

Cours de probabilités

ÉLÉMENTS DE CALCUL DES PROBABILITÉS

SOMMAIRE

1	NOTION D'EVENEMENT ET DE PROBABILITE	6
1.1	PRESENTATION	6
1.2	PROBABILITE SUR UN ESPACE FINI	6
1.3	PROBABILITE SUR UN ESPACE QUELCONQUE	7
1.4	RAPPEL SUR L'ALGEBRE DES ENSEMBLES	8
1.5	APPLICATIONS : QUATRE COROLLAIRES DES AXIOMES	9
1.5.1	Corollaire 1 : probabilité d'un événement impossible	9
1.5.2	Corollaire 2 : probabilité de l'événement complémentaire	9
1.5.3	Corollaire 3 : théorème des probabilités totales	9
1.5.4	Corollaire 4 : théorème de l'inclusion	9
1.6	EXERCICES	10
1.6.1	Énoncés	10
1.6.2	Solutions	11
2	PROBABILITES CONDITIONNELLES & INDEPENDANCE	13
2.1	PROBABILITES CONDITIONNELLES	13
2.1.1	Définition	13
2.1.2	Interprétation par fréquence relative	13
2.1.3	Cas particuliers	13
2.2	ÉVENEMENTS INDEPENDANTS	14
2.2.1	Indépendance de deux événements	14
2.2.2	Interprétation par fréquence relative	14
2.3	THEOREME DES PROBABILITES TOTALES & THEOREME DE BAYES	15
2.4	EXERCICES	17
2.4.1	Énoncés	17
2.4.2	Solutions	18
3	VARIABLE ALEATOIRE REELLE	20
3.1	NOTION DE VARIABLE ALEATOIRE REELLE	20
3.1.1	Introduction	20
3.1.2	Définition	21
3.2	FONCTION DE REPARTITION	22
3.2.1	Définition	22
3.2.2	Propriétés	25
3.2.3	Interprétation par fréquence relative	25
3.3	FONCTION DENSITE DE PROBABILITE	26
3.3.1	Définition générale	26
3.3.2	Propriétés de la densité	28
3.4	VARIABLE ALEATOIRE CONTINUE	28
3.4.1	Définition	28
3.4.2	Interprétation fréquentielle de la densité	29
3.5	VARIABLE ALEATOIRE DISCRETE	30
3.6	EXEMPLES DE LOI DE PROBABILITE	31
3.6.1	Lois discontinues	31
3.6.2	Lois continues	32
3.7	TRANSFORMATION D'UNE VARIABLE ALEATOIRE REELLE	35
3.7.1	Transformation de la densité	35
3.7.2	Exemples	37
3.7.3	Cas particulier d'une transformation constante sur un intervalle	38
3.7.4	Cas particulier d'une transformation discontinue en un point	40
3.8	EXERCICES	42
3.8.1	Énoncés	42
3.8.2	Solutions	43
4	VALEURS TYPQUES D'UNE VARIABLE ALEATOIRE REELLE	48
4.1	ESPERANCE MATHEMATIQUE	48

4.1.1	Définition	48
4.1.2	Propriétés.....	48
4.1.3	Interprétation fréquentielle de l'espérance	49
4.1.4	Espérance d'une fonction certaine de la variable aléatoire	50
4.2	VALEUR EQUIPROBABLE (OU MEDIANE)	51
4.3	VALEUR LA PLUS PROBABLE	51
4.4	PERCENTILE OU VALEUR A P%	51
4.5	MOMENTS D'UNE VARIABLE ALEATOIRE	51
4.5.1	Définition	51
4.5.2	Moment du second ordre – variance et écart-type.....	51
4.5.3	Théorème de Bienaymé-Tchebychev	53
4.5.4	Moment centré.....	54
4.5.5	Moments absolus et moments généralisés.....	55
4.6	MOMENTS DES REPARTITIONS COURANTES	55
4.7	EXERCICES	57
4.7.1	Énoncés.....	57
4.7.2	Solutions.....	58
5	FONCTION CARACTERISTIQUE D'UNE VARIABLE ALEATOIRE REELLE	61
5.1	DEFINITION	61
5.2	PROPRIETES.....	62
5.2.1	Existence et borne supérieure	62
5.2.2	Symétrie.....	62
5.2.3	Variable centrée	63
5.2.4	Formule d'inversion.....	63
5.2.5	Théorème fondamental.....	63
5.3	EXEMPLES DE FONCTION CARACTERISTIQUE	63
5.3.1	Exemple 1 : variable aléatoire indicatrice.....	63
5.3.2	Exemple 2 : densité exponentielle	64
5.3.3	Exemple 3 : densité gamma pour n entier.....	64
5.4	THEOREME DES MOMENTS.....	65
5.5	EXERCICES	66
5.5.1	Énoncés.....	66
5.5.2	Solutions.....	67
6	VARIABLES ALEATOIRES REELLES BIDIMENSIONNELLES	69
6.1	GENERALITES.....	69
6.2	COUPLE DE VARIABLES DISCRETES.....	69
6.2.1	Introduction.....	69
6.2.2	Définitions pour une variable discrète	73
6.2.3	Loi de probabilité du couple	73
6.2.4	Lois marginales.....	74
6.2.5	Lois conditionnelles ou lois liées.....	74
6.2.6	Indépendance	75
6.3	VARIABLES ALEATOIRES REELLES BIDIMENSIONNELLE CONTINUE	76
6.3.1	Fonction de répartition bidimensionnelle	76
6.3.2	Fonctions de répartition marginales.....	76
6.3.3	Valeurs asymptotiques de la fonction de répartition.....	77
6.3.4	Calcul de la probabilité d'un rectangle	77
6.3.5	Fonction de densité bidimensionnelle	78
6.3.6	Interprétation infinitésimale de la densité de probabilité	79
6.3.7	Utilisation de la densité.....	79
6.3.8	Volume sous la densité.....	80
6.3.9	Relation entre la fonction de répartition et la densité bidimensionnelle.....	80
6.3.10	Densité marginale.....	80
6.3.11	Lois conditionnelles.....	82
6.3.12	Variable aléatoires indépendantes.....	83
6.4	VARIABLES ALEATOIRES GAUSSIENNES.....	83
6.4.1	Définition	83
6.4.2	Variables aléatoires gaussiennes indépendantes	84

6.5	TRANSFORMATION DE DEUX VARIABLES ALEATOIRES.....	84
6.5.1	Introduction.....	84
6.5.2	Variable aléatoire 1D fonction de deux variables aléatoires 1D.....	85
6.5.3	Application 1 : somme de deux variables aléatoires.....	86
6.5.4	Application 2 : Opération Min et Max.....	89
6.5.5	Application 3 : Rapport et produit.....	90
6.5.6	Variable aléatoire 2D fonctions de deux variables aléatoires 1D.....	91
6.5.7	Applications 1 : transformation linéaire.....	93
6.5.8	Applications 2 : transformation $Z = \sqrt{X^2 + Y^2}$ et $W = Y/X$	93
6.5.9	Applications 3 : Variables aléatoires auxiliaires.....	94
6.6	EXERCICES.....	95
6.6.1	Énoncés.....	95
6.6.2	Solutions.....	98
7	ESPERANCE, FONCTION CARACTERISTIQUE ET MOMENTS POUR 2 VARIABLES ALEATOIRES	102
7.1	ESPERANCE D'UNE FONCTION	102
7.2	FONCTION CARACTERISTIQUE	103
7.2.1	Définition	103
7.3	THEOREME DES MOMENTS.....	103
7.4	MOMENTS POUR DEUX VARIABLES ALEATOIRES	103
7.4.1	Définition	103
7.4.2	Covariance.....	105
7.4.3	Coefficient de corrélation.....	106
7.4.4	Corrélation et indépendance.....	108
7.5	EXERCICES.....	110
7.5.1	Énoncés.....	110
7.5.2	Solutions.....	111
8	VARIABLE ALEATOIRE N DIMENSIONNELLE	113
8.1	FONCTION DE REPARTITION ET DE DENSITE	113
8.1.1	Définitions.....	113
8.1.2	Propriétés.....	113
8.2	TRANSFORMATIONS	114
8.2.1	Une fonction de n variables	114
8.2.2	N fonctions de n variables.....	114
8.3	PROBABILITE CONDITIONNELLE ET INDEPENDANCE	115
8.3.1	Probabilité conditionnelle.....	115
8.3.2	Variables aléatoires indépendantes	115
8.4	ESPERANCE MATHEMATIQUE ET FONCTION CARACTERISTIQUE	115
8.4.1	Espérance mathématique	115
8.4.2	Fonction caractéristique	116
8.4.3	Non corrélation, Orthogonalité et Indépendance.....	116
8.4.4	Application : Introduction au problème d'estimation d'une variable aléatoire.....	117
8.5	MATRICE DE CORRELATION ET DE VARIANCE-COVARIANCE.....	119
8.5.1	Définitions.....	119
8.5.2	Matrices de corrélation et de variance-covariance	120
8.5.3	Propriétés de Γ_Z (et de C_Z).....	121
8.5.4	Interprétation probabiliste des propriétés de Γ_Z	122
8.6	APPLICATION AUX VECTEURS REELS GAUSSIENS	123
8.6.1	Définitions.....	123
8.6.2	Propriétés d'un vecteur réel gaussien	124
8.6.3	Vecteur complexe gaussien	126
8.7	EXERCICES	128
8.7.1	Énoncés.....	128
9	NOTION DE CONVERGENCE, LOI DES GRANDS NOMBRES ET THEOREME DE LA LIMITE CENTRALE	130
9.1	CONVERGENCE D'UNE SUITE DE VARIABLES ALEATOIRES.....	130

9.1.1	<i>Définitions</i>	130
9.1.2	<i>Différents modes de convergence</i>	131
9.2	EXEMPLES DE CONVERGENCE EN LOI	133
9.2.1	<i>Approximation à la loi binomiale</i>	133
9.3	GENERALISATION : THEOREME DE LA LIMITE CENTRALE (OU DE LAPLACE-LIAPOUNOFF).....	136
9.4	AUTRE EXEMPLE DE CONVERGENCE : THEOREME DE BERNOULLI, LOI DES GRANDS NOMBRES.....	137
9.4.1	<i>Théorème de Bernoulli</i>	137
9.4.2	<i>Loi des grands nombres</i>	138
9.5	EXERCICES	139
9.5.1	<i>Énoncés</i>	139
9.5.2	<i>Corrigés</i>	139

Chacun connaît, pour les utiliser dans le langage courant, les notions de probabilité et de probable que l'on oppose à celles de certitude et de certain.

D'un point de vue subjectif, la probabilité exprime le degré de certitude qu'un événement se réalise, degré que les hommes tirent des faits qu'ils connaissent.

Toutefois cette notion intuitive apparaît comme insuffisante pour conduire à une définition de la probabilité.

CHAPITRE 1. Notion d'événement et de probabilité

1.1 Présentation

Historiquement, le calcul des probabilités s'est développé à partir du XVII^e siècle autour des problèmes de jeux dans des situations où le nombre de cas possibles est fini. Les développements plus récents concernant des espaces non nécessairement finis nécessitent les outils techniques de la théorie de la mesure.

1.2 Probabilité sur un espace fini

On s'intéresse à une expérience aléatoire qui conduit à la réalisation d'un seul résultat parmi un nombre fini de résultats : $\omega_1, \omega_2, \dots, \omega_n$. On note Ω l'ensemble des résultats possibles :

$$\Omega = \{ \omega_1, \omega_2, \dots, \omega_n \}$$

On définit la fréquence d'apparition du résultat ω_k comme le rapport du nombre n_k d'expériences dont le résultat est ω_k sur le nombre total d'expériences n :

$$f_k = \frac{n_k}{n}$$

Si on mesure la fréquence d'apparition du résultat ω_k au cours d'un grand nombre de répétitions de l'expérience, on constate qu'elle fluctue de moins en moins. La limite p_k de f_k lorsque n tend vers $+\infty$ correspond à la notion intuitive de probabilité.

On appelle événement une partie A de Ω . La fréquence de A c'est-à-dire la proportion d'expériences dont le résultat est dans A est égale à : $\sum_{\omega_k \in A} f_k$.

On est donc amené à associer la probabilité $\sum_{\omega_k \in A} p_k$ à l'événement A . Comme la fréquence de

Ω est 1, on pose la définition suivante :

Définition : Une probabilité p sur un ensemble fini $\{\omega_1, \omega_2, \dots, \omega_n\}$ est une suite finie de nombres réels positifs p_1, p_2, \dots, p_n tels que :

$$\sum_{1 \leq k \leq n} p_k = 1$$

On attribue à toute partie A de Ω le nombre $\sum_{\omega_k \in A} p_k$ qui est appelé la probabilité de A .

Cette notion d'événement est à la base du calcul des probabilités. Elle permet de décrire toutes des résultats possibles en utilisant le vocabulaire d'une théorie générale, celle des ensembles, en établissant une correspondance entre :

Résultat élémentaire d'une épreuve \Leftrightarrow élément de Ω ,

Événement \Leftrightarrow sous-ensemble de Ω ,

En particulier, l'événement impossible sera l'ensemble vide \emptyset et l'événement certain sera l'ensemble Ω .

Exemple 1 : Considérons l'expérience consistant à jeter un dé et l'événement $A = \{\text{Sortie d'un As}\}$. Le nombre de résultats possibles est : $n = 6$. Le nombre de résultats favorables à A est : $n_A = 1$. La probabilité de A est alors : $P(A) = \frac{1}{6}$

Cette définition apparaît comme extrêmement simple et pratique dans tous les cas où l'on sait effectuer un dénombrement des résultats possibles (jeux, urnes).

Dès que l'expérience se complique, il convient d'apporter le plus grand soin à ce dénombrement pour éviter toute ambiguïté.

Exemple 2 : Considérons maintenant, pour illustrer cette dernière remarque, un jeu avec deux dés, un dé bleu et un dé rouge, parfaitement symétriques (non pipés). Notons A l'événement suivant est :

$A = \{\text{La somme des faces des deux dés est égale à 7}\}$

1ère solution : Avec deux dés, la somme des faces peut être égale à (2, 3, 4 11, 12) : soit 11 cas possibles. La valeur 7 est un résultat parmi ces 11 résultats possibles.

La probabilité de A est alors : $P(A) = \frac{1}{11}$

Ce résultat est faux car il y a plusieurs façons d'obtenir certains résultats, par exemple : $3 + 4 = 2 + 5 = 7$.

2ème solution : En fait, chaque résultat est une paire de chiffres. Il y a donc $6 \times 6 = 36$ cas possibles et la valeur 7 de la somme des deux dés peut être obtenue de 6 façons :

(1,6) (2,5) (3,4) (4,3) (5,2) (6,1)

Ainsi, la probabilité de A est : $P(A) = \frac{6}{36} = \frac{1}{6}$

On retiendra que le dénombrement doit s'effectuer à partir de cas **également possibles**.

1.3 Probabilité sur un espace quelconque

La pratique du calcul des probabilités montre que la définition précédente devient vite insuffisante, par exemple dès qu'il s'agit de modéliser des temps d'attente. On ne peut pas étudier avec un espace de cardinal fini une expérience aléatoire aussi simple que : « on lance un dé jusqu'à la première obtention d'un six ».

La probabilité p telle que nous allons la définir ci-dessous est une fonction qui associe un nombre compris entre 0 et 1 à un événement et censé mesurer les chances de réalisation de cet événement. Pour des raisons sortant du cadre de ce cours, il n'est pas toujours possible d'attribuer une probabilité à chaque partie de Ω . En d'autres termes, p ne peut pas être considérée comme une application de l'ensemble $P(\Omega)$ de toutes les parties de Ω dans $[0,1]$ mais comme une fonction ayant un domaine de définition F généralement plus petit que $P(\Omega)$. Voici les propriétés qu'il est raisonnable d'exiger de F :

$$\Omega \in F,$$

$$A \in F \Rightarrow \bar{A} \in F,$$

$$(A_n)_{n \geq 1} \in F \Rightarrow \sum_{n \geq 1} A_n \in F,$$

On dit que F est stable par passage au complémentaire et par union dénombrable. En passant au complémentaire, on montre que :

$$\emptyset \in F,$$

$$(A_n)_{n \geq 1} \in F \Rightarrow \prod_{n \geq 1} A_n \in F,$$

On parvient ainsi à une définition abstraite mais féconde de l'expérience faisant appel à la théorie de la mesure et généralement notée (Ω, F, p) où Ω est l'espace des résultats, F un

sous-espace de Ω (corps Borélien) obtenu par réunion dénombrable ou intersection dénombrable des sous-ensembles de Ω qui sont des événements (sous-ensembles auxquels on peut attribuer une probabilité) et p une mesure sur F satisfaisant les trois axiomes de la définition suivante :

Définition : On dit que p est une probabilité si p est une application de F dans l'ensemble des réels qui à chaque événement A de F attribue un nombre $p(A)$ que l'on appelle probabilité de l'événement (ou mesure sur A) et qui satisfait aux 3 axiomes suivants :

Axiome 1 : $p(A) \geq 0$,

Axiome 2 : $p(\Omega) = 1$,

Axiome 3 : $A \in F, B \in F, A \cdot B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

Cette propriété devant s'étendre à une suite infinie d'événements disjoints :

$$A_i \in F, i \neq j \quad A_i \cdot A_j = \emptyset \Rightarrow p\left(\sum_{i \geq 1} A_i\right) = \sum_{i \geq 1} p(A_i)$$

Pour notre part, nous nous contenterons dans ce cours d'utiliser la correspondance entre ensemble et événements après avoir rappelé les définitions les plus classiques relatives à l'algèbre des ensembles.

1.4 Rappel sur l'algèbre des ensembles

On utilisera les notations suivantes :

La réunion : $(+)$

L'intersection : $(.)$

L'ensemble total : (Ω)

L'ensemble vide : (\emptyset)

Le complémentaire de A : (\bar{A})

Lois commutatives : $A + B = B + A$ et $A \cdot B = B \cdot A$

Lois associatives : $(A + B) + C = A + (B + C)$ et $(A \cdot B) \cdot C = A \cdot (B \cdot C)$

Lois distributives : $A \cdot (B + C) = A \cdot B + A \cdot C$ et $A + (B \cdot C) = (A + B) \cdot (A + C)$

Lois complémentaires : $A + \bar{A} = \Omega, A \cdot \bar{A} = \emptyset, A + \Omega = \Omega,$

: $A \cdot \Omega = A, A + \emptyset = A, A \cdot \emptyset = \emptyset$

Lois d'involution : $\overline{(\bar{A})} = A$

Lois idempotentes : $A + A = A$ et $A \cdot A = A$

Lois de Morgan : $\overline{(A + B)} = \bar{A} \cdot \bar{B} \quad \overline{(A \cdot B)} = \bar{A} + \bar{B}$

1.5 Applications : quatre corollaires des axiomes

1.5.1 Corollaire 1 : probabilité d'un événement impossible

La probabilité d'un événement impossible (ensemble vide) est nulle : $p(\emptyset) = 0$

Démonstration :

Pour démontrer ce résultat, il suffit de calculer de deux manières la probabilité de la réunion $A + \emptyset = A$.

D'une part, on a : $p(A + \emptyset) = p(A)$. D'autre part, les événements A et \emptyset étant disjoints ($A \cdot \emptyset = \emptyset$), on trouve, d'après l'axiome 3 : $p(A + \emptyset) = p(A) + p(\emptyset)$. En identifiant les deux résultats, on obtient le résultat recherché.

1.5.2 Corollaire 2 : probabilité de l'événement complémentaire

La probabilité d'un événement est égale au complément à 1 de la probabilité de son inverse (complémentaire) : $p(A) = 1 - p(\bar{A}) \leq 1$

Démonstration :

On part de la relation $A + \bar{A} = \Omega$. D'une part, on a, d'après l'axiome 2 : $p(A + \bar{A}) = p(\Omega) = 1$. D'autre part, on a, d'après l'axiome 3 : $p(A + \bar{A}) = p(A) + p(\bar{A})$ car $A \cdot \bar{A} = \emptyset$. En identifiant ces deux relations, on trouve $p(A) = 1 - p(\bar{A})$.

1.5.3 Corollaire 3 : théorème des probabilités totales

$$p(A + B) = p(A) + p(B) - p(A \cdot B)$$

Démonstration :

D'une part, écrivons $A+B$ comme réunion d'ensembles disjoints :

$$A + B = A + B \cdot \Omega = A + B \cdot (A + \bar{A}) = A + A \cdot B + \bar{A} \cdot B = A \cdot (\Omega + B) + \bar{A} \cdot B = A + \bar{A} \cdot B$$

Or, comme $A \cdot (\bar{A} \cdot B) = \emptyset$, on peut appliquer l'axiome 3 :

$$p(A + B) = p(A) + p(\bar{A} \cdot B)$$

D'autre part, B peut s'écrire sous la forme : $B = B \cdot (A + \bar{A}) = A \cdot B + \bar{A} \cdot B$.

Comme $(A \cdot B) \cdot (\bar{A} \cdot B) = \emptyset$, en appliquant l'axiome 3, on trouve : $p(B) = p(A \cdot B) + p(\bar{A} \cdot B)$

En éliminant la probabilité $p(\bar{A} \cdot B)$ entre les deux relations précédentes, on trouve l'expression cherchée.

1.5.4 Corollaire 4 : théorème de l'inclusion

$$A \subset B \Rightarrow p(A) \leq p(B)$$

Démonstration :

On a : $A + B = B = A + \bar{A} \cdot B$.

Comme $A \cdot (\bar{A} \cdot B) = \emptyset$, en appliquant l'axiome 3 : $p(B) = p(A) + p(\bar{A} \cdot B)$. D'après l'axiome 1, on trouve le résultat du théorème

1.6 Exercices

1.6.1 Énoncés

Permutations - Combinaisons

1. Dans combien d'ordres différents peut-on placer 4 pneus interchangeables sur 4 roues d'une voiture ? Et 5 pneus sur 4 roues ?
2. Trouver le nombre de « mélodies » (suite de notes différentes) de 3 notes qui peuvent être conçues à partir de 8 notes différentes si aucune note n'est répétée ?
3. On prend 5 cartes au hasard pour faire une main de poker dans un jeu de 32 cartes. Combien de mains différentes sont-elles possibles ?
4. On désire transmettre de l'information codée en disposant 6 lampes électriques de couleur en ligne horizontale. Combien de messages différents sont-ils possibles en changeant les positions de 3 lumières rouges, 2 vertes, une jaune ? (elles sont toutes allumées simultanément).
5. De combien de façons peut-on former un comité de 5 personnes à choisir parmi 7 hommes et 5 femmes, si ce comité doit être constitué de 3 hommes et de 2 femmes ? Et d'au moins 2 hommes et au moins une femme ?

Ensembles

6. L'ensemble fondamental Ω comprend tous les nombres entiers positifs. Les ensembles A, B, C sont définis par :

$$A = \{1, 2, 3, 6, 7, 10\} \quad B = \{3, 4, 8, 10\} \quad C = \{n \text{ entier} \mid n > 5\}$$

Définir les ensembles suivants: $A + B$, $A.B$, $A.\bar{B}$, $A.C$, $B.C$, \bar{C} , $A + B + C$.

7. L'ensemble fondamental Ω comprend tous les nombres réels x. De plus, on a :

$$A = \{2 \leq x \leq 5\} \quad B = \{3 \leq x \leq 6\}.$$

Trouver les ensembles : $A + B$, $A.B$, \bar{A} , \bar{B} , $A.\bar{B}$.

8. Un ensemble fini a n éléments. Combien peut-on former de sous-ensembles avec ces éléments, y compris l'ensemble lui-même et l'ensemble vide ?

Examiner les cas $n = 1, 2, \dots$

Probabilités

9. Parmi les fonctions suivantes, quelles sont celles qui définissent une probabilité sur l'ensemble $\Omega = \{a_1, a_2, a_3\}$?

$$P(a_1) = \frac{1}{2} \quad P(a_2) = \frac{1}{3} \quad P(a_3) = \frac{1}{4},$$

$$P(a_1) = \frac{1}{2} \quad P(a_2) = -\frac{1}{2} \quad P(a_3) = 1,$$

$$P(a_1) = \frac{1}{5} \quad P(a_2) = \frac{2}{5} \quad P(a_3) = \frac{2}{5},$$

$$P(a_1) = \frac{1}{3} \quad P(a_2) = 0 \quad P(a_3) = \frac{2}{3}.$$

10. Trois étudiants participent à une course. Les étudiants A et B ont la même probabilité de gagner et chacun d'eux a deux fois plus de chance que l'étudiant C de gagner. Calculer la probabilité pour que B ou C gagne.

11. On tire un nombre entier au hasard compris entre 1 et 50. Calculer la probabilité pour que le nombre obtenu soit divisible par 5, soit un nombre premier, se termine par un 2.

12. Une classe est formée par 10 garçons dont la moitié a les yeux marrons et par 20 filles dont la moitié a également les yeux marrons. Calculer la probabilité pour qu'une personne tirée au hasard soit un garçon ou ait les yeux marrons.

13. On jette 50 fois un dé. Le tableau suivant donne les 6 résultats possibles, ainsi que leur fréquence d'apparition.

Nombre	1	2	3	4	5	6
Fréquence	7	9	8	7	9	10

Calculer la fréquence relative de l'événement correspondant à l'apparition d'un 4, d'un nombre impair, d'un nombre premier.

1.6.2 Solutions

Permutations - Combinaisons

1. Avec 4 pneus sur 4 roues : 24 ; Avec 5 pneus sur 4 roues : 120
2. Nombre de mélodies : 336
3. Nombre de mains différentes : 201.376
4. Nombre de messages : 60
5. Nombre de comités de 3 hommes et 2 femmes : 350. Au moins 2 hommes et au moins une femme : 735

Ensembles.

6. $A + B = (1, 2, 3, 4, 6, 7, 8, 10)$ $A \cdot B = (3, 10)$ $A \cdot \bar{B} = (1, 2, 6, 7)$ $A \cdot C = (6, 7, 10)$
 $B \cdot C = (8, 10)$ $\bar{C} = (1, 2, 3, 4, 5)$ $A + B + C = \Omega - \{5\}$
7. $A + B = (2 \leq x < 6)$ $A \cdot B = (3 < x \leq 5)$ $\bar{A} = \{(x < 2) \cup (5 < x)\}$ $\bar{B} = \{(x < 3) \cup (6 < x)\}$
 $A \cdot \bar{B} = (2 \leq x \leq 3)$
8. Il y a 2^n sous-ensembles.

Probabilités.

9. Non : on a $\sum_i P(a_i) \neq 1$; Non : on a $P(a_2) < 0$; oui ; oui.
10. $P(B \text{ ou } C) = 3/5$
11. divisible par 5 : 1/5 ; nombre premier : 3/10 ; se termine par un 2 : 1/10
12. 2/3
13. $\text{Fr}(4) = 7/50$; $\text{Fr}(\text{impair}) = 24/50$; $\text{Fr}(\text{premier}) = 26/50$

CHAPITRE 2. Probabilités conditionnelles & Indépendance

2.1 Probabilités conditionnelles

2.1.1 Définition

On considère deux événements A et M et l'on suppose que l'événement M s'est produit.

Le fait que l'événement M se soit produit constitue une information a priori. S'il existe une liaison entre A et M , cette information va modifier la probabilité de l'événement A .

Cette nouvelle valeur de la probabilité de A s'appelle « La probabilité de A conditionnelle à M » et l'on pose par définition :

$P(A/M) = \frac{P(A.M)}{P(M)} \quad \text{avec } P(M) \neq 0$	2-1
---	-----

On vérifie en effet que la probabilité conditionnelle satisfait aux trois axiomes de base (exercice) :

- La probabilité conditionnelle est un nombre positif,
- La probabilité de l'événement conditionnel certain est égale à 1,
- La probabilité conditionnelle de la réunion de deux événements conditionnellement disjoints ($A.B.M = \emptyset$) est égale à la somme de leurs probabilités conditionnelles respectives.

2.1.2 Interprétation par fréquence relative

Notre première définition de la probabilité va nous permettre de préciser le sens de la probabilité conditionnelle. On répète n fois une expérience E sur laquelle on a défini deux événements A et M et l'on désigne par :

- $n(A)$ le nombre de fois où l'événement A se réalise,
- $n(M)$ le nombre de fois où l'événement M se réalise,
- $n(A.M)$ le nombre de fois où A et M se sont réalisés simultanément.

D'après la définition par fréquence relative :

$$P(A) = \frac{n(A)}{n}; \quad P(M) = \frac{n(M)}{n}; \quad P(A.M) = \frac{n(A.M)}{n}$$

D'après la définition de la probabilité conditionnelle :

$$P(A/M) = \frac{\frac{n(A.M)}{n}}{\frac{n(M)}{n}} = \frac{n(A.M)}{n(M)}$$

La probabilité de A n'est plus définie par rapport au nombre total d'épreuves n mais par rapport au nombre d'épreuves $n(M)$ où l'événement M s'est produit.

2.1.3 Cas particuliers

- A est inclus dans M : $A \subset M$

$$\text{Donc : } A.M = A \text{ et } P(A/M) = \frac{P(A)}{P(M)} \geq P(A)$$

Si M s'est réalisé, A a plus de chance de se réaliser.

- M est inclus dans A : $M \subset A$

$$\text{donc : } A.M = M \text{ et } P(A/M) = 1 \quad \text{si } M \text{ s'est réalisé, } A \text{ est certain.}$$

2.2 Événements indépendants

L'indépendance est une notion fondamentale du calcul des probabilités. Elle caractérise l'absence de liaison entre deux événements.

2.2.1 Indépendance de deux événements

Deux événements sont dits indépendants si la probabilité de leur intersection (probabilité qu'ils se réalisent simultanément) est égale au produit de leurs probabilités.

$P(A.B) = P(A).P(B)$	2-2
----------------------	-----

Attention à ne pas confondre événements disjoints et indépendants. En effet si deux événements sont disjoints, cela signifie qu'ils ne peuvent pas se produire simultanément et la probabilité de leur intersection est nulle.

A l'aide des axiomes, on en déduit (Exercice) :

$$P(A.\bar{B}) = P(A).P(\bar{B}) \quad (A \text{ et } \bar{B} \text{ indépendants})$$

$$P(\bar{A}.B) = P(\bar{A}).P(B) \quad (\bar{A} \text{ et } B \text{ indépendants})$$

$$P(\bar{A}.\bar{B}) = P(\bar{A}).P(\bar{B}) \quad (\bar{A} \text{ et } \bar{B} \text{ indépendants})$$

D'après la définition de la probabilité conditionnelle, si A et B sont indépendants on trouve :

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

La réalisation de A (respectivement B) ne nous apprend rien sur celle possible de B (respectivement A).

2.2.2 Interprétation par fréquence relative.

Reprenons l'exemple du paragraphe (2.1.3). Nous avons par définition :

$$P(A) = \frac{n(A)}{n} \quad P(A/M) = \frac{n(A.M)}{n(M)}$$

$$\text{Si } A \text{ et } M \text{ sont indépendants : } \frac{n(A)}{n} = \frac{n(A.M)}{n(M)}$$

Ce qui signifie que si les événements A et M sont indépendants, la fréquence relative de l'événement A est la même si l'on considère le nombre total d'épreuves ou simplement celles pour lesquelles l'événement M s'est produit.

Exemple :

On prend un jeu de 52 cartes. Les deux événements A et B suivants sont-ils indépendants ?

$$A = \{\text{on tire une dame}\} \quad B = \{\text{on tire un cœur}\}$$

En appliquant la définition par dénombrement :

$$P(A) = \frac{4}{52} \quad P(B) = \frac{13}{52} = \frac{1}{4} \quad \Rightarrow \quad P(A).P(B) = \frac{1}{52}$$

D'autre part : $A.B = \{\text{on tire une dame de cœur}\}$, soit :

$$P(A.B) = \frac{1}{52}$$

Donc, les deux événements considérés sont indépendants.

2.3 Théorème des probabilités totales & Théorème de BAYES

On considère N événements (M_1, M_2, \dots, M_N) complets mutuellement exclusifs (disjoints), c'est-à-dire tels que :

$$\sum_{i=1}^N M_i = \Omega$$

$$i \neq j \Rightarrow M_i \cdot M_j = \emptyset$$

Le théorème des probabilités totales nous dit que la probabilité d'un événement A quelconque est donnée par :

$$P(A) = \sum_{i=1}^N P(A | M_i) \cdot P(M_i) \quad 2-3$$

En appliquant la définition des probabilités conditionnelles et le théorème des probabilités totales, on montre le théorème de BAYES :

$$P(M_i | A) = \frac{P(A | M_i) \cdot P(M_i)}{\sum_{i=1}^N P(A | M_i) \cdot P(M_i)} \quad 2-4$$

Dans ces équations, on note :

- $P(M_i)$: probabilité a priori de M_i ,
- $P(A | M_i)$: probabilité de A conditionnelle à M_i
- $P(M_i | A)$: probabilité a posteriori de M_i (probabilité conditionnelle à A).

Démonstration du théorème des probabilités totales :

Compte tenu des hypothèses sur les M_i , on peut écrire :

$$P(A) = P(A \cdot \Omega) = P(A \cdot \sum_{i=1}^N M_i)$$

Or, pour $i \neq j$, les M_i sont disjoints et donc :

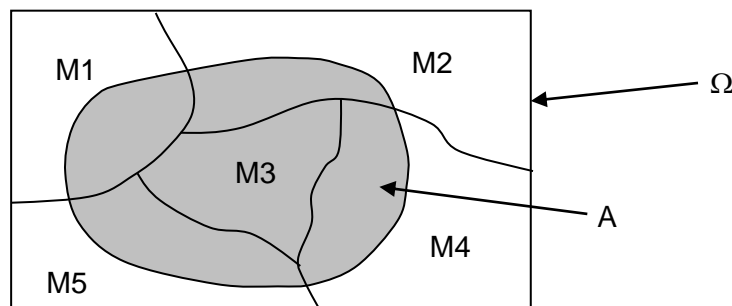
$$i \neq j \Rightarrow (A \cdot M_i) \cdot (A \cdot M_j) = \emptyset$$

En appliquant l'axiome 3, on obtient la relation cherchée :

$$P(A) = \sum_{i=1}^N P(A \cdot M_i) = \sum_{i=1}^N P(A | M_i) \cdot P(M_i)$$

On montrera sans difficulté (exercice) que ce théorème est encore valable si :

$$A \subset \sum_{i=1}^N M_i \subset \Omega.$$



Interprétation du théorème de BAYES :

Le théorème des probabilités totales répond à la question : étant donné un événement A pouvant résulter de plusieurs causes M_1, M_2, \dots, M_N , s'excluant mutuellement, quelle est la probabilité de A connaissant les probabilités élémentaires $p(M_i)$ et les probabilités conditionnelles de A à chaque cause M_i , soit $p(A|M_i)$?

Le théorème de BAYES répond à la question inverse : supposons que l'événement A se soit produit et que l'on sache qu'il pouvait résulter de plusieurs causes M_1, M_2, \dots, M_N , s'excluant mutuellement. Quelle est dans ces conditions la probabilité que ce soit précisément la cause M_i qui l'ait produit ?

Application

Considérons l'expérience qui consiste à tirer une pièce dans un lot de 4 500 pièces dont 800 sont fausses et 3 700 bonnes.

Ces pièces sont réparties dans 4 boîtes identiques B1, B2, B3, B4 de la façon suivante :

Boite	Total	Bonnes	Fausse
B1	2 000	1 600	400
B2	500	300	200
B3	1 000	900	100
B4	1 000	900	100

Soit F l'événement auquel on s'intéresse : $F = \{\text{tirer une pièce fausse}\}$.

Pour tirer une pièce, il faut d'abord faire le choix d'une boîte. Nous allons nous intéresser à différentes probabilités :

Probabilité de choisir une boîte donnée (ou de tirer la pièce dans cette boîte).

Soit l'événement $B_i = \{\text{on tire la pièce dans la boîte } B_i\}$ défini pour i dans $\{1,2,3,4\}$. Toutes les pièces étant contenues dans la réunion des 4 boîtes, on a :

$$\sum_{i=1}^4 B_i = \Omega$$

De plus, la pièce ne peut provenir que d'une boîte, ce qui s'exprime par :

$$B_i \cdot B_j = \emptyset, \quad i \neq j$$

Si l'on tire une pièce dans une boîte, on ne tire pas la pièce dans une autre (événements disjoints). Donc, d'après les axiomes 2 et 3 :

$$P\left(\sum_{i=1}^4 B_i\right) = P(\Omega) = \sum_{i=1}^4 P(B_i) = 1$$

Supposons que le choix d'une boîte soit équiprobable ce qui s'exprime par :

$$P(B_1) = P(B_2) = P(B_3) = P(B_4) = \frac{1}{4}$$

Probabilité conditionnelle

On pose, par définition (1.2.2), que la probabilité de tirer une pièce fausse dans une boîte donnée est égale à la répartition des pièces qu'elle contient. Soit :

$$P(F/B_1) = \frac{400}{2000} = 0,2 \quad P(F/B_2) = \frac{200}{500} = 0,4$$

$$P(F/B_3) = \frac{100}{1000} = 0,1 \quad P(F/B_4) = \frac{100}{1000} = 0,1$$

Probabilité de tirer une pièce fausse

Le théorème des probabilités totales nous permet de calculer la probabilité de tirer une pièce fausse :

$$P(F) = \sum_{i=1}^4 P(F | B_i) \cdot P(B_i) = \frac{1}{4} \cdot (0,2 + 0,4 + 0,1 + 0,1) = 0,2$$

Application du théorème de Bayes:

Sachant que la pièce tirée est défectueuse, quelle est la probabilité pour qu'elle provienne, par exemple, de la boîte B_2 ?

Nous avons trouvé :

$$P(F | B_2) = 0,4 \quad P(B_2) = 0,25 \quad P(F) = \sum_{i=1}^4 P(F | B_i) \cdot P(B_i) = 0,2$$

D'où la probabilité a posteriori :

$$P(B_2 | F) = \frac{0,4 \times 0,25}{0,2} = 0,5$$

Ainsi, la probabilité a posteriori $P(B_2 | F)$ d'avoir tiré la pièce dans la boîte B_2 (sachant que la pièce tirée était défectueuse) est deux fois plus grande que la probabilité a priori $P(B_2)$ de tirer la pièce dans la boîte B_2 .

2.4 Exercices

2.4.1 Énoncés

1. Une urne contient N boules qui diffèrent seulement par leur couleur. Il y a R boules rouges. Deux boules sont tirées au hasard dans l'urne, une à la fois et sans les remplacer. Démontrer que la probabilité totale que la deuxième boule soit rouge est R/N .

2. Une urne contient 3 boules blanches et 2 boules rouges, soit : (b_1, b_2, b_3, r_1, r_2) On tire au hasard 2 boules successivement. Quelle est la probabilité pour que la première boule soit blanche et que la deuxième boule soit rouge ?

3. Un appel téléphonique se produit dans l'intervalle $[0, T]$ au hasard. Soit t l'instant aléatoire où il se produit.

Trouver la probabilité $P(\{t_0 \leq t \leq t_1 \mid t \geq t_0\})$, c'est-à-dire la probabilité que l'appel se produise dans l'intervalle $[t_0, t_1]$, étant donné qu'il ne s'est pas produit avant t_0 .

4. Un train et Monsieur X arrivent à la même gare indépendamment et au hasard entre 9 et 10 heures. Le train s'arrête pendant 10 minutes. Monsieur X désire prendre le train mais il attend seulement t_0 minutes.

Déterminer la durée t_0 telle que la probabilité pour que Monsieur X prenne le train soit égale à 0,5.

5. Si A et B sont indépendants, démontrer que A et le \bar{B} sont aussi,

6. Les événements A et B ont tous les deux une probabilité non nulle. Démontrer que les propriétés ci-dessous sont vraies ou fausses :

- Si A et B sont disjoints, ils sont indépendants,
- Si A et B sont indépendants, ils sont disjoints.

7. Un événement A a la probabilité p de se produire (et q de ne pas se produire) indépendamment du fait qu'il s'est produit ou non à des épreuves précédentes.

Avec n épreuves, déterminer les probabilités suivantes (mettre sous la forme la plus simple) :

- l'événement A ne se produit pas à la n ème épreuve,
- l'événement A ne se produit pas du tout sur l'ensemble des n épreuves,

- l'événement A se produit seulement à la n ème épreuve,
- l'événement A se produit exactement 1 fois,
- l'événement A se produit au moins 1 fois,
- l'événement A se produit exactement 2 fois,
- l'événement A se produit au moins 2 fois,
- l'événement A se produit au plus 2 fois,
- l'événement A se produit un nombre pair de fois (utiliser les développements de $(p+q)^n$ et $(p-q)^n$ en distinguant les cas où n est pair ou impair).

8. On dispose de trois cartes. La première a deux faces rouges. La seconde a deux faces blanches. La troisième a une face rouge et l'autre blanche.

On tire une carte et on constate que la face visible est rouge. Quelle est la probabilité pour que l'autre face soit rouge ?

9. On considère 3 urnes (U_1, U_2, U_3) fermées à clef. L'urne U_1 contient 6 boules noires et 3 boules blanches. L'urne U_2 contient 6 boules noires et 9 boules blanches. L'urne U_3 contient 3 boules noires et 3 boules blanches.

Les clefs des urnes sont enfermées dans une boîte qui contient une clef pour U_1 , 2 clefs pour U_2 et 1 clef pour U_3 .

On tire une clef au hasard dans la boîte, on ouvre l'urne correspondante et on extrait une boule.

- a) Quelle est la probabilité pour que cette boule soit noire ?
- b) On constate que la boule est noire. Quelle est la probabilité pour qu'elle ait été tirée de l'urne U_1 ?

2.4.2 Solutions

$$1. \quad P(Br) = \frac{R-1}{N-1} \cdot \frac{R}{N} + \frac{R}{N-1} \cdot \frac{N-R}{N} = \frac{R}{N}$$

$$2. \quad P(B1 \text{ blanche et } B2 \text{ rouge}) = 0,3$$

$$3. \quad P(A|M) = \frac{t_1 - t_0}{T - t_0}$$

$$4. \quad t_0 = 26,8 \text{ minutes}$$

$$5. \quad P(A) = P(A.B + A.\bar{B}) = P(A).P(B) + P(A.\bar{B}) \text{ et } P(\bar{B}) = 1 - P(B),$$

$$6. \quad \text{Si } A.B = \emptyset \text{ ils ne sont pas indépendants} \quad \text{Si } A \text{ et } B \text{ sont indépendants } A.B \neq \emptyset$$

$$7. \quad q = 1-p; \quad q^n; \quad pq^{n-1}; \quad npq^{n-1}; \quad 1-q^n; \quad Pn(2) \text{ avec } Pn(k) = C_n^k p^k q^{n-k}; \quad 1-q^n - npq^{n-1}; \\ Pn(0) + Pn(1) + Pn(2); \quad \frac{1}{2}(1 + (-1)^n (p-q)^n)$$

$$8. \quad P(2^{\text{ème}} \text{ face rouge}) = 2/3$$

$$9. \quad a) P(\text{boule noire}) = 59/120 \quad b) P(U_1 | R) = 20/59.$$

CHAPITRE 3. Variable aléatoire réelle

3.1 Notion de variable aléatoire réelle

3.1.1 Introduction

Jusqu'à présent, nous avons considéré une expérience E qui définissait un espace Ω formé par tous les résultats possibles.

En général, ces résultats étaient des objets identifiés dans l'expérience comme la sortie de « pile ou face », après le jet d'une pièce ou le tirage d'une boule de couleur donnée dans une urne. Parfois même, ces résultats pouvaient être numériques comme dans un jeu avec des dés.

De toute façon, couleur, face ou chiffre n'étaient que des intermédiaires servant à identifier les résultats. Nous nous proposons maintenant d'identifier les résultats ω_i d'une expérience en leur associant des nombres x_i : $x_i = X(\omega_i)$

Exemple 1 : Considérons l'expérience qui consiste à jeter un dé de poker (pour éviter toute ambiguïté). Il y a 6 résultats possibles. A chaque résultat, nous allons associer un nombre.

Résultats : ω_i	Variable réelle : $x_i = X(\omega_i)$
As	1
Roi	2
Dame	3
Valet	4
Dix	5
Neuf	6

Par exemple, le résultat $\omega_i = \{\text{sortie d'une dame}\}$ est équivalent à la donnée $x = 3$.

Exemple 2 : On jette une pièce de monnaie trois fois de suite. Il y a huit résultats possibles correspondant aux différentes combinaisons de pile ou face.

Supposons que l'événement auquel on s'intéresse représente le nombre de fois où pile est sortie. Nous allons choisir la correspondance $x = X(\omega)$ de telle sorte que la variable x , qui est fonction des résultats représente effectivement l'événement choisi

Résultats : ω	Variable réelle $x = X(\omega)$
$\omega_1 = \{ P P P \}$	3
$\omega_2 = \{ P P F \}$	2
$\omega_3 = \{ P F P \}$	2
$\omega_4 = \{ P F F \}$	1
$\omega_5 = \{ F P P \}$	2
$\omega_6 = \{ F P F \}$	1
$\omega_7 = \{ F F P \}$	1
$\omega_8 = \{ F F F \}$	0

Plusieurs résultats ont ainsi la même image par $X(\cdot)$; par exemple ω_2 , ω_3 et ω_5 car $X(\omega_2) = X(\omega_3) = X(\omega_5)$ (cf. tableau précédent).

3.1.2 Définition

La variable aléatoire X apparaît ainsi comme une application, définie sur Ω (espace des résultats) et à valeur dans \mathbb{R} (droite réelle).

On supposera en particulier que tous les résultats ω_i ont une image x_i par $X(\cdot)$, deux résultats pouvant avoir une même image.

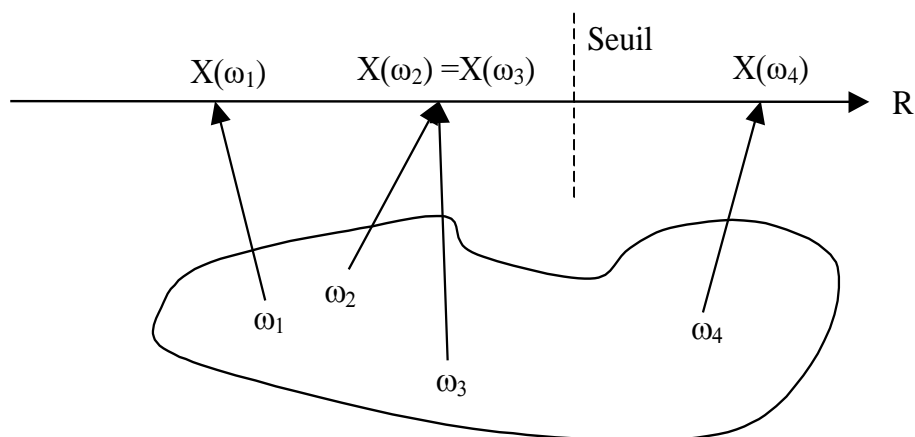


Figure 3.1 :

On voudrait maintenant associer une probabilité à la variable aléatoire X .

Dans Ω , un événement est représenté par une collection arbitraire de résultats ω_i formant un sous-ensemble auquel on attribue une probabilité.

Comme $X(\cdot)$ est une application de Ω dans \mathbb{R} , il suffit de faire correspondre les sous-ensembles de \mathbb{R} aux sous-ensembles de Ω pour en connaître la probabilité.

Par exemple, si A est un sous-ensemble de Ω ($A \subset \Omega$) avec la probabilité $P(A)$, $X(A)$ l'ensemble des x qui sont l'image de A par $X(\cdot)$ dans \mathbb{R} aura pour probabilité $P(A)$. Ceci revient à écrire :

$$P(X(A)) = P(\{\omega \text{ tels que } X(\omega) \in X(A)\})$$

Sur la droite réelle, les sous-ensembles sont des unions d'intervalles. Pour décrire ces sous-ensembles, on utilise des sous-ensembles de base de la forme $\{X < x\}$. Explicitons cette condition : $\{X < x\}$ signifie l'ensemble de tous les résultats ω tels que leur image par $X(\cdot)$ soit inférieure à la valeur x . En d'autres termes :

$$\{X < x\} = \{\omega | X(\omega) < x\}.$$

Sur la figure 3.1 précédente où l'on a placé un seuil x arbitraire, on trouve que la condition $\{X < x\}$ correspond au sous-ensemble $\{\omega_1, \omega_2, \omega_3\}$ de Ω .

Ceci nous conduit à donner une définition très générale d'une variable aléatoire :

Définition : Une variable aléatoire est une application de Ω , l'espace des résultats, dans \mathbb{R} , la droite réelle, telle que :

- l'ensemble des points ω défini par la condition $\{X(\omega) < x\}$, notée $\{X < x\}$ soit un événement quel que soit x ,
- la probabilité d'avoir $\{X = \pm\infty\}$ est nulle : $P(\{X = +\infty\}) = P(\{X = -\infty\}) = 0$.

Pratiquement, cela signifie que X est une variable aléatoire si l'on connaît la probabilité d'avoir $\{X < x\}$ quelle que soit la valeur de x réel.

Remarque 1 : on peut de la même manière définir une variable aléatoire complexe $Z(\omega)$ en considérant deux applications $(X(\omega), Y(\omega))$ de Ω dans \mathbb{R} et en posant :

$$Z(\omega) = X(\omega) + j Y(\omega)$$

Remarque 2 : Probabilité d'un intervalle.

Soit à déterminer la probabilité de l'événement C défini par : $C = \{x_1 \leq X < x_2\}$. X étant une variable aléatoire, on peut définir les événements A et B par :

$$A = \{X < x_1\} \quad B = \{X < x_2\}$$

On peut écrire que :

$$B = A + C = \{X < x_1\} + \{x_1 \leq X < x_2\}$$

Ces ensembles étant disjoints, l'axiome 3 (additivité) s'applique et on en déduit :

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

$$\text{D'où : } P(\{x_1 \leq X < x_2\}) = P(B) - P(A) = P(\{X < x_2\}) - P(\{X < x_1\})$$

3.2 Fonction de répartition

3.2.1 Définition

On appelle fonction de répartition d'une variable aléatoire réelle la probabilité de l'événement $\{X < x\}$ considérée comme une fonction du réel x . Cette fonction est généralement notée $F_X(x)$ et vaut par définition :

$$F_X(x) = p\{X < x\}$$

Dans cette équation, X désigne la variable aléatoire et x le seuil.

Exemple 1 :

Reprenons l'exemple du dé de poker. Ce dé étant supposé parfaitement symétrique, on attribue à chacune des figures la probabilité $1/6$.

Les différentes valeurs de la variable aléatoire X associée à chacune des différentes figures sont données dans le tableau suivant :

Résultats	Probabilité	Variable réelle
ω	$P(X = x)$	$x = X(\omega)$
As	1/6	1
Roi	1/6	2
Dame	1/6	3
Valet	1/6	4
Dix	1/6	5
Neuf	1/6	6

Pour déterminer la fonction de répartition, il faut définir les résultats ω_i correspondant à la condition $\{X(\omega_i) < x\}$ pour différentes valeurs de x . Les différentes valeurs de cette fonction sont données dans le tableau suivant.

Résultats ω_i correspondant à la condition $\{X(\omega_i) < x\}$		
Valeur du seuil x	résultats	Valeur du $F_X(x)$
$x \leq 1$	\emptyset , pas de résultats	$F_X(x) = 0$
$1 < x \leq 2$	As	$F_X(x) = 1/6$
$2 < x \leq 3$	As, Roi	$F_X(x) = 2/6$
$3 < x \leq 4$	As, Roi, Dame	$F_X(x) = 3/6$
$4 < x \leq 5$	As, Roi, Dame, Valet	$F_X(x) = 4/6$
$5 < x \leq 6$	As, Roi, Dame, Valet, Dix	$F_X(x) = 5/6$
$x > 6$	As, Roi, Dame, Valet, Dix, Neuf	$F_X(x) = 1$

Dans ce cas, on peut représenter $F_X(x)$ analytiquement sous la forme :

$$F_X(x) = \frac{1}{6} \sum_{i=1}^6 H(x-i)$$

Dans cette équation, H est l'échelon Heaviside défini par:

$$H(x) = 0 \text{ pour } x \leq 0$$

$$H(x) = 1 \text{ pour } x > 0$$

On voit que la fonction de répartition a l'allure d'une fonction en escalier (cf. figure 3.2).

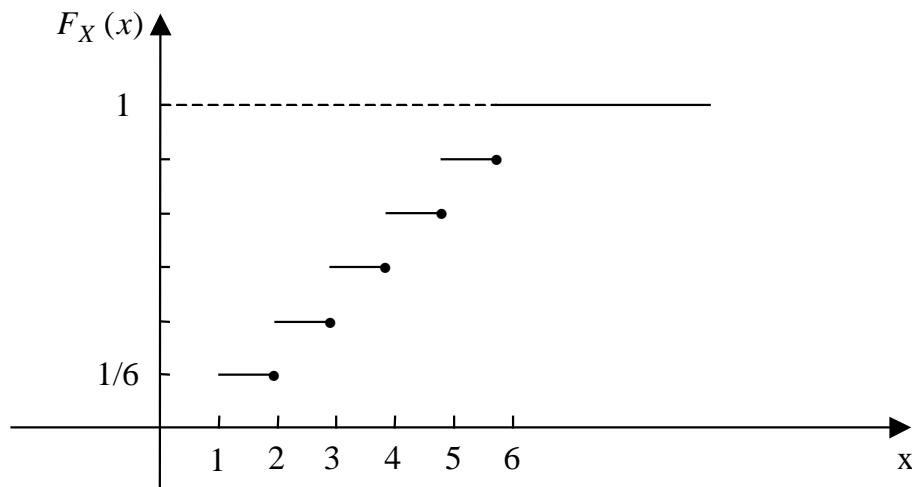


Figure 3.2

Exemple 2 :

Considérons maintenant l'exemple de la pièce jetée trois fois. On suppose cette pièce symétrique et l'on attribue une égale probabilité aux événements pile et face. Donc (conséquence de l'axiome 2) :

$$P(F) = P(P) = \frac{1}{2}$$

Les épreuves étant indépendantes, chaque résultat de l'événement $A = \{\text{pièce jetée trois fois}\}$, a la probabilité : $P(\{\text{pièce jetée trois fois}\}) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$

Déterminons la fonction de répartition de la variable X , égale au nombre de fois où pile est sortie.

Les valeurs sont données dans le tableau suivant.

Résultats ω_i correspondant à la condition $\{X(\omega_i) < x\}$		
Valeur du seuil x	Résultats	Valeur du $F_X(x)$
$x \leq 0$	\emptyset , pas de résultats	$F_X(x) = 0$
$0 < x \leq 1$	1 résultat	$F_X(x) = 1/8$
$1 < x \leq 2$	4 résultats (1+3)	$F_X(x) = 4/8$
$2 < x \leq 3$	7 résultats (1+3+3)	$F_X(x) = 7/8$
$x > 3$	8 résultats	$F_X(x) = 1$

$$F_X(x) = \frac{1}{8}H(x) + \frac{3}{8}H(x-1) + \frac{3}{8}H(x-2) + \frac{1}{8}H(x-3)$$

Cette fonction de répartition est présentée sur la figure 3.3.

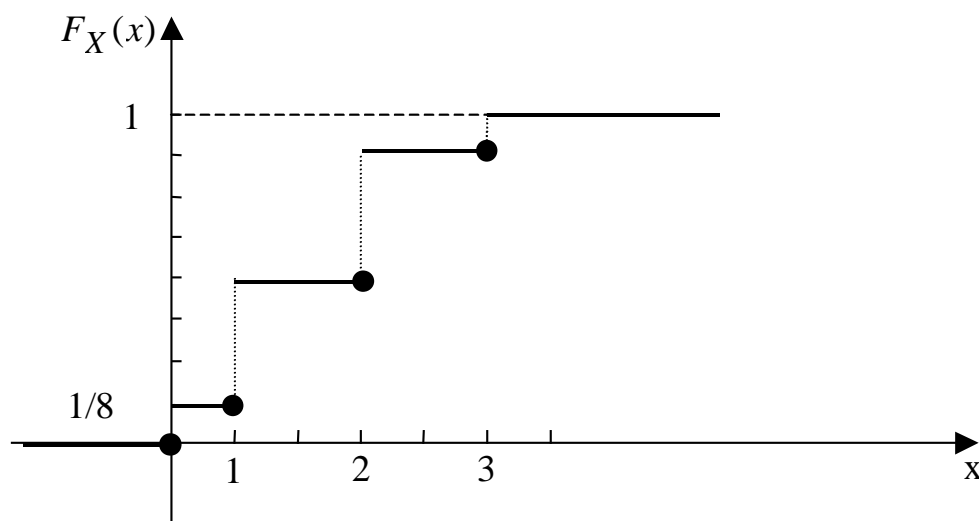


Figure 3.3.

