètre est :

$$\hat{\lambda} = \frac{1}{\bar{x}} \tag{9.10}$$

$\bar{x}$ est la moyenne empirique, $\bar{x} = \frac{1}{n}\sum_i x_i$.

9.3.5 Loi gamma

La loi γ_m décrit l'intervalle de temps entre le premier et le dernier évènement d'une suite de $(m+1)$ évènements successifs. On peut la considérer comme une généralisation de la loi exponentielle $\gamma_1(\lambda) \equiv \mathcal{E}(\lambda)$.

Définie dans $\mathbb{R}^+$, sa fonction de densité est paramétrée par m et λ :

$$f(x) = \frac{\lambda^m}{\Gamma(m)} x^{m-1} e^{-\lambda x} \quad , x \geq 0 \tag{9.11}$$

Définition 12 (Fonction $\Gamma()$). *La fonction $\Gamma()$, ou fonction Gamma d'Euler, est définie*[2] *par l'intégrale* $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$.

Plus prosaïquement, concernant les cas les plus souvent rencontrés, pour un entier m, elle est calculée à l'aide de

$$\Gamma(m) = (m-1) \times (m-2) \times \cdots \times \Gamma(1)\,,\ avec\ \Gamma(1) = 1$$
$$\Gamma(\frac{m}{2}) = (\frac{m}{2} - 1) \times (\frac{m}{2} - 2) \times \cdots \times \Gamma(\frac{1}{2})\,,\ avec\ \Gamma(\frac{1}{2}) = \sqrt{\pi}$$

Elle vérifie la relation de récurrence $\Gamma(m) = (m-1)\Gamma(m-1)$

[1] Pour une lecture approfondie du sujet, nous conseillons la lecture de Daumas, F., "Méthodes de normalisation des données", RSA, (30)4, 1982. Accessible en ligne sur le serveur `http://www.numdam.org`

[2] Pour une description détaillée, voir `http://fr.wikipedia.org/wiki/Fonction_gamma`

Les caractéristiques de la loi $\gamma_m(\lambda)$ sont :

$$
\begin{aligned}
E(X) &= \frac{m}{\lambda} \\
V(X) &= \frac{m}{\lambda^2} \\
\gamma_1 &= \frac{2}{\sqrt{m}} \\
\gamma_2 &= \frac{6}{m}
\end{aligned}
$$

9.4 Loi de Weibull

9.4.1 Description

Utilisée également en fiabilité, la loi de Weibull, par rapport à la loi exponentielle, introduit une information supplémentaire : le régime d'utilisation de la pièce.

La loi de Weibull est très utilisée dans les sciences de l'ingénieur[3].

9.4.2 Définition analytique

La v.a. X est définie dans $\mathbb{R}^+$. La fonction de densité dépend de 2 paramètres α et λ :

$$f(x) = \lambda \alpha x^{\alpha-1} e^{-\lambda x^{\alpha-1}} \tag{9.12}$$

Le paramètre α s'interprète de la manière suivante :

- $\alpha < 1$, la pièce évaluée est en rodage, elle s'améliore au fil du temps (un moteur neuf d'un véhicule par exemple) ;
- $\alpha = 1$, nous sommes en régime stationnaire, la pièce est utilisée normalement et ne subit pas de modifications de ses caractéristiques ;
- $\alpha > 1$, nous sommes en vieillissement accéléré, la pièce est utilisée dans des conditions extrêmes (en compétition par exemple).

9.4.3 Caractéristiques de la loi

Les caractéristiques de la loi sont les suivantes :

$$
\begin{aligned}
E(X) &= \lambda^{-\frac{1}{\alpha}} \Gamma(1+\frac{1}{\alpha}) \\
V(X) &= \lambda^{-\frac{2}{\alpha}} [\Gamma(1+\frac{2}{\alpha}) - \Gamma^2(1+\frac{1}{\alpha})] \\
\gamma_1 &= [\Gamma(1+\frac{2}{\alpha}) - \Gamma^2(1+\frac{1}{\alpha})]^{-3} [\Gamma(1+\frac{3}{\alpha}) - 3\Gamma(1+\frac{2}{\alpha})\Gamma(1+\frac{1}{\alpha}) + 2\Gamma^3(1+\frac{1}{\alpha})] \\
\gamma_2 &= [\Gamma(1+\frac{2}{\alpha}) - \Gamma^2(1+\frac{1}{\alpha})]^{-4} [\Gamma(1+\frac{4}{\alpha}) - 4\Gamma(1+\frac{3}{\alpha})\Gamma(1+\frac{1}{\alpha}) + 6\Gamma(1+\frac{2}{\alpha})\Gamma^2(1+\frac{1}{\alpha}) - 3\Gamma^4(1+\frac{1}{\alpha})] - 3
\end{aligned}
$$

Les coefficients d'asymétrie et d'aplatissement dépendent des paramètres de la fonction. Dans le figure 9.5, nous donnons une illustration de la fonction de densité.

