STATISTIQUE DESCRIPTIVE ET PROBABILITE

Prof. Auguste AMAN
UFR de Mathématiques et Informatique
Université de Cocody-Abidjan

Table des matières

Introduction

Chapitre 1

Statistique descriptive

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.

1.1 Les données statistiques

1.1.1 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 $(x_1; x_2; \cdots; x_n)$ 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 elèves 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.*

1.1.2 Les types de variables

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

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'hiérachie entre les differentes modalités ; exemple : sexe, couleur des yeux.
- les varibles **qualitatives ordinales** : les differentes modalités peuvent être ordonnées de manière naturelle ; exemple : la mention au baccalauréat, la frequence d'une activité.

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

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 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*

1.1.3 Les variables qualitatives : tableaux de fréquence et représentation graphique

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.

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 $C_1, \ldots, C_k$ 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 **ordinale** : $\phi_j = f_1 + f_2 + \cdots + f_j$ la frequence relative cumulée pour cette modalité
(avec la convention : $\phi_0 = 0$). Elle n'a de sens que si la variable est qualitative ordinale et si les modalités $C_1, C_2, \ldots, C_k$ sont ordonnées suivant l'odre croissant naturel (ou hiérachique ascendant) qui règne parmi ces modalités.

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

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

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

Représentation graphique : diagrammes en secteurs et diagrammes en tuyaux d'orgue

1. **Diagrammes en secteurs** : chaque modalité est représentée par un secteur de disque dont l'angle est proportionnel à la fréquence de la modalité (ou pourcentage), l'angle 360 degrés équivalent à la fréquence relative 1 (ou au pourcentage 100%).

   Exemple 5.

2. **Diagramme en tuyaux d'orgue** : en abscisse sont disposées les différentes modalités aux quelles on associe des rectangles espacés entre eux, de largeur constente, dont la hauteur (en ordonnée) sont proportionnelle à l'effectif ou à la fréquence relative de chaque modalité. Preciser 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 6.

Remarque 3. *Cette representation graphique est plus adaptée dans le cas d'une variable qualitative ordinale car elle rend conte de la structure d'ordre entre les modalités, disposée de gauche à droite en ordre croissant. C'est imposible de suggérer une structure d'ordre dans un diagramme en secteur.*

1.1.4 Les variables quantitatives discrètes

Exemple 7. *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.

Tableaux de distribution de fréquences

Exemple 8. *Pour la variable P, on remplie le tableau suivant :*

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

On note $v_1, v_2, \ldots, v_k$ les k valeurs différentes que peut prendre la variable (on n'en rencontrera pas pas d'exemple dans ce cour, 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 $\Phi_j = f_1 + \cdots + f_j$ la j-ème fréquence relative cumulée (avec la convention :

$\Phi_0 = 0$). On résume habituelement les données comme dans le tableau-type suivant :

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

Représentation graphique

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 : pointure.

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 9. *Pointure*

1.1.5 Les variables quantitatives continues

Exemple 10. *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.

Tableaux de distribution de fréquences-fréquences cumulées

Les données brutes de la variable pour chaque individu sont notées $x_1, \ldots, x_n$. 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 amoncellemment 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 : $[b_0, b_1[$, $\quad [b_1, b_2[, \ldots, \quad [b_{k-1}, b_k[$ (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 $\Phi_j = f_1 + \cdots + f_j$ 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 11. *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 4.
- *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 5. *Contrairement au cas d'une variable qualitative ou discrète, ce tableau représente une perte d'information par rapport aux données brutes*

Représentation graphique

Histogramme Sur l'axe des abscisses sont placées les bornes des classes en respectant une échelle. Pour chaque classe, on élève un rectangle de hauteur proportionnelle à la densité de proportion.
Exemple de taille T :

Remarque 6. *On représente la* ***densité de proportion*** *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 conrespondant à la classe* $[b_j, b_{j-1}[$ *l'aire est*

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

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 $F(x)$:

$$F(x) = f_1 + f_2 + \cdots + f_{j-1} + (x - b_j) \times d_j = \phi_j + (x - b_{j-1}) \times d_j$$

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.

1.2 Résumés numériques d'une variable quantitative

Dans ce chapitre, X désigne une variable quantitative.

1.2.1 Paramètre de tendance centrale

Le mode

Le mode rend compte de l'endroit où les données sont le plus concentrées.
Pour une variable **discrète**, le mode est la où les valeurs de la variable qui correspond(ent) à l'**effectif maximal** (ou à la fréquence maximal).
Pour une variable **continue** regroupée en classe, le mode est là où les classes de **densité de proportion maximale**.