3.2.2 Propriétés

La fonction de répartition d'une variable aléatoire est une fonction :

- 1) bornée et normalisée: $0 \leq F_X(x) \leq 1$, Sa limite en $-\infty$ est 0 et en $+\infty$ est 1.
- 2) monotone non décroissante,
- 3) continue à gauche.

Démonstration :

1) $F_X(x)$ est nécessairement comprise entre zéro et 1 puisqu'elle représente une probabilité et les valeurs aux bornes résultent de la définition d'une variable aléatoire.

2) Monotone non décroissante. En effet, si $x_1 < x_2$, alors on a: $\{X < x_1\} \subset \{X < x_2\}$ ce qui conduit (corollaire 4) à : $F_X(x_1) = P(\{X < x_1\}) \leq F_X(x_2) = P(\{X < x_2\})$

3) Continue à gauche : c'est-à-dire que $F(x)$ tend vers $F(x_0)$ quand x tend vers x_0 par valeurs inférieures (à gauche de x_0).

En considérant des intervalles disjoints, on peut écrire

$$F_X(x_0) - F_X(x_0 - \varepsilon) = P(\{x_0 - \varepsilon \leq X < x_0\}) \text{ avec } \varepsilon > 0.$$

Lorsque ε tend vers 0, l'ensemble $[x_0 - \varepsilon, x_0[$ tend, puisqu'il ne contient pas x_0 , vers l'ensemble vide de probabilité nulle et $F_X(x_0 - \varepsilon)$ tend vers $F_X(x_0)$.

Remarque 1 :

Il faut bien distinguer les inégalités strictes ($<$) et larges (\leq) quand on fait usage des fonctions de répartitions.

Il peut arriver que l'ensemble des ω tels que $X(\omega) = x_0$ ne soit pas vide. Il existe alors une probabilité non nulle d'avoir $X = x_0$. Soit : $P(\{X = x_0\}) = p_0$.

Dans ce cas, $F_X(x_0 + \varepsilon) \xrightarrow{\varepsilon \rightarrow 0} F_X(x_0) + p_0$ expression qui montre que $F_X(x)$ n'est pas continue à droite de x_0 si $p_0 \neq 0$. C'est précisément le cas des exemples précédents et chaque fois qu'on a affaire avec une variable discrète.

Inversement si une variable aléatoire est continue, sa fonction de répartition est continue (à droite et à gauche)

Remarque 2 :

Si une fonction $G(x)$ satisfait aux 3 propriétés précédentes, on peut toujours trouver une expérience E et une variable X telles que $G(x)$ en soit la fonction de répartition.

On peut ainsi parler de variable aléatoire et de fonction de répartition sans avoir à préciser l'expérience sur laquelle elles sont définies.

En attribuant la probabilité zéro aux intervalles qui ne correspondent pas aux valeurs possibles de X , on peut toujours représenter une variable aléatoire comme un point sur l'axe réel tout entier.

3.2.3 Interprétation par fréquence relative.

Pour donner un sens plus concret à la notion de fonction de répartition d'une variable aléatoire, nous en voyons ici une interprétation à l'aide des fréquences relatives.

Espace dénombrable fini :

On jette un dé et l'on choisit comme variable aléatoire le nombre marqué sur sa face extérieure. Cette épreuve est répétée n fois et l'on compte le nombre de fois $n(x)$ où les résultats sont tels que $X < x$.

Dans ce cas : $F_X(x) = P(\{X < x\}) \cong \frac{n(x)}{n}$

Cette fonction aura sensiblement l'allure de celle qu'on aurait obtenue en posant a priori que la probabilité des événements élémentaires était égale à $1/6$ (à condition que le dé ne soit pas pipé). On a fait ici une mesure de la fonction de répartition.

Espace continu - fonction temporelle :

On considère maintenant un espace continu $\Omega = \{\text{tout } x \text{ réel}\}$. Supposons que notre variable aléatoire soit une fonction temporelle que l'on observe sur une durée T . Ce pourrait être, par exemple, la tension de bruit en sortie d'un amplificateur (cf. figure 3.4).

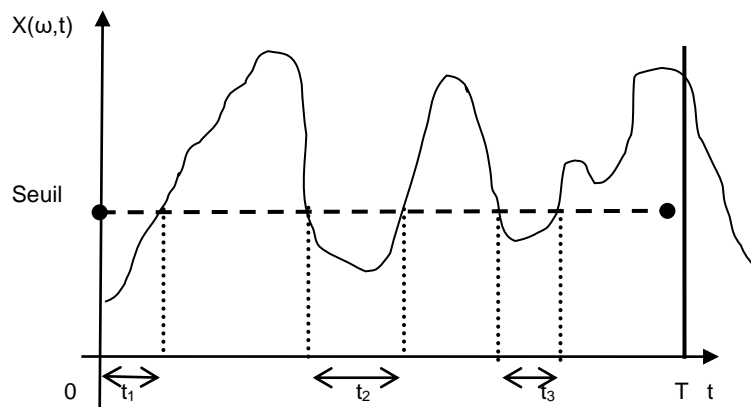


Figure 3.4

On fait la somme des durées t_i pendant lesquelles la variable X (la tension de bruit) est inférieure ou égale au seuil x ,

d'où on déduit : $F_X(x) = P(\{X < x\}) = \frac{\sum t_i}{T}$

C'est encore une méthode possible de mesurer des fonctions de répartition (avec quelques réserves qui seront vues dans le cours relatif aux fonctions aléatoires).

3.3 Fonction densité de probabilité

3.3.1 Définition générale

On appelle densité de probabilité d'une variable aléatoire X , la dérivée par rapport à x (le seuil) de sa fonction de répartition.

$f_X(x) = \frac{dF_X(x)}{dx}$	4-1
-------------------------------	-----

Si l'on prend soin de considérer cette dérivée au sens des distributions, cette définition reste valable, que la variable X soit continue, discrète ou mixte. Pratiquement, trois cas peuvent se présenter (cf. figure 4.1) :

- en x_1 : $F_X(x)$ est continue et dérivable,
- en x_2 : $F_X(x)$ est continue mais non dérivable au sens classique. Elle admet, en fait, deux dérivées : une à droite et une à gauche de x_2 ,
- en x_3 : $F_X(x)$ est discontinue. Elle n'a pas de dérivée au sens classique.

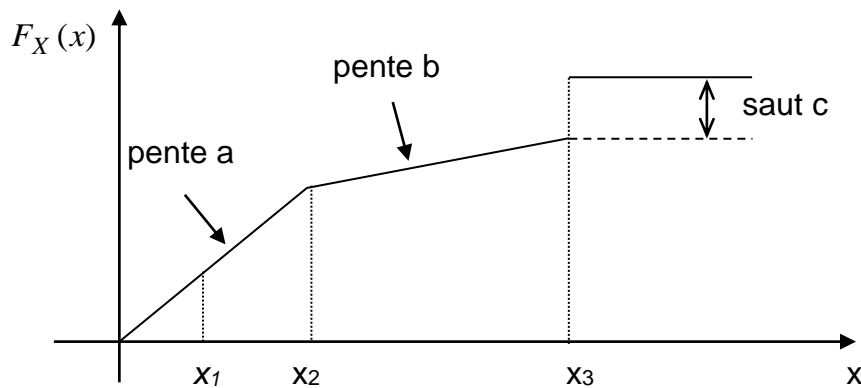


Figure 4.1

En x_3 , on peut écrire, au sens des distributions :

$$f_X(x = x_3) = [F_X(x_3^+) - F_X(x_3)] \cdot \delta(x - x_3) = c \cdot \delta(x - x_3)$$

avec $\delta(x)$: distribution de Dirac

$c = F_X(x_3^+) - F_X(x_3)$: saut de la fonction au point x_3 ,

La densité correspondante est présentée sur la figure 4.2.

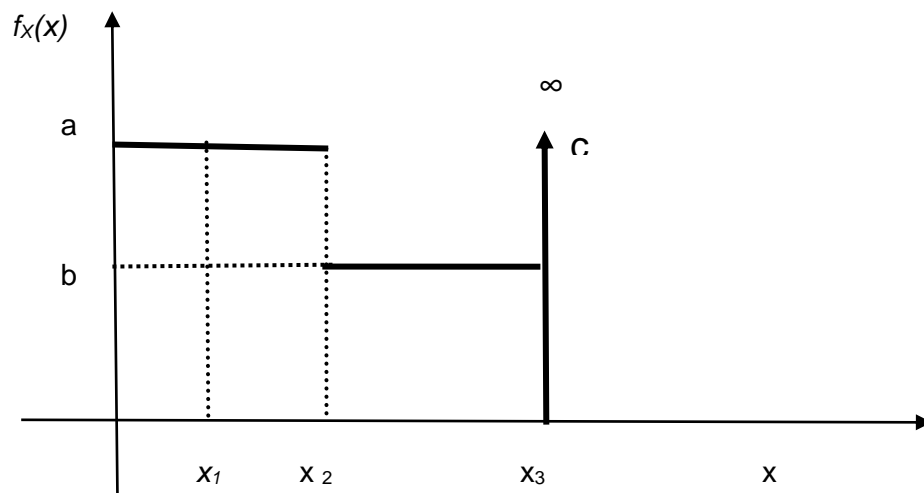


figure 4.2

3.3.2 Propriétés de la densité

1) La surface sous la densité est normalisée.

La densité étant la dérivée de la fonction de répartition,

on a : $F_X(x) = \int_{-\infty}^x f_X(u) du$ et en particulier :

$F_X(+\infty) = \int_{-\infty}^{+\infty} f_X(u) du = 1$	4-2
---	-----

2) La fonction de densité est non négative

$f_X(x) \geq 0$ car c'est la dérivée d'une fonction monotone non décroissante.

Si une fonction $f(x)$ est non négative et sa somme égale à l'unité, elle peut être considérée comme la densité de probabilité d'une variable aléatoire.

3.4 Variable aléatoire continue

3.4.1 Définition.

On dit qu'une variable aléatoire réelle est continue si sa fonction de répartition est continue et dérivable au sens usuel.

En utilisant l'axiome d'additivité, on peut écrire :

$$P(\{x_1 \leq X < x_2\}) = P(\{X < x_2\}) - P(\{X < x_1\})$$

$$P(\{x_1 \leq X < x_2\}) = F_X(x_2) - F_X(x_1)$$

$$P(\{x_1 \leq X < x_2\}) = \int_{-\infty}^{x_2} f_X(u) du - \int_{-\infty}^{x_1} f_X(u) du$$

$P(\{x_1 \leq X < x_2\}) = \int_{x_1}^{x_2} f_X(u) du$	4-3
--	-----

Remarque : Cette relation est encore valable sur l'intervalle fermé en vertu de l'hypothèse de la continuité de $F_X(x)$.

Maintenant, posons : $x_1 = x$ et $x_2 = x + dx$

on trouve alors : $P(\{x \leq X < x + dx\}) = \int_x^{x+dx} f_X(u) du$

Soit la densité pouvant être considéré comme constante sur l'intervalle :

$P(\{x \leq X < x + dx\}) = f_X(x) dx$	4-4
--	-----

$f_X(x)dx$ représente la probabilité pour la variable X d'appartenir à l'intervalle $[x, x+dx]$.

La densité peut s'interpréter comme le facteur de proportionnalité existant entre la probabilité sur un intervalle et la longueur de cet intervalle. Ce facteur dépend du seuil, c'est-à-dire de la position de l'intervalle considéré.

Il ressort de cette présentation que :

- la fonction de densité $f_X(x)$ n'est pas une probabilité
- la probabilité en un point (sur un intervalle de longueur nulle) est nulle si la densité est continue :

$$P(\{X = x\}) = \lim_{dx \rightarrow 0} f_X(x) dx = 0$$

3.4.2 Interprétation fréquentielle de la densité.

Au chapitre précédent, nous avons donné un sens physique à la notion de fonction de répartition en considérant une variable aléatoire représentée par une fonction temporelle.

La mesure du temps pendant lequel cette fonction restait en dessous du seuil x par rapport au temps total d'observation donnait une représentation fréquentielle de la fonction de répartition $F_X(x)$.

Reprenons cet exemple pour la fonction de densité mais en considérant deux seuils séparés de dx .

La mesure du temps où la fonction temporelle reste comprise entre les seuils x et $x+dx$ conduit à la fonction de densité.

$$f_X(x) dx = \frac{\sum \Delta t_i}{T}$$

En faisant varier le seuil, on obtiendra une représentation complète de la densité (cf. figure 4.3).

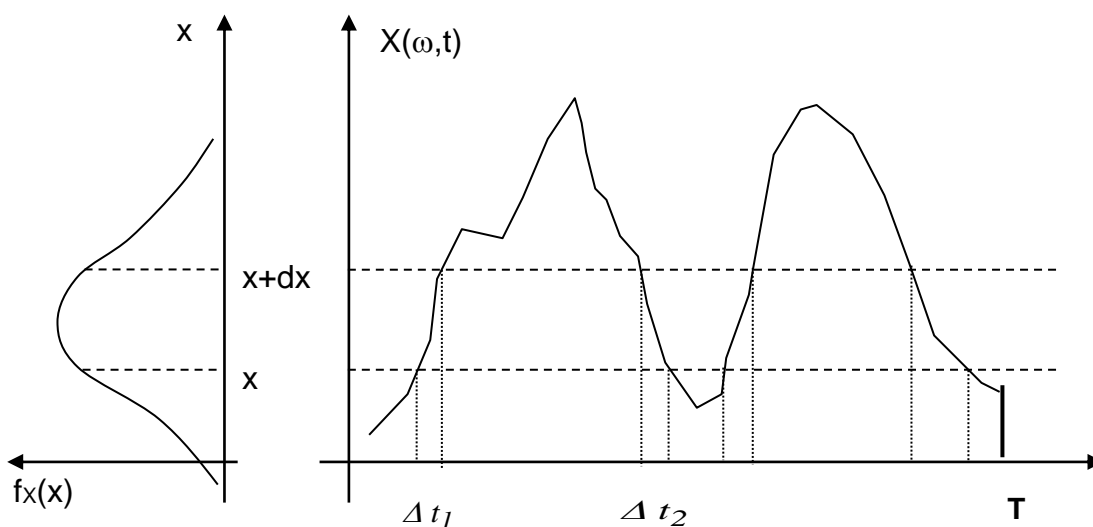


Figure 4.3

3.5 Variable aléatoire discrète

Nous considérons maintenant le cas où la variable aléatoire est définie sur un espace dénombrable (fini ou infini).

La variable X est susceptible de prendre un certain nombre de valeurs discrètes $X = x_i$ avec des probabilités p_i :

$$P(X = x_i) = p_i \neq 0 \text{ avec la condition } \sum_i p_i = 1.$$

Ce cas correspond à la plupart des exemples donnés jusqu'à présent. La fonction de répartition est alors discontinue en forme d'escalier.

En dérivant cette fonction on obtient des distributions de Dirac dont le poids p_i est égal au saut de la fonction au point x_i considéré.

On peut écrire : $F_X(x) = \sum_i P(X = x_i) \cdot H(x - x_i)$ avec $H(x)$: la fonction échelon de Heaviside (cf. figure 4.4).

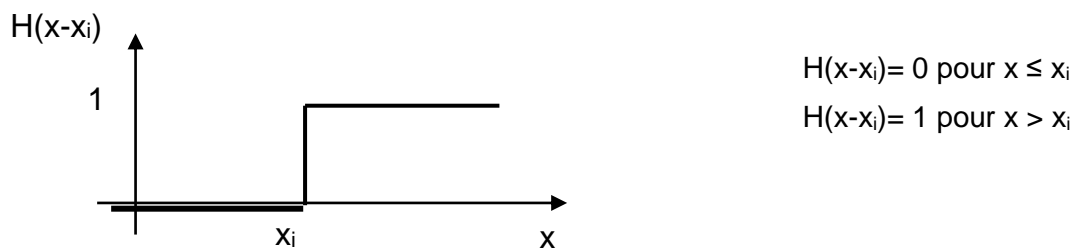


Figure 4.4

par dérivation : $f_X(x) = \sum_i P_i \cdot \delta(x - x_i)$

Comme précédemment la fonction $f_X(x)$ n'est pas une probabilité. Elle est nulle presque partout et infinie aux points $\{x_i\}$.

On peut l'interpréter de la manière suivante : Supposons que la probabilité p_i est répartie de manière uniforme sur l'intervalle $[x_i, x_i + \varepsilon]$,

sur cet intervalle, la densité est égale à p_i/ε (cf. figure 4.6).

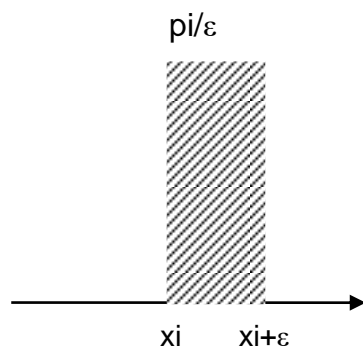


figure 4.5

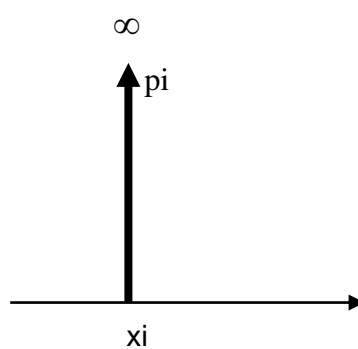


figure 4.6

Faisons tendre ε vers zéro.

Toute la probabilité se concentre en x_i et la densité tend alors vers l'infini

A la limite, on obtient une distribution de Dirac (cf. figure 4.6).

Exemple : Jeu de pile ou face

On définit les probabilités suivantes :

$$\begin{cases} P(\text{pile}) = p \\ P(\text{face}) = q \end{cases}$$

Sur ces bases, on définit la variable indicatrice suivante :

$$\begin{cases} X(\omega_i) = 1 & \text{si } \omega_i = \text{Pile} \\ X(\omega_i) = 0 & \text{si } \omega_i = \text{Face} \end{cases} \quad \text{et la relation } p + q = 1.$$

La fonction de répartition s'écrit (cf. figure 4.7) : $F_X(x) = q \cdot H(x) + p \cdot H(x-1)$

et la densité (cf. figure 4.8) : $f_X(x) = q \cdot \delta(x) + p \cdot \delta(x-1)$

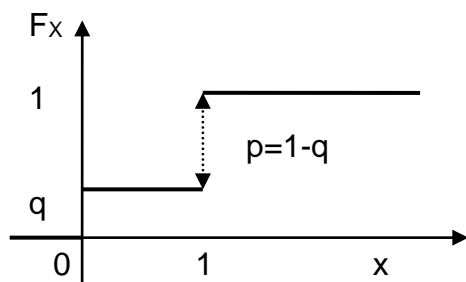


Figure 4.7

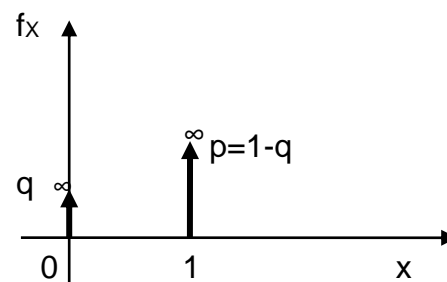


Figure 4.8

3.6 Exemples de loi de probabilité

Nous donnons ici la définition des lois de probabilités les plus courantes. Les propriétés de ces lois seront étudiées plus loin et leur intérêt apparaîtra encore plus tard.

3.6.1 Lois discontinues

1. Loi de Bernoulli

La loi de Bernoulli est la loi de la variable aléatoire X prenant deux valeurs 0 et 1 :

$$P\{X = 1\} = p \quad P\{X = 0\} = q = 1 - p$$

2. Loi binomiale

La loi binomiale est la loi de la variable aléatoire X prenant les valeurs entières 0, 1, ..., n :

$$0 \leq k \leq n \quad P\{X = k\} = C_n^k p^k q^{n-k}$$

3. Loi de Poisson

La loi de Poisson est la loi de la variable aléatoire X prenant les valeurs entières 0, 1, ... :

$$k \geq 0 \quad P(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

3.6.2 Lois continues

1. Loi uniforme (en densité)

La loi uniforme est la loi de la variable aléatoire X prenant ses valeurs dans un intervalle $[a,b]$ dont la densité est constante sur cet intervalle :

$$f_X(x) = \frac{1}{b-a} \quad \text{pour } a \leq x \leq b$$

$$f_X(x) = 0 \text{ ailleurs}$$

2. Loi normale (ou de Gauss)

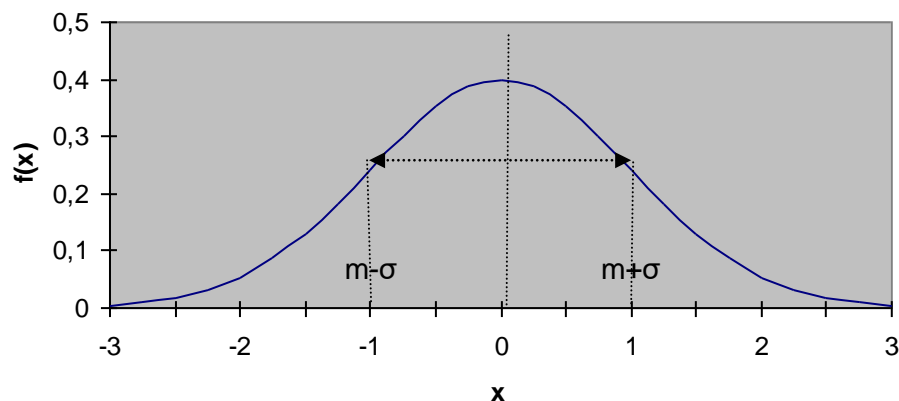
Cette loi est fondamentale en probabilité et dans toutes les branches qui s'y rattachent (statistiques, fonctions aléatoires). Son universalité dans l'interprétation des phénomènes physiques s'explique par un théorème que nous verrons en fin de cours, le théorème de la limite centrale.

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad \text{pour } x \in \mathbb{R}$$

Cette fonction de densité est symétrique par rapport à la droite d'abscisse m . Le paramètre σ correspond à l'abscisse des points d'inflexion par rapport à l'axe de symétrie. Si l'on désigne par h_0 son amplitude maximale, pour $x=m$, on a sensiblement :

$$f_X(m \pm \sigma) \cong 0,6 h_0$$

figure 4.9:Gauss



Les deux paramètres m et σ définissent entièrement la densité. Nous verrons que m est la moyenne de la répartition et σ son écart type.

Si $m=0$, la loi est dite normale centrée

Si $m=0$ et si $\sigma=1$, la loi est dite normale réduite. Dans ce cas :

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \text{pour } x \in \mathbb{R}$$

4. Loi exponentielle (ou première loi de Laplace)

La loi exponentielle est la loi d'une variable aléatoire X réelle positive dont la densité est :

$$f_X(x) = \alpha e^{-\alpha x} \quad \alpha > 0; x \geq 0$$

Cette loi est très utilisée dans les statistiques pour les problèmes du type conversations téléphoniques, files d'attente, gestion de stocks.

5. Loi de Rayleigh (ou normale circulaire ou deuxième loi de Laplace)

$$f_X(x) = \frac{x}{\alpha^2} e^{-\frac{x^2}{2\alpha^2}} \quad \text{pour } x \geq 0$$

Cette loi est très utilisée dans les problèmes de détection en présence de bruits (dans les télécommunications et dans les radars en particulier).

6. Loi de Cauchy

$$f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2} \quad \text{pour } x \in \mathbb{R}$$

Résultat du quotient de deux variables gaussiennes, elle présente la particularité d'avoir une variance infinie.

7. Loi du Chi Carré (χ^2)

$$f_X(x) = \frac{(\chi^2)^{\frac{n-2}{2}} \cdot e^{-\frac{\chi^2}{2}}}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \quad \text{pour } \chi^2 \geq 0 \quad \Gamma(.) \text{ fonction Gamma}$$

$$\text{avec } \Gamma(n) = \int_0^{+\infty} t^{n-1} e^{-t} dt \quad \Gamma(n) = (n-1)! \text{ pour } n \text{ entier}$$

Cette loi est fondamentale en statistique. Elle représente la loi de la variable obtenue en faisant la somme de n variables gaussiennes indépendantes élevées au carré comme nous le verrons au chapitre 12.

Exercice : montrer que les fonctions données ci-dessus sont bien des densités de probabilité. Calculer leur fonction de répartition.

Le tableau ci-après présente la fonction de répartition et la fonction de densité des principales lois de probabilité.

Il indique également l'espérance mathématique et l'écart type de la variable correspondante dont nous verrons la définition au chapitre suivant

Loi	Répartition $F_X(x)$	Densité $f_X(x)$	Espérance Écart type
Binomiale $X=k$ entier	$\sum_{k=0}^n C_n^k p^k q^{n-k} H(x-k)$ $0 \leq k \leq n \quad p+q=1$	$\sum_{k=0}^n C_n^k p^k q^{n-k} \delta(x-k)$	$E = np$ $\sigma = \sqrt{npq}$
Poisson $X=k$ entier ≥ 0	$e^{-\lambda} \sum_{k=0}^n \frac{\lambda^k}{k!} H(x-k)$	$e^{-\lambda} \sum_{k=0}^n \frac{\lambda^k}{k!} \delta(x-k)$	$E = \lambda$ $\sigma = \sqrt{\lambda}$
Uniforme $x_1 \leq X \leq x_2$	$\frac{x-x_2}{x_2-x_1} (H(x-x_1) - H(x-x_2))$	$\frac{1}{x_2-x_1} (\delta(x-x_1) - \delta(x-x_2))$	$E = \frac{x_1+x_2}{2}$ $\sigma = \frac{x_2-x_1}{2\sqrt{3}}$
Gauss $X \in R$	$\frac{1}{2} + G(y)$ avec $y = \frac{x-m}{\sigma}$ $G(y) = \frac{1}{\sqrt{2\pi}} \int_0^y e^{-\frac{t^2}{2}} dt$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$	$E = m$ $\sigma = \sigma$
Laplace $X \geq 0$ (exponentielle)	$1 - e^{-\alpha x}, \alpha > 0, x \geq 0$	$\alpha e^{-\alpha x}, \alpha > 0, x \geq 0$	$E = 1/\alpha$ $\sigma = 1/\alpha$
Double Exponentielle $X \in R$	$\frac{1}{2} e^{\alpha x}, x \leq 0$ $1 - \frac{1}{2} e^{-\alpha x}, x \geq 0, \alpha > 0$	$\frac{\alpha}{2} e^{-\alpha x }$ $x \geq 0, \alpha > 0$	$E = 0$ $\sigma = \sqrt{2}/\alpha$
Rayleigh $X \geq 0$	$1 - e^{-\frac{x^2}{2\alpha^2}}, x \geq 0, \alpha > 0$	$\frac{x}{\alpha^2} e^{-\frac{x^2}{2\alpha^2}}, x \geq 0, \alpha > 0$	$E = \alpha \sqrt{\frac{\pi}{2}}$ $\sigma = \alpha \sqrt{2 - \frac{\pi}{2}}$
Maxwell $X \geq 0$	$2G(\frac{x}{\alpha}) - \sqrt{\frac{2}{\pi}} \alpha R(x), \alpha > 0, x \geq 0$ $R(x)$ densité de Rayleigh	$\frac{1}{\alpha^3} \sqrt{\frac{2}{\pi}} x^2 e^{-\frac{x^2}{2\alpha^2}}, x \geq 0, \alpha > 0$	$E = \alpha \frac{2\sqrt{2}}{\sqrt{\pi}}$ $\sigma = \alpha \sqrt{3 - \frac{8}{\pi}}$
Cauchy $X \in R$	$\frac{1}{2} + \frac{1}{\pi} \text{Arctg}(\frac{x}{\alpha})$	$\frac{\alpha}{\pi} \frac{1}{\alpha^2 + x^2}$	$E = 0$ $\sigma = \infty$
Student $X \in R$	n entier, nombre de degrés de liberté	$\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2}) [1 + \frac{t^2}{n}]^{\frac{n+1}{2}}}$	$E = 0$
Chi Carré $\chi^2 \geq 0$	n entier, nombre de degrés de liberté	$\frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} e^{-\frac{\chi^2}{2}}$	$E = n$ $\sigma = \sqrt{2n}$

3.7 Transformation d'une variable aléatoire réelle

Une des questions fondamentales qui se pose dans l'application de la théorie des probabilités est celle de la transformation des variables aléatoires.

3.7.1 Transformation de la densité

Soit g une fonction réelle certaine (non aléatoire) de la variable x . Soit également X une variable aléatoire de densité $f_X(x)$.

On désire connaître la loi de probabilité de la variable Y déduite de X par la relation :

$$Y = g(X)$$

En d'autres termes, on cherche la probabilité de l'événement infinitésimal : $\{y \leq Y \leq y + dy\}$

1^{er} cas : la relation entre x et y est biunivoque

La fonction g est strictement monotone. Supposons la croissante. Dans ce cas :

$$x \leq X \leq x + dx \Leftrightarrow y \leq Y \leq y + dy$$

En tenant compte de la définition d'une densité de probabilité :

$$P\{x \leq X \leq x + dx\} = f_X(x) \cdot dx$$

$$P\{y \leq Y \leq y + dy\} = f_Y(y) \cdot dy$$

En différenciant : $y = g(x) \Rightarrow dy = g'(x) \cdot dx, dy > 0$

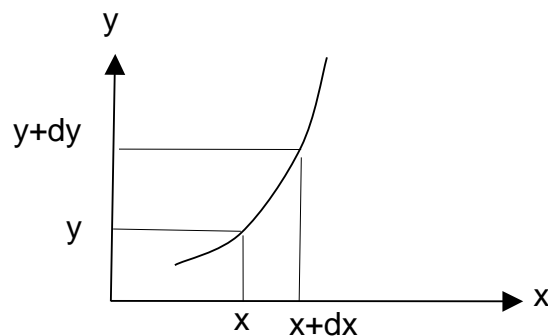
D'où l'expression des probabilités :

$$P\{y \leq Y \leq y + dy\} = f_Y(y) \cdot dy = f_X(x) \cdot dx \quad (5-1)$$

D'où l'on tire l'expression de la densité de Y :

$$f_Y(y) = f_X(x) \cdot \frac{dx}{dy} = f_X(g^{-1}(y)) \cdot [g^{-1}(y)]' \quad (5-2)$$

où $y = g^{-1}(x)$ représente la fonction inverse de $y = g(x)$.



Exemple de fonction de transformation biunivoque croissante

Si la fonction g est décroissante, on a :

$$dy = g'(x) \cdot dx > 0 \Rightarrow dx < 0$$

$$x + dx \leq X \leq x \Leftrightarrow y \leq Y \leq y + dy$$

Soit en probabilité (positive) :

$$f_Y(y) \cdot dy = -f_X(x) \cdot dx$$

Quel que soit le signe de la dérivée, on peut écrire :

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{dx}{dy} \right| = f_X(g^{-1}(y)) \cdot |(g^{-1}(y))'| \quad (5-3)$$

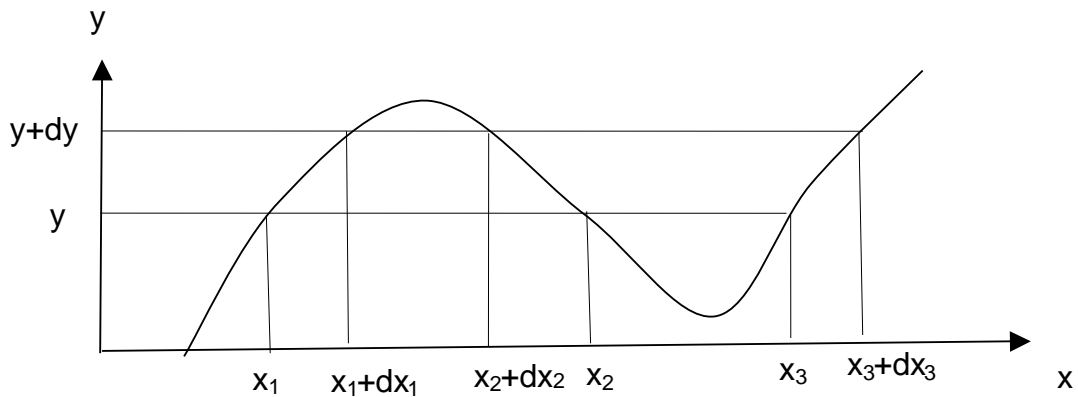
1^{er} cas : la relation entre x et y n'est pas biunivoque

Il peut exister plusieurs valeurs x_i de x correspondant à une même valeur de y telle que :

$$y = g(x_i) \quad i = 1, 2, \dots$$

Supposons, par exemple, qu'il y ait 3 valeurs de x correspondant à une même valeur de y

Désignons par x_1, x_2 et x_3 les abscisses des points d'intersection de la courbe $y = g(x)$ avec la droite d'ordonnée y et par x_1+dx_1, x_2+dx_2 et x_3+dx_3 les abscisses des points d'intersection avec la droite d'ordonnée $y+dy$ (avec $dy > 0$) comme sur la figure suivante :



Exemple de transformation non biunivoque

Pour que Y appartienne à l'intervalle $[y, y+dy]$, il faut et il suffit, que l'on ait l'une des conditions suivantes :

$$X \in [x_1, x_1 + dx_1] \quad dx_1 > 0$$

$$X \in [x_2 + dx_2, x_2] \quad dx_2 < 0$$

$$X \in [x_3, x_3 + dx_3] \quad dx_3 > 0$$

Lorsque dy est proche de 0, ces conditions sont exclusives. D'où, d'après l'axiome relatif à la probabilité d'événements disjoints :

$$P\{y \leq Y \leq y + dy\} = P\{x_1 \leq X \leq x_1 + dx_1\} + P\{x_2 + dx_2 \leq X \leq x_2\} + P\{x_3 \leq X \leq x_3 + dx_3\}$$

Ce qui s'écrit d'une manière générale en utilisant les densités :

$$f_Y(y) \cdot |dy| = \sum_i f_X(x_i) \cdot |dx_i|$$

D'où la densité de Y :

$$f_Y(y) = \sum_i f_X(x_i) \cdot \left| \frac{dx}{dy} \right|_{x=x_i} = \sum_i f_X(g_i^{-1}(y)) \cdot |g_i^{-1}(y)|' \quad (5-4)$$

Dans cette équation, les x_i sont les racines de l'équation $x = g^{-1}(y)$ à y fixé.

3.7.2 Exemples

1) Transformation est de la forme (détecteur quadratique) : $y = x^2$

La variable aléatoire X a pour densité $f_X(x)$. Calculer la densité de la variable Y déduite de X par la relation certaine (détecteur quadratique) :

$$y = x^2$$

La fonction inverse s'écrit :

$$x = \pm\sqrt{y}$$

Il y a toujours deux racines pour y positif. Calculons les pentes en dérivant :

$$dy = 2 \cdot x \cdot dx$$

$$[g_i^{-1}(y)]' = \frac{dx}{dy} = \pm \frac{1}{2\sqrt{y}}$$

D'où l'expression de la densité (attention à ne pas oublier les valeurs absolues)

$$f_Y(y) = \frac{1}{2\sqrt{y}} \cdot [f_X(\sqrt{y}) + f_X(-\sqrt{y})]$$

Application :

La variable X suit une loi de Rayleigh, dont la densité de probabilité est donnée par :

$$f_X(x) = \frac{x}{\sigma^2} \cdot \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad x \geq 0$$

Comme x est positif ou nul, on peut considérer que $f_X(x) = 0$ pour x négatif et il vient pour tout y positif :

$$f_Y(y) = \frac{1}{2\sqrt{y}} \cdot \frac{\sqrt{y}}{\sigma^2} \cdot \exp\left(-\frac{(\sqrt{y})^2}{2\sigma^2}\right) = \frac{1}{2\sigma^2} \cdot \exp\left(-\frac{y}{2\sigma^2}\right)$$

C'est une densité exponentielle.

2) Transformation est de la forme : $Y = \sin(X)$

Supposons que la variable X soit uniforme sur $[-\pi, \pi]$

Pour tout y donné dans l'intervalle $[-1, 1]$, l'équation $y = \sin(x)$ a deux racines x_1 et x_2 donnée par :

$$x_1 = \text{Arc sin}(y)$$

$$x_2 = \pi - \text{Arc sin}(y)$$

En ces deux points, les pentes sont :

$$\frac{dy}{dx} = \cos x \Rightarrow \frac{dx}{dy} = \frac{1}{\cos x} = \pm \frac{1}{\sqrt{1-y^2}}$$

D'où l'expression de la densité de Y :

$$f_Y(y) = \frac{1}{\sqrt{1-y^2}} \cdot [f_X(x_1) + f_X(x_2)]$$

Or la densité de X est uniforme sur $[-\pi, \pi]$, soit :

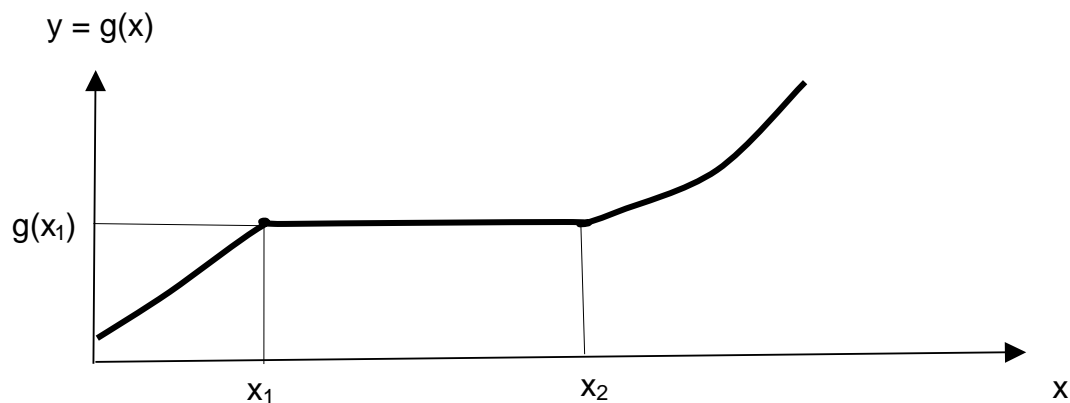
$$f_X(x_1) = f_X(x_2) = \frac{1}{2\pi}$$

D'où la densité de Y, pour y dans l'intervalle $[-1, 1]$:

$$f_Y(y) = \frac{1}{\pi} \cdot \frac{1}{\sqrt{1-y^2}}$$

3.7.3 Cas particulier d'une transformation constante sur un intervalle

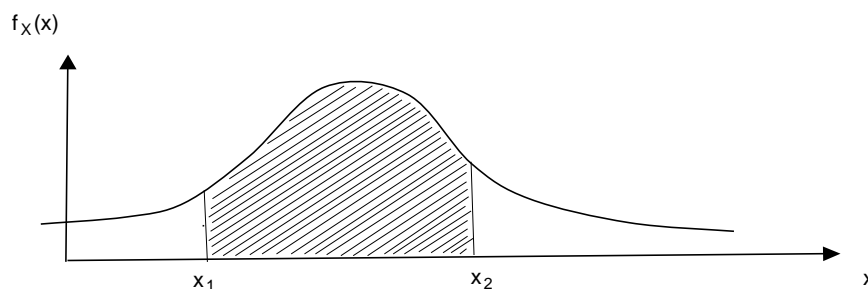
Supposons la fonction g continue, croissante pour tout x sauf sur l'intervalle $[x_1, x_2]$ où elle est constante comme sur la figure suivante :



Exemple fonction constante sur un intervalle

Dans ce cas, toutes les valeurs de x de l'intervalle $[x_1, x_2]$ se transforment en une seule valeur de y et la probabilité d'avoir $Y = g(x_1)$ est égale à la probabilité d'avoir $x_1 \leq X \leq x_2$:

$$P\{Y = g(x_1)\} = P\{x_1 \leq X \leq x_2\} = \int_{x_1}^{x_2} f_X(x) \cdot dx$$



Cette probabilité est répartie sur un intervalle de longueur nulle. On aura donc une distribution de Dirac dans la fonction de densité $f_Y(y)$ au point $Y = g(x_1)$ et une discontinuité dans la fonction de répartition $F_Y(y)$.

$$F_Y\{y = g(x_1^+)\} = F_Y\{y = g(x_1^-)\} + P\{Y = g(x_1)\}$$

Un limiteur d'amplitude ou un redresseur simple alternance sont des exemples physiques de telles transformations.

Considérons, par exemple, la variable aléatoire X uniformément répartie sur l'intervalle $[-1, 1]$ et prenons comme transformation $g(x)$ le redressement simple alternance c'est-à-dire :

$$y = x \quad \text{si} \quad x \geq 0$$

$$y = 0 \quad \text{si} \quad x \leq 0$$

On a pour $x \leq 0$:

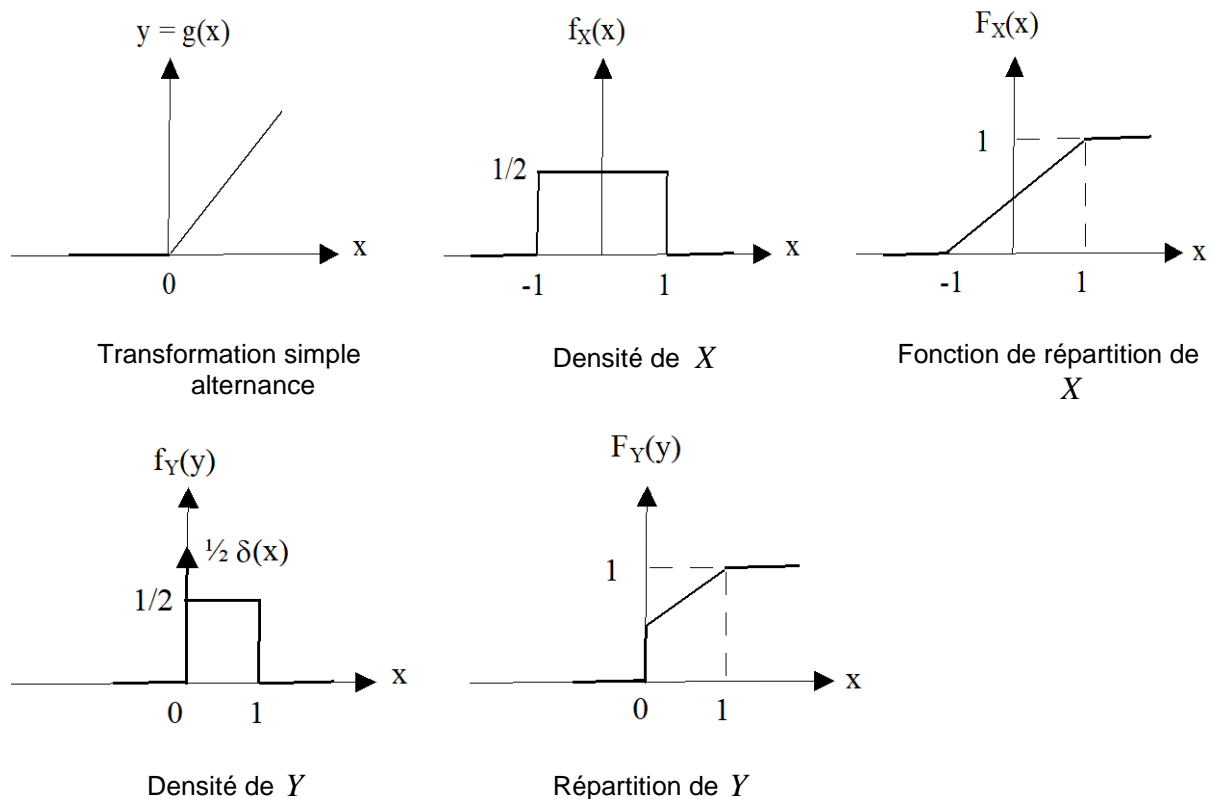
$$P\{Y = 0\} = P\{X \leq 0\} = \frac{1}{2} \int_{-1}^0 dx = \frac{1}{2}$$

Et pour $x \geq 0$:

$$f_Y(y) = f_X(x) = \frac{1}{2}$$

D'où les expressions de la densité et de la fonction de répartition de la variable aléatoire Y :

$$\begin{cases} f_Y(y) = \frac{1}{2} \cdot \delta(y) + \frac{1}{2} & , \quad 0 \leq y \leq 1 \\ F_Y(y) = \frac{1}{2} + \frac{y}{2} & , \quad 0 \leq y \leq 1 \end{cases}$$

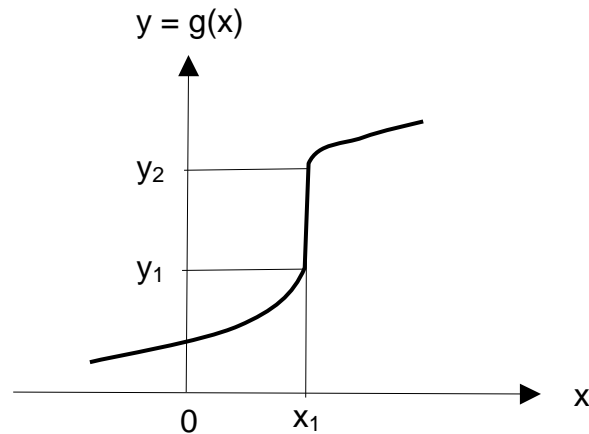


3.7.4 Cas particulier d'une transformation discontinue en un point

On suppose que la variable aléatoire X est du type continu et que $g(x)$ a une discontinuité en $g(x_1)$ telle que :

$$y_1 = g(x_1^-)$$

$$y_2 = g(x_1^+)$$



La variable X étant continue, on a :

$$P\{y_1 \leq Y \leq y_2\} = P\{X = x_1\} = 0$$

Autrement dit, la densité de Y est nulle et sa répartition est constante sur l'intervalle $[y_1, y_2]$.

Un exemple classique de transformation qui concilie ces deux cas particuliers est celui de convertisseur analogique/numérique qui transforme une variable continue en une variable discrète.

TRANSFORMATION	DENSITÉ DE PROBABILITÉ
Linéaire $Y = aX + b$	$f_Y(y) = \frac{1}{ a } \cdot f_X\left(\frac{y-b}{a}\right)$
Hyperbolique $Y = \frac{a}{X}$	$f_Y(y) = \frac{ a }{y^2} \cdot f_X\left(\frac{a}{y^2}\right)$
Quadratique $Y = a \cdot X^2$	$y/a \geq 0$ $f_Y(y) = \frac{1}{2 \cdot \sqrt{a \cdot y}} \cdot \left[f_X\left(\sqrt{\frac{y}{a}}\right) + f_X\left(-\sqrt{\frac{y}{a}}\right) \right]$
Redressement double alternance $Y = X $	$y \geq 0$ $f_Y(y) = f_X(y) + f_X(-y)$
Redressement simple alternance $Y = \frac{X + X }{2}$	$y \geq 0$ $f_Y(y) = f_X(y) + f_X(0) \cdot \delta(x)$
$Y = a \cdot \sin(X + \theta)$	$f_Y(y) = \frac{1}{\sqrt{a^2 - y^2}} \sum_{n=-\infty}^{+\infty} f_X(x_n)$ avec $\theta + x_n = \arcsin\left(\frac{y}{a}\right) + 2 \cdot n \cdot \pi \quad y \leq a$ En particulier, si $f_X(x) = \frac{1}{2 \cdot \pi}$, $a=1$ et $x \in [-\pi, \pi]$ $f_Y(y) = \frac{1}{\pi \cdot \sqrt{1 - y^2}}$
$Y = a \cdot \operatorname{tg}(X)$	$f_Y(y) = \frac{ a }{a^2 + y^2} \sum_{n=-\infty}^{+\infty} f_X(x_n)$ avec $x_n = \operatorname{arctg}\left(\frac{y}{a}\right) + n \cdot \pi$ En particulier, si $f_X(x) = \frac{1}{2 \cdot \pi}$, $a=1$ et $x \in [-\pi, \pi]$ $f_Y(y) = \frac{1}{\pi} \frac{1}{1 + y^2}$

3.8 Exercices

3.8.1 Énoncés

Fonctions de répartition

1. Deux dés équilibrés sont jetés simultanément. Posons que la variable aléatoire X est égale à la somme des chiffres présentés sur les dés.

Déterminer la fonction de répartition de la variable aléatoire X .

2. On considère l'expérience aléatoire constituée de 4 épreuves indépendantes. A chaque épreuve, la probabilité d'un événement A est $p = 1/3$ de se produire. On définit la variable X comme le nombre total de fois où l'événement A s'est produit au cours des 4 épreuves.

Décrire l'ensemble Ω en fonction de A et \bar{A} . Esquisser $F_X(x)$. Calculer $P(\{1 < X < 3\})$ et $P(\{1 \leq X < 3\})$.

3. Quelle restriction faut-il imposer à la variable X pour que la fonction $\cos(x)$ puisse représenter une fonction de répartition ?

Densité de probabilité

4. Donner l'expression de la fonction de densité relative aux exercices 1, 2 et 3 précédents.

5. La variable X a une densité uniforme sur l'intervalle $[-1, +3]$. Calculer et représenter sa fonction de répartition. Quelle est la probabilité des événements $A=\{0 \leq X \leq 2\}$, $B=\{1 \leq X \leq 3\}$, $C=\{0 \leq X \leq 3\}$. Retrouver $P(C)$ en fonction de $P(A)$ et $P(B)$ à partir de la formule donnée au chapitre I

6. La variable X a pour densité $f_X(x) = \frac{\alpha}{2} e^{-\alpha|x|}$ $\alpha > 0$; $x \in \mathbb{R}$. Calculer sa fonction de répartition $F_X(x)$. Vérifier que $F_X(x)$ satisfait les 3 propriétés d'une telle fonction. Calculer la probabilité de l'événement $\{|X| < \alpha\}$

Transformation d'une variable aléatoire réelle

7. La variable X est uniformément répartie sur l'intervalle $[0,1]$. Déterminer la densité et la fonction de répartition de la variable Y définie par les transformations :

a. $y = 3x + 1$

b. $y = -3x + 1$

c. $y = x^2$

d. $y = \ln(x)$

e. $y = \ln(x^2)$

f. $y = \exp(x)$

8. La variable X a une densité exponentielle

$$f_X(x) = \frac{1}{\alpha} \cdot \exp\left(-\frac{x}{\alpha}\right) \quad x \geq 0$$

Calculer la densité de la variable Y définie par : $y = \sqrt{x}$

9. Soit X une variable uniformément répartie sur l'intervalle $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. On considère la nouvelle variable $Y = \tan(X)$.

- Déterminer la densité de la variable aléatoire Y .
- Que devient cette densité si X est uniformément répartie sur l'intervalle $[-\pi, \pi]$?

10. La variable X est uniformément répartie sur l'intervalle $[-2c, 2c]$

- Déterminer la densité et la fonction de répartition de la variable X
- On considère la transformation :

$$Y = x \quad \text{si} \quad |x| \geq c$$

$$Y = 0 \quad \text{si} \quad |x| < c$$

Représenter cette fonction.

- Calculer la probabilité de l'événement $(Y = 0)$
- Déterminer la densité et la fonction de répartition de la variable Y

3.8.2 Solutions

Fonctions de répartition

1.

$F_X(x) =$	0	Pour	$x \leq 2$
	1/36		$2 < x \leq 3$
	3/36		$3 < x \leq 4$
	6/36		$4 < x \leq 5$
	10/36		$5 < x \leq 6$
	15/36		$6 < x \leq 7$
	21/36		$7 < x \leq 8$
	26/36		$8 < x \leq 9$
	30/36		$9 < x \leq 10$
	33/36		$10 < x \leq 11$
	35/36		$11 < x \leq 12$
	1		$12 < x$

$$\begin{aligned}
 F_X(x) = & \frac{1}{36} [H(x-2) + H(x-12)] + \frac{2}{36} [H(x-3) + H(x-11)] \\
 & + \frac{3}{36} [H(x-4) + H(x-10)] + \frac{4}{36} [H(x-5) + H(x-9)] \\
 & + \frac{5}{36} [H(x-6) + H(x-8)] + \frac{6}{36} H(x-5)
 \end{aligned}$$

2. $\Omega = \{\bar{A}\bar{A}\bar{A}\bar{A}\bar{A}; (\bar{A}\bar{A}\bar{A}\bar{A}A), \dots, (AAAAA)\}$. Le cardinal de cet ensemble est 2^4 .

$F_X(x) =$	0	pour	$x \leq 0$
	16/81		$0 < x \leq 1$
	48/81		$1 < x \leq 2$
	72/81		$2 < x \leq 3$
	80/81		$3 < x \leq 4$
	1		$x > 4$

$$F_X(x) = \frac{16}{81}H(x) + \frac{32}{81}H(x-1) + \frac{24}{81}H(x-2) + \frac{8}{81}H(x-3) + \frac{1}{81}H(x-4)$$

$$P(1 < X < 3) = 24/81 \quad P(1 \leq X < 3) = 56/81$$

3. $F_X(x) = \cos(x)$ pour $X \in \left[\frac{3 \cdot \pi}{2} + 2 \cdot k \cdot \pi; 2 \cdot \pi + 2 \cdot k \cdot \pi \right]$, k fixé

Densité de probabilité

4. Respectivement pour chacun des exercices 1, 2 et 3 :

$$f_X(x) = \frac{1}{36}[\delta(x-2) + \delta(x-12)] + \frac{2}{36}[\delta(x-3) + \delta(x-11)] \\ + \frac{3}{36}[\delta(x-4) + \delta(x-10)] + \frac{4}{36}[\delta(x-5) + \delta(x-9)] \\ + \frac{5}{36}[\delta(x-6) + \delta(x-8)] + \frac{6}{36}\delta(x-7)$$

$$f_X(x) = \frac{16}{81}\delta(x) + \frac{32}{81}\delta(x-1) + \frac{24}{81}\delta(x-2) + \frac{8}{81}\delta(x-3) + \frac{1}{81}\delta(x-4)$$

$$f_X(x) = -\sin(x) \text{ pour } \frac{3\pi}{2} + 2k\pi \leq x \leq (k+1)2\pi, k \text{ fixé}$$

5. $F_X(x) = \int_{u=-1}^x \frac{du}{4} = \frac{x+1}{4}$ pour $-1 \leq x \leq 3$ $P(A) = \frac{1}{2}$ $P(B) = \frac{1}{2}$ $P(C) = \frac{3}{4}$

$$P(C) = P(A+B) = P(A) + P(B) - P(A \cdot B) = 1 - \frac{1}{4} = \frac{3}{4}$$

6. $F_X(x) = \frac{1}{2}e^{\alpha x}$ pour $x \leq 0$ $F_X(x) = 1 - \frac{1}{2}e^{-\alpha x}$ pour $x \geq 0$

$$F_X(-\infty) = 0, F_X(x) \text{ est continue } \forall x, F_X(+\infty) = 1$$

$$P\{|X| < \alpha\} = 1 - e^{-\alpha^2}$$

Transformation d'une variable aléatoire réelle

7.

a. $f_Y(y) = \frac{1}{3}$ $F_Y(y) = \frac{y-1}{3}$ si $1 \leq y \leq 4$

b. $f_Y(y) = \frac{1}{3} \quad F_Y(y) = \frac{y+2}{3} \quad \text{si } -2 \leq y \leq 1$

c. $f_Y(y) = \frac{1}{2 \cdot \sqrt{y}} \quad F_Y(y) = \sqrt{y} \quad \text{si } 0 \leq y \leq 1$

d. $f_Y(y) = e^y \quad F_Y(y) = e^y \quad \text{si } y \leq 0$

e. $f_Y(y) = \frac{1}{2} \cdot \exp\left(\frac{y}{2}\right) \quad F_Y(y) = \exp\left(\frac{y}{2}\right) \quad \text{si } y \leq 0$

f. $f_Y(y) = \frac{1}{y} \quad F_Y(y) = \ln(y) \quad \text{si } 1 \leq y \leq e$

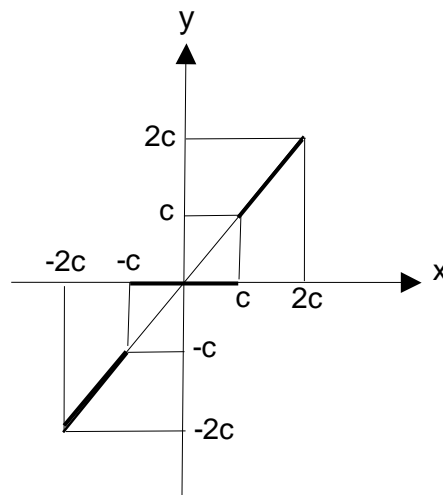
8. $f_Y(y) = \frac{2y}{\alpha} \exp\left(-\frac{y^2}{\alpha}\right) \quad \text{si } y \geq 0$

9. a. $f_Y(y) = \frac{1}{\pi} \cdot \frac{1}{1+y^2}$ b. La densité est identique

10.

a. $f_X(x) = \frac{1}{4c} \quad F_X(x) = \frac{x+2c}{4c} \quad \text{si } -2c \leq x \leq 2c$

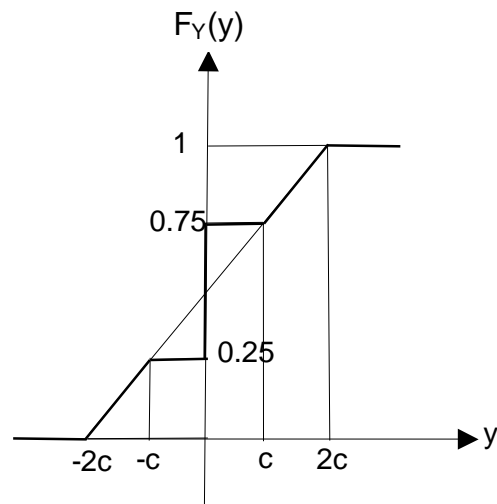
b.



c. $P\{Y = 0\} = F_X(c) - F_X(-c) = \frac{1}{2}$

d. $-2 \cdot c < y \leq -c \quad f_Y(y) = \frac{1}{4 \cdot c} \quad F_Y(y) = F_X(y)$
 $-c < y < 0 \quad f_Y(y) = 0 \quad F_Y(y) = F_X(-c)$

$y = 0$	$f_Y(y) = \frac{1}{2} \cdot \delta(y)$	$F_Y(y) = F_X(-c)$
$0 < y < c$	$f_Y(y) = 0$	$F_Y(y) = F_X(c)$
$c < y < 2 \cdot c$	$f_Y(y) = \frac{1}{4 \cdot c}$	$F_Y(y) = F_X(y)$



CHAPITRE 4. Valeurs typiques d'une variable aléatoire réelle

Une variable aléatoire est entièrement définie par la donnée de sa fonction de répartition ou par celle de sa densité. Dans la pratique, ces fonctions ne sont pas toujours très maniables, ni même connues ; il est alors très intéressant d'essayer de décrire, d'une manière plus ou moins précise, une variable aléatoire par la donnée de paramètres certains, éventuellement accessibles à la mesure comme nous le verrons dans le cours de statistique.

