PROBABILITÉS ET STATISTIQUES

M. Le Gonidec

Table des matières

Première partie

PROBABILITÉS

Chapitre 1

Dénombrement

1.1 Ensembles finis

Définitions 1.1. Un ensemble Ω non vide est un **ensemble fini** s'il existe un entier naturel n tel qu'il existe une bijection de $\{1, \ldots, n\}$ dans Ω. Dans ce cas, on dit que le **cardinal** de E est n et on écrit $|\Omega| = n$ ou $\mathrm{Card}(\Omega) = n$ ou. On dit aussi que Ω est un n-ensemble.

Le cardinal de Ω est le nombre d'éléments de Ω.

Par convention, l'ensemble vide est fini et $|\emptyset| = 0$.

Un ensemble E est **dénombrable** s'il existe une bijection de $\mathbb{N}$ dans E.

Propriétés 1.2. *Soit Ω un ensemble fini (resp. dénombrable).*

1. *Si A est un ensemble tel qu'il existe une bijection de Ω vers A. Alors A est un ensemble fini et $|\Omega| = |A|$ (resp. dénombrable),*
2. *Tout sous-ensemble A de Ω est un ensemble fini (resp. dénombrable), et $|A| \leq |\Omega|$,*

Remarques : Il n'y a que dans les cadre des ensembles finis que $A \subsetneq \Omega$ implique $|A| < |\Omega|$.

Les ensembles $\mathbb{N}$, $\mathbb{Z}$ et $\mathbb{Q}$ sont dénombrables. $\mathbb{R}$ ne l'est pas ($\mathbb{R}$ comme tout souus intervalle non vide de $\mathbb{R}$ a la puissance du continu).

1.2 Méthodes de base pour dénombrer...

Proposition 1.3. Cardinal d'une union disjointe

Quels que soient les ensembles finis, deux à deux disjoints, alors $\bigcup_{k=1}^{n} A_k$ est fini et

$$\left| \bigcup_{k=1}^{n} A_k \right| = \sum_{k=1}^{n} |A_k| .$$

Corollaire 1.4. *Quel que soit le sous-ensemble A du n-ensemble E, on a $\left|\overline{A}\right| = n - |A|$ où $\overline{A}$ désigne le complémentaire de A dans E.*

Proposition 1.5. Cardinal d'une union quelconque

Cas de 2 ensembles : *Quels que soient les ensembles finis A et B, leur réunion $A \cup B$ et leur intersection sont des ensembles finis, et leurs cardinaux vérifient*

$$|A \cup B| = |A| + |B| - |A \cap B| .$$

Cas général : Formule de Poincaré

Quelle que soit la famille finie d'ensembles finis A_1, A_2, ..., A_n, l'ensemble $\bigcup_{k=1}^{n} A_k$ est fini et

$$\left| \bigcup_{k=1}^{n} A_k \right| = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} |A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}| .$$

Proposition 1.6. Cardinal d'une produit cartésien d'ensembles *Soient A et B deux ensembles finis. Le cardinal du produit cartésien $A \times B$ est*

$$|A \times B| = |A| \cdot |B| .$$

La généralisation de ce théorème au produit cartésien de k ensembles finis A_1, A_2, ..., A_k donne :

$$|A_1 \times A_2 \times \ldots \times A_k| = \prod_{i=1}^{k} |A_i| .$$

Corollaire 1.7. Nombre de p-uplet d'un ensemble fini
le nombre de p-uplets d'un n-ensemble E est n^p.

(le nombre de p-uplets d'élements de E est le cardinal de $E^p = E \times E \times \cdots \times E$ p fois.)

