

**COURS
DE
PROBABILITÉS ET STATISTIQUE DESCRIPTIVE**

Table des matières

1	Les données statistiques	3
1.	Les variables statistiques-éléments de vocabulaire	3
2.	Les types de variables	3
2.1.	Variables qualitatives	3
2.2.	Variables quantitatives	4
2	Tableaux statistiques	5
1.	Les variables qualitatives	5
1.1.	Tableaux de distribution de fréquence absolues, relatives et cumulées	5
2.	Les variables quantitatives discrètes	6
2.1.	Tableaux de distribution de fréquences	6
3.	Les variables quantitatives continues	7
3.1.	Tableaux de distribution de fréquences-fréquences cumulées	7
3	Représentations graphiques	9
1.	Les variables qualitatives	9
2.	Les variables quantitatives	10
2.1.	Les variables quantitatives discrètes	10
2.2.	Les variables quantitatives continues	11
4	Résumés numériques d'une variable statistique	12
1.	Paramètre de tendance centrale	12
1.1.	Le mode	12
1.2.	La moyenne	13
1.3.	La médiane	14
1.4.	Quantiles	16
1.5.	Utilisation des paramètres de tendance centrale	17
2.	Paramètres de dispersion	17
2.1.	L'étendue	17
2.2.	L'intervalle inter-quartile	18

2.3.	La variance et l'écart-type	18
3.	Changement de variable linéaire ou affine - Variable centrée réduite	19
3.1.	Changement de variable linéaire ou affine	19
3.2.	Variable centrée réduite	20
4.	Boîtes à moustaches	20
5	Liaison entre deux variables	22
1.	Généralités	22
1.1.	Tableau de contingence	22
1.2.	Distribution marginale	23
1.3.	Distribution conditionnelle	23
2.	Liaison linéaire entre deux variables quantitatives	24
2.1.	Moyennes et variances marginales	25
2.2.	Moyennes et variances conditionnelles	25
2.3.	Covariance	25
2.4.	Coefficient de corrélation	26
2.5.	Régression linéaire	27
2.6.	Régression linéaire après transformation d'une variable	28
3.	Liaison entre deux variables qualitatives	28
3.1.	Mesure de la liaison entre deux variables qualitatives	28
3.2.	Représentation graphique	29
4.	liaison entre variable qualitative et une variable quantitative	30
4.1.	Classement des données et distributions marginales	30
6	Denombrement	32
1.	Cardinal d'un ensemble fini	32
2.	Principes de comptage	32
2.1.	Principe additif	32
2.2.	Principe multiplicatif	33
3.	Arrangements	33
3.1.	Arrangements avec répétition	33
3.2.	Arrangements sans répétition	34
3.3.	Permutation	34
4.	Combinaisons	35
4.1.	Binôme de Newton	35
5.	Quel modèle choisir ?	35

7	Espace probabilisé	37
1.	Expérience aléatoire	37
2.	Probabilité	39
3.	Modélisation d'une expérience aléatoire	43
4.	Probabilités conditionnelles, indépendance	43
4.1.	Probabilité conditionnelle	43
4.2.	Indépendance	45
8	Variables aléatoires réelles	47
1.	Généralités	47
2.	Vecteurs aléatoires	49
2.1.	Vecteurs aléatoires discrètes	49
2.2.	Vecteurs aléatoires à densité	50
2.3.	Matrice de variance-covariance	51
2.4.	Fonction caractéristique	52
2.5.	Vecteurs gaussiens	52
2.6.	Généralités	52
2.7.	Lois dérivées	53
3.	Convergence	54
3.1.	Convergence de variables aléatoires	54
3.2.	Lois des grands nombres	55
3.3.	Théorème Central Limite	55
4.	Approximation	56

PARTIE I : STATISTIQUE DESCRIPTIVE

Introduction

Recueillir et analyser les données sont les deux objectifs fondamentaux de la Statistique. Pour parvenir à cela, il faut suivre les étapes suivantes :

1. La collecte des données : définir l'objet étudié, les variables statistiques mises en cause, le questionnaire et fabriquer l'échantillon représentatif (sondage, plan d'expériences...)
2. Une fois les données collectées et corrigées, les visualiser sous forme de tableaux ou graphes et les résumer grâce à des paramètres qui permettent de dégager les caractéristiques essentielles du phénomène étudié (statistique descriptive, analyse des données)
3. L'étape de la modélisation (statistique inférentielle) est de fournir des résultats relatifs à une population à partir de mesures statistiques réalisées sur des échantillons. La statistique inférentielle fournit des éléments permettant de spécifier du mieux possible, à partir de l'échantillon observé, le modèle probabiliste qui a engendré les données. Nous entendrons le terme de modèle dans le sens d'une formalisation mathématique supposée reproduire de manière approchée la réalité d'un phénomène dans le but d'en reproduire le fonctionnement pour permettre de comprendre, de prédire et/ou d'agir.

Les méthodes statistiques sont utilisées dans de nombreux domaines tels que l'ingénierie (contrôle de qualité de fabrication...), la médecine (expérimentation de nouveaux traitements...), les sciences économiques et sociales, l'économetrie, la démographie, et bien d'autres.

Chapitre 1

Les données statistiques

1. Les variables statistiques-éléments de vocabulaire

On observe un **échantillon** composé de n individus appartenant à une même **population** de taille N . Chaque individu de l'échantillon est observé à travers des caractéristiques, caractères ou indicateurs appelés **variables**. Une série statistique $(x_1; x_2; \dots; x_n)$ est la suite des valeurs prises par une ou plusieurs variables pour chacun des individus de l'échantillon. Chaque valeur prise par une ou plusieurs variables est appelé une **modalité**.

Exemple 1. *Un questionnaire est distribué à 150 personnes dans la cour d'un établissement secondaire. Il comporte diverses questions. La population = l'ensemble des élèves de cet établissement. L'échantillon = les étudiants ayant répondu au questionnaire. Un individu est une personne interrogée. Les variables correspondent aux questions posées : l'âge, la taille, la couleur des yeux, etc.*

En statistique descriptive la population c'est l'ensemble des individus effectivement étudiés, sans chercher à étendre les constatations faites à une population plus vaste, ce qui relève de la statistique inférentielle.

2. Les types de variables

Le type d'une variables depend de la nature de ses modalités. On distingue plusieurs types de variables :

2.1. Variables qualitatives

Une variable est dite qualitative lorsque les réponses possibles à la question posée, ou les modalités, ne correspondent pas à une quantité mesurable par un nombre mais appartiennent à un groupe de catégories.

Exemple 2. *le sexe, la couleur des yeux, la mention au baccalauréat, la fréquence d'une activité (jamais, rarement, parfois, souvent, très souvent).*

on distingue :

- les variables **qualitatives nominales** : il n'y a pas d'hierarchie entre les différentes modalités ; exemple : sexe, couleur des yeux.
- les variables **qualitatives ordinales** : les différentes modalités peuvent être ordonnées de manière naturelle ; exemple : la mention au baccalauréat, la fréquence d'une activité, niveau d'études scolaires : école primaire < 1er cycle < CAP < BEP < Bac < BTS < DEUG <

Remarque 1.1. *Certaines variables nominales peuvent être désignées par un code numérique, qui n'a pas de valeur quantité. Exemple : le code postal, le sexe (1 = garçon, 2 = fille)*

2.2. Variables quantitatives

Les réponses correspondent à des quantités mesurables et sont données sous forme de nombre.

On distingue :

- Les variables quantitatives discrètes : elles prennent leurs valeurs dans un ensemble discret, le plus souvent fini. Exemple : le nombre d'enfants, la pointure du pied.
- les variables quantitatives continues : elles peuvent prendre toutes les valeurs d'un intervalle réel. Exemple : la taille des individus, une note à un examen.

Remarque 1.2. *L'âge peut être vu et traité comme une variable quantitative discrète ou continue suivant la précision que l'on choisit et le nombre de valeurs qu'il prend au sein de la population. Il peut également exister des variables basées sur l'âge qui sont qualitatives. Si dans un sondage on pose la question "quelle est votre tranche d'âge parmi les possibilités suivantes : - de 25 ans, entre 25 et 45, entre 40 et 60 et +60 ans", on peut voir la variable "tranche d'âge" comme une variable qualitative ordinale*

Chapitre 2

Tableaux statistiques

1. Les variables qualitatives

Exemple 3. On s'intéresse à la variable "couleur des yeux" sur un groupe de 20 personnes. On code chaque modalité de la manière suivante : M =marron, V =vert, N =noir, B =bleu. On obtient la série statistique suivante :

$M, V, M, M, M, M, M, N, M, N, M, M, B, M, M, M, B, M, M, M.$

1.1. Tableaux de distribution de fréquence absolues, relatives et cumulées

Exemple 4. Pour l'exemple précédent, on remplit le tableau suivant :

Couleur des yeux	M	V	N	B	Total
Effectif					
Proportion					

Tableau-type : On choisit une notation pour la variable, par exemple : X . n désigne le nombre d'individus dans l'échantillon. on note $C_1, \dots, C_k$ les k modalités de la variable. Pour $1 \leq j \leq k$, on note

- n_j l'effectif associé à la modalité C_j (le nombre d'individus pour lesquels la valeur prise par la variable est C_j),
- $f_j = n_j/n$ la fréquence relative ou proportion associée à cette modalité,
- et si la variable est qualitative **ordinaire** : $N_j = n_1 + n_2 + \dots + n_j$ resp. $\phi_j = f_1 + f_2 + \dots + f_j$ la fréquence absolue (effectif) cumulée croissante resp. la fréquence relative cumulée croissante pour cette modalité (avec la convention : $\phi_0 = 0$). Elle n'a de sens que si la variable est qualitative ordinaire et si les modalités $C_1, C_2, \dots, C_k$ sont ordonnées suivant l'ordre croissant naturel (ou hiérarchique ascendant) qui règne parmi ces modalités. Exemple : niveau d'études scolaires : école primaire < 1er cycle < CAP < BEP < Bac < BTS < DEUG <

Le tableau suivant est un tableau-type qui permet de résumer les données.

Variable X	C_1	C_2	$\dots$	C_k	Totales
Fréquence absolue ou effectif	n_1	n_2	$\dots$	n_k	n
Fréquence relative ou proportion	$f_1 = n_1/n$	$f_2 = n_2/n$	$\dots$	$f_k = n_k/n$	1
Fréquence relative cumulée*	$\phi_1 = f_1$	$\phi_2 = f_1 + f_2$	$\dots$	$\phi_k = 1$	pas de sens

*Attention : uniquement dans le cas de variables qualitatives ordinales.

2. Les variables quantitatives discrètes

Exemple 5. On s'intéresse à la variable "pointure" (que l'on notera P) sur un groupe de 20 personnes. On obtient la série statistique suivante :

39, 43, 38, 39, 39, 42, 44, 44, 48, 40, 44, 43, 41, 37, 39, 38, 45, 41, 44, 44.

2.1. Tableaux de distribution de fréquences

Exemple 6. Pour la variable P , on remplit le tableau suivant :

P	37	38	39	40	41	42	43	44	45	46	47	48
Effectif												
Proportion												
Proportion cumulée												

On note $v_1, v_2, \dots, v_k$ les k valeurs différentes que peut prendre la variable avec $v_i < v_j$ si $i < j$ (on n'en rencontrera pas pas d'exemple dans ce cours, mais une variable discrète peut prendre une infinité de valeurs). Pour $1 \leq j \leq n$, on note n_j l'effectif des individus pour lesquels la variable prend la valeur v_j . On note f_j la fréquence relative ou proportion pour la valeur v_j et $\Phi_j = f_1 + \dots + f_j$ la j -ème fréquence relative cumulée (avec la convention : $\Phi_0 = 0$). On résume habituellement les données comme dans le tableau-type suivant :

Valeurs prises par la variable	v_1	v_2	$\dots$	v_k	Total
Fréquence absolue	n_1	n_2	$\dots$	n_k	n
Fréquence relative	$f_1 = n_1/n$	$f_2 = n_2/n$	$\dots$	$f_k = n_k/n$	1
Fréquence relative cumulée ↗	$\Phi_1 = f_1$	$\Phi_2 = f_1 + f_2$	$\dots$	$\Phi_k = 1$	pas de sens

On définit de même pour la valeur v_j la fréquence cumulée décroissante :

$$G_j = \frac{1}{n}(n_j + \dots + n_k) = f_j + \dots + f_k.$$

La quantité $N_j^d = n_j + \dots + n_k$ est appelée effectif cumulé décroissant.

3. Les variables quantitatives continues

Exemple 7. On s'intéresse à la taille, notée T et exprimée en mètre, de 20 individus. On a obtenu la série statistique suivante :

1,72; 1,87; 1,66; 1,73; 1,64; 1,77; 1,80; 1,81; 1,60; 1,78; 1,83; 1,75; 1,70; 1,58; 1, 68; 1,66; 1,93; 1,75; 1,80; 1,85.

3.1. Tableaux de distribution de fréquences-fréquences cumulées

Les données brutes de la variable pour chaque individu sont notées $x_1, \dots, x_n$. Elle peuvent prendre n'importe quelle valeur dans un interval de $\mathbb{R}$ et il est très rare d'avoir deux fois la même valeur pour deux individus différents. Il serait donc unitile de tracer un diagramme en bâton comme dans le cas d'une variable discrète : il consisterait en un amoncellement illisible de bâton de hauteur $1/n$. On choisir donc de faire un **Regroupement en classe**.

- L'intervalle où la variable prend ses valeurs est divisé en k classes : $[b_0, b_1[, [b_1, b_2[, \dots, [b_{k-1}, b_k[$ (il est possible d'avoir des bornes infinies).
- Pour $1 \leq j \leq n$, on note n_j l'effectif associé à la classe $[b_{j-1}, b_j[, f_j = n_j/n$ la fréquence relative associé à cette classe et $\Phi_j = f_1 + \dots + f_j$ la j -ème fréquence cumulée (avec la convention $\Phi_0 = 0$)
- On note $a_j = b_j - b_{j-1}$ l'amplitude de la classe $[b_{j-1}, b_j[$.
- On note $d_j = f_j/a_j$ la densité de proportion pour la classe $[b_{j-1}, b_j[$.

Exemple 8. de la taille

T	[1, 50; 1, 65[	[1, 65; 1, 70[	[1, 70; 1, 75[	[1, 75; 1, 80[	[1, 80; 1, 85[	[1, 85; 2, 00[
Effectif						
Proportion						
Proportion cumulée						
Amplitude						
Densité de proportion						

Remarque 2.1. – la densité de la proportion permet de comparer les effectifs dans chaque classe en tenant compte de la taille de ces classes (cf. la notion de densité de la population en géographie).