Les valeurs typiques généralement utilisées pour décrire une variable aléatoire sont :

- l'espérance mathématique (ou valeur moyenne) (Notation US : mean or average value)
- la valeur médiane (ou équiprobable, US : median)
- la valeur la plus probable (US : mode)
- les percentiles (ou à P%)

Enfin d'une manière plus générale, les moments et en particulier le moment centré d'ordre deux (ou variance).

4.1 Espérance mathématique

4.1.1 Définition

On appelle espérance mathématique de la variable aléatoire X la quantité :

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f_X(x) \cdot dx \quad (6-1)$$

Cette définition est valable pour une variable discrète comme pour une variable continue. Si X est de type valeur aléatoire discrète, sa densité s'écrit :

$$f_X(x) = \sum_i P_i \cdot \delta(x - x_i)$$

$$E(X) = \int_{-\infty}^{+\infty} x \cdot \sum_i P_i \cdot \delta(x - x_i) \cdot dx = \sum_i P_i \cdot \int x \delta(x - x_i) dx$$

$$E(X) = \sum_i P_i \cdot x_i \quad (6-2)$$

4.1.2 Propriétés

Si la densité est symétrique par rapport à une valeur a , l'espérance est égale à cette valeur (sous réserve de convergence de l'intégrale).

$$E(X) = a \text{ si } f_X(x) = f_X(2a - x)$$

L'espérance n'est pas nécessairement finie. On rencontre en effet des lois qui conduisent à des valeurs infinies de la moyenne.

Exemple : si la variable aléatoire suit une loi de Cauchy :

$$f_X(y) = \frac{1}{\pi} \cdot \frac{1}{1 + x^2}$$

$$E(X) = \frac{1}{\pi} \cdot \int_{-\infty}^{+\infty} \frac{x}{1 + x^2} \cdot dx$$

Cette intégrale est divergente et l'espérance n'est pas définie.

L'espérance a la même dimension que la variable aléatoire : si X est un temps, $E(X)$ est aussi un temps.

Dans le cas d'une variable discrète, l'espérance peut être égale à une valeur qui n'est pas prise par la variable.

Exemple:

Dans une expérience avec une pièce on prend :

$X = +1$ si le résultat est pile avec $P\{X = +1\} = 0.5$

$X = -1$ si le résultat est face avec $P\{X = -1\} = 0.5$

L'espérance est donnée par : $E(X) = \sum_i P_i \cdot x_i = 0.5 \times (-1) + 0.5 \times (+1) = 0$

L'espérance est nulle alors que la variable aléatoire X ne peut prendre que les valeurs ± 1 .

L'espérance est parfois notée \bar{x} , bien que cette terminologie soit plutôt réservée à la moyenne arithmétique de n observations expérimentales :

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

Cette dernière expression se rattache à la notion fréquentielle de probabilité.

4.1.3 Interprétation fréquentielle de l'espérance

On considère un espace probabilisé dénombrable fini de dimension N .

Il y a N résultats possibles $\{\omega_i\}_{i=1\dots N}$, auxquels on fait correspondre N valeurs numériques par l'application :

$x_i = \varphi(\omega_i)$, ce qui définit la variable aléatoire discrète $X = x_i$, où $i = 1\dots N$.

On répète n fois une épreuve pour obtenir n résultats.

Certains résultats pouvant être identiques, il faut compter le nombre de fois où un même résultat apparaît ce qui donne :

Résultat	Se produit	Ce qui donne
ω_1	k_1 fois	k_1 fois, $x_1 = \varphi(\omega_1)$
ω_2	k_2 fois	K_2 fois, $x_2 = \varphi(\omega_2)$
...
ω_N	k_N fois	k_N fois, $x_N = \varphi(\omega_N)$
N résultats	n fois	

En prenant la moyenne arithmétique, on obtient :

$$\bar{x} = \frac{k_1 \cdot x_1 + k_2 \cdot x_2 + \dots + k_N \cdot x_N}{n}$$

Expression qui peut s'écrire encore

$$\bar{x} = \left(\frac{k_1}{n}\right) \cdot x_1 + \left(\frac{k_2}{n}\right) \cdot x_2 + \dots + \left(\frac{k_N}{n}\right) \cdot x_N$$

Les facteurs $\left(\frac{k_i}{n}\right)$ représentent la fréquence relative de la valeur x_i qui peut être assimilée à la probabilité d'avoir $X = x_i$ quand n devient grand.

$$P_i = P\{X = x_i\} \approx \frac{k_i}{n}$$

Ce qui donne

$$\bar{x} \approx \sum_{i=1}^N P_i \cdot x_i = E(X)$$

On retrouve la définition de l'espérance pour une variable discrète.

4.1.4 Espérance d'une fonction certaine de la variable aléatoire

Nous avons vu qu'il est possible de définir une nouvelle variable aléatoire Y à partir de la variable aléatoire X par l'intermédiaire d'une fonction certaine g en posant que :

$$Y = g(X)$$

On se propose ici de calculer l'espérance de la variable aléatoire Y sans avoir à calculer explicitement sa fonction de densité $f_Y(y)$. Par définition, on peut écrire que :

$$E(Y) = \int_{-\infty}^{+\infty} y \cdot f_Y(y) \cdot dy = E\{g(X)\}$$

Si $g(X)$ est une fonction monotone croissante, les densités sont liées par la relation :

$$f_Y(y) \cdot dy = f_X(x) \cdot dx$$

En effectuant le changement de variable $y = g(x)$ dans l'intégrale de définition, on obtient :

$$E(Y) = \int_{-\infty}^{+\infty} g(x) \cdot f_X(x) \cdot \frac{dx}{dy} \cdot dy = \int_{-\infty}^{+\infty} g(x) \cdot f_X(x) \cdot dx \quad (6-3)$$

On obtiendrait le même résultat avec une fonction $g(X)$ monotone décroissante (exercice).

On démontre, sans difficulté, que cette expression est encore valable si l'équation $y = g(x)$ a plusieurs racines pour y donné.

Nous retiendrons, c'est un résultat important, que pour calculer l'espérance d'une variable aléatoire Y déduite de X par la relation certaine $y = g(x)$, on peut :

- soit calculer sa densité $f_Y(y)$ et en déduire son espérance $E(Y)$
- soit calculer directement $E(g(X))$ en utilisant $f_X(x)$

4.2 Valeur équiprobable (ou médiane)

C'est la valeur x_e qui a autant de chances d'être dépassée que non dépassée :

$$P(X < x_e) = P(X > x_e) = 0.5 = F_X(x_e)$$

Cette valeur ne doit pas être confondue avec l'espérance (on a $E(X) = m_1 = x_e$ lorsque la densité est symétrique).

4.3 Valeur la plus probable

C'est l'abscisse du maximum de la courbe de densité de probabilité (cas continu) ou celle de la valeur x_i qui a la plus grande probabilité (cas discret).

Remarquons que cette valeur n'est pas nécessairement unique.

4.4 Percentile ou valeur à P%

C'est la valeur x_p qui a la probabilité $\frac{P}{100}$ d'être non dépassée.

$$P(X < x_p) = F(x_p) = \int_{-\infty}^{x_p} f_X(x) \cdot dx = \frac{P}{100}$$

Dans la pratique on s'intéresse souvent aux quartiles qui sont les valeurs à 25 %, 50 % (la médiane) et 75 %.

4.5 Moments d'une variable aléatoire

4.5.1 Définition

La notion de moment généralise celle d'espérance.

On appelle moment d'ordre n d'une variable aléatoire l'espérance mathématique de la fonction : $g(X) = X^n$, où n est un entier.

$m_n = \int_{-\infty}^{+\infty} x^n \cdot f_X(x) \cdot dx$	(6-4)
--	-------

En particulier, avec ces notations :

- $m_0 = 1$,
- $m_1 = E(X)$

4.5.2 Moment du second ordre – variance et écart-type

Le moment du second ordre est très important. Il correspond dans les processus physiques à la notion de puissance. De plus, quand il est centré, il donne une mesure de la dispersion d'une variable aléatoire autour de sa valeur moyenne.

On a, par définition :

$$m_2 = \int_{-\infty}^{+\infty} x^2 f_X(x) dx, \quad m_2 \text{ n'est jamais négatif}$$

Définition :

Pour évaluer la dispersion de la variable aléatoire X autour de sa moyenne, nous définissons une nouvelle variable aléatoire Y :

$$Y = X - m_1, \text{ où } m_1 = E(X)$$

Dans ce cas, $E(Y) = 0$ et la variable est centrée. La variance est alors définie comme le moment du second ordre de Y :

$$Var(X) = E(Y^2) = \sigma_X^2$$

Soit pour une variable continue :

$$\sigma_X^2 = E((X - m_1)^2) = \int_{-\infty}^{+\infty} (x - m_1)^2 \cdot f_X(x) \cdot dx \quad (6-5)$$

et pour une variable discrète :

$$\sigma_X^2 = \sum_i (x_i - m_1)^2 \cdot P_i \quad (6-6)$$

Relations utiles

En développant le carré on obtient :

$$\sigma_X^2 = Var(X) = E(X^2) - E^2(X) = m_2 - m_1^2 \quad (6-7)$$

Si c est une constante,

$$Var(cX) = c^2 Var(X) \quad (6-8)$$

La racine carrée de la variance s'appelle l'écart type (US : standard deviation) :

$$\sigma_X = \sqrt{E(X^2) - E^2(X)} \quad (6-9)$$

Exercice :

Déterminer la moyenne et la variance de la variable réduite : $Y = \frac{X - m_1}{\sigma_X}$

Exemple :

a) Cas continu : on considère la densité suivante

$$f_X(x) = \frac{1}{b\sqrt{2\pi}} \cdot \exp\left(-\frac{(x-a)^2}{2 \cdot b^2}\right)$$

Calculer sa moyenne et sa variance.

La densité $f_X(x)$ est symétrique par rapport à a : $f_X(x+a) = f_X(-x+a)$.

Donc $m_1 = E(X) = a$.

La fonction $f_X(x)$ étant une densité, on a :

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{(x-a)^2}{2 \cdot b^2}\right) \cdot dx = b\sqrt{2\pi}$$

En dérivant par rapport à b , on obtient :

$$\int_{-\infty}^{+\infty} \frac{(x-a)^2}{b^3} \cdot \exp\left(-\frac{(x-a)^2}{2 \cdot b^2}\right) \cdot dx = \sqrt{2\pi}$$

$$\text{Soit : } b^2 = \frac{1}{b \cdot \sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-a)^2 \cdot \exp\left(-\frac{(x-a)^2}{2 \cdot b^2}\right) \cdot dx$$

$$b^2 = E((x-a)^2) = \sigma_X^2.$$

On aura reconnu une densité gaussienne et ce qui précède justifie l'écriture habituelle :

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(x-m_1)^2}{2 \cdot \sigma^2}\right)$$

b) Cas discret : on considère une expérience avec un événement A dont la probabilité de réalisation est p :

$$P(A) = p, P(\bar{A}) = q = 1 - p$$

et la variable aléatoire indicatrice X qui prend les valeurs :

$$X = 1 \quad \text{si } A \quad P(X = 1) = p$$

$$X = 0 \quad \text{si } \bar{A} \quad P(\bar{A}) = q$$

D'après l'équation (6-4) on a :

$$E(X) = \sum_{i=1}^2 x_i \cdot P_i = 0 \cdot q + 1 \cdot p = p$$

$$E(X^2) = \sum_{i=1}^2 x_i^2 \cdot P_i = 0^2 \cdot q + 1^2 \cdot p = p$$

Soit pour la variance d'après (6-7)

$$\sigma_X^2 = p - p^2 = p(1 - p) = pq$$

4.5.3 Théorème de Bienaymé-Tchebychev

Ce théorème s'exprime sous la forme d'une inégalité qui met en évidence le rôle fondamental que jouent la moyenne et la variance dans la description d'une variable aléatoire

$\text{Soit } P\{ X - m_1 \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2} \quad \varepsilon > 0 \quad (6-10)$

Cette inégalité signifie que la probabilité pour qu'une variable aléatoire s'écarte de sa moyenne de plus de ε est d'autant plus faible que sa variance est petite.

Elle n'est évidemment significative que pour $\varepsilon > \sigma$.

Démonstration :

D'après la définition de la variance on peut écrire :

$$\sigma_X^2 = E((X - m_1)^2) = \int_{-\infty}^{+\infty} (x - m_1)^2 \cdot f_X(x) \cdot dx$$

Décomposons l'intégrale en trois parties :

$$\sigma_X^2 = \underbrace{\int_{-\infty}^{m_1 - \varepsilon}}_{I_1 \geq 0} + \underbrace{\int_{m_1 - \varepsilon}^{m_1 + \varepsilon}}_{I_2 \geq 0} + \underbrace{\int_{m_1 + \varepsilon}^{+\infty} (x - m_1)^2 \cdot f_X(x) \cdot dx}_{I_3 \geq 0}$$

La fonction $(x - m_1)^2 f_X(x)$ étant positive, les trois intégrales du second membre sont positives et l'on peut minorer σ^2 par :

$$\sigma_X^2 \geq \int_{-\infty}^{m_1 - \varepsilon} + \int_{m_1 + \varepsilon}^{+\infty} (x - m_1)^2 \cdot f_X(x) \cdot dx$$

Sur l'intervalle de définition de ces deux intégrales, on a encore :

$$(x - m_1)^2 \geq \varepsilon^2$$

D'où une nouvelle minoration :

$$\sigma_X^2 \geq \varepsilon^2 \cdot \left[\int_{-\infty}^{m_1 - \varepsilon} f_X(x) \cdot dx + \int_{m_1 + \varepsilon}^{+\infty} f_X(x) \cdot dx \right]$$

L'expression entre crochets représente la probabilité de l'événement $\{|X - m_1| \geq \varepsilon\}$.

D'où la relation : $\frac{\sigma^2}{\varepsilon^2} \geq P\{|X - m_1| \geq \varepsilon\}$

On pose parfois $\varepsilon = t\sigma$ et l'inégalité devient :

$$P\{|X - m_1| \geq t \cdot \sigma\} \leq \frac{1}{t^2}$$

Avec $t = 10$ on trouve que, quelle que soit la loi de la variable aléatoire X , la probabilité pour qu'elle prenne une valeur supérieure à 10σ est inférieure à 1%.

4.5.4 Moment centré

La variance peut être considérée comme le moment centré d'ordre deux. On a obtenu une relation entre ce moment et les deux premiers moments non centrés, qui était :

$$Var(X) = E(X^2) - E^2(X)$$

Nous allons étendre cette relation aux moments d'ordre n . Désignons par μ_n les moments centrés.

On a par définition :

$$\mu_n = E((X - m_1)^n) = \int_{-\infty}^{+\infty} (x - m_1)^n \cdot f_X(x) \cdot dx$$

A l'aide des résultats précédents on peut écrire :

$$\mu_0 = 1$$

$$\mu_1 = 0$$

$$\mu_2 = m_2 - m_1^2$$

D'une manière plus générale, on obtient, en développant sous le signe somme l'expression $(x - m_1)^n$ et en intégrant terme à terme,

$$\mu_n = \sum_{r=0}^n C_n^r \cdot (-1)^r \cdot m_1^r \cdot m_{n-r} \quad C_n^r = \frac{n!}{r!(n-r)!}$$

De la même façon on obtient les moments non centrés en fonction des moments centrés en posant :

$$X^n = [(X - m_1) + m_1]^n$$

$$E(X^n) = m_n = \int_{-\infty}^{+\infty} [(x - m_1) + m_1]^n \cdot f_X(x) \cdot dx$$

$$m_n = \sum_{r=0}^n C_n^r \cdot m_1^r \cdot \mu_{n-r}$$

4.5.5 Moments absolus et moments généralisés

On utilise parfois d'autres types de moments et en particulier les moments absolus et généralisés.

Les moments absolus sont définis par :

$$M_n = E(|X|^n) = \int_{-\infty}^{+\infty} |x|^n \cdot f_X(x) \cdot dx$$

et les moments généralisés par :

$$m_{n,a} = E((X - a)^n)$$

où a n'est plus le moment du 1^{er} ordre mais une constante arbitraire.

4.6 Moments des répartitions courantes

On démontrera, après des calculs parfois longs mais sans difficulté les expressions suivantes.

1) Densité uniforme

$$f_X(x) = \frac{1}{2 \cdot c}, \text{ pour } x \in [-c, c]$$

$$m_n = \begin{cases} \frac{c^n}{n+1} & \text{si } n \text{ est pair} \\ 0 & \text{si } n \text{ est impair} \end{cases}$$

En particulier : $m_1 = \mu_1 = 0$, et $m_2 = \mu_2 = \frac{c^2}{3}$

2) Densité gaussienne centrée

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(x-m_1)^2}{2 \cdot \sigma^2}\right)$$

$$m_n = \begin{cases} 1 \cdot 3 \cdot \dots \cdot (n-1) \cdot \sigma^n & \text{si } n \text{ est pair} \\ 0 & \text{si } n \text{ est impair} \end{cases}$$

En particulier : $m_1 = 0$, $m_2 = \sigma^2$, $m_3 = 0$, $m_4 = 3\sigma^4$

3) Densité de Rayleigh

$$f_X(x) = \frac{x}{\alpha^2} \cdot \exp\left(-\frac{x^2}{2 \cdot \alpha^2}\right) \quad x \geq 0$$

$$m_n = \begin{cases} 1 \cdot 3 \cdot \dots \cdot n \cdot \alpha^n \cdot \sqrt{\frac{\pi}{2}} & \text{si } n \text{ est impair} \\ 2^{\frac{n}{2}} \left(\frac{n}{2}\right)! \cdot \alpha^n & \text{si } n \text{ est pair} \end{cases}$$

En particulier : $m_1 = \alpha\sqrt{\frac{\pi}{2}}$ et $m_2 = 2\alpha^2$

4.7 Exercices

4.7.1 Énoncés

1) Déterminer la valeur la plus probable x_p , la médiane x_e , l'espérance m_1 et la variance s^2 de la densité ci-dessous.

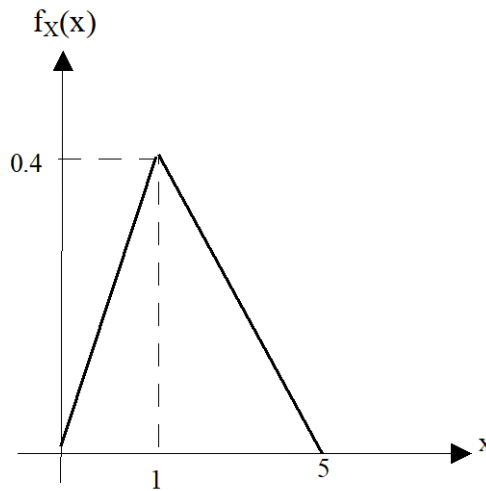


Figure 1

- 2) On considère la v.a. continue X pour laquelle $E(X) = 1$ et $E(X^2) = 2$
- Déterminer l'espérance et la variance de la v.a. $Y = 2X - 3$,
 - On suppose que X est uniformément répartie sur l'intervalle $[-2, 4]$
 - Retrouver $E(X)$ et $E(X^2)$,
 - Déterminer la densité de $Y = 2X - 3$,
 - En déduire l'espérance et la variance de Y .
- 3) On lance 3 fois de suite une pièce truquée pour laquelle $P(\text{face}) = \frac{2}{3}$ et $P(\text{pile}) = \frac{1}{3}$.
- Déterminer la loi de probabilité de la variable X égale au nombre de fois où l'événement face se produit.
 - En déduire l'espérance et la variance de X .
- 4) On considère une expérience aléatoire susceptible de donner deux résultats, A avec la probabilité $p(A) = p$ et \bar{A} avec la probabilité $p(\bar{A}) = q = 1 - p$.
On définit sur cette expérience la v.a. X par :
- $X = 1$, si $\omega_1 = A$
- $X = 0$, si $\omega_1 = \bar{A}$
- Calculer l'espérance et la variance de X .
 - On considère la v.a. de fonction de répartition $F_X(x)$.

Avec x_0 un nombre arbitraire, on forme la fonction :

$$g(x) = \begin{cases} 1 & \text{si } x < x_0 \\ 0 & \text{si } x \geq x_0 \end{cases}$$

Montrer que si $Y = g(X)$, $E(Y) = F_X(x_0)$.

5)

a. Utiliser le théorème de Tchebychev pour minorer la probabilité qu'une v.a. s'écarte, en valeur absolue, de sa moyenne d'une valeur supérieure à 3 fois son écart-type.

b. Calculer exactement la probabilité de cet événement sachant que la variable suit une loi de Gauss avec $F_X(x = m + 3\sigma) = 0,9987$.

4.7.2 Solutions

1)

$$x_p = 1$$

$$x_e = 5 - \sqrt{10} \approx 1.84$$

$$m_1 = E(X) = 2$$

$$\text{Var}(X) = 7/6 \approx 1.17$$

2)

$$\text{a. } E(Y) = 2 \cdot E(X) - 3 = -1 ; \text{Var}(Y) = 4 \cdot V(X) = 4$$

$$\text{b. } f_Y(y) = \frac{1}{12} \quad -7 \leq Y \leq 5. \text{ On retrouve } E(Y) = -1 \text{ et } V(Y) = 4$$

3)

$$f_X(x) = \frac{1}{27} [\delta(x) + 6 \cdot \delta(x-1) + 12 \cdot \delta(x-2) + 8 \cdot \delta(x-3)]$$

$$E(X) = 2$$

$$\text{Var}(X) = 2/3$$

4)

$$\text{a. } f_X(x) = (1-p) \cdot \delta(x) + p \cdot \delta(x-1)$$

$$E(X) = p$$

$$\text{Var}(X) = p \cdot (1-p)$$

$$\text{b. } f_Y(y) = (1 - F_X(x_0)) \cdot \delta(x) + F_X(x_0) \cdot \delta(x-1)$$

$$E(Y) = F_X(x_0)$$

5)

a. D'après Tchebychev : $P\{|X - m| \geq 3 \cdot \sigma\} \leq \frac{1}{9} \approx 0.11$

b. $P\{|X - m| \geq 3 \cdot \sigma\} = 2 \cdot [1 - F_X(x = m + 3 \cdot \sigma)] = 0.0026$

CHAPITRE 5. Fonction caractéristique d'une variable aléatoire réelle

5.1 Définition

On appelle fonction caractéristique de la variable aléatoire réelle X , la fonction complexe de t définie par :

$$\begin{cases} R & \rightarrow & C \\ t & \mapsto & \varphi_X(t) = E\{e^{jtx}\} \end{cases} \quad (7-1)$$

La fonction caractéristique apparaît comme l'espérance d'une fonction complexe $Z = e^{jtx}$ de la variable aléatoire réelle X .

Si X est une variable aléatoire continue de densité $f_X(x)$, on a :

$$\varphi_X(t) = \int_{-\infty}^{+\infty} e^{jtx} f_X(x) dx \quad (7-2)$$

Si X est une variable aléatoire discrète on posera, par définition :

$$\begin{aligned} \varphi_X(t) &= \sum_i P_i \cdot e^{j \cdot t \cdot x_i} \\ \text{avec } P_i &= \text{Prob}\{X = x_i\} \end{aligned} \quad (7-3)$$

Remarquons que si on écrit la densité en utilisant les distributions et en appliquant la définition générale :

$$\begin{aligned} f_X(x) &= \sum_i P_i \delta(x - x_i) \\ \varphi_X(t) &= \int_{-\infty}^{+\infty} e^{jtx} \cdot \sum_i P_i \delta(x - x_i) dx \end{aligned}$$

soit, en permutant les opérateurs de sommation discrète et d'intégration :

$$\varphi_X(t) = \sum_i P_i \int_{-\infty}^{+\infty} e^{jtx} \cdot \delta(x - x_i) dx$$

La propriété d'échantillonnage de la distribution de Dirac donne :

$$\int_{-\infty}^{+\infty} e^{jtx} \cdot \delta(x - x_i) dx = e^{jtx_i}$$

$$\text{et } \varphi_X(t) = \sum_i P_i e^{jtx_i}$$

Là encore, l'introduction des distributions permet d'uniformiser les définitions.

Très souvent, les valeurs x_i prises par une variable aléatoire discrète X sont multiples d'un même pas c :

$$X = x_i = i \cdot c \text{ pour } i = 0, 1, 2, \dots$$

Dans ce cas, si l'on change t en $t + \frac{2\pi}{c}$ on obtient :

$$\varphi_X\left(t + \frac{2\pi}{c}\right) = \sum_i e^{j\left(t + \frac{2\pi}{c}\right)ic} \cdot P_i$$

$$\varphi_X\left(t + \frac{2\pi}{c}\right) = \varphi_X(t)$$

La fonction caractéristique est dans ce cas périodique, de période : $\frac{2\pi}{c}$.

Remarque 1 :

En prenant comme définition de la transformée de Fourier d'une fonction $g(x)$ l'expression :

$$F[g(x)] = \tilde{g}(y) = \int_{-\infty}^{+\infty} g(x) e^{-j2\pi yx} dx$$

On trouve que :

$$\varphi_X(t) = \tilde{f}_X\left(-\frac{t}{2\pi}\right)$$

La fonction caractéristique d'une variable aléatoire est la transformée de Fourier, au signe près, de sa fonction de densité.

Remarque 2 :

On introduit parfois, pour des raisons de commodités, une seconde fonction caractéristique (la précédente étant la première). Cette seconde fonction est définie par :

$\phi_X(t) = \log[\varphi_X(t)]$	(7- 4)
----------------------------------	---------

5.2 Propriétés

5.2.1 Existence et borne supérieure

La fonction caractéristique existe toujours. En effet, on a, par définition :

$$\varphi_X(t) = \int_{-\infty}^{+\infty} e^{jtx} f_X(x) dx$$

$$|\varphi_X(t)| = \left| \int_{-\infty}^{+\infty} e^{jtx} f_X(x) dx \right| \leq \int_{-\infty}^{+\infty} |e^{jtx} f_X(x)| dx$$

$$|\varphi_X(t)| \leq \int_{-\infty}^{+\infty} f_X(x) dx = 1 \text{ car } f_X(x) \text{ est une densité.}$$

On en déduit, en particulier, que $\varphi_X(t)$ est bornée.

$$|\varphi_X(t)| \leq \varphi_X(0) = 1$$

5.2.2 Symétrie

A l'aide de l'intégrale de définition, on obtient facilement :

$$\varphi_X^*(t) = \varphi_X(-t)$$

En développant la fonction caractéristique en module et phase, cette relation donne :

$$\varphi_X(t) = \rho(t)e^{j\psi(t)} = \rho(-t)e^{-j\psi(-t)}, \quad \rho(t) = |\varphi_X(t)|$$

- Le module $\rho(t)$ est une fonction paire.
- La phase $\psi(t)$ est une fonction impaire.

5.2.3 Variable centrée

Soit X une variable aléatoire de moyenne m_1 . La variable centrée correspondante s'obtient en posant :

$$Y = X - m_1$$

La fonction caractéristique devient :

$$\varphi_Y(t) = E\{e^{jt(X-m_1)}\}$$

$$\varphi_Y(t) = e^{-jm_1 t} \varphi_X(t) \quad (7-5)$$

5.2.4 Formule d'inversion

Si $\varphi_X(t)$ est la fonction caractéristique d'une variable aléatoire, le théorème de réciprocity de Fourier permet de trouver la densité :

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \varphi_X(t) e^{-jxt} dt \quad (7-6)$$

Pratiquement, cela signifie que la connaissance de la fonction caractéristique suffit à déterminer une variable aléatoire.

5.2.5 Théorème fondamental

Nous donnons ici l'énoncé d'un théorème, sans démonstration, qui nous sera utile à la fin du cours, dans l'étude de la convergence en loi des variables aléatoires.

Théorème :

Soit $\varphi_{X_n}(t)$ une suite de fonctions caractéristiques tendant, lorsque n tend vers l'infini, vers une fonction caractéristique $\varphi_X(t)$.

Alors, la fonction de répartition $F_{X_n}(t)$ relative à $\varphi_{X_n}(t)$ tend vers la fonction de répartition $F_X(t)$ relative à $\varphi_X(t)$.

5.3 Exemples de fonction caractéristique

5.3.1 Exemple 1 : variable aléatoire indicatrice

Cherchons la fonction caractéristique de la variable indicatrice X définie par :

$$X = \begin{cases} 1 & \text{avec } P\{X=1\} = p \\ 0 & \text{avec } P\{X=0\} = q \end{cases} \text{ et } p + q = 1$$

L'équation (7-3) relative aux variables aléatoires discrètes donne :

$$\varphi_X(t) = \sum_{i=0}^I P_i e^{jtx_i} = qe^{jt0} + pe^{jt1}$$

$$\varphi_X(t) = pe^{jt} + q$$

5.3.2 Exemple 2 : densité exponentielle

La densité exponentielle est définie par :

$$f_X(x) = \alpha e^{-\alpha x}, \alpha > 0, x \geq 0$$

Soit, pour la fonction caractéristique :

$$\varphi_X(t) = \int_0^{+\infty} \alpha e^{-\alpha x} e^{jtx} dx$$

$$\varphi_X(t) = \varphi_{\text{exp}}(t) = \frac{\alpha}{\alpha - jt}$$

5.3.3 Exemple 3 : densité gamma pour n entier

Cette densité est une généralisation de la précédente.

$$f_X(x) = \frac{\alpha^{n+1}}{n!} x^n e^{-\alpha x}, \alpha > 0, x \geq 0$$

De plus, en posant :
$$\begin{cases} n &= 2k \\ \alpha &= \frac{I}{2} \\ x &= \chi^2 \end{cases}$$
, on retrouve la loi du χ^2 définie au chapitre 4.

Avec l'exemple précédent, nous avons trouvé :

$$\int_0^{+\infty} e^{-\alpha x} e^{jtx} dx = (\alpha - jt)^{-1}$$

En différenciant cette expression par rapport à α , on obtient :

$$\text{pour } n=1, \int x e^{-\alpha x} e^{jtx} dx = 1.(\alpha - jt)^{-2}$$

$$\text{pour } n=2, \int x^2 e^{-\alpha x} e^{jtx} dx = 1.2.(\alpha - jt)^{-3}$$

.....

$$\text{pour } n=n, \int x^n e^{-\alpha x} e^{jtx} dx = n!(\alpha - jt)^{-(n+1)}$$

et en définitive :

$$\varphi_{\gamma_n}(t) = \frac{\alpha^{n+1}}{(\alpha - jt)^{n+1}}$$

Remarquons que cette expression s'écrit encore :

$$\varphi_{\gamma_n}(t) = \left(\frac{\alpha}{\alpha - jt} \right)^{n+1} = [\varphi_{\text{exp}}(t)]^{n+1}$$

Nous reviendrons sur cette remarque dans l'étude de variable à n dimensions.

5.4 Théorème des moments

Considérons une variable aléatoire X de fonction caractéristique $\varphi_X(t)$. Supposons que $\varphi_X(t)$ soit développable en série de Mac Laurin :

$$\varphi_X(t) = \sum_{n=0}^{+\infty} \frac{\varphi_X^{(n)}(0)}{n!} t^n \quad (7-7)$$

D'autre part, la définition de la fonction caractéristique donne :

$$\varphi_X(t) = E\{e^{jtx}\} = \int_{-\infty}^{+\infty} f_X(x) e^{jtx} dx$$

Développons cette fois e^{jtx} en série entière et intégrons terme à terme, il vient :

$$\varphi_X(t) = \int f_X(x) \left[1 + jtx + \frac{(jtx)^2}{2!} + \dots + \frac{(jtx)^n}{n!} + \dots \right] dx$$

$$\varphi_X(t) = \sum_{n=0}^{+\infty} \int f_X(x) \frac{(jtx)^n}{n!} dx = \sum_{n=0}^{+\infty} \frac{(jt)^n}{n!} \int f_X(x) x^n dx$$

Or, $\int f_X(x) x^n dx$ n'est autre que le moment d'ordre n de la variable aléatoire X .

$$\varphi_X(t) = \sum_{n=0}^{+\infty} \frac{m_n}{n!} (jt)^n \quad (7-8)$$

En identifiant (7-7) et (7-8), on obtient le théorème des moments :

$$\left. \frac{d^n \varphi_X(t)}{dt^n} \right|_{t=0} = j^n m_n \quad (7-9)$$

La valeur à l'origine de la dérivée nième de la fonction caractéristique est égale au moment d'ordre n de la variable aléatoire multiplié par j^n .

Ce résultat est valable quel que soit n :

- si le développement de la fonction caractéristique en série est convergent,
- **et** si tous les moments existent.

Il est valable à l'ordre $n=N$ si les moments m_n existent pour $n \leq N$. On obtient ce résultat par dérivation sous le signe somme.

Application : Déterminer la moyenne et l'écart type d'une variable aléatoire distribuée suivant une loi binomiale.

Loi binomiale : $P\{X = k\} = C_n^k p^k q^{n-k}$, avec $0 \leq X \leq n$, $p + q = 1$

La fonction caractéristique s'écrit : $\varphi_X(t) = (pe^{jt} + q)^n$

D'où : $\varphi_X'(t) = n(pe^{jt} + q)^{n-1} \cdot jpe^{jt}$

$\varphi_X'(0) = jnp = jm_1$ d'après (7-9) et $m_1 = np$.

De même, pour le moment du second ordre :

$$\varphi_X''(t) = jnp[(n-1)(pe^{jt} + q)^{n-2} \cdot jpe^{2jt} + (pe^{jt} + q)^{n-1} je^{jt}]$$

$$\varphi_X''(0) = j^2 np[(n-1)p + 1]$$

$$\varphi_X''(0) = j^2 [n^2 p^2 + npq] = j^2 m_2, \text{ d'après (7-9).}$$

$$m_2 = n^2 p^2 + npq$$

Donc la variance $\text{Var}\{X\} = \sigma_X^2 = m_2 - m_1^2 = npq$ et l'écart type : $\sigma_X = \sqrt{npq}$

5.5 Exercices

5.5.1 Énoncés

1)

- a. Déterminer la fonction caractéristique d'une variable aléatoire répartie suivant une loi de Poisson de paramètre λ .

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad X = 0, 1, \dots$$

- b. Utiliser $\varphi_X(t)$ pour calculer l'espérance et l'écart type de X .

2) X est une variable aléatoire continue admettant pour fonction caractéristique : $\varphi_X(t) = e^{-a|t|}$.

- a. Déterminer la densité de X .
- b. Peut-on utiliser $\varphi_X(t)$ pour calculer les moments de X ?

3)

- a. Calculer la fonction caractéristique $\varphi_{X_1}(t)$ de la v.a. X_1 uniformément répartie sur l'intervalle $[a - c; a + c]$.

- b. Même question pour la v.a. X_2 de densité $f_{X_2}(x) = \frac{-|x| + b}{b^2}, -b \leq x \leq b$.
 $= 0$, ailleurs

- c. Retrouver $\varphi_{X_2}(t)$ à partir de $\varphi_{X_1}(t)$ et de considérations sur les densités de X_1 et de X_2 .

5.5.2 Solutions

1)

a. $\varphi_X(t) = e^{\lambda(e^{it} - 1)}$

b. $m_1 = \lambda, m_2 = \lambda(\lambda + 1), \sigma = \sqrt{\lambda}$

2)

a. $f_X(x) = \frac{a}{\pi} \frac{1}{a^2 + x^2}$, loi de Cauchy

b. les moments de $f_X(x)$ n'existent pas et, de plus, $\varphi_X(t)$ n'est pas dérivable à l'origine.

3)

a. $\varphi_{X_1}(t) = \frac{\sin(ct)}{ct} e^{jat}$

b. $\varphi_{X_2}(t) = \left(\frac{\sin\left(\frac{bt}{2}\right)}{\frac{bt}{2}} \right)^2$

c. en posant $a=0$ et $b=2c$ on trouve que la densité de X_2 se déduit de celle de X_1 par convolution (notée \otimes).

$$f_{X_2}(x_2) = f_{X_1}(x_1) \otimes f_{X_1}(x_1), \text{ d'où } \varphi_{X_2}(t) = \varphi_{X_1}(t) \varphi_{X_1}(t) = \left(\frac{\sin(ct)}{ct} \right)^2$$

CHAPITRE 6. Variables aléatoires réelles bidimensionnelles

6.1 Généralités

Nous avons étudié, jusqu'à présent, le cas simple d'une seule variable aléatoire ou ce que l'on peut encore appeler une variable à une dimension.

Nous avons vu que cette variable était entièrement définie par la donnée d'une des fonctions suivantes :

- $F_X(x)$ sa fonction de répartition
- $f_X(x)$ sa fonction de densité
- $\varphi_X(t)$ sa fonction caractéristique

Cette variable peut encore être décrite par la donnée de paramètres certains : les moments, en supposant qu'ils existent.

Enfin, nous avons envisagé la création de nouvelles variables par l'intermédiaire de fonctions certaines : $Y = g(X)$

Toutes ces notions vont être reprises ici pour deux variables aléatoires, nous allons donc substituer, à la variable X , le couple ordonné (X, Y) . Ceci va constituer le cas bidimensionnel.

6.2 Couple de variables discrètes

Pour bien mettre en évidence l'extension que permet l'introduction d'une deuxième dimension, nous commençons par le cas discret en traitant un exemple.

6.2.1 Introduction

L'expérience consiste à jeter deux dés de couleur différente. Le premier dé est bleu, le second est rouge.

Les résultats possibles sont identifiés, dans Ω , par les chiffres apparaissant sur la face supérieure de chaque dé.

$$\omega_n = (\text{face du dé rouge, face du dé bleu})$$

$$\omega_n = (k, l) \quad k, l = 1, 2, \dots, 6$$

Il y a 36 résultats élémentaires ω_n possibles. Si les dés sont identiques et symétriques, on choisit de leur attribuer la probabilité :

$$P\{\omega_n\} = \frac{1}{36}, \quad n = 1 \text{ à } 36.$$

Définition des variables aléatoires :

Pour définir notre couple de variables aléatoires, il faut se donner deux applications de Ω dans l'ensemble des réels. Nous prendrons :

$$X(\omega_n) = X(k, l) = k = x_i$$

$$Y(\omega_n) = Y(k, l) = k + l = y_j$$

Soulignons tout de suite que :

- la variable X peut prendre les valeurs : $x_i = 1, 2, \dots, 6$ pour $i=1$ à 6
- la variable Y peut prendre les valeurs : $y_j = 2, 3, \dots, 12$ pour $j=1$ à 11

Mais le couple ordonné (X, Y) ne peut pas prendre les valeurs de n'importe quelle combinaison (x_i, y_j) . Par exemple les couples $(4, 3)$ ou $(2, 10)$ ne seront jamais rencontrés au cours d'une épreuve comme le montre le diagramme suivant :

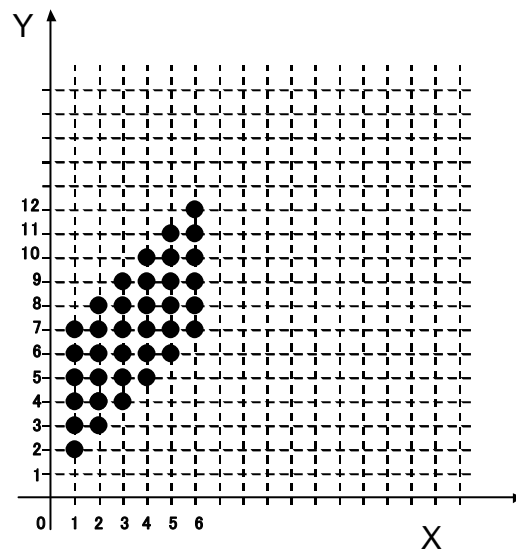


Figure 2 Espace de définition du couple (X,Y)

Tableau de contingence :

Il nous reste à déterminer la probabilité affectée à chaque couple possible.

$$P\{X = x_i \text{ et } Y = y_j\} = P\{x_i, y_j\} = p_{i,j}$$

Il suffit, pour déterminer $p_{i,j}$ de remarquer que l'ensemble des deux applications choisies établit une relation biunivoque (bijection) entre les résultats (ω_n) et les couples (x_i, y_j) .

$$\text{Donc : } p_{i,j} = P\{\omega_n\} = \frac{1}{36}.$$

On peut retrouver ce résultat en utilisant les probabilités composées :

$$P\{X = x_i \text{ et } Y = y_j\} = P\{Y = y_j | X = x_i\} P\{X = x_i\}$$

Or : $P\{X = x_i\} = P\{X = k\} = \frac{1}{6}$. D'autre part, quand $X = k$, la probabilité pour que Y prenne

la valeur $y_j = k + l$ est égale à la probabilité d'obtenir l sur le second dé, c'est-à-dire $\frac{1}{6}$.

D'où le résultat.

Nous pouvons maintenant dresser un tableau définissant la loi de probabilité du couple (X, Y). C'est ce qu'on appelle un tableau de contingence représenté ci-dessous.

Il est bien entendu qu'en tous les points du plan (X, Y) où $p_{i,j}$ n'est pas défini, on a $p_{i,j} = 0$ (événements impossibles).

X	x ₁ =1	x ₂ = 2	x ₃ =3	x ₄ =4	x ₅ =5	x ₆ =6	PY _j
Y							
y ₁ =2	1/36	0	0	0	0	0	1/36

y₂=3	1/36	1/36	0	0	0	0	2/36
y₃=4	1/36	1/36	1/36	0	0	0	3/36
y₄=5	1/36	1/36	1/36	1/36	0	0	4/36
y₅=6	1/36	1/36	1/36	1/36	1/36	0	5/36
y₆=7	1/36	1/36	1/36	1/36	1/36	1/36	6/36
y₇=8	0	1/36	1/36	1/36	1/36	1/36	5/36
y₈=9	0	0	1/36	1/36	1/36	1/36	4/36
y₉=10	0	0	0	1/36	1/36	1/36	3/36
y₁₀=11	0	0	0	0	1/36	1/36	2/36
y₁₁=12	0	0	0	0	0	1/36	1/36
PX_i	6/36	6/36	6/36	6/36	6/36	6/36	1

Tableau de contingence

Loi de probabilité pour le couple (X, Y)

L'examen du tableau va nous permettre de faire apparaître trois types de lois de probabilités.

1) La loi de probabilité du couple (X, Y)

C'est la loi qui correspond aux différents éléments du tableau. Par exemple :

$$P\{X = x_1 \text{ et } Y = y_1\} = P\{X = 1 \text{ et } Y = 1\} = p_{1,1} = \frac{1}{36}$$

$$P\{X = x_2 \text{ et } Y = y_1\} = P\{X = 2 \text{ et } Y = 1\} = p_{2,1} = 0$$

2) Les deux lois marginales

Loi marginale en X :

On considère la première colonne du tableau et l'on somme toutes les probabilités dans cette colonne. On obtient ainsi la probabilité d'avoir $X = x_1$ quel que soit Y. En effet :

$$P(X = x_1) = P((X = x_1 \text{ et } Y = y_1) \text{ ou } (x_1, y_2) \text{ ou } \dots (x_1, y_{11}))$$

Mais ces événements sont disjoints et l'axiome d'additivité donne :

$$P(X = x_1) = PX_1 = \sum_j p_{1,j} = \frac{6}{36}.$$

En étendant cette opération à toutes les colonnes, c'est-à-dire à toutes les valeurs de X possibles, on engendre la loi de probabilité marginale de X. Dans le cas présent, cette loi est uniforme :

$$\begin{aligned} PX_i &= \frac{1}{6} & i = 1 \text{ à } 6 \\ PX_i &= 0 & \text{ailleurs} \end{aligned}$$

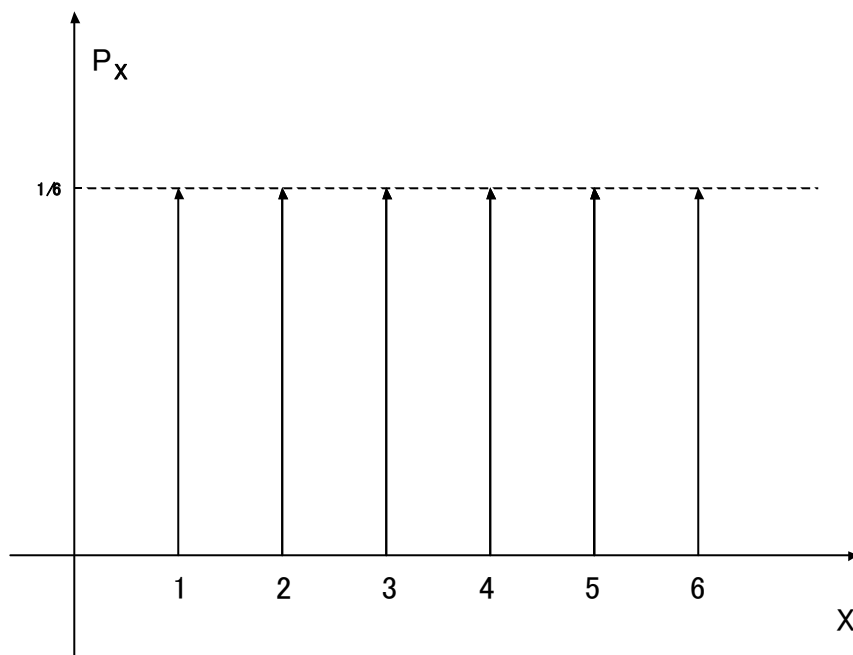


Figure 4.4 Loi de la variable X

Loi marginale en Y :

On effectue la même opération en considérant cette fois chaque ligne du tableau. Par exemple :

$$P\{Y = y_1\} = PY_1 = \sum_i p_{i,1} = p_{1,1} = \frac{1}{36}$$

$$P\{Y = y_6\} = PY_7 = \sum_i p_{i,6} = \frac{6}{36}$$

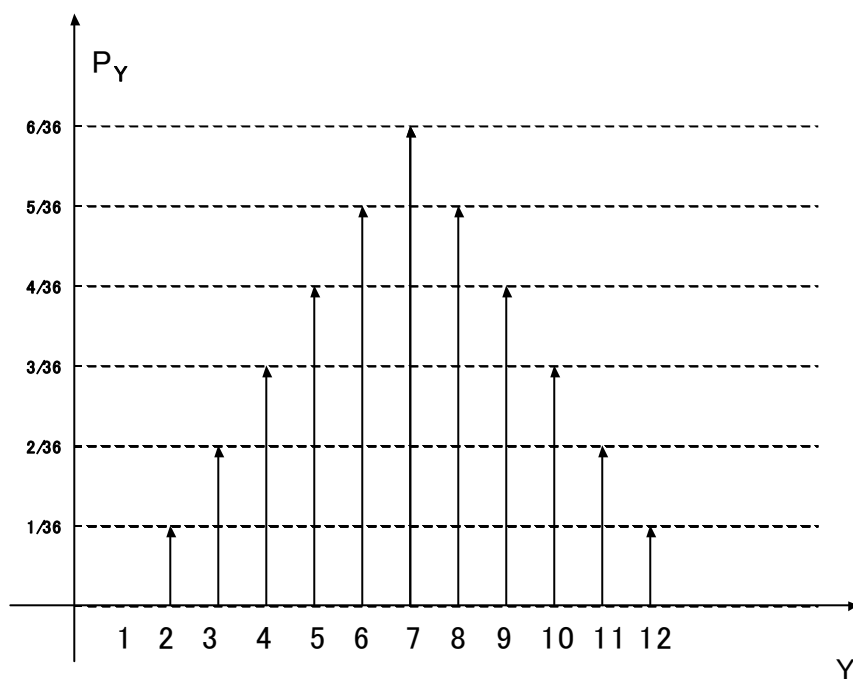


Figure 3

Loi de la variable Y

La loi marginale en Y a une forme triangulaire.

3) Les lois liées ou conditionnelles

On se fixe une valeur de X , par exemple $X = x_3 = 3$, et on s'intéresse à la probabilité conditionnelle de Y quand X a la valeur 3.

Pour chaque élément de la colonne considérée, on peut écrire, par définition :

$$P\{Y = y_j | X = 3\} = \frac{P\{Y = y_j \text{ et } X = 3\}}{P\{X = 3\}}$$

Or, la probabilité d'avoir $X = 3$ est donnée par la loi marginale en X :

$$P\{X = 3\} = PX_3 = \frac{1}{6}$$

Pour chaque terme, on obtient :

$$P\{Y = y_j | X = 3\} = \frac{p_{3,j}}{PX_3} = 6 p_{3,j}$$

On trouve ainsi :

$$P\{Y = y_j | X = 3\} = \frac{1}{6} \quad \text{de } j = 3 \text{ à } 8 \\ = 0 \quad \text{ailleurs}$$

Le raisonnement peut s'étendre à toutes les valeurs de X ; ce qui définit les lois de Y conditionnelles à X .

On définit, de la même façon, les lois de X conditionnelles à Y . Par exemple, la loi de X conditionnelle à $Y = y_3 = 4$ s'écrit :

$$P\{X | Y = 4\} = \frac{\frac{1}{36}}{\frac{36}{36}} = \frac{1}{3} \quad \text{pour } X = 1, 2, 3 \\ = 0 \quad \text{ailleurs}$$

6.2.2 Définitions pour une variable discrète

Une variable aléatoire discrète bidimensionnelle est définie par le couple ordonné (X, Y) où X et Y sont deux variables aléatoires simples pouvant prendre les valeurs :

$$X = x_i, \quad i = 1, 2, \dots$$

$$Y = y_j, \quad j = 1, 2, \dots$$

6.2.3 Loi de probabilité du couple

C'est la loi bidimensionnelle, elle est définie par :

$P\{X = x_i \text{ et } Y = y_j\} = p_{i,j}$	(8- 1)
--	--------

où $p_{i,j}$ représente la probabilité d'avoir simultanément $X = x_i$ et $Y = y_j$.

Pour respecter les axiomes, le nombre $p_{i,j}$ doit satisfaire aux conditions :

$0 \leq p_{i,j} \leq 1$ $\sum_i \sum_j p_{i,j} = 1$	(8- 2)
--	--------

En sommant les $p_{i,j}$ relatifs aux valeurs x_i, y_j telle que $X < x$ et $Y < y$ on obtient la fonction de répartition bidimensionnelle.

$$F_{X,Y}(x, y) = P\{X < x, Y < y\} = \sum_{X < x} \sum_{Y < y} p_{i,j}$$

6.2.4 Lois marginales

Loi marginale en X :

On considère l'événement $\{X=x_i\}$. Cet événement est réalisé avec les couples :

$$\{X = x_i\} = \{(x_i, y_1) \text{ ou } (x_i, y_2) \dots, (x_i, y_j) \dots\} = \bigcup_{j=1}^{+\infty} \{(x_i, y_j)\}$$

Or, tous ces événements sont disjoints, donc d'après l'axiome 3 :

$$P\{X = x_i\} = \sum_j P(X = x_i \text{ et } Y = y_j)$$

$P\{X = x_i\} = \sum_j p_{i,j} = PX_i$	(8- 3)
--	--------

La correspondance $(X = x_i, PX_i)$ définit la loi de probabilité marginale de X . D'après les propriétés de $p_{i,j}$, on a :

$$0 \leq PX_i \leq 1$$

$$\sum_i PX_i = 1$$

La répartition marginale est dans ce cas identique à celle d'une variable à une dimension :

$$F_X(x) = \sum_{x_i < x} PX_i$$

Loi marginale en Y :

La loi marginale en Y se définit de la même façon par la correspondance : (y_j, PY_j)

$P\{Y = y_j\} = \sum_i p_{i,j} = PY_j$ $F_Y(y) = \sum_{y_j < y} PY_j$	(8- 4)
---	--------

6.2.5 Lois conditionnelles ou lois liées

Contrairement aux lois marginales qui ne sont que deux, on peut définir autant de lois conditionnelles qu'il y a de valeurs possibles pour les variables X ou Y .

Loi conditionnelle de X à Y fixé :

C'est la probabilité de la variable aléatoire X en supposant que Y a une valeur figée y_j .

D'après la définition de la probabilité conditionnelle de deux événements, on peut écrire :

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i \text{ et } Y = y_j)}{P(Y = y_j)}, \text{ soit :}$$

$$P(X = x_i | Y = y_j) = \frac{p_{i,j}}{\sum_i p_{i,j}} = \frac{p_{i,j}}{PY_j} \quad (8-5)$$

La correspondance $\left(x_i(y_j), \frac{p_{i,j}}{PY_j}\right)$ définit la loi de probabilité de la variable X conditionnelle à l'événement $Y = y_j$.

Remarque :

L'expression précédente fait bien apparaître le principe qui préside à la définition d'une probabilité conditionnelle.

En effet, si l'on suppose que Y a déjà, par exemple, la valeur y_1 , les seuls événements possibles sont les couples (x_i, y_1) pour lesquels toutes les valeurs x_i sont associées à l'unique valeur y_1 .

Dans ces conditions, l'événement certain est représenté par la réunion de ces couples exclusivement. Il est alors nécessaire de normaliser les probabilités $p_{i,1}$ pour respecter l'axiome 2 qui dit que la probabilité de l'événement certain doit être égale à 1. C'est précisément ce qu'exprime la formule (8-5).

Loi conditionnelle de Y à X fixé :

On obtiendrait de la même façon :

$$P(Y = y_j | X = x_i) = \frac{p_{i,j}}{\sum_j p_{i,j}} = \frac{p_{i,j}}{PX_i} \quad (8-6)$$

La correspondance $\left(y_j(x_i), \frac{p_{i,j}}{PX_i}\right)$ définit la loi de la probabilité de la variable Y conditionnelle à l'événement $X = x_i$.

Pour ces deux derniers cas, les lois de répartition conditionnelles deviennent :

$$F_X(x | Y = y_j) = \frac{\sum_{x_i < x} p_{i,j}}{PY_j}$$

$$F_Y(y | X = x_i) = \frac{\sum_{y_j < y} p_{i,j}}{PX_i}$$

6.2.6 Indépendance

Les deux variables X et Y sont indépendantes si quelles que soient i, j :

$$P\{X = x_i \text{ et } Y = y_j\} = P\{X = x_i\}P\{Y = y_j\}$$

$$p_{i,j} = PX_i \cdot PY_j, \quad \forall i, j \quad (8-7)$$

Autrement dit, la probabilité du couple (X, Y) est égale au produit des probabilités marginales.

Dans l'exemple du jeu de dés, les variables X et Y n'étaient pas indépendantes car la relation (8- 7) n'était pas vérifiée pour toutes les valeurs possibles de i et j .

Lorsque les variables sont indépendantes les lois conditionnelles s'identifient aux deux lois marginales.

$$P\{X = x_i | Y = y_j\} = PX_i, \forall y_j$$

$$P\{Y = y_j | X = x_i\} = PY_j, \forall x_i$$

6.3 Variables aléatoires réelles bidimensionnelle continue

Dans ce paragraphe, on considère un couple (X, Y) de deux variables aléatoires réelles. On va définir les lois bidimensionnelles associées.