Méthode 1 : il peut arriver que les éléments qui sont à dénombrer dans un arbre des choix ne soient pas les éléments d'un produit cartésien $A \times B$ mais plutôt les éléments d'une union disjointe de la forme $\bigcup_{i=1}^{n_A} (\{a_i\} \times B_i)$ où les ensembles B_i dépendent du choix de l'élément a_i, alors

$$\left| \bigcup_{i=1}^{n_A} (\{a_i\} \times B_i) \right| = \sum_{i=1}^{n_A} |\{a_i\} \times B_i| = \sum_{i=1}^{n_A} |B_i| .$$

★ **Exercice 1.8.** Soit E un ensemble fini de cardinal n. Donner le nombre de couples (A, B) de parties de E telles que $A \subset B$. Vérifier pour $n = 2$.

Méthode 2 : Lorsqu'un ensemble à dénombrer ne s'exprime pas simplement en termes d'union disjointe ou quelconque (somme ou formule de Poincaré, choix parallèles) ou de produit cartésien (produit, choix successifs), il peut être efficace de dénombrer un autre ensemble lié au premier puis d'adapter le résultat obtenu.

En particulier, lorsqu'il ne faut pas tenir compte de l'ordre dans lequel les choix successifs interviennent, on peut introduire un ordre artificiel ou, lorsqu'il y a des objets qui ne doivent pas être distingués, on peut cependant les considérer comme distincts dans un premier temps.

★ **Exercice 1.9.** En utilisant cette méthode, donner :

1. le nombre de poignées de mains échangées lorsque n personnes se rencontrent,
2. le nombre de configurations différentes lorsque 4 convives s'assoient autour d'une table ronde (à rotation près).

1.3 Choisir p objets parmi n : arrangements et combinaisons...

Soit $(n, p) \in \mathbb{N}^* \times \mathbb{N}^*$.

Considérons un n-ensemble Ω avec lequel on constitue des p-listes, c'est à dire des « listes » de p éléments choisis dans Ω. Les règles de formation de ces listes peuvent varier :

- choix ordonné : deux p-listes sont identiques si elles sont formés des mêmes éléments, dans le même ordre,
- choix non ordonné : deux p-listes sont identiques si elles sont formés avec mêmes éléments (multiplicité prise en compte),
- choix avec répétition possible : un élément de E peut apparaitre plusieurs fois dans la liste.
- choix sans répétition : un élément de E ne peut apparaitre qu'une seule fois dans la liste.

Il y a donc quatre types de p-listes, étudiées dans les quatre sections ci-dessous.

1.3.1 Nombre de listes ordonnées avec répétition possible

Définition 1.10. Soit E un n-ensemble. On appelle p-liste ordonnée avec répétition possible un élément de E^p.

Proposition 1.11. *Soit E un n-ensemble.*
Le nombre de p-listes ordonnées avec répétition possible est n^p.

(Ces listes sont en fait les p-uplets d'éléments de E et leur nombre est une conséquence du théorème du cardinal d'un produit cartésien.)

★ **Exercice 1.12.** Combien de mots de longueur 3 peut-on écrire avec un alphabet de 26 lettres ?

Remarque 1.13. Soit un p-ensemble F et un n-ensemble E. Si on désigne par E^F l'ensemble des applications de F dans E, il existe une correspondance bijective entre chacune des p-listes ordonnées avec répétition possible et un des graphes d'une application $f : F \to E$, d'où :

Le nombre d'applications d'un p-ensemble F dans un n-ensemble E est :

$$|E^F| = |E|^{|F|} = n^p.$$

On peut notamment se servir de cela pour montrer que le nombre de parties d'un p-ensemble est $|\{0, 1\}|^p = 2^p$.

★ **Exercice 1.14.** Utiliser la remarque précédente pour montrer que le nombre de parties d'un p-ensemble est 2^p.

★ **Exercice 1.15.** Determiner le nombre de couples (A, B) de parties d'un n-ensemble E tels que $A \subset B \subset E$.

1.3.2 Nombre de listes ordonnées sans répétition

Définition 1.16. Soit E un n-ensemble.