Exemple 13. *Pointure, taille.*

La moyenne

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

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

Exemple 14. *pointure, taille*

Cas d'une variale 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 :

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

Exemple 15. *Pointure.*

Cas d'une variable continue regroupée en classes : la variable X est regroupée das les classes $[b_{j-1}, b_j[$ $(1 \leq j \leq n)$, les fréquences relatives associées à ces classe 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 représentant des classes(leur centre) : $c_j = (b_j + b_{j-1})/2$, par la formule :

$$\bar{x}_{app} = f_1 c_1 + f_2 c_2 + \cdots + f_k c_k = \sum_{i=1}^{n} f_j c_j$$

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.*

La médiane

"En gros", le calcul de la médiane revient à ranger les observations par ordre croissant et trouver un point au-dessus duquel se situent 50% des observations.

a) Cas d'une variable discrète.
- Si n est **impair**, la médiane est la $\frac{n+1}{2}$-ième observation.
- Si n est **pair**, il y'a plusieur façons convenables de définir la médiane. Nous choisirons la suivante : la médiane est la plus petite valeur observée v_j telle que l'effectif cumulé en v_j dépasse $n/2$. Autrement dit, c'est la plus petite valeur v_j pour laquelle la proportion cumulée dépasse $\frac{1}{2}$.

Remarque 7. *Cette définition est encore vraie pour n* impair.
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 Q_2 de l'équation :

$$F(Q_2) = 0,5$$

où F est la fonction de répartion empirique de la variable. On sait que cette solution existe parce que F est continue, et $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to +\infty} F(x) = 1$. Si de plus F est strictement croissante, la solution Q_2 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 exatement $0,5$, 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 frequence 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 :

$$F(Q_2) = \Phi_{j-1} + (Q_2 - b_j) \times d_j = 0,5$$

d'où

$$Q_2 = \frac{0,5 - \phi_{j-1}}{d_j} + b_{j-1}$$

Ou encore, en terme de b_j et de F :

$$Q_2 = \frac{0,5 - F(b_{j-1})}{F(b_j) - F(b_{j-1})} \times (b_j - b_{j-1}) + b_{j-1}$$

Cette méthode peut se traduire graphiquement ent 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

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 inferieur 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 class b_j telle que la proportion cumulée sur la classe $[b_{j-1}, b_j[$ est exatement 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 :

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

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

Utilisation des paramètres de tendance centrale

Robustesse

La médiane est plus **robuste** que la moyenne : une ou plusieurs données erronnées 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 :

1.2.2 Paramètres de dispersion

L'étendue

Soit x_{min} la plus petite observation et x_{max} la plus grande. On définie **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"*

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 mésure de dispersion.

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

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 nombr 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 le cas d'une variable continue pour laquelle on dispose seulement des **données regroupées en classes**, on peut faire un calcule 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 toujour positive ou nullle. Elle est nulle si et seulement si toutes les observation sonts identiques :

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

– L'untité de la variance est l'untié de X au carré.

L'ecart-type σ_X est défini par :

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

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

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

1.2.3 Changement de variable linéaire ou affine - Variable centrée réduite

Changement de variable linéaire ou affine

On considère une variable quantative X et on lui faire 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
$Y=aX$	$y_i=ax_i$	$\bar{y}=a\bar{x}$	$Var(Y)=a^2Var(X)$	$\sigma_Y\lvert a\rvert\sigma_X$
$Y=X+b$	$y_i=x_i+b$	$\bar{y}=\bar{x}+b$	$Var(Y)=Var(X)$	$\sigma_Y=\sigma_X$
$Y=aX+b$	$y_i=ax_i+b$	$\bar{y}=a\bar{x}+b$	$Var(Y)=a^2Var(X)$	$\sigma_Y=\lvert a\rvert\sigma_X$

Exemple 24.

Variable centrée réduite

On considére 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 ?*

1.2.4 Boîtes a 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 : $b_- = Q_1 - 1,5(Q_3 - Q_1)$ et $b_+ = Q_3 + 1,5(Q_3 - Q_1)$. 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 $[Q_1, m_{sup}]$
- 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.

Exemple 27. *Boîte a 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, saut si elles sont fournies en plus.

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

1.3 Laison entre deux variables

1.3.1 Laison linéaire entre deux variables quantitatives

On concidère le couple de variables (X, Y). On dispose d'observation de ce couple de variables sur un échantillon de taille n : pour chaque individu on connaît le couple d'observation (x_i, y_i).

Covariance

Définition 1. *On définit la* **covariance** *de X et de$Y$ par :*

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}[(x_i - \bar{x})(y_i - \bar{y})].$$

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

Remarque 8. *Lien avec la variance :* $Cov(X,Y) = Var(X)$

Remarque 9. *Formule pratique :*

$$Cov(X,Y) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i y_i\right) - \bar{x}\bar{y}.$$

Propriété 1. *Changement d'échelle : soient a, b, c, d des constantes réelles. On a*

$$Cov(aX + b, cY + d) = acCov(X, Y).$$

Proposition 1. *Expression de la variance d'une somme de variables :*

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y).$$

Proposition 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 $Var(X + aY) \geq 0$:

$$\begin{aligned} Var(X + aY) &= Var(X) + Var(aY) + 2Cov(X, aY) \\ &= Var(X) + a^2 Var(Y) + 2aCov(X, Y) \quad \text{par la propriéte 1} \\ &\geq 0 \end{aligned} \tag{1.1}$$

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 1.1 est nul. Dans ce cas, l'équation admet une racine double :

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

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

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

Dans le second cas,

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

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

Coefficient de corrélation

Définition 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 2. *il découle de la proposition 2 que*

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

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* $a < 0$.