[3] Voir par ex. `http://www.si.ens-cachan.fr/ressource/r34/r34.htm`

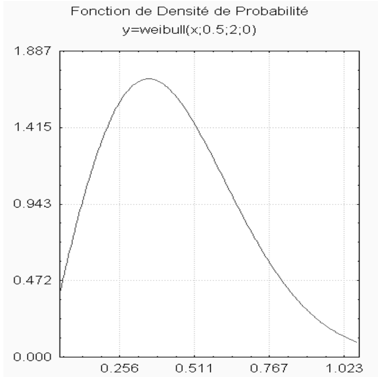

Fig. 9.5. Distribution de la loi de Weibull

9.4.4 Estimation des paramètres

Il n'y a pas d'estimation directe des paramètres de la loi de Weibull, il faut procéder en deux temps :

1. Tout d'abord, déterminer $\hat{\alpha}$ en résolvant l'équation

$$1 + \hat{\nu}^2 = \frac{\Gamma(1 + \frac{2}{\hat{\alpha}})}{\Gamma^2(1 + \frac{1}{\hat{\alpha}})}$$

   où $\hat{\nu} = \frac{\hat{\sigma}}{\bar{x}}$ est le coefficient de variation empirique ;

2. Puis en déduire $\hat{\lambda}$ en appliquant la formule

$$\hat{\lambda} = \left(\frac{\Gamma(1 + \frac{1}{\hat{\alpha})}}{\bar{x}} \right)^{\hat{\alpha}}$$

L'expression de $\hat{\alpha}$ semble rédhibitoire, on peut se contenter d'en obtenir une estimation simplifiée par tâtonnement ou par une approche graphique.

Remarque 27 (La fonction $ln\Gamma()$). Si la fonction $\Gamma()$ est assez ardue à calculer, surtout pour les valeurs non-usuelles (ni un entier, ni un demi-entier), la fonction $\ln \Gamma()$ est très souvent implémentée dans les bibliothèques de calcul. Pour prendre un exemple parmi tant d'autres, dans le tableur EXCEL, il s'agit de la fonction LNGAMMA().

9.5 Loi de Pareto

9.5.1 Description

La loi de Pareto s'applique pour les distributions tronquées. Prenons un exemple de la vie courante, en France, la borne basse du salaire horaire est forcément le SMIC, il ne peut pas en être autrement.

La loi de Pareto permet de tenir compte de cette contrainte en restreignant le domaine de définition de la v.a. X [4].

---

[4] Voir `http://fr.wikipedia.org/wiki/Distribution_de_Pareto`

9.5.2 Définition analytique

La loi possède 2 paramètres, $\alpha > 0$ et c qui introduit la contrainte $x > c$. Le domaine de définition de X est $]c, +\infty[$

La fonction de densité est monotone décroissante, elle s'écrit

$$f(x) = \frac{\alpha}{c}\left(\frac{c}{x}\right)^{\alpha+1} \tag{9.13}$$

La fonction de répartition est directement obtenue avec

$$F(x) = 1 - \left(\frac{c}{x}\right)^{\alpha} \tag{9.14}$$

9.5.3 Caractéristiques de la loi

$$\begin{aligned} E(X) &= \frac{\alpha}{\alpha - 1} \text{ , pour } \alpha > 1 \\ V(X) &= \frac{\alpha}{(\alpha-1)^2(\alpha-2)}c^2 \text{ , pour } \alpha > 2 \\ E[X^k] &= \frac{\alpha}{\alpha - c}c^k \text{ , pour } \alpha > k \end{aligned}$$

9.5.4 Estimation des paramètres

L'estimation des paramètres par la méthode du maximum de vraisemblance pour un échantillon de taille n permet d'obtenir

$$\begin{aligned} \hat{c} &= x_{min} - \frac{1}{n} \\ \hat{\alpha} &= \left[\frac{1}{n}\sum_{i=1}^{n} \ln \frac{x_i}{\hat{c}}\right]^{-1} \end{aligned}$$

9.6 Loi de Gumbel

9.6.1 Description

La loi de Gumbel est utilisée pour modéliser les valeurs extrêmes.

Par exemple, pour définir de manière adéquate la puissance d'un serveur, on s'intéresse au nombre maximal d'accès simultanés à un site web dans une journée, observé sur plusieurs jours.

9.6.2 Définition analytique

La loi n'est pas paramétrée, sa fonction de densité (Figure 9.6) est :

$$f(x) = e^{-x-e^{-x}} \tag{9.15}$$

La fonction de répartition est obtenue aisément :

$$F(x) = e^{-e^{-x}} \tag{9.16}$$

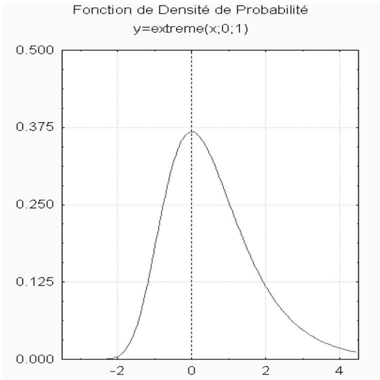

Fig. 9.6. Distribution de la loi de Gumbel ($\mu = 0$, $\beta = 1$)

9.6.3 Caractéristiques de la loi

$$
\begin{aligned}
E(X) &= 0.5772156649 \text{ ...constante d'Euler, notée } \gamma \text{ dans la littérature} \\
V(X) &= \frac{\pi^2}{6} \\
\gamma_1 &= 1.29857 \\
\gamma_2 &= 5.4
\end{aligned}
$$

La distribution n'est pas symétrique, elle est étalée à droite.