– Dans le cas de classes qui ont toutes les même longueur, il n'est pas nécessaire de calculer la densité de proportion, il est suffisant d'étudier les fréquences relatives ou absolues (qui sont directement proportionnelle a la densité de proportion).

Tableau-type

Variable X	$[b_0, b_1[$	$[b_1, b_2[$	$\dots$	$[b_{k-1}, b_k[$	Total
Fréq. relative	$f_1 = n_1/n$	$f_2 = n_2/n$	$\dots$	$f_k = n_k/n$	1
Fréq. relative cumulée	$\Phi_1 = f_1$	$\Phi_2 = f_1 + f_2$	$\dots$	$\Phi_k = 1$	
Amplitude	$a_1 = b_1 - b_0$	$a_2 = b_2 - b_1$	$\dots$	$a_k = b_{k-1} - b_k$	
Densité de proportion	$d_1 = f_1/a_1$	$d_2 = f_2/a_2$	$\dots$	$d_k = f_k/a_k$	

Remarque 2.2. *Contrairement au cas d'une variable qualitative ou discrète, ce tableau représente une perte d'information par rapport aux données brutes*

Chapitre 3

Représentations graphiques

1. Les variables qualitatives

Pour une variable ou caractère qualitatif, on utilise principalement trois types de représentation graphique : le diagramme en bâtons, la représentation par tuyaux d'orgue et la représentation par secteurs.

- **Diagramme en bâtons** : en abscisse sont disposées les différentes modalités, de façon arbitraire aux quelles on associe des segments espacés entre eux dont les longueurs (en ordonnée) sont proportionnelles à l'effectif ou à la fréquence relative de chaque modalité. Préciser le nom des axes, le nom du graphique et la source des informations

Nous appelons polygone statistique, ou diagramme polygonal, la ligne obtenue en joignant les sommets des bâtons.

Exemple :

Caractère : catégorie socio-professionnelle.

Ouvriers = O, Cadre moyen = CM, Cadre supérieur = CS.

caractère	O	CM	CS
Effectifs	20	10	5

- **Diagramme en tuyaux d'orgue** : en abscisse sont disposées les différentes modalités, de façon arbitraire aux quelles on associe des rectangles espacés entre eux, de largeur constante, dont la hauteur (en ordonnée) sont proportionnelle à l'effectif ou à la fréquence relative de chaque modalité. Préciser le nom des axes, le nom du graphique et la source des informations. Dans le cas d'une variable qualitative ordinale, on peut également construire le diagramme en tuyau d'orgue des effectifs ou des proportions cumulés.

Exemple 9.

- **Diagrammes en secteurs** : chaque modalité est représentée par un secteur de disque dont l'angle est proportionnel à l'effectif ou à la fréquence de la modalité (ou pourcentage).

Ces diagrammes conviennent très bien pour des données politiques ou socio-économiques.

Dans un diagramme circulaire (cercle complet), l'effectif total ou la fréquence relative 1 (ou le pourcentage 100%) correspond à l'angle 360° .

Pour représenter les données sur un diagramme semi-circulaire (demi-cercle), il suffira de calculer les mesures des secteurs angulaires par rapport à 180° .

Exemple 10.

2. Les variables quantitatives

Il existe deux types de représentation graphique d'une distribution statistique à caractère quantitatif :

- Le diagramme différentiel correspond à une représentation des effectifs ou des fréquences.
- Le diagramme intégral correspond à une représentation des effectifs cumulés, ou des fréquences cumulées.

2.1. Les variables quantitatives discrètes

- **Diagramme différentiel** : diagramme en bâtons

On trace un graphique avec

- sur l'axe des abscisses les différentes valeurs prises par la variable, placées en **respectant une échelle**,
- en ordonné les fréquences relatives ou les fréquences absolues.
- Pour chaque valeur v_j on construit un bâton vertical à l'abscisse v_j , de hauteur proportionnel a la fréquence de la valeur v_j .

Exemple : peinture.

Nous appelons polygone statistique, ou diagramme polygonal, la ligne obtenue en joignant les sommets des bâtons.

- **Diagramme intégral** : courbe en escaliers des effectifs cumulés ou des fréquences cumulées.

Fonction de répartition empirique

La fonction de répartition empirique permet de décrire la série statistique de manière complète. Elle est définie sur $\mathbb{R}$ et prend ses valeurs dans $[0, 1]$. Pour x dans $\mathbb{R}$, elle est définie par

$$F(x) = \begin{cases} 0 & \text{si } x < v_1 \\ \Phi_j & \text{si } v_j \leq x < v_{j+1} \\ 1 & \text{si } v_k \leq x \end{cases}$$

Exemple 11. *Pointure*

2.2. Les variables quantitatives continues

- **Diagramme différentiel** : histogramme des densités.

Sur l'axe des abscisses sont placées les bornes des classes représentant les modalités en respectant une échelle. Pour chaque classe, on élève un rectangle de hauteur (ordonnée) proportionnelle à la densité de proportion ou d'effectif.

Exemple de taille T :

Remarque 3.1. On représente la **densité de proportion ou d'effectif** et non pas les fréquences relatives ou absolues.

Consequence 1. L'aire d'un rectangle est proportionnelle à la fréquence (absolues ou relatives) de la classe correspondante. En effet, pour le rectangle correspondant à la classe $[b_j, b_{j+1}[$ l'aire est

$$(b_j - b_{j-1}) \times d_j = f_j.$$

Dans la pratique, on utilise la règle de construction suivante :

Vérifier si les amplitudes des différentes classes sont identiques.

- Si les amplitudes sont identiques, on représente sur l'axe des abscisse les classes par des segments de même longueur. On associe à chaque classe un rectangle dont la hauteur est proportionnelle à l'effectif ou à la fréquence.

- Si les amplitudes sont non identiques, on choisit une unité d'amplitude U et on construit l'histogramme de telle sorte que la hauteur du rectangle de la classe $[b_{j-1}, b_j[$ soit proportionnelle à l'effectif par unité d'amplitude $\frac{n_i}{a_i}U$ associé.

- **Diagramme intégral** : courbe cumulative des effectifs ou des fréquences.

La courbe cumulative des fréquences doit représenter la fonction de répartition de la variable statistique.

Fonction de répartition empirique

Pour x une valeur dans l'intervalle $[b_{j-1}, b_j[$, on approche la proportion d'individus pour lesquels la variable est inférieure ou égale à x par l'aire de l'histogramme entre les abscisses b_{j-1} et x notée $F(x)$:

$$F(x) = f_1 + f_2 + \dots + f_{j-1} + (x - b_{j-1}) \times d_j = \Phi_{j-1} + (x - b_{j-1}) \times d_j$$

On a ainsi définie une fonction Φ qui vaut 0 sur $] -\infty, b_0[$ et 1 sur $[b_1, +\infty[$. Elle vaut Φ_j en b_j . Sur $[b_{j-1}, b_j[$, cette fonction, affine par morceaux, est appelée **fonction de répartition empirique**.

Exemple 12. Fonction de répartition empirique de la variable T .

Chapitre 4

Résumés numériques d'une variable statistique

1. Paramètre de tendance centrale

1.1. Le mode

Le mode rend compte de l'endroit où les données sont le plus concentrées.

Le mode, noté Mo , est la modalité la plus fréquente ou dominante dans la population i.e. celle qui admet la plus grande fréquence : $f(Mo) = \max_{i \in [1, k]} (f_i)$.

Il est parfaitement défini pour une variable qualitative ou une variable quantitative discrète.

Pour une variable quantitative continue regroupée en classe, nous parlons de classe modale : c'est la classe dont la densité de fréquence est maximum.

Si les classes ont même amplitude la densité est remplacée par l'effectif ou la fréquence et nous retrouvons la définition précédente.

Nous définissons le mode, pour une variable quantitative continue, en tenant compte des densités de fréquence des 2 classes adjacentes par la méthode suivante :

$$Mo = x_m + a \times \frac{\Delta i}{\Delta i + \Delta s}$$

avec

x_m : limite inférieure de la classe d'effectif (par unité d'amplitude) maximal

a : l'amplitude de la classe modale

Δi : Ecart d'effectif (par unité d'amplitude) entre la classe modale et la classe inférieure la plus proche

Δs : Ecart d'effectif (par unité d'amplitude) entre la classe modale et la classe supérieure la plus proche

Exemple 13. *Pointure, taille.*

Remarque :

Lorsque les classes adjacentes à la classe modale ont des densités de fréquences égales, le mode coïncide avec le centre de la classe modale.

Le mode dépend beaucoup de la répartition en classes.

Une variable statistique peut présenter plusieurs modes locaux : on dit alors qu'elle est plurimodale. Cette situation est intéressante : elle met en évidence l'existence de plusieurs sous-populations, donc l'hétérogénéité de la population étudiée.

1.2. La moyenne

On note $\{x_1, x_2, \dots, x_n\}$ la série statistique. La moyenne est définie par :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

Exemple 14. *pointure, taille*

Cas d'une variable discrète : si $v_1, v_2, \dots, v_k$ sont les k valeurs prises par la variable X , n_j l'effectif et f_j la fréquence relative correspondant à la valeur v_j , on peut réécrire :

$$\bar{x} = \frac{n_1 v_1 + n_2 v_2 + \dots + n_k v_k}{n} = \frac{1}{n} \sum_{i=1}^k n_i v_i = \sum_{i=1}^k f_i v_i$$

Exemple 15. *Pointure.*

Cas d'une variable continue regroupée en classes : la variable X est regroupée dans les classes $[b_{j-1}, b_j[$ ($1 \leq j \leq n$), les fréquences relatives associées à ces classes sont notées f_j , $1 \leq j \leq n$. Lorsque les données brutes ne sont plus accessibles et qu'on ne dispose que des données regroupées en classe, on calcule une **moyenne approchée** grâce à des représentants des classes (leur centre) : $c_j = (b_j + b_{j-1})/2$, par la formule :

$$\bar{x}_{app} = f_1 c_1 + f_2 c_2 + \dots + f_k c_k = \sum_{i=1}^k f_i c_i$$

Exemple : calcul d'une moyenne approchée de la variable "taille" à partir du groupement en classes.

Propriétés de la moyenne : si on fait le changement de variable $Y = aX + b$ (traduction sur la série statistiques : $y_i = ax_i + b, 1 \leq i \leq n$), alors

$$\bar{y} = a\bar{x} + b$$

Exemple 16. *calcul de la taille moyenne en mètres.*

1.3. La médiane

La médiane Me correspond au centre des valeurs observées classées par ordre croissant

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)},$$

ou à la valeur pour laquelle 50% des valeurs observées sont supérieures et 50% sont inférieures.

a) Cas d'une variable discrète :

- si n est impair, la médiane est la $\frac{n+1}{2}$ -ième valeur observée : $Me = x_{(\frac{n+1}{2})}$.
- si n est pair, une médiane est une valeur quelconque entre la $\frac{n}{2}$ -ième valeur observée : $x_{(\frac{n}{2})}$ et la $\frac{n}{2} + 1$ -ième valeur observée : $x_{(\frac{n}{2}+1)}$. On parle donc d'intervalle médian. On peut prendre comme médiane $x_{(\frac{n}{2})}$ ou $x_{(\frac{n}{2}+1)}$. Mais il peut être commode de prendre le milieu : $Me = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}$.

La détermination de la médiane se fait donc à l'aide des effectifs cumulés, des proportions cumulées ou de la fonction de répartition empirique (graphiquement).

Exemple 17. *pointure*

b) Cas d'une variable continue.

La médiane est définie comme la solution de l'équation :

$$F(Me) = 0,5$$

où F est la fonction de répartition empirique de la variable. On sait que cette solution existe parce que F est continue, et $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow +\infty} F(x) = 1$. Si de plus F est strictement croissante, la solution Me est unique. la méthode pratique est la suivante :

1. S'il existe une borne de classe b_j telle que la proportion cumulée sur la classe $[b_{j-1}, b_j[$ est exactement 0,5, alors la **médiane** est ce b_j .
2. Sinon, alors il existe une classe $[b_{j-1}, b_j[$ telle que

$$F(b_{j-1}) < 0,5 < F(b_j).$$

Cette classe est la première sur laquelle la fréquence cumulée dépasse 0,5. Pour $x \in [b_{j-1}, b_j[$, $F(x) = \Phi_{j-1} + (x - b_{j-1}) \times d_j$. Mais en particulier :

$$F(Me) = \Phi_{j-1} + (Me - b_{j-1}) \times d_j = 0,5$$

d'où

$$Me = \frac{0,5 - \Phi_{j-1}}{d_j} + b_{j-1}$$

Ou encore, en terme de b_j et de F :

$$Me = \frac{0,5 - F(b_{j-1})}{F(b_j) - F(b_{j-1})} \times (b_j - b_{j-1}) + b_{j-1}$$

Cette méthode peut se traduire graphiquement en utilisant le graphe de la fonction de répartition empirique et le théorème de Thalès.

Exemple 18. *médiane de la variable "taille", regroupée en classes.*

Méthode graphique avec la fonction de répartition empirique

1.4. Quantiles

a) cas d'une variable continue

Soit X une variable quantitative continue, de fonction de répartition empirique F . On suppose qu'on dispose de la répartition en classe des observations.

Le Quantile d'ordre p de X est la solution notée q_p de :

$$F(q_p) = p.$$

Cela signifie qu'une proportion d'environ p des observations est inférieure à q_p et qu'une proportion d'environ $1 - p$ des données est supérieure à q_p .

Quantiles particuliers

- Quartiles : quantiles correspondant aux proportions multiples de 0,25 (un quart). On note Q_1 le premier quartile, qui correspond à $q_{0,25}$, Q_3 le troisième quartile, qui correspond à $q_{0,75}$. La médiane est le deuxième quartile $Q_2 = q_{0,5}$.
- Déciles : quantiles correspondant aux proportions multiples de 0,1 : $q_{0,1}$ (premier décile), $q_{0,2}$ (deuxième décile), etc.
- Percentiles ou centiles : quantiles correspondant aux proportions multiples de 0,01. Par exemple, le 65ème percentile est le quantile $q_{0,65}$.

Calcul du quantile q_p : même méthode que pour le calcul de la médiane.