On appelle p-liste ordonnée sans répétition possible un éléments $(e_1, \ldots, e_p)$ de E^p pour lequel les e_i sont distincts ($e_i = e_j \Longrightarrow i = j$). On parle souvent d'**arrangement de p objets parmi n.**

Proposition 1.17. *Soit E un n-ensemble et soit $0 \leq p \leq n$.*

Le nombre de p-listes ordonnées sans répétition d'éléments de E est

$$A_n^p = n(n-1)\cdots(n-p+1) = \frac{n!}{(n-p)!}$$

Si $p > n$, on pose par convention $A_n^p = 0$.

Remarque : Il existe une correspondance bijective entre chacune des listes ci-dessus et une des applications injectives f du p-ensemble F dans le n-ensemble E, d'où :

Proposition 1.18. *Le nombre d'injections d'un p-ensemble F dans un n-ensemble E est A_n^p si $p \leq n$, 0 sinon.*

Exemples 1.19.

1. Nombre de mots de p lettres distinctes choisies dans un alphabet de n lettres.
2. Nombre de tiercés possibles dans une course de chevaux.

1.3.3 Nombre de listes non ordonnées sans répétition

Définition 1.20. Soit E un n-ensemble.

Une p-liste non ordonnée sans répétition de E est un sous ensemble de E^p à p éléments. On parle souvent de **combinaison de p objets parmi n.**

Proposition 1.21. *Soit E un n-ensemble et soit $0 \leq p \leq n$.*

Le nombre de p-listes non ordonnées sans répétition de E est

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

Si $p > n$, on pose par convention $\binom{n}{p} = 0$.

Exemples 1.22.

1. Nombre de mains de 5 cartes choisies dans un jeu de 32.
2. Nombre d'échantillons de taille p prélevés dans une population de taille n.

Remarques 1.23.

1. $\binom{n}{p}$ se lit "p parmi n".
2. L'ancienne notation de $\binom{n}{p}$ était C_n^p. On la trouve encore dans d'anciens ouvrages.
3. Pour trouver le nombre de combinaisons, il suffit de diviser le nombre d'arrangements par le nombre de permutations de chaque p-liste.
4. Un p-ensemble est une p-liste non ordonnée sans répétition. D'où le corollaire immédiat :

Proposition 1.24. *Le nombre de sous-ensembles à p éléments d'un n-ensemble E est $\binom{n}{p}$ si $p \leqslant n$, 0 sinon.*

1.3.4 Nombre de listes non ordonnées avec répétition possible

Définition 1.25. Soit E un n-ensemble.

Une p-liste non ordonnée avec répétition possible est appellée **combinaison avec repétition de p objets parmi n**

Proposition 1.26. *Le nombre de p-listes non ordonnées avec répétition possible, les éléments étant choisis dans un n-ensemble E est*

$$\binom{n+p-1}{p} = \frac{(n+p-1)!}{p!(n-1)!}.$$

Exemple 1.27. Nombre de manière de ranger p paires de chaussettes (identiques) dans n tiroirs.

★ **Exercice 1.28.** Démontrer la Proposition précédente.

1.3.5 Résumé

Nombre de façons de choisir p objets parmi n :

	Avec répétition possible	Sans répétition $(p \leq n)$
avec ordre	n^p	$A_n^p = \frac{n!}{(n-p)!}$
sans ordre	$\binom{n+p-1}{p} = \frac{(n+p-1)!}{p!(n-1)!}$	$\binom{n}{p} = \frac{A_n^p}{p!} = \frac{n!}{p!(n-p)!}$

1.4 Permutations et anagrammes

Les problèmes de permutations et d'anagrammes sont en fait des cas particuliers de n-listes parmi n objets, à la différence que certains objets parmis les n disponibles peuvent éventuellement être identiques (cas des anagrammes).

Définition 1.29. Soit E un n-ensemble, on appelle **permutation** de E, toute bijection de E dans lui-même.