9.6.4 Loi généralisée de Gumbel

La loi généralisée de Gumbel est paramétrée par μ, un paramètre de localisation, et β, un paramètre d'échelle. La fonction de densité de la v.a. X devient alors, avec $z = e^{-\frac{x-\mu}{\beta}}$ pour simplifier l'écriture,

$$
f(x) = \frac{ze^{-z}}{\beta} \tag{9.17}
$$

Avec

$$
\begin{aligned}
E(X) &= \mu + \beta\gamma \\
V(X) &= \frac{\pi^2}{6}\beta^2
\end{aligned}
$$

La loi standard de Gumbel que nous avons présentée en premier lieu correspond aux paramètres $\mu = 0$ et $\beta = 1$.

10

Test d'adéquation à une loi

10.1 La démarche de modélisation

10.1.1 Objectifs de la modélisation

La modélisation d'un ensemble d'observations consiste à mettre en évidence les déterminants sous-jacents à l'origine de la formation de ces données. Dans notre cas, il s'agit d'identifier la loi de probabilité et les paramètres associés.

Remarque 28 (Statistique inférentielle). Nous entrons dans un cadre différent, celui de la *statistique inférentielle* que l'on appelle également *statistique mathématique*. Nous nous appuyons sur les données observées pour inférer sur les caractéristiques de la population[1], la loi théorique dans le cas présent. Nous nous contenterons d'en donner une description simplifiée et parcellaire de manière à comprendre les tests d'adéquation dans ce support. La statistique inférentielle fait partie d'un domaine à part, très vaste, qui fera l'objet d'autres supports.

En modélisant :
- Nous pouvons produire une description plus concise des données, et finalement une meilleure connaissance de la réalité.
- Nous délimitons mieux la probabilité de survenue des valeurs de la v.a. étudiée : ses caractéristiques, moyenne, écart-type ; les valeurs les plus probables, les valeurs qui ont très faible probabilité d'apparaître , etc.
- D'une certaine manière, cela nous permet aussi de prédire les valeurs des observations futures ou tout du moins la plage de ces valeurs en précisant la probabilité de leur apparition.
- Nous pouvons mesurer l'efficacité d'une action en comparant les paramètres de la loi de distribution avant et après l'opération (ex. on veut comparer la probabilité de survenue d'accidents à un carrefour, avant et après avoir installé une nouvelle signalisation).
- etc.

[1] voir par exemple `http://www.iut-lannion.fr/LEMEN/MPDOC/STAT/chap3/prech3.htm`

10.1.2 Les principales étapes

Dans le cas du test d'adéquation à une loi, la séquence des opérations est toujours la même[2] :

1. A partir de la description du problème, nous **définissons la loi** qui semble la mieux appropriée pour décrire les données.
2. Nous essayons alors d'en **estimer les paramètres**.
3. Puis nous validons la modélisation en **vérifiant la compatibilité** de la distribution théorique (loi de probabilité) avec la distribution empirique (observée).

Il faut bien s'entendre sur la validation. Il ne s'agit pas de certifier la loi de probabilité mais d'évaluer la compatibilité des données observées avec la loi théorique. Pour ce faire, nous devons mettre en place une mesure objective, une mesure qui quantifierait le désaccord entre les deux distributions. Par la suite, nous devrons déterminer statistiquement si cet écart est *faible*, essentiellement dû au hasard, les fluctuations d'échantillonnage ; ou s'il est *fort*, révélant une incompatibilité, et donc une inadéquation du modèle (loi théorique) proposé.

Dans la section qui vient, nous mettons en avant une mesure quantifiant l'écart entre 2 distributions et nous déterminons sa loi de probabilité[3]. Puis, dans les sections suivantes, nous mettrons en oeuvre le test d'adéquation sur des données simulées.

10.2 Test d'adéquation et loi du χ^2

10.2.1 Indicateur de distance, la statistique du χ^2

Nous voulons quantifier l'écart entre une distribution empirique et une distribution théorique que l'on a choisie.

Dans un premier temps, nous devons calculer les paramètres de la loi théorique. Nous utilisons à cet effet les estimateurs associés aux lois.

Dans un second temps, nous subdivisons les n valeurs observées de X en K intervalles de regroupement, on parle également de *classes* de regroupement. La largeur des intervalles n'est pas forcément constante. Chaque classe est définie par sa borne haute et sa borne basse. Posons o_k l'effectif associé à la classe Δ_k.

Puisque nous disposons maintenant des paramètres de la loi théorique, et que chaque intervalle est définie par ses bornes, nous pouvons calculer les probabilités p_k d'appartenir à chaque intervalle en utilisant la fonction de répartition. L'effectif théorique associé à l'intervalle Δ_k est égal à $t_k = n \times p_k$.

La statistique du χ^2_c (pour χ^2 calculé) quantifiant l'écart entre la distribution théorique et empirique s'écrit :

$$\chi^2_c = \sum_{k=1}^{K} \frac{(o_k - t_k)^2}{t_k} \tag{10.1}$$

[2] Un description de la démarche globale de modélisation en traitement de données est disponible sur le site du NIST "NIST/SEMATECH e-Handbook of Statistical Methods", `http://www.itl.nist.gov/div898/handbook/pmd/section4/pmd41.htm`

[3] Une description simplifiée, très accessible et pédagogique, du test d'adéquation du χ^2 est disponible sur le web, voir : `http://www.mission-laique.com/pedagogie/pdf/math56/am56_p59_72.pdf`

Par définition, $\chi_c^2 \geq 0$. Si $\chi_c^2 = 0$, la distribution théorique cadre parfaitement avec la distribution empirique. Si $\chi_c^2 > 0$, il y a des différences.