Lorque 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 laison 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 laison entre X et Y.

Régression linéaire

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

$$y_i = ax_i + b + \epsilon_i, \quad i = 1, \ldots, n, \tag{1.2}$$

où a et b sont des coefficients réels. Le terme ϵ_i désigne un *bruit*, c'est à dire une pertubation supposée *petite*. Dans ce cour, on ne cherchera pas à donner un sens précis a 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 (1.2), au sens du critère des moindres carrés. On cherche

$$\min_{a,b} \sum_{i=1}^n (y_i - b - ax_i)^2. \tag{1.3}$$

La solution, qui s'obtient en annulant les dérivées partielles de la fonction de (a,b) qui est minimisée en (1.3), est

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

où $\bar{x}$ et $\bar{y}$ désigne les moyennes respective 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 :

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

Régression linéaire aprés transformation d'une variable

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

1.3.2 Liaison entre deux variables qualitatives

On observe une série statistique $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ composée de n couples d'observations d'un couple de variables qualitatives (X, Y). On suppose que X a I modalités notées $C_1, \ldots, C_I$ et Y a J modalitées notées $D_1, \ldots, D_J$. 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).

Table de contingence

Dans la table de contingence, on regroupe les effectifs n_{ij}. On peut compléter la table de contingence en ajoutant les totaux en et en colones.
On note $n_{i.} = n_{i1} + \cdots + n_{IJ} = \sum_{j=1}^{J} n_{ij}$ le total sur la ligne i de la table de contingence, $n_{.j} = n_{1j} + \cdots + n_{IJ} = \sum_{i=1}^{I} n_{ij}$ le total sur la colonne j de la table de contigence.

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

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

Remarque 10. *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.*

Distribution marginale

La distribution marginale de la variable X est la donnée des **effectifs marginaux** $n_{1.}, \ldots, n_{I.}$. 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	$\ldots$	C_I	Total
Effectif	$n_{1.}$	$\ldots$	$n_{I.}$	n
Proportion	$f_{1.} = n_{1.}/n$	$\ldots$	$f_{I.} = n_{I.}/n$	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

Distribution conditionlle

a) Profils-lignes

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

$Y_{\vert X=C_i}$	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 X	D_1	D_2	$\dots$	D_J	Total
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-colonnes, que l'on peut présenter dans le tableau suivant :

Y $X_{\vert Y}$	D_1	D_2	$\dots$	D_J
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-colonnes du couple de variables "Situation professionnelle" et "Situation vis-à-vis de l'immigration".*

Mesure de la liaison entre deux variables qualitatives

Compairaison 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,

$$n_{ij} = \frac{n_{i.}n_{.j}}{n}. \tag{1.4}$$

Réciproquement, si (1.4) a lieu, alors il y a indépendance entre X et Y.

preuve :

b) 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 3. *La **distance du** χ^2 observée sur la série statistique $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ 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 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 laison 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 4. *On appelle **contribution au** χ^2 du couple de modalités (C_i, D_j) et (X,Y) la quantité* $\frac{\left(n_{ij} - \frac{n_{i.}n_{.j}}{n}\right)^2}{\frac{n_{i.}n_{.j}}{n}}$.

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

Définition 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 6. *Le coefficient C de Cramer est défini par :*

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

Propriété 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 trés faible. Des valeurs proches de 1 signifient qu'il y a une forte liaison forte entre X et Y.*
- *Ce coefficient, qui varie entre 0 et 1, permet de comparer la laison 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.*

Représentation graphique

a) Distribution joint

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".*

1.3.3 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 : $C_1, \ldots, C_I$
- Y quantitative, discrète ou continue,avec donnée brutes ou regroupées en classes.

Exemple 38.

Classement des données et distributions marginales

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

On note $n_{1.}, \ldots, n_{I.}$ 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 :

$$\begin{array}{c}
(x_{11}, y_{11}), (x_{12}, y_{12}), \ldots, (x_{1n_{1.}}, y_{1n_{1.}}) = (C_1, y_{11}), (C_1, y_{12}), \ldots, (C_1, y_{1n_{1.}}) \\
(x_{21}, y_{21}), (x_{22}, y_{22}), \ldots, (x_{2n_{2.}}, y_{2n_{2.}}) = (C_2, y_{21}), (C_2, y_{22}), \ldots, (C_2, y_{2n_{2.}}) \\
\ldots \\
(x_{I1}, y_{I1}), (x_{I2}, y_{I2}), \ldots, (x_{In_{I.}}, y_{In_{I.}}) = (C_I, y_{I1}), (C_I, y_{I2}), \ldots, (C_I, y_{In_{I.}})
\end{array}$$

Chapitre 2

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.