Proposition 1.30. *Soit E un n-ensemble. Le nombre de permutations de E est $n!$.*

Définitions 1.31. On appelle **alphabet**, un ensemble fini de signes quelconques appelés **lettres**.

On appelle **mot**, toute liste ordonnée, et avec répétition possible, d'éléments de cet alphabet.

On appelle **anagramme** toute permutation des lettres d'un mot.

Remarques 1.32. La définition d'alphabet donnée ici est une généralisation qui recouvre le sens de ces mots employés couramment en français. Par exemple, $\{\circ, \star\}$ est l'alphabet constitué des lettres $\circ$ et $\star$. Un alphabet est constitué de signes distincts (c'est un ensemble !) et non ordonnés. Cependant, comme pour notre alphabet de vingt-six lettres habituel, on peut employer un "ordre conventionnel" pour *ranger* les mots. Mais deux alphabets constitués des mêmes signes donnés dans un ordre différent sont identiques. $\circ \star \circ$ et $\circ \circ \star$ sont deux mots différents construits avec l'alphabet précédent et s'il on "ordonne" l'alphabet en posant $\circ \leq \star$, alors dans le "dictionnaire", $\circ \circ \star$ se trouvera avant $\circ \star \circ$.

Pour la notion d'anagramme donnée ici, par exemple, $\circ \star \circ$ et $\circ \circ \star$ sont deux des anagrammes de $\circ \star \circ$. Contrairement au sens qu'a anagramme en français (*niche* est un anagramme de *chien*), dans cette généralisation, un anagramme n'a pas nécessairement un sens et le mot lui-même fait partie de ses anagrammes (puisque l'application identique est évidemment une permutation).

Proposition 1.33. Nombre d'anagrammes formés de p lettres distinctes

Soit un alphabet de n lettres et un sous-ensemble de p lettres données de cet alphabet.

Le nombre d'anagrammes de taille p formés de ces p lettres distinctes est $\begin{cases} 0 & \text{si } p > n, \\ p! & \text{si } p \leq n. \end{cases}$

Ce résultat est évident : Une fois les p lettres différentes choisies, c'est le nombre de permutations possibles entre ces p lettres.

Proposition 1.34. Nombre d'anagrammes d'un mot donné

Soit $E = \{e_1, e_2, \ldots, e_n\}$ un alphabet de n lettres.

Soit $m = m_1 m_2 \ldots m_p$ un mot de p lettres sur l'alphabet E. Soit α_i le nombre d'occurrence de la lettre e_i dans m.

Le nombre d'anagrammes du mot m est donné par

$$\frac{p!}{\prod_{i=1}^{n} \alpha_i!} = \frac{(\sum_{i=1}^{n} \alpha_i)!}{\prod_{i=1}^{n} \alpha_i!}$$

Remarque : lorsque $p = n$ et les $\alpha_i = 1$, alors on retrouve bien le nombre de permutation d'un n-ensemble.

★ **Exercice 1.35.** Démontrer la Proposition précédente.

Chapitre 2

Mesures de probabilités et variables aléatoires

2.1 Mesures de Probabilité

2.1.1 Un soupçon de théorie de la mesure, et après, promis, on en fait plus...

Soit Ω un ensemble, fini ou infini, discret ou continu.

Définition 2.1. Une **tribu** (ou **σ-algèbre**) est une partie $\mathcal{T} \subset \mathcal{P}(\Omega)$ qui verifie :
- $\Omega \in \mathcal{T}$,
- $\mathcal{T}$ est stable par passage au complémentaire,
- $\mathcal{T}$ est stable par union dénombrable.

Les élément d'une tribu seront appelés **évenements**.

La **tribu engendrée par une classe** $\mathcal{C} \subset \mathcal{P}(\Omega)$, notée $\sigma(\mathcal{C})$ est la plus petite tribu contenant tous les éléments de $\mathcal{C}$.