Il est évident que la valeur seuil 0 n'est pas du tout crédible dans ce type d'analyse. A cause des fluctuations d'échantillonnage, il est quasiment impossible que la loi observée et la loi théorique concorde parfaitement. Ce serait suspect d'ailleurs, laissant à imaginer que les données ont été manipulées.

Il nous faut proposer une procédure pour définir de manière raisonnée le seuil à partir duquel le χ_c^2 calculé est suffisamment important pour que la décision de rejeter l'hypothèse de compatibilité puisse être prise ; ou inversement, en deçà de quelle valeur nous pouvons considérer que l'écart est négligeable, imputable aux fluctuations d'échantillonnage. Pour rester dans un cadre probabiliste, il nous faut déterminer la loi de distribution de l'écart mesuré.

Remarque 29 (Contraintes sur la constitution des classes). Pour que la loi du χ^2 soit valide, il y a deux contraintes sur la constitution des classes. Tout d'abord, (1) les classes doivent être "assez nombreuses" ($K \geq 8$ selon certaines sources, mais cela dépend aussi du nombre de paramètres estimés). Mais a contrario, (2) il faut également que les effectifs théoriques dans chaque classe soit assez élevé ($t_k \geq 5$, a-t-on coutume de dire, ou tout du moins 80% des t_k doivent être ≥ 5). Si cette dernière condition n'est pas remplie, nous devons procéder à des regroupements de classes, au risque de buter sur la première contrainte[4].

10.2.2 Loi du χ^2

La loi du χ^2 est très utilisée, non pas pour modéliser des phénomènes observés, mais pour mettre en oeuvre des tests statistiques. Elle est formée par l'addition de lois normales centrées et réduites indépendantes. Elle est paramétrée par ν dit *degrés de liberté*. C'est une loi asymétrique, étalée à droite (Figure 10.1). Elle devient symétrique à mesure que le nombre de degrés de liberté augmente.

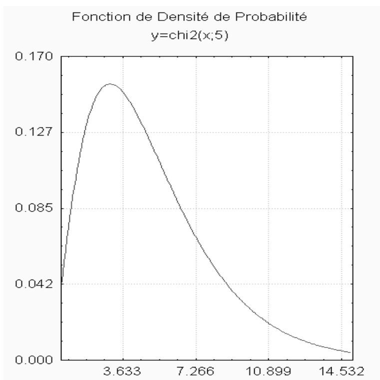

Fig. 10.1. Distribution de la loi du $\chi^2(5)$, à 5 degrés de liberté

Dans le test d'adéquation, le nombre de degrés de liberté est défini par :

[4] Pour une discussion sur les contraintes du test d'adéquation, voir `http://fr.wikipedia.org/wiki/Test_de_Khi-2`

$$\nu = K - r - 1 \tag{10.2}$$

où r est, pour la loi de probabilité modélisée, le nombre de paramètres estimés à partir des données.

Exemple 29. S'il s'agit de la loi normale, $r = 2$ car nous devons estimer la moyenne et l'écart-type ; pour la loi de Poisson, $r = 1$ puisque nous ne devons estimer qu'un seul paramètre ; etc.

Dans la théorie des tests, nous devons introduire un risque qui est une probabilité de se tromper. En effet, nous travaillons sur des phénomènes entachés de fluctuations. Fonder des décisions certaines sur de données observées est complètement illusoire. Il paraît plus raisonnable de prendre des décisions en évaluant au mieux les risques encourus.

Définition 13 (Risque de première espèce). *Dans le test d'adéquation, nous confrontons* 2 *hypothèses :* H_0*, l'hypothèse nulle, la modélisation des données avec la loi est licite ;* H_1*, l'hypothèse alternative, la loi n'est pas compatible avec les données. Dans ce cadre, nous sommes particulièrement sensibles à un risque (le risque de première espèce), c'est la probabilité* α *de décider que la loi modélisée est en désaccord avec les données observées alors qu'en réalité la loi est licite, on note* $\alpha = P(\textit{décider}\, H_1/H_0\, \textit{est vraie})$*. On parle aussi de probabilité de rejeter à tort l'hypothèse nulle.*

Dans la pratique statistique, il existe une série de valeurs ($\alpha = 0.01; 0.05; 0.1; \ldots$) communément utilisées. Elles nous permettent de spécifier notre niveau d'exigence par rapport aux résultats.

10.2.3 Mise en oeuvre du test d'adéquation

Pour mettre en oeuvre le test, nous devons dans un premier temps fixer le risque α (par ex. $\alpha = 0.05$). Ce risque α que l'on s'est choisit permet de subdiviser en 2 parties le domaine de définition du χ^2.

1. On appelle *région d'acceptation* (R.A.) l'intervalle $0 \leq \chi^2_c \leq \chi^2_{1-\alpha}(\nu)$. Lorsque le χ^2_c calculé est dans cet intervalle, nous concluons que les distributions théoriques et observées sont compatibles.
2. On appelle *région critique ou région de rejet* (R.C.) l'intervalle $\chi^2_{1-\alpha}(\nu) < \chi^2_c < +\infty$ qui permet de conclure à l'incompatibilité entre les distributions.