6.3.1 Fonction de répartition bidimensionnelle

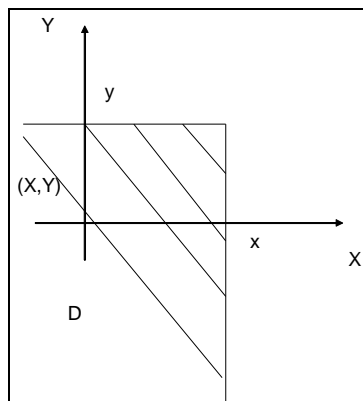
Pour les variables aléatoires scalaires, « monodimensionnelles », la fonction de répartition est définie par la probabilité de l'événement $\{X < x\}$:

$$F_X(x) = P(\{X < x\}).$$

De la même manière on définira la fonction de répartition du couple de variables aléatoires (X, Y) par la probabilité de l'intersection des événements $\{X < x\}$ et $\{Y < y\}$:

$F_{X,Y}(x, y) = P(\{X < x\} \cap \{Y < y\}) = P(X < x \text{ et } Y < y)$	9-1
--	-----

Si nous utilisons un système de coordonnées cartésiennes pour représenter les variables aléatoires X et Y , la fonction de répartition $F_{X,Y}(x, y)$ représente alors la probabilité pour le couple (X, Y) (ou vecteur à deux composantes) d'appartenir au domaine D représenté ci-dessous :

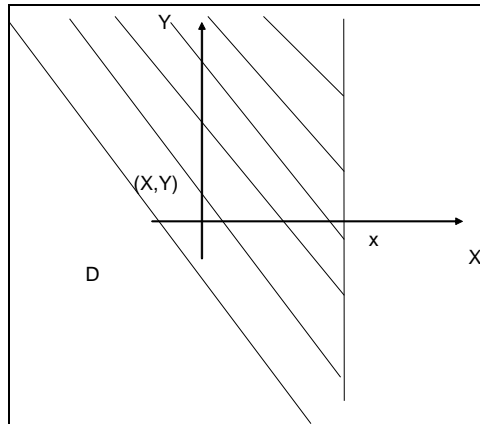


6.3.2 Fonctions de répartition marginales

La fonction de répartition en X est définie par la probabilité de l'événement $\{X < x\}$ quelle que soit la valeur prise par Y , ce qui correspond à :

$F_X(x) = P(\{X < x\}) = P(\{X < x \text{ et } Y < +\infty\}) = F_{X,Y}(x, +\infty)$	9-2
--	-----

C'est la probabilité du domaine hachuré la figure suivante :

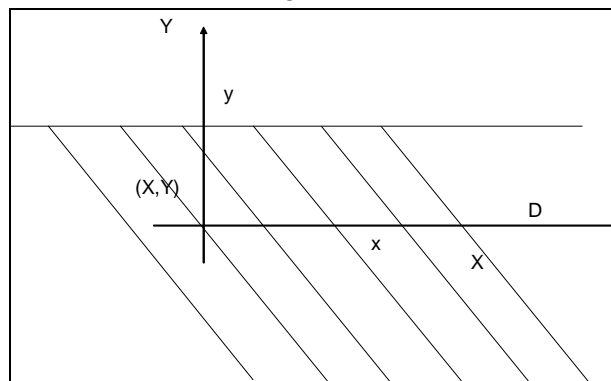


La fonction de répartition en Y se définit de la même manière. C'est la probabilité pour le couple (X, Y) d'appartenir au domaine D situé en dessous la droite d'ordonnée $Y = y$

$$F_Y(y) = F_{X,Y}(+\infty, y)$$

9-3

C'est la probabilité du domaine hachuré la figure suivante :



6.3.3 Valeurs asymptotiques de la fonction de répartition

En considérant la définition de la fonction de répartition on établit les relations suivantes :

$$F_{X,Y}(x, -\infty) = 0$$

$$F_{X,Y}(-\infty, y) = 0$$

$$F_{X,Y}(+\infty, +\infty) = 1$$

9-4

Éléments de démonstration :

- On peut en effet écrire :

$$\{X < x \text{ et } Y < y\} = \{X < x\} \cap \{Y < y\} \subset \{Y < y\}$$

et donc :

$$0 \leq F_{X,Y}(x, y) \leq F_Y(y)$$

- Lorsque y tend vers $-\infty$ on a $F_Y(-\infty) = 0$ par définition d'une variable aléatoire, d'où le résultat : $F_{X,Y}(x, -\infty) = 0$

- On a vu que $F_Y(y) = F_{X,Y}(+\infty, y)$, lorsque y tend vers $+\infty$, $F_{X,Y}(+\infty, +\infty) = F_Y(+\infty) = 1$

6.3.4 Calcul de la probabilité d'un rectangle

Calcul de la probabilité : $P(\{x_1 \leq X < x_2, Y < y\})$

En considérant les sous-ensembles dans le plan (X, Y) , on peut écrire :

$$\{X < x_2, Y < y\} = \{X < x_1, Y < y\} + \{x_1 \leq X < x_2, Y < y\}$$

Comme les sous-ensembles du second membre sont disjoints, en appliquant les axiomes 1 et 3, on obtient :

$$P(\{x_1 \leq X < x_2, Y < y\}) = F_{X,Y}(x_2, y) - F_{X,Y}(x_1, y) \geq 0 \quad 9-5$$

On en déduit également : $F_{X,Y}(x_2, y) \geq F_{X,Y}(x_1, y)$, pour $x_2 \geq x_1$

Cette expression montre que la fonction de répartition est monotone non décroissante en x.

Calcul de la probabilité : $P(\{X < x, y_1 \leq Y < y_2\})$

On trouve de la même manière :

$$P(\{X < x, y_1 \leq Y < y_2\}) = F_{X,Y}(x, y_2) - F_{X,Y}(x, y_1) \geq 0 \quad 9-6$$

On en déduit également : $F_{X,Y}(x, y_2) \geq F_{X,Y}(x, y_1)$, pour $y_2 \geq y_1$

Cette expression montre que la fonction de répartition est monotone non décroissante en y

Calcul de la probabilité : $P(\{x_1 \leq X < x_2, y_1 \leq Y < y_2\})$

On cherche à exprimer en fonction de $F_{X,Y}(x, y)$ la probabilité du couple (X, Y) d'appartenir au domaine : $D : \{X \in [x_1, x_2[\text{ et } Y \in [y_1, y_2[\}$

Considérons l'événement :

$$\{X < x_2, Y < y_2\} = \{X < x_2, y_1 \leq Y < y_2\} + \{X < x_2, Y < y_1\}$$

Le premier terme du second membre se développe comme :

$$\{X < x_2, y_1 \leq Y < y_2\} = \{x_1 \leq X < x_2, y_1 \leq Y < y_2\} + \{X < x_1, y_1 \leq Y < y_2\}$$

Soit en probabilité d'événements disjoints :

$$F_{X,Y}(x_2, y_2) = P(\{X < x_2, y_1 \leq Y < y_2\}) + F_{X,Y}(x_2, y_1)$$

$$F_{X,Y}(x_2, y_2) = P(\{x_1 \leq X < x_2, y_1 \leq Y < y_2\}) + P(\{X < x_1, y_1 \leq Y < y_2\}) + F_{X,Y}(x_2, y_1)$$

$$F_{X,Y}(x_2, y_2) = P(\{x_1 \leq X < x_2, y_1 \leq Y < y_2\}) + F_{X,Y}(x_1, y_2) - F_{X,Y}(x_1, y_1) + F_{X,Y}(x_2, y_1)$$

$$P(\{x_1 \leq X < x_2, y_1 \leq Y < y_2\}) = F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1) + F_{X,Y}(x_1, y_1) \quad 9-7$$

6.3.5 Fonction de densité bidimensionnelle

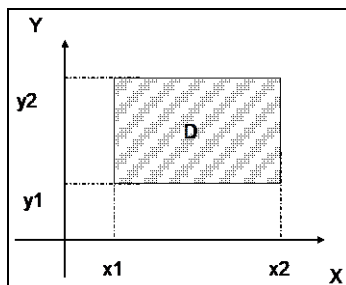
La densité bidimensionnelle s'obtient en différentiant deux fois la fonction de répartition par rapport à x, puis à y (on suppose que cette différentielle existe).

$$f_{X,Y}(x, y) = \frac{\partial}{\partial y} \left(\frac{\partial F_{X,Y}(x, y)}{\partial x} \right) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial y \partial x} \quad 9-8$$

6.3.6 Interprétation infinitésimale de la densité de probabilité

Intégrons cette densité sur le domaine rectangulaire $D: \{[x_1, x_2][y_1, y_2]\}$

$$I = \iint_D f(x, y) dx dy = \iint_D \frac{\partial^2 F_{X,Y}(x, y)}{\partial y \partial x} dx dy \quad 9-9$$



En intégrant, par exemple, par rapport à y puis par rapport à x, il vient :

$$I = \int_{x_1}^{x_2} \left[\frac{\partial F_{X,Y}(x, y)}{\partial x} \right]_{y_1}^{y_2} dx = \int_{x_1}^{x_2} \left(\frac{\partial F_{X,Y}(x, y_2)}{\partial x} - \frac{\partial F_{X,Y}(x, y_1)}{\partial x} \right) dx$$

$$I = [F_{X,Y}(x, y_2)]_{x_1}^{x_2} - [F_{X,Y}(x, y_1)]_{x_1}^{x_2} = F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1) + F_{X,Y}(x_1, y_1)$$

$$\int_{x_1}^{x_2} \int_{y_1}^{y_2} f_{X,Y}(x, y) dx dy = F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1) + F_{X,Y}(x_1, y_1) \quad 9-10$$

Le second membre de cette équation est égal à $P(\{x_1 \leq X < x_2, y_1 \leq Y < y_2\}) \leq 0$.

Si nous posons :

$$x_1 = x, x_2 = x + dx$$

$$y_1 = y, y_2 = y + dy$$

Dans le domaine infinitésimal $dS = dx dy$, la densité $f_{X,Y}(x, y)$ peut être considérée comme constante et on peut écrire :

$$P(\{x \leq X < x + dx, y \leq Y < y + dy\}) = f_{X,Y}(x, y) dx dy \quad 9-11$$

Ce qui signifie que la probabilité du couple de variable (X, Y) d'appartenir à l'élément de surface dS au voisinage du point de coordonnées (x, y) est proportionnelle à la valeur de la fonction de densité bidimensionnelle en ce point.

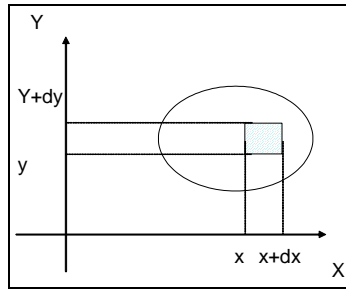
Remarquons, en particulier, que pour une variable aléatoire continue, la probabilité d'avoir $\{X = x, Y = y\}$ est nulle car dS est nulle

6.3.7 Utilisation de la densité

L'équation 9-11 permet de calculer la probabilité pour une variable aléatoire d'appartenir à un domaine D quelconque.

Il suffit en effet de décomposer D en éléments de surface dS disjoints et d'appliquer l'axiome 3 du chapitre 1. On trouve ainsi :

$$P(\{(X, Y) \in D\}) = \int_D f_{X,Y}(x, y) dx dy \quad 9-12$$



6.3.8 Volume sous la densité

Posons :

$$z = f_{X,Y}(x, y)$$

Cette équation définit une surface dans une représentation cartésienne à 3 dimensions. L'intégrale double 9-12 représente alors le volume situé sous la densité.

6.3.9 Relation entre la fonction de répartition et la densité bidimensionnelle

Reprenons l'équation 9-10 et posons $x_1 = -\infty, y_1 = -\infty$, d'après les propriétés de la fonction de répartition, on a :

$$F_{X,Y}(x, -\infty) = 0, F_{X,Y}(-\infty, y) = 0, F_{X,Y}(-\infty, -\infty) = 0,$$

Ce qui nous donne :

$$\int_{-\infty}^{x_2} \int_{-\infty}^{y_2} f_{X,Y}(x, y) dx dy = F_{X,Y}(x_2, y_2) \quad 9-13$$

On retrouve la fonction de répartition.

Dans 9-13, posons $x_2 = +\infty, y_2 = +\infty$, il vient :

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx dy = F_{X,Y}(+\infty, +\infty) = 1 \quad 9-14$$

Cette expression nous montre que le volume sous la densité est unitaire. De plus, la fonction $f_{X,Y}(x, y)$ est toujours positive ou nulle car la fonction de répartition est une fonction monotone non décroissante en x et y .

Réciproquement, si une fonction $f_{X,Y}(x, y)$ satisfait ces deux conditions, elle peut être considérée comme représentative d'une densité de probabilité à 2 dimensions.

6.3.10 Densité marginale

Pour obtenir les densités marginales, deux méthodes sont possibles.

- Soit on dérive les fonctions de répartition marginales définies en 9-2, 9-3 :

$$\frac{dF_X(x)}{dx} = \frac{dF_{X,Y}(x, +\infty)}{dx} = f_X(x)$$

$$\frac{dF_Y(y)}{dy} = \frac{dF_{X,Y}(+\infty, y)}{dy} = f_Y(y)$$

- Soit on intègre la densité bidimensionnelle de manière à éliminer la variable non désirée . Pour le démontrer revenons à la relation, très utile et à connaître, entre probabilité et densité de probabilité

$$f_X(x)dx = P\{x \leq X < x+dx\} = P\{x \leq X < x+dx \text{ et } Y < +\infty\} = \int_x^{x+dx} \int_{y=-\infty}^{+\infty} f_{X,Y}(u,y) du dy$$

Sur l'intervalle différentiel x , $x+dx$ l'expression sous le signe somme peut être considérée comme constante et on trouve :

$$f_X(x)dx = \left[\int_{y=-\infty}^{+\infty} f_{X,Y}(u,y) du dy \right] dx$$

En définitive, les densités marginales peuvent être indifféremment définies par les expressions :

$$\frac{dF_{X,Y}(x,+\infty)}{dx} = f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y) dy \quad 9-15$$

$$\frac{dF_{X,Y}(+\infty,y)}{dy} = f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y) dx \quad 9-16$$

Exemple : On considère la fonction

$$f_{X,Y}(x,y) = \begin{cases} kxe^{-y} & 0 \leq x \leq 1 \text{ et } y \geq 0 \\ 0 & \text{ailleurs} \end{cases}$$

Cette fonction étant une fonction positive ($k > 0$), elle peut représenter une densité de probabilité. Pour qu'elle soit une densité, il faut que son intégrale sur son domaine de définition soit égale à 1. La fonction de répartition est égale à :

$$F_{X,Y}(x,y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u,v) du dv = k \int_{-\infty}^x \int_{-\infty}^y ue^{-v} du dv$$

En tenant compte des bornes du domaine, on a :

$$F_{X,Y}(x,y) = \begin{cases} \int_{u=0}^x \int_{v=0}^y ue^{-v} du dv = \int_{u=0}^x u du \int_{v=0}^y e^{-v} dv = K \frac{x^2}{2} (1 - e^{-y}) \end{cases}$$

$$F_{X,Y}(+\infty,+\infty) = F_{X,Y}(1,+\infty) = \frac{K}{2} = 1,$$

La constante K vaut 2. En résumé :

$$f_{X,Y}(x,y) = \begin{cases} 2xe^{-y} & \text{pour } x \in [0,1], y \in [0,+\infty[\\ 0 & \text{ailleurs} \end{cases}$$

$$F_{X,Y}(x,y) = \begin{cases} x^2(1 - e^{-y}) & \text{pour } x \in [0,1], y \in [0,+\infty[\\ 0 & \text{ailleurs} \end{cases}$$

On vérifiera en utilisant les différentes méthodes exposées précédemment que l'on trouve :

$$F_X(x) = F_{X,Y}(x,+\infty) = x^2 \quad f_X(x) = 2x \text{ sans oublier } 0 \leq x \leq 1$$

$$F_Y(y) = F_{X,Y}(1,y) = 1 - e^{-y} \quad f_Y(y) = e^{-y} \text{ sans oublier } y \geq 0$$

6.3.11 Lois conditionnelles

Considérons les deux événements A et M suivants :

$$A = \{Y < y\}, \quad M = \{x_1 \leq X < x_2\}$$

La probabilité conditionnelle $p(A|M)$ représente la probabilité d'avoir $\{Y < y\}$, sachant que X appartient à l'intervalle $[x_1, x_2[$. Elle est donnée par :

$$P(\{A|M\}) = \frac{P(A \cdot M)}{P(M)}$$

D'après 9-5, on a : $P(A \cdot M) = P(\{x_1 \leq X < x_2, Y < y\}) = F_{X,Y}(x_2, y) - F_{X,Y}(x_1, y)$

D'après 9-3, on a : $P(\{M\}) = P(\{x_1 \leq X < x_2\}) = F_X(x_2) - F_X(x_1)$

On en déduit la fonction de répartition conditionnelle en Y :

$$F_Y(y|x_1 \leq X < x_2) = P(\{A|M\}) = \frac{F_{X,Y}(x_2, y) - F_{X,Y}(x_1, y)}{F_X(x_2) - F_X(x_1)} \quad 9-17$$

Le numérateur de cette équation peut s'écrire en fonction de la densité de probabilité :

$$F_{X,Y}(x_2, y) - F_{X,Y}(x_1, y) = \int_{-\infty}^y \int_{x_1}^{x_2} f(u, v) du dv$$

En dérivant par rapport à y, on obtient la densité conditionnelle :

$$f_Y(y|x_1 \leq X < x_2) = \frac{dF_Y(y|x_1 \leq X < x_2)}{dy} = \frac{\int_{x_1}^{x_2} f_{X,Y}(x, y) dx}{F_X(x_2) - F_X(x_1)} \quad 9-18$$

Posons que : $x_1 = x, x_2 = x + \Delta x$. En supposant Δx petit, on peut écrire :

$$f_Y(y|x \leq X < x + \Delta x) \approx \frac{f_{X,Y}(x, y) \Delta x}{\Delta x \cdot \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy}$$

Faisons tendre Δx vers 0, il vient :

$$f_Y(y|X = x) = \frac{f_{X,Y}(x, y)}{\int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy} = \frac{f_{X,Y}(x, y)}{f_X(x)} \quad 9-19$$

Expression qui définit la densité conditionnelle de Y à X fixé.

On obtient de même l'expression qui définit la densité conditionnelle de X à Y fixé :

$$f_X(x|Y = y) = \frac{f_{X,Y}(x, y)}{\int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx} = \frac{f_{X,Y}(x, y)}{f_Y(y)} \quad 9-20$$

6.3.12 Variable aléatoires indépendantes

On dit que les variables aléatoires X et Y sont indépendantes si les événements $\{X < x\}$ et $\{Y < y\}$ sont indépendants, soit :

$$P(X < x \text{ et } Y < y) = P(X < x)P(Y < y)$$

Cette relation s'exprime avec les fonctions de répartition :

$$F_{X,Y}(x, y) = F_X(x)F_Y(y)$$

En dérivant par rapport à x et à y , on trouve pour la densité :

$f_{X,Y}(x, y) = f_X(x)f_Y(y)$	9-21
--------------------------------	------

Si deux variables aléatoires simples sont indépendantes alors leur densité de probabilité bidimensionnelle est égale au produit de leurs densités marginales

6.4 Variables aléatoires gaussiennes

6.4.1 Définition

Une variable aléatoire (X, Y) est gaussienne si sa densité de probabilité est de la forme :

$$f_{X,Y}(x, y) = e^{-\varphi(x, y)} \quad \varphi(x, y) = ax^2 + bxy + cy^2 + dx + ey + K$$

Elle ne dépend que de 5 paramètres, car la constante K est fixée par la condition de normalisation :

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx dy = 1$$

Il est d'usage de représenter cette densité sous la forme pratique suivante :

$f_{X,Y}(x, y) = A e^{-\frac{1}{2(1-r^2)}\varphi(x, y)}$ $A = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}$ $\varphi(x, y) = \frac{(x-m_1)^2}{\sigma_1^2} - \frac{2r(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2}$	9-22
---	------

Les 5 paramètres sont, dans ce cas m_1 , σ_1 , m_2 , σ_2 et r .

Les densités de probabilités marginales sont obtenues en intégrant la densité de probabilité conjointe (loi bidimensionnelle du couple) par rapport à x ou y . Par exemple, la loi marginal de Y est donnée par :

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx$$

Faisons le changement de variable :

$$u = \frac{x-m_1}{\sigma_1}, du = \frac{dx}{\sigma_1} \quad v = \frac{y-m_2}{\sigma_2}$$

La loi marginal de Y est donnée par :

$$f_Y(y) = \int_{-\infty}^{+\infty} A e^{-\frac{u^2 - 2ruv + v^2}{2(1-r^2)}} \sigma_1 du = A\sigma_1 e^{-\frac{v^2}{2}} \int_{-\infty}^{+\infty} e^{-\frac{(u-rv)^2}{2(1-r^2)}} du$$

Faisons le changement de variable :

$$z = \frac{u - rv}{\sqrt{1-r^2}}, \quad dz = \frac{du}{\sqrt{1-r^2}}$$

Il vient :

$$\int_{-\infty}^{+\infty} e^{-\frac{(u-rv)^2}{2(1-r^2)}} du = \sqrt{1-r^2} \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} dz = \sqrt{1-r^2} \sqrt{2\pi}$$

Et la densité s'écrit :

$$f_Y(y) = A \sigma_1 e^{-\frac{v^2}{2}} \sqrt{2\pi(1-r^2)} \quad \text{avec} \quad A = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \quad \text{et} \quad v = \frac{y-m_2}{\sigma_2}$$

Soit finalement : $f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-m_2)^2}{2\sigma_2^2}}$ On reconnaît la loi de Gauss à une dimension.

On obtient de la même manière la densité marginale en X : $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-m_1)^2}{2\sigma_1^2}}$

Ces relations montrent que

- m_1, σ_1 : sont l'espérance et l'écart type de la variable aléatoire X .
- m_2, σ_2 : sont l'espérance et l'écart type de la variable aléatoire Y

Remarques :

- Si la densité bidimensionnelle est gaussienne alors les densités marginales le sont aussi (la réciproque n'est généralement pas valable)
- Le dernier paramètre r a disparu de la définition des densités marginales. Il définit le coefficient de corrélation entre X et Y . Nous verrons sa définition et son interprétation dans les chapitres suivants.

6.4.2 Variables aléatoires gaussiennes indépendantes

Si les variables aléatoires X et Y sont indépendantes, la densité bidimensionnelle est égale au produit des densités marginales.

Donc si X et Y sont gaussiennes et indépendantes, la variable aléatoire (X, Y) est gaussienne et on vérifie dans ce cas que $r = 0$.

Réciproquement, si le couple (X, Y) a une densité de probabilité conjointe gaussienne, et si $r = 0$ alors les variables aléatoires X et Y sont indépendantes.

Pour des variables aléatoires **gaussiennes**, la non corrélation implique l'indépendance.

Nous verrons plus tard que l'indépendance implique toujours la non corrélation.

La réciproque n'est assurée que dans le cas où les deux variables aléatoires sont gaussiennes.

6.5 Transformation de deux variables aléatoires

6.5.1 Introduction

Avec une variable aléatoire à une dimension, nous avons étudié la possibilité de définir une nouvelle variable par l'intermédiaire d'une fonction certaine, nous généralisons maintenant cette notion au cas bidimensionnel.

Nous commençons par le cas particulier, très utile dans la pratique, qui consiste à définir une variable unique à partir d'un couple de variables aléatoires.

6.5.2 Variable aléatoire 1D fonction de deux variables aléatoires 1D

On considère :

- un couple de deux variables aléatoires (X,Y) admettant une densité de probabilité conjointe $f_{X,Y}(x,y)$
- une fonction certaine $z = g(x,y)$

On se propose de déterminer la densité de probabilité et la fonction de répartition de la variable aléatoire Z déduite du couple (X,Y) par la relation : $Z = g(X,Y)$

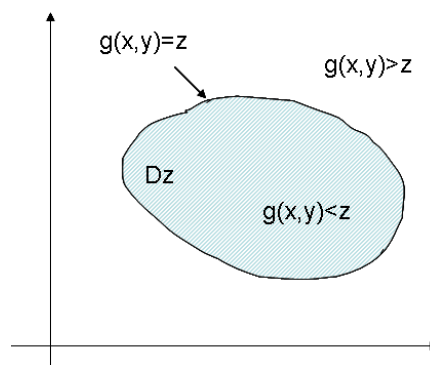
Calcul de la fonction de répartition

Pour connaître la fonction de répartition de Z , il faut calculer la probabilité de l'événement :

$$\{Z < z\} = \{g(X,Y) < z\}$$

La condition $g(X,Y) < z$ définit dans le plan (X,Y) un certain domaine D_z pour z fixé. On peut donc écrire :

$F_Z(z) = P(Z < z) = P((X,Y) \in D_z) = \iint_{D_z} f_{X,Y}(x,y) dx dy$	10-1
---	------



Calcul de la densité de probabilité

La densité peut naturellement se calculer à partir de la fonction de répartition :

$$f_Z(z) = \frac{dF_Z(z)}{dz}$$

On peut également la calculer directement sur le domaine. On considère le domaine D_z défini comme précédemment :

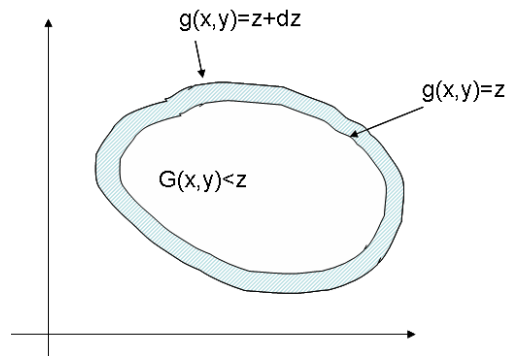
$$D_z = \{(x,y); g(x,y) < z\}$$

Appliquons à z une variation dz . Le domaine devient :

$$D_z + dz = \{(x,y); g(x,y) < z + dz\}$$

Soit $\Delta D_z = "D_z + dz" - "D_z"$ le domaine différentiel défini par

$$\Delta D_z = \{(x,y); z \leq g(x,y) < z + dz\}$$



En intégrant $f_{X,Y}(x, y)$ sur ΔD_z , on obtient :

$$f_Z(z)dz = P(z \leq Z < z + dz) = P((X, Y) \in \Delta D_z)$$

$$f_Z(z)dz = \iint_{\Delta D_z} f_{X,Y}(x, y) dx dy$$

10-2

6.5.3 Application 1 : somme de deux variables aléatoires

C'est l'opération fondamentale avec deux variables aléatoires. La variable Z est définie par :

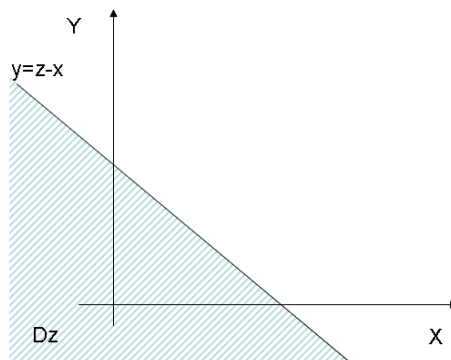
$$Z = X + Y$$

Fonction de répartition

Suivant la méthode décrite au §6.5.2, on définit le domaine D_z pour z fixé par :

$$x + y < z.$$

C'est la partie du plan des (X, Y) situés à gauche de la droite d'équation $y = z - x$



La fonction de répartition se calcule comme :

$$F_Z(z) = \iint_{D_z} f_{X,Y}(x, y) dx dy$$

Pour calculer cette intégrale double, on fixe y et on fait varier x de $-\infty$ à $z - y$, puis on intègre en y de $-\infty$ à $+\infty$:

$$F_Z(z) = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z-y} f_{X,Y}(x, y) dx \right) dy$$

10-3

En intégrant d'abord en y , puis en x , on aurait obtenu :

$$F_Z(z) = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z-x} f_{X,Y}(x, y) dy \right) dx$$

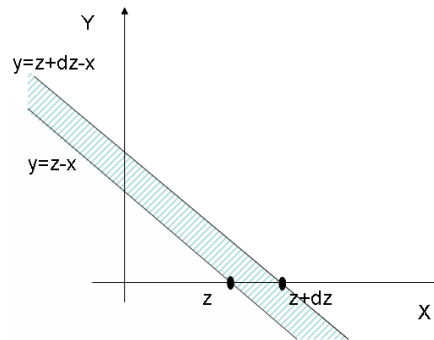
10-4

Densité de probabilité

On considère le domaine ΔD_z compris entre les droites :

$$x + y = z$$

$$x + y = z + dz$$



Comme précédemment, on fixe, par exemple, x et on intègre en y , mais cette fois entre les deux droites

$$f_Z(z)dz = \iint_{\Delta D_z} f_{X,Y}(x, y) dx dy$$

$$f_Z(z)dz = \int_{-\infty}^{+\infty} \left(\int_{z-x}^{z-x+dz} f_{X,Y}(x, y) dy \right) dx$$

Sur l'intervalle différentiel dz la densité de probabilité du couple est constante et il vient :

$$f_Z(z) = \int_{-\infty}^{+\infty} f_{X,Y}(x, z-x) dx \quad 10-5$$

En fixant y et en intégrant entre les deux droites on trouve :

$$f_Z(z) = \int_{-\infty}^{+\infty} f_{X,Y}(z-y, y) dy \quad 10-6$$

Cas de variables indépendantes

La fonction de densité prend une forme particulièrement simple lorsque les variables aléatoires sont indépendantes.

Dans ce cas $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ et les équations 10-5 et 10-6 deviennent :

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)dx$$

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(z-y)f_Y(y)dy$$

On reconnaît une équation de convolution entre les lois marginales.

$$f_Z(z) = f_X \otimes f_Y(z) = \int_{-\infty}^{+\infty} f_X(z-y)f_Y(y)dy \quad 10-7$$

Application au cas discret

On considère l'expérience consistant à jeter un dé deux fois de suite.

Les variables aléatoires X et Y sont définies par :

$X = i$, résultat du premier jet, $i = 1, 2, \dots, 6$ et $P(\{X = i\}) = \frac{1}{6}$

$Y = j$, résultat du second jet, $j = 1, 2, \dots, 6$ et $P(\{Y = j\}) = \frac{1}{6}$

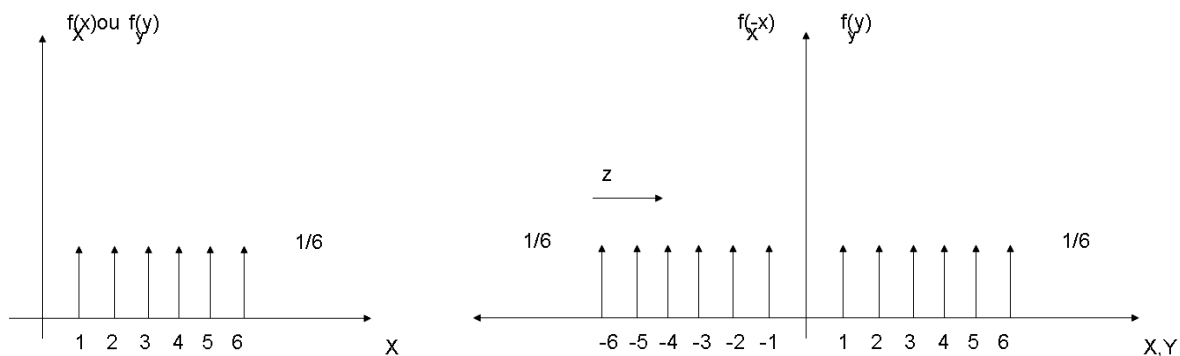
Les deux jets étant indépendants, la densité bidimensionnelle du couple (X, Y) est donnée par le produit des densités marginales.

$$P(\{X = i \text{ et } Y = j\}) = P(\{X = i\})P(\{Y = j\}) = \frac{1}{36}$$

On désire calculer la densité de probabilité de la variable Z définie par : $Z = X + Y$, par la méthode de convolution et la méthode générale.

- Calcul par convolution.

La convolution suppose le retournement d'une fonction suivi par une translation. On multiplie alors les deux fonctions et on intègre le produit obtenu. Cette opération est présentée dans les figures qui suivent.



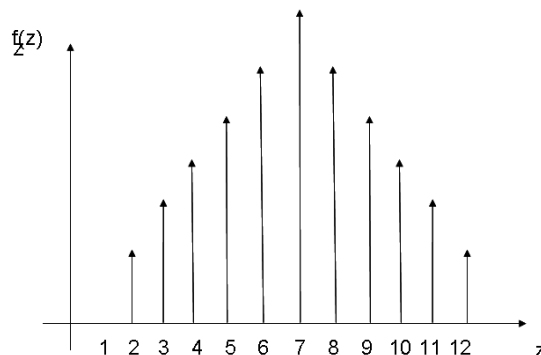
La première valeur de $f_z(z)$ est obtenue pour $z = 2$. Il y a deux raies de Dirac en correspondance et la probabilité correspondante est, après le produit,

$$P(\{Z = 2\}) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$$

Avec une translation de $z = 3$, il y a deux termes communs à chaque densité, d'où après produit et sommation :

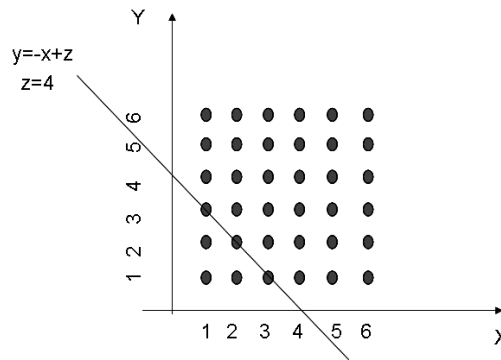
$$P(\{Z = 3\}) = 2 \cdot \frac{1}{36} = \frac{2}{36}$$

On obtient, en définitive, une densité triangulaire déjà rencontrée.



- calcul par la méthode générale

Représentons la densité bidimensionnelle



et sommions les masses de probabilité le long de la droite d'équation $y = -x + z$.

Par exemple, avec $z=4$, on trouve $P(Z=4) = 3/36$

En faisant varier z de 2 à 12 on retrouve la densité précédente.

6.5.4 Application 2 : Opération Min et Max

Fonction de répartition de $Z = \text{Max}(X, Y)$

La variable Z est définie par :

$$Z = X \text{ si } X \geq Y$$

$$Z = Y \text{ si } Y \geq X$$

Pour calculer la probabilité de l'événement $\{Z < z\}$ il faut déterminer le domaine D_z tel que $\text{Max}(X, Y) < z$. Ceci suppose que les deux coordonnées (x, y) sont inférieures à z .

On a $P(\text{Max}(X, Y) < z) = P(X < z \text{ et } Y < z)$

$F_Z(z) = F_{X,Y}(z, z)$	10-8
--------------------------	------

Si les variables sont indépendantes :

$$F_Z(z) = F_X(z)F_Y(z)$$

En fiabilité, la variable aléatoire Z correspond au temps de panne de deux systèmes montés en parallèle.

Fonction de répartition de $Z = \text{Min}(X, Y)$

La variable Z est définie par :

$$Z = X \text{ si } X \leq Y$$

$$Z = Y \text{ si } Y \leq X$$

Pour calculer la probabilité de l'événement $\{Z < z\}$, il faut déterminer le domaine D_z tel que $\text{Min}(X, Y) < z$. Pour qu'un point (x, y) du plan satisfasse à cette condition, il faut et il suffit qu'une de ses coordonnées soit inférieure à z . Donc il suffit que : $x < z$ ou $y < z$.

Tous les points de coordonnées (x, y) telles que $\begin{matrix} x < z \text{ et } y \in [-\infty, +\infty] \\ y < z \text{ et } x \in [-\infty, +\infty] \end{matrix}$ répondent à la question.

$$P(\text{Min}(X, Y) < z) = P(X < z \text{ ou } Y < z) = P(X < z) + P(Y < z) - P(X < z \text{ et } Y < z)$$

C'est-à-dire :

$$F_Z(z) = F_X(z) + F_Y(z) - F_{X,Y}(z, z) \quad 10-9$$

Si X et Y sont des variables aléatoires indépendantes on a :

$$F_Z(z) = F_X(z) + F_Y(z) - F_X(z)F_Y(z)$$

La variable aléatoire Min est très utilisée dans les problèmes de fiabilité. Si X et Y représentent les temps de panne de deux systèmes. La variable Z correspond au temps de panne des deux systèmes placés en série.

6.5.5 Application 3 : Rapport et produit.

Fonction de répartition et densité de probabilité de $Z = Y / X$

Pour calculer la probabilité de l'événement, $\{Z < z\}$ il faut déterminer le domaine D_Z tel que $\frac{y}{x} < z$, qui peut s'écrire comme :

$$y < xz \text{ si } x > 0$$

$$y > xz \text{ si } x < 0$$

Le domaine D_Z se décompose en deux parties suivant le signe de z . La fonction de répartition est donnée par :

$$F_Z(z) = \iint_{D_Z} f_{X,Y}(x, y) dx dy$$

$$F_Z(z) = \int_0^{+\infty} \left(\int_{-\infty}^{zx} f_{X,Y}(x, y) dy \right) dx + \int_{-\infty}^0 \left(\int_{zx}^{+\infty} f_{X,Y}(x, y) dy \right) dx \quad 10-10$$

On obtient la densité en intégrant $f_Z(z) dz = \iint_{\Delta D_Z} f_{X,Y}(x, y) dx dy$ sur le domaine ΔD_Z défini par

les droites $y = xz$ et $y = x(z + dz)$:

$$f_Z(z) dz = \int_0^{+\infty} \left(\int_{zx}^{(z+dz)x} f_{X,Y}(x, y) dy \right) dx + \int_{-\infty}^0 \left(\int_{(z+dz)x}^{zx} f_{X,Y}(x, y) dy \right) dx$$

On remarque que, par exemple, $\int_{zx}^{(z+dz)x} f_{X,Y}(x, y) dy = \int_{zx}^{zx+xdz} f_{X,Y}(x, y) dy \approx f_{X,Y}(x, zx) x dz$

$$\text{Il vient alors : } f_Z(z) dz = \int_0^{+\infty} dx x f_{X,Y}(x, zx) - \int_{-\infty}^0 dz x f_{X,Y}(x, zx) dx$$

Et finalement.

$$f_Z(z) = \int_{-\infty}^{+\infty} |x| f_{X,Y}(x, zx) dx \quad 10-11$$

Fonction de répartition et densité de probabilité de $Z = XY$

En appliquant la même méthode, on trouve que le domaine D_Z est compris entre deux hyperboles équilatères. La séparation des cas $y > 0$ et $y < 0$ donne :

$$F_Z(z) = \int_0^{+\infty} \left(\int_{-\infty}^{z/y} f_{X,Y}(x,y) dx \right) dy + \int_{-\infty}^0 \left(\int_{-\infty}^{+\infty} f_{X,Y}(x,y) dx \right) dy \quad 10-12$$

Et pour la densité :

$$f_Z(z) = \int_{-\infty}^{+\infty} \frac{1}{|y|} f_{X,Y}\left(\frac{z}{y}, y\right) dy \quad 10-1.$$

6.5.6 Variable aléatoire 2D fonctions de deux variables aléatoires 1D

On considère :

- une variable aléatoire bidimensionnelle continue dont on connaît la fonction de répartition ou la densité de probabilité.
- deux fonctions réelles et certaines de deux variables réelles

$$z = g(x, y) \text{ et } w = h(x, y)$$

A l'aide de ces deux fonctions on construit deux nouvelles variables aléatoires $Z = g(X, Y)$ et $W = h(X, Y)$.

On cherche la fonction de répartition et la densité bidimensionnelle du couple (Z, W) connaissant celle du couple (X, Y)

Fonction de répartition

Comme précédemment, nous définissons le domaine $D_{Z,W}$ comme l'ensemble des points (x, y) satisfaisant cette fois à la double condition :

$$D_{Z,W} : \left\{ \begin{array}{l} g(x, y) < z \\ h(x, y) < w \end{array} \right\} \text{ où } z \text{ et } w \text{ sont des valeurs fixées.}$$

L'événement équivalent est $\{Z < z \text{ et } W < w\} = \{(X, Y) \in D_{Z,W}\}$ et en prenant la probabilité des deux membres de cette équation :

$$F_{Z,W}(z, w) = \iint_{D_{Z,W}} f_{X,Y}(u, v) du dv \quad 10-13$$

Densité de probabilité

Nous définissons le domaine différentiel $\Delta D_{Z,W}$ comme l'ensemble des points (x, y) :

$$\Delta D_{Z,W} := \{(x, y) \text{ tels que } z \leq g(x, y) < z + dz \text{ et } w \leq h(x, y) < w + dw\}$$

En intégrant la densité sur ce domaine, on obtient :

$$f_{Z,W}(z, w) dz dw = \iint_{\Delta D_{Z,W}} f_{X,Y}(x, y) dx dy \quad 10-14$$

Dans le cas où la relation entre (z, w) et (x, y) est biunivoque, on peut considérer la densité comme constante sur $\Delta D_{Z,W}$ ce qui permet d'écrire :

$$f_{Z,W}(z, w) dz dw = f_{X,Y}(x, y) dx dy$$

Pour obtenir la densité recherchée, il faut utiliser la notion de Jacobien qui détermine la correspondance entre les surfaces différentielles.

On démontre alors que $dzdw = |J(x, y)|dxdy$ avec $J(x, y) = \begin{vmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \end{vmatrix}$

Ce qui donne pour la densité :

$f_{z,w}(z, w) = \frac{f_{x,y}(x, y)}{ J(x, y) }$	10-15
---	-------

Remarque : Comme dans le cas d'une fonction inverse d'une variable, on peut considérer l'opération inverse qui fait passer de l'aire $dzdw$ à l'aire $dxdy$

$$dxdy = |J(z, w)|dzdw$$

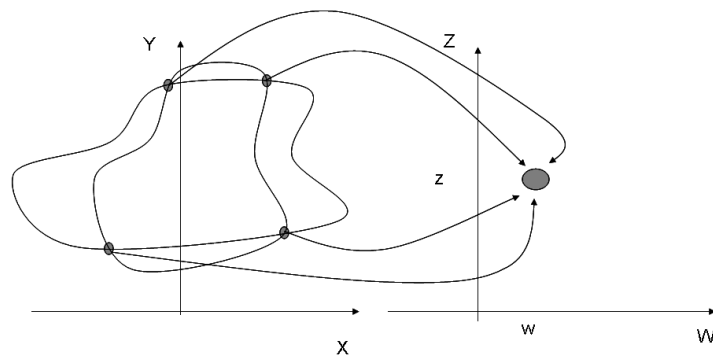
Avec cette fois comme déterminant : $J(z, w) = \begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial w} \end{vmatrix}$

Les deux Jacobiens sont liés par la relation : $J(x, y)J(z, w) = 1$

Lorsque la transformation est univoque, la transformation du couple (x, y) en (z, w) n'est pas bijective, la formule précédente ne peut plus s'appliquer.

Considérons maintenant le cas où plusieurs couples (x_i, y_i) ont pour image par (g, h) un même point (z, w) .

Ces points correspondent aux intersections des courbes $g(x, y) = z$ et $h(x, y) = w$ pour z et w donnés :



Si il y a n racines (n couples (x_i, y_i)), on peut écrire pour les événements

$$\{z \leq Z < z + dz, w \leq W < w + dw\} = \sum_i \{x_i \leq X < x_i + dx_i, y_i \leq Y < y_i + dy_i\}$$

Or, en (X, Y) les événements sont disjoints donc, en prenant la probabilité de chacun des membres de cette équation, on obtient :

$$f_{Z,W}(z,w)dzdw = \sum_i f_{X,Y}(x_i, y_i)dx_i dy_i$$

Et encore en utilisant le jacobien :

$f_{Z,W}(z,w) = \sum_i \frac{f_{X,Y}(x_i, y_i)}{ J(x_i, y_i) }$	10-16
---	-------

Où les couples (x_i, y_i) sont les racines de $g(x, y) = z$ et $h(x, y) = w$

6.5.7 Applications 1 : transformation linéaire

Les variables Z et W sont liées à X et Y par des relations linéaires :

$$Z = g(X, Y) = aX + bY$$

$$W = h(X, Y) = cX + dY$$

Les droites d'équations :

$$\begin{cases} z = ax + by \\ w = cx + dy \end{cases}$$

se coupent en général en un point. La relation est biunivoque.

On peut écrire réciproquement :

$$\begin{cases} x = U(z, w) = \alpha z + \beta w \\ y = V(z, w) = \chi z + \delta w \end{cases}$$

Le Jacobien s'écrit

$$|J(z, w)| = \begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial w} \end{vmatrix} = \begin{vmatrix} \alpha & \beta \\ \chi & \delta \end{vmatrix} = |\alpha\delta - \beta\chi|$$

On trouve alors la densité du couple (Z, W) :

$$f_{Z,W}(z, w) = \frac{f_{X,Y}(\alpha z + \beta w, \chi z + \delta w)}{|\alpha\delta - \beta\chi|}$$

Si $\alpha\delta = \beta\chi$, les droites sont parallèles et les variables Z et W sont liées par une relation linéaire.

La densité du couple (Z, W) devient singulière. Elle est nulle partout sauf sur la droite d'équation $\beta w = \delta z$ où elle est infinie. C'est un cas très particulier que nous retrouverons au chapitre suivant à propos de la corrélation de deux variables aléatoires.

6.5.8 Applications 2 : transformation $Z = \sqrt{X^2 + Y^2}$ et $W = Y/X$

Commençons par déterminer les racines des équations : $z = \sqrt{x^2 + y^2}$ et $w = \frac{y}{x}$,

Ce sont les intersections du cercle de rayon z et de la droite d'équation $y = wx$.

On trouve deux racines :

$$\begin{cases} x_I = \frac{z}{\sqrt{I+w^2}} \\ y_I = \frac{zw}{\sqrt{I+w^2}} \end{cases} \text{ et } \begin{cases} x_2 = -x_1 \\ y_2 = -y_1 \end{cases}$$

Le calcul du Jacobien donne :

$$J(x,y) = \begin{vmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \\ -\frac{y}{x^2} & \frac{I}{x} \end{vmatrix} = \frac{I + \left(\frac{y}{x}\right)^2}{\sqrt{x^2+y^2}}$$

$$\text{Donc : } J(x_1, y_1) = J(x_2, y_2) = \frac{1+w^2}{z}$$

Ce qui nous donne pour la densité : $f_{Z,W}(z,w) = \frac{z}{1+w^2} (f_{X,Y}(x_1, y_1) + f_{X,Y}(x_2, y_2))$.

Application au cas gaussien.

Supposons que les variables aléatoires X et Y sont gaussiennes, indépendantes, centrées et de mêmes variances : $f_{X,Y}(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$,

En remplaçant dans l'expression précédente, on obtient :

$$f_{Z,W}(z,w) = \frac{I}{\pi} \cdot \frac{I}{1+w^2} \left(\frac{z}{\sigma^2} e^{-\frac{z^2}{2\sigma^2}} \right) \text{ avec } z \geq 0 \text{ et } w \in \mathbb{R}$$

Expression qui montre après calcul des densités marginales que :

- les variables Z et W sont indépendantes ($f_{Z,W}(z,w) = f_Z(z)f_W(w)$)
- Z suit une loi de Rayleigh
- W suit une loi de Cauchy

6.5.9 Applications 3 : Variables aléatoires auxiliaires

La méthode générale pour deux fonctions de deux variables s'applique aussi au cas d'une seule fonction et conduit parfois à des solutions plus simples.

On pose dans ce cas $\begin{cases} Z = g(X,Y) \\ W = X \end{cases}$

Quand on a obtenu la densité bidimensionnelle en (Z,W) suivant la méthode générale, il suffit d'intégrer par rapport à W pour avoir la densité de Z :

$$f_Z(z) = \int f_{Z,W}(z,v) dv$$

Par exemple pour la somme de deux variables aléatoires on pose :

$$\begin{cases} Z = X + Y \\ W = X \end{cases} \text{ Soit } \begin{cases} Y = Z - W \\ X = W \end{cases}$$

Le Jacobien est égal à 1 et il vient d'après 10-14 : $f_{Z,W}(z, w) = f_{X,Y}(w, z - w)$

$$\text{d'où } f_Z(z) = \int f_{X,Y}(v, z - v) dv$$

6.6 Exercices

6.6.1 Énoncés

Loi bidimensionnelle discrète

1). On tire deux cartes au hasard dans une boîte qui contient cinq cartes numérotées 1, 1, 2, 2, 3. Soit X la somme et Y le plus grand des deux nombres tirés.

Définir la loi du couple (X, Y) . (Représenter le tableau de contingence).

2). On lance une pièce et un dé. La variable X prend la valeur +1 si le résultat de la pièce est pile, -1 dans le cas contraire.

La variable Y est égale au nombre qui apparaît sur la face supérieure du dé.

Donner la loi du couple (X, Y) .

3). On lance trois fois de suite une pièce de monnaie. Sur cette expérience, on définit la variable bi-dimensionnelle (X, Y) par les applications suivantes :

$X = 0$ si le résultat de la première épreuve est face et $X = 1$ si c'est pile.

Y est égale au nombre de fois où face est sortie.

- Donner l'expression des lois marginales en X et en Y .
- Loi du couple (X, Y) ?
- Loi de Y conditionnelle à $x = 0$?

Loi bidimensionnelle continue

1) La variable aléatoire (X, Y) est distribuée uniformément sur le domaine D :

$$D = \{(x_1, x_2), (y_1, y_2)\} = \{x \in [0, 2] \times y \in [0, 5]\}$$

- déterminer la densité $f_{X,Y}(x, y)$
- en déduire les densités marginales
- donner l'expression de la fonction de répartition $F_{X,Y}(x, y)$ pour les valeurs de seuils suivants
 - a) $(x, y) \in D$
 - b) $0 < x < 2, y > 5$
 - c) $x > 2, 0 < y < 5$
 - d) $x \geq 2, y \geq 5$
 - e) $x \leq 0, y \leq 0$

2) On considère la fonction de densité donnée par :

$$f_{X,Y}(x,y) = \begin{cases} Ay(1-x) & 0 \leq x \leq 1, 0 \leq y \leq 2xy \\ 0 & \text{ailleurs} \end{cases}$$

- Déterminer le coefficient A
- Déterminer les lois marginales, les variables aléatoires X et Y sont-elles indépendantes.
- Trouver la probabilité de l'événement : $\left\{ X \leq \frac{1}{2} \right\}$

3) Le couple de variables aléatoires (X,Y) a pour densité $f_{X,Y}(x,y)$ et pour fonction de répartition $F_{X,Y}(x,y)$.

- Donner l'expression de la fonction de répartition et de la densité de probabilité du couple (X,Y) conditionnellement à l'événement $\{X \geq 0, Y \geq 0\}$
- Donner la forme explicite de la densité conditionnelle lorsque X et Y sont normales, indépendantes, de moyenne nulle et de même écart type

Loi bidimensionnelle continue

1)

$$- f_{X,Y}(x,y) = \begin{cases} \frac{1}{10} & (x,y) \in D \\ 0 & \text{ailleurs} \end{cases}$$

$$- f_X(x) = \begin{cases} \frac{1}{2} & 0 \leq x < 2 \\ 0 & \text{ailleurs} \end{cases}, f_Y(y) = \begin{cases} \frac{1}{5} & 0 \leq y < 5 \\ 0 & \text{ailleurs} \end{cases}$$

$$-$$

- $F_{X,Y}(x,y) = \frac{xy}{10}$, pour $(x,y) \in D$
- $F_{X,Y}(x,y) = \frac{x}{2}$, pour $0 < x < 2, y > 5$
- $F_{X,Y}(x,y) = \frac{y}{5}$, pour $x > 2, 0 < y < 5$
- $F_{X,Y}(x,y) = 1$, pour $x \geq 2, y \geq 5$
- $F_{X,Y}(x,y) = 0$, pour $x \leq 0, y \leq 0$

2)

$$a) A \text{ tel que } \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x,y) dx dy = 1, A = 6$$

$$b) f_X(x) = \begin{cases} 12(1-x)x^2 & 0 < x < 1 \\ 0 & \text{ailleurs} \end{cases}, f_Y(y) = \begin{cases} \frac{3y}{4}(2-y)^2 & 0 < y < 2 \\ 0 & \text{ailleurs} \end{cases}$$

$$c) P\left(\left\{X \leq \frac{1}{2}\right\}\right) = F_X\left(\frac{1}{2}\right) = \frac{5}{16}$$

3)

$$a) F_{X,Y}(x,y / X \geq 0, Y \geq 0) = \frac{P(0 \leq X < x, 0 \leq Y < y)}{P(X \geq 0, Y \geq 0)},$$

$$F_{X,Y}(x,y / X \geq 0, Y \geq 0) = \frac{F_{X,Y}(x,y) - F_{X,Y}(0,y) - F_{X,Y}(x,0) + F_{X,Y}(0,0)}{F_{X,Y}(+\infty, +\infty) - F_{X,Y}(0, +\infty) - F_{X,Y}(+\infty, 0) + F_{X,Y}(0,0)}$$

$$F_{X,Y}(x,y / X \geq 0, Y \geq 0) = \frac{F_{X,Y}(x,y) - F_{X,Y}(0,y) - F_{X,Y}(x,0) + F_{X,Y}(0,0)}{1 - F_Y(0) - F_X(0) + F_{X,Y}(0,0)}$$

$$f_{X,Y}(x,y / X \geq 0, Y \geq 0) = \frac{\partial^2 F_{X,Y}(x,y / X \geq 0, Y \geq 0)}{\partial x \partial y} = \frac{f_{X,Y}(x,y)}{1 - F_Y(0) - F_X(0) + F_{X,Y}(0,0)}$$

$$b) f_{X,Y}(x,y / X \geq 0, Y \geq 0) = \frac{2}{\pi \sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$$

Transformation de variables

1) Les variables X et Y sont indépendantes et ont une densité de probabilité de Rayleigh :

$$f_X(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, x \geq 0$$

$$f_Y(y) = \frac{y}{\sigma^2} e^{-\frac{y^2}{2\sigma^2}}, y \geq 0$$

Déterminer la densité de probabilité et la fonction de répartition de la variable aléatoire

$$Z = \frac{X}{Y}$$

2) Les variables X et Y sont indépendantes et ont pour densité de probabilité :

$$f_X(x) = ae^{-ax}, x \geq 0, a > 0$$

$$f_Y(y) = ae^{-ay}, y \geq 0, a > 0$$

Déterminer la fonction de répartition, la densité de probabilité et l'espérance des variables aléatoires suivantes :

- a) $Z = X + Y$
- b) $U = X - Y$
- c) $\begin{cases} W = X - Y \text{ si } X \geq Y \\ W = 0 \text{ si } X < Y \end{cases}$
- d) $T = \frac{Y}{X}$

3) Le Jacobien peut être considéré comme un facteur de multiplication d'aire ; c'est à dire que si $z = g(x, y)$ et $w = h(x, y)$ on a :

$$ds_1 = dz dw = |J(x, y)| dx dy$$

$$ds_2 = dx dy = |J(z, w)| dz dw$$

Appliquer cette notion aux fonctions suivantes :

$$z = 2x \text{ et } w = 3y$$

Vérifiez vos résultats graphiquement avec

$$T = \{D_{z,w} | 2 < z \leq 4 ; 3 < w \leq 6\}$$

4) On considère les transformations de variables suivantes :

$$X = U + V \text{ et } Y = V - U^2$$

b) calculer le Jacobien $J(u, v)$

c) un triangle T dans le plan (u, v) a comme sommets: $(0,0)$, $(2,0)$, $(0,2)$, trouver le domaine S correspondant dans le plan (x, y)

d) calculer la surface S par intégration dans le domaine (x, y)

Vérifier en intégrant sur T en (u, v)

6.6.2 Solutions

1)

Y \ X	2	3	4	5	P Y
1	0,1	0	0	0	0,1
2	0	0,4	0,1	0	0,5
3	0	0	0,2	0,2	0,4
P X	0,1	0,4	0,3	0,2	1

2) Les variables sont indépendantes, donc :

$$P(X = x_i \text{ et } Y = y_j) = P(X = x_i)P(Y = y_j)$$

X \ Y	1	2	3	4	5	6	P X
1	1/12	1/12	1/12	1/12	1/12	1/12	1/2
-1	1/12	1/12	1/12	1/12	1/12	1/12	1/2
P Y	1/6	1/6	1/6	1/6	1/6	1/6	1

3)

Y	0	1	2	3
P(Y/x=)	0	1/4	2/4	1/4

X	Y	0	1	2	3	P X
0		0	1/6	2/8	1/8	1/2
1		1/8	2/8	1/8	0	1/2
	P Y	1/8	3/8	3/8	1/8	1

Transformation de variables

$$1) F_Z(z) = \frac{z^2}{z^2 + \left(\frac{a}{b}\right)^2}, z \geq 0$$

$$f_Z(z) = 2 \left(\frac{a}{b}\right)^2 \frac{z}{\left(z^2 + \left(\frac{a}{b}\right)^2\right)^2}, z \geq 0$$

2)

$$a). F_Z(z) = 1 + \frac{b}{a-b} e^{-az} - \frac{a}{a-b} e^{-bz}, z \geq 0$$

$$f_Z(z) = \frac{ab}{a-b} (e^{-az} - e^{-bz}), z \geq 0$$

$$E(Z) = \frac{a+b}{a-b}$$

$$b). F_U(u) = \frac{a}{a+b} e^{bu}, u \leq 0, F_U(u) = 1 - \frac{b}{a+b} e^{-au}, u \geq 0$$

$$f_U(u) = \frac{ab}{a+b} e^{bu}, u \leq 0, f_U(u) = \frac{ab}{a+b} e^{-au}, u \geq 0$$

$$E(U) = \frac{b-a}{ab}$$

$$c). F_W(w) = 1 + \frac{b}{a+b} e^{-aw}, w \geq 0$$

$$f_W(w) = \frac{ab}{a+b} e^{-aw} + \frac{a}{a+b} \delta(w)$$

$$E(W) = \frac{b/a}{a+b}$$

$$d). F_T(t) = \frac{at}{at+b}, t \geq 0$$

$$f_T(t) = \frac{a}{at+b}, t \geq 0$$

$$E(T) = +\infty$$

3) $J(x, y) = 6, J(z, w) = 1/6,$

$dS_1 = 6, dS_2 = 1/6$

4)

a). $J(u, v) = 1 + 2u,$

b). $D(x, y)$ est compris à l'intérieure des courbes $y=x$; $y=-x^2$; $x=2$

c). $S=14/3$

CHAPITRE 7. Espérance, fonction caractéristique et moments pour 2 variables aléatoires

7.1 Espérance d'une fonction

Cette question a déjà été étudiée avec une fonction d'une variable $y = g(x)$. Nous avons montré qu'il n'était pas nécessaire de connaître explicitement la densité de la variable aléatoire Y pour calculer son espérance.