Dans ce cours, on ne fera appel qu'à 2 tribus simples :
- Si Ω est dénombrable, la tribu discrète : $\mathcal{T} = \mathcal{P}(\Omega)$, i.e. l'ensemble des parties de Ω, engendrée par les singletons ;
- Si $\Omega = \mathbb{R}$, la tribu borélienne de $\mathbb{R}$: $\mathcal{B}(\mathbb{R})$, i.e. la tribu qui peut être engendrée soit par :
  - la famille des intervalles $\{]a, b[,\ a < b \in \mathbb{R}\}$,
  - la famille des intervalles $\{]-\infty, t],\ t \in \mathbb{R}\}$,
  - la famille des intervalles $\{]-\infty, t[,\ t \in \mathbb{R}\}$,
  - la famille des intervalles $\{]t, +\infty[,\ t \in \mathbb{R}\}$,
  - la famille des intervalles $\{[t, +\infty[,\ t \in \mathbb{R}\}$

Définition 2.2. Un espace Ω muni d'une tribu $\mathcal{T}$ est appelé **espace mesurable** ou **espace probabilisable**, selon le contexte.

C'est sur ces espaces qu'on va pouvoir définir des mesures, et en particulier des mesures de probabilité...

2.1.2 Voculabulaire ensembliste Vs Vocabulaire probabiliste

Soit $(\Omega, \mathcal{T})$ un espace probabilisable. Ci-dessous, on résume le vocabulaire probabiliste usuel, avec son equivalent ensembliste.

Vocabulaire ensembliste	Notation	Vocabulaire probabiliste
Ensemble vide	$\emptyset$	Évènement impossible
Univers	Ω	Événement certain
Deux sous-ensembles de $\mathcal{T}$	A, B	Deux événements
A est un singleton	$A = \{\omega\}$	A est un événement élémentaire
A inclus dans B	$A \subset B$	A entraîne B
Complémentaire de A	$\overline{A}$ ou ${}^cA$	Contraire de A
Intersection de A et B	$A \cap B$	A et B (conjonction)
Réunion de A et B	$A \cup B$	A ou B (disjonction inclusive)
Complémentaire de B dans A	$A \setminus B$	A et non B
Différence symétrique de A et B	$A \triangle B$	Soit A, soit B (disjonction exclusive)
A et B sont disjoints	$A \cap B = \emptyset$	A et B sont incompatibles
Partition de Ω	$(A_i)_{i \in I}$	Système complet d'événements

2.1.3 Mesures de probabilité

Définition 2.3. Une **mesure de probabilité** sur un espace mesurable $(\Omega, \mathcal{T})$ est une mesure positive $\mathbb{P}$, de masse totale égale à 1, i.e. une fonction $\mathbb{P} : \mathcal{T} \to \mathbb{R}_+$ qui satisfait les 3 conditions suivantes :

- $\mathbb{P}(\emptyset) = 0$,
- $\forall (A_i)_{i \in I} \in \mathcal{T}$ disjoints, $\mathbb{P}\left(\cup_{i \in I} A_i\right) = \sum_{i \in I} \mathbb{P}(A_i)$,
- $\mathbb{P}(\Omega) = 1$.

Un triplet $(\Omega, \mathcal{T}, \mathbb{P})$ est appelé **espace de probabilités**.

Proposition 2.4. Quelques propriétés immédiates d'une mesure de probabilité :
Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace de probabilité et A, B et $(A_n)_{n \geq 0}$ des evenements de $\mathcal{T}$. On a :

1. $\mathbb{P}({}^cA) = 1 - \mathbb{P}(A)$,
2. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$,
3. $A \subset B \Longrightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$,
4. $\mathbb{P}\left(\cup_{n \geq 0} An\right) \leq \sum_{n \geq 0} \mathbb{P}(A_n)$,
5. *si* $A_n \xrightarrow[n \to \infty]{} A$ *avec* $A_n \subset A_{n+1}$, *alors* $\mathbb{P}(A_n) \xrightarrow[n \to \infty]{} \mathbb{P}(A)$,
6. *si* $A_n \xrightarrow[n \to \infty]{} A$ *avec* $A_{n+1} \subset A_n$, *alors* $\mathbb{P}(A_n) \xrightarrow[n \to \infty]{} \mathbb{P}(A)$.