Le seuil s, que l'on appelle *valeur critique ou seuil critique*, qui permet de définir l'acceptation ou le rejet dépend bien évidemment de α. Formellement, il est définit de la manière suivante :

$$P(\chi^2(\nu) > s) = \alpha \tag{10.3}$$

s est donc le quantile d'ordre $1 - \alpha$ de la loi du χ^2 à ν degrés de liberté :

$$s = \chi^2_{1-\alpha}(\nu) \tag{10.4}$$

Dans la figure 10.2, nous montrons le quantile d'ordre $1 - \alpha$ pour une loi du χ^2 à 5 degrés de liberté.

Les seuils critiques peuvent être lues dans les tables statistiques que l'on trouve dans tous les bons ouvrages[5]. Les logiciels de statistique les fournissent également.

[5] On peut trouver d'excellentes références sur le Web, voir par ex. `http://courses.wcupa.edu/rbove/eco252/252suppkey.htm`

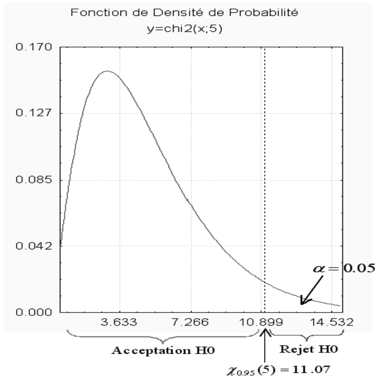

Fig. 10.2. Exemple de seuil critique $\chi^2_{0.95}(5)$ pour un risque $\alpha = 0.05$

Remarque 30 (La p-value). Souvent, les logiciels calculent directement la p-value α'. Elle correspond à la probabilité pour que la loi du $\chi^2(\nu)$ dépasse le χ^2_c calculé c.-à-d. $\alpha' = P(\chi^2(\nu) > \chi^2_c)$.

Dans ce cas, on compare directement la *p-value* au risque du test, on décide de rejeter l'hypothèse de compatibilité (hypothèse H_0) si $\alpha' < \alpha$. Cette règle de décision est totalement équivalente à celle définie par le seuil critique.

10.2.4 Exemple : adéquation avec la loi de Poisson

Dans un pays où il n'y a pas saisons, pas de jours fériés, ni de vacances, on vous demande de vous poster à un carrefour à un horaire donné et de noter le nombre de conducteurs qui omettent de mettre leur ceinture de sécurité. Vous vous postez au carrefour pendant 300 jours consécutifs. Vous êtes maintenant à la tête d'une série de 300 observations. On vous demande de déterminer la loi de distribution de la v.a. "absence de ceinture".

Statistique descriptive

Afficher directement le tableau des données ne sert absolument à rien. Il nous faut résumer l'information. Le plus simple est de construire l'histogramme de fréquence pour se donner une idée de la configuration des données (Figure 10.3).

La distribution est plus ou moins symétrique. La valeur la plus basse observée est 7, la plus élevée est 30. La valeur la plus fréquente, le mode, est $x = 19$.

En tous cas, elle ne présente pas des caractéristiques "gênantes". Si la distribution était bi-modale par exemple, aucune des lois dans la panoplie vue plus haut ne conviendrait.

Choix de la loi théorique et calcul des paramètres

Nous étudions la survenue d'un événement dans un intervalle de temps. Nous pensons à une loi de Poisson. Elle possède un seul ($r = 1$) paramètre λ que l'on estime à l'aide de la moyenne arithmétique.

$$\hat{\lambda} = \bar{x} = \frac{5371}{300} = 17.9$$

L'hypothèse que nous voulons vérifier est donc : la v.a. X "absence de ceinture de sécurité" est-elle compatible avec une loi de Poisson $\mathcal{P}(17.9)$?

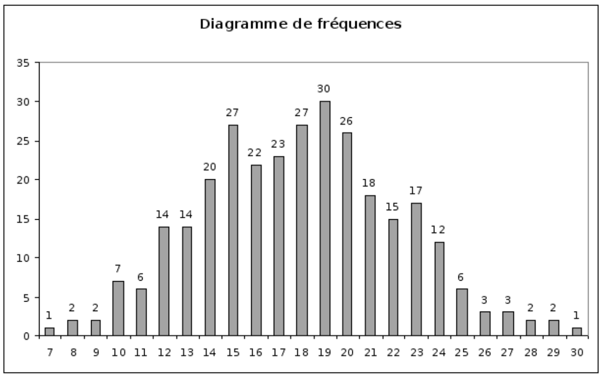

Fig. 10.3. Diagramme de fréquences des données "absence de ceinture"

Calcul de la statistique et conclusion

Dans un premier temps, sans prendre de précautions particulières, nous formons le tableau des effectifs observées. A partir de la loi $\mathcal{P}(17.9)$, nous calculons les fréquences relatives associées à chaque valeur. Nous produisons également les effectifs théoriques (Figure 10.4).

x	Effectifs empiriques	Fréquences théoriques	Effectifs théoriques
7	1	0.0020	0.6
8	2	0.0044	1.3
9	2	0.0088	2.6
10	7	0.0157	4.7
11	6	0.0255	7.6
12	14	0.0380	11.4
13	14	0.0524	15.7
14	20	0.0669	20.1
15	27	0.0799	24.0
16	22	0.0894	26.8
17	23	0.0941	28.2
18	27	0.0936	28.1
19	30	0.0882	26.4
20	26	0.0789	23.7
21	18	0.0673	20.2
22	15	0.0547	16.4
23	17	0.0426	12.8
24	12	0.0318	9.5
25	6	0.0227	6.8
26	3	0.0157	4.7
27	3	0.0104	3.1
28	2	0.0066	2.0
29	2	0.0041	1.2
30	1	0.0024	0.7
Total	300	0.9958	298.7

Fig. 10.4. Tableau des effectifs observés et théoriques

Le nombre de classes est $K = 24$, cela convient. En revanche, dans plusieurs cas, les effectifs théoriques sont faibles ($t_k < 5$), ce qui peut fausser les calculs.