Celle-ci était alors donnée par $E(Y) = E(g(X)) = \int g(x)f_X(x)dx$

On démontre de la même manière que si une variable Z est définie par : $Z = g(X, Y)$. Son espérance est donnée par :

$$E(Z) = E(g(X, Y)) = \iint g(x, y)f_{X,Y}(x, y)dxdy \quad 11-5$$

Pour une variable discrète, cette expression devient ($P_{i,j}$ est la densité bidimensionnelle) :

$$E(Z) = E(g(X, Y)) = \sum_{i,j} g(x_i, y_j)P_{i,j} \quad 11-6$$

Compte tenu de la linéarité des équations de définition (11.1) et (11.2) on peut écrire pour $g(x, y) = \sum_k \lambda_k g_k(x, y)$

$$E(\sum_k \lambda_k g_k(x, y)) = \sum_k \lambda_k E(g_k(x, y))$$

Par exemple, avec la somme pondérée de 2 variables, on obtient :

$$E(Z) = E(aX + bY) = aE(X) + bE(Y)$$

Cette relation est valable que X et Y soient indépendantes ou non.

Espérance conditionnelle bidimensionnelle

Nous avons vu au chapitre IX la définition de la densité conditionnelle de Y à $X = x$.

$$f_Y(y|X=x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

On en déduit l'expression de l'espérance mathématique de Y conditionnelle à $X = x$ qui est une fonction du paramètre x

$$E(Y|X=x) = \int y f_Y(y|X=x) dy = \frac{\int y f_{X,Y}(x, y) dy}{f_X(x)} \quad 11-7$$

D'une manière plus générale, pour une fonction $g(X, Y)$

$$E(g(X, Y)|X=x) = \int g(x, y) f_Y(y|X=x) dy = \frac{\int g(x, y) f_{X,Y}(x, y) dy}{f_X(x)} \quad 11-8$$

et symétriquement :

$$E(g(X, Y)|Y=y) = \int g(x, y) f_X(x|Y=y) dx = \frac{\int g(x, y) f_{X,Y}(x, y) dx}{f_Y(y)}$$

Un résultat utile qui découle de ces deux dernières expressions et si $g(x, y) = g_1(x)g_2(y)$ est :

$$E(g_1(X)g_2(Y)|X=x) = g_1(X)E(g_2(Y)|X=x)$$

Toutes ces expressions servent plus particulièrement, en statistique, dans les problèmes de régression.

7.2 Fonction caractéristique

7.2.1 Définition

On appelle fonction caractéristique de la variable aléatoire (X, Y) l'espérance de la fonction $\exp j(t_1 x + t_2 y)$

$$\varphi_{X,Y}(t_1, t_2) = E(e^{jt_1 x + jt_2 y}) \quad 11-9$$

Où t_1 et t_2 sont des variables réelles et certaines.

$$\varphi_{X,Y}(t_1, t_2) = \iint e^{jt_1 x + jt_2 y} f_{X,Y}(x, y) dx dy \quad 11-10$$

C'est au signe près la transformée de Fourier d'une fonction de deux variables.

La densité $f_{X,Y}(x, y)$ étant absolument intégrable, le théorème de réciprocité s'applique et on a

$$f_{X,Y}(x, y) = \frac{1}{(2\pi)^2} \iint e^{-(jt_1 x + jt_2 y)} \varphi_{X,Y}(t_1, t_2) dt_1 dt_2$$

Notons que les fonctions caractéristiques à une dimension correspondant aux densités marginales s'obtiennent par les relations :

$$\varphi_X(t_1) = \varphi_{X,Y}(t_1, 0)$$

$$\varphi_Y(t_2) = \varphi_{X,Y}(0, t_2)$$

et que $\varphi_{X,Y}(0, 0) = 1$

7.3 Théorème des moments

Comme pour le cas d'une variable, la fonction caractéristique à deux dimensions peut être développée en série de Mac LAURIN au voisinage de l'origine $(0, 0)$ et l'on obtient :

$$\varphi_{X,Y}(t_1, t_2) = 1 + jE(X)t_1 + jE(Y)t_2 - \frac{1}{2}E(X^2)t_1^2 - \frac{1}{2}E(Y^2)t_2^2 - E(XY)t_1 t_2 + \dots$$

7.4 Moments pour deux variables aléatoires

7.4.1 Définition

Le moment d'ordre n d'une variable aléatoire à une dimension a été défini par la relation

$$m_n = E(X^n) = \int x^n f_X(x) dx$$

Nous étendons maintenant cette définition au cas d'une variable aléatoire à deux dimensions.

Moments non centrés

Avec deux dimensions les moments peuvent dépendre de deux paramètres entiers l et k . On pose par définition

$$m_{l,k} = E(X^l Y^k) = \iint x^l y^k f_{X,Y}(x, y) dx dy \quad 11-11$$

Pour une variable discrète, cette expression devient

$$m_{l,k} = E(X^l Y^k) = \sum_i \sum_j x_i^l y_j^k P_{i,j}$$

11-12

Exemple : considérons l'expérience définie par le jet d'une pièce et d'un dé. On pose

$$x_i = i \quad i = 1, \dots, 6 \quad P(X = x_i) = P_i = \frac{1}{6}$$

$$y_j = +1 \quad j = 1 \quad \text{si le résultat est pile} \quad P(Y = y_j) = P_j = \frac{1}{2}$$

$$y_j = -1 \quad j = 2 \quad \text{si le résultat est face} \quad P(Y = y_j) = P_j = \frac{1}{2}$$

Les deux variables sont indépendantes. Il y a 12 résultats élémentaires (nombre, face) de probabilité :

$$P_{i,j} = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}$$

calculons $E(X), E(Y), E(XY)$

$$m_{1,0} = E(X) = \sum_{i=1}^6 \sum_{j=1}^2 x_i^1 y_j^0 = \frac{1}{12} \sum_{i=1}^6 2x_i = 3,5$$

$$m_{0,1} = E(Y) = \sum_{i=1}^6 \sum_{j=1}^2 x_i^0 y_j^1 = \frac{1}{12} \sum_{i=1}^6 x_i^0 \cdot 0 = 0$$

$$m_{1,1} = E(XY) = \sum_{i=1}^6 \sum_{j=1}^2 x_i^1 y_j^1 = \frac{1}{12} \sum_{i=1}^6 x_i \cdot 0 = 0$$

Ordre des moments :

L'ordre des moments est égal à la somme des indices $n = l + k$

Il y a ainsi deux moments du premier ordre ($n = 1$)

$$m_{1,0} = E(X^1 Y^0) = \iint x f_{X,Y}(x, y) dx dy = \int x f_X(x) dx = E(X)$$

$$m_{0,1} = E(X^0 Y^1) = \iint y f_{X,Y}(x, y) dx dy = E(Y)$$

Trois moments du second ordre ($n = 2$)

$$m_{2,0} = E(X^2 Y^0) = \iint x^2 f_{X,Y}(x, y) dx dy = \int x^2 f_X(x) dx = E(X^2)$$

$$m_{0,2} = E(X^0 Y^2) = \iint y^2 f_{X,Y}(x, y) dx dy = E(Y^2)$$

$$m_{1,1} = E(X^1 Y^1) = \iint xy f_{X,Y}(x, y) dx dy$$

Et, de manière plus générale, $n + 1$ moments d'ordre n

Moments centrés

Par définition le moment centré d'ordre $n=l+k$ s'écrit :

$$\mu_{l,k} = E((X - m_{1,0})^l (Y - m_{0,1})^k) = \iint (x - m_{1,0})^l (y - m_{0,1})^k f_{X,Y}(x, y) dx dy$$

11-13

On a bien sûr :

$$\mu_{1,0} = \mu_{0,1} = 0$$

On utilise plus particulièrement les moments centrés du deuxième ordre

$$\mu_{2,0} = E((X - m_{1,0})^2) = \sigma_X^2 \quad \text{La variance de } X$$

$$\mu_{0,2} = E((Y - m_{0,1})^2) = \sigma_Y^2 \quad \text{La variance de } Y$$

$$\mu_{1,1} = E((X - m_{1,0})(Y - m_{0,1})) = \sigma_{X,Y} \quad \text{La covariance de } X, Y$$

7.4.2 Covariance

Le passage à deux dimensions introduit un nouveau moment centré du deuxième ordre : la covariance.

$$\mu_{1,1} = E((X - m_{1,0})(Y - m_{0,1})) = \sigma_{X,Y} \quad 11-14$$

Nous avons vu que la variance s'exprimait en fonction des moments non centrés :

$$\mu_{2,0} = E((X - m_{1,0})^2) = \sigma_X^2 = E(X^2) - (E(X))^2 = m_{2,0} - m_{1,0}^2$$

$$\mu_{0,2} = E((Y - m_{0,1})^2) = \sigma_Y^2 = E(Y^2) - (E(Y))^2 = m_{0,2} - m_{0,1}^2$$

Il en est de même pour la covariance. On trouve, en développant le produit et en prenant l'espérance dans (11-10)

$$\mu_{1,1} = \sigma_{X,Y} = E(XY) - E(X)E(Y) = m_{1,1} - m_{1,0}m_{0,1} \quad 11-15$$

Propriétés

- contrairement aux variances en X et Y , qui sont toujours positives, la covariance peut être positive ou négative.

- il existe entre les différents moments du second ordre une inégalité importante : l'inégalité de SCHWARTZ.

$$E^2(XY) \leq E(X^2)E(Y^2) \quad 11-16$$

Ce qui donne pour des variables non centrées :

$$m_{1,1}^2 = m_{2,0}m_{0,2} \quad 11-17$$

Et pour des variables centrées :

$$\begin{aligned} \mu_{1,1}^2 &= \mu_{2,0}\mu_{0,2} \\ \sigma_{XY}^2 &= \sigma_X^2\sigma_Y^2 \end{aligned} \quad 11-18$$

Ces inégalités découlent du raisonnement suivant : a étant un nombre réel quelconque, on a : $E((Y + aX)^2) > 0$ quel que soit a , comme intégrale d'une fonction positive

Développons cette expression

$$E((Y + aX)^2) = E(Y^2) + a^2E(X^2) + 2aE(XY) > 0$$

C'est un trinôme du second degré en a qui doit être positif quel que soit a ;

donc son discriminant est négatif, c'est-à-dire

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

7.4.3 Coefficient de corrélation

7.4.3.1 Définition

On appelle coefficient de corrélation des variables aléatoires X et Y le nombre r quotient de la covariance par le produit des écart type de X et Y :

$$r = \frac{E((X - m_{1,0})(Y - m_{1,0}))}{\sqrt{E((X - m_{1,0})^2)E((Y - m_{0,1})^2)}} = \frac{\mu_{1,1}}{\mu_{2,0}\mu_{0,2}} = \frac{\sigma_{X,Y}}{\sigma_X\sigma_Y} \quad 11-19$$

7.4.3.2 Propriétés.

Le coefficient de corrélation exprime la dépendance linéaire entre les variables X et Y . L'inégalité de SCHWARTZ, pour les moments centrés, (11 – 14) donnent immédiatement $r^2 = 1$

$$-1 \leq r \leq 1 \quad 11-20$$

7.4.3.3 Valeurs particulières de r

Premier cas : $r^2 = 1$

Nous allons montrer que lorsque le coefficient de corrélation r est maximal et égal à l'unité, en valeur absolue, les deux variables aléatoires X et Y sont liées par une relation linéaire.

Cette démonstration sera faite de façon tout à fait générale pouvant servir d'introduction à la notion d'estimation statistique.

Le problème est alors le suivant : étant donné deux variables aléatoires X et Y , on se propose, connaissant X , d'approcher la variable Y par une fonction linéaire de X telle que l'erreur en moyenne quadratique entre Y et cette fonction soit minimale.

On prend donc comme fonction de X : $g(X) = aX + b$ où a et b sont deux constantes réelles.

L'erreur est alors donnée par la variable aléatoire : $Z = Y - (aX + b)$ et l'on doit déterminer a et b telles que $E(Z^2)$ soit minimale.

$$E(Z^2) = E((Y - (aX + b))^2)$$

Supposons a fixé et isolons b

$$E(Z^2) = E((Y - aX) - b)^2 = E((Y - aX)^2 + b^2 - 2b(Y - aX))$$

En appliquant l'algèbre de l'espérance on obtient

$$E(Z^2) = E((Y - aX)^2) + b^2 - 2bE(Y - aX)$$

En dérivant, on trouve que ce trinôme du second degré en b est minimal pour

$$b = E(Y - aX)$$

Soit :

$$\begin{aligned} b &= E(Y) - aE(X) \\ b &= m_{0,1} - am_{1,0} \end{aligned} \quad 11-21$$

On peut déjà donner une interprétation à ce résultat : la meilleure estimée d'une variable aléatoire (ici, $(Y - aX)$) par une constante est sa valeur moyenne.

Remplaçons b par sa valeur :

$$E(Z^2) = E((Y - m_{0,1}) - a(X - m_{1,0}))^2 = E(Y'^2 - aX'^2)$$

avec $Y' = Y - m_{0,1}$ et $X' = X - m_{1,0}$ variables centrées

et développons l'espérance :

$$E(Z^2) = E(Y'^2) + a^2 E(X'^2) - 2aE(X'Y')$$

Le même raisonnement nous indique que ce trinôme est minimal pour

$$a = \frac{E(X'Y')}{E(X'^2)} = \frac{\mu_{1,1}}{\sigma_X^2} = \frac{r\sigma_Y}{\sigma_X}$$

Finalement, on trouve que la meilleure estimation linéaire de Y par X , en moyenne quadratique, est donnée par la fonction :

$$g(x) = \frac{r\sigma_Y}{\sigma_X} (x - m_{1,0}) + m_{0,1} \quad 11-22$$

Dans ce cas, en remplaçant a par sa valeur, l'erreur $E(Z^2)$ est donnée par (11-19)

$$E(Z^2) = \sigma_Y^2(1 - r^2) \quad 11-23$$

Si $r^2 = 1$ on a $E(Z^2) = 0$, c'est-à-dire par définition $E(Z^2) = \int z^2 f_Z(z) dz = 0$

Comme le terme sous le signe somme est positif, cette condition ne peut être réalisée qu'avec

$$f_Z(z) = \delta(z)$$

La densité de la variable aléatoire Z se réduit à une raie de Dirac à l'origine, ce qui signifie que Z est nul avec une probabilité égale à 1.

L'estimation de Y par $g(X)$ est alors parfaite et les deux variables sont liées par la relation :

$$Y = \pm \frac{\sigma_Y}{\sigma_X} (X - m_{1,0}) + m_{0,1}$$

Ou de manière plus symétrique

$$\frac{Y - m_{0,1}}{\sigma_Y} = \pm \frac{X - m_{1,0}}{\sigma_X} \quad 11-24$$

Réciproquement : en appliquant la formule de définition de r , on trouve sans difficulté que si Y est de la forme $Y = aX + b$, le coefficient de corrélation de X et Y est égal à 1.

En conclusion, on peut dire qu'une condition nécessaire et suffisante pour que deux variables aléatoires soient liées linéairement, est que leur coefficient de corrélation soit égal à 1 en valeur absolue.

Deuxième cas : $r^2 = 0$

Lorsque le coefficient de corrélation est nul, on dit que les variables sont non corrélées.

$$r = 0 \text{ si } E(XY) = E(X)E(Y)$$

L'espérance du produit est égale au produit des espérances.

Application : variance d'une somme

Nous savons que l'espérance d'une somme est égale à la somme des espérances

$$E(aX + bY) = aE(X) + bE(Y)$$

On obtient le même genre de relation avec les variances lorsque les variables sont décorrélées.

$$\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) = a^2 \sigma_X^2 + b^2 \sigma_Y^2$$

La démonstration est immédiate en appliquant la définition car la non corrélation annule le terme $(X - m_{1,0})(Y - m_{0,1})$.

Remarque : la corrélation est en quelque sorte une mesure de la dépendance linéaire entre deux variables. Inversement, la corrélation entre deux variables peut être faible alors que ces variables sont très liées l'une à l'autre. Ce qui veut dire que, si une liaison existe, elle ne s'apparente pas à une forme linéaire comme le montre l'exemple suivant.

On considère une variable aléatoire X dont la densité est symétrique.

$$f_X(x) = f_X(-x) \text{ donc, } E(X) = 0$$

On définit la variable Y par $Y = |X|$, Y est parfaitement liée à X

Calculons le moment centré $\mu_{1,1}$ pour le couple (X, Y) .

$$\mu_{1,1} = E(XY) - E(X)E(Y)$$

Comme :

$$E(X) = 0 \text{ et } E(XY) = \int x|x|f_X(x)dx$$

Le terme sous l'intégrant étant une fonction impaire, cette intégrale est nulle et $r = 0$

La corrélation entre X et Y est nulle bien que les deux variables soient tout à fait dépendantes.

7.4.4 Corrélation et indépendance

Deux variables aléatoires sont stochastiquement indépendantes (indépendantes au sens des probabilités) si leur densité bidimensionnelle est égale au produit des densités marginales.

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

Dans ces conditions, on a quels que soient l et k ,

$$m_{l,k} = E(X^l Y^k) = E(X^l)E(Y^k)$$

$$m_{l,k} = m_{l,0}m_{0,k}$$

En particulier pour $l = k = 1$, $m_{1,1} = m_{1,0}m_{0,1}$ Ce qui signifie que $r = 0$.

Donc si deux variables aléatoires sont indépendantes, elles sont aussi décorrélées.

Réciproque : la réciproque est généralement fausse car la non corrélation ($r = 0$) entraîne simplement :

$$m_{1,1} = m_{1,0}m_{0,1}$$

Le seul cas où la non corrélation entraîne l'indépendance est celui des variables aléatoires gaussiennes.

La densité bidimensionnelle pour deux variables normales centrées est donnée par (cf. IX.3).

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}\left(\frac{x^2}{\sigma_X^2} - \frac{2rxy}{\sigma_X\sigma_Y} + \frac{y^2}{\sigma_Y^2}\right)}$$

Par le calcul, on vérifie que $E(XY) = r\sigma_X\sigma_Y$ ce qui justifie la notation.

En posant $r = 0$, $f_{XY}(x, y)$ se met sous la forme d'un produit de deux fonctions en x et y qui sont les densités marginales des variables aléatoires X et Y

$$f_{XY}(x, y) = f_X(x)f_Y(y) \text{ expression qui démontre l'indépendance.}$$

Une autre forme pratique pour exprimer l'indépendance de deux variables est de passer par la densité conditionnelle.

Si $f_Y(y|X=x)$ est indépendante de X , alors les variables sont indépendantes.

On a en effet :

$$f_Y(y|X=x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

Si X et Y sont indépendantes,

$$f_{XY}(x,y) = f_X(x)f_Y(y) \text{ et } f_Y(y|X=x) = f_Y(y)$$

Orthogonalité

Deux variables aléatoires sont dites orthogonales si $E(XY) = 0$

Avant d'en terminer avec ce chapitre, résumons les liaisons qui peuvent exister entre deux variables aléatoires et la terminologie correspondante.

Conditions

Si

$$f_{XY}(x,y) = f_X(x)f_Y(y)$$

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

$$E(XY) = 0$$

$$r \pm 1$$

Terminologie

Les variables X et Y sont

indépendantes et donc non corrélées

non corrélées

orthogonales

liées linéairement

7.5 Exercices

7.5.1 Énoncés

1) Calculer $E(X)$, $E(Y)$, la covariance de X et Y et leur coefficient de corrélation dans le cas de l'exercice 1 du chapitre 8.

2) Vérifier que $\text{Cov}(XY) = 0$ dans le cas de l'exercice 2 du chapitre 8.

3) Calculer $E(X)$, $E(Y)$, $\text{Cov}(XY)$, $r(XY)$ dans le cas de l'exercice 3 du chapitre 8.

4) X et Y sont deux variables indépendantes également réparties sur l'intervalle $[0, a]$. On considère les variables

$$S = X + Y$$

$$T = XY$$

Calculer le coefficient de corrélation r_1 entre X et S , r_2 entre X et T .

5) Les variables X et Y sont normales, centrées, d'écart type σ_X et σ_Y et de coefficient de corrélation r .

On considère le changement de variables

$$Z = X \cos \theta - Y \sin \theta$$

$$W = X \sin \theta + Y \cos \theta$$

Déterminer la valeur de θ qui rend Z et W indépendantes.

Que devient cette valeur quand $\sigma_X = \sigma_Y$?

Donner l'expression de Z et W dans ce cas.

7.5.2 Solutions

$$1) \quad \begin{array}{ll} E(X)=3,6 & E(Y)=2,3 \\ \text{Cov}(X, Y)=0,52 & r(X, Y)=0,9 \end{array}$$

$$2) \quad \begin{array}{ll} E(X)=0 & E(X, Y)=0 \\ \text{Cov}(X, Y)=0 & \end{array}$$

$$3) \quad \begin{array}{ll} E(X)=1/2 & E(Y)=3/2 \\ \text{Cov}(X, Y)=-1/4 & r(X, Y)=-1/3 \end{array}$$

4) En utilisant l'algèbre de l'espérance

$$r(X, S) = \frac{\sigma_X}{\sigma_S} = \frac{1}{\sqrt{2}} \quad r(X, T) = \frac{E(X)}{\sqrt{\sigma_X^2 + E^2(X)}} = \frac{\sqrt{3}}{7}$$

5)

$$\theta_0 \text{ tel que } \text{tg}(2\theta_0) = 2r \frac{\sigma_X \sigma_Y}{\sigma_Y^2 - \sigma_X^2}$$

$$\text{pour } \sigma_X = \sigma_Y \Rightarrow \theta_0 = \frac{\pi}{4}$$

$$\text{et } Z = \frac{1}{\sqrt{2}}(X - Y)$$

$$X = \frac{1}{\sqrt{2}}(X + Y)$$

CHAPITRE 8. Variable Aléatoire n dimensionnelle

Dans ce chapitre, nous allons généraliser à n variables aléatoires toutes les notions introduites pour deux variables.

En particulier, nous envisagerons l'extension des définitions aux cas suivants :

- Fonction de répartition et de densité
- Transformation pour des fonctions certaines
- Probabilité conditionnelle et indépendance
- Espérance mathématique

Les résultats obtenus seront ensuite illustrés par une introduction à la théorie des mesures.

8.1 Fonction de répartition et de densité

8.1.1 Définitions

La fonction de répartition n-dimensionnelle définit la probabilité de l'intersection des événements primordiaux $\{X_i < x_i\}$, i variant de 1 à n, où X_i désigne la variable de rang i et x_i le seuil relatif à cette variable :

$$F_{X_1 \dots X_n}(x_1, x_2, \dots, x_n) = P\{X_1 < x_1 \text{ et } X_2 < x_2, \dots, X_n < x_n\} \quad 12-1$$

Par analogie avec le cas bidimensionnel, nous aurons pour la densité

$$f_{X_1 \dots X_n}(x_1, x_2, \dots, x_n) = \frac{\partial^n F_{X_1 \dots X_n}(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \dots \partial x_n} \quad 12-2$$

8.1.2 Propriétés

La fonction de répartition est positive, bornée et l'hyper volume sous la fonction de densité est normalisé :

$$F_{X_1 \dots X_n}(x_1 = +\infty, x_2 = +\infty, \dots, x_n = +\infty) = \int_{R^n} f_{X_1 \dots X_n}(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n = 1$$

La fonction de répartition est une fonction non décroissante de chacune des variables.

La densité est une fonction positive ou nulle : $f_{X_1 \dots X_n}(x_1, x_2, \dots, x_n) \geq 0$

La probabilité pour que la variable (X_1, X_2, \dots, X_n) se trouve dans l'hyper volume différentiel $dx_1 dx_2 \dots dx_n$ centré sur le point (x_1, x_2, \dots, x_n) est proportionnelle à la densité en ce point

$$P\{(X_1, X_2, \dots, X_n) \in \Delta D_n(x_i, x_i + dx_i)\} = f_{X_1 \dots X_n}(x_1, x_2, \dots, x_n) dx_1, dx_2, \dots, dx_n, \quad 12-3$$

Lois marginales

On passe des densités et répartition à n dimensions aux lois marginales en éliminant les variables non désirées. Dans le cas de la densité, on intègre par rapport aux variables que l'on veut éliminer et dans celui de la répartition en prenant un seuil égal à $+\infty$

Par exemple, avec $n=4$, on obtiendra les lois marginales (x_2, x_3) en éliminant les variables x_1 et x_4

$$f_{X_2 X_3}(x_2, x_3) = \iint f_{X_1 X_2 X_3 X_4}(x_1, x_2, x_3, x_4) dx_1 dx_4$$

$$F_{X_2 X_3}(x_2, x_3) = F_{X_1 X_2 X_3 X_4}(+\infty, x_2, x_3, +\infty)$$

8.2 Transformations

8.2.1 Une fonction de n variables

On a souvent à considérer le passage de n variables aléatoires à une seule variable par l'intermédiaire d'une fonction certaine.

Soit Y cette nouvelle variable : $Y = g(X_1, X_2, \dots, X_n)$

Pour obtenir la densité de Y, il faut déterminer le domaine ΔD_y pour y donné dans l'espace à n dimensions qui donne tous les points (x_1, x_2, \dots, x_n) telle que soit satisfaite la condition

$$y \leq g(x_1, x_2, \dots, x_n) < y + dy$$

d'où l'on tire :

$$\begin{aligned} f_Y(y)dy &= P\{X_1, X_2, \dots, X_n \in \Delta D_y\} \\ f_Y(y)dy &= \int_{\Delta D_y} f_{X_1 \dots X_n}(x_1, x_2, \dots, x_n) dx_1, dx_2, \dots, dx_n \end{aligned} \quad 12-4$$

8.2.2 N fonctions de n variables

On considère n fonctions certaines, définies sur R^n , à valeur dans R.

$$y_k = g_k(x_1, x_2, \dots, x_n) \quad k = 1, \dots, n \quad 12-5$$

Pour trouver la densité de la variable (Y_1, Y_2, \dots, Y_n) on utilise le Jacobien de la transformation

$$f_{Y_1 \dots Y_n}(y_1, y_2, \dots, y_n) = \frac{f_{X_1 \dots X_n}(x_1, x_2, \dots, x_n)}{|J(x_1, x_2, \dots, x_n)|} \quad 12-6$$

Avec :

$$J(y_1, y_2, \dots, y_n) = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \dots & \frac{\partial y_1}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial y_n}{\partial x_1} & \frac{\partial y_n}{\partial x_2} & \dots & \frac{\partial y_n}{\partial x_n} \end{vmatrix} \quad 12-7$$

S'il existe plusieurs points $(x_1^r, x_2^r, \dots, x_n^r) \in R^n$ qui satisfont aux relations (12-5) pour $(y_1, y_2, \dots, y_n) \in R^n$ donnés la densité en Y s'écrira:

$$f_{Y_1 \dots Y_n}(y_1, y_2, \dots, y_n) = \sum_r \frac{f_{X_1 \dots X_n}(x_1^r, x_2^r, \dots, x_n^r)}{|J(x_1^r, x_2^r, \dots, x_n^r)|} \quad 12-8$$

8.3 Probabilité conditionnelle et indépendance

8.3.1 Probabilité conditionnelle

Avec deux variables X_1 et X_2 nous avons établi par un processus de limite (cf. IX-2) l'expression de la densité de probabilité de X_2 lorsque X_1 avait une valeur donnée x_1

$$f_{X_2}(x_2|X_1=x_1) = \frac{f_{X_1X_2}(x_1, x_2)}{f_{X_1}(x_1)}$$

$$\text{avec } f_{X_1}(x_1) = \int f_{X_1X_2}(x_1, x_2) dx_2$$

De la même manière, nous pouvons écrire, en considérant les $(n-k)$ événements conditionnels définis par :

$$X_j = x_j \text{ pour } j = k+1, k+2, \dots, n$$

$$f_{X_1 \dots X_k}(x_1, x_2, \dots, x_k | x_{k+1}, \dots, x_n) = \frac{f_{X_1 \dots X_n}(x_1, x_2, \dots, x_k, x_{k+1}, \dots, x_n)}{f_{X_{k+1} \dots X_n}(x_{k+1}, x_{k+2}, \dots, x_n)} \quad 12-9$$

8.3.2 Variables aléatoires indépendantes

Les variables X_1, \dots, X_n sont indépendantes dans leur ensemble si les événements primordiaux $\{X_i < x_i\}$, i variant de 1 à n , le sont.

Ce qui donne :

$$F_{X_1 \dots X_n}(x_1, x_2, \dots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \dots F_{X_n}(x_n) \quad 12-10$$

d'où l'on déduit la relation équivalente :

$$f_{X_1 \dots X_n}(x_1, x_2, \dots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \dots f_{X_n}(x_n) \quad 12-11$$

Les fonctions de répartition et de densité n -dimensionnelles sont égales aux produits des fonctions marginales correspondantes.

Il est alors évident que si les n variables sont indépendantes dans leur ensemble, elles sont aussi indépendantes k à k ($k < n$).

Il est à remarquer que la réciproque n'est pas valable (par exemple, l'indépendance de trois variables deux à deux n'entraîne pas l'indépendance des trois).

8.4 Espérance mathématique et fonction Caractéristique

8.4.1 Espérance mathématique

On posera, par définition, pour l'espérance d'une fonction:

$$E(g(x_1, x_2, \dots, x_n)) = \int_{R^n} g(x_1, x_2, \dots, x_n) f_{X_1 \dots X_n}(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n \quad 12-12$$

La linéarité de l'intégrale donne :

$$E(\sum_k \lambda_k g_k(x_1, x_2, \dots, x_n)) = \sum_k \lambda_k E(g_k(x_1, x_2, \dots, x_n)) \quad 12-13$$

Moments :

Les moments correspondent à l'espérance d'une fonction $g()$ particulière s'exprimant sous la forme d'un produit de puissances entières de chaque variable.

$m_{k_1, k_2, \dots, k_n} = E(X_1^{k_1} X_2^{k_2} \dots X_n^{k_n})$ avec $k_1 + k_2 + \dots + k_n = r$ l'ordre du moment.

Par exemple avec trois variables, $m_{3,2,1} = E(X_1^3 X_2^2 X_n^1)$ est un des moments d'ordre 6.

8.4.2 Fonction caractéristique

C'est encore une application de l'espérance pour une nouvelle fonction $g()$ particulière :

$$g(x_1, x_2, \dots, x_n) = e^{j(t_1 x_1 + t_2 x_2 + \dots + t_n x_n)}$$

$$\phi_{X_1, \dots, X_n}(x_1, x_2, \dots, x_n) = E(e^{j(t_1 x_1 + t_2 x_2 + \dots + t_n x_n)}) \quad 12-14$$

où les t_j sont des variables réelles certaines.

En particulier, si les X_i sont indépendantes,

$$\phi_{X_1, \dots, X_n}(t_1, \dots, t_n) = \phi_{X_1}(t_1) \phi_{X_2}(t_2) \dots \phi_{X_n}(t_n) \quad 12-15$$

Si, dans ce cas, on pose $t_1 = t_2 = \dots = t_n = t$, on trouve

$$\phi_{X_1, \dots, X_n}(t, \dots, t) = \phi_{X_1}(t) \phi_{X_2}(t) \dots \phi_{X_n}(t) = \prod_1^n \phi_{X_i}(t)$$

C'est la fonction caractéristique de la variable somme $Z = \sum_i X_i$ car

$$\phi_Z(t) = E(e^{jtz}) = E(e^{j(t \sum x_i)}) = E(\prod_i e^{jtx_i}) = \prod \phi_{X_i}(t) \quad 12-16$$

De plus, si tous les X_i ont la même loi

$$\phi_Z(t) = (\phi_{X_i}(t))^n \quad 12-17$$

Cette expression nous sera très utile dans l'étude des suites de variables aléatoires et en particulier pour la démonstration du théorème de la limite centrale.

8.4.3 Non corrélation, Orthogonalité et Indépendance

Les n variables aléatoires X_1, \dots, X_n sont non corrélées si leur covariance (par paire) est nulle, c'est-à-dire si :

$$E((X_i - m_i)(X_j - m_j)) = 0 \quad \forall i, j \quad i \neq j \quad 12-18$$

ou ce qui est équivalent $E(X_i X_j) = E(X_i)E(X_j) = m_i m_j \quad \forall i, j \quad i \neq j$

Si les X_i sont non corrélés, le théorème de la variance relatif à la somme pondérée de deux variables se généralise (le démontrer): si $Z = \sum_i \lambda_i X_i$

$$Var(Z) = \sigma_Z^2 = \lambda_i^2 Var(X_i) = \lambda_i^2 \sigma_{X_i}^2 \quad 12-19$$

Les variables aléatoires X_i sont orthogonales si $E(X_i X_j) = 0 \quad \forall i, j \quad i \neq j$

Dans ce cas, on peut écrire

$$E((\sum_i X_i)^2) = \sum_i E(X_i^2)$$

12-20

8.4.4 Application : Introduction au problème d'estimation d'une variable aléatoire

8.4.4.1 Estimation d'un paramètre

Nous considérons ici n variables aléatoires ayant toutes la même loi de probabilité. Cette situation se présente dans la pratique chaque fois que l'on entreprend de mesurer une certaine grandeur en répétant le nombre d'épreuves (la mesure) pour améliorer la précision.

En d'autres termes, on considère que la grandeur mesurée est affectée d'une erreur et que cette erreur donne à la mesure un caractère aléatoire susceptible d'être décrit par une loi de probabilité $f_X(x)$

En statistique, on décrit cette situation (n mesures ayant la même loi) en disant que l'on prend un échantillon de taille n dans une population parente non exhaustive (la répétition des épreuves ne modifie pas les caractéristiques de la population observée).

L'échantillon à n valeurs x_1, x_2, \dots, x_n qui représente les valeurs prises par n variables aléatoires X_1, X_2, \dots, X_n

A partir de ces mesures on veut faire une estimation de la variable aléatoire X c'est-à-dire connaître la loi $f_X(x)$ qui est la même pour toutes les variables X_i .

C'est généralement trop demandé et on va se contenter d'estimer l'espérance et la variance de la variable X .

8.4.4.2 Moyenne arithmétique d'un échantillon

Considérons un échantillon constitué de n variables aléatoires de même loi de probabilité et introduisant une nouvelle variable \bar{X} définie par :

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

12-21

C'est-à-dire la moyenne arithmétique de ces n variables.

Toutes les variables X_i ont les mêmes moments puisqu'elles ont, par hypothèse, la même densité $f_X(x)$. Posons :

$$E(X_i) = m$$

$$Var(X_i) = E((X_i - m)^2) = \sigma_X^2$$

Nous admettrons également que les X_i sont non corrélés :

$$E((X_i - m_i)(X_j - m_j)) = 0 \quad \forall i, j \quad i \neq j$$

Sous ces hypothèses, on trouve sans difficulté l'espérance et la variance de la variable \bar{X}

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i)$$

$$Var(\bar{X}) = Var\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n Var(X_i)$$

C'est-à-dire

$$\begin{aligned} E(\bar{X}) &= m \\ \text{Var}(\bar{X}) &= \frac{\sigma_X^2}{n} \end{aligned} \quad 12-22$$

La variance de \bar{X} nous donne une idée de la dispersion de cette variable autour de la moyenne m qui est également celle de X .

Pour interpréter ce résultat nous utiliserons le théorème de TCHEBYCHEV.

$$P(|\bar{X} - m| \geq \varepsilon) \leq \frac{\text{Var}(\bar{X})}{\varepsilon^2} = \frac{\sigma_X^2}{n\varepsilon^2} \quad 12-23$$

Cette inégalité montre que la probabilité pour que l'estimateur \bar{X} s'écarte de la moyenne m de X de plus de ε tend vers 0 quand la taille n de l'échantillon (le nombre de mesures) tend vers l'infini.

8.4.4.3 Variance de l'échantillon

Introduisons maintenant une nouvelle variable aléatoire \bar{V} transformée des n variables X_i par la relation :

$$\bar{V} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad 12-24$$

Cette variable (qui représente la moyenne arithmétique des carrés des écarts par rapport à la moyenne estimée) va nous servir à estimer la variance de X comme \bar{X} nous a servi à estimer l'espérance m de X

$$\text{Prenons l'espérance de } \bar{V} : E(\bar{V}) = \frac{1}{n} \sum_{i=1}^n E((X_i - \bar{X})^2)$$

Il faut donc calculer $E((X_i - \bar{X})^2)$

X_i et \bar{X} ayant la même espérance m faisons apparaître des variables centrées en ajoutant et en retranchant m

$$\begin{aligned} E((X_i - \bar{X})^2) &= E(((X_i - m) - (\bar{X} - m))^2) \\ &= E((X_i - m)^2) + E((\bar{X} - m)^2) - 2E((X_i - m)(\bar{X} - m)) \end{aligned}$$

$$E((X_i - m)^2) = \text{Var}(X_i) = \sigma_X^2$$

De ces trois termes, deux nous sont connus :

$$E((\bar{X} - m)^2) = \text{Var}(\bar{X}) = \frac{\sigma_X^2}{n}$$

Calculons le dernier :

$$(X_i - m)(\bar{X} - m) = (X_i - m) \left(\frac{1}{n} \sum_{j=1}^n X_j - m \right) = (X_i - m) \frac{\sum_{j=1}^n (X_j - m)}{n}$$

En prenant l'espérance de cette dernière expression, les termes en $(X_i - m)(X_j - m)$ pour $i \neq j$ disparaissent compte tenu de l'hypothèse de décorrélation (12-18) et il reste :

$$E((X_i - m)(\bar{X} - m)) = \frac{E((X_i - m)^2)}{n} = \frac{\sigma_X^2}{n}$$

En définitive, on obtient $E((X_i - \bar{X})^2) = \sigma_X^2 + \frac{\sigma_X^2}{n} - 2\frac{\sigma_X^2}{n} = \frac{n-1}{n}\sigma_X^2$

soit pour l'espérance de \bar{V}

$E(\bar{V}) = \frac{1}{n} \sum_{i=1}^n E((X_i - \bar{X})^2) = \frac{n-1}{n} \sigma_X^2$	12-25
---	-------

La grandeur qui nous intéresse est σ_X^2 , la variance de X , avec $\sigma_X^2 = \frac{n}{n-1} E(\bar{V})$

On trouve ainsi que la moyenne de \bar{V} n'estime pas directement $Var(X)$. On dit que \bar{V} est un estimateur biaisé de la variance de X .

Si $n \rightarrow \infty$, $\bar{V} \rightarrow \sigma_X^2$ on dit (cf. cours de statistique) que \bar{V} est un estimateur asymptotiquement non biaisé de la variance de X

Notre étude n'est pas tout à fait complète car pour connaître la précision avec laquelle \bar{V} estime σ_X^2 . Il faudrait encore calculer la variance de \bar{V}

C'est-à-dire $E((\bar{V} - E(\bar{V}))^2) = E((\bar{V} - \frac{n-1}{n}\sigma_X^2)^2)$

C'est l'application de ces résultats aux variables normales (gaussiennes) qui conduit aux statistiques en chi carré (cf. IV-5).

Remarque : Pour obtenir un estimateur non biaisé de la variance de X , posons $\bar{V}' = \alpha \bar{V}$

Déterminons α tel que $E(\bar{V}') = \sigma_X^2$

$E(\bar{V}') = \alpha E(\bar{V}) = \alpha \frac{n-1}{n} \sigma_X^2$ d'après ce qui précède

d'où $\alpha = \frac{n}{n-1}$ et $\bar{V}' = \alpha \bar{V} = \frac{n}{n-1} \cdot \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$

l'estimateur non biaisé de la variance de X , si son espérance est inconnue et estimée par \bar{X} , a donc pour expression

$\bar{V}' = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$	12-26
---	-------

Vous pouvez vérifier que c'est bien cette expression qui est utilisée sur votre calculatrice pour estimer l'écart type d'une distribution statistique.

8.5 Matrice de corrélation et de variance-covariance

La représentation d'une variable aléatoire à n dimensions par un vecteur aléatoire conduit à la notion de matrice de corrélation et de matrice de variance-covariance pour des vecteurs à espérance nulle.

Cette notion est importante, car elle intervient directement dans l'expression de la densité de probabilité d'un vecteur aléatoire gaussien et permet la démonstration simple de ses propriétés. De plus, elle est très utilisée dans les méthodes modernes de traitement du signal (filtrage spatial, modélisation auto-régressive de processus).

8.5.1 Définitions

Vecteurs aléatoires complexes

Pour plus de généralités, nous considérons le cas d'un vecteur aléatoire complexe Z de dimension n . La variable complexe Z est un vecteur colonne de composantes $Z_j = X_j + iY_j$ ou X_j et Y_j sont deux variables aléatoires réelles.

$$\vec{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix}$$

Espérance de \vec{Z}

On désigne l'espérance du vecteur \vec{Z} par

$$E(\vec{Z}) = \begin{bmatrix} E(Z_1) \\ E(Z_2) \\ \vdots \\ E(Z_n) \end{bmatrix} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix} = \vec{m} \quad 12-27$$

8.5.2 Matrices de corrélation et de variance-covariance

On appelle matrice de corrélation du vecteur aléatoire \vec{Z} la matrice Γ_Z définie par :

$$\Gamma_Z = E(\vec{Z} \vec{Z}^+) \quad 12-28$$

Dans cette équation, \vec{Z}^+ représente le vecteur transposé conjugué du vecteur \vec{Z} :

$$\vec{Z}^+ = (\vec{Z}^T)^*$$

$$\text{et } \Gamma_Z = E \left\{ \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix} \cdot [Z_1^* \ Z_2^* \ \dots \ Z_n^*] \right\}$$

Γ_Z est une matrice carrée, ($n \times n$), dont les éléments $\gamma_{i,j}$ ont pour expression

$$\gamma_{i,j} = E(Z_i Z_j^*) \quad \begin{matrix} \text{pour } i = j & \gamma_{i,j} = E(|Z_i|^2) \\ \text{pour } i \neq j & \gamma_{i,j} = E(Z_i Z_j^*) \end{matrix} \quad 12-29$$

On obtient la matrice de variance covariance C_Z du vecteur \vec{Z} en considérant le vecteur centré $\vec{Z} - \vec{m}$

$$C_Z = E(\vec{Z} - \vec{m} (\vec{Z} - \vec{m})^+) = \Gamma_Z - \vec{m} \vec{m}^+ \quad 12-30$$

C_Z est une matrice carrée, ($n \times n$), dont les éléments $c_{i,j}$ ont pour expression

$$c_{i,j} = E\left((Z_i - m_i)(Z_j - m_j)^*\right)$$

$$\text{pour } i = j \quad c_{i,i} = E\left(|Z_i|^2 - |m_i|^2\right) = \sigma_{Z_i}^2 \text{ la variance de } Z_i$$

$$\text{pour } i \neq j \quad c_{i,j} = E\left(Z_i Z_j^*\right) - E(Z_i)E(Z_j^*) \text{ la covariance de } Z_i, Z_j$$

8.5.3 Propriétés de Γ_Z (et de C_Z)

Toutes les propriétés qui vont être énoncées sont bien sûr valables pour des variables centrées et donc pour C_Z

Propriété 1 : Γ_Z est une matrice hermitienne

Propriété 2 : Γ_Z est une matrice définie non-négative

De plus, si Γ_Z est de rang n, elle est définie positive

Propriété 3 : Γ_Z est diagonalisable

Propriété 4 : les valeurs propres de Γ_Z sont réelles, positives ou nulles

Les propriétés 2, 3, 4 sont, en fait, des propriétés générales des matrices hermitiennes (et des matrices symétriques pour Z réelle)

Démontrons ces propriétés

Propriété 1 : Γ_Z est une matrice hermitienne, c'est-à-dire $\gamma_{i,j} = \gamma_{j,i}^*$

$$\text{on a en effet } \gamma_{j,i}^* = E\left(\left(Z_j Z_i^*\right)^*\right) = E\left(Z_j^* Z_i\right) = \gamma_{i,j}$$

Les éléments de Γ_Z , symétriques par rapport à la diagonale principale, sont conjugués et les éléments de la diagonale principale sont réels, positifs ou nuls.

Propriété 2 : Γ_Z est définie non-négative.

Une matrice est dite définie non-négative si la forme quadratique Q associée à cette matrice est définie non-négative, c'est-à-dire si :

$$Q = \sum_i \sum_j \gamma_{i,j} u_i u_j^* \geq 0$$

où les u_i sont des composantes d'un vecteur certain $\vec{U}^T = (u_1, u_2, \dots, u_n)^T \in C^n$

En effet $Q(\vec{U})$ s'écrit matriciellement

$$Q(\vec{U}) = \vec{U}^+ \Gamma_Z \vec{U} = \vec{U}^+ E\left(\vec{Z} \vec{Z}^+\right) \vec{U}$$

$$Q(\vec{U}) = E\left(\vec{U}^+ \vec{Z} \vec{Z}^+ \vec{U}\right) \text{ car } \vec{U} \text{ est certain}$$

$$Q(\vec{U}) = E\left(\left(\vec{U}^+ \vec{Z}\right)\left(\vec{U}^+ \vec{Z}\right)^+\right)$$

$$Q(\vec{U}) = E\left(\left|\vec{U}^+ \vec{Z}\right|^2\right) = E\left(\left|\sum_{i=1}^n u_i Z_i\right|^2\right) \geq 0$$

De plus, si Γ_Z est de rang n , c'est-à-dire si son déterminant est non nul, elle est dite définie positive, c'est-à-dire que Q est positive et ne s'annule que si tous les u_i sont nuls.

Propriété 3 : Γ_Z est diagonalisable.

On démontre que si une matrice est hermitienne (ou symétrique pour une matrice réelle), il existe au moins une matrice P orthonormée, telle que la matrice D définie par $D = P^{-1} \Gamma_Z P$ soit une matrice diagonale.

$$D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

Les λ_i qui sont les valeurs propres de D sont également les valeurs propres de Γ_Z

Rappel (utile pour la suite): si P est orthonormée (ou unitaire), alors son inverse est égale à sa transposée conjuguée $P^{-1} = P^+$

Propriété 4 : les valeurs propres de Γ_Z sont réelles, positives ou nulles.

Cette propriété se déduit directement des deux précédentes.

Posons $\vec{U} = P\vec{V}$

P étant une matrice unitaire qui diagonalise Γ_Z il vient

$$\begin{aligned} Q(\vec{V}) &= (P\vec{V})^+ \Gamma_Z P\vec{V} \\ Q(\vec{V}) &= \vec{V}^+ P^+ \Gamma_Z P\vec{V} = \vec{V}^+ P^{-1} \Gamma_Z P \vec{V} \quad \text{car } P^+ = P^{-1} \\ Q(\vec{V}) &= \vec{V}^+ D \vec{V} = \sum_{i=1}^n \lambda_i v_i^2 \geq 0 \quad \forall \vec{V} \end{aligned}$$

Si D est de dimension n , toutes les valeurs propres λ_i sont différentes de 0.

La forme quadratique $Q(\vec{V})$ étant positive quels que soient les v_i on en déduit que les valeurs propres d'une matrice de corrélation, ou de variance covariance, sont nécessairement positives.

8.5.4 Interprétation probabiliste des propriétés de Γ_Z .

Il est intéressant d'interpréter les résultats qui découlent des propriétés des matrices à symétrie hermitienne en terme de propriétés des vecteurs aléatoires.

Soit \vec{W} le vecteur aléatoire déduit de \vec{Z} par la transformation linéaire : $\vec{W} = P^{-1} \vec{Z}$. Calculons la matrice de corrélation de \vec{W} :

$$\begin{aligned} \Gamma_W &= E(\vec{W} \vec{W}^+) = E(P^{-1} \vec{Z} (P^{-1} \vec{Z})^+) \\ \Gamma_W &= E(P^{-1} \vec{Z} \vec{Z}^+ P) \quad \text{car } P^{-1} = P^+ \\ \Gamma_W &= P^{-1} \Gamma_Z P = D \end{aligned}$$

Donc la matrice de corrélation de \vec{W} est diagonale et les valeurs propres λ_i de D (donc de Γ_Z) sont les moments du second ordre des composantes W_i du vecteur aléatoire \vec{W}

$$\lambda_i = E(W_i W_i^*) = E(|W_i|^2)$$

Ce sont des quantités positives. De plus, les composantes de \vec{W} sont orthogonales car puisque Γ_W est diagonale, cela signifie que :

$$E(W_i W_j^*) = 0 \quad \forall \quad i \neq j$$

Si Γ_W est de dimension $r < n$ (et donc également Γ_Z) nous aurons $n - r$ valeurs propres λ_i nulles. C'est-à-dire que les variables aléatoires W_i , pour $i = n-r$ à n , sont telles que $E(|W_i|^2) = 0$. Elles sont donc, presque sûrement, nulles.

Interprétons ce résultat en revenant au vecteur \vec{Z} .

La matrice Γ_Z , qui est de même rang que Γ_W , est de rang $n - r$.

Comme $\vec{W} = P^{-1} \vec{Z}$ cela signifie qu'il existe $n - r$ relations linéaires entre les variables Z_i de la forme :

$$\sum_{j=1}^n \alpha_{i,j} Z_j = W_i = 0 \quad \text{pour } i = n-r+1 \text{ à } n$$

$\alpha_{i,j}$ étant les composantes de la matrice P^{-1} .

Notons inversement que, si Γ_Z est de rang n , nous avons démontré qu'on peut toujours par une transformation linéaire non singulière transformer un ensemble de n variables aléatoires non liées par une relation linéaire en n variables aléatoires orthogonales.

8.6 Application aux vecteurs réels gaussiens

8.6.1 Définitions

Soit \vec{X} un vecteur réel gaussien d'espérance \vec{m} et soit

$$C_X = E\left((\vec{X} - \vec{m})(\vec{X} - \vec{m})^T\right) \quad \text{sa matrice de variance-covariance de composantes :}$$

$$c_{i,j} = E((X_i - m_i)(X_j - m_j)) = E(X_i X_j) - E(X_i)E(X_j)$$

$$c_{i,j} = r_{i,j} \sigma_i \sigma_j \quad \text{pour } i \neq j$$

$$c_{i,i} = \sigma_i^2 \quad \text{pour } i = j$$

$$\sigma_i^2 \quad \text{variance de } X_i$$

$$r_{i,j} \quad \text{coefficient de corrélation de } X_i, X_j$$

\vec{X} est un vecteur réel gaussien si :

- Sa fonction de densité est de la forme

$$f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} |C_X|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (\vec{X} - \vec{m})^T C_X^{-1} (\vec{X} - \vec{m})\right) \quad 12-31$$

avec : $\vec{X} \in R^n$

C_X^{-1} matrice inverse de C_X

- ou si sa fonction caractéristique est de la forme

$$\varphi_{X_1, X_2, \dots, X_n}(u_1, u_2, \dots, u_n) = E\left(e^{j\vec{U}^T \vec{X}}\right) = \exp\left(j\vec{U}^T \vec{m} - \frac{1}{2}\vec{U}^T C_X \vec{U}\right)$$

avec \vec{U} un vecteur certain $\in R^n$

Exemple :

Appliquons les formules précédentes au cas à deux dimensions. Pour retrouver les notations habituelles posons

$$X = X_1 \text{ et } Y = X_2$$

$$\vec{X} = \begin{bmatrix} X \\ Y \end{bmatrix} \quad E(\vec{X}) = \begin{bmatrix} m_X \\ m_Y \end{bmatrix}$$

$$C = \begin{bmatrix} \sigma_X^2 & r\sigma_X\sigma_Y \\ r\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$$

$$|C| = (1 - r^2)\sigma_X^2\sigma_Y^2$$

En résolvant l'équation $CC^{-1} = I$ on trouve sans difficulté l'expression de C^{-1}

$$C^{-1} = \frac{1}{(1 - r^2)\sigma_X^2\sigma_Y^2} \begin{bmatrix} \sigma_X^2 & -r\sigma_X\sigma_Y \\ -r\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$$

et en remplaçant dans l'expression de la densité on obtient

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-r^2}} \exp\left(-\frac{1}{2(1-r^2)}\left(\frac{(x-m_X)^2}{\sigma_X^2} - \frac{2(x-m_X)(y-m_Y)}{2\sigma_X\sigma_Y} + \frac{(y-m_Y)^2}{\sigma_Y^2}\right)\right)$$

On retrouve l'expression (9-22)

le calcul de la fonction caractéristique est immédiat avec $\vec{U} = \begin{bmatrix} u \\ v \end{bmatrix}$

$$\varphi_{X,Y}(u,v) = \exp\left(j(um_X + vm_Y) - \frac{1}{2}(\sigma_X^2 u^2 + 2r\sigma_X\sigma_Y uv + \sigma_Y^2 v^2)\right)$$

8.6.2 Propriétés d'un vecteur réel gaussien

Propriété 1 : si les composantes X_i du vecteur \vec{X} sont mutuellement décorréées, elles ont indépendantes.

Si les X_i sont mutuellement décorréées, on a $c_{i,j} = 0 \quad \forall i \neq j$

donc C est diagonale et son inverse C^{-1} l'est également.

$$C^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \frac{1}{\sigma_n^2} \end{bmatrix}$$

En remplaçant dans (12-32) on vérifie que la densité n dimensionnelle s'écrit :

$$f_{X_1 \dots X_n}(x_1, x_2, \dots, x_n) = \prod_{i=1}^n \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(x_i - m_i)^2}{2\sigma_i^2}} = \prod_{i=1}^n f_{X_i}(x_i)$$

Elle est égale au produit des densités marginales des X_i donc les composantes du vecteur \vec{X} sont indépendantes dans leur ensemble.

Propriété 2:

Toutes les lois déduites de $f_{X_1 \dots X_n}(x_1, x_2, \dots, x_n)$ et d'ordre inférieur à n sont gaussiennes.

On obtient la fonction caractéristique correspondant à un sous-ensemble de dimension r ($r < n$) des variables en annulant les composantes u_i correspondantes.

La fonction caractéristique des variables restantes est encore de forme gaussienne, d'où la propriété.

Exemple

$$\phi_{X_1, X_2}(u_1, u_2) = E(e^{j(u_1 x_1 + u_2 x_2)})$$

$$\phi_{X_1, X_2}(u_1, u_2) = \phi_{X_1, X_2, \dots, X_n}(u_1, u_2, 0, \dots, 0)$$

$$\phi_{X_1, X_2}(u_1, u_2) = \exp(j(u_1 m_1 + u_2 m_2)) \exp\left(-\frac{1}{2} [u_1, u_2] \begin{bmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}\right)$$

Propriété 3:

Le caractère gaussien est conservé après une transformation linéaire non singulière.