★ **Exercice 2.5.** Démontrer à l'aide de la définition, les propriétés énoncées dans la Proposition précédente.

Proposition 2.6. *Une mesure de probabilité sur $(\Omega, \mathcal{T})$ est caractérisée par les valeurs qu'elle prend sur une famille d'évènements génératrice de la tribu $\mathcal{T}$.*

Plus précisement, si $(\Omega, \mathcal{T})$ est un espace probabilisable et $\mathbb{P}_1$ et $\mathbb{P}_2$ sont deux mesures de probabilité telles que $\mathbb{P}_1(A) = \mathbb{P}_2(A)$ pour tout élement d'une classe $\mathcal{C}$ qui engendre $\mathcal{T}$, alors $\mathbb{P}_1(A) = \mathbb{P}_2(A)$ pour tout élement de $\mathcal{T}$.

★ Exercice 2.7. De l'importance de la tribu...

On lance un dé. Soit $\Omega = \{1, 2, 3, 4, 5, 6\}$ l'ensemble des valeurs possibles du résultat du lancé.

1. On muni Ω de la tribu discrète $\mathcal{T} = \mathcal{P}(\Omega)$ et de la probabilité uniforme $\mathbb{P}$. Calculer la probabilité de le résultat du lancer soit 1 ou 2 puis la probabilité que le résultat du lancer soit un nombre premier.
2. On définit la classe d'évènement $\mathcal{C} = \{\{1\}, \{2\}\}$. Soit $\tilde{\mathbb{P}}$ une mesure de probabilité définie sur $\mathcal{C}$ par $\tilde{\mathbb{P}}(\{1\}) = \frac{1}{6}$ et $\tilde{\mathbb{P}}(\{2\}) = \frac{1}{6}$.

   Sur quels ensembles de $\mathcal{T}$ peut-on définir $\tilde{\mathbb{P}}$? Pour chacun de ces ensembles, donner la valeur de leur mesure de probabilité.

   Peut-on calculer la probabilité de le résultat du lancer soit 1 ou 2 ? Peut-on calculer la probabilité de le résultat du lancer soit un nombre premier ?

Moralité 1 : Pour pouvoir calculer la probabilité d'un ensemble, il faut que ce soit un élément de la tribu sur laquelle est définie la probabilité !

★ Exercice 2.8. De l'importance de la tribu... la suite

On lance deux dés. Soit $\Omega = \{1, 2, 3, 4, 5, 6\}$ l'ensemble des valeurs possibles du résultat de chacun des deux lancers.

On cherche à comprendre le comportement aléatoire de la somme S des résultats des 2 dés. S est à valeurs dans $\Omega' = \{2, 3, \ldots, 12\}$ muni de la tribu discrète $\mathcal{T}' = \mathcal{P}(\Omega')$.

1. On munit Ω de la tribu discrète $\mathcal{T} = \mathcal{P}(\Omega)$ et de la probabilité uniforme $\mathbb{P}$. Calculer la probabilité des évenements suivants : $E = \{S = 3\}$ et $F = \{S = 4\}$.
2. On définit la classe d'évènement $\mathcal{C} = \{\{1\}, \{2\}\}$. Soit $\tilde{\mathbb{P}}$ une mesure de probabilité définie sur $\mathcal{C}$ par $\tilde{\mathbb{P}}(\{1\}) = \frac{1}{6}$ et $\tilde{\mathbb{P}}(\{2\}) = \frac{1}{6}$. On munit Ω de la tribu $\sigma(\mathcal{C})$.

   Peut-on calculer $\tilde{\mathbb{P}}(E)$ et $\tilde{\mathbb{P}}(F)$?