De plus, nous constatons que la somme des fréquences relatives théoriques est inférieure à 1. Il y a des cas non recensés dans ce tableau. Dans le cas présent, il s'agit essentiellement des valeurs extrêmes.

Nous devons effectuer des regroupements et compléter notre tableau (Figure 10.5). Le tableau comporte maintenant $K = 19$ classes. Moins de 80% des classes ont un effectif théorique sensiblement inférieur à 5.

x	Effectifs empiriques	Fréquences théoriques	Effectifs théoriques
inf. ou égal à 9	5	0.0162	4.5
10	7	0.0157	4.7
11	6	0.0255	7.6
12	14	0.0380	11.4
13	14	0.0524	15.7
14	20	0.0669	20.1
15	27	0.0799	24.0
16	22	0.0894	26.8
17	23	0.0941	28.2
18	27	0.0936	28.1
19	30	0.0882	26.4
20	26	0.0789	23.7
21	18	0.0673	20.2
22	15	0.0547	16.4
23	17	0.0426	12.8
24	12	0.0318	9.5
25	6	0.0227	6.8
26	3	0.0157	4.7
27 et +	8	0.0266	8.0
Total	300	1.0000	300.0

Fig. 10.5. Tableau des effectifs observés et théoriques après regroupement

La confrontation des distributions empiriques et théoriques peut être représentée dans un diagramme de fréquences (Figure 10.6).

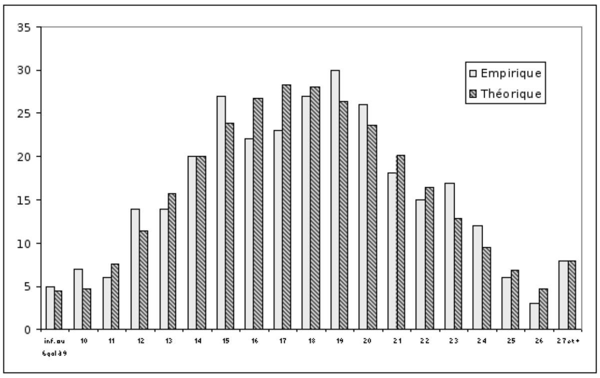

Fig. 10.6. Diagramme des fréquences empiriques et théoriques

L'étape suivante consiste à calculer les écarts entre les distributions empiriques et théoriques. Dans un tableur, il s'agit tout simplement de rajouter une nouvelle colonne, puis d'en faire la somme (Figure 10.7).

La statistique du χ^2 est

$$\chi^2_c = 8.4$$

x	Effectifs empiriques	Fréquences théoriques	Effectifs théoriques	(Obs - Thq)^2/Thq
inf. ou égal à 9	5	0.0162	4.5	0.0
10	7	0.0157	4.7	1.1
11	6	0.0255	7.6	0.4
12	14	0.0380	11.4	0.6
13	14	0.0524	15.7	0.2
14	20	0.0669	20.1	0.0
15	27	0.0799	24.0	0.4
16	22	0.0894	26.8	0.9
17	23	0.0941	28.2	1.0
18	27	0.0936	28.1	0.0
19	30	0.0882	26.4	0.5
20	26	0.0789	23.7	0.2
21	18	0.0673	20.2	0.2
22	15	0.0547	16.4	0.1
23	17	0.0426	12.8	1.4
24	12	0.0318	9.5	0.6
25	6	0.0227	6.8	0.1
26	3	0.0157	4.7	0.6
27 et +	8	0.0266	8.0	0.0
Total	300			8.4

Fig. 10.7. Tableau des effectifs observés et théoriques, calcul de la statistique du χ^2

Nous fixons notre risque à $\alpha = 0.05$, le nombre de degrés de liberté est $\nu = 19 - 1 - 1 = 17$, ce qui nous permet d'obtenir un seuil critique égal à $\chi_{0.95}(17) = 27.58$.

Le χ^2_c est largement inférieur à $\chi_{0.95}(17)$, nous pouvons affirmer que notre distribution est compatible avec une loi de Poisson $\mathcal{P}(17.9)$ au risque de $\alpha = 0.05$[6].

Si nous raisonnons en termes de *p-value*, $\alpha' = 0.957$, largement supérieure au risque α. L'hypothèse nulle ne peut pas être rejetée. Les décisions concordent.

10.2.5 Exemple : adéquation avec la loi Normale

Plus étonnant est ce second exemple. Nous voulons tester maintenant si la distribution observée est compatible avec une distribution normale. Pourquoi ? Il ne s'agit plus d'une loi de Poisson finalement ?

Il faut faire très attention à ce stade. Répétons-le, le test d'adéquation ne vise pas à déterminer la prétendue *bonne* loi qui régirait les données. N'allez surtout pas demander au conducteur contrevenant s'il est un processus de Poisson ou pas. Le rôle du test d'adéquation est simplement de vérifier la compatibilité de la loi théorique avec la distribution observée. Charge au statisticien d'exploiter les résultats par la suite, aux fins de prédiction, d'explication, etc.