Soit \vec{X} un vecteur gaussien, de moyenne \vec{m} et de matrice de variance-covariance C_X

On considère la transformation linéaire non singulière qui fait correspondre à \vec{X} le vecteur \vec{Y} (de même dimension que \vec{X}) défini par : $\vec{Y} = A\vec{X}$ où A est une matrice réelle ($n.n$) de déterminant non nul.

Calculons la fonction caractéristique de \vec{Y}

$$\phi_{Y_1, Y_2, \dots, Y_n}(u_1, u_2, \dots, u_n) = E(e^{j\vec{U}^T \vec{Y}}) = E(e^{j\vec{U}^T A \vec{X}})$$

Soit \vec{V} le vecteur réel certain tel que $\vec{V}^T = \vec{U}^T A$ et donc $\vec{V} = A^T \vec{U}$

\vec{X} étant gaussien, on peut écrire

$$E\left(e^{j\vec{V}^T \vec{X}}\right) = \exp\left(j\vec{V}^T \vec{m}_X - \frac{1}{2} \vec{V}^T C_X \vec{V}\right)$$

$$\phi_{X_1, X_2, \dots, X_n}(u_1, u_2, \dots, u_n) = E\left(e^{j\vec{U}^T \vec{X}}\right) = \exp\left(j\vec{U}^T \vec{m} - \frac{1}{2} \vec{U}^T C_X \vec{U}\right)$$

soit pour la fonction caractéristique de \vec{Y}

$$\phi_{X_1, X_2, \dots, X_n}(u_1, u_2, \dots, u_n) = \exp\left(j\vec{U}^T A \vec{m}_X - \frac{1}{2} \vec{U}^T A C_X A^T \vec{U}\right)$$

Expression qui démontre que le vecteur \vec{Y} est gaussien

- de moyenne $E(\vec{Y}) = A E(\vec{X}) = A \vec{m}_x$
- de matrice de variance covariance $C_Y = A C_X A^T$

Propriété 4 :

On peut toujours, par une transformation linéaire non singulière passer d'un ensemble de n variables aléatoires gaussiennes, qui ne sont pas linéairement dépendantes, à un ensemble de n nouvelles variables aléatoires gaussiennes statistiquement indépendantes.

Cette propriété se déduit immédiatement de l'hermiticité de la matrice de covariance et de la forme particulière de la densité n dimensionnelle.

Nous avons vu en XII.7.3 qu'on pouvait, par une transformation linéaire non singulière, diagonaliser la matrice de covariance C_X et donc son inverse C_X^{-1} . Si, de plus, les variables X_i sont gaussiennes dans leur ensemble, alors elles sont indépendantes (propriété 3), d'où la propriété 4.

Notons pour terminer que la loi de probabilité d'un vecteur gaussien est entièrement déterminée si on connaît ses statistiques du premier et du second ordre.

8.6.3 Vecteur complexe gaussien

Soit $\vec{Z}^T = (Z_1 Z_2, \dots, Z_n)$ un vecteur complexe de composantes Z_i

$$Z_i = X_i + jY_i \quad i = 1, n$$

Pour décrire les propriétés statistiques du vecteur $Z^T = (Z_1 Z_2, \dots, Z_n)$ il faut se donner la loi de densité $f_{X_1 \dots X_n}(x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n)$ **des 2 n variables aléatoires qui le composent.**

En toute généralité, le vecteur \vec{Z} sera un vecteur gaussien si sa densité à 2 n dimensions est de forme gaussienne.

Pour simplifier les notations, nous supposons que \vec{Z} est un vecteur centré et nous poserons

$$\vec{W}^T = (X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_n)$$

$$w^T = (x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n)$$

$$\vec{k}^T = (u_1, u_2, \dots, v_1, v_2, \dots, v_n)$$

la matrice de covariance de \vec{W} s'écrit

$$C_W = E(WW^T) = \begin{bmatrix} C_X & C_{XY} \\ C_{YX} & C_Y \end{bmatrix}$$

avec

$$C_X = E\left(\vec{X}\vec{X}^T\right) \text{ matrice de covariance de } \vec{X}$$

$$C_Y = E\left(\vec{Y}\vec{Y}^T\right) \text{ matrice de covariance de } \vec{Y}$$

$$C_{XY} = E\left(\vec{X}\vec{Y}^T\right) \text{ matrice de covariance de } \vec{X} \text{ et } \vec{Y}$$

$$C_{YX} = E\left(\vec{Y}\vec{X}^T\right) \text{ matrice de covariance de } \vec{Y} \text{ et } \vec{X}$$

si \vec{Z} est gaussien centré, sa fonction de densité (d'ordre $2n$) s'écrit d'après (12-32)

$$f_{\vec{Z}}(\vec{z}) = f_{\vec{W}}(\vec{w}) = \frac{1}{(2\pi)^n |C_{\vec{W}}|^{1/2}} \exp\left(-\frac{1}{2} \vec{W}^T C_{\vec{W}}^{-1} \vec{W}\right) \quad 12-33$$

et sa fonction caractéristique

$$\varphi_{\vec{Z}}(\vec{z}) = \varphi_{\vec{W}}(\vec{k}) = \exp\left(-\frac{1}{2} (\vec{k}^T C_{\vec{W}}^{-1} \vec{k})\right) \quad 12-34$$

Remarque importante : on notera que la matrice de covariance utilisée n'est pas la matrice hermitienne C_Z introduite précédemment (12-30).

En effet, cette matrice à coefficients complexes est de dimension n .

Elle ne permet donc pas de représenter la loi d'un vecteur gaussien complexe de C^n dans toute sa généralité.

Exemple :

Prenons le cas le plus simple où \vec{Z} n'a qu'une composante.

$$\vec{Z} = Z_1 = X + jY$$

$$C_w \text{ est de la forme } \begin{bmatrix} \sigma_X^2 & r\sigma_X\sigma_Y \\ r\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$$

Par contre, la matrice de covariance de \vec{Z} se réduit à un seul terme

$$C_Z = \sigma_X^2 + \sigma_Y^2$$

Cependant, dans les méthodes modernes de traitement du signal, on utilise les notations suivantes pour décrire la loi de probabilité et la fonction caractéristique d'un vecteur gaussien complexe centré

$$f_{\vec{Z}}(\vec{z}) = \frac{1}{\sqrt{(2\pi)^{2n} |C_{\vec{Z}}|}} \exp\left(-\frac{\vec{z}^+ C_{\vec{Z}}^{-1} \vec{z}}{2}\right) \quad 12-35$$

$$\varphi_{\vec{Z}}(\vec{\Omega}) = \exp\left(-\frac{\vec{\Omega}^+ C_{\vec{Z}} \vec{\Omega}}{2}\right) \quad \text{avec } \vec{\Omega} = \vec{U} + j\vec{V}$$

Il est clair que ces expressions ne peuvent convenir qu'en imposant certaines propriétés statistiques au vecteur \vec{Z}

Ces conditions sont les suivantes (théorème de Goodman) $C_{XX} = C_{YY}$
 $C_{XY} = -C_{YX}$

Les vecteurs \vec{X} et \vec{Y} parties réelles et imaginaires de \vec{Z} ont la même matrice de covariance et leurs matrices d'intercorrélation sont de signes contraires.

Plus concrètement, les variables étant supposées centrées,

$$E(X_i^2) = E(Y_i^2) = \sigma_i^2 \quad 12-36$$

$$E(X_i Y_i) = 0$$

et pour $i \neq j$

$$E(X_i X_j) = E(Y_i Y_j) = r_{ij} \sigma_i \sigma_j$$

$$E(X_i Y_j) = -E(X_j Y_i)$$

Ces relations sont très importantes. L'expression (12-35) est utilisée pour caractériser les propriétés statistiques du bruit thermique à bande étroite que l'on retrouve dans tous les problèmes de "communication".

8.7 Exercices

8.7.1 Énoncés

1) On considère les variables aléatoires X_1, X_2, X_3 indépendantes

-Démontrer que : $E(X_1^k X_2^m X_3^n) = E(X_1^k) E(X_2^m) E(X_3^n)$

-Ces trois variables sont à densité uniforme sur l'intervalle $(-0,5 ; +0,5)$.

Calculer les quatre premiers moments de la variable $Z = X_1 + X_2 + X_3$

2) La variable z est définie par $Z = \sum_{i=1}^n X_i$

Toutes les variables X_i ayant la même loi

-Trouver une expression pour la variance de Z en fonction de σ_X^2 , la variance des X_i , n et le coefficient de corrélation r en admettant que :

$$r_{X_i Y_j} = r \quad \forall i, j \quad i \neq j$$

-Dans le cas où les variables X_i sont indépendantes et de type (0,1)

$$P(X_i = 0) = p \quad P(X_i = 1) = 1 - p = q$$

Déterminer la fonction caractéristique de Z, en déduire la densité de Z.

3) La variable X peut prendre les valeurs + ou - 1 avec la même probabilité. Pour déterminer expérimentalement l'espérance de X on réalise n tirages indépendants de cette variable et on prend comme estimateur:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{avec} \quad X_i = \pm 1$$

Déterminer la valeur minimale de n , en utilisant l'inégalité de Tchebychev, pour avoir une probabilité supérieure à 0,9 que la valeur trouvée approche $E(X)$ à $\pm 0,1$ près.

4) Les trois composantes V_x, V_y, V_z de la vitesse V d'une particule sont des variables aléatoires indépendantes gaussiennes centrées et de variance σ^2

$$\sigma^2 = \frac{KT}{m} \quad (\text{théorie cinétique}).$$

1. Calculer la densité de la variable $V = \sqrt{V_x^2 + V_y^2 + V_z^2}$ (passer en coordonnées sphériques).
2. Calculer $E(V)$, $E(V^2)$, et la valeur la plus probable V_m .
3. Calculer la densité de l'énergie de la particule $W = \frac{1}{2} m V^2$
4. Calculer $E(W)$, σ_W^2

CHAPITRE 9. Notion de convergence, loi des grands nombres et théorème de la limite centrale

Nous avons introduit trois définitions de la probabilité au début de ce cours :

- Par dénombrement
- Par fréquence relative
- Axiomatique

Toute notre étude s'est développée en s'appuyant sur la définition axiomatique qui nous a permis de calculer des probabilités d'événements (ou des densités de probabilité) de manière cohérente à partir de probabilités a priori, connues ou supposées connues

La définition par fréquence relative présentait à la fois un avantage et un inconvénient. L'avantage, c'était précisément de permettre la mesure de la probabilité a priori d'un événement. L'inconvénient, c'était l'imprécision de cette définition qui faisait intervenir une notion de limite.

Pour lever cette imprécision, nous allons introduire la notion de convergence pour une suite de variables aléatoires.

C'est l'étude de la convergence qui va nous indiquer la relation existant entre la définition a priori (par dénombrement) de la probabilité d'un événement et la fréquence relative de cet événement dans une expérience constituée d'un grand nombre d'épreuves.

C'est là que se fait la liaison entre les probabilités et la statistique.

9.1 Convergence d'une suite de variables aléatoires.

9.1.1 Définitions

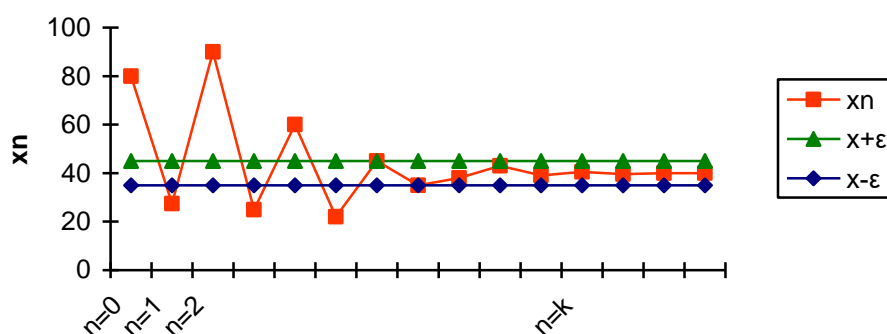
En analyse classique, on appelle suite une séquence ordonnée de nombres : $x_1, x_2, \dots, x_n, \dots$

On dit que la suite de terme général x_n tend vers une limite x quand n tend vers l'infini si :

$$\forall \varepsilon > 0, \exists N, n > N \Rightarrow |x_n - x| < \varepsilon$$

On dit alors que la suite converge et qu'elle diverge dans le cas contraire.

figure 13.1



Cas d'une variable aléatoire

Considérons maintenant une suite de variables aléatoires $X_1, X_2, \dots, X_n, \dots$

Chaque variable est susceptible de prendre un certain nombre de valeurs qui dépendent du résultat de l'épreuve.

Supposons, pour simplifier, que chaque variable puisse prendre k valeurs discrètes.

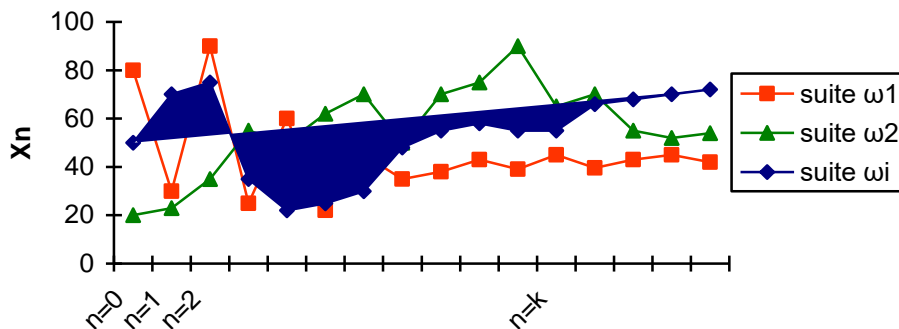
La variable de rang i , X_i , peut prendre k valeurs suivant le résultat : $X_i = x_{i,j} \quad j = 1, k$

On peut alors former k^n suites de nombres possibles pour une suite de n variables aléatoires.

Une suite particulière $x_n(\omega_l)$, ($l = 1$ à k^n) sera représentée par une séquence ordonnée de nombres $x_{i,j}$ pour i variant de 1 à n : j étant pour chaque i un nombre compris entre 1 et k .

Par exemple : $x_{1,3}, x_{2,5}, x_{3,9}, \dots, x_{n,5}$

figure 13.2



9.1.2 Différents modes de convergence

Nous examinons maintenant les divers modes de convergence possibles en allant du plus restrictif au moins restrictif.

9.1.2.1 Convergence partout

Il est évident que la limite d'une suite de variables aléatoires dépend du résultat composé. Cela signifie simplement que la limite est une variable aléatoire. Pour chaque résultat composé ω_k on a, quand n tend vers l'infini,

$$\left. \begin{array}{l} X_n(\omega_1) \rightarrow X(\omega_1) \\ X_n(\omega_2) \rightarrow X(\omega_2) \\ \dots\dots\dots \\ X_n(\omega_l) \rightarrow X(\omega_l) \\ \dots\dots\dots \end{array} \right\} \text{valeurs de la variable } X$$

Nous dirons que la suite de variables aléatoires converge partout si :

$$\forall \omega, \lim_{n \rightarrow \infty} |X_n(\omega) - X(\omega)| = 0 \quad 13-1$$

En d'autres termes, quelle que soit la suite numérique formée à partir des valeurs prises par les variables aléatoires X_n , cette suite est convergente.

Cette convergence de toutes les suites est en général trop stricte et nous étudions maintenant d'autres modes de convergence.

9.1.2.2 Convergence presque partout

On dit encore convergence presque sûre ou avec probabilité 1. La suite X_n tend vers X presque partout si :

$$P(\lim_{n \rightarrow \infty} (X_n - X) = 0) = 1 \quad 13-2$$

En d'autres termes, quel que soit le résultat ω considéré, la probabilité de l'événement suivant est égale à 1 :

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$$

Remarque : si cette condition est réalisée, cela ne signifie pas pour autant que l'ensemble des résultats ω donnant une suite divergente est l'ensemble vide. C'est d'ailleurs ce qui distingue ce mode de convergence de la convergence partout.

Par exemple, nous avons vu qu'avec une variable continue les événements du type $\{X=x\}$ ont une probabilité nulle. Or, l'ensemble qui correspond à ces événements n'est pas vide puisqu'il contient tous les éléments x .

En d'autres termes, il est possible qu'une suite particulière $(\omega_p)_p$ soit divergente. Sa probabilité de tirage est alors nulle.

9.1.2.3 Convergence en moyenne quadratique

C'est un des modes les plus utilisés en théorie, car il permet de revenir aux notions classiques de convergence.

La suite X_n converge vers X en moyenne quadratique si :

$$\lim_{n \rightarrow \infty} E((X_n - X)^2) = 0 \quad 13-3$$

Plus précisément si :

$$\forall \varepsilon > 0, \exists N(\varepsilon), n > N \Rightarrow E((X_n - X)^2) < \varepsilon \quad 13-4$$

9.1.2.4 Convergence en probabilité

La suite de variables aléatoires X_n converge en probabilité vers la variable X si :

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P_n = \lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1 \quad 13-5$$

Plus précisément :

$$\forall \varepsilon, \eta > 0, \exists N(\varepsilon, \eta), n > N \Rightarrow P(|X_n - X| < \varepsilon) \geq 1 - \eta \quad 13-6$$

Ce qui s'écrit encore, en considérant l'événement complémentaire,

$$\forall \varepsilon, \eta > 0, \exists N(\varepsilon, \eta), n > N \Rightarrow P(|X_n - X| \geq \varepsilon) \leq \eta \quad 13-7$$

Il ne faut pas confondre cette convergence en probabilité avec la convergence presque partout. On prend ici la limite de la probabilité alors que dans la convergence presque partout on prenait la probabilité de la limite.

9.1.2.5 Convergence en loi

Ce mode de convergence est le moins restrictif de tous. On désigne par $F_{X_n}(x)$ la fonction de répartition de la variable aléatoire X_n et $F_X(x)$ par celle de la variable aléatoire X . On dit que la suite X_n tend vers X en loi si :

$$\forall x \quad \lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x) \quad 13-8$$

De plus, on peut démontrer que, $g(x)$ étant une fonction continue, si $F_{X_n}(x)$ converge vers $F_X(x)$, la suite $\int g(x) dF_n(x)$ converge vers $\int g(x) dF_X(x)$ et réciproquement.

En particulier si

nous aurons la convergence des fonctions caractéristiques

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x) \Leftrightarrow \varphi_{X_n}(t) \xrightarrow{n \rightarrow \infty} \varphi_X(t)$$

13-9

C'est un critère commode que nous utiliserons dans les paragraphes suivants.

9.1.2.6 Comparaison des différents modes de convergence

En utilisant les définitions on peut comparer les différents modes de convergence exposés. Les résultats de ces comparaisons apparaissent dans la figure ci-dessous :

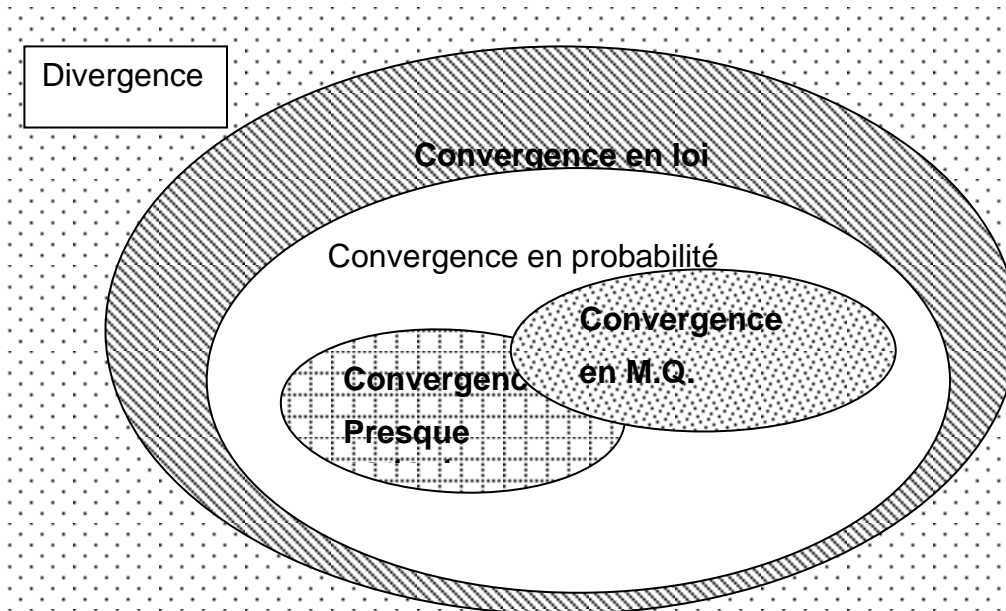


Figure 13.3

Ces résultats sont difficiles à établir, sauf dans le cas de la convergence quadratique qui entraîne la convergence en probabilité. La démonstration repose sur l'inégalité de Tchebychev. Pour la variable $X_n - X$, on peut écrire :

$$P(|X_n - X| > \varepsilon) \leq \frac{E((X_n - X)^2)}{\varepsilon^2}$$

S'il y a convergence en moyenne quadratique, on pourra toujours trouver un n qui rende $E((X_n - X)^2) < \varepsilon^2 \eta$, d'où le résultat.

9.2 Exemples de convergence en loi

9.2.1 Approximation à la loi binomiale.

On considère une expérience avec un événement A ayant la probabilité p de se produire. On lui associe la variable X (variable indicatrice) par l'application :

$$\begin{cases} X = 1 & \text{si } A \text{ est réalisé} \\ X = 0 & \text{si non} \end{cases}$$

Calculons l'espérance et la variance de cette variable

$$E(X) = 0 \cdot q + 1 \cdot p = p \quad \text{avec } q = 1 - p$$

$$E(X^2) = 0^2 \cdot q + 1^2 \cdot p = p$$

$$\sigma_X^2 = p - p^2 = p(1 - p) = pq$$

Sa fonction caractéristique est donnée par :

$$\phi_X(t) = pe^{jt} + q$$

On répète n fois la même expérience de manière indépendante et on considère la variable Z_n définie par :

$$Z_n = \sum_{i=1}^n X_i$$

On a alors d'après (12.22) :

$$E(Z_n) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = np$$

$$\text{Var}(Z_n) = \sigma_{Z_n}^2 = \sum_{i=1}^n \text{Var}(X_i) = npq$$

Sa fonction caractéristique est donnée par :

$$\phi_{Z_n}(t) = (pe^{jt} + q)^n$$

La variable Z_n ainsi construite représente le nombre de fois où l'événement A s'est produit en n épreuves. Elle suit donc une loi binomiale avec :

$$\forall k, 0 \leq k \leq n, \quad P(Z_n = k) = C_n^k p^k q^{n-k}$$

1) Tendance vers la loi de Poisson

On veut maintenant trouver une approximation de la loi binomiale dans le cas où n devient grand et p petit, de telle sorte que le produit np reste fini. On prendra pour fixer les idées :

$$1 < np < 10$$

Ce type d'approximation est intéressant car le coefficient C_n^k est parfois difficile à calculer. Le réel λ étant fixé, posons :

$$p = \frac{\lambda}{n}$$

Avec ces notations, la fonction caractéristique s'écrit :

$$\phi_{Z_n}(t) = (pe^{jt} + q)^n = \left(\frac{\lambda}{n}e^{jt} + 1 - \frac{\lambda}{n}\right)^n = \left(\frac{\lambda}{n}(e^{jt} - 1) + 1\right)^n$$

Cherchons la limite de cette expression quand n tend vers l'infini. On a :

$$\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n = e^x$$

Par conséquent :

$$\lim_{n \rightarrow \infty} \phi_{Z_n}(t) = \exp(\lambda(e^{jt} - 1)) = \phi_Z(t)$$

Ce qui n'est autre que la fonction caractéristique de la loi de Poisson de paramètre λ :

$$P(Z = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad k = 0, 1, \dots$$

Comme la convergence des fonctions caractéristiques entraîne celle des fonctions de répartition (voir le paragraphe 9.1.2.5), nous avons là un exemple de convergence en loi.

La suite des variables aléatoires Z_n distribuées suivant une loi binomiale et qui tend vers une variable Z distribuée suivant une loi de Poisson.

Pratiquement, on retiendra que l'on peut approximer la loi binomiale par la loi de Poisson de paramètre $n \cdot p$ si n est grand et p petit avec : $1 < np < 10$.

2) Tendance vers la loi normale

Nous considérons à nouveau la variable Z_n distribuée suivant la loi binomiale et nous allons chercher une approximation en loi pour Z_n distribuée quand n tend vers l'infini mais sans faire d'hypothèse sur p .

Quand n tend vers l'infini, la moyenne $n \cdot p$, et la variance $n \cdot p \cdot q$ de Z_n tendent vers l'infini pour p fixé. Pour avoir une expression plus maniable nous utiliserons la variable réduite U_n :

$$U_n = \frac{Z_n - E(Z_n)}{\sigma_{Z_n}} = \frac{Z_n - n \cdot p}{\sqrt{n \cdot p \cdot q}}$$

Calculons la fonction caractéristique de U_n :

$$\phi_{U_n}(t) = E(e^{j \cdot u \cdot t}) = E(e^{j \cdot \frac{Z_n - n \cdot p}{\sqrt{n \cdot p \cdot q}} \cdot t}) = e^{-\frac{j \cdot n \cdot p \cdot t}{\sqrt{n \cdot p \cdot q}}} E(e^{\frac{j \cdot Z_n \cdot t}{\sqrt{n \cdot p \cdot q}}}) = e^{-\frac{j \cdot n \cdot p \cdot t}{\sqrt{n \cdot p \cdot q}}} \phi_{Z_n}\left(\frac{t}{\sqrt{n \cdot p \cdot q}}\right)$$

En remplaçant $\phi_{Z_n}(t)$ par sa valeur :

$$\phi_{U_n}(t) = e^{-\frac{j \cdot n \cdot p \cdot t}{\sqrt{n \cdot p \cdot q}}} (p \cdot e^{\frac{j \cdot t}{\sqrt{n \cdot p \cdot q}}} + q)^n$$

Prenons le logarithme de cette expression

$$\text{Log}(\phi_{U_n}(t)) = -\frac{j \cdot n \cdot p \cdot t}{\sqrt{n \cdot p \cdot q}} + n \cdot \text{Log}(p \cdot e^{\frac{j \cdot t}{\sqrt{n \cdot p \cdot q}}} + q)$$

Développons l'exponentielle :

$$e^{\frac{j \cdot t}{\sqrt{n \cdot p \cdot q}}} = 1 + \frac{j \cdot t}{\sqrt{n \cdot p \cdot q}} - \frac{t^2}{2 \cdot n \cdot p \cdot q} + \dots$$

$$\text{Log}(\phi_{U_n}(t)) = -j \frac{np}{\sqrt{npq}} t + n \text{Log}\left(1 + \frac{jp}{\sqrt{npq}} t - \frac{t^2}{2nq} + \dots\right)$$

En développant le logarithme au second ordre en $\frac{1}{\sqrt{n}}$:

$$\text{Log}(\phi_{U_n}(t)) = -\frac{t^2}{2} + \varepsilon\left(\frac{t^3}{\sqrt{n}}\right)$$

Dans cette équation, le terme $\varepsilon\left(\frac{t^3}{\sqrt{n}}\right)$ tend vers 0 quand n tend vers l'infini. En définitive :

$$\phi_{U_n}(t) \xrightarrow{n \rightarrow \infty} e^{-\frac{t^2}{2}}$$

On reconnaît la fonction caractéristique d'une loi normale réduite, ce qui revient à dire que la loi de Z_n tend vers une loi normale, de moyenne $n \cdot p$ et d'écart type $\sqrt{n \cdot p \cdot q}$.

Remarque : ce résultat n'est pas contradictoire avec l'approximation précédente. Il se trouve que si p est petit, il faut considérer de très grandes valeurs de n pour que l'approximation par la loi normale soit suffisamment précise. De plus, nous allons voir maintenant que la loi de Poisson tend également vers une loi normale quand n vers l'infini.

9.3 Généralisation : théorème de la limite centrale (ou de Laplace-Liapounoff)

Le résultat précédent peut être considéré comme un cas particulier d'un théorème beaucoup plus général.

On considère une suite de variables aléatoires indépendantes $X_1, X_2, \dots, X_n, \dots$ pourvues de densités respectives : $f_{X_1}(x_1), f_{X_2}(x_2), \dots, f_{X_i}(x_i), \dots, f_{X_n}(x_n)$

On suppose l'existence de $E(X_i)=m_i$ et de $\text{Var}(X_i)=\sigma_i^2$ et on forme la somme : $Z_n = \sum_{i=1}^n X_i$

On sait que dans ce cas :

$$E(Z_n) = m_{Z_n} = \sum_{i=1}^n m_i$$

$$\text{Var}(Z_n) = \sigma_{Z_n}^2 = \sum_{i=1}^n \sigma_i^2$$

Le théorème de la limite centrale nous dit que, sous certaines conditions générales, la densité de la variable Z_n tend vers une loi normale de paramètre m_{Z_n} et $\sigma_{Z_n}^2$, soit :

$$\lim_{n \rightarrow \infty} f_{Z_n}(z) = \frac{1}{\sigma_{Z_n} \sqrt{2\pi}} \exp\left(-\frac{(z - m_{Z_n})^2}{2\sigma_{Z_n}^2}\right)$$

Nous démontrons ce théorème, dans un cas particulier, en considérant n variables aléatoires identiques, indépendantes et de même loi. Alors :

$$\forall i, \quad E(X_i) = m, \quad \text{Var}(X_i) = \sigma^2$$

$$E(Z_n) = n \cdot m \quad \text{Var}(Z_n) = n \cdot \sigma^2$$

On forme la variable réduite :

$$U_n = \frac{Z_n - E(Z_n)}{\sigma_{Z_n}} = \frac{Z_n - n \cdot m}{\sigma \cdot \sqrt{n}}$$

La fonction caractéristique de U_n s'écrit :

$$\phi_{U_n}(t) = e^{\frac{-j \cdot n \cdot m \cdot t}{\sigma \cdot \sqrt{n}}} \phi_{Z_n}\left(\frac{t}{\sigma \cdot \sqrt{n}}\right)$$

La variable Z_n étant la somme de n variables indépendantes et de même loi, on a :

$$\phi_{Z_n}\left(\frac{t}{\sigma \cdot \sqrt{n}}\right) = \left(\phi_X\left(\frac{t}{\sigma \cdot \sqrt{n}}\right)\right)^n$$

Développons $\phi_X\left(\frac{t}{\sigma \cdot \sqrt{n}}\right)$ (chapitre 5.4 équation 7-8) :

$$\phi_X\left(\frac{t}{\sigma \cdot \sqrt{n}}\right) = 1 + j \frac{mt}{\sigma \sqrt{n}} - \frac{(m\sigma)^2 t^2}{2n\sigma^2} + \dots$$

$$\text{Log}(\phi_{U_n}(t)) = -j \frac{nm}{\sigma \sqrt{n}} t + n \text{Log}\left(1 + \frac{jm}{\sigma \sqrt{n}} t - \frac{(m2 + \sigma2)t^2}{2n\sigma2} + \dots\right)$$

En développant le logarithme au second ordre en $\frac{t}{\sqrt{n}}$, on obtient :

$$\text{Log}(\phi_{U_n}(t)) = -\frac{t^2}{2} + \varepsilon\left(\frac{t^3}{\sqrt{n}}\right)$$

Lorsque n tend vers l'infini, on reconnaît le logarithme de la fonction caractéristique de la loi normale réduite.

Entre autres conditions, une condition nécessaire pour que le théorème de la limite centrale s'applique est : $\int x^\alpha f_{X_i}(x) dx < Cste$ pour $\alpha < 2$

Dans le cas de la démonstration précédente, il suffit, lorsque toutes les variables ont la même loi, que $E(X_i) = m$ et $\text{Var}(X_i) = \sigma^2$ existent.

Par exemple, supposons que les X_i suivent une loi de Cauchy

$$f_{X_i} = \frac{1}{\pi} \cdot \frac{1}{1+x^2} \quad x \in \mathbb{R}$$

la fonction caractéristique est de la forme $\phi_{X_i}(t) = e^{-|t|} \quad t \in \mathbb{R}$

la variable $Z_n = \sum_{i=1}^n X_i$ a pour fonction caractéristique, compte tenu de l'indépendance des X_i

$$\phi_{Z_n}(t) = (\phi_{X_i}(t))^n = e^{-n|t|} \quad t \in \mathbb{R}$$

Déterminons la densité de Z_n par inversion de la fonction caractéristique (équation 7-6)

$$f_{Z_n}(z) = \frac{1}{2\pi} \int \phi_{Z_n}(t) e^{jzt} dt$$

$$f_{Z_n}(z) = \frac{1}{2\pi} \int e^{-n|t|} e^{jzt} dt$$

$$f_{Z_n}(z) = \frac{1}{2\pi} \left(\int_{-\infty}^0 e^{(n+jz)t} dt + \int_0^{+\infty} e^{(-n+jz)t} dt \right)$$

$$f_{Z_n}(z) = \frac{1}{n\pi} \frac{1}{1 + \left(\frac{z}{n}\right)^2} \quad z \in \mathbb{R}$$

C'est encore une loi de Cauchy donc le théorème de la limite centrale ne s'applique pas car $\int x^\alpha f_{X_i}(x) dx$ diverge pour $\alpha \geq 1$

9.4 Autre exemple de convergence : théorème de Bernoulli, loi des grands nombres

9.4.1 Théorème de Bernoulli

Reprenons la variable indicatrice du paragraphe 9.2.1

$$X_i = 1 \quad \text{si } A \text{ est réalisé avec } P(A) = p$$

$$X_i = 0 \quad \text{si non} \quad \text{avec } P(\bar{A}) = q = 1 - p$$

Après n épreuves indépendantes, la variable somme $Z_n = \sum_{i=1}^n X_i$ suit une loi binomiale avec $E(Z_n) = n \cdot p$ et $\text{Var}(Z_n) = n \cdot p \cdot q$. Considérons la variable :

$$U_n = \frac{Z_n}{n}$$

On a alors :

$$E(U_n) = p, \text{Var}(U_n) = \frac{pq}{n}$$

La variable Z_n prend la valeur k si l'événement A s'est produit k fois au cours des n épreuves et U_n prend la valeur : $\frac{k}{n}$. Le rapport $\frac{k}{n}$ représente ainsi la fréquence relative de l'événement A .

On voudrait maintenant savoir si la fréquence relative est susceptible de représenter la probabilité p de l'événement A observé.

C'est l'objet du théorème de Bernoulli qui dit que la fréquence relative converge en probabilité vers $P(A)$, la probabilité a priori de A , quand le nombre d'épreuves tend vers l'infini.

En utilisant l'inégalité de Tchebychev, on peut écrire :

$$P\{|U_n - E(U_n)| > \varepsilon\} \leq \frac{\sigma_{U_n}^2}{\varepsilon^2}$$

$$P\{|U_n - p| > \varepsilon\} \leq \frac{pq}{n\varepsilon^2}$$

Cette expression montre que la probabilité pour que la variable aléatoire U_n s'écarte de $p=p(A)$ de plus de ε tend vers 0 quand n tend vers l'infini

Exemple : Combien faut-il faire d'épreuves avec une pièce symétrique ($p = q = 0,5$) pour que la fréquence relative de l'événement pile soit comprise entre 0,4 et 0,6 avec une probabilité de 0,9 ?

L'inégalité de Tchebychev donne :

$$P(|U_n - 0.5| > \varepsilon) \leq \frac{100}{4n} \leq 0.1$$

$$d'où \quad n \geq 250$$

9.4.2 Loi des grands nombres

La loi faible des grands nombres est une généralisation du théorème de Bernoulli.

On considère la variable \bar{X} , moyenne arithmétique de n variables X_i décorrélées et de même loi.

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

Nous avons vu que sous ces hypothèses :

$$E(\bar{X}) = E(X_i) = m$$

$$\text{Var}(\bar{X}) = \frac{\text{Var}(X_i)}{n} = \frac{\sigma_X^2}{n}$$

L'inégalité de Tchebychev donne : $\lim_{n \rightarrow \infty} P(|\bar{X} - m| > \varepsilon) = 0$ ou, ce qui est équivalent :

$$\lim_{n \rightarrow \infty} P(|\bar{X} - m| \leq \varepsilon) = 1$$

La moyenne arithmétique des X_i converge en probabilité vers leur espérance mathématique.

Remarquons que nous avons également la convergence en moyenne quadratique :

$$E((\bar{X} - m)^2) = \frac{\sigma_X^2}{n} \xrightarrow{n \rightarrow \infty} 0$$

La loi forte des grands nombres (Borel) fournit le même résultat mais en considérant une convergence presque sûre :

$$P(\lim_{n \rightarrow \infty} |\bar{X} - m| = 0) = 1$$

9.5 Exercices

9.5.1 Énoncés

1) Z étant une variable aléatoire quelconque, démontrer que

$$P(|Z| > \varepsilon) \leq \frac{E(Z^2)}{\varepsilon^2}$$

En déduire que la convergence en M.Q. entraîne la convergence en probabilité.

2) La variable X compte le nombre de fois où un événement A s'est produit en n épreuves indépendantes.

$$2.1) \text{ Avec } P(A) = p \text{ et } \varepsilon > 0 \text{ démontrer que } P\left(\left|\frac{X}{n} - p\right| > \varepsilon\right) \leq \frac{1}{4n\varepsilon^2}$$

2.2) Une pièce est jetée n fois. La variable X représente le nombre de fois où pile s'est produit.

Déterminer le plus petit n pour lequel l'inégalité de Tchebychev implique $P(0.4 < X < 0.6) > 0.9$

$$2.3) \text{ Calculer } n \text{ en supposant que } \frac{X}{n} \text{ suit une loi normale.}$$

On donne $P(|U| < 1.64) = 0.9$ où U est une variable aléatoire normale réduite.

3) Soit X une variable aléatoire de densité constante sur $(0, \frac{\pi}{2})$

3.1) Calculer les deux premiers moments de la variable $Y = \cos(X)$

3.2) En déduire par l'application de la loi des grands nombres une évaluation statistique de l'intégrale

$$I = \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \cos(x) dx$$

3.3) A partir de combien de tirage de X aura-t-on avec 9 chances sur 10 une évaluation de I à 5 %.

On donne $P(|U| < 1.64) = 0.9$, U normale réduite.

9.5.2 Corrigés

1) Démonstration identique à celle du théorème de Tchebychev

En posant $Z = X_n - X$ on en déduit que

$$P(|X_n - X| > \varepsilon) \leq \frac{E((X_n - X)^2)}{\varepsilon^2}$$

Par hypothèse le second membre tend vers 0 quand $n \rightarrow \infty$ d'où la convergence en probabilité.

$\frac{X}{n}$ représente la fréquence relative de l'événement A.

2)

$$2.1) \text{ On sait que } E\left(\frac{X}{n}\right) = p \quad \text{et que } \text{var}\left(\frac{X}{n}\right) = \frac{p(1-p)}{n} \leq \frac{1}{4n} \text{ pour } 0 \leq p \leq 1$$

$$\text{D'où en appliquant la relation de Tchebychev } P\left(\left|\frac{X}{n} - p\right| > \varepsilon\right) \leq \frac{1}{4n\varepsilon^2}$$

$$2.2) \text{ Dans ce cas on a } p = q = \frac{1}{2} \quad \text{et } \varepsilon = 0.1$$

$$\text{On choisit } n \text{ pour que } P\left(\left|\frac{X}{n} - \frac{1}{2}\right| < 0.1\right) \geq 1 - \frac{1}{4n(0.1)^2} \geq 0.9$$

Soit $n \geq 250$

$$2.3) \text{ Il faut que } 1.64 \cdot \frac{1}{4n} = 0.1 \quad \text{soit } n = 68$$

Cet exemple montre que la borne de Tchebychev est très large. Ce qui est normal car elle est valable pour toute loi de probabilité.

3)

$$3.1) \quad E(Y) = \frac{2}{\pi} \quad E(Y^2) = \frac{1}{2} \quad \sigma_Y^2 = \frac{1}{2} - \frac{4}{\pi^2}$$

3.2) Estimation, à partir d'un tirage équiprobable des x_i , sur l'intervalle $(0, \frac{\pi}{2})$

$$\text{par } Un = \frac{1}{n} \sum_{i=1}^n Y_i = \frac{1}{n} \sum_{i=1}^n \cos(X_i)$$

$$\text{avec } E(U_n) = E(Y) = \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \cos(x) dx = \frac{2}{\pi} I \rightarrow I_n = \frac{\pi}{2} U_n$$

3.2) n tel que

$$1.64 \frac{\sigma_{I_n}}{E(I_n)} = 1.64 \frac{\pi \sigma_Y}{2 \sqrt{n}} = 0.05 \rightarrow n \geq 252$$

Statistiques Descriptives

II Statistique descriptive

- ✓ **Mots clés** : distribution statistique,
{diagramme en barre, histogramme},{moyenne, médiane, mode},
{variance, écart type}
- ✓ **Présenter** des données statistiques....
- ✓ Savoir **calculer et interpréter** les différents paramètres statistiques

Une nécessité...

1	1	1	3	5	1	1	1	3	5
0	2	0	2	3	0	2	0	2	3
5	3	2	2	1	5	3	2	2	1
0	0	2	1	1	0	0	2	1	1
1	1	1	3	5	1	1	1	3	5
0	2	0	2	3	0	2	0	2	3
5	3	2	2	1	5	3	2	2	1
0	0	2	1	1	0	0	2	1	1
1	1	1	3	5	1	1	1	3	5
0	2	0	2	3	0	2	0	2	3
5	3	2	2	1	5	3	2	2	1
0	0	2	1	1	0	0	2	1	1
1	1	1	3	5	1	1	1	3	5
0	2	0	2	3	0	2	0	2	3
5	3	2	2	1	5	3	2	2	1
0	0	2	1	1	0	0	2	1	1
1	1	1	3	5	1	1	1	3	5
0	2	0	2	3	0	2	0	2	3

800 valeurs, c'est tout de même fastidieux à manipuler...

Notations sur les séries statistiques

$\{...\}$ = collection d'éléments

$\{x_i\}_{1 \leq i \leq n}$ est un raccourci pour $\{x_1, x_2, \dots, x_i, \dots, x_n\}$

Je le noterai parfois simplement $\{x_i\}_n$ ou $\{x_i\}$

Les notations où interviennent des majuscules se réfèrent à des recensements (directement sur toute la population). Ainsi : $\{X_i\}_N$

Méthodes

Plusieurs techniques sont disponibles :

- ▶ Classification des données
- ▶ Visualisation graphique
- ▶ Quantification à l'aide de paramètres statistiques

Difficulté croissante avec le nombre de variables

Séries statistiques

Série statistique simple = ensemble de données relatives à **une** variable mesurée sur un échantillon ou une population d'éléments

Série statistique double = ensemble de **couples** de données relatives à **deux** variables mesurées sur un échantillon ou une population d'éléments

Série statistique multiple = ensemble de **multiplets** de données relatives à **plusieurs** variables mesurées sur un échantillon ou une population d'éléments

II – 1

Statistique descriptive d'une série statistique simple

II – 1 a)

Hiérarchisation

Méthode de classement

Il est plus commode de regrouper les données en quelques **classes** plus maniables.

Par exemple, pour les résultats du sondage

Nombre d'objet en verre jeté	Nombre de réponses
0	145
1	213
2	335
3	179
4	78
5	50
6et+	0

La stratégie de classification dépend du **type de variable**

Variables quantitatives, variables qualitatives

Variable quantitative = Variable dont les valeurs possibles sont comparables et que l'on peut formuler de manière numérique

Ex : Nombre d'enfants, volume,...

Variable qualitative = Variable non quantitative. Elle se réfère souvent à une caractéristique (espèce, genre).

Ex : Carottes, navets,...

Variables discrètes, Variables continues

Variable discrète = Variable dont les valeurs possibles sont discontinues, c'est dire séparées.

Ex : Nombre d'enfants, d'objets,...

Variable continue = Variable dont les valeurs peuvent passer continûment

Ex : Masse, volume, concentration,...

Intervalles de classe

On définit un **critère** :

- Variable quantitative = 1 **intervalle de classe**
=

[borne inférieure borne supérieure]

La valeur centrale est l'**indice de classe**

- Variable qualitative
1 critère qualitatif = 1 classe

Méthode de classement

<i>Variable</i>	Quantitative	Qualitative
Discrète	Valeurs/Intervalle de classe	Qualité
Continue	Intervalle de classe	X

Quantifier le contenu des classes

Effectif = **Fréquence absolue** (d'une classe) = f =
Nombre d'éléments appartenant à la classe

Fréquence relative (d'une classe) = f_{rel} =
Effectif rapporté à l'effectif total de l'échantillon
(n)
 $f_{rel} = f/n$

Pourcentage = fréquence relative exprimée en
%

$$100 \times f_{rel} = 100 \times f/n$$

Distributions statistiques

On crée ainsi une
distribution statistique
formée d'une série de couples
(**intervalle de classe** , **fréquence**)

Tableaux de distribution de fréquences

Un tableau montrant les couples
(critère, fréquences)
s'appelle un
tableau de distribution de fréquences

Nombre d'objet en verre jeté	Nombre de réponses
0	145
1	213
2	335
3	179
4	78
5	50
6et+	0

Notations sur les distributions

Une distribution est donnée par une série de couples
(indice, effectif)

Conformément aux notations des séries statistiques,
je les noterai sous la forme

$$\{v_i, f_i\}_{1 \leq i \leq D}$$

Contrairement aux séries statistiques, il est implicite
que les valeurs sont rangées par **ordre croissant**

$$v_i < v_{i+1}$$

Propriétés des distributions

La classification doit être complète.
Notamment, on doit retrouver pour toute
distribution

$$\{v_i, f_i\}_{1 \leq i \leq D}$$

construite sur la série statistique $\{x_i\}_{1 \leq i \leq n}$

$$\sum_{i=1}^D f_i = n$$

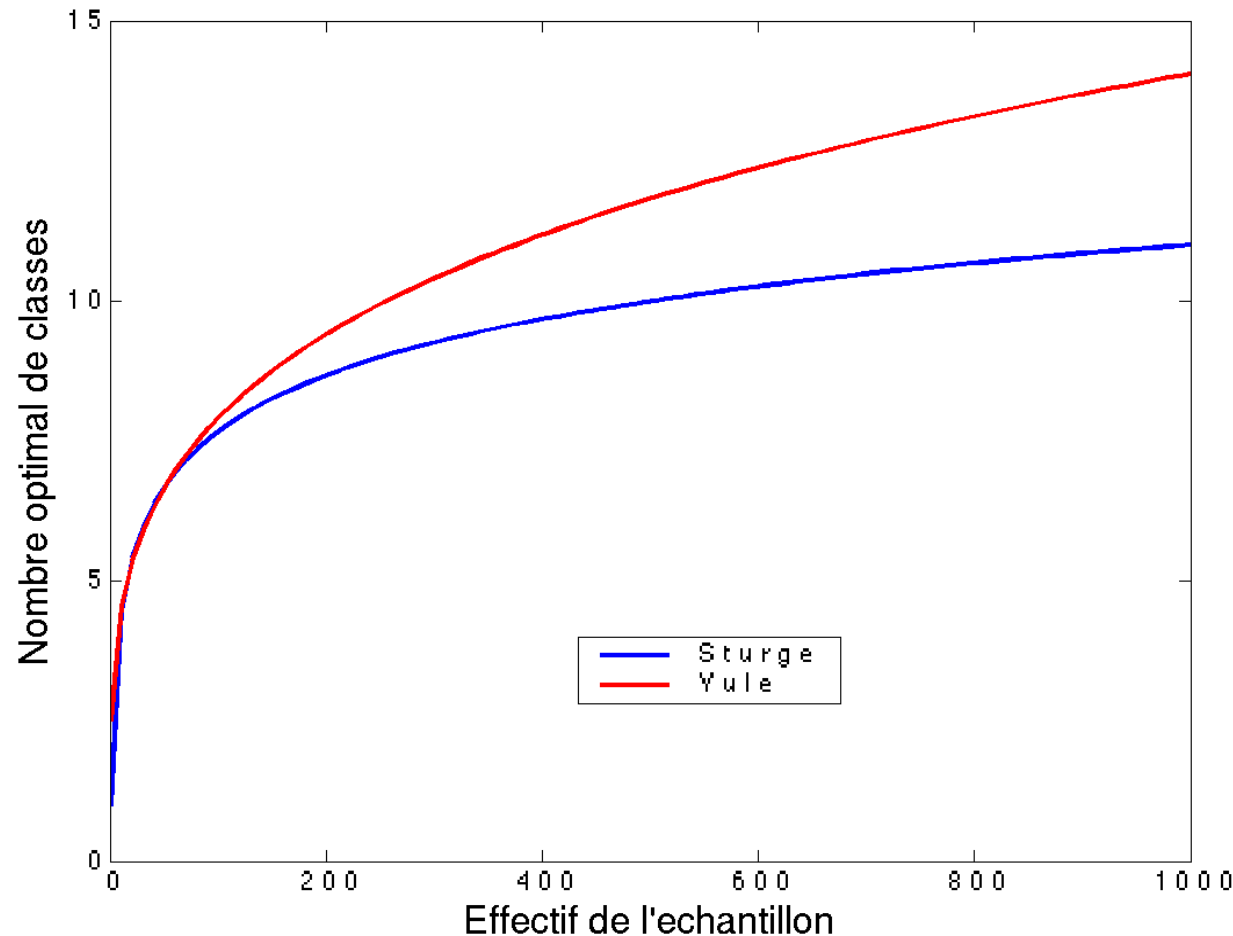
Le nombre des éléments dispersés dans les
classes
est égal à l'effectif initial

Optimisation du choix de l'intervalle de classe

Dans le cas des variables continues,
le choix des intervalles de classe est délicat :

- **Trop petits:** le nombre de classes est trop grand pour être maniable
- **Trop grands:** des détails sont dissimulés au sein d'une même classe

Optimisation du nombre de classes



(Variable continue)
Règles empiriques

Règle de Sturge

$$D = 1 + \frac{10}{3} \log_{10} n$$

Règle de Yule

$$D = \frac{10}{4} n^{\frac{1}{4}}$$

Optimisation du choix de l'intervalle de classe

(variables continues)

La plupart des études sont réalisées avec :
Des intervalles de classes de longueur aussi
égales que possible
Les classes de fréquence nulle sont évitées

Distributions cumulées

Une distribution cumulée $\{v_i, f_{cum,i}\}_{1 \leq i \leq D}$
dérivée de la distribution $\{v_i, f_i\}_{1 \leq i \leq D}$:

- ◆ A les mêmes intervalles de classe
- ◆ Les fréquences cumulées sont la somme de la fréquence de la classe et des fréquences de toutes les classes la précédant

$$f_{cum,i} = \sum_{j=1}^{j=i} f_j$$

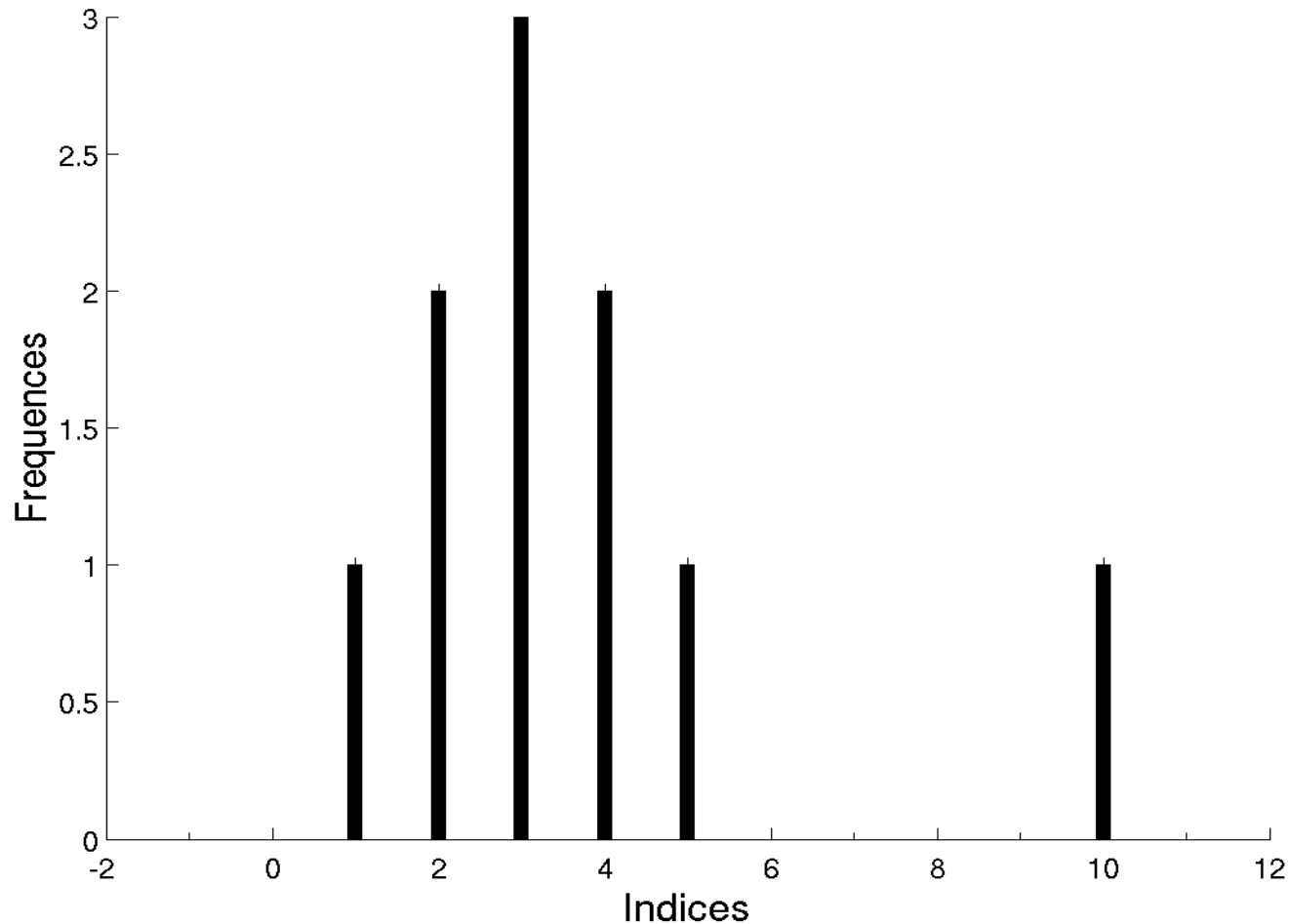
Distributions cumulées

Nombre d'objet en verre jeté	Nombre de réponses	Nombre de réponses cumulées
0	145	145
1	213	358
2	335	693
3	179	872
4	78	950
5	50	1000
6et+	0	1000

II – 1 b)

Représentations graphiques

Diagramme en bâtons

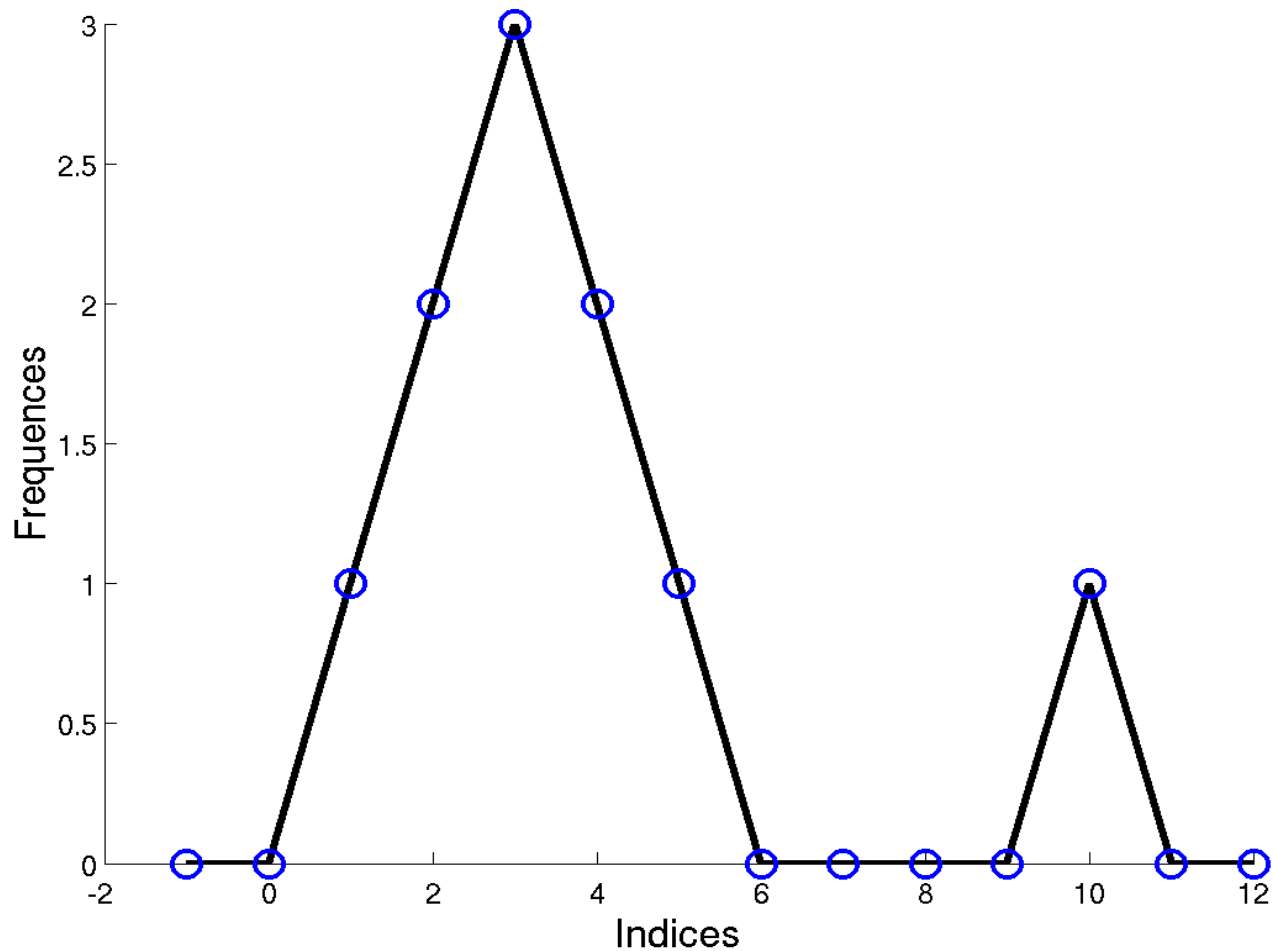


Préférentiellement pour des variables
discrètes

Diagramme en bâtons

<i>Variable</i>	Quantitative	Qualitative
Discrète	Valeurs/Intervalle de classe	Qualité
Continue	Intervalle de classe	X

Polygone de fréquence

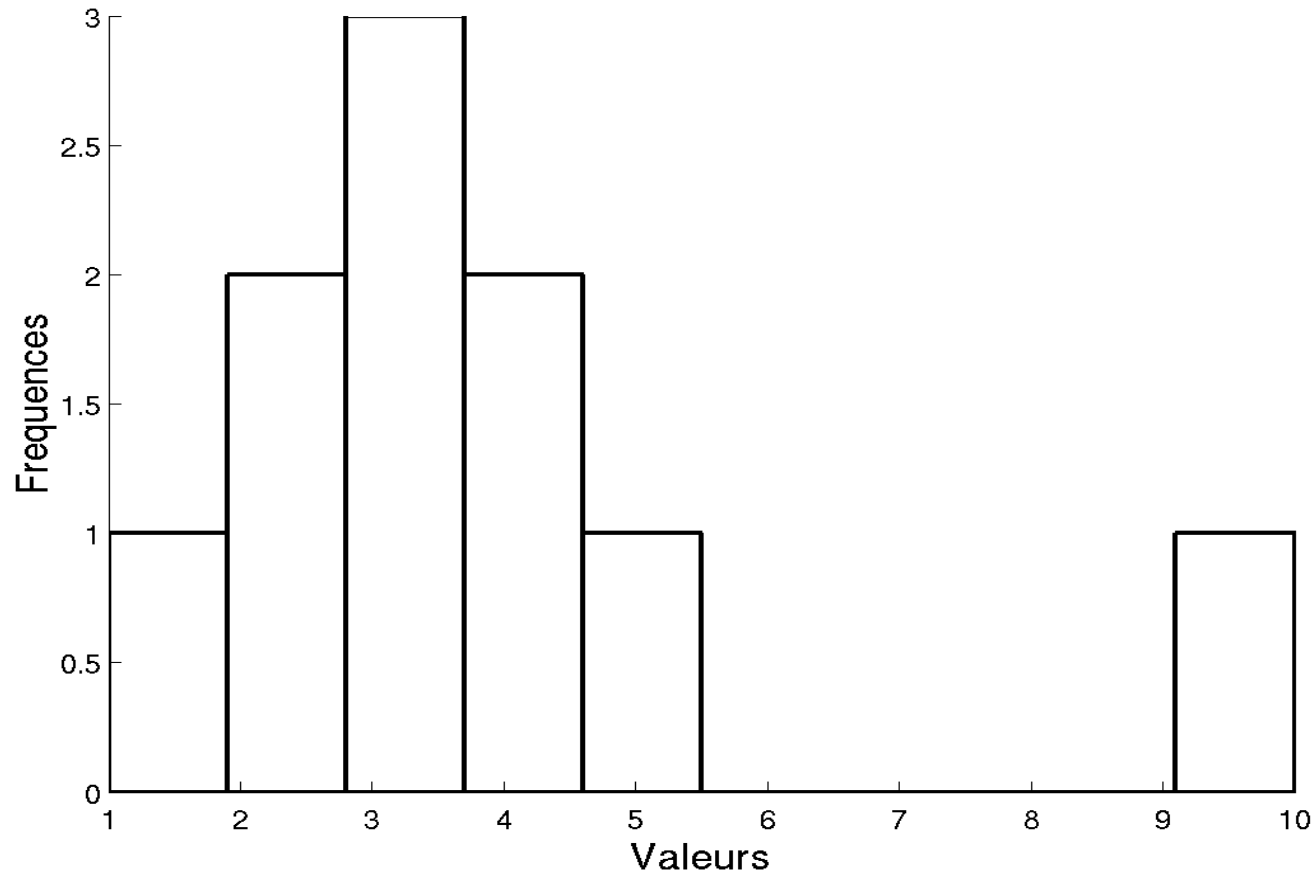


Préférentiellement pour des variables
discrètes

Polygone de fréquence

<i>Variable</i>	Quantitative	Qualitative
Discrète	Valeurs/Intervalle de classe	Qualité
Continue	Intervalle de classe	X

Histogramme



Préférentiellement pour des variables continues
histos = tissu

Histogramme

<i>Variable</i>	Quantitative	Qualitative
Discrète	Valeurs/Intervalle de classe	Qualité
Continue	Intervalle de classe	X

II – 1 c)

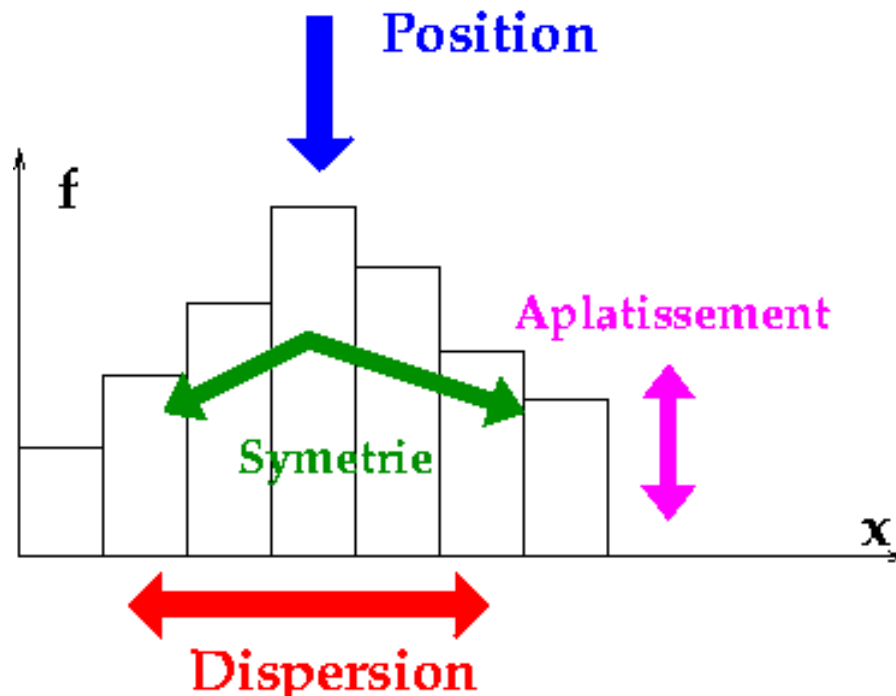
Paramètres statistiques

Paramètres statistiques

L'approche est différente suivant que la variable est quantitative ou qualitative

- **Qualitative** : on s'intéresse à la répartition des éléments dans les classes
- **Quantitative** : on essaie de dériver des valeurs indépendantes de la classification.