2.1 Cardinal d'un ensemble fini

Définition 7. *Un ensemble E non vide est dit fini s'il existe un entier n et une bijection de $\{1, 2, \ldots, n\}$ sur E. Lorsqu'il existe, l'entier n est unique et est noté $Card(E)$. C'est le cardinal ou le nombre d'éléments de E*

Définition 8. *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 3.

1. *Si $\bar{A}$ est le complémentaire de A dans E alors*
$$Card(\bar{A}) = Card(E) - Card(A).$$
2. *$Card(A \cup B) = Card(A) + Card(B) - Card(A \cap B)$.*
3. *Si $A \cap B = \emptyset$ alors $Card(A \cup B) = Card(A) + Card(B)$*
4. *$Card(A \times B) = Card(A) \times Card(B)$*

2.2 Principes de comptage

2.2.1 Principe additif

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

- $A_i \cap A_j = \emptyset$ pour $i \neq j$
- $E = A_1 \cup A_2 \cup \ldots \cup A_n$.

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 $A_1, A_2, \ldots, A_n$ de cet ensemble, où les cardinaux des ensembles A_i sont plus faciles à déterminer. Il ne reste alors qu'à faire la somme des differents 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.*

2.2.2 Principe multiplicatif

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

$$n_1 \times n_2 \times \ldots \times n_p.$$

2.3 Arrangements

2.3.1 Arrangements avec répétition

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

Définition 9. *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 4. *Le nombre d'arrangements avec répétition de p objets parmi n est* n^p.

En effet, on a n possibilités pour chaque place, soit $n \times n \times \ldots \times n = n^p$ 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.*

2.3.2 Arrangements sans répétition

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

Définition 10. *Un arrangement de p éléments de E est une partie ordonnée de p éléments (distincts) de E. Cela revient à prendre p objets distincts dans E en tenant compte de l'ordre dans lequel on les choisit.*

Proposition 5. *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 :

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

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 $Card(E) = 20$. *Un tiercé correspond à 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.*

2.3.3 Permutation

Soit E un ensemble fini à n éléments.

Définition 11. *Une permutation de E est un arrangement des n éléments de E. Cela revient à prendre les n éléments de E en tenant compte de l'ordre dans lequel on les choisit.*

Proposition 6. *Le nombre de permutations d'un ensemble E à n éléments est*

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

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 :

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

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.*

2.4 Combinaisons

Définition 12. *Une combinaison de p éléments de E est une partie non ordonnée de E formée de p éléments. Cela revient à prendre p objets dans E sans tenir compte de l'ordre dans lequel on les choisit.*

Proposition 7. *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 simultanement p boules parmi n constitue une combinaison de p éléments parmi n éléments. Il y a C_n^p possibilités.*

2.4.1 Binôme de Newton

Proposition 8. *Soient a et b deux nombres réels et n un entier naturel non nul, alors :*

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

2.5 Quel modèle choisir ?

- 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 avec répétition**.
- 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 3

Espace probabilisé

3.1 Expérience aléatoire

Activité 1. On lance un dé équilibré à 6 faces numérotées 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 à 7 "
6. On considère les ensembles suivants :
   A est l'ensemble "obtenir un nombre au moins égal à 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 égal à 4".
   (b) Déterminer l'ensemble "obtenir un multiple de 2 et un nombre au moins égal à 4".
   (c) Déterminer l'ensemble "obtenir un multiple de 2 et le chiffre 5".

Définition 13. *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 à des résultats différents.*

Remarque 11. *Avant toute expérimentation, on peut décrire l'ensemble des résultats possibles de l'expérience aléatoire.*

Définition 14. *Soit $\mathcal{E}$ une expérience aléatoire. On appelle univers, et l'on note souvent Ω, l'ensemble des résultats possibles de $\mathcal{E}$. On supposera que Ω est fini et non vide.*

Définition 15. *On appelle évènement toute partie A de Ω.*

Remarque 12. *1. L'évènement $A = \Omega$ est appelé évènement certain. Il se réalise toujours.*
2. L'évènement $A = \emptyset$ est appelé évènement impossible. Il ne se réalise jamais.
3. L'évènement $A = \{\omega\}$ constitué d'un seul élément de Ω est appelé évènement élémentaire.

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

- l'union ; l'évènement $A \cup B$ se réalise si A se réalise ou B se réalise

- l'intersction ; $A \cap B$ 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 à nouveau sa couleur.*

1. *Déterminer l'univers des éventualités de cette expérience.*
2. *Citer un évènement élémentaire et un évènement non élémentaire*
3. *Soit A l'évènement : "les deux boules sont de même couleur", B l'évènement : "obtenir une boule bleue et une boule verte ", et C l'évènement : "obtenir d'abord une boule rouge"*
   (a) *Déterminer l'évènement contraire de A*
   (b) *Déterminer l'évènement : "A et B "; "A et C" puis l'évènement "A ou C", Les évènements A et B sont-ils incompatibles ?*