1. S'il existe une borne de classe b_j telle que la proportion cumulée sur la classe $[b_{j-1}, b_j[$ est exactement p , autrement dit : $F(b_j) = p$, alors q_p .
2. Sinon, alors il existe une classe $[b_{j-1}, b_j[$ telle que

$$F(b_{j-1}) < p < F(b_j).$$

Cette classe est la première sur laquelle la fréquence cumulée dépasse p . Pour $x \in [b_{j-1}, b_j[$, $F(x) = \Phi_{j-1} + (x - b_{j-1}) \times d_j$. Mais en particulier :

$$F(q_p) = \Phi_{j-1} + (q_p - b_{j-1}) \times d_j = p$$

D'où

$$q_p = \frac{p - F(b_{j-1})}{F(b_j) - F(b_{j-1})} \times (b_j - b_{j-1}) + b_{j-1}$$

Ou encore, en terme des b_j et de F :

$$q_p = \frac{p - F(b_{j-1})}{F(b_j) - F(b_{j-1})} \times (b_j - b_{j-1}) + b_{j-1}$$

Exemple 19. *troisième quartile de la variable "taille"*

b) cas d'une variable discrète

Comme pour la médiane, il existe diverses manières de définir les quantiles d'une loi discrète :

comme la fonction de répartition empirique n'est pas continue mais a des paliers, elle ne prend pas toutes les valeurs entre 0 et 1. Pour une proportion p fixée, on cherche donc une valeur x telle que $F(x)$ s'approche, en un certain sens, de p . Nous choisissons la définition suivante :

$$q_p = \begin{cases} v_1 & \text{lorsque } 0 < p \leq \Phi_1 = f_1 \\ v_2 & \text{lorsque } \Phi_1 < p \leq \Phi_2 \\ \dots, \\ v_j & \text{lorsque } \Phi_{j-1} < p \leq \Phi_j \\ \dots, \\ v_k & \text{lorsque } p = \Phi_k = 1 \end{cases}$$

ou $q_p = x_{(\lceil np \rceil)}$ où $\lceil np \rceil$ est le plus petit entier tel que $np \leq \lceil np \rceil$.

Exemple 20. *troisième quartile de la variable "pointure".*

1.5. Utilisation des paramètres de tendance centrale

Robustesse

La médiane est plus **robuste** que la moyenne : une ou plusieurs données erronées ne font pratiquement, voire pas du tout, changer la médiane, alors qu'elles peuvent affecter considérablement la moyenne.

Assymétrie

La comparaison de la médiane et de la moyenne permet de détecter des assymétries de données :

Si la distribution des valeurs est symétrique, la valeur de la médiane est proche de la valeur de la moyenne arithmétique. $Me \simeq \bar{x}$.

De façon générale on a :

- $Mo = Me = \bar{x} \implies$ distribution symétrique,
- $Mo < Me < \bar{x} \implies$ distribution dissymétrique à gauche,
- $Mo > Me > \bar{x} \implies$ distribution dissymétrique à droite

2. Paramètres de dispersion

Il est possible que deux variables statistiques aient la même valeur centrale mais complètement différentes du point de vue de la concentration ou dispersion des valeurs observées autour de cette valeur centrale. Il est donc nécessaire de trouver des mesures permettant d'apprécier la dispersion d'une série statistique ou d'une distribution observée.

2.1. L'étendue

Soit x_{min} la plus petite observation et x_{max} la plus grande. On définit **l'étendue** $e = x_{max} - x_{min}$. Elle a la même unité que l'unité de la variable. Elle n'est pas très informative car elle ne tient pas

du tout compte de la répartition des données à l'intérieur de l'intervalle $[x_{min}, x_{max}]$.

Exemple 21. *étendu de la variable "taille"*

2.2. L'intervalle inter-quartile

On appelle **intervalle inter-quartile** l'intervalle $[Q_1, Q_3]$, qui contient environ 50% des observations. La **distance inter-quartile** $Q_3 - Q_1$ est une mesure de dispersion.

Exemple 22. *intervalle inter-quartile de la variable "taille".*

2.3. La variance et l'écart-type

La **variance** est définie par :

$$Var(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

L'expression suivante est la plus pratique pour le calcul de la variance :

$$Var(X) = \left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right) - (\bar{x})^2$$

Preuve : en développant le carré dans la définition de la variance.

Pour une variable **quantitative discrète** en prenant la valeur v_j un nombre n_j de fois ou (ou avec la fréquence f_j), pour $1 \leq j \leq k$:

$$\begin{aligned} Var(X) &= \frac{1}{n} \sum_{j=1}^k n_j (v_j - \bar{x})^2 = \sum_{j=1}^k f_j (v_j - \bar{x})^2 \\ &= \left(\frac{1}{n} \sum_{j=1}^k n_j v_j^2 \right) - (\bar{x})^2 = \left(\sum_{j=1}^k f_j v_j^2 \right) - (\bar{x})^2 \end{aligned}$$

Dans le cas d'une variable continue pour laquelle on dispose seulement des **données regroupées en classes**, on peut faire un calcul approché similaire à celui de la moyenne approchée $\bar{x}_{app}$. On calcule une valeur approchée de la variance, notée $Var_{app}(X)$. Toutes les expressions qui suivent sont équivalentes.

$$\begin{aligned} Var_{app}(X) &= \frac{1}{n} \sum_{j=1}^k n_j (c_j - \bar{x}_{app})^2 = \sum_{j=1}^k f_j (c_j - \bar{x}_{app})^2 \\ &= \left(\frac{1}{n} \sum_{j=1}^k n_j c_j^2 \right) - (\bar{x}_{app})^2 = \left(\sum_{j=1}^k f_j c_j^2 \right) - (\bar{x}_{app})^2 \end{aligned}$$

où c_j est le centre de la j -ème classe, dotée de l'effectif n_j (ou de la fréquence relative f_j).

Propriétés de la variance

- La variance est toujours positive ou nulle. Elle est nulle si et seulement si toutes les observations sont identiques :

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \Leftrightarrow \forall i, x_i - \bar{x} = 0$$

- L'unité de la variance est l'unité de X au carré.

L'écart-type σ_X est défini par :

$$\sigma_X = \sqrt{\text{Var}(X)}$$

Propriété : l'unité de σ_X est l'unité de X .

Plus σ_X est grand plus les modalités sont dispersées.

Exemple 23. *variance et écart-type de la variable "pointure", de la variable "taille".*

Le coefficient de variation

La comparaison des dispersions de deux séries statistiques peut se faire grâce aux écart-types lorsque ses séries ont des moyennes du même ordre de grandeur et ne contiennent pas de valeurs aberrantes. Dans le cas contraire, on peut utiliser le coefficient de variation défini par

$$CV = \frac{\sigma_X}{\bar{x}}.$$

Ce paramètre est une mesure relative de dispersion et permet une interprétation plus appropriée. On l'exprime en général en pourcentage.

3. Changement de variable linéaire ou affine - Variable centrée réduite

3.1. Changement de variable linéaire ou affine

On considère une variable quantitative X et on lui fait subir une application affine qui la transforme en une variable Y . a et b sont des constantes réelles

Nouvelle variable Y	Observations y_i	Moyenne de Y	Variance de Y	Ecart-type de Y
$Y = aX$	$y_i = ax_i$	$\bar{y} = a\bar{x}$	$\text{Var}(Y) = a^2\text{Var}(X)$	$\sigma_Y = a \sigma_X$
$Y = X + b$	$y_i = x_i + b$	$\bar{y} = \bar{x} + b$	$\text{Var}(Y) = \text{Var}(X)$	$\sigma_Y = \sigma_X$
$Y = aX + b$	$y_i = ax_i + b$	$\bar{y} = a\bar{x} + b$	$\text{Var}(Y) = a^2\text{Var}(X)$	$\sigma_Y = a \sigma_X$

Exemple 24.

3.2. Variable centrée réduite

On considère une variable X de moyenne $\bar{x}$ et de variance $Var(x)$, d'écart-type $\sigma_X = \sqrt{Var(X)}$. On définit une nouvelle variable

$$Y = \frac{X - \bar{x}}{\sigma_X}$$

Elle est sans unité. Cette variable est appelée **variable centrée réduite associée** à X . En effet, elle est :

- **centrée** : $\bar{y} = \frac{\bar{x} - \bar{x}}{\sigma_X} = 0$.
- **réduite** : $Var(Y) = \frac{Var(X)}{Var(X)} = 1$.

Quand on transforme une variable en la variable centrée réduite associée, on retire à cette variable toute l'information concernant son échelle ou unité, et sa *localisation*. Il ne reste plus que des informations sur la **forme** de la distribution. Cette transformation permet de comparer plusieurs variables sur le plan de la forme, même si ce sont des variables exprimées dans des échelles différentes ou qui ont des moyennes complètement différentes.

Exemple 25. Variable centrée réduite associée à la variable "pointure", à la variable "taille".

Autre utilisation : Etant donné un individu i pour lequel la variable prend la valeur x_i , on peut situer cet individu dans l'ensemble des observations en calculant son écart à la moyenne réduit :

$$\frac{x_i - \bar{x}}{\sigma_X}$$

Exemple 26. quel est l'écart à la moyenne, mesuré en écart-types, d'un individu mesurant 177 cm ?

4. Boîtes à moustaches

La boîte à moustaches est une représentation graphique qui permet de visualiser les quartiles ainsi que la dispersion des données et de repérer les données extrêmes ou *outliers*. Elle se fait couramment pour les variables quantitatives continues ou pour les variables quantitatives discrètes prenant un grand nombre de valeurs différentes. En revanche, elle n'a pas beaucoup d'intérêt pour une variable discrète prenant peu de valeurs différentes.

Elle est constituée :

- d'une **boîte** dont les bornes sont les premier et troisième quartile Q_1 et Q_3 . A l'intérieur de la boîte figure la médiane Q_2 .
- de **moustaches**. On définit tout d'abord deux bornes : $m_- = Q_1 - 1,5(Q_3 - Q_1)$ et $m_+ = Q_3 + 1,5(Q_3 - Q_1)$. On note m_{inf} la plus petite observation supérieure à m_- , et m_{sup} la plus grande observation inférieure à m_+ . Soit :

$$m_{inf} = \min\{x_i : x_i \geq m_-\}$$

$$m_{sup} = \max\{x_i : x_i \leq m_+\}$$

La moustache inférieure est le segment $[m_{inf}, Q_1]$. La moustache supérieure, de la même manière, est le segment $[Q_3, m_{sup}]$

- des **données extrêmes** éventuelles : les observations qui sont en dehors de la boîte et des moustaches, c'est à dire : supérieures à m_+ ou inférieures à m_- . On place ces données une à une quand on en dispose.



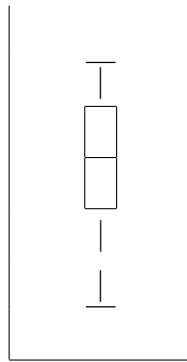
Remarque :

- Une boîte et des moustaches courtes indiquent que la série est assez concentrée autour de sa médiane.

Au contraire une boîte et des moustaches longues indiquent que la série est assez dispersée.

L'examen de la boîte à moustaches permet d'avoir une idée de la symétrie de la distribution selon que la boîte et les moustaches sont symétriques ou, au contraire, de plus petite amplitude à gauche (asymétrie à gauche) ou à droite (asymétrie à droite).

- La représentation peut aussi se faire verticalement, d'où l'appellation de "boîte à pattes".



Exemple 27. Boîte à moustache de la variable "taille" à partir de la série statistique de 20 observations.

Dans le cas où on ne dispose pas des données brutes mais seulement des données regroupées en classes, on utilise les extrémités b_0 et b_k de la première et de la k -ème classe.

- la limite inférieure m_{inf} de la moustache inférieure est $\max\{m_-, b_0\}$ et la limite supérieure m_{sup} de la moustache supérieure est $\min\{m_+, b_k\}$.
- On ne peut pas placer les données extrêmes, sauf si elles sont fournies en plus.

Exemple 28. Boîte à moustaches de la variable "taille" à partir des données regroupées.

Chapitre 5

Liaison entre deux variables

1. Généralités

On observe une série statistique $\{(x_1, y_1), \dots, (x_n, y_n)\}$ composée de n couples d'observations d'un couple de variables (X, Y) . On suppose que X a I modalités notées $C_1, \dots, C_I$ et Y a J modalités notées $D_1, \dots, D_J$. Pour $1 \leq i \leq I$ et $1 \leq j \leq J$, on note n_{ij} l'effectif des couples d'observations égaux à (C_i, D_j) .

1.1. Tableau de contingence

Dans le tableau de contingence, on regroupe les effectifs n_{ij} . On peut compléter le tableau de contingence en ajoutant les totaux en lignes et en colonnes.

On note $n_{i.} = n_{i1} + \dots + n_{iJ} = \sum_{j=1}^J n_{ij}$ le total sur la ligne i de la table de contingence,

$n_{.j} = n_{1j} + \dots + n_{IJ} = \sum_{i=1}^I n_{ij}$ le total sur la colonne j de la table de contingence.

Y	D_1	D_2	$\dots$	D_J	Total
X					
C_1	n_{11}	n_{12}	$\dots$	n_{1J}	$n_{1.}$
C_2	n_{21}	n_{22}	$\dots$	n_{2J}	$n_{2.}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
C_I	n_{I1}	n_{I2}	$\dots$	n_{IJ}	$n_{I.}$
Total	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.J}$	n

Exemple 29. L'INSEE fournit les données suivantes relatives à la situation professionnelle des personnes habitant en France en 2006, immigrées ou non immigrées.

<i>situation quant à l'immigration</i>	<i>Immigrés</i>	<i>Non immigrés</i>	<i>Ensemble</i>
<i>Situation professionnelle</i>			
<i>Actif ayant un emploi</i>	2223906	23895180	26119096
<i>Chômeur</i>	559201	2845339	3404540
<i>Retraité ou préretraités</i>	963333	11901857	12865190
<i>Elèves, étudiants, stagiaire</i>	321533	4999097	5320630
<i>Femme ou homme au foyer</i>	486427	1926779	2413206
<i>Autres inactifs</i>	583016	12480429	13063445
<i>Ensemble</i>	5137416	58048681	63186098

Remarque 5.1. La définition d'un immigré selon le Haut conseil à l'immigration, utilisée pour cette étude, est une personne née étrangère à l'étranger et résidant en France.

1.2. Distribution marginale

La distribution marginale de la variable X est la donnée des **effectifs marginaux** $n_{1.}, \dots, n_{I.}$. C'est la distribution de la variable X . On peut la présenter dans un tableau et calculer les fréquences ($f_{i.} = n_{i.}/n$), qui sont les proportions associée à chaque modalité de la variable X . On peut calculer de même la distribution marginale de la variable Y .