Dans cet exemple, nous avions constaté dans le diagramme de fréquence (Figure 10.3) que la distribution empirique était mono-modale, quasiment symétrique ($\hat{\gamma}_1 = 0.11$). L'idée de la confronter à la loi normale n'est pas totalement farfelue.

Suivant la procédure habituelle, nous avons tout d'abord calculé les estimations[7] des paramètres $r = 2$ de la loi normale $\hat{\mu} = \bar{x} = 17.9$ et $\hat{\sigma} = 4.27$.

Nous pouvons dès lors élaborer le tableau de calcul. Il est très important de noter que chaque ligne du tableau définit un intervalle dorénavant. Par ex., la ligne correspondant à la valeur 18 correspond en fait à l'intervalle $17 < x \leq 18$. La dernière colonne nous sert à calculer la statistique du test (Figure 10.8) :

---

[6] Pour être tout à fait franc, je ne suis pas trop étonné, les données ont été générées sous le logiciel R (`http://www.r-project.org/`) avec une loi de Poisson $\mathcal{P}(18)$. Une incompatibilité aurait été assez inquiétante.

[7] Nous utilisons bien l'estimateur non-biaisé de l'écart type décrit en 9.2.

x	Effectifs empiriques	Fréquences théoriques	Effectifs théoriques	(Obs - Thq)^2/Thq
inf. ou égal à 9	5	0.0186	5.6	0.1
10	7	0.0136	4.1	2.1
11	6	0.0209	6.3	0.0
12	14	0.0305	9.1	2.6
13	14	0.0420	12.6	0.2
14	20	0.0549	16.5	0.8
15	27	0.0679	20.4	2.2
16	22	0.0796	23.9	0.1
17	23	0.0883	26.5	0.5
18	27	0.0927	27.8	0.0
19	30	0.0923	27.7	0.2
20	26	0.0869	26.1	0.0
21	18	0.0775	23.3	1.2
22	15	0.0655	19.6	1.1
23	17	0.0524	15.7	0.1
24	12	0.0397	11.9	0.0
25	6	0.0284	8.5	0.8
26	3	0.0193	5.8	1.3
27 et +	8	0.0290	8.7	0.1
Total	300			13.2

Fig. 10.8. Tableau de calcul de la statistique du χ^2, test d'adéquation à la loi normale

$$\chi_c^2 = 13.2$$

Nous conservons le risque $\alpha = 0.05$, le nombre de degrés de liberté devient $\nu = 19 - 2 - 1 = 16$. Le seuil critique du test est égal à $\chi_{0.95}(16) = 26.29$.

L'hypothèse de compatibilité à la loi normale ne peut être rejetée. Quoi de plus naturel finalement ? Nous avions vu plus tôt que la loi de Poisson convergeait vers la loi normale lorsque le paramètre λ était élevé, et qu'en pratique, l'approximation était correcte dès que $\lambda \geq 18$. Nous sommes tout à fait dans ce cas de figure.

Si on s'intéresse à la *p-value*, elle est égale à $\alpha' = 0.659$. La décision reste très tranchée, elle l'est un peu moins par rapport à la modélisation à l'aide de la loi de Poisson.

10.3 Test d'adéquation de Kolmogorov-Smirnov

Le test du χ^2 convient quelle que soit la nature de la v.a. étudiée. Si on veut tester l'adéquation à une loi multinomiale d'une v.a. discrète, elle convient. Si on veut tester une v.a. continue, il suffit de la découper en intervalles et elle peut s'appliquer également. Mais dans ce cas elle est peu puissante. En effet, intervertir les lignes du tableau de calcul n'a pas d'incidence sur les résultats. Or cette interversion n'a aucun sens si la v.a. est continue, les modalités (les intervalles dans ce cas) sont ordonnées. Modifier cet ordre devrait altérer, ou même rendre caduque, les conclusions du test d'adéquation.

Lorsque la v.a. X est continue, ou tout du moins prend des valeurs ordonnées, le test de Kolmogorov-Smirnov (K.S.) est une alternative plus puissante au test du χ^2. Elle s'appuie sur la comparaison de la fonction de répartition empirique avec la fonction de répartition théorique construite à partir de la loi modélisée. Pour que ce rapprochement ait un sens, il est évident que la v.a. doit être ordonnée.

En revanche, et c'est bien là l'inconvénient de cette approche par rapport au test du χ^2, les paramètres de la loi sont censés tous connus et non pas estimés à partir de l'échantillon[8,9].

Exemple 30. Reprenons notre exemple ci-dessus. Nous savons que les données ont été générées à partir d'un générateur de nombre pseudo-aléatoires suivant une loi $\mathcal{P}(18)$. En passant à la loi normale, ses paramètres sont donc $\mu = 18$ et $\sigma = \sqrt{18} = 4.2426$. Nous testerons donc l'adéquation de la distribution empirique avec la loi $\mathcal{N}(18, 4.2426)$.

10.3.1 Test de Kolmogorov-Smirnov

Soit $F_n(x)$ la fonction de répartition empirique observées sur les n observations x_i. A partir de la loi et de ses paramètres, nous disposons de la fonction de répartition théorique $F_0(x)$ que nous pouvons appliquer sur chaque observation x_i.