3.2 Probabilité

Définition 16. *Soit Ω l'univers associé à une expérience aléatoire. On appelle tribu, tout sous-ensemble $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ vérifiant*

1. $\Omega \in \mathcal{A}$
2. *Si $A \in \mathcal{A}$ alors $\bar{A} = \Omega \setminus A \in \mathcal{A}$*
3. *Pour toute famille fini ou denombrable $(A_i)_{i\in I}$ d'éléments de $\mathcal{A}$,*

$$\bigcup_{i\in I} A_i \in \mathcal{A}.$$

Définition 17. *Soient Ω l'univers associé à une expérience aléatoire et $\mathcal{A}$ une tribu sur Ω. Alors $(\Omega, \mathcal{A})$ est appelé espace pobabilisable. Un élément de $\mathcal{A}$ est appelé évènement.*

Définition 18. *Soit $(\Omega, \mathcal{A})$ un espace probabilisable. On appelle probabilité sur $(\Omega, \mathcal{A})$ une application $\mathbb{P} : \mathcal{A} \to [0, 1]$ telle que*

1. $\mathbb{P}(\Omega) = 1$
2. *Pour toute famille finie ou dénombrable $(A_i)_{i\in I}$ d'éléments de $\mathcal{A}$ deux à deux disjoints ($i \neq j \Rightarrow A_i \cap A_j = \emptyset$),*

$$\mathbb{P}(\bigcup_{i\in I} A_i) = \sum_{i\in I} \mathbb{P}(A_i).$$

Définition 19. *On appelle espace probabilisé le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ où Ω est un ensemble, $\mathcal{A}$ est une tribu sur Ω et $\mathbb{P}$ est une probabilité sur $\mathcal{A}$.*

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 ?

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'évènement "obtenir un nombre au moins égal à 4" et B="obtenir un multiple de 2 "

1. Calculer la probabilité $\mathbb{P}(A)$ de l'évènement A et $\mathbb{P}(B)$ de l'évènement 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'évènement C= "obtenir un nombre impair "
   (b) Calculer $\mathbb{P}(B) + \mathbb{P}(C)$

Remarque 13. *Hypothèse d'équiprobabilité :* ***les évènements élémentaires ont tous la même probabilité.***
Dans ce cas, nous avons pour tout évènement A, nous avons

$$A = \bigcup_{\omega \in A} \{\omega\}.$$

Alors, nous obtenons

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

Application 2. *Soit A et B deux évènements 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 $\mathbb{P}(\bar{A})$*

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 ?*

3.3 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 appelées "évènements" ; lorsque l'univers Ω est fini ou dénombrable, chaque partie de l'univers peut être consideré comme un évènement ; la famille des évènements est donc l'ensemble des parties de Ω noté $\mathcal{P}(\Omega)$
3. une probabilité $\mathbb{P}$.

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

3.4 Probabilités conditionnelles, indépendance

3.4.1 Probabilité conditionnelle

Activité 4. 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 évènements suivants : A="l'élève a 18 ans "
   B ="l'élève est une fille"
   C=" 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 $\frac{\mathbb{P}(A\cap B)}{\mathbb{P}(A)}$.

Théorème 1. *Soit une expérience aléatoire $\mathcal{E}$ d'univers Ω fini, $\mathbb{P}$ une probabilité sur Ω et B un évènement tel que $\mathbb{P}(B) \neq 0$. L'application*

$$\begin{aligned} \mathbb{P}_B : \mathcal{P} &\longrightarrow [0,1] \\ A &\longmapsto \mathbb{P}_B(A) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} \end{aligned}$$

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

Définition 20. *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.*

Définition 21 (Système complet d'événements)**.** *On dit qu'une famille $(B_k)_{1\leq k\leq n}$ est un système complet d'évènements lorsque :*

- $\forall\,(i,j) \in \lfloor 1,n\rfloor^2$, $(i \neq j \Rightarrow B_i \cap B_j =$ *(On dit alors que les B_k, $1\cdots n$, sont deux a deux disjoints ou incompatibles)*
- $\bigcup_{k=1}^{n} B_k = \Omega$.

Autrement dit, les B_k, $1,\cdots,n$, constituent une partition de Ω.

Théorème 2 (Probabilité totales)**.** *Soit $(\Omega,\mathcal{P}(\Omega),\mathbb{P})$ est espace probabilisé.*
Si une famille $(B_k)_{1\leq k\leq n}$ est un système complet d'évènements alors pour tout évènement 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 1 (Formule de Bayes)**.** *Soit $(\Omega,\mathcal{P}(\Omega),\mathbb{P})$ est espace probabilisé.*
Soit $(B_k)_{1\leq k\leq n}$ est un systeme complet d'évènements
Soit A un évènement tel que $\mathbb{P}(A) \neq 0$. Alors

$$\forall\, j) \in \lfloor 1,n\rfloor,\ \ \mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\sum_{k=1} n\mathbb{P}(A|B_k)\mathbb{P}(B_k)}$$