Distribution marginale de X :

X	C_1	$\dots$	C_I	Total
Effectif	$n_{1.}$	$\dots$	$n_{I.}$	n
Proportion	$f_{1.} = n_{1.}/n$	$\dots$	$f_{I.} = n_{I.}/n$	1

Distribution marginale de Y :

Y	D_1	$\dots$	D_I	Total
Effectif	$n_{.1}$	$\dots$	$n_{.J}$	n
Proportion	$f_{.1} = n_{.1}/n$	$\dots$	$f_{.J} = n_{.J}/n$	1

Exemple 30. Situation professionnelle de la population en France en 2006

1.3. Distribution conditionnelle

a) Profils-lignes

La distribution conditionnelle de Y sachant la modalité de C_i de X est la distribution dont les proportions sont données dans le tableaux suivant :

$Y _{X=C_i}$	D_1	$\dots$	D_I	Total
Proportion	n_{i1}/n_i	$\dots$	n_{iJ}/n_i	1

Une telle distribution est appelée profil-ligne. L'ensemble des profils-lignes peut être présenté dans un tableau :

Y_X	D_1	D_2	$\dots$	D_J	Total
X					
C_1	$n_{11}/n_{1.}$	$n_{12}/n_{1.}$	$\dots$	$n_{1J}/n_{1.}$	1
C_1	$n_{21}/n_{2.}$	$n_{22}/n_{2.}$	$\dots$	$n_{2J}/n_{2.}$	1
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	
C_I	$n_{I1}/n_{I.}$	$n_{I2}/n_{I.}$	$\dots$	$n_{IJ}/n_{I.}$	1

Exemple 31. *Distribution conditionnelle de la variable " Situation quant à l'immigration " sachant la modalité " Actifs ayant un emploi " en France en 2006, ou : situation quant à l'immigration des actifs ayant un emploi en France en 2006.*

b) Profils-colones

De même, l'ensemble des distributions conditionnelles de X sachant les modalités de Y est l'ensemble des profils-colones, que l'on peut présenter dans le tableau suivant :

Y	D_1	D_2	$\dots$	D_J
$X_{ Y}$				
C_1	$n_{11}/n_{.1}$	$n_{12}/n_{.2}$	$\dots$	$n_{1J}/n_{.J}$
C_1	$n_{21}/n_{.1}$	$n_{22}/n_{.2}$	$\dots$	$n_{2J}/n_{.J}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
C_I	$n_{I1}/n_{.1}$	$n_{I2}/n_{.2}$	$\dots$	$n_{IJ}/n_{.J}$
Total	1	1	1	1

Exemple 32. *Ensemble des profils-colones du couple de variables "Situation professionnelle" et "Situation vis-à-vis de l'immigration".*

2. Laison linéaire entre deux variables quantitatives

Les valeurs distinctes de X et Y sont représentées respectivement par $x_1, \dots, x_p$ et $y_1, \dots, y_q$

On désigne par n_{ij} l'effectif associé à l'observation (x_i, y_j) et par f_{ij} la fréquence associée :

	y_1	$\dots$	y_j	$\dots$	y_q	Total
x_1	n_{11}	$\dots$	n_{1j}	$\dots$	n_{1q}	$n_{1.}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$		$\vdots$	$\vdots$
x_i	n_{i1}	$\dots$	n_{ij}	$\dots$	n_{iq}	$n_{i.}$
$\vdots$	$\vdots$		$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_p	n_{p1}	$\dots$	n_{pj}	$\dots$	n_{pq}	$n_{p.}$
Total	$n_{.1}$	$\dots$	$n_{.j}$	$\dots$	$n_{.q}$	n

Avec

$$n_{i.} = \sum_{j=1}^q n_{ij}; \quad n_{.j} = \sum_{i=1}^p n_{ij}; \quad \sum_{i=1}^p \sum_{j=1}^q n_{ij} = \sum_{i=1}^p n_{i.} = \sum_{j=1}^q n_{.j} = n$$

2.1. Moyennes et variances marginales

$$X : \quad \bar{x} = \frac{1}{n} \sum_{i=1}^p n_{i.} x_i, \quad s^2(X) = \frac{1}{n} \sum_{i=1}^p n_{i.} (x_i - \bar{x})^2$$

$$Y : \quad \bar{y} = \frac{1}{n} \sum_{j=1}^q n_{.j} y_j, \quad s^2(Y) = \frac{1}{n} \sum_{j=1}^q n_{.j} (y_j - \bar{y})^2$$

2.2. Moyennes et variances conditionnelles

$$X|Y = y_j : \quad \bar{x}_j = \frac{1}{n_{.j}} \sum_{i=1}^p n_{ij} x_i, \quad s_j^2(X) = \frac{1}{n_{.j}} \sum_{i=1}^p n_{ij} (x_i - \bar{x}_j)^2.$$

$$Y|X = x_i : \quad \bar{y}_i = \frac{1}{n_{i.}} \sum_{j=1}^q n_{ij} y_j, \quad s_i^2(Y) = \frac{1}{n_{i.}} \sum_{j=1}^q n_{ij} (y_j - \bar{y}_i)^2.$$

2.3. Covariance

Définition 5.1. On définit la **covariance** de X et de Y par :

$$\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^q n_{ij} [(x_i - \bar{x})(y_j - \bar{y})].$$

L'unité dans est exprimée la covariance est le produit des unités de X et de Y .

Remarque 5.2. Lien avec la variance : $\text{Cov}(X, X) = \text{Var}(X)$

Remarque 5.3. Formule pratique :

$$\text{Cov}(X, Y) = \left(\frac{1}{n} \sum_{i=1}^p \sum_{j=1}^q n_{ij} x_i y_j \right) - \bar{x} \bar{y}.$$

Propriété 5.1. Changement d'échelle : soient a, b, c, d des constantes réelles. On a

$$\text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y).$$

Proposition 5.1. Expression de la variance d'une somme de variables :

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).$$

Proposition 5.2. *Inégalité de Cauchy-Schwarz :*

$$\|Cov(X, Y)\| \leq \sigma_X \sigma_Y.$$

Preuve : Pour tout réelle a , on peut développer grâce à la proposition 1 la quantité $Var(X + aY) \geq 0$:

$$\begin{aligned} Var(X + aY) &= Var(X) + Var(aY) + 2Cov(X, aY) \\ &= Var(X) + a^2 Var(Y) + 2aCov(X, Y) \quad \text{par la propriété 1} \\ &\geq 0 \end{aligned} \quad (5.1)$$

Le polynôme du second degré en a étant de signe constant, son discriminant est négatif ou nul :

$$4(Cov(X, Y))^2 - 4Var(X)Var(Y) \leq 0,$$

d'où l'égalité recherchée.

Remarquons au passage que le cas d'égalité se produit lorsque le discriminant de l'équation 5.1 est nul. Dans ce cas, l'équation admet une racine double :

$$\begin{aligned} a &= -\frac{2Cov(X, Y)}{2Var(Y)} = -\frac{Cov(X, Y)}{Var(Y)} \\ &= \begin{cases} -\frac{\sigma_X}{\sigma_Y} & \text{si } Cov(X, Y) = +\sigma_X \sigma_Y \\ \frac{\sigma_X}{\sigma_Y} & \text{si } Cov(X, Y) = -\sigma_X \sigma_Y \end{cases} \end{aligned}$$

Dans le premier cas, cela signifie que $X - \frac{\sigma_X}{\sigma_Y}Y$ a une variance nulle, donc est une constante, d'où

$$X = \frac{\sigma_X}{\sigma_Y} + \text{constante}.$$

Dans le second cas,

$$X = \frac{\sigma_X}{\sigma_Y}Y + \text{constante}.$$

Ces deux cas sont les seuls cas d'égalité dans la proposition 2. Ils correspondent au fait que les variables X et Y s'obtiennent l'une à partir de l'autre par une application affine.

2.4. Coefficient de corrélation

Définition 5.2. *Le coefficient de corrélation $r(X, Y)$ est défini par :*

$$r(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}.$$

C'est un coefficient sans unité. Sa valeur absolue est invariante par translation et changement d'échelle des variables : pour toutes constantes réelles $a \neq 0$, b , $c \neq 0$, d ,

$$r(aX + b, cY + d) = \frac{ac}{|ac|} r(X, Y).$$

Propriété 5.2. *il découle de la proposition 2 que*

$$-1 \leq r(X, Y) \leq 1.$$

De plus, les cas de l'égalité sont les suivantes :

$r(X, Y) = 1$ si et seulement si les deux variables satisfont une relation affine du type $Y = aX + b$ avec $a > 0$.

$r(X, Y) = -1$ si et seulement si les deux variables satisfont une relation affine du type $Y = aX + b$ avec $a < 0$.

Lorsque le nuage des points (x_i, y_i) est exactement situé sur une droite (cas idéal), on est dans la situation où $r(X, Y) = \pm 1$. Lorsque $r(X, Y)$ est proche de ± 1 (pour fixer les idées : $|r(X, Y)| \geq 0,8$), alors il y'a une liaison linéaire importante entre X et Y . Lorsqu'au contraire $r(X, Y)$ est proche de 0, alors il n'existe pas de relation linéaire entre X et Y . Attention, il peut y avoir quand même un autre type de liaison entre X et Y .

2.5. Régression linéaire

On suppose à présent que les observations du couple de variable (X, Y) satisfont une relation de la forme suivante,

$$y_i = ax_i + b + \epsilon_i, \quad i = 1, \dots, n, \quad (5.2)$$

où a et b sont des coefficients réels. Le terme ϵ_i désigne un *bruit*, c'est à dire une perturbation supposée *petite*. Dans ce cours, on ne cherchera pas à donner un sens précis à la mesure de ce bruit.

Disposant des observations $(x_i, y_i)_{i=1}^n$ du couple (X, Y) , on cherche à trouver les coefficients a et b qui permettent le mieux d'ajuster les données à une relation du type (5.2), au sens du critère des moindres carrés. On cherche

$$\min_{a,b} \sum_{i=1}^n (y_i - b - ax_i)^2. \quad (5.3)$$

La solution, qui s'obtient en annulant les dérivées partielles de la fonction de (a, b) qui est minimisée en (5.3), est

$$\begin{aligned} \hat{a} &= \frac{Cov(X, Y)}{Var(X)}, \\ \hat{b} &= \bar{y} - \hat{a}\bar{x}, \end{aligned}$$

où $\bar{x}$ et $\bar{y}$ désigne les moyennes respectives de X et Y . La droite des moindres carrés est la droite d'équation : $y = \hat{a}x + \hat{b}$. On peut remarquer qu'elle passe toujours par le barycentre $(\bar{x}, \bar{y})$ du nuage de points. Sa pente peut aussi s'écrire à l'aide du coefficient de corrélation : $\hat{a} = r(X, Y) \frac{\sigma_X}{\sigma_Y}$.

Prediction

Pour une valeur x_0 de la variable X qui ne fait pas partie des observations, on peut faire une

prédiction de la valeur correspondante de Y en calculant l'ordonnée du point d'abscisse x_0 sur la droite des moindres carrés :

$$y_0 = \hat{a}x_0 + \hat{b}$$

2.6. Régression linéaire après transformation d'une variable

On suppose que les observations $(x_i, y_i)_{i=1}^n$ satisfont une relation de type

$$y_i = af(x_i) + b + \epsilon_i,$$

Pour une certaine fonction f donnée et de bruit ϵ_i . On peut estimer les coefficients de la droite de régression de Y sur $f(X)$ par la méthode décrite auparavant.

3. Liaison entre deux variables qualitatives

3.1. Mesure de la liaison entre deux variables qualitatives

Comparaison qualitative des profils-lignes ou des profils-colones

Il y'a indépendance stricte entre X et Y lorsque tous les profils-lignes sont identiques. Il sont dans ce cas tous identiques à la distribution marginal de Y .

De la même manière, l'indépendance a lieu lorsque tous les profils-colonnes sont égaux à la distribution marginale de X .

Ceci implique : pour tous i, j ,

$$n_{ij} = \frac{n_{i.}n_{.j}}{n}. \quad (5.4)$$

Réciproquement, si (5.4) a lieu, alors il y a indépendance entre X et Y .

preuve :

La distance du χ^2 pour mesurer l'écart à l'indépendance

Dans la pratique, cette indépendance stricte ne s'observe jamais sur un échantillons. On peut être plus ou moins éloigné de cette situation parfaite. La distance du χ^2 d'écart à l'indépendance permet de mesurer le degré de dépendance entre X et Y . Elle se base sur la comparaison entre n_{ij} et $\frac{n_{i.}n_{.j}}{n}$.

Définition 5.3. La **distance du χ^2** observée sur la série statistique $\{(x_1, y_1), \dots, (x_n, y_n)\}$ est définie par

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \left(\frac{\left(n_{ij} - \frac{n_{i.}n_{.j}}{n} \right)^2}{\frac{n_{i.}n_{.j}}{n}} \right)$$

Exemple 33. Distance du χ^2 pour mesurer l'écart à l'indépendance entre les variables "situation quant à l'immigration" et "situation professionnelle" en France 2006.

Propriété 5.3. – la grandeur $\chi^2 = 0$ si il y a indépendance stricte entre X et Y .

- la grandeur χ^2 est d'autant plus élevée que la liaison est forte : il existe alors des cellules (i, j) avec une écart important $n_{ij} - \frac{n_{i.}n_{.j}}{n}$.
- l'inégalité suivante est toujours vérifiée :

$$\frac{\chi^2}{n} \leq \min\{I - 1, J - 1\}.$$

Définition 5.4. On appelle **contribution au χ^2** du couple de modalités (C_i, D_j) et (X, Y) la quantité $\frac{(n_{ij} - \frac{n_{i.}n_{.j}}{n})^2}{\frac{n_{i.}n_{.j}}{n}}$.

Plus la contribution est forte, plus la liaison entre les modalités C_i et D_j est importante.

Définition 5.5. L'association entre les modalités C_i et D_j est dite **positive** si $n_{ij} - \frac{n_{i.}n_{.j}}{n} > 0$. Elle est **négative** si $n_{ij} - \frac{n_{i.}n_{.j}}{n} < 0$.

Exemple 34. Liaison entre la modalité "Elèves, étudiants, stagiaires" de la variable "Situation professionnelle" et la modalité "Immigrés" de la variable "Situation quant à l'immigration".

Définition 5.6. Le coefficient C de Cramer est défini par :

$$C = \sqrt{\frac{\chi^2}{n \cdot \min\{I - 1, J - 1\}}}.$$

Propriété 5.4. – $0 \leq C \leq 1$

- $C = 0$ lorsqu'il y a indépendance. De petites valeurs de C signient que la liaison entre X et Y est très faible. Des valeurs proches de 1 signifient qu'il y a une forte liaison entre X et Y .
- Ce coefficient, qui varie entre 0 et 1, permet de comparer la liaison entre plusieurs couples de variables.