Moralité 2 : Si on a une application $S : \Omega \to \Omega'$, Pour pouvoir calculer la probabilité que $S \in A$ pour $A \in \mathcal{T}'$, il faut que $S^{-1}(A)$ soit un élément de la tribu sur laquelle est définie la probabilité : il faut que S soit ce qu'on appelle mesurable...

C'est tout le principe des variables aléatoires...

2.2 Variables aléatoires

2.2.1 variables aléatoires et lois de probabilité

Définition 2.9. Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace de probabilité et $(\Omega', \mathcal{T}')$ un espace mesurable.

On appelle **variable aléatoire** (**v.a.**) une fonction

$$X : \Omega \to (\Omega', \mathcal{T}')$$

qui est $\mathcal{T}$-mesurable, i. e. : $X : \forall A \in \mathcal{T}'$, $X^{-1}(A) := \{w \in \Omega \mid X(w) \in A\} \in \mathcal{T}$.

1. Si $X(\Omega)$ est un ensemble fini, on parle de **variable aléatoire finie**,
2. Si $X(\Omega)$ est un ensemble (fini ou) dénombrable on parle de **variable aléatoire discrète**,
3. Si $X(\Omega)$ est un ensemble (fini ou) dénombrable de $\mathbb{N}$, on parle de **variable aléatoire entière**,
4. Si $X(\Omega)$ est un sous-ensemble de $\mathbb{R}$, on parle de **variable aléatoire réelle (v.a.r)**.
5. si $X(\Omega)$ est un sous-ensemble de $\mathbb{R}^n$, on parle de **variable aléatoire n-dimensionnelle** ou de **vecteur aléatoire réel**.

Remarque 2.10. Forme générale des v.a.r. discrètes (écriture fonctionnelle)

- **v.a.r. finies** : X est de la forme $\sum_{i=1}^{N} \alpha_i \mathbf{1}_{A_i}$ où $N \in \mathbb{N}$, $\alpha_i \in \mathbb{R}$, $A_i \in \mathcal{T}$,
- **v.a.r. discrètes** : X est de la forme $\sum_{i=1}^{+\infty} \alpha_i \mathbf{1}_{A_i}$ où $\alpha_i \in \mathbb{R}$, $A_i \in \mathcal{T}$,
- **v.a.r entières** : X est de la forme $\sum_{n=1}^{+\infty} n\mathbf{1}_{A_i}$ où $\alpha_i \in \mathbb{R}$, $A_i \in \mathcal{T}$.

Proposition 2.11. *Soient X, Y, et $(X_n)_{n\geq 0}$ des v. a. r.*
Les fonctions suivantes sont également des v. a. r. :

1. *X^+, X^- et $-X$,*
2. *$\sup(X,Y)$ et $\inf(X,Y)$,*
3. *$\sup_{n\geq 0} X_n$, $\inf_{n\geq 0} X_n$, $\limsup_{n\geq 0} X_n$, $\liminf_{n\geq 0} X_n$ et $\lim_{n\to+\infty} X_n$ si elle existe,*
4. *$X+Y$, XY, $\frac{X}{Y}$ si elle existe, et d'une manière générale, toute fonction $\phi(X,Y)$ où ϕ est mesurable.*

Définition 2.12. La **loi de probabilité d'une variable aléatoire** X est une mesure qui permet de donner, pour tout ensemble $A \subset X(\Omega)$, la proportion de la population vérifiant $X(\omega) \in A$, notée $\mathbb{P}(X \in A)$ ou $\mathbb{P}_X(A)$:

Si deux v.a. X et Y ont la même loi de probabilité, on dit quelles sont **équidistribuées**. On note alors $X \sim Y$.

Cas discrèt : Soit X une v.a. discrète à valeurs dans E dénombrable.

La loi de X est donnée par la famille de nombres $\{\mathbb{P}(X = k) \ : \