Le test de Kolmogorov-Smirnov repose sur la comparaison de ces deux fonctions de répartitions, plusieurs statistiques peuvent être définies :

$$
\begin{aligned}
K_n^+ &= \sup_x (F_n(x) - F_0(x)) \\
K_n^- &= \sup_x (F_0(x) - F_n(x)) \\
K_n &= \sup_x |F_0(x) - F_n(x)|
\end{aligned}
$$

En pratique, à partir des n observations x_i, nous calculons :

$$
\begin{aligned}
K_n^+ &= \max_i (\frac{i-1}{n} - F_0(x_i)) \\
K_n^- &= \max_i (F_0(x_i) - \frac{i}{n}) \\
K_n &= \max(K_n^+, K_n^-)
\end{aligned}
$$

Les lois de ces statistiques ont été tabulées sous leur forme exacte et asymptotique[10]. Il suffit, pour accepter ou rejeter l'adéquation, de comparer la statistique calculée avec le seuil critique lu dans la table.

En ce qui concerne la statistique K_n, lorsque n est grand ($n \geq 80$) et pour un risque $\alpha \leq 0.2$, il est possible d'approximer le seuil critique à partir de la formule :

$$
d_\alpha \approx \frac{\sqrt{-\frac{1}{2} \ln \frac{\alpha}{2}}}{\sqrt{n}} \tag{10.5}
$$

La région critique du test (rejet de l'hypothèse de compatibilité) est définie par :

$$
R.C. : K_n \geq d_\alpha
$$

[8] Dans le cas spécifique du test d'adéquation à la loi normale, des modifications peuvent être introduites pour que le test de Kolmogorov reste applicable alors que les paramètres μ et σ sont estimés à partir de l'échantillon

[9] Est-ce vraiment un inconvénient dans la pratique ? Généralement le choix de la loi repose sur une expertise. De la même manière le choix des paramètres de la loi peut s'appuyer sur les connaissances du domaine, la littérature, les résultats dans des travaux similaires, etc.

[10] Voir par exemple Tassi, page 317

10.3.2 Mise en oeuvre du test

Les calculs sont facilement mis en oeuvre dans un tableur. Il nous faut, dans un premier temps, trier les données selon les x_i croissants. Puis nous formons deux nouvelles colonnes correspondant à $\frac{i-1}{n}$ et $\frac{i}{n}$. Enfin, nous créons une dernière colonne avec la fonction de répartition théorique $F_0(x_i)$ (Figure 10.9)

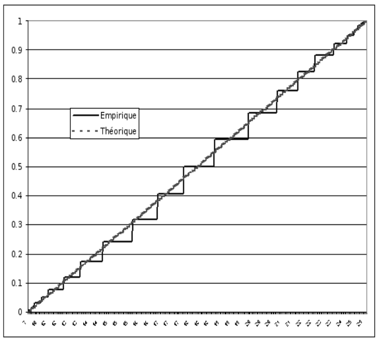

Fig. 10.9. Juxtaposition des fonctions de répartition théorique $F_0(x)$ et empirique $\frac{i-1}{n}$

Exemple 31. Pour illustrer notre propos, avec la loi normale et les paramètres ci-dessus, le tableur EXCEL[11] fournit $F_0(10) = 0.02967$, etc.

Les calculs fournissent les résultats suivants :

$$K_n^+ = 0.0669$$
$$K_n^- = 0.0398$$
$$K_n = 0.0669$$

Le seuil critique pour un risque $\alpha = 0.05$ est :

$$d_{0.05} = \frac{\sqrt{-\frac{1}{2}\ln\frac{0.05}{2}}}{\sqrt{300}} = 0.0784$$

Nous constatons que $K_n < d_{0.05}$, nous ne sommes pas dans la région critique du test, l'hypothèse de compatibilité de la distribution empirique avec la loi normale ne peut être rejetée.

Remarque 31 (Cohérence entre le test de K.S. et le test du χ^2). Dans cet exemple, le test de Kolmogorov-Smirnov est cohérent avec les conclusions du test d'adéquation du χ^2. Tout est pour le mieux. **S'il y avait désaccord, le test de Kolmogorov-Smirnov concluant au rejet de l'adéquation. Étant plus puissant, c.-à-d. il détecte mieux l'hypothèse alternative lorsqu'elle est vraie, nous accorderons plus de crédit à sa décision.**

[11] =LOI.NORMALE(10,18,RACINE(18),1)

Littérature

1. Aïvazian, S., Enukov, I., Mechalkine, L., *Eléments de modélisation et traitement primaire des données*, Mir, 1986.
2. Grais, B., *Méthodes Statistiques*, Dunod, 3ème édition, 2003.
3. Jolion, J.M., *Probabilités et Statistique*, `http://rfv.insa-lyon.fr/~jolion/STAT/poly.html`
4. Le Men, J.F., *Statistiques*, `http://www.iut-lannion.fr/LEMEN/MPDOC/STAT/presstat.htm`
5. Pelat D., *Bruits et Signaux (introduction aux méthodes de traitements des données) : statistique des variables aléatoires*, `http://www.lpthe.jussieu.fr/DEA/pelat.ps`
6. Saporta, G., *Probabilités, Analyse des données et Statistique*, Technip, 2ème édition, 2006.
7. Tassi, P., *Méthodes statistiques*, Economica, 1992.
8. Veysseyre R., *Aide Mémoire. Statistique et probabilités pour l'ingénieur*, Dunod, 2ème édition, 2006.