Exemple 35. Calcul du C de Cramer pour mesurer l'écart à l'indépendance entre les variables "Situation quant à l'immigration" et "Situation professionnelle" en France en 2006.

3.2. Représentation graphique

a) Distribution jointe

Exemple 36. Diagramme en barres de la distribution jointe des variables "Situation quant à l'immigration" et "Situation professionnelle".

b) Distribution conditionnelle

Exemple 37. Diagramme en barres de la distribution de la variable "Situation professionnelle" sachant la variable "Situation quant à l'immigration".

4. liaison entre variable qualitative et une variable quantitative

On observe des couples $\{(x_i, y_i), 1 \leq i \leq n\}$ d'observations du couple de variable (X, Y) avec :

- X qualitative à I modalités : $C_1, \dots, C_I$
- Y quantitative, discrète ou continue, avec donnée brutes ou regroupées en classes.

Exemple 38.

4.1. Classement des données et distributions marginales

La distribution marginale de X est la distribution associée à la série statistique $(x_1, \dots, x_n)$ (variable qualitative). La distribution marginale de Y est la distribution associée à la série statistique $(y_1, \dots, y_n)$ (variable quantitative). On note $\bar{y}$ la moyenne marginale de la variable Y et de σ_Y^2 sa variance marginale.

On note $n_1, \dots, n_I$ les effectifs marginaux de la variable X . C'est-à-dire : n_1 est l'effectif des observations pour lesquelles X prend la modalité C_1 , etc...on peut regrouper les couples d'observations (x_i, y_i) qui comportent la même modalité x_i . Après regroupement, on obtient la nouvelle énumération :

$$\begin{aligned} (x_{11}, y_{11}), (x_{12}, y_{12}), \dots, (x_{1n_1}, y_{1n_1}) &= (C_1, y_{11}), (C_1, y_{12}), \dots, (C_1, y_{1n_1}) \\ (x_{21}, y_{21}), (x_{22}, y_{22}), \dots, (x_{2n_2}, y_{2n_2}) &= (C_2, y_{21}), (C_2, y_{22}), \dots, (C_2, y_{2n_2}) \\ &\dots \\ (x_{I1}, y_{I1}), (x_{I2}, y_{I2}), \dots, (x_{In_I}, y_{In_I}) &= (C_I, y_{I1}), (C_I, y_{I2}), \dots, (C_I, y_{In_I}) \end{aligned}$$

PARTIE II : DENOMBREMENT ET PROBABILITÉ
--

Chapitre 6

Denombrement

Le dénombrement consiste à déterminer le nombre d'éléments d'un ensemble fini. Ce chapitre fournit des méthodes de dénombrement particulièrement utiles en probabilités.

1. Cardinal d'un ensemble fini

Définition 6.1. *Un ensemble E non vide est dit fini s'il existe un entier n et une bijection de $\{1, 2, \dots, n\}$ sur E . Lorsqu'il existe, l'entier n est unique et est noté $\text{Card}(E)$. C'est le cardinal ou le nombre d'éléments de E*

Définition 6.2. *Un ensemble E est dit dénombrable s'il existe une bijection de $\mathbb{N}$ sur E . Un ensemble E est dit infini non dénombrable s'il n'est ni fini, ni dénombrable.*

Soit E un ensemble fini et A, B deux parties de E .

Proposition 6.1. 1. Si $\bar{A}$ est le complémentaire de A dans E alors

$$\text{Card}(\bar{A}) = \text{Card}(E) - \text{Card}(A).$$

2. $\text{Card}(A \cup B) = \text{Card}(A) + \text{Card}(B) - \text{Card}(A \cap B).$

3. Si $A \cap B = \emptyset$ alors $\text{Card}(A \cup B) = \text{Card}(A) + \text{Card}(B)$

4. $\text{Card}(A \times B) = \text{Card}(A) \times \text{Card}(B)$

2. Principes de comptage

2.1. Principe additif

Soit E un ensemble fini et $A_1, A_2, \dots, A_n$ des parties de E constituant une partition de E , c'est dire,

- $A_i \cap A_j = \emptyset$ pour $i \neq j$

- $E = A_1 \cup A_2 \cup \dots \cup A_n$.

Alors nous avons $Card(E) = \sum_{i=1}^n Card(A_i)$.

Lorsqu'on veut dénombrer un ensemble fini E , on peut trouver une partition $A_1, A_2, \dots, A_n$ de cet ensemble, où les cardinaux des ensembles A_i sont plus faciles déterminer. Il ne reste alors qu'à faire la somme des différents cardinaux obtenus.

Exemple 39. *J'ai dans ma bibliothèque 50 livres de mathématiques en français et 40 livres de mathématiques en anglais (et aucun dans une autre langue). Je peux donc y choisir un livre de mathématiques de $50 + 40 = 90$ façons différentes.*

2.2. Principe multiplicatif

Si une situation correspond p choix successifs ayant chacun respectivement $n_1, n_2, \dots, n_p$ possibilités alors le nombre total de possibilités est

$$n_1 \times n_2 \times \dots \times n_p.$$

3. Arrangements

3.1. Arrangements avec répétition

Soit $p \in \mathbb{N}^*$ et E un ensemble fini n éléments.

Définition 6.3. *Un arrangement avec répétition de p éléments (ou p -liste) de E est une partie ordonnée de p éléments de E non nécessairement distincts. Cela revient à prendre p objets dans E en tenant compte de l'ordre dans lequel on les choisit, et en pouvant prendre plusieurs fois le même.*

Proposition 6.2. *Le nombre d'arrangements avec répétition de p objets parmi n est n^p .*

En effet, on a n possibilités pour chaque place, soit $n \times n \times \dots \times n = n^p$ possibilités d'arrangement d'après le principe multiplicatif.

Exemple 40. *Combien y a-t-il de numéros de téléphone commençant par 08 ?*

Un numéro de téléphone est constitué de 8 chiffres. Les 6 numéros qui suivent le "08" sont des arrangements avec répétitions de 6 éléments de l'ensemble

$$\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, \}.$$

Il y en a $10^6 = 1000000$ possibilités.

Exemple 41. *Tirer successivement p boules, en les remettant chaque fois dans l'urne, et en tenant compte de l'ordre de sortie des numéros constitue un arrangement avec répétition de p éléments parmi n . Il y a n^p possibilités.*

3.2. Arrangements sans répétition

Soit $p \in \mathbb{N}^*$ et E un ensemble fini n éléments.

Définition 6.4. *Un arrangement de p éléments de E est une partie ordonnée de p éléments (distincts) de E . Cela revient prendre p objets distincts dans E en tenant compte de l'ordre dans lequel on les choisit.*

Proposition 6.3. *Le nombre d'arrangements de p objets parmi n est*

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

Nous avons n possibilités pour la première place, $n-1$ possibilités pour la deuxième place, $n-2$ possibilités pour la troisième place, ..., $(n-(p-1))$ possibilités pour la dernière place. D'après le principe multiplicatif, le nombre total de possibilités est :

$$\begin{aligned} A_n^p &= n \times (n-1) \times (n-2) \times \dots \times (n-(p-1)) \\ &= \frac{n!}{(n-p)!}. \end{aligned}$$

Exemple 42. Le tiercé. *Une course de chevaux comporte 20 partants. Combien peut-il y avoir de résultats possibles de tiercés dans l'ordre ?*

Soit E l'ensemble des numéros des chevaux. On a $\text{Card}(E) = 20$. Un tiercé correspond un arrangement de 3 éléments de E , il y en a $A_{20}^3 = 6840$ possibilités.

Exemple 43. *Tirer successivement p boules sans remise en tenant compte de l'ordre de sortie des numéros constitue un arrangement de p éléments parmi n . Il y a A_n^p possibilités.*

3.3. Permutation

Soit E un ensemble fini n éléments.

Définition 6.5. *Une permutation de E est un arrangement des n éléments de E . Cela revient prendre les n éléments de E en tenant compte de l'ordre dans lequel on les choisit.*

Proposition 6.4. *Le nombre de permutations d'un ensemble E n éléments est*

$$n! = n \times (n-1) \times \dots \times 2 \times 1.$$

Nous avons n possibilités pour la première place, $n-1$ possibilités pour la deuxième place, $n-2$ possibilités pour la troisième place, ..., 1 possibilités pour la dernière place. D'après le principe multiplicatif, le nombre total de possibilités est :

$$n! = n \times (n-1) \times (n-2) \times \dots \times 1 = A_n^n$$

Exemple 44. De combien de façons peut-on répartir 7 personnes sur 7 chaises ?

Désignons par $p_1, p_2, p_3, p_4, p_5, p_6, p_7$ les 7 personnes et posons

$$E = \{p_1, p_2, p_3, p_4, p_5, p_6, p_7\}.$$

Une répartition peut se voir comme une permutation de E , il y en a $7! = 5040$.

Exemple 45. Une urne contient n boules distinctes. Tirer successivement les n boules en tenant compte de l'ordre de sortie des boules constitue une permutation de n éléments. Il y a $n!$ possibilités.

4. Combinaisons

Définition 6.6. Une combinaison de p éléments de E est une partie non ordonnée de E formée de p éléments. Cela revient prendre p objets dans E sans tenir compte de l'ordre dans lequel on les choisit.

Proposition 6.5. Le nombre de combinaisons possibles de p objets pris parmi n est

$$C_n^p = \frac{n!}{p!(n-p)!} = \frac{A_n^p}{p!}.$$

Exemple 46. Quel est le nombre de comités de 3 personnes que l'on peut élire dans une assemblée de 20 personnes.

Le nombre de comités possibles est le nombre de combinaisons de 3 personnes parmi 20, soit $C_{20}^3 = 1140$

Exemple 47. Tirer simultanément p boules parmi n constitue une combinaison de p éléments parmi n éléments. Il y a C_n^p possibilités.

4.1. Binôme de Newton

Proposition 6.6. Soient a et b deux nombres reals et n un entier naturel non nul, alors :

$$(a+b)^n = \sum_{k=1}^n C_n^k a^k b^{n-k}.$$

5. Quel modèle choisir ?

- Si l'énoncé contient le mot **successif**, il faut tenir compte de tous les ordres dans lesquels on peut obtenir un événement donné. On doit souvent multiplier par le nombre d'ordres possibles le résultat trouvé pour un ordre déterminé.
- Si l'énoncé contient les mots "**successif et avec remise**", cela signifie que l'ordre dans lequel on considère les éléments a de l'importance et qu'un élément peut éventuellement être répété. Le modèle mathématique est la p -liste ou arrangement avec répétition.

- Si l'énoncé contient les mots **successif et sans remise**, cela signifie que l'ordre dans lequel on considère les éléments a de l'importance mais que tous les éléments considérés sont distincts (ou qu'il n'y a pas de répétition d'éléments). Le modèle mathématique est l'arrangement.
- Si l'énoncé contient le mot **simultanément**, cela signifie que l'ordre dans lequel on considère les éléments n'a pas d'importance. Le modèle mathématique est la combinaison.

Chapitre 7

Espace probabilisé

1. Expérience aléatoire

Activité 1. On lance un dé équilibré à 6 faces numérotés de 1 à 6 et on note le nombre figurant sur la face supérieure du dé après l'arrêt.

1. Quel est l'ensemble de résultats possibles ?
Peut-on prévoir avant le lancer les résultats possibles ?
2. Déterminer l'ensemble des résultats pairs.
3. Déterminer l'ensemble des possibilités d'avoir le chiffre "2" ?
4. Déterminer l'ensemble des possibilités d'avoir le chiffre "7" ?
5. Déterminer l'ensemble de résultats "obtenir un nombre inférieur 7 "

6. On considère les ensembles suivants :

A est l'ensemble "obtenir un nombre au moins gal 4 "

B est l'ensemble "obtenir un multiple de 2"

C est l'ensemble "obtenir le chiffre 5"

- (a) Déterminer l'ensemble "obtenir un multiple de 2 ou un nombre au moins gal 4".
- (b) Déterminer l'ensemble "obtenir un multiple de 2 et un nombre au moins gal 4".
- (c) Déterminer l'ensemble "obtenir un multiple de 2 et le chiffre 5".

Correction de l'activité 1. On lance un dé équilibré de 6 faces numrotées de 1 6 et on note le nombre figurant sur la face supérieure du dé après l'arrêt.

1. L'ensemble des résultats possibles appel univers des possibles est

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

On ne peut pas prévoir l'avance le résultat.

2. L'ensemble des résultats pairs est $\{2, 4, 6\}$

3. Nous avons une possibilité d'avoir le chiffre 2 est $\{2\}$.
4. Ce résultat est impossible.
5. L'ensemble des résultats "obtenir un nombre inférieur 7" est $\{1, 2, 3, 4, 5, 6\}$.
6. Nous avons $A = \{4, 5, 6\}$, $B = \{2, 4, 6\}$ et $C = \{5\}$.
 - (a) On note D l'ensemble "obtenir un multiple de 2 ou un nombre au moins égal 4" :
 $D = A \cup B = \{2, 4, 5, 6\}$.
 - (b) On note E l'ensemble "obtenir un multiple de 2 et un nombre au moins égal 4".
 $E = A \cap B = \{4, 6\}$
 - (c) On note F l'ensemble "obtenir un multiple de 2 et le chiffre 5"
 Cet ensemble est $F = A \cap C = \emptyset$.

Définition 7.1. Une expérience $\mathcal{E}$ est qualifiée d'aléatoire si on ne peut pas prévoir par avance son résultat et si, répétée dans des conditions identiques, elle peut donner lieu des résultats différents.

Remarque 7.1. Avant toute expérimentation, on peut décrire l'ensemble des résultats possibles de l'expérience aléatoire.

Définition 7.2. Soit $\mathcal{E}$ une expérience aléatoire. On appelle univers, et l'on note souvent Ω , l'ensemble des résultats possibles de $\mathcal{E}$. Si Ω est non vide. On notera $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω . Dans toute la suite de ce chapitre, on supposera que Ω est fini.

Définition 7.3. On appelle événement associé à une expérience aléatoire, toute partie A de Ω .

- Remarque 7.2.**
1. L'événement $A = \Omega$ est appel événement certain. Il se réalise toujours.
 2. L'événement $A = \emptyset$ est appel événement impossible. Il ne se réalise jamais.
 3. L'événement $A = \{\omega\}$ constitué d'un seul élément de Ω est appel événement élémentaire.

Les événements étant des ensembles, on utilisera 3 opérateurs définies sur les ensembles :

- l'union ; l'événement $A \cup B$ se réalise si A se réalise ou B se réalise
- l'intersection ; $A \cap B$ se réalise si A se réalise et B se réalise
- le complémentaire ; $\bar{A}$ se réalise si A ne réalise pas.