3.4.2 Indépendance

Définition 22. *Deux évènements 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 3. *Deux évènements 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 évènements : $A=$ "face apparaît au premier lancer " et $B=$"pile apparaît au deuxième lancer" sont-ils indépendants ?*
2. *Les évènements : $C=$"le même côté apparaît deux fois" et $D=$" le nombre d'apparition de " face" est diffèrent de deux " sont-ils indépendants ?*

Proposition 9. *Soit $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ est espace probabilisé.*

1. *Si A et B sont des événements indépendants, alors A et $\overline{B}$ aussi.*
2. *Si [(A et B sont indépendants), (A et C sont indépendants) et ($B \subset C$)] alors A et $C \setminus B$ sont indépendants.*
3. *Tout événement A est indépendant de l'événement impossible (resp. certain).*
4. *Si $(A_k)_{1 \leq k \leq n}$ est une famille d'evenements deux a deux incompatibles et si A est independant de chaque A_k, $1 \leq k \leq n$, alors A est indépendant de $\bigcup_{k1}^{n} A_k$.*

Chapitre 4

Variables aléatoires

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

4.1 Définitions

Définition 23. *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 14.

- *Si $(E, \mathcal{B}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, X est une variable aléatoire réelle.*
- *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 absolument continue.*

Définition 24. *Soit X une variable aléatoire définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans l'espace probabilisable $(E, \mathcal{B})$. On appelle loi de probabilité de X, la loi de 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 25. *La fonction de répartition de la v.a.r X est définie par :*

$$\begin{aligned} F : \mathbb{R} &\longrightarrow [0, 1] \\ x &\longrightarrow F(x) = \mathbb{P}(X \leq x). \end{aligned}$$

Propriété 5.

a) F est une fonction non décroissante

b) $\lim\limits_{x \to -\infty} F(x) = 0$ et $\lim\limits_{x \to +\infty} F(x) = 1$.

c) F est continue à droite.

Définition 26. *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 15. *X est une variable aléatoire 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.

4.2 Moments d'une variable aléatoire réelle

Théorème 4. *(Formule de transfert) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire réelle définie sur Ω, et soit $g : \mathbb{R} \to \mathbb{R}$.*

1. *Si X est discrète :*
$$\mathbb{E}(g(X)) = \sum_{x \in X(\Omega)} g(x)\mathbb{P}(X = x).$$

2. *Si X est une variable réelle absolument continue de densité f*
$$\mathbb{E}(g(X)) = \int_{\mathbb{R}} g(x)f(x)dx.$$

Proposition 10. *(Inegalité de Markov)*
Soient X une v.a.r telle que $\mathbb{E}(|X|) < +\infty$. Alors pour tout $c \in \mathbb{R}_+$

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

Proposition 11. *(Inegalité de Tchebychev)*
Soient X une v.a.r $\mathbb{E}(X^2) < +\infty$ est définie. Alors pour tout $c > 0$

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

4.3 Lois usuelles

4.3.1 Lois discrètes

Loi uniforme sur $\{1, \ldots, N\}$, $N \in \mathbb{N}^*$, $\mathcal{U}_N$

$$X \backsim \mathcal{U}_N \iff \begin{cases} X(\Omega) = \{1, \ldots, N\} \\ P(X = k) = \frac{1}{N}, \quad \forall k \in X(\Omega) \end{cases}$$

$$E(X) = \frac{n+1}{2}$$

et

$$var(X) = \frac{n^2 - 1}{12}.$$

Exemple 48. *Soit X le résultat d'un lancer de dé non truqué : alors $\forall i \in X(\Omega) = \{1, 2, 3, 4, 5, 6\}$, $P(X = i) = \frac{1}{6}$; X suit la loi uniforme $\mathcal{U}_6$.*

Loi de Bernouilli $\mathcal{B}(1,p)$ $p \in]0,1[$

$$X \backsim \mathcal{B}(1,p) \Longleftrightarrow \begin{cases} X(\Omega) = \{0,1\} \\ P(X=1) = p, \quad P(X=0) = 1-p \end{cases}$$

$$E(X) = p$$

$$var(X) = p(1-p).$$

La fonction caractéristique est

$$\phi_X(t) = (1 - p + pe^{it}).$$

Cette variable modélise l'issue d'une expérience où l'on ne s'intéresse qu'au "succès" ou à l'"echec" de l'expérience.

Exemple 49. *Lancer d'une pièce de monnaie (pile ou face), qualité d'un produit (bon ou defectueux), sondage elctoral (pour ou contre).*

Loi binomiale $\mathcal{B}(n,p)$ $n \in \mathbb{N}^*$ $p \in]0,1[$

On réalise n fois successivement et d'une manière indépendante une expérience aléatoire de Bernouilli. La variable aléatoire égale au nombre de succès obtenus au cours des n épreuves suit la loi binomiale $\mathcal{B}(n,p)$.

$$X \backsim \mathcal{B}(n,p) \Longleftrightarrow \begin{cases} X(\Omega) = \{0,\dots,n\} \\ P(X=k) = C_n^k p^k (1-p)^{n-k}, \quad \forall k \in X(\Omega) \end{cases}$$

$$E(X) = np$$

$$var(X) = np(1-p).$$

La fonction caractéristique est

$$\phi_X(t) = (1 - p + pe^{it})^n.$$

Cette loi modélise une succession de "succès" et d'"échecs", p étant la probabilité du succès.