Variables quantitatives

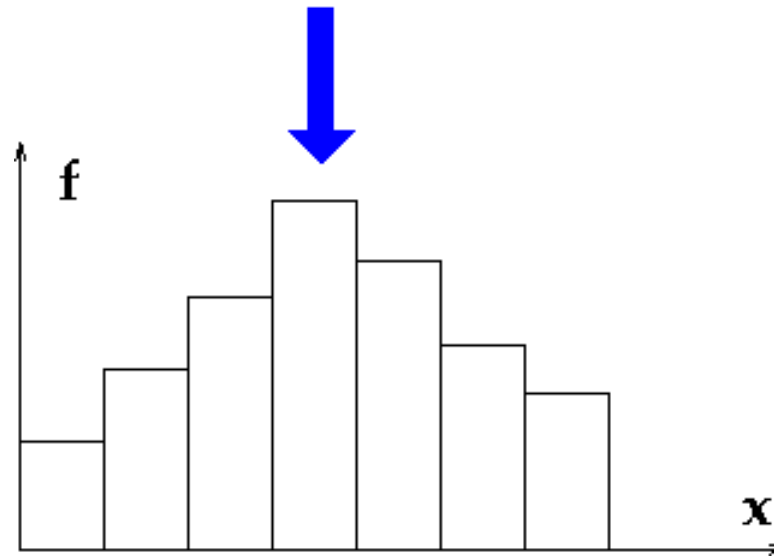


Dans l'ordre :

- (1) Position
- (2) Dispersion
- (3) Symétrie
- (4) Aplatissement

...

Paramètres de position



- Moyenne
- Médiane
- Mode

Moyenne arithmétique

Moyenne arithmétique =
moyenne (d'une série statistique $\{x_i\}$) =
somme des valeurs de la série rapportée à
son nombre d'éléments (= effectif, ici noté N)

$$\mu_x = \bar{X} = \langle X \rangle = \frac{x_1 + x_2 + \dots + x_j + \dots + x_N}{N} = \frac{\sum_{j=1}^N x_j}{N}$$

Moyenne arithmétique

Il existe d'autre type de moyenne.
Par exemple, la **moyenne géométrique** :

$$(x_1 \times x_2 \times \dots \times x_j \times \dots \times x_N)^{1/N} = \left(\prod_{j=1}^N x_j \right)^{1/N}$$

Mais la moyenne arithmétique présente l'immense avantage que la somme des écart à la moyenne sont alors nuls.

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

Moyenne et distributions

Même si la série statistique $\{x_i\}_N$ a été réorganisée en distribution, il est possible de retrouver la valeur de la moyenne à partir des valeurs de la distribution

$$\{v_i, f_i\}_{1 \leq i \leq D}$$

$$\bar{x} = \frac{\sum_{i=1}^D v_i f_i}{D}$$

C'est la formulation du **barycentre** des indices pondérés par les fréquences

Moyenne arithmétique

Avantages

- ✓ Simple à calculer
- ✓ Linéarité : $\overline{ax} = a \times \bar{x}$
- ✓ Additivité : $\overline{x + y} = \bar{x} + \bar{y}$
- ✓ La somme des écarts à la moyenne est plus faible que la somme des écarts à la médiane ou au mode

Désavantages

- ✗ Sensibilité aux valeurs extrêmes
(ex: {2, 10, 3, 3, 5, 3, 4, 1, 4, 2})
- ✗ Si la distribution est dissymétrique, la moyenne représente mal la valeur centrale

Médiane

Médiane = valeur de la variable qui sépare la série statistique en deux groupes d'égale effectif.

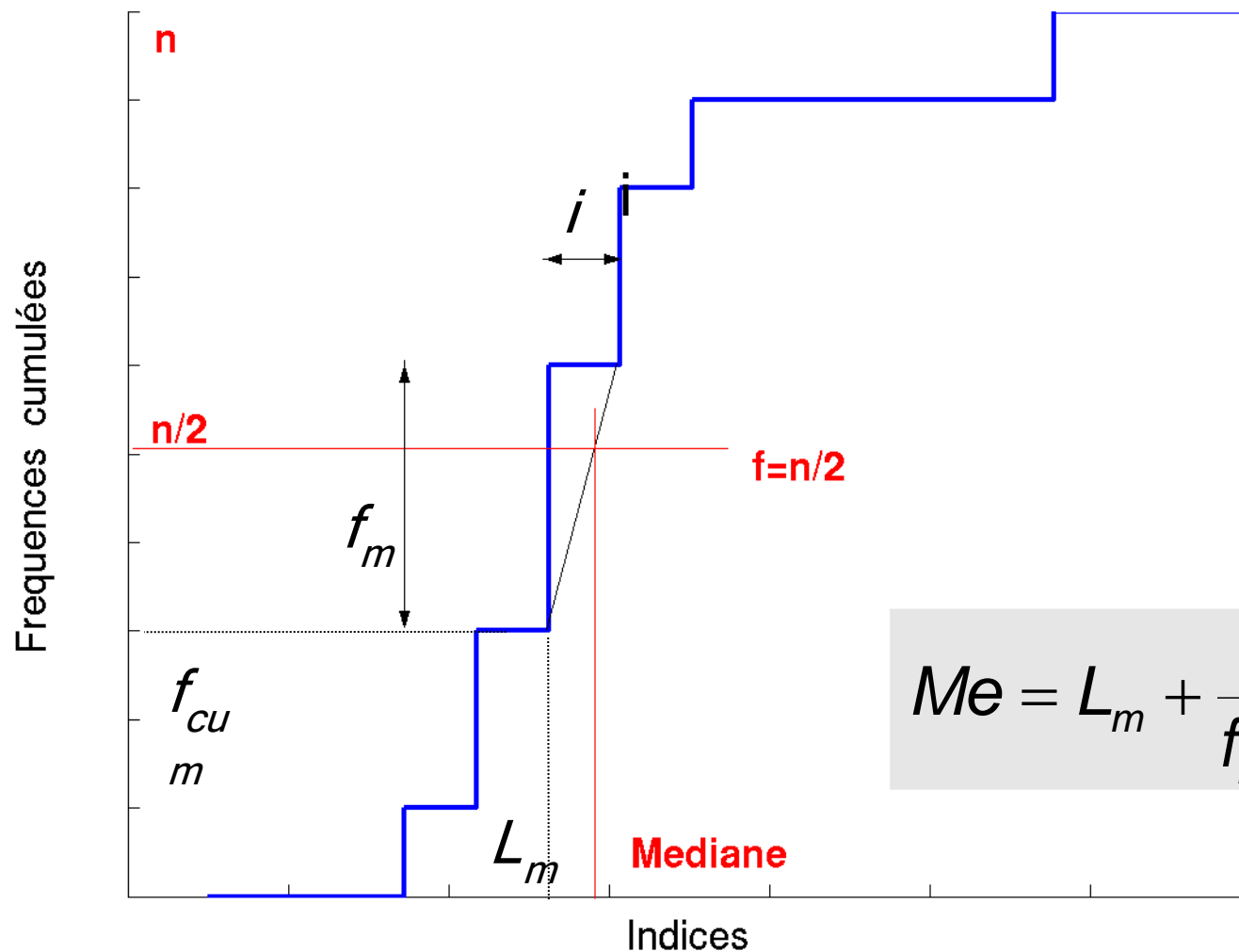
En pratique :

- 1) On classe les données par ordre croissant
- 2) La médiane est la valeur qui se trouve au milieu des données triées

ex: {2,10,3,3,5,3,4,1,4,2}

Médiane d'une distribution

Elle se détermine à partir des
fréquences cumulées.



Médiane

Avantages

✓ Peu sensibles aux valeurs extrêmes.

✓ Linéarité :

$$Me(ax) = a \times Me(x)$$

Désavantages

✗ Se prête mal aux calculs :

$$Me(x + y) \neq Me(x) + Me(y)$$

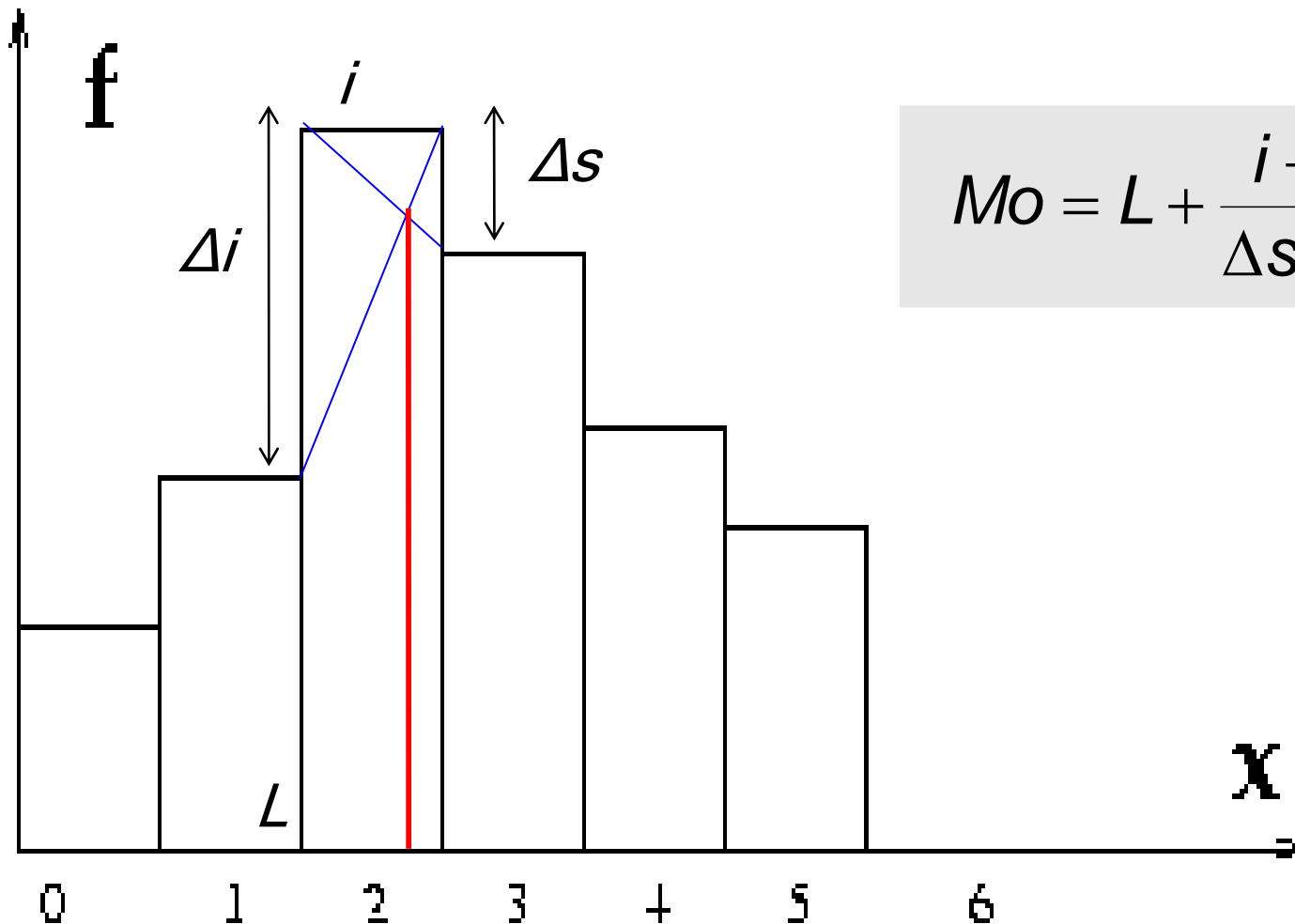
Mode

Mode = indice de la classe ayant la fréquence la plus élevée.

En pratique :

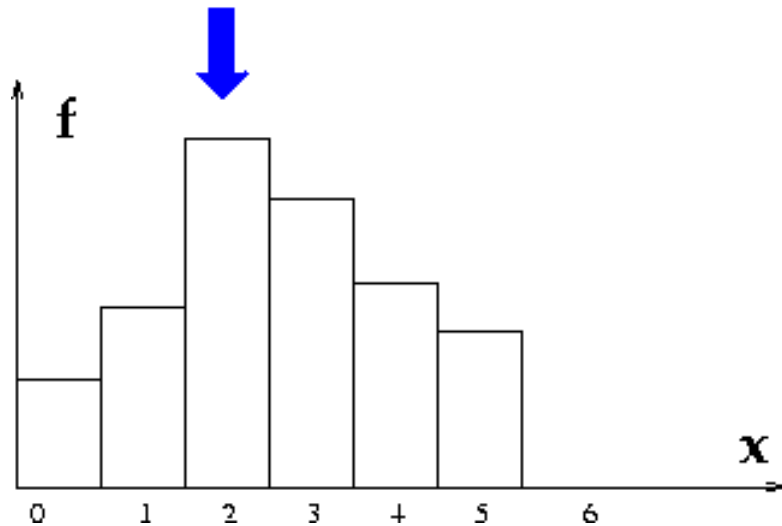
- 1) On trace l'histogramme
- 2) On recherche le maximum

Mode d'une distribution

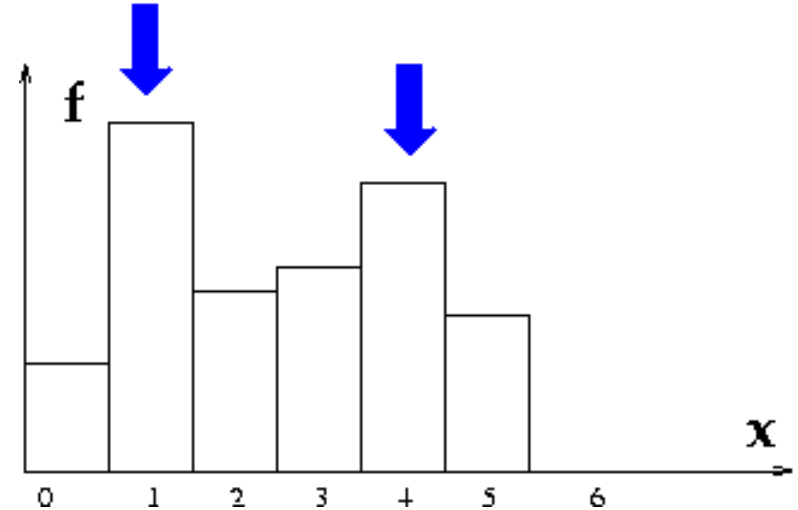


Distributions monomodales, bimodales,...

Monomodale



Bimodale



À quoi ressemblera une distribution **multimodale** ?

Mode

Avantages

- ✓ Faible sensibilité aux valeurs extrêmes
- ✓ Si la population est très hétérogène (p.ex. distribution bimodale), il vaut mieux deux modes qu'une moyenne ou qu'une médiane

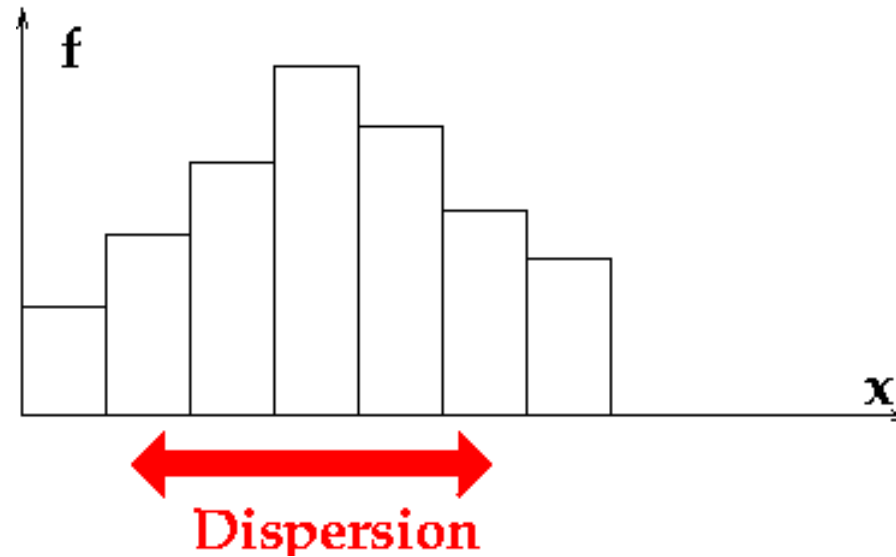
Désavantages

- ✗ Extrême sensibilité aux choix des intervalles de classe
- ✗ Ne se prête pas aux calculs.

$$Mo(ax) \neq a \times Mo(x)$$

$$Mo(x + y) \neq Mo(x) + Mo(y)$$

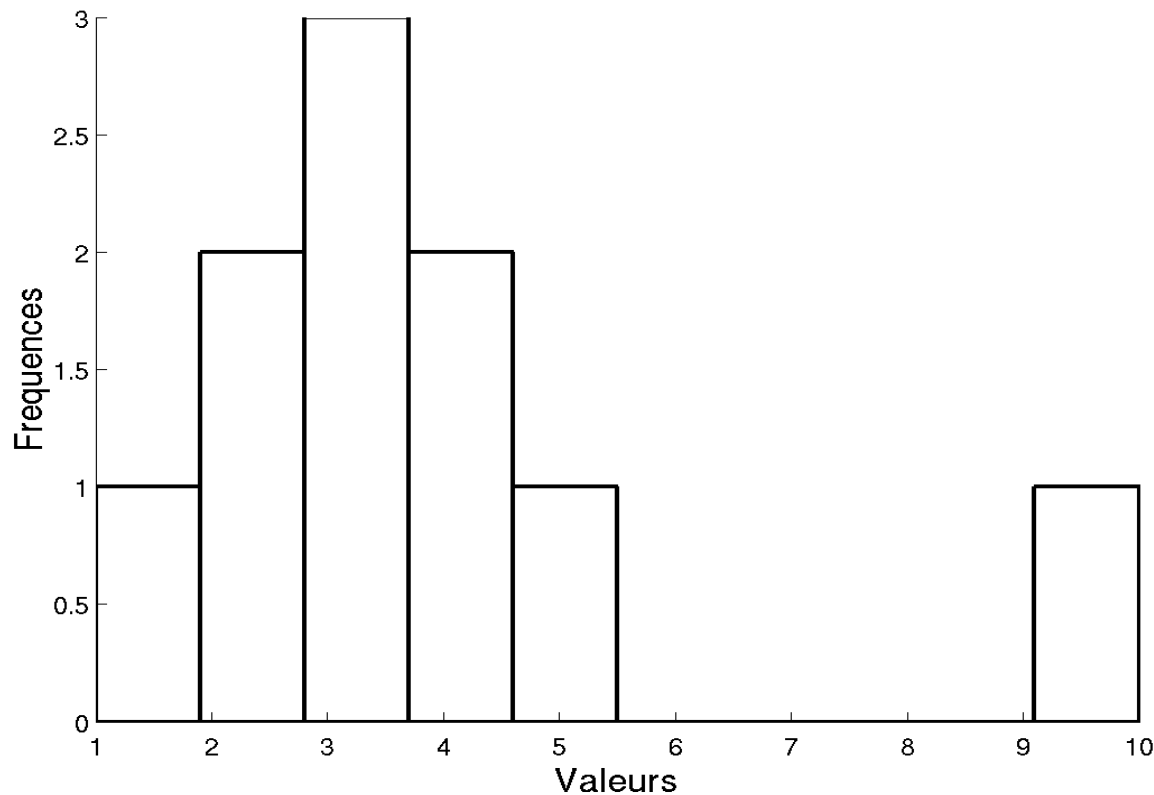
Paramètres de dispersion



- ▶ Amplitude
- ▶ Quartiles
- ▶ Variance
- ▶ Écart type

Amplitude

Amplitude = **Étendue** =
écart entre la valeur maximale et la valeur
minimale de la distribution.



Étendue

Position	Amplitude
Dispersion	Quartiles
	Variance
	Écart-type

Amplitude

Avantages

- ✓ Facile à établir

Désavantages

- ✗ Extrême sensibilité aux choix des intervalles de classe
- ✗ Ne se prête pas aux calculs.

Quartiles

Quartile à 25% = valeur de la variable qui délimite 25% des premières données de la série statistique classée par ordre croissant

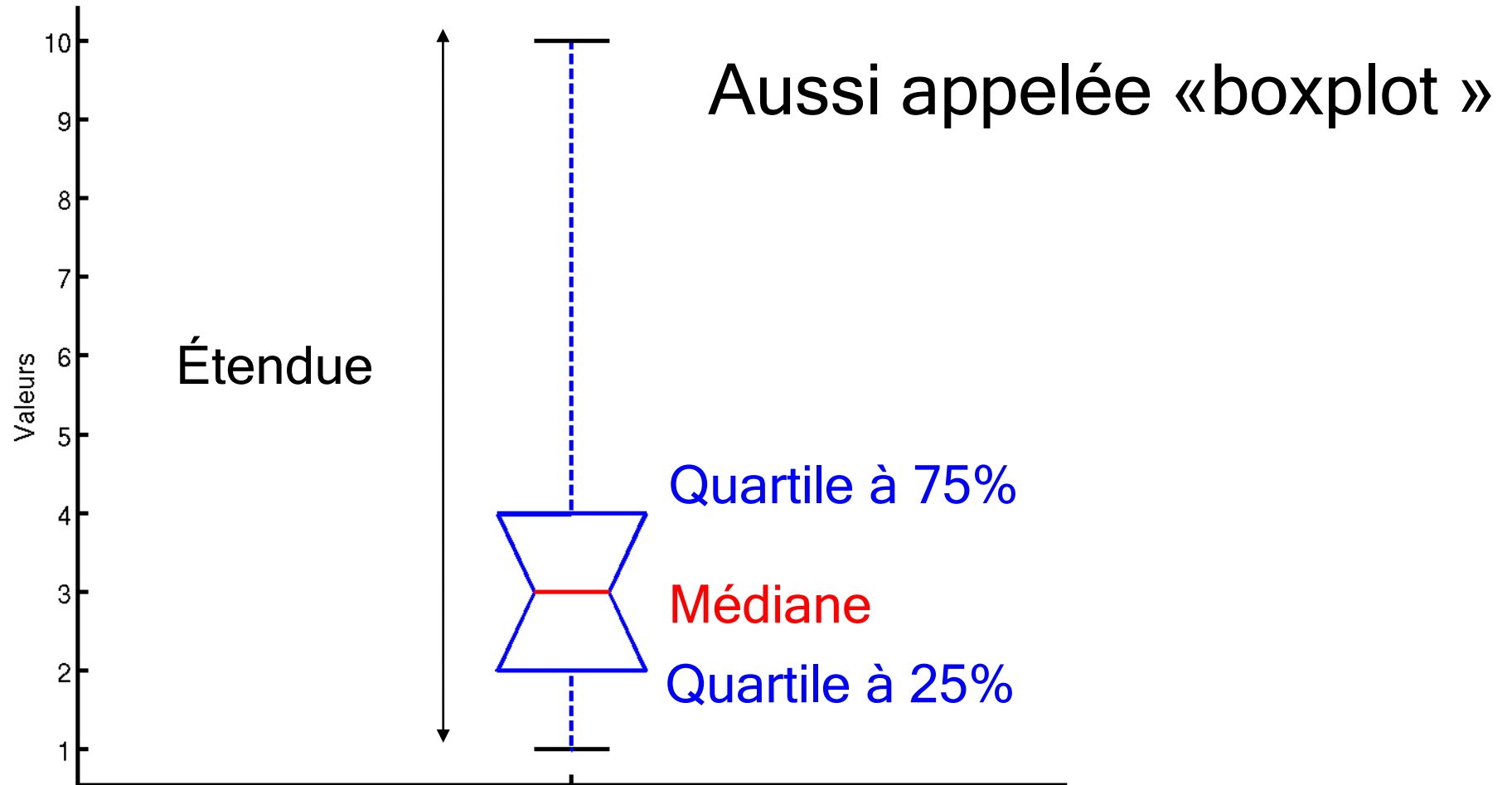
Quartile à 75% = valeur de la variable qui sépare 75% des premières données de la série statistique classée par ordre croissant

Intervalle inter-quartile =

[Quartile à 25% Quartile à 75%]

Position	Amplitude
Dispersion	Quartiles
	Variance
	Écart-type

Boîte à moustaches



Quartiles

Avantages

- ✓ Peu sensibles aux valeurs extrêmes.
- ✓ Echelle : $Q(ax) = a \times Q(x)$
- ✓ Les écarts des deux quartiles à 25% et 75% donnent une idée de l'asymétrie

Désavantages

- ✗ Il faut calculer 2 valeurs
- ✗ Se prête mal aux calculs :
 $Q(x + y) \neq Q(x) + Q(y)$
- ✗ Extrapolation difficile de l'échantillon à la population

Variance d'une population

Variance (d'une **population** $\{X_i\}$) =
moyenne des carrés des écarts des valeurs à
la moyenne de la population.

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

Variance d'un échantillon

Variance (d'un échantillon $\{x_i\}$) =
somme des carrés des écarts des valeurs à la
moyenne de l'échantillon, ramenée au
nombre de **degrés de liberté** de l'échantillon
(**n-1**, si n est l'effectif de l'échantillon).

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

Et non n!

Position	Amplitude
Dispersion	Quartiles
	Variance
	Écart-type

Pourquoi cette différence ?

La variance d'un échantillon est optimisée pour approcher aux mieux la variance de la population.

Diviser par n et non par $n-1$ introduit un **biais statistique**.

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

Et non n !

Par exemple, supposons que l'**échantillon** a un seul élément $n=1$.
On ne peut alors pas remonter la variance de la population.
D'ailleurs $s_x^2 = 0/0$ est indéterminé.

Si par contre, on parle d'une **population** à un élément, sa variance est nulle.
D'ailleurs elle vaut $\sigma^2 = 0/1 = 0$.

Position	Amplitude
Dispersion	Quartiles
	Variance
	Écart-type

Pourquoi cette différence ?

Nous reverrons ce problème dans la partie sur la statistique inductive.

Nous expliquerons alors ce qu'est un degré de liberté et pourquoi c'est $(n-1)$ et non n qui permet d'obtenir une **estimation non biaisée** de la variance de la population.

Position	Amplitude
Dispersion	Quartiles
	Variance
	Écart-type

Calcul de la variance

En pratique :

On remplace l'expression de

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

par

$$\sum_{i=1}^N x_i^2 - N\bar{x}^2$$

Position	Amplitude
Dispersion	Quartiles
	Variance
	Écart-type

Écart-type

Écart-type = racine carrée de la variance
(homogène à une valeur)

Population

$$\sigma = \sqrt{\sigma^2}$$

Échantillon

$$s_x = \sqrt{s_x^2}$$

Position	Amplitude
Dispersion	Quartiles
	Variance
	Écart-type

Écart-type

Avantages

- ✓ Simple à calculer
- ✓ Echelle : $s_{ax} = a \times s_x$
- ✓ Ajout d'une constante :
 $s_{a+x} = s_x$

Désavantages

Écart-type

Si la distribution est symétrique,
on observe *approximativement* ⁽¹⁾

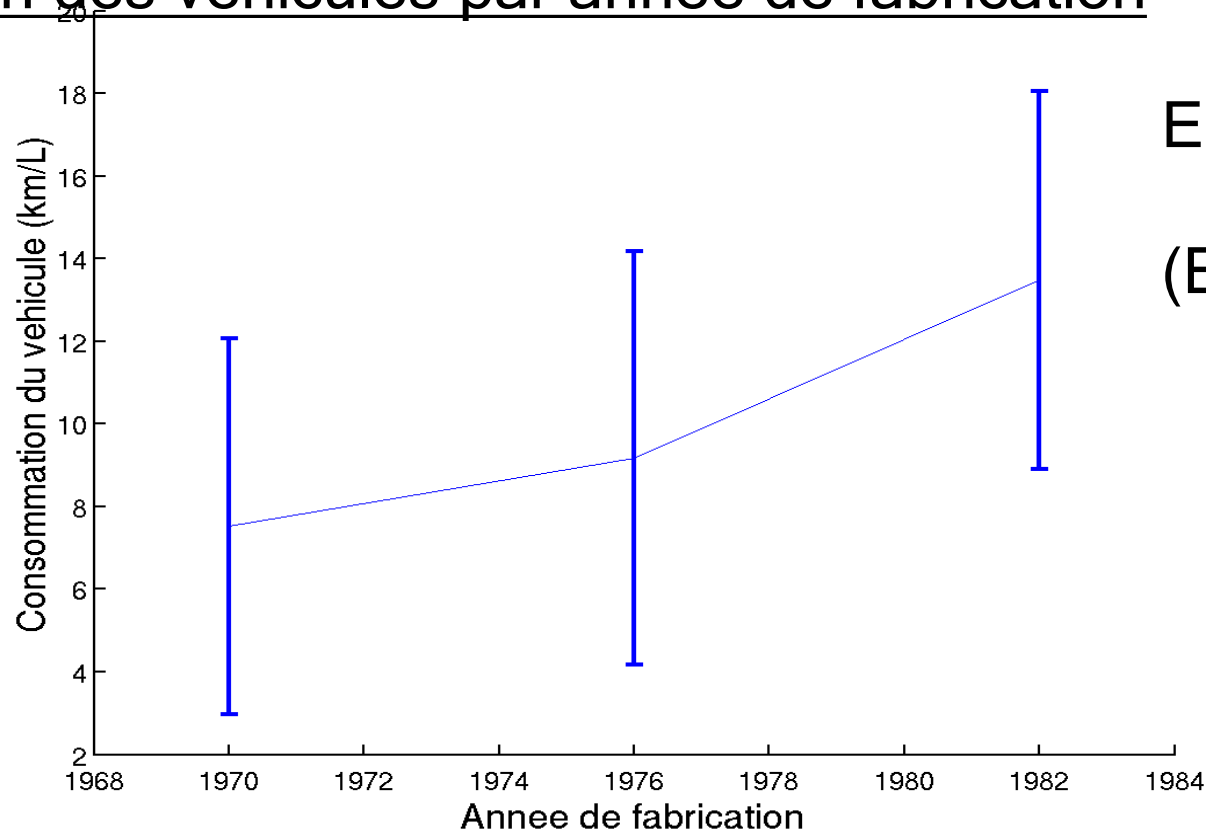
- Que **68%** des valeurs sont dans $[\langle x \rangle - \sigma, \langle x \rangle + \sigma]$
- Que **95%** des valeurs sont dans $[\langle x \rangle - 2\sigma, \langle x \rangle + 2\sigma]$
- Que **99%** des valeurs sont dans $[\langle x \rangle - 3\sigma, \langle x \rangle + 3\sigma]$

(1) On verra que ces valeurs dérivent en fait
des propriétés d'une loi normale

Barres d'erreur

Les deux informations {moyenne + écart-type} peuvent être données simultanément par des **barres d'erreur**.
Il faut préciser alors l'incertitude (5%, 1%, ...) associée.

Consommation des véhicules par année de fabrication



Erreur à $\pm 2\sigma$

(Erreur à 95%)

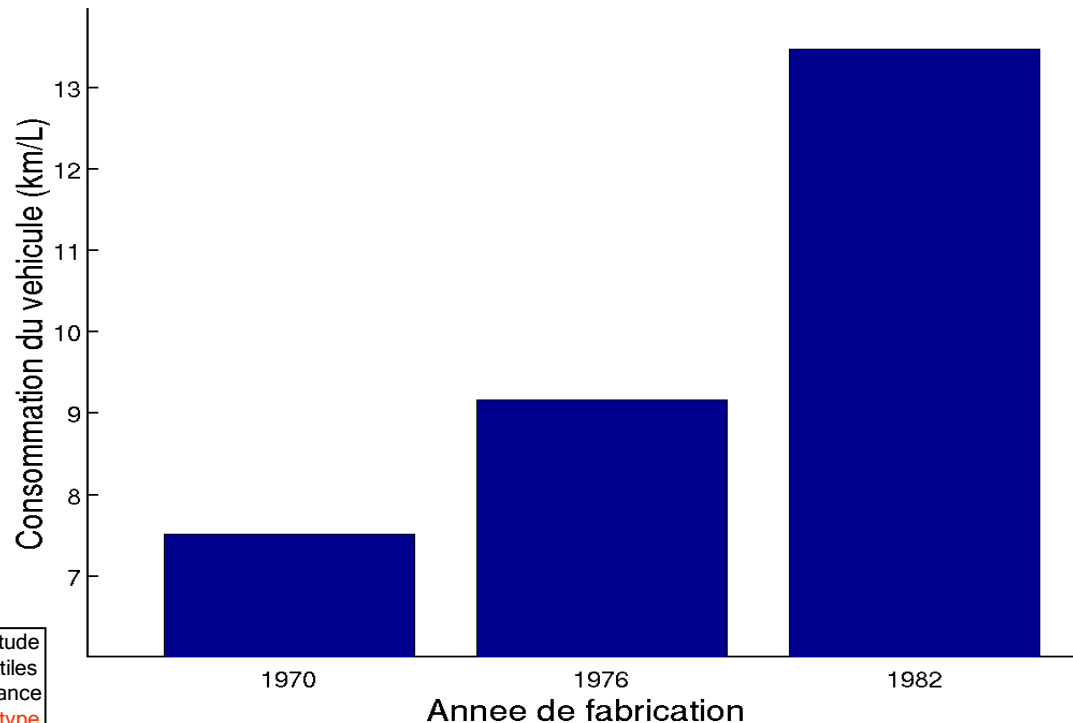
Grosso modo

Position	Amplitude
Dispersion	Quartiles
	Variance
	Écart-type

Attention aux représentations graphiques

Voici les mêmes données présentées différemment :

Consommation des véhicules par année de fabrication



Le graphe est trompeur :

- ✗ Pas de paramètre de dispersion
- ✗ Utilisation d'une échelle verticale différente

Position	Amplitude
Dispersion	Quartiles
	Variance
	Écart-type

Coefficient de variation

Coefficient de variation =

Rapport de la variance à la moyenne (en %)

$$CV = \frac{s_x}{\bar{x}} \times 100$$

Sans unité, il permet de comparer
des distributions de fréquences d'unités différentes

Mais il devient mal défini
si la moyenne est proche de zéro

Standardisation

Centrage

$$X \longrightarrow Y = X - \bar{X}$$

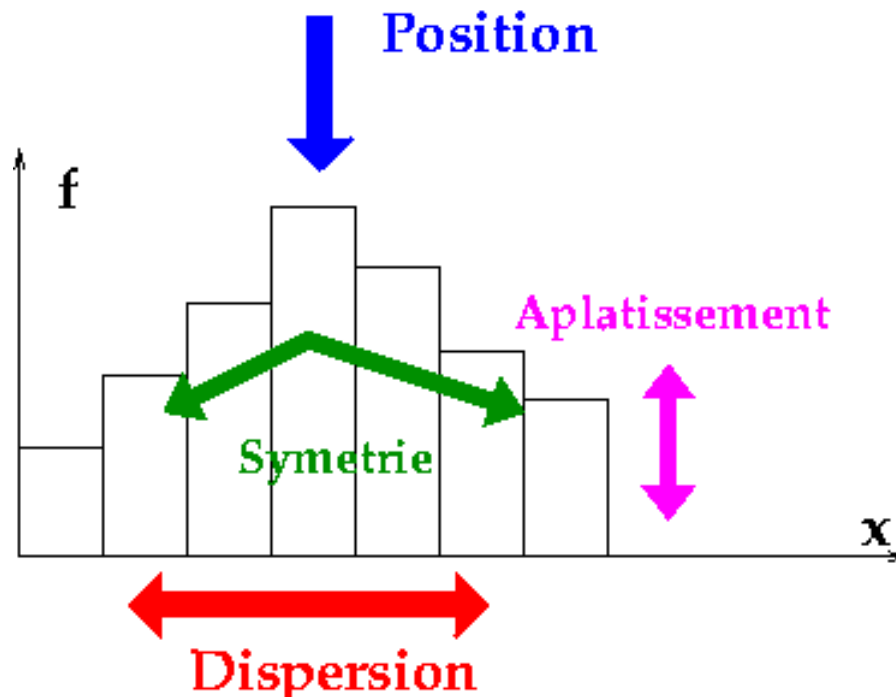
$$\bar{Y} = 0 \quad \sigma_Y = \sigma_X$$

Standardisation

$$X \longrightarrow Z = \frac{X - \bar{X}}{\sigma_X}$$

$$\bar{Z} = 0 \quad \sigma_Z = 1$$

Paramètres d'ordres supérieurs



(3) Asymétrie

(4) Aplatissement

...

Moments d'ordres supérieurs

Moment d'ordre 3

Population

$$M_3 = \frac{\sum_{i=1}^N (X_i - \bar{X})^3}{N}$$

Échantillon

$$m_{3x} = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n-1}$$

Moment d'ordre 4

Population

$$M_4 = \frac{\sum_{i=1}^N (X_i - \bar{X})^4}{N}$$

Échantillon

$$m_{4x} = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n-1}$$

Coefficient d'asymétrie

Population

$$\alpha_3 = \frac{M_3}{\sigma^3}$$

Échantillon

$$\alpha_3 = \frac{m_{3x}}{s_x^3}$$

$$\alpha_3 < 0$$

$$0$$

$$\alpha_3 > 0$$

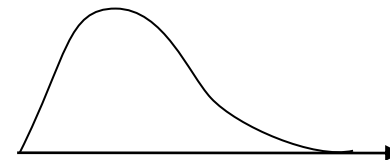
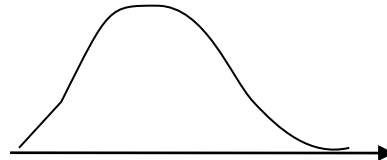
$$\alpha_3$$

Asymétrique à gauche

Symétrique

Asymétrique à droite

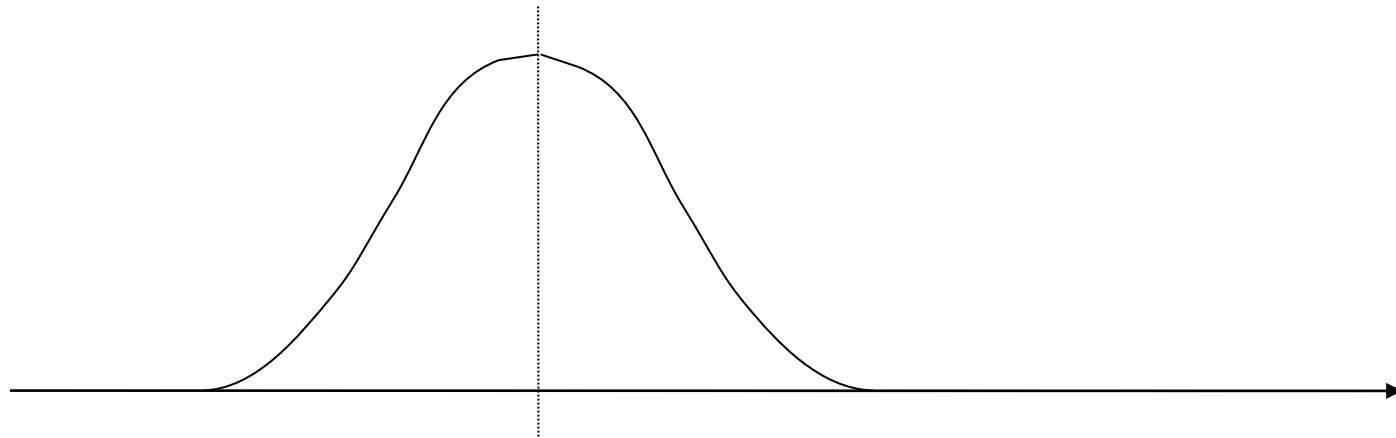
Sans
Dimension



Courbe symétrique

Moyenne=Médiane=Mode

$$\alpha_3=0$$



$\langle X \rangle$

Me

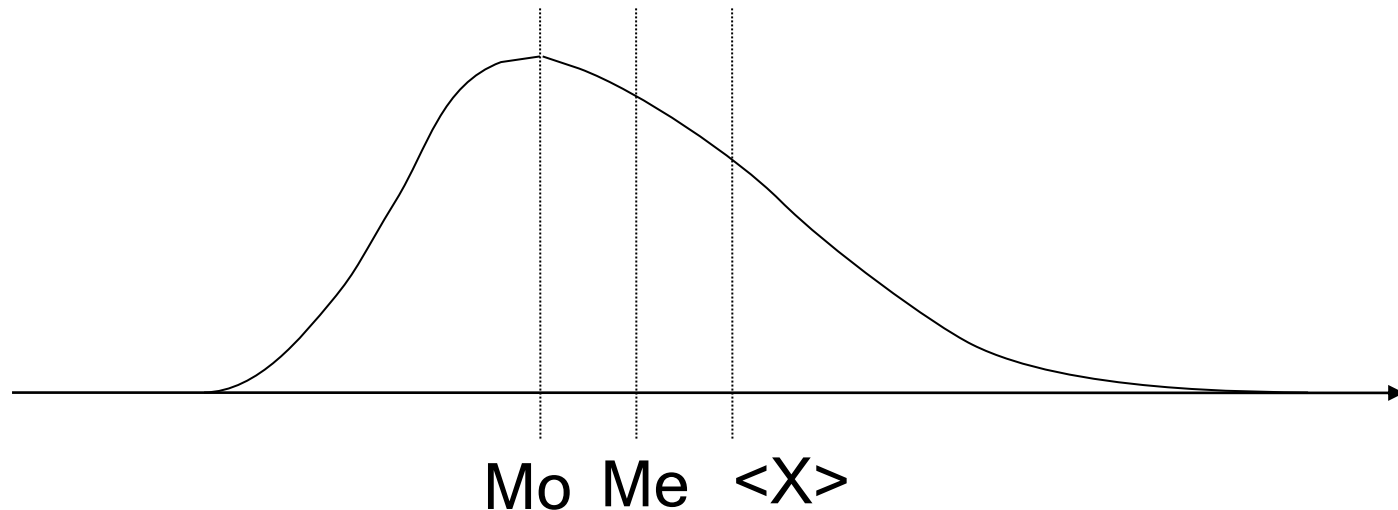
Mo

Courbe asymétrique à droite

La courbe est allongée vers la droite

Moyenne > Médiane > Mode

$$\alpha_3 > 0$$

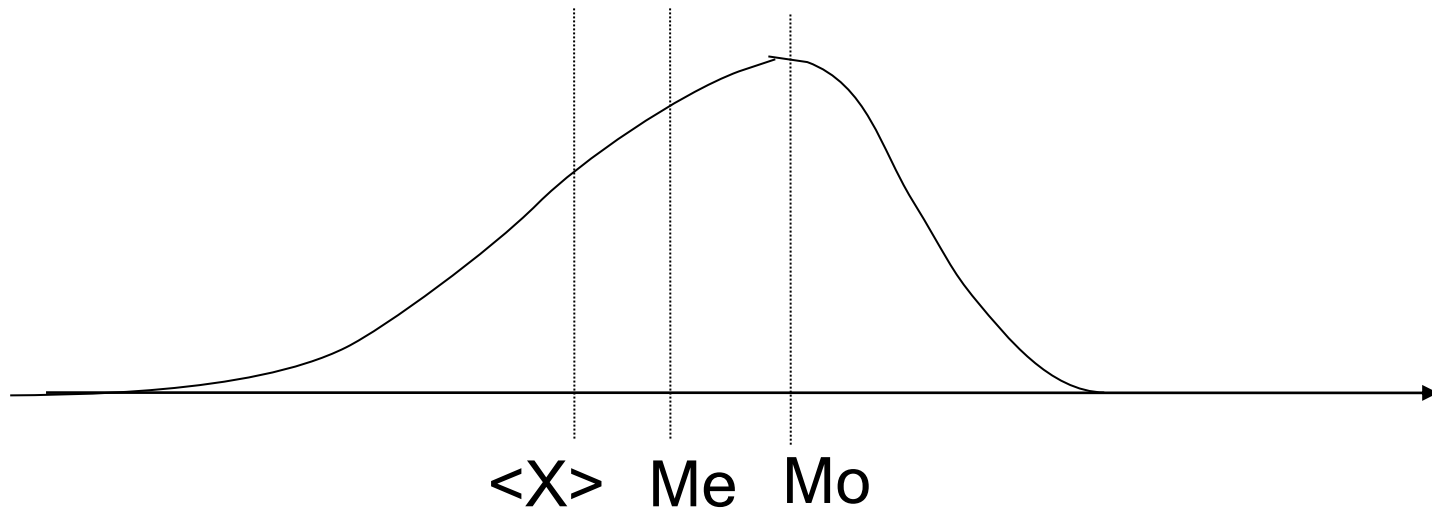


Courbe asymétrique à gauche

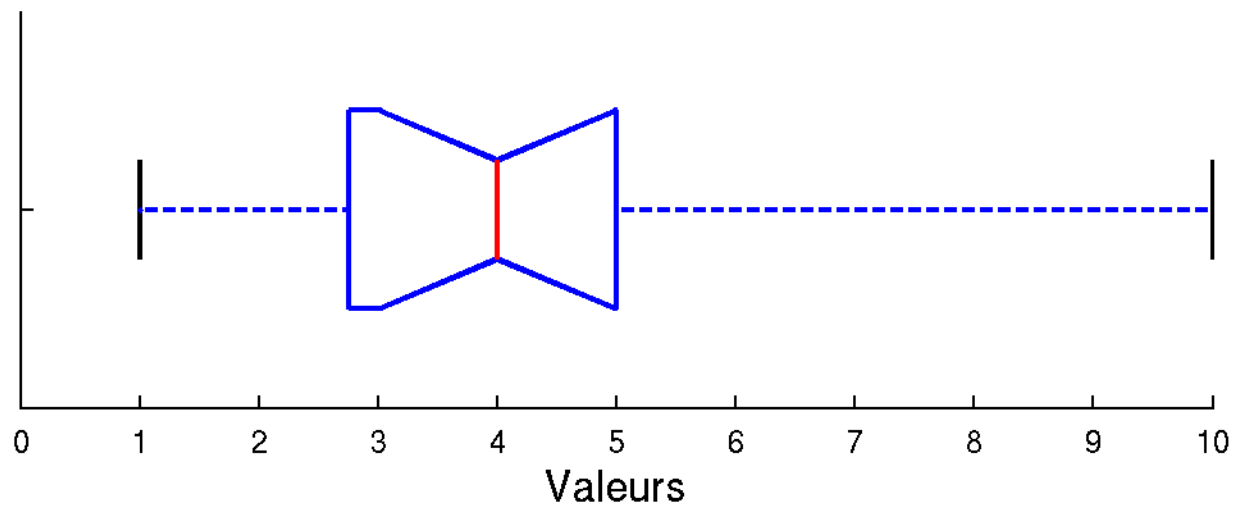
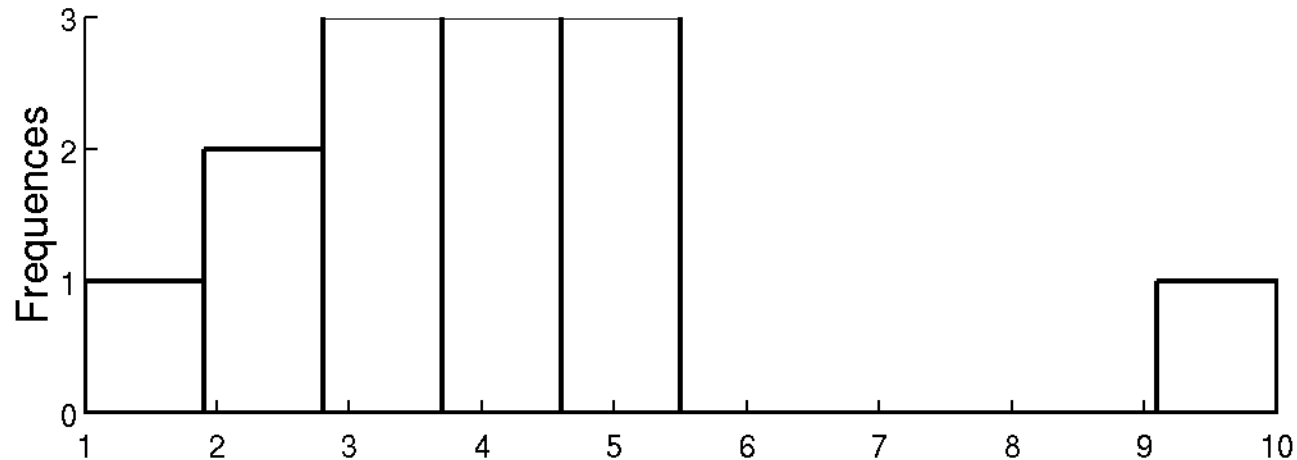
La courbe est allongée vers la gauche

Moyenne < Médiane < Mode

$$\alpha_3 < 0$$



Visualisation graphique de l'asymétrie



Aplatissement

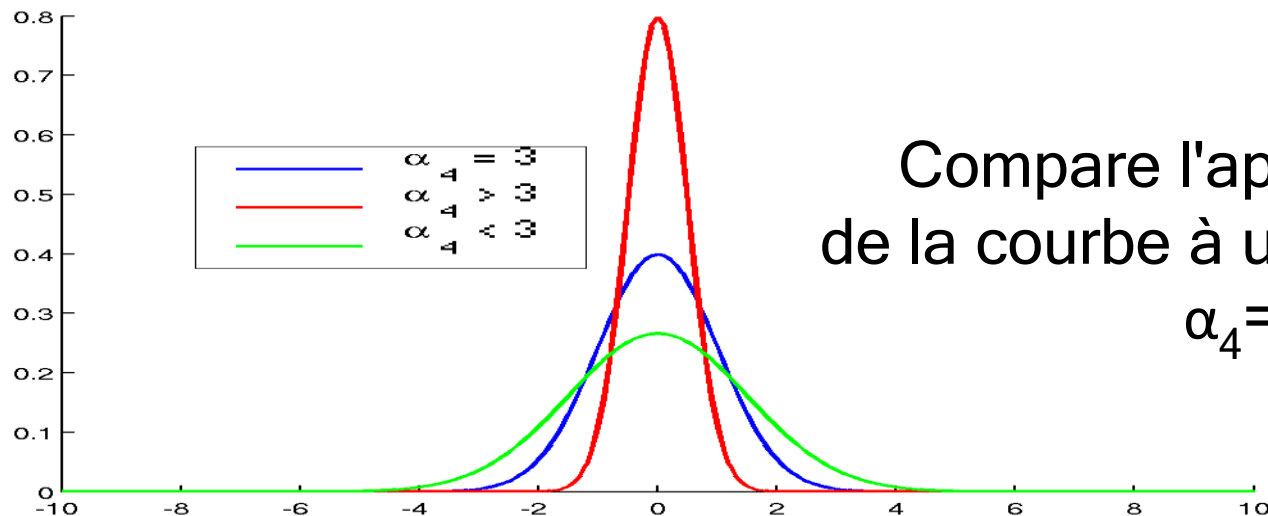
Coefficient d'aplatissement = kurtose

Population

$$\alpha_4 = \frac{M_4}{\sigma^4}$$

Échantillon

$$\alpha_4 = \frac{m_{4x}}{s_x^4}$$



Compare l'aplatissement
de la courbe à une loi normale
 $\alpha_4=3$

Présenter ses résultats

(a) Variable qualitative ou variable quantitative discrète prenant peu de valeurs

Tableau de distributions de valeurs

(b) Variable qualitative ou variable quantitative discrète prenant beaucoup de valeurs

Diagramme en bâtons

Présenter ses résultats

(c) Variable continue ou variable discrète prenant un grand nombre de valeurs

Précisez

(1) Les valeurs centrales (moyenne)

(2) La dispersion de vos résultats.