Application 1. Un sac contient trois boules de couleurs différentes ; une boule rouge, une boule verte et une boule bleue. On tire au hasard une boule du sac, on note sa couleur. On replace la boule dans le sac et on retire au hasard en notant nouveau sa couleur.

1. Déterminer l'univers des éventualités de cette expérience.
2. Citer un événement élémentaire et un événement non élémentaire
3. Soit A l'événement : "les deux boules sont de même couleur", B l'événement : "obtenir une boule bleue et une boule verte", et C l'événement : "obtenir d'abord une boule rouge"

- (a) Déterminer l'événement contraire de A
- (b) Déterminer l'événement : " A et B "; " A et C " puis l'événement " A ou C ", Les événements A et B sont-ils incompatibles ?

Correction de l'exercice d'application 1. On note :

R = "la couleur de la boule tirée est rouge"

B = "la couleur de la boule tirée est bleue"

V = "la couleur de la boule tirée est verte"

1. L'événement élémentaire est un couple (C_1, C_2) où C_1 représente la couleur de la première boule tirée et C_2 la couleur de la deuxième. L'univers des possibles est

$$\Omega = \{(R, R), (R, V), (V, R), (R, B), (B, R), (V, V), (V, B), (B, V), (B, B)\}.$$

2. (R, V) est un événement élémentaire ; $\{(R, R), (R, V)\}$ est un événement non élémentaire.
3. Nous avons

$$A = \text{"les deux boules sont de même couleur"} = \{(R, R), (V, V), (B, B)\}$$

$$B = \text{"obtenir une boule bleue et une boule verte"} = \{(B, V), (V, B)\}$$

$$C = \text{"obtenir d'abord une boule rouge"} = \{(R, R), (R, V), (R, B)\}.$$

- (a) L'événement contraire de A est

$$\bar{A} = \{(R, V), (V, R), (R, B), (B, R), (V, B), (B, V)\}.$$

- (b) Nous avons :

$$A \cap B = \emptyset$$

$$A \cap C = \{(R, R)\}$$

$$A \cup C = \{(R, R), (V, V), (B, B), (R, V), (R, B)\}$$

Les événements A et B sont incompatibles.

Définition 7.4. Soit Ω l'univers d'une expérience aléatoire, et $\mathcal{P}(\Omega)$ l'ensemble de parties de Ω . Alors, le couple $(\Omega, (\Omega))$ est appelé espace probabilisable.

2. Probabilité

Activité 2. Un sac contient trois boules de couleurs différentes ; une boule rouge, une boule verte et une boule bleue. On tire au hasard une boule du sac, on note sa couleur. On replace la boule dans le sac et on retire en notant à nouveau sa couleur. Chaque boule a la même chance d'être tirée.

1. Quel est le nombre de résultats possibles ?
2. Quelle est la fréquence d'apparition du couple (R, B) ?
3. Quelle est la fréquence d'apparition de chaque couple de l'univers.
Faire la somme de tous les résultats obtenus ?
4. Quelle est la fréquence d'apparition de deux boules de même couleur ?

Correction de l'activité 2. *Un sac contient trois boules de couleurs différentes ; une boule rouge, une boule verte et une boule bleue. On tire au hasard une boule du sac, on note sa couleur. On replace la boule dans le sac et on retire en notant nouveau sa couleur. Chaque boule a la même chance d'être tirée.*

1. $\text{card}(\Omega) = 9$.
2. La fréquence d'apparition du couple (R, B) est $\frac{1}{\text{card}(\Omega)} = \frac{1}{9}$.
3. La fréquence d'apparition de chaque couple de l'univers est $\frac{1}{9}$. De plus,

$$\frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} = 1$$

4. Tirer deux boules de même couleur revient tirer un élément de

$$A = \{(R, R), (V, V), (B, B)\}.$$

La fréquence d'apparition de deux boules de même couleur est $\frac{3}{9}$.

Définition 7.5. *On appelle probabilité sur l'univers Ω d'une expérience aléatoire l'application*

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

telle que

- $\mathbb{P}(\Omega) = 1$
- pour tout sous-ensemble $\{A_1, \dots, A_n, \dots\} \subset \mathcal{P}(\Omega)$ deux deux disjoints, nous

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Propriété 7.1. 1. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

2. si $A \cap B = \emptyset$ alors $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

3. $\mathbb{P}(A) + \mathbb{P}(\bar{A}) = 1$

Activité 3. On lance un dé truqué numéroté de 1 à 6 tel que $P_1 = P_2 = P_3 = P_4 = P_5 = \frac{1}{7}$ et $P_6 = \frac{2}{7}$ où P_i est la probabilité d'apparition du numéro i , $i \in \{1, 2, 3, 4, 5, 6\}$. Soit A l'événement "obtenir un nombre au moins égale à 4" et B ="obtenir un multiple de 2 "

1. Calculer la probabilité des événements A et B .
2. Déterminer les ensembles $A \cap B$ et $A \cup B$ puis calculer $\mathbb{P}(A \cap B)$ et $\mathbb{P}(A \cup B)$
3. Comparer $\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ et $\mathbb{P}(A \cup B)$
4. (a) Calculer la probabilité de l'événement $C = \text{"obtenir un nombre impair"}$
(b) Calculer $\mathbb{P}(B) + \mathbb{P}(C)$

Correction de l'activité 3. Soit A l'événement "obtenir un nombre au moins gal 4" et $B = \text{"obtenir un multiple de 2"}$

1. Nous avons $A = \{4, 5, 6\} = \{4\} \cup \{5\} \cup \{6\}$. Les événements $\{4\}$, $\{5\}$ et $\{6\}$ sont deux deux disjoints, nous obtenons

$$\mathbb{P}(A) = \mathbb{P}(\{4\}) + \mathbb{P}(\{5\}) + \mathbb{P}(\{6\}) = P_4 + P_5 + P_6 = \frac{4}{7}$$

De même, nous avons $B = \{2, 4, 6\}$ et $\mathbb{P}(B) = \frac{1}{7} + \frac{1}{7} + \frac{2}{7} = \frac{4}{7}$.

2. $A \cap B = \{4, 6\}$ et $A \cup B = \{2, 4, 5, 6\}$ et

$$\mathbb{P}(A \cap B) = \frac{3}{7} \quad \mathbb{P}(A \cup B) = \frac{5}{7}.$$

3. $\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(A \cup B)$

4. (a) $C = \{1, 3, 5\}$

$$(b) \mathbb{P}(C) = \frac{3}{7} \text{ et } \mathbb{P}(B) + \mathbb{P}(C) = 1.$$

Activité 4. Soit A et B deux événements tels que $\mathbb{P}(A) = 0.45$; $\mathbb{P}(B) = 0.60$ et $\mathbb{P}(A \cup B) = 0.80$ calculer $\mathbb{P}(A \cap B)$ et $\mathbb{P}(\bar{A})$

Correction de l'activité 4. 1. Nous avons

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Par suite

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B) = 0.25.$$

2. $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A) = 0.55$

Remarque 7.3. Une expérience se déroule dans les conditions équiprobables si tous les événements élémentaires ont la même probabilité de se réaliser. Dans ce cas, nous avons pour tout événement A ,

$$\begin{aligned} \mathbb{P}(A) &= \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = \frac{1}{\text{card}(\Omega)} \sum_{\omega \in A} 1 \\ &= \frac{\text{card}(A)}{\text{card}(\Omega)} \\ \mathbb{P}(A) &= \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}. \end{aligned}$$

Exercice de fixation 1. Dans un jeu de 32 cartes il y'a 4 As, on tire au hasard 4 cartes de ce jeu.

1. Calculer la probabilité d'obtenir 2 As.
2. Quelle est la probabilité de n'avoir aucun As ?
3. Quelle est la probabilité de tirer au moins un As ?

Correction de l'exercice de fixation 1. Dans un jeu de 32 cartes il y'a 4 As, on tire au hasard 4 cartes de ce jeu. Le nombre total de possibilités est le nombre de combinaisons de 4 cartes parmi 32 :

$$\text{card}(\Omega) = C_{32}^4.$$

1. Soit l'événement $A = \text{"obtenir 2 As dans le tirage"}$. Nous avons :

- C_4^2 possibilités de tirer 2 As parmi 4
- C_{28}^2 possibilités de tirer les 2 cartes restantes parmi 28.

D'après le principe multiplicatif, le nombre de possibilités d'obtenir 2 As dans le tirage est

$$\text{card}(A) = C_4^2 \times C_{28}^2.$$

La probabilité d'obtenir 2 As est donc

$$\mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{C_4^2 \times C_{28}^2}{C_{32}^4}.$$

2. Soit l'événement $B = \text{"n'avoir aucun As"}$. Nous avons C_{28}^4 possibilités d'obtenir un tirage sans aucun As, soit $\text{card}(B) = C_{28}^4$. La probabilité de n'avoir aucun As est donc

$$\mathbb{P}(B) = \frac{C_{28}^4}{C_{32}^4}.$$

3. Soit l'événement $C = \text{"avoir au moins un As"}$. L'événement contraire de C est B . Ainsi, nous obtenons

$$\mathbb{P}(C) = 1 - \mathbb{P}(B) = 1 - \frac{C_{28}^4}{C_{32}^4}.$$

Deuxième méthode : L'événement $C = \cup_{i=1}^4 C_i$ où l'événement $C_i = \text{"avoir exactement } i \text{ As avec } i \in \{1, 2, 3, 4\}$. Nous avons

- C_4^i possibilités de tirer i As parmi 4
- C_{28}^{4-i} possibilités de tirer les $4 - i$ cartes restantes parmi 28.

D'après le principe multiplicatif, le nombre de possibilités d'obtenir exactement i As dans le tirage est

$$\text{card}(A) = C_4^i \times C_{28}^{4-i}.$$

De plus C_1, C_2, C_3 et C_4 sont deux deux incompatibles. Ce qui implique que

$$\begin{aligned} \mathbb{P}(C) &= \mathbb{P}(C_1) + \mathbb{P}(C_2) + \mathbb{P}(C_3) + \mathbb{P}(C_4) \\ &= \frac{C_4^1 \times C_{28}^3 + C_4^2 \times C_{28}^2 + C_4^3 \times C_{28}^1 + C_4^4}{C_{32}^4} \end{aligned}$$

3. Modélisation d'une expérience aléatoire

Lors de la modélisation d'une expérience aléatoire $\mathcal{E}$, on est amené à choisir :

1. un univers Ω
2. une famille de parties de Ω . Dans le cas où l'univers Ω est fini, on considère $\mathcal{P}(\Omega)$
3. une probabilité $\mathbb{P}$.

Le triplet $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ est appelé espace probabilisé.

4. Probabilités conditionnelles, indépendance

4.1. Probabilité conditionnelle

Activité 5. Dans une classe de Terminale D de 36 élèves, 23 ont 18 ans, 29 sont des filles et 17 filles ont 18 ans. On choisit au hasard un élève de cette classe.

1. Calculer la probabilité des événements suivants : $A = \text{"l'élève a 18 ans"}$, $B = \text{"l'élève est une fille"}$, $C = \text{"l'élève est une fille de 18 ans"}$
2. Si l'élève est une fille, quelle est la probabilité pour qu'elle ait 18 ans ?
3. Comparer le résultat de la question 2 et $\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.

Correction de l'activité 5. Première méthode :

	18 ans	Autres	Total
Filles	17	12	29
Garçons	6	1	7
Total	23	13	36

1. Nous obtenons

$$\begin{aligned}\mathbb{P}(A) &= \frac{23}{36} \\ \mathbb{P}(B) &= \frac{29}{36} \\ \mathbb{P}(A \cap B) &= \frac{17}{36}.\end{aligned}$$

2. D'après le tableau, c'est parmi les 29 filles qu'on cherche celles qui ont 18 ans :

$$\mathbb{P}(C) = \frac{17}{29}$$

3. Nous obtenons

$$\begin{aligned}\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} &= \frac{\frac{17}{36}}{\frac{29}{36}} = \frac{17}{29} \\ \mathbb{P}(C) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}\end{aligned}$$

Deuxième méthode : arbres de choix. La somme des probabilités des branches issues d'un même noeud est 1. La probabilité de l'événement correspond à un trajet est le produit des probabilités des différentes branches composant ce trajet. En dehors des branches du premier niveau, les probabilités indiquées sont des probabilités conditionnelles.

Théorème 7.1. Soit une expérience aléatoire $\mathcal{E}$ d'univers Ω , $\mathbb{P}$ une probabilité sur Ω et B un événement tel que $\mathbb{P}(B) \neq 0$. L'application

$$\mathbb{P}_B : \mathcal{P} \longrightarrow [0, 1]$$

$$A \longmapsto \mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

est une probabilité sur Ω . $\mathbb{P}_B(A)$ se lit probabilité de A sachant B

Définition 7.6. L'application $\mathbb{P}_B$ ainsi définie s'appelle "probabilité conditionnelle sachant B ". La quantité $\mathbb{P}_B(A)$ est parfois notée $\mathbb{P}(A|B)$.

Exercice de fixation 2. Une urne contient trois boules rouges et deux boules blanches. On tire successivement avec remise deux boules de l'urne en notant leur couleur. Calculer la probabilité d'avoir deux boules de même couleur sachant que la première boule est rouge.

Correction de l'exercice de fixation 2. Le cardinal de l'univers est le nombre d'arrangements avec répétition d'un ensemble 2 éléments dans un ensemble 5 éléments, soit $\text{card}(\Omega) = 5^2$.

Soit A l'événement "avoir deux boules de même couleur"; $A = A_1 \cup A_2$ où A_1 "avoir deux boules rouges" et A_2 "avoir deux boules blanches"; $\text{card}(A_1)$ est le nombre d'arrangements avec répétitions d'un ensemble 2 éléments dans un ensemble 3 éléments soit $\text{card}(A_1) = 3^2$; $\text{card}(A_2)$ est le nombre d'arrangements avec répétitions d'un ensemble 2 éléments dans un ensemble à 2 éléments soit $\text{card}(A_2) = 2^2$; Par suite

$$\text{card}(A) = 3^2 + 2^2.$$

Soit B l'événement "la première boule tirée est rouge"; nous avons 3^1 possibilités de tirer une boule rouge au premier tirage et 5^1 possibilités de tirer une boule au second tirage, soit

$$\text{card}(B) = 3^1 \times 5^1.$$

$A \cap B$ = "les deux boules tirées sont rouges"; $\text{card}(A \cap B) = 3^2$. Nous obtenons

$$\mathbb{P}(A \cap B) = \frac{3^2}{5^2}$$

$$\mathbb{P}(B) = \frac{3^1 \times 5^1}{5^2}.$$

Nous déduisons que

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{3^2}{3^1 \times 5^1}.$$

Définition 7.7 ((Système complet d'événement)). On dit qu'une famille $(B_k)_{1 \leq n}$ est un **système complet d'événements** lorsque :

1. $\forall (i, j) \in [1; n]^2, (i \neq j) B_i \cap B_j = \emptyset$ (On dit alors que les $(B_k)_{1 \leq n}$, sont deux à deux disjoints ou incompatibles)
2. $\bigcup_{k=1}^n B_k = \Omega$.