Propriété 6. *Si $X_1 \backsim \mathcal{B}(n_1,p)$ et $X_2 \backsim \mathcal{B}(n_2,p)$ avec X_1 et X_2 indépendantes alors $X_1+X_2 \backsim \mathcal{B}(n_1+n_2,p)$.*

Loi hypergéométrique $\mathcal{H}(N,n,p)$

Soit une population de N individus parmi lesquels une proportion p (donc Np individus) possède un caractère. Il s'agit par exemple de la proportion des individus qui souffrent d'une maladie, ou de la proportion des pièces défectueuses dans un grand lot de fabrication. On prélève un échantillon de n individus parmi cette population (le tirage pouvant s'effectuer d'un seul coup ou au fur et à mesure mais sans remise). On note X la variable aléatoire égale au nombre d'individus de l'échantillon possédant le caractère envisagé. La loi de X est appelée loi hypergéométrique de paramètre N, n, p et notée $\mathcal{H}(N,n,p)$:

$$X \backsim \mathcal{H}(N,n,p) \Longleftrightarrow \begin{cases} X(\Omega) = \{max(0, n-(1-p)N), min(Np,n)\} \\ P(X=k) = \frac{C_{Np}^k C_{(1-p)N}^{n-k}}{C_N^n}, \quad \forall k \in X(\Omega) \end{cases}.$$

$$E(X) = np.$$

Propriété 7. *Quand $N \to +\infty$ avec n et p fixés, alors $\mathcal{H}(N,n,p)$ converge en loi vers $\mathcal{B}(n,p)$ (En pratique $\frac{n}{N} < 1$).*

Loi géométrique $\mathcal{G}(p)$, $p \in]0,1[$ C'est la loi du nombre d'essais (ou épreuves) nécessaires pour faire apparaître un évènement de probabilité p. C'est le cas de nombre d'examens necessaires pour réussir une épreuve en supposant que la probabilité de réussir à chaque passage de l'examen est de type p et que les résultats sont indépendants d'un examen vers un autre. Soit la variable X égale le nombre d'essais avant d'obtenir le premier succès :

$$X \backsim \mathcal{G}(p) \Longleftrightarrow \begin{cases} X(\Omega) = \mathbb{N}^* \\ P(X=k) = p(1-p)^{k-1}, \ \ \forall k \in X(\Omega) \end{cases}.$$

$$E(X) = \frac{1}{p}$$

$$var(X) = \frac{1-p}{p^2}.$$

Exemple 50. *On effectue des lancers indépendants d'une pièce, dont la probabilité d'obtenir face est p, jusqu'à l'obtention d'un "face". On note X la v.a.r égale au nombre de lancers nécessaires. On dit également que X est le temps d'attente du premier "face".*

Loi de Poisson $\mathcal{P}(\lambda)$ $\lambda > 0$

Pour modéliser des phénomènes rares (nombre d'accidents d'avion, nombre d'appels téléphoniques pendant un certain temps, nombre de pièces défectueuses dans une commande importante, nombre de suicides par an dans un pays donné...), on utilise la loi de Poisson (de paramètre $\lambda > 0$) :

$$X \backsim \mathcal{P}(\lambda) \Longleftrightarrow \begin{cases} X(\Omega) = \mathbb{N} \\ P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}, \ \ \forall k \in X(\Omega) \end{cases}$$

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

La fonction caractéristique est

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

Propriété 8. *Si X_1 et X_2 sont deux variables aléatoires de Poisson $\mathcal{P}(\lambda_1)$, $\mathcal{P}(\lambda_2)$ respectivement, indépendantes, alors $X_1 + X_2 \backsim \mathcal{P}(\lambda_1 + \lambda_2)$. (Ceci est vrai pour une somme finie quelconque de v.a de Poisson indépendantes)*

Propriété 9. *Soit X une variable aléatoire suivant une loi de Poisson $\mathcal{P}(\lambda)$. Alors la variable aléatoire $\dfrac{X-\lambda}{\sqrt{\lambda}}$ converge en loi vers la loi normale centrée réduite $\mathcal{N}(0,1)$, lorsque λ tend vers l'infini.*

Propriété 10. *$X \backsim \mathcal{B}(n,p)$. Quand $n \to +\infty$ et $p \to 0$ tel que $np \to \lambda$. Alors X converge en loi vers $\mathcal{P}(\lambda)$.*

4.3.2 Lois à densité

Loi uniforme

On dit que la v.a.r continue suit une loi uniforme sur l'intervalle $[a,b]$ si sa fonction densité est