Et si vous avez de la place, montrez un histogramme.

Si vous devez comparer votre échantillon à d'autres échantillons, indiquez les barres d'erreurs ou utilisez un diagramme à moustaches

Les sondages disent-ils n'importe quoi ?

Sondage Ifop - L'EXPRESS / BFM / I-TELEVISION

Le 11 avril 2002

Fiche technique:

Échantillon de 1004 personnes, représentatif de la population française de 18 ans et plus, inscrites sur les listes électorales. La représentativité de l'échantillon a été assurée par la méthode des quotas (sexe, âge, profession du chef de famille) après stratification par région et catégorie d'agglomération. Les interviews ont eu lieu par téléphone au domicile des personnes interrogées. Du 5 au 6 avril 2002. La notice de ce sondage est consultable auprès de la Commission des Sondages.

Question:

Si dimanche prochain devait se dérouler le premier tour de l'élection présidentielle, pour lequel des candidats suivants y aurait-il le plus de chances que vous votiez ?

	5-6 avril (%)	28-29 mars (%)	Évolution (%)
Lionel JOSPIN	17.6	20.5	-3
Jacques CHIRAC	21	21	=
Jean-Marie LE PEN	13	10	+3
...

Commentez.

Variables qualitatives

- ▶ Mode
- ▶ Richesse
- ▶ Diversité de Shannon
 - ▶ Régularité
- ▶ Diversité de Simpson

Mode

Mode =
Classe de plus grande fréquence

Qualitatif	Richesse
Quantitatif	Diversité de Shannon
	Régularité
	Diversité de Simpson

Richesse

Richesse =

nombre de classes dans lesquelles se répartissent les éléments de l'échantillon

Similaire à l'**étendue**,
elle mesure l'ampleur de la gamme
des valeurs prises par les éléments de l'échantillon

Avec nos notations $\{v_i, f_i\}_{1 \leq i \leq D}$: Rich = D

Exemple : On inventorie les espèces présente dans une haie bordant un champ. La richesse est le nombre d'espèces compté.

Qualitatif	Richesse
Quantitatif	Diversité de Shannon
	Régularité
	Diversité de Simpson

Diversité de Shannon

Diversité =

Estimateur de la façon dont les éléments se répartissent entre les différentes catégories.

L'indice de diversité le plus utilisé est celui de
Shannon :

$$H' = - \sum_{k=1}^{Rich} f_k \log(f_k)$$

qui comme l'entropie en thermodynamique est maximal quand la répartition est la plus uniforme possible

Régularité

La diversité a le défaut de dépendre de la richesse
(du nombre de catégories)

Régularité =

Rapport de la diversité au logarithme de la richesse

$$R = \frac{- \sum_{k=1}^{Rich} f_k \log(f_k)}{\log(Rich)}$$

Qualitatif	Richesse
Quantitatif	Diversité de Shannon
	Régularité
	Diversité de Simpson

Diversité de Simpson

La diversité de Shannon d'un échantillon est **biaisée** :
si l'échantillon est de taille trop faible,
elle représente mal celle de la population

La **diversité de Simpson** corrige ce défaut :

$$D = 1 - \sum_{k=1}^{Rich} \frac{f_k (f_k - 1)}{n(n-1)}$$

n est le nombre d'élément de l'échantillon

Elle correspond à la probabilité que deux éléments
tirés au hasard dans l'échantillon
soient dans la même catégorie.

Elle est souvent utilisée en écologie

II – 2

**Statistique descriptive
d'une série statistique
double**

Séries statistiques doubles

La série statistique double se présente sous la forme d'une liste de couples de données :

X1	Y1
X2	Y2
X3	Y3
X4	Y4
X5	Y5
X6	Y6
...	...

On notera $\{x_i, y_i\}_{1 \leq i \leq n}$

II – 2 a)

Hiérarchisation

Tableaux de corrélation

L'équivalent du tableau de distribution de fréquences pour une série statistique double est un
tableau de corrélation

	Longueur d'ailes de mésanges noires		
	55-60	61-65	66-70
Mâle adulte	0	40	23
Mâle immature	0	46	2
Femelle adulte	6	62	1
Femelle immature	17	33	0

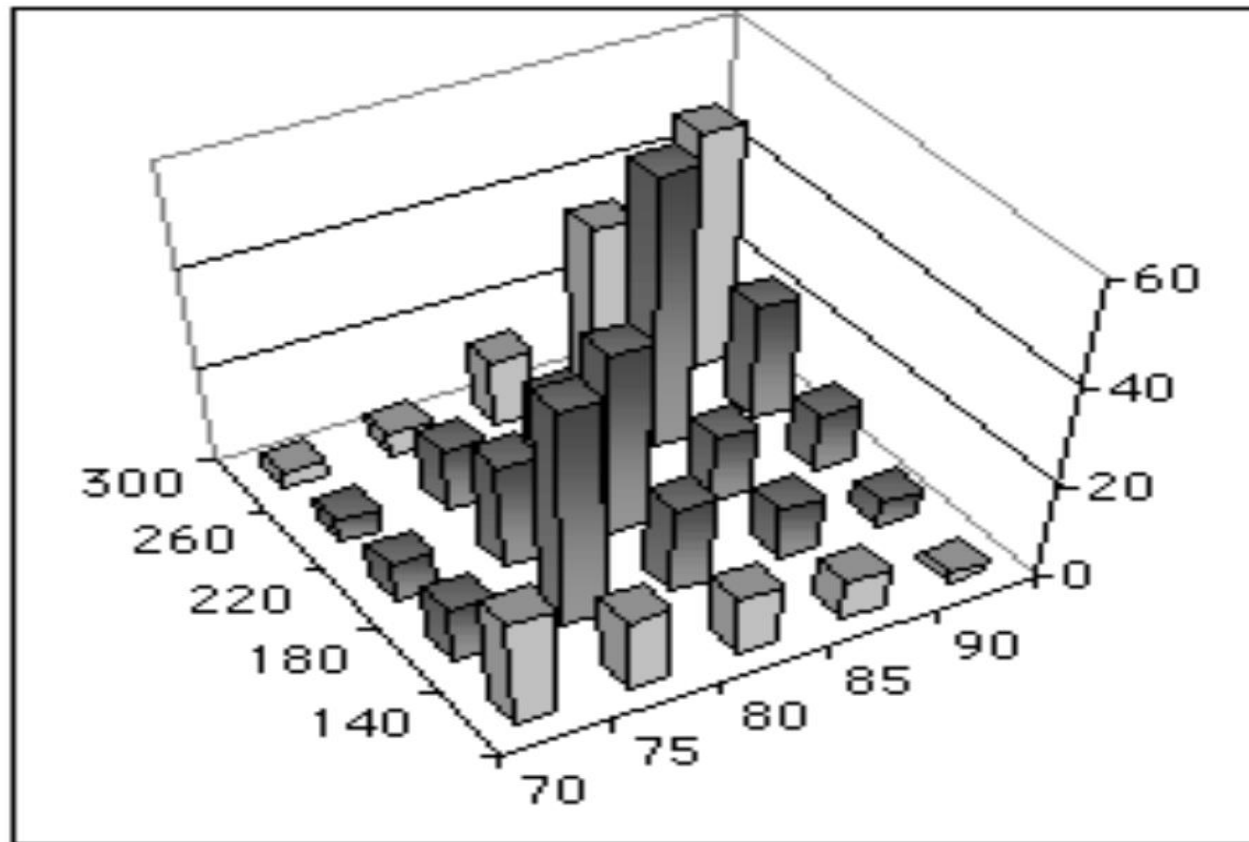
Dans le cas de deux variables qualitatives, on parle aussi de **tableau de contingence**

II – 2 b)

Représentations graphiques

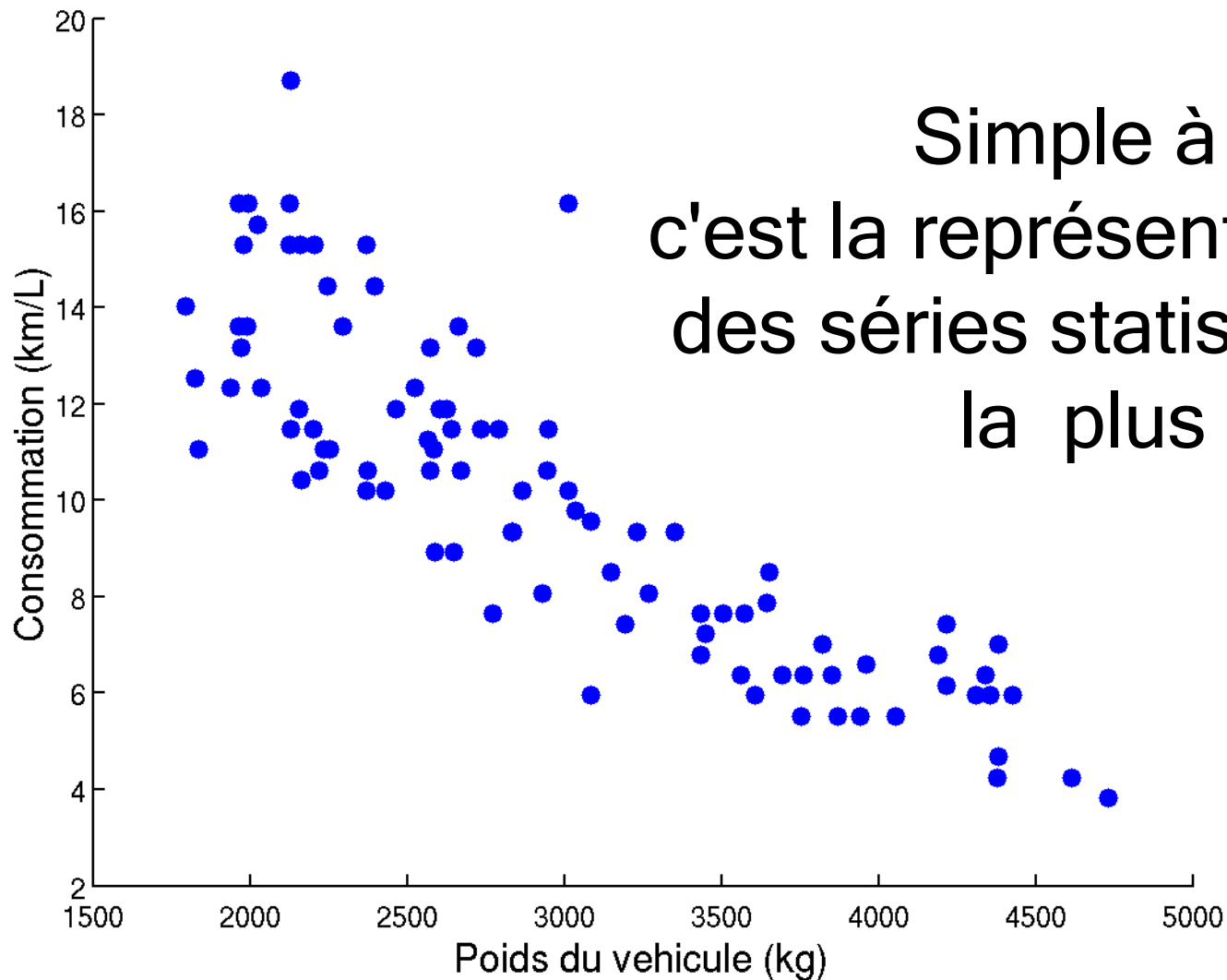
Stéréogramme

Généralisation 3D de la notion d'histogramme



Elle s'avère toutefois fastidieuse
à tracer et à interpréter

Diagramme de dispersion



II – 2 c)

Paramètres statistiques

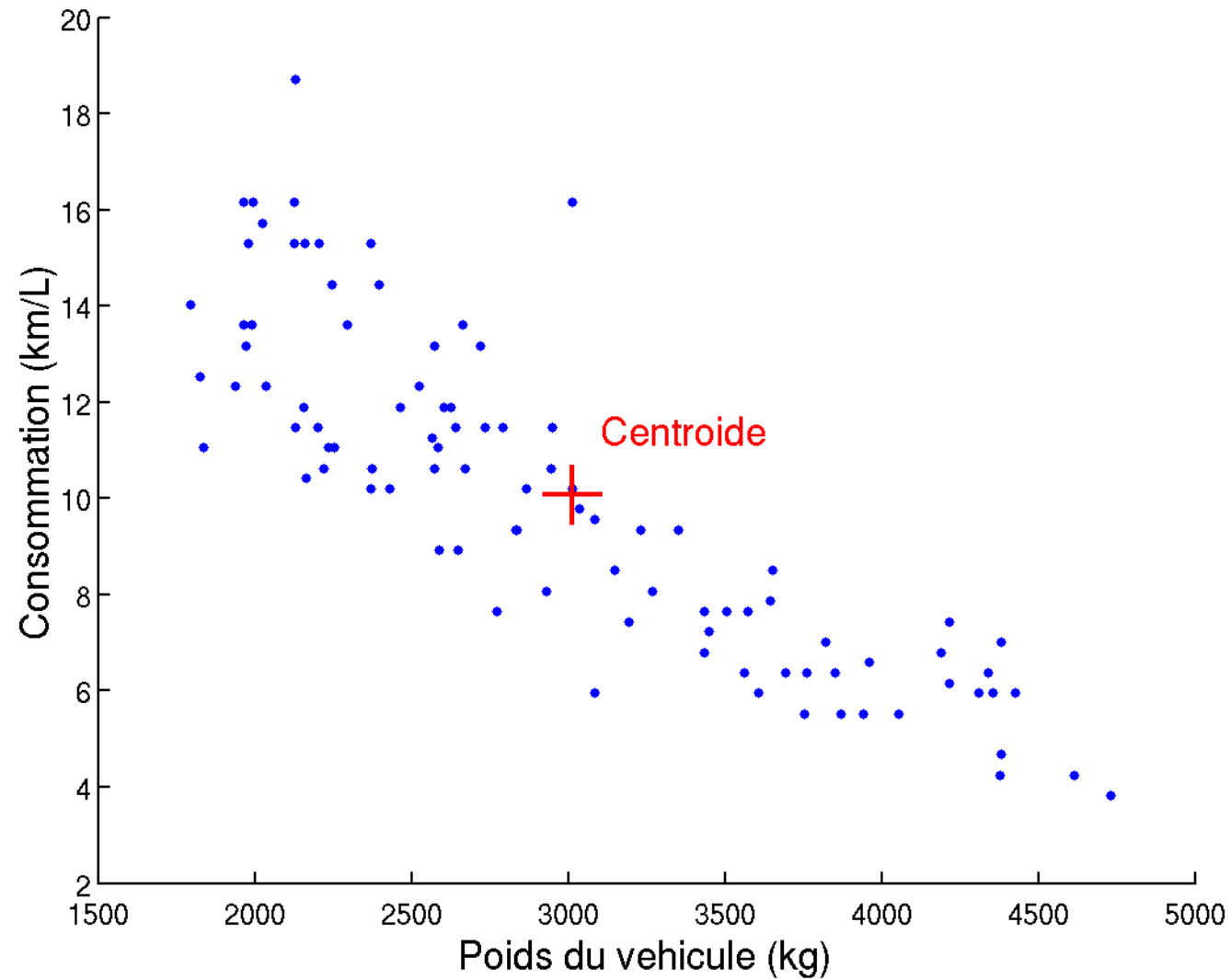
Centre de gravité

Centre de gravité =
centroïde (d'une série statistique $\{x_i, y_i\}$) =
point dont les coordonnées sont les moyennes
de chacun des coordonnées

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_N}{N} = \frac{\sum_{i=1}^N X_i}{N}$$

$$\bar{Y} = \frac{Y_1 + Y_2 + \dots + Y_N}{N} = \frac{\sum_{i=1}^N Y_i}{N}$$

Centre de gravité



Position	Moyenne
Dispersion	Médiane
	Mode

Médiane

La médiane n'a pas de sens
pour une série statistique double
car elle demande de classer préalablement
tous les éléments de la série.

Or, il n'y a pas de relation d'ordre naturelle en
2D

Mode

Mode =

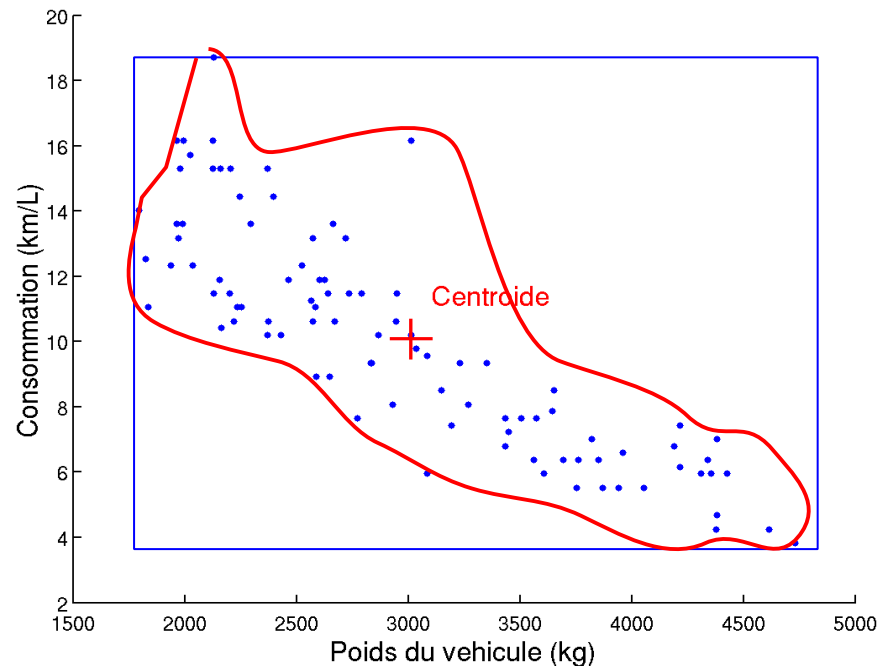
zone du plan correspondant au maximum de densité d'éléments.

Ce paramètre a les mêmes défauts que le mode d'une série statistique unidimensionnelle.

Amplitude

L'amplitude est mal définie.

Il est en effet possible de trouver **plusieurs zones du plan** qui délimitent les données.



On peut se restreindre à un **rectangle** de dimension
[amplitude(X) amplitude(Y)]
même si ce n'est pas toujours satisfaisant

Position	Amplitude
Dispersion	Quartiles
	(Co)variances

Quartiles

Pour les mêmes raisons
que dans le cas de la médiane,
on ne peut pas définir de quartiles

Variances

Comme pour les séries simples, il est possible de calculer la variance pour chacune des coordonnées de la série

Population

$$\sigma_X^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N}$$

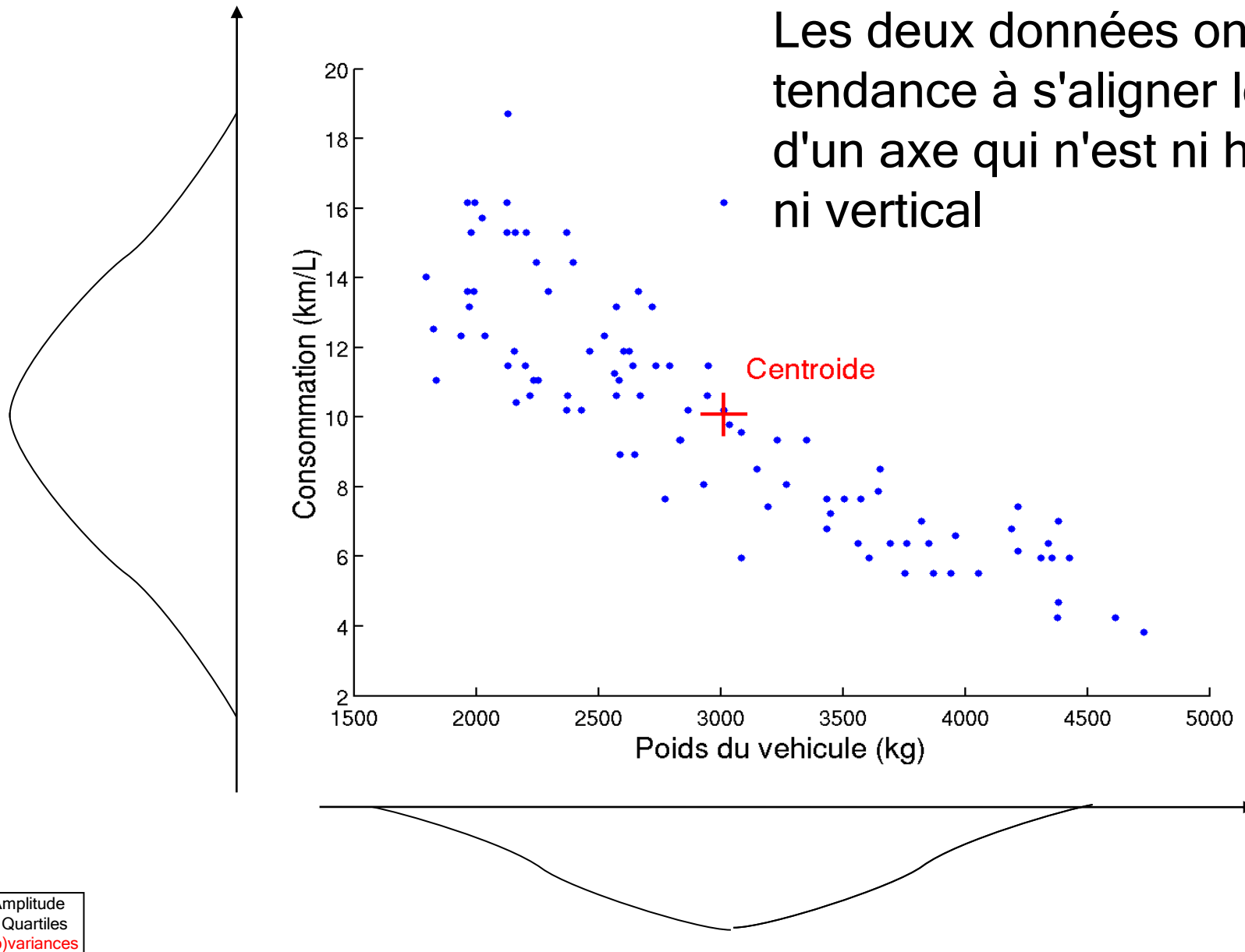
$$\sigma_Y^2 = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N}$$

Échantillon

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

$$s_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$$

Mais est-ce satisfaisant ?



Covariance

On introduit un nouveau paramètre pour traduire cette tendance d'une **variation** **conjointe** des deux données :
la **covariance**

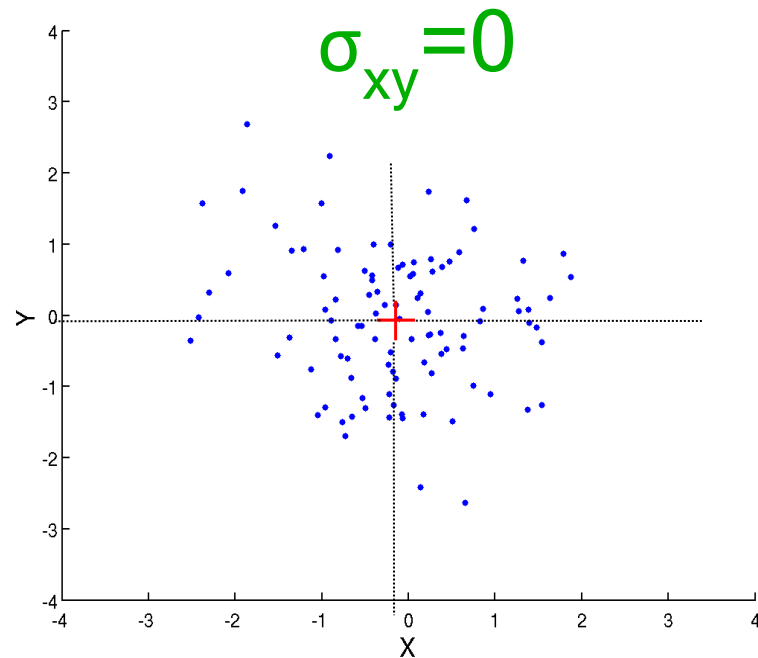
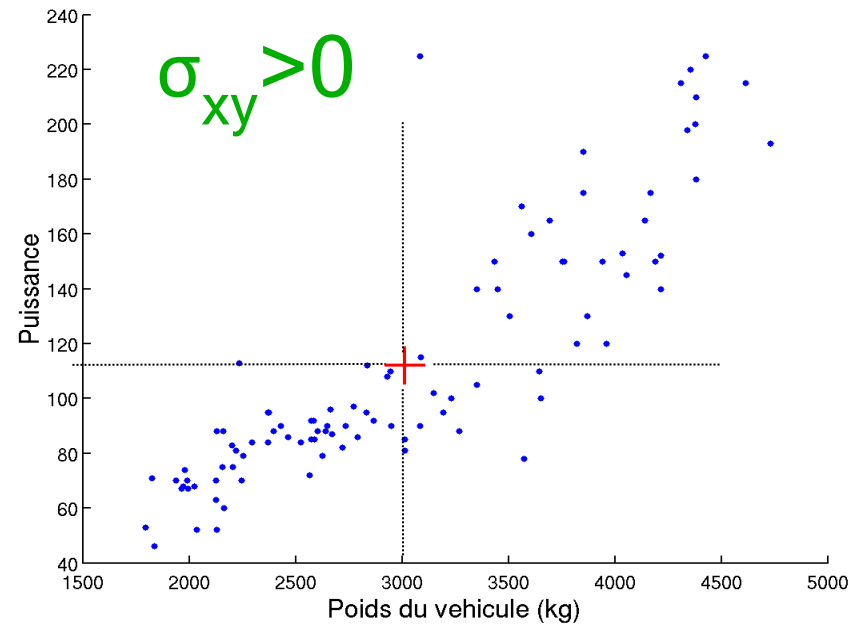
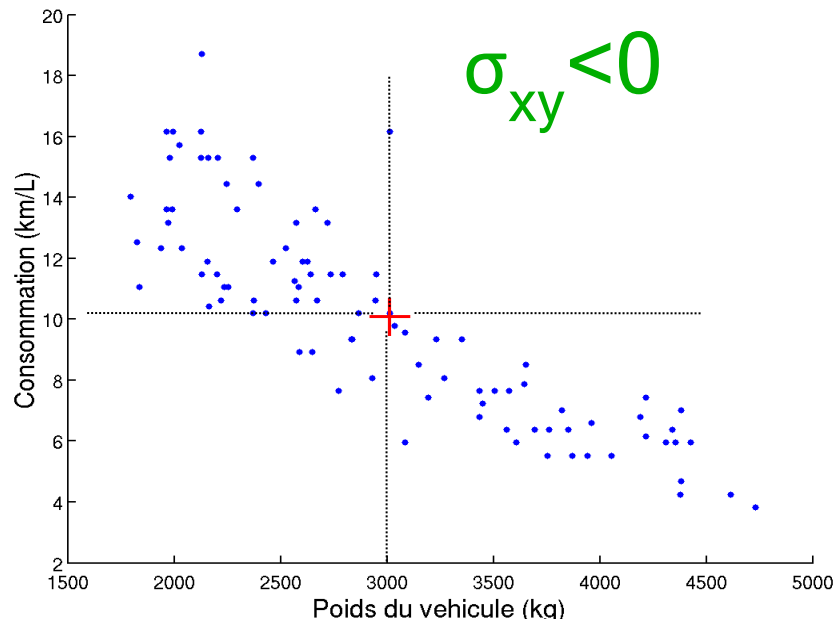
Population

$$\sigma_{XY} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{N}$$

Échantillon

$$s_{XY} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

Covariance



Position	Amplitude
Dispersion	Quartiles
	(Co)variances

Coefficient de corrélation

Coefficient de corrélation linéaire de
Spearman =

Rapport de la covariance sur les écarts-types
de chacune des variables

$$R = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

$R = (\text{intensité du couplage } X \text{ et } Y) / (\text{« bruit » sur } X \text{ et } Y)$

R est compris entre -1 et 1

Corrélation \neq Causalité

La mise en évidence d'une **corrélation** entre deux facteurs **ne démontre pas** l'existence d'une relation de **causalité** entre ces facteurs.

Exemple : The bell curve

La causalité se démontre par **expérience directe**.

Matrice de covariance

(1) La variance est un cas particulier de covariance avec $X=Y$

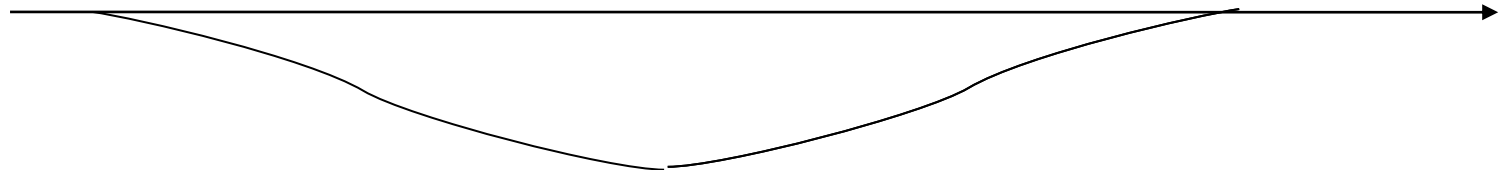
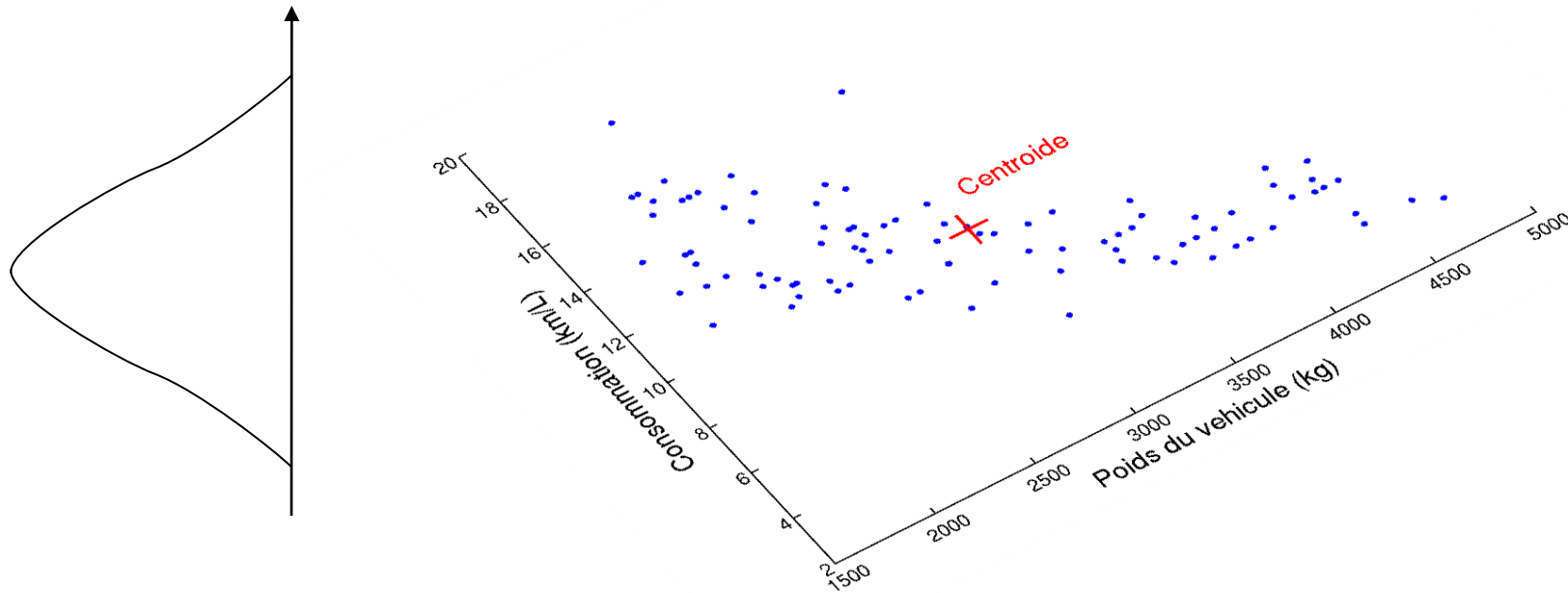
$$(2) \sigma_{XY} = \sigma_{YX}$$

Il est donc tentant de combiner ces informations dans une seule structure :
la matrice de covariance

$$\begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix}$$

Matrice de covariance

En tournant les axes (=changement de coordonnées), on arrive à une situation où le nuage de points s'aligne sur les nouveaux axes verticaux et horizontaux.



Position	Amplitude
Dispersion	Quartiles
	(Co)variances

Matrice de covariance

Les deux variables ont ainsi tendance à varier indépendamment et non plus conjointement.

Quantitativement : $\sigma_{XY} = 0$

Nous avons donc **diagonalisé**
la matrice de covariance

$$\begin{bmatrix} \sigma_{X'}^2 & 0 \\ 0 & \sigma_{Y'}^2 \end{bmatrix}$$

Présenter ses résultats

(a) Variable qualitative ou variable quantitative discrète prenant peu de valeurs

Tableau de distributions de valeurs

(b) Variable qualitative ou variable quantitative discrète prenant beaucoup de valeurs

Diagramme de dispersion en précisant le centroïde et la matrice de covariance

III Probabilités

- ✓ **Mots clés** : tirage aléatoire, densité de probabilité
- ✓ **Savoir calculer** ces valeurs
- ✓ **Connaître** les principales fonctions statistiques

LA STATISTIQUE INTRODUCTION



LA STATISTIQUE: QUELLE EST CETTE DISCIPLINE ?

Le Petit Robert

- Etude méthodique des faits sociaux par des procédés numériques (classements, dénombrements, inventaires chiffrés, recensements) destinée à renseigner et à **aider les gouvernements**.

Larousse

- Ensemble des méthodes qui ont pour objet la **collecte**, le **traitement** et l'**interprétation** de **données** d'observation relatives à un groupe d'individus ou d'unités.

Le Trésor de la Langue Française

- **Branche des mathématiques** ayant pour objet l'analyse (généralement **non exhaustive**) et l'interprétation de données quantifiables.



OBJECTIFS...

**Le but essentiel de la statistique
et
le rôle du statisticien**

sont précisément d'éviter le manque d'objectivité !



HISTOIRE DE LA STATISTIQUE

- **Statistique** : du latin *statisticum*
(ce qui se rapporte à l'Etat)
 - **Gottfried Achenwall (1746)** :
Premier enseignement de la statistique (Marburg,
Allemagne)
- En fait, origine plus ancienne :
- mot déjà utilisé dans un texte administratif de **Colbert**
(vers 1666)



LA STATISTIQUE DESCRIPTIVE

Au début,

la statistique a consisté à observer des faits

- **XIV^e siècle** : début des enregistrements des actes civils
(naissances, mariages, décès)

→ **Statistique descriptive**



LA STATISTIQUE INFÉRENTIELLE

- **XVII^e siècle** : Probabilités → estimations, prévisions

Extrapolation à partir d'une partie de la population
(W. Petty: estimation de la population londonienne, 1686)

→ **Statistique inférentielle**

Juger d'après un échantillon:

→ problème de la **représentativité**



LE XX^E SIÈCLE

Date cruciale : 3 novembre 1936

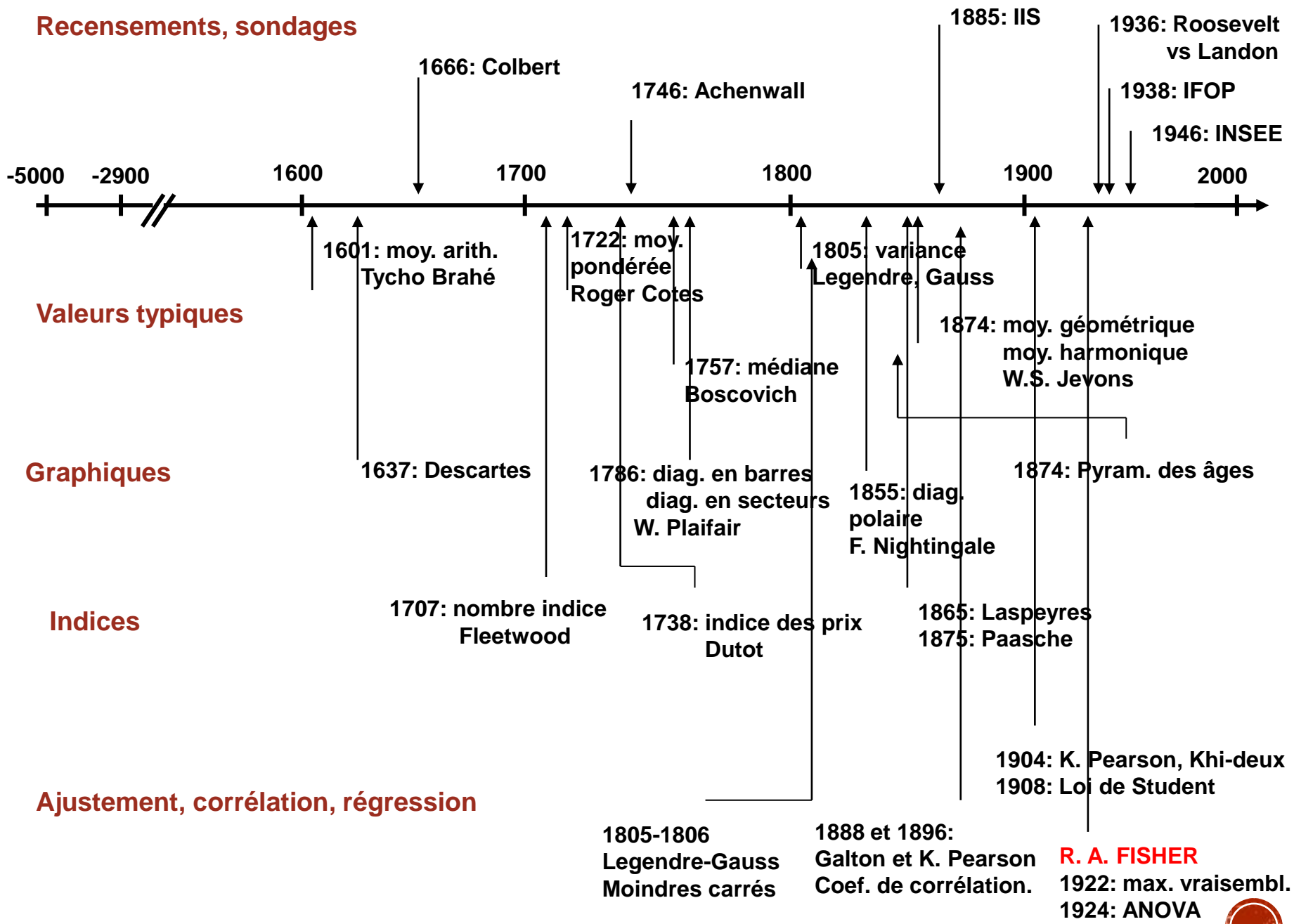
**Election présidentielle américaine
F. Roosevelt *versus* G. Landon**

- **Magazine *Literary Digest* : 2000 000 personnes (téléphone)
→ Landon**
- **G. Gallup : 3 000 personnes
→ Roosevelt**

Principe de la validité d'un sondage accepté

- **Avènement des ordinateurs**





LA STATISTIQUE NOUS DIT COMMENT:

- effectuer les mesures
- extraire l'**information** des données
- appréhender l'**incertitude** et revendiquer le **droit à l'erreur**
- **quantifier** le risque d'erreur



DANS LE BUT DE:

***CONNAÎTRE, C'EST MESURER* (LÉON BRUNSCHVIG)**

- Trouver et décrire une **relation**
 - Risque cardio-vasculaire et tabac

- Prendre une **décision**
 - Efficacité d'un médicament

- **Prévoir** et **planifier**
 - Budget prévisionnel (commune, gouvernement...)



L'INFORMATION

Matière première du XXI^e siècle
Sa production et son exploitation



STATISTICIENS



LE ROLE SOCIAL DU STATISTICIEN

Statisticiens → **Interlocuteurs privilégiés des décideurs**

- dans **tous** les secteurs d'activité :
(politique, économique, scientifique, industriel ...)

- à **tous** les niveaux :

(collecte de données, conception des systèmes d'information,
contrôle de la production, analyse et restitution des données, etc.)



LE STATISTICIEN SAIT QUE ...

- Les modèles (les hypothèses) validés ne sont que **provisaires**
- *Tous les modèles sont faux ; certains sont utiles*
(George Box)
- *Corrélation n'est pas causation*



LA VARIABILITE

❖ La variabilité : un caractère universel

En général, un critère de qualité dans une population

Exception : l'industrie

Qualité \equiv  de la variabilité

❖ Le statisticien doit savoir :

- appréhender
- analyser
- « gérer » la variabilité

❖ La variabilité est une entrave à l'inférence statistique



LES DOMAINES DE LA STATISTIQUE

- **Etude Série statistique univariée et bivariée**
- **L'analyse des données → Le data mining**
- **Les plans d'expériences**
- **Les sondages**
- **L'estimation**
- **Les tests statistiques**
- **Les séries chronologiques**
- **La modélisation**



LA METHODE EXPERIMENTALE

- **Observation** d'un phénomène
- **Formulation** d'une **hypothèse** (une théorie)
- **Vérification** de cette hypothèse

Exemple: Introduction à l'étude de la médecine expérimentale (Claude Bernard, 1865)



LES SONDAGES

- **Qualités d'un échantillon :**

- **la taille** (sondages d'opinion : $n \cong 1000$)

- **la représentativité** **sinon** → **biais**

→ **énormes précautions dans le choix de l'échantillon**

Rôle prépondérant du hasard



L'ANALYSE DES DONNEES

- **Data is the new oil** (Gerd Leonhard, 2006)
- Cette **matière première brute**
peut contenir de l'**information**
qu'il faut savoir extraire

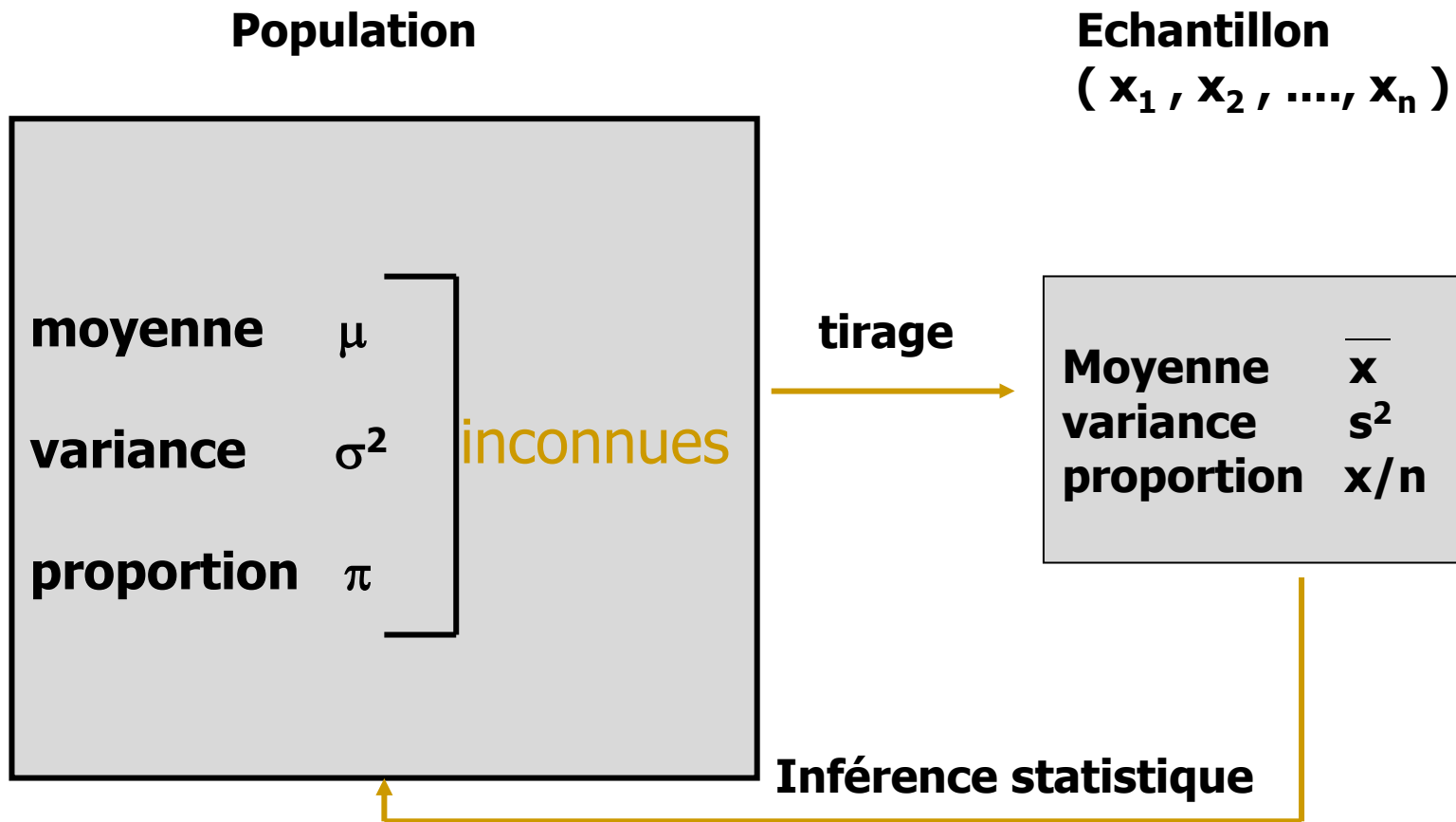


L'INFERENCE STATISTIQUE

- **Données observées : réalisation d'un phénomène aléatoire**
- **C'est un résultat parmi «beaucoup» d'autres possibilités**
- **«Diversité potentielle des résultats possibles»**
- **Ne pas en tenir compte → risque de conclusions erronées**
- **Remède: Statistique inférentielle et Théorie des probabilités**



L'INFERENCE STATISTIQUE



LA MODELISATION

- **En général, une observation dépend
d'un grand nombre de facteurs**
- **Cause**s multifactorielless
- **Les effets des facteurs ne sont pas simplement
additifs**
(présence d'interactions)



LA MODELISATION (SUITE)

Le statisticien va tenter :

- d'identifier les **facteurs prépondérants**
- évaluer leur importance relative
- expliciter le **lien** entre ces facteurs «causaux»
et le caractère étudié, à l'aide d'un

modèle statistique

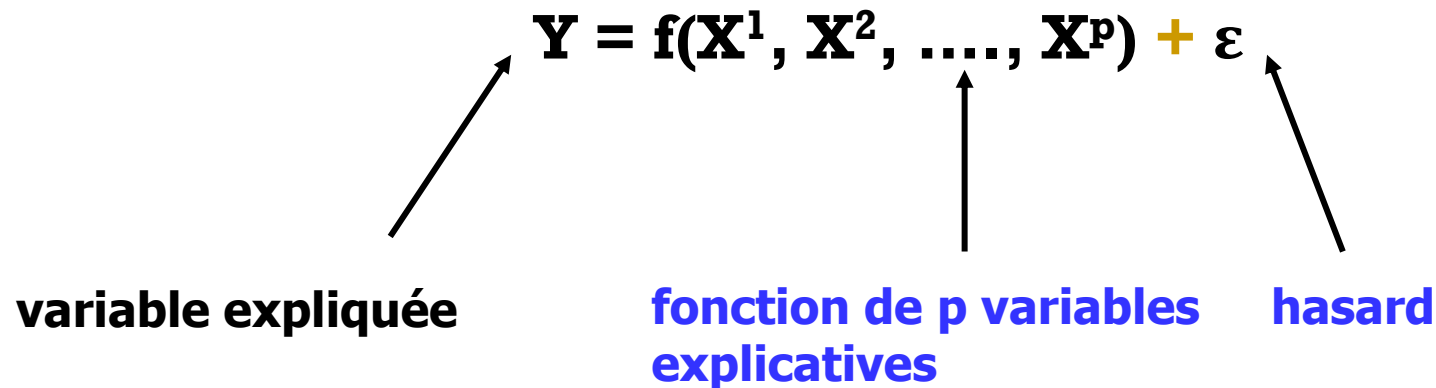


MODELE STATISTIQUE

L'univers est partagé en deux parties:

$$Y = f(X^1, X^2, \dots, X^p) + \varepsilon$$

variable expliquée fonction de p variables explicatives hasard



Le **hasard**: la somme de nos ignorances, selon Laplace



LE PRIX D'UNE VOITURE

- **variable expliquée : prix d'une voiture**
- **variables explicatives : cylindrée, puissance, vitesse, largeur, longueur, poids ...**

$$\begin{array}{lcl} \text{prix} = 1334 + 130,0 \text{ puissance} + 12,4 \text{ poids} \\ (\text{€}) & (\text{CV DIN}) & (\text{kg}) \end{array}$$



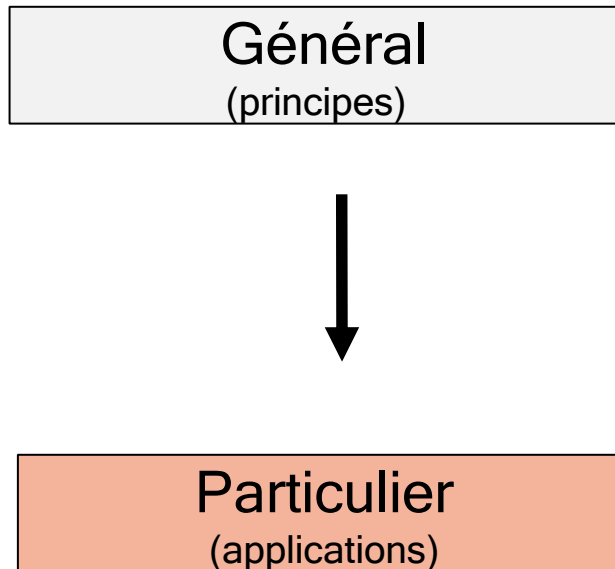
DÉMARCHE STATISTIQUE

- 1) On cherche à caractériser un phénomène qui concerne une certaine **population**
- 2) On ramène ce phénomène à la mesure d'une ou plusieurs **variables**, mesuré sur un **élément**.
- 3) Malheureusement, on ne peut pas faire les mesures sur toute la population (ce serait alors un **recensement**). On se restreint à un sous-ensemble, l'**échantillon**
- 4) On obtient un ensemble de valeurs, appelée la **série statistique**.
- 5) Le statisticien synthétise les données (**statistique descriptive**)
- 6) Le statisticien généralise les résultat de l'échantillon à toute la population (**statistique inductive/ inférentielle**)

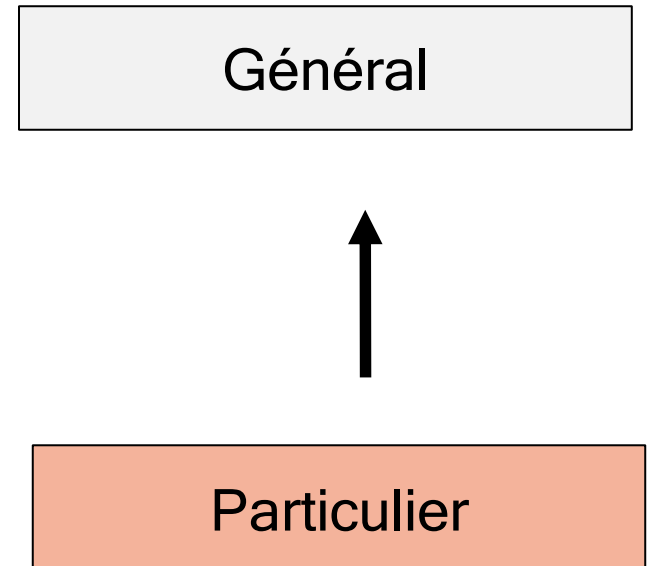


INDUCTION - DÉDUCTION

Déduction



Induction



DÉFINITIONS

Population = ensemble sur lequel porteront les **conclusions** de l'étude.

Échantillon = Sous-ensemble de la population
dans lequel seront collectées les **données** de l'étude.

Variable = Quantité **mesurée** lors de l'étude.



DÉJÀ APPARAISSENT LES PREMIERS PROBLÈMES...

- En quoi l'échantillon est-il **représentatif** de la population ?

Tirage aléatoire = l'échantillon est pris au hasard pour éviter d'introduire un **biais statistique**

- Comment mesurer les variables?



OBJETIFS DU COURS

Le but du cours est que vous maîtrisiez le travail interprétatif du statisticien :

Statistique descriptive

Analyse de données

Statistique inductive



STATISTIQUE DESCRIPTIVE

- **Mots clés** : distribution statistique, {diagramme en barre, histogramme}, {moyenne, médiane, mode}, {variance, écart type}
- **Présenter** des données statistiques....
- Savoir **calculer et interpréter** les différents paramètres statistiques



NOTATIONS SUR LES SÉRIES STATISTIQUES

$\{...\}$ = collection d'éléments

$\{x_i\}_{1 \leq i \leq n}$ est un raccourci pour
 $\{x_1, x_2, \dots, x_i, \dots, x_n\}$

Je le noterai parfois simplement $\{x_i\}_n$ ou $\{x_i\}$

Les notations où interviennent des majuscules se réfèrent à des recensements (directement sur toute la population). Ainsi :

$\{X_i\}_N$



MÉTHODES

Plusieurs techniques sont disponibles :

- ▶ **Classification** des données
- ▶ **Visualisation graphique**
- ▶ **Quantification** à l'aide de paramètres statistiques

Difficulté croissante avec le nombre de variables



SÉRIES STATISTIQUES

Série statistique simple = ensemble de données relatives à **une** variable mesurée sur un échantillon ou une population d'éléments

Série statistique double = ensemble de **couples** de données relatives à **deux** variables mesurées sur un échantillon ou une population d'éléments

Série statistique multiple = ensemble de **multiplats** de données relatives à **plusieurs** variables mesurées sur un échantillon ou une population d'éléments