Autrement dit les $B_k, k = 1, \dots, n$, constituent une partition de Ω .

Exemple 7.1. Soit A une partie non trivial de Ω . Alors la paire $\{A, \bar{A}\}$ est un système complet d'événement.

Théorème 7.2 ((Probabilité totale)). Soit $(\Omega, \mathcal{P}, \mathbb{P})$ un espace probabilisé et $(B_k)_{1 \leq n}$ un système complet d'événement. Alors pour tout événement A , on a :

$$\mathbb{P}(A) = \sum_{k=1}^n \mathbb{P}(A \cap B_k) = \sum_{k=1}^n \mathbb{P}(A|B_k)\mathbb{P}(B_k)$$

Corollaire 7.1 ((Formule de Bayes)). Soit $(\Omega, \mathcal{P}, \mathbb{P})$ un espace probabilisé et $(B_k)_{1 \leq n}$ un système complet d'événement. Alors pour tout événement A , on a $\forall j \in [1; n]$,

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\sum_{k=1}^n \mathbb{P}(A|B_k)\mathbb{P}(B_k)}.$$

4.2. Indépendance

Définition 7.8. Deux événements A et B de probabilités non nulles sont **indépendants** lorsque la réalisation (ou non) de l'un n'a pas d'influence sur la probabilité de réalisation de l'autre : $\mathbb{P}_A(B) = \mathbb{P}(B)$ ou $\mathbb{P}_B(A) = \mathbb{P}(A)$.

Théorème 7.3. Deux événements A et B sont indépendants si et seulement si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

Exercice de fixation 3. On lance une pièce de monnaie non truquée deux fois de suite et on note le couple de côtés qui apparaît.

1. Les événements : $A =$ "face apparaît au premier lancer " et $B =$ "pile apparaît au deuxième lancer " sont-ils indépendants ?
2. Les événements : $C =$ "le même côté apparaît deux fois" et $D =$ " le nombre d'apparition de " face " est différent de deux " sont-ils indépendants ?

Correction de l'exercice de fixation 3. On lance une pièce de monnaie non truquée deux fois de suite et on note le couple de côtés qui apparaît. L'univers est

$$\Omega = \{(P, P), (P, F), (F, F), (F, P)\}$$

1. $A = \{(F, F), (F, P)\}$ $B = \{(P, P), (F, P)\}$ $A \cap B = \{(F, P)\}$. Nous avons

$$\mathbb{P}(A \cap B) = \frac{1}{4} \quad \mathbb{P}(A) = \frac{2}{4} \quad \mathbb{P}(B) = \frac{2}{4}.$$

On note que $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. On déduit que A et B sont indépendants.

2. $C = \{(F, F), (P, P)\}$ $D = \{(P, P), (F, P), (P, F)\}$ $C \cap D = \{(P, P)\}$. Nous avons

$$\mathbb{P}(C \cap D) = \frac{1}{4} \quad \mathbb{P}(C) = \frac{2}{4} \quad \mathbb{P}(D) = \frac{3}{4}.$$

On note que $\mathbb{P}(C \cap D) \neq \mathbb{P}(C)\mathbb{P}(D)$. On déduit que C et D ne sont pas indépendants.

Chapitre 8

Variables aléatoires réelles

On fait une expérience aléatoire qui est traduite par l'espace probabilisé $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$. Maintenant on s'intéresse à certaines conséquences de cette expérience.

1. Généralités

Définition 8.1. Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $(E, \mathcal{B})$ un espace probabilisable. On appelle variable aléatoire toute application X définie sur Ω à valeurs dans E telle que

$$\forall A \in \mathcal{B}, \quad X^{-1}(A) = \{\omega \in \Omega; X(\omega) \in A\} \in \mathcal{A}.$$

Remarque 8.1. 1. (i) Si $(E, \mathcal{B}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, X est une variable aléatoire réelle

2. (ii) Soit X une variable aléatoire réelle. Si $X(\Omega)$ est un sous-ensemble fini ou infini dénombrable de $\mathbb{R}$, alors la v.a.r X est dite discrète. Sinon, elle est dite continue.

Définition 8.2. Soit X une variable aléatoire définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ valeurs dans l'espace probabilisable $(E, \mathcal{B})$. On appelle loi de probabilité de X , la probabilité $\mathbb{P}_X$ définie sur $\mathcal{B}$ par :

$$\forall A \in \mathcal{B}, \quad \mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(X \in A).$$

Définition 8.3. Soit X une variable aléatoire réelle définie sur $(\Omega, \mathcal{A}, \mathbb{P})$. On appelle fonction de répartition de la v.a.r X , la fonction F définie par :

$$F : \mathbb{R} \longrightarrow [0, 1] \\ x \longrightarrow F(x) = \mathbb{P}(X \leq x) = \mathbb{P}_X([-\infty, x]).$$

Propriété 8.1. 1. (i) F est une fonction non décroissante

2. [b] $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$.

3. (c) F est continue à droite et limitée à gauche.

Définition 8.4. Soit X une variable aléatoire réelle. Supposons que la fonction de répartition F soit continue et strictement croissante. Pour $0 \leq \alpha \leq 1$; on note x_α l'unique nombre réel vérifiant

$$F(x_\alpha) = \mathbb{P}(X < x_\alpha) = \alpha.$$

On dit X_α est le quantile d'ordre α .

Remarque 8.2. Pour connaître la loi d'une variable aléatoire discrète X , il faut connaître l'ensemble de ses valeurs possibles, et la probabilité avec laquelle elle réalise chaque valeur i.e

$$X(\Omega) = \{x_1, \dots, x_n, \dots\} \subset \mathbb{R} \text{ et } P(X = x_i).$$

Pour toute fonction h , on définit l'espérance de $h(X)$ par

$$\mathbb{E}[h(X)] = \sum_{i=1}^n h(x_i)P(X = x_i).$$

En particulier

si $h(x) = |x|^p$, $p \geq 1$ alors on parle de moment d'ordre p de la v.a. X . Le moment d'ordre 1 est appelé l'espérance de X .

si $h(x) = |x - \mathbb{E}(X)|^p$, $p \geq 1$ alors on parle de moment centré d'ordre p de la v.a. X . Le moment centré d'ordre 2 est appelé la variance de X .

Remarque 8.3. 1. (i) X une variable aléatoire continue est absolument continue s'il existe une fonction définie sur $\mathbb{R}$, positive, vérifiant $\int_{-\infty}^{\infty} f(t)dt = 1$, telle que $\forall x \in \mathbb{R}$,

$$\mathbb{P}(X < x) = \int_{-\infty}^x f(t)dt.$$

La fonction f est appelée densité de probabilité de la variable aléatoire X .

2. (ii) Pour toute fonction continue h , on définit l'espérance de $h(X)$ par

$$E[h(X)] = \int_{-\infty}^{\infty} h(x)f(x)dx.$$

En particulier

si $h(x) = |x|^p$, $p \geq 1$ alors on parle de moment d'ordre p de la v.a. X . Le moment d'ordre 1 est appelé l'espérance de X .

si $h(x) = |x - \mathbb{E}(X)|^p$, $p \geq 1$ alors on parle de moment centré d'ordre p de la v.a. X . Le moment centré d'ordre 2 est appelé la variance de X .

Remarque 8.4. La loi d'une variable aléatoire est complètement déterminée via sa fonction de répartition, ou via sa densité de probabilité.

Proposition 8.1. (*Inégalité de Markov*)

Soient X une v.a.r telle que $\mathbb{E}(|X|) < +\infty$. Alors pour tout $c \in \mathbb{R}_+$

$$\mathbb{P}(|X| > c) \leq \frac{\mathbb{E}[|X|]}{c}.$$

Proposition 8.2. (*Inégalité de Tchebychev*)

Soient X une v.a.r $\mathbb{E}(X^2) < +\infty$ est définie. Alors pour tout $c > 0$

$$\mathbb{P}(|X - \mathbb{E}(X)| > c) \leq \frac{\text{Var}(X)}{c^2}.$$

2. Vecteurs aléatoires

Définition 8.5. On appelle vecteur aléatoire de dimension n un n -uplet $X = (X_1, \dots, X_n)$ où chaque X_i est une variable aléatoire réelle.

Définition 8.6. On appelle fonction de répartition de $X = (X_1, \dots, X_n)$ la fonction $F_X : \mathbb{R}^n \rightarrow [0, 1]$ définie par

$$F_X(x_1, \dots, x_n) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n).$$

2.1. Vecteurs aléatoires discrets

Définition 8.7. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Un vecteur aléatoire X de dimension n , est dit discret si $X(\Omega)$ est un sous ensemble fini ou infini dénombrable de $\mathbb{R}^n$.

Définition 8.8. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et X un vecteur aléatoire discret de dimension n . Pour décrire la loi $\mathbb{P}_X$ de X , on donne :

- $X(\Omega)$.
- pour tout $x = (x_1, \dots, x_n) \in X(\Omega)$:

$$\mathbb{P}_X(X = x) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n).$$

Nous nous limitons au cas $n = 2$. Les définitions et propriétés des couples s'étendent facilement aux vecteurs n dimensionnels.

Proposition 8.3. Soit $X = (X_1, X_2)$ un vecteur aléatoire discrète de dimension 2. Alors :

- $X_1(\Omega) = \{x_1 \in \mathbb{R} : \exists x_2 \in \mathbb{R} \text{ tel que } (x_1, x_2) \in X(\Omega)\}.$
- $X_2(\Omega) = \{x_2 \in \mathbb{R} : \exists x_1 \in \mathbb{R} \text{ tel que } (x_1, x_2) \in X(\Omega)\}.$
- pour tout $x_1 \in X_1(\Omega)$, $\mathbb{P}(X_1 = x_1) = \sum_{x_2 \in X_2(\Omega)} \mathbb{P}(X_1 = x_1, X_2 = x_2).$
- pour tout $x_2 \in X_2(\Omega)$, $\mathbb{P}(X_2 = x_2) = \sum_{x_1 \in X_1(\Omega)} \mathbb{P}(X_1 = x_1, X_2 = x_2)$

Définition 8.9. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, X et Y deux variables aléatoires réelles discrètes définies sur Ω , X et Y sont indépendantes si et seulement si

$$\forall x \in X(\Omega), \forall y \in Y(\Omega), \mathbb{P}(\{X = x\} \cap \{Y = y\}) = \mathbb{P}(X = x)\mathbb{P}(Y = y).$$

Définition 8.10. On appelle loi conditionnelle de Y sachant $\{X = x\}$ la probabilité

$$\mathbb{P}(Y = y|X = x) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)}, \quad \forall y \in Y(\Omega).$$

Définition 8.11. On appelle espérance conditionnelle de Y sachant $\{X = x\}$, l'espérance conditionnelle de Y sachant $\{X = x\}$:

$$\mathbb{E}(Y|X = x) = \sum_{y \in Y(\Omega)} y \mathbb{P}(Y = y|X = x).$$

Définition 8.12.

$$\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(Y).$$

2.2. Vecteurs aléatoires à densité

Définition 8.13. On dit que $X = (X_1, \dots, X_n)$ admet une densité $f_X : \mathbb{R}^n \rightarrow \mathbb{R}_+$ ssi pour tout Borelien $B \in \mathbb{B}(\mathbb{R}^n)$ on peut exprimer la probabilité de l'événement $\{X \in B\}$ par une intégrale sur B de f_X :

$$P(X \in B) = \int_B f_X(x_1, \dots, x_n) dx_1 \dots dx_n$$

avec

$$\int_{\mathbb{R}^n} f_X(x_1, \dots, x_n) dx_1 \dots dx_n = 1.$$

Remarque 8.5. Si la fonction de répartition F_X est différentiable, on peut déterminer la densité en dérivant F_X :

$$f_X(x_1, \dots, x_n) = \frac{\partial^n F_X(x_1, \dots, x_n)}{\partial x_1 \dots \partial x_n}$$

Définition 8.14. Si X est un vecteur aléatoire dans $\mathbb{R}^n$ admettant une densité f_X alors tout sous-vecteur Y de X de dimension $k \leq n$ admet une densité qu'on obtient en intégrant f_X par rapport aux composantes qui ne figurent pas dans Y . On appelle cette densité la densité marginale de Y .

Remarque 8.6. Cas particulier ($n = 2$). Soit $X = (X_1, X_2)$ un vecteur aléatoire de dimension 2 de densité f_X . La densité marginale de X_1 est $f_{X_1}(x_1) = \int_{\mathbb{R}} f_X(x_1, x_2) dx_2$ et celle de X_2 est $f_{X_2}(x_2) = \int_{\mathbb{R}} f_X(x_1, x_2) dx_1$

Définition 8.15. Soit $Z = (X, Y)$ un vecteur aléatoire dans $\mathbb{R}^m \times \mathbb{R}^n$ admettant une densité f_Z . Soient f_X et f_Y les densités marginales des vecteurs X et Y telles que $f_Y(y) > 0$.

- On appelle densité conditionnelle de X sachant $Y = y$ la densité donnée par

$$f_{X|Y=y}(x) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)}.$$

- On appelle espérance conditionnelle de X sachant $Y = y$ la quantité

$$\mathbb{E}(X|Y = y) = \int_{\mathbb{R}} x f_{X|Y=y}(x) dx.$$

Proposition 8.4. Soient X et Y deux v.a. avec densité jointe $f_{(X,Y)}(x, y)$. Alors X et Y sont indépendantes ssi il existe deux applications g, h telles que $f_{(X,Y)}(x, y) = g(x)h(y)$ pour tout couple (x, y) tel que $f_{(X,Y)}(x, y) > 0$.

2.3. Matrice de variance-covariance

Définition 8.16. Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire. Si $X_1, \dots, X_n$ admettent toutes un moment d'ordre 1, alors le vecteur

$$\mathbb{E}(X) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_n))$$

est appelé espérance de X .

Définition 8.17. Soient X_1 et X_2 deux v.a.r admettant toutes des moments d'ordre 2. Alors

$$\text{Cov}(X_1, X_2) = \mathbb{E}[(X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2))] = \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2)$$

est appelée la covariance entre X_1 et X_2 .