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a,b] \\ 0 & \text{sinon} \end{cases}$$

$$E(X) = \frac{b+a}{2}$$

$$var(X) = \frac{(b-a)^2}{2}.$$

Si $a = -b$, la fonction caractéristique est

$$\phi_X(t) = \frac{\sin at}{at}.$$

Loi normale On dit que X suit une loi normale de paramètre (m, σ^2) avec $m \in \mathbb{R}$, $\sigma^2 \in \_+^*$ si sa densité de probabilité est

$$f_X(x)\frac{1}{\sigma\sqrt{2\pi}}\exp\{-\frac{1}{2}(\frac{x-m}{\sigma})^2\}, \quad x \in \mathbb{R}.$$

La fonction caractéristique est

$$\phi_X(t) = \exp\left\{itm - \frac{t^2\sigma^2}{2}\right\}.$$

Propriété 11. *(i) $X \backsim \mathcal{N}(m, \sigma^2)$ alors $E(X) = m$ et $var(X) = \sigma^2$.*
(ii) $X \backsim \mathcal{N}(m, \sigma^2)$ si et seulement si la v.a.r $\frac{X-m}{\sigma} \backsim \mathcal{N}(0, 1)$.

Loi gamma $\gamma(a, \rho)$, $a > 0$, $\rho > 0$ X suit une loi gamma de paramètre a et ρ si sa densité

$$f(x) = \frac{\rho^a}{\Gamma(a)}e^{-\rho x}x^{a-1}\mathbb{1}_{\mathbb{R}_+}(x)$$

où

$$\Gamma(a) = \int_0^\infty e^{-x}x^{a-1}dx.$$

$$E(X) = \frac{a}{\rho}$$

$$var(X) = \frac{a}{\rho^2}.$$

La fonction caractéristique est

$$\phi_X(t) = \frac{1}{(1-\frac{it}{\rho})^a}$$

Proposition 12. *Soient X et Y deux variables aléatoires indépendantes suivant respectivement $\gamma(a_1, \rho)$ et $\gamma(a_2, \rho)$. Alors $X + Y$ suit une loi $\gamma(a_1 + a_2, \rho)$.*

Loi exponentielle Si $a = 1$ la loi $\gamma(1, \rho) = \mathcal{E}(\rho)$ est appelé loi exponentielle de paramètre $\rho > 0$ et a pour densité de probabilité

$$f(x) = \rho e^{-\rho x}\mathbb{1}_{\mathbb{R}_+}(x)$$

La fonction caractéristique est

$$\phi_X(t) = \frac{\lambda}{\lambda - it}.$$

Cette loi de probabilité est fortement utilisée pour décrire les durées de vie (par exemple la durée de vie des transistors electroniques).

4.3.3 Lois dérivées

Loi du Chi-deux à n dégrés de liberté $\chi^2(n)$, où $n \in \mathbb{N}^*$

Si $X_1, \ldots, X_n$ sont des variables aléatoires indépendantes identiquement distribuées telle que $X_i \backsim \mathcal{N}(0,1)$, alors

$$Y = \sum_{i=1}^{n} X_i^2 \backsim \chi^2(n)$$

a pour densité

$$f_Y(y) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} e^{-\frac{x}{2}} x^{\frac{n}{2}-1} \mathbb{1}_{\mathbb{R}_+}(x).$$

Théorème 5. *Si X est un vecteur gaussien de dimension n, $\mathcal{N}(m,\Sigma)$, la variable aléatoire $Y = (X - m)^t\Sigma^{-1}(X-m)$ suit une loi du $\chi^2(n)$*

Théorème 6. *Soient X et Y deux variables aléatoires indépendante suivant respectivement deux lois du $\chi^2(n)$ et $\chi^2(m)$. La variable aléatoire $X+Y$ suit une loi $\chi^2(n+m)$.*

Propriété 12. *Si Y suit une loi $\chi^2(n)$*

1. *$\frac{Y-n}{\sqrt{2n}}$ converge en loi vers la loi $\mathcal{N}(0,1)$ quand n tend vers l'infini.*
2. *$\sqrt{2Y} - \sqrt{2n-1}$ converge en loi vers la loi $\mathcal{N}(0,1)$ quand n tend vers l'infini.*

Loi de Fisher-Snedecor $F_{n,m}$

Soient X et Y deux variable aléatoires indépendantes suivant respectivement les lois $\chi^2(n)$ et $\chi^2(m)$. La variable aléatoire $F_{n,m} = \frac{X/n}{Y/m}$ suit une loi de Fisher-Snedecor à n et m dégrés de liberté.

Loi de Student

Soient X et Y deux variables aléatoires indépendantes suivant respectivement $\mathcal{N}(0,1)$ et $\chi^2(n)$.
On appelle loi de Student n dégrés de liberté la loi suivie par le rapport

$$T_n = \frac{X}{\sqrt{\frac{Y}{n}}}.$$

$$E(T_n) = 0$$
$$var(T_n) = \frac{n}{n-2}, \quad n > 2.$$

Théorème 7. *T_n converge en loi vers $\mathcal{N}(0,1)$ lorsque n tend vers l'infini.*