Définition 8.18. Si X_1 et X_2 admettent des moments d'ordre 2 telles que $\text{Var}(X_1)\text{Var}(X_2) > 0$, alors

$$\rho(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1)}\sqrt{\text{Var}(X_2)}}.$$

Définition 8.19. Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire tel que chaque composante admet un moment d'ordre 2. On appelle matrice de variance-covariance la matrice symétrique positive :

$$\Sigma_X = \mathbb{E}[(X - \mathbb{E}(X))(X - \mathbb{E}(X))^t] = (\text{Cov}(X_i, X_j))_{1 \leq i, j \leq n}$$

Théorème 8.1. Si $X_1, \dots, X_n$ sont indépendantes alors

1. $\mathbb{E}(X_1 \dots X_n) = \mathbb{E}(X_1) \dots \mathbb{E}(X_n)$
2. $\Sigma_X = \text{diag}(\text{Var}(X_1), \dots, \text{Var}(X_n))$

2.4. Fonction caractéristique

Définition 8.20. Si X est une variable alatoire valeurs dans $\mathbb{R}^n$, sa fonction caractéristique est $\phi_X : \mathbb{R}^n \longrightarrow \mathbb{C}$ défini par

$$\phi_X(u) = E(e^{i\langle u, X \rangle})$$

où $\langle u, X \rangle$ désigne le produit scalaire de $X = (X_1, \dots, X_n)^t$ et $u = (u_1, \dots, u_n)^t$:

$$\langle u, X \rangle = \sum_{i=1}^n u_i X_i$$

Propriété 8.2. 1. ϕ_X est bornée : $\forall u \in \mathbb{R}^n, |\phi_X(u)| \leq 1, \phi_X(0) = 1$.

2. ϕ_X est continue

3. $\phi_X(-u) = \bar{\phi}_X(u)$

4. Si $X_1, \dots, X_n$ sont indépendantes alors

$$\phi_{\sum_{i=1}^n X_i}(u) = \prod_{i=1}^n \phi_{X_i}(u).$$

5. $X_1, \dots, X_n$ sont indépendantes $\iff \phi_{(X_1, \dots, X_n)^t}(u) = \prod_{i=1}^n \phi_{X_i}(u)$.

6. X est une variable valeurs dans $\mathbb{R}^n$ et A une matrice p lignes et n colonnes, B un vecteur de dimension n , alors

$$\phi_{AX+B}(u) = e^{i\langle u, B \rangle} \phi_X(A^t \cdot u), \quad \forall u \in \mathbb{R}^n.$$

7. Si $E(|X|^k) < \infty$, ϕ_X est k fois continûment différentiable sur $\mathbb{R}^n$ et

$$\frac{\partial^k \phi_X(t)}{\partial x_{i_1} \partial x_{i_2} \dots \partial x_{i_k}} = i^k E(e^{i\langle t, X \rangle} X_{i_1} X_{i_2} \dots X_{i_k}).$$

Théorème 8.2. (Thorme d'inversion)

Si X est une variable alatoires valeurs dans $\mathbb{R}^n$, et si ϕ_X est une fonction intégrable sur $\mathbb{R}^n$, alors X admet une densité f_X telle que

$$f_X(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{-i\langle t, x \rangle} \phi_X(t) dt.$$

2.5. Vecteurs gaussiens

2.6. Généralités

Définition 8.21. Soit $X = (X_1, \dots, X_n)^t$ un vecteur alatoire de dimension n . On dit que X est un vecteur gaussien si toute combinaison linéaire de ses composantes suit une loi gaussienne.

Proposition 8.5. 1. Si X est un vecteur gaussien alors ses composantes sont gaussiennes. La réciproque est fautive. En effet, soient Y et ε deux variables alatoires indépendantes telles que Y suit la loi gaussienne centrée réduite $\mathcal{N}(0, 1)$ et que $P(\varepsilon = 1) = P(\varepsilon = -1) = 1/2$. Soit $W = \varepsilon Y$. Alors (Y, W) n'est pas un vecteur gaussien car $Y+W$ n'est pas une variable gaussienne ($P(Y+W=0) = 1/2$). Pourtant chacune des composantes est une v.a.r gaussienne

2. Soient $X_1, \dots, X_n$ des variables aléatoires réelles gaussiennes. Si elles sont indépendantes alors le

vecteur $X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ est gaussien.

3. Soit X un vecteur gaussien. Les composantes de X sont indépendantes si et seulement si Σ_X est une matrice diagonale

4. Si X est un vecteur gaussien, sa fonction caractéristique est

$$\phi_X(u) = \exp\{iu^t E(X) - \frac{1}{2}u^t \Sigma_X u\}, \quad \forall u \in \mathbb{R}^n,$$

où Σ_X est la matrice de covariance de X .

Définition 8.22. Si $\det(\Sigma_X) = 0$, on dit que X est dit d'gnr.

Si $\det(\Sigma_X) > 0$, on dit que X est dit non-d'gnr.

Proposition 8.6. Si $X \sim \mathcal{N}(m, \Sigma_X)$ est non-dégénéré alors

$$f_X(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det(\Sigma_X)}} \exp\left\{-\frac{1}{2}(x-m)^t \Sigma_X^{-1}(x-m)\right\} \mathbb{I}_{\mathbb{R}^n}(x).$$

Théorème 8.3. (Théorème de Cochran)

Soit $X_1, \dots, X_n$ n variables aléatoires indépendantes identiquement distribuées de même loi $\mathcal{N}(m, \sigma^2)$, $\sigma > 0$. Alors

$$1. \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \sim \mathcal{N}(m, \frac{\sigma^2}{n})$$

$$2. \frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2 \text{ avec}$$

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

3.

$$\sqrt{n} \frac{(\bar{X}_n - m)}{S_n} \sim T_{n-1}$$

4. $\bar{X}_n$ et S_n^2 sont indépendantes

2.7. Lois dérivées

1. Loi du χ^2 : soit $X_1, \dots, X_n$ des variables aléatoires indépendantes identiquement distribuées de loi $\mathcal{N}(0, 1)$, alors

$$X \sum_{i=1}^n X_i^2$$

suit par définition une loi du χ^2 n dgrs de lib. On note $X \rightsquigarrow \chi^2(n)$.

2. Loi de student : soient Y et Z deux variables alatoires indépendantes telles $Y \rightsquigarrow \chi^2(n)$ et $Z \rightsquigarrow \mathcal{N}(0, 1)$. On appelle loi de Student n degrs de libert la loi suivie par la variable alatoire relle

$$X = \frac{Z}{\sqrt{Y/n}}.$$

On note $X \rightsquigarrow T(n)$. La densité de la loi de Student est une fonction symétrique. $T(n)$ converge en loi vers une $\mathcal{N}(0, 1)$ quand $n \rightarrow +\infty$.

3. Loi de Fisher-Snedecor : soient Y_1 et Y_2 deux variables alatoires relles indépendantes telles que : $Y_1 \rightsquigarrow \chi^2(n_1)$ et $Y_2 \rightsquigarrow \chi^2(n_2)$. On appelle loi de Fisher-Snedecor n_1 et n_2 degrs de libert la loi suivie par la variable alatoire relle

$$X = \frac{Y_1/n_1}{Y_2/n_2}.$$

On note $X \rightsquigarrow F(n_1, n_2)$.

3. Convergence

3.1. Convergence de variables alatoires

Soit $(X_n)_{n \geq 1}$ une suite de variables alatoires relles définies sur le mme espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

Définition 8.23. (X_n) converge en probabilité vers X si

$$\forall \epsilon > 0, \quad \lim_{n \rightarrow \infty} \mathbb{P}\{|X_n - X| > \epsilon\} = 0.$$

Théorème 8.4. Si $\mathbb{E}(X_n) \rightarrow m$ et $\text{Var}(X_n) \rightarrow 0$, quand $n \rightarrow +\infty$ alors X_n converge en probabilité vers m .

Définition 8.24. On dit qu'une suite de variables alatoires X_n valeurs dans $\mathbb{R}^d$ converge en loi vers une variable alatoire X si pour toute fonction continue et bornée sur $\mathbb{R}^d$,

$$\mathbb{E}(f(X_n)) \longrightarrow \mathbb{E}(f(X))$$

quand n tend vers l'infini.

Proposition 8.7. X_n converge en loi vers X , si F_n désignant la suite de fonctions de répartition de la suite X_n , et F tant la fonction de répartition de X ,

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

en tout point x de continuité de F .

Proposition 8.8. (Théorème de Paul Lévy)

Soit X_n une suite de variables alatoires, et on note $\phi_n(u)$ la fonction caractéristique de X_n :

a) Si X_n converge en loi vers X , et si X admet $\phi(u)$ pour fonction caractéristique, alors

$$\lim_{n \rightarrow +\infty} \phi_n(u) = \phi(u), \quad \forall u \in \mathbb{R}^d.$$

b) Si $\phi_n(u)$ converge simplement vers une certaine fonction $\phi(u)$ et si $\phi(u)$ est continue en 0, alors $\phi(u)$ est la fonction caractéristique d'une variable aléatoire X telle que X_n converge en loi vers X .

Proposition 8.9. 1. Si $X_n \xrightarrow{\text{loi}} X$ et $Y_n \xrightarrow{P} a$ ($a \in \mathbb{R}$) alors

$$(a) \quad X_n + Y_n \xrightarrow{\text{loi}} X + a,$$

$$(b) \quad X_n Y_n \xrightarrow{\text{loi}} aX$$

$$(c) \quad \frac{X_n}{Y_n} \xrightarrow{\text{loi}} \frac{X}{a} \quad (a \neq 0)$$

2. Si $f : \mathbb{R}^d \rightarrow \mathbb{R}$ est continue, alors

$$X_n \xrightarrow{P} X \implies f(X_n) \xrightarrow{P} f(X)$$

$$X_n \xrightarrow{\text{loi}} X \implies f(X_n) \xrightarrow{\text{loi}} f(X).$$

3.2. Lois des grands nombres

Proposition 8.10. Soit (X_n) une suite de variables aléatoires indépendantes identiquement distribuées, de moyenne m et de variance σ^2 . Alors $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ converge en probabilité vers m .

3.3. Théorème Central Limite

Proposition 8.11. Soit (X_n) une suite de variables aléatoires indépendantes identiquement distribuées, de moyenne m et de variance σ^2 . Alors $\frac{\sqrt{n}(\bar{X}_n - m)}{\sigma}$ converge en loi vers $\mathcal{N}(0, 1)$.

Resultats importants :

- Cas σ^2 est connue :

$$\frac{\sqrt{n}(\bar{X}_n - m)}{\sigma} \xrightarrow{\text{loi}} \mathcal{N}(0, 1) \quad \text{quand } n \rightarrow +\infty.$$

En pratique, nous avons

$$\frac{\sqrt{n}(\bar{X}_n - m)}{\sigma} \rightsquigarrow \mathcal{N}(0, 1) \quad n \geq 30.$$

Pour $n \leq 30$, le résultat reste vrai seulement si X suit une loi normale.

- Cas σ^2 est inconnue :

- si X suit une loi normale $\mathcal{N}(m, \sigma^2)$ alors

$$\frac{\sqrt{n}(\bar{X}_n - m)}{S_n} \rightsquigarrow T(n-1) \quad \text{loi de Student } n-1 \text{ dgrs de libert}$$

De plus,

$$T(n) \rightarrow \mathcal{N}(0, 1)$$

En pratique pour $n \geq 30$, $T(n) \rightsquigarrow \mathcal{N}(0, 1)$.

- si X ne suit pas une loi normale la loi suivie par $\frac{\sqrt{n}(\bar{X}_n - m)}{S_n}$ n'est pas connue. Mais, nous avons

$$\frac{\sqrt{n}(\bar{X}_n - m)}{S_n} \xrightarrow{\text{loi}} \mathcal{N}(0, 1) \quad \text{quand } n \rightarrow +\infty.$$

4. Approximation

- Convergence en loi d'une variable alatoire hypergomtrique vers une loi binomiale : soit $X_N \rightsquigarrow \mathcal{H}(N, M, n)$; si $\lim_{N \rightarrow +\infty} \frac{M}{N} = p$, alors

$$X_N \overset{\text{loi}}{\approx} \mathcal{B}(n, p)$$

Dans la pratique, le remplacement de $X_N \rightsquigarrow \mathcal{H}(N, M, n)$ par $X \rightsquigarrow \mathcal{B}(n, M/N)$ est excellent ds que $20n < N$, $10n < M$, $10n < N - M$.

- Convergence en loi d'une variable alatoire binomiale vers une loi Poisson : Si une variable alatoire $X_n \rightsquigarrow \mathcal{B}(n, p)$, Si $\lim_{n \rightarrow \infty} np = \lambda > 0$ alors

$$X_n \overset{\text{loi}}{\approx} \mathcal{P}(\lambda)$$

Dans la pratique, le remplacement de $X_n \rightsquigarrow \mathcal{B}(n, p)$ par $X \rightsquigarrow \mathcal{P}(np)$ est bon ds que $n > 30$, $np < 10$.

- Convergence en loi d'une variable alatoire binomiale vers une Normale : $X \rightsquigarrow \mathcal{B}(n, p)$ et $U_n = \frac{X_n - np}{\sqrt{np(1-p)}}$. Quand $n \rightarrow \infty$, $U_n \xrightarrow{\text{loi}} \mathcal{N}(0, 1)$.

Dans la pratique, le remplacement de $X \rightsquigarrow \mathcal{B}(n, p)$ par $X \rightsquigarrow \mathcal{N}(np, np(1-p))$ est bon ds que $np > 10$, $n > 20$, $n(1-p) > 10$.

- Convergence en loi d'une variable alatoire de Poisson vers une Normale : soit $X_\lambda \rightsquigarrow \mathcal{P}(\lambda)$ et $U_\lambda = \frac{X_\lambda - \lambda}{\sqrt{\lambda}} \xrightarrow{\text{loi}} \mathcal{N}(0, 1)$ quand $\lambda \rightarrow \infty$

Dans la pratique, le remplacement de $X_\lambda \rightsquigarrow \mathcal{P}(\lambda)$ par $X_\lambda \rightsquigarrow \mathcal{N}(\lambda, \lambda)$ est bon si $\lambda > 10$

- Convergence en loi d'une variable alatoire de Student n dgrs de libert vers une loi Normale :

$$T(n) \xrightarrow{\text{loi}} \mathcal{N}(0, 1), \text{ quand } n \rightarrow \infty$$

Dans la pratique, le remplacement de T_n par $U \rightsquigarrow \mathcal{N}(0, 1)$ est bon ds que $n > 60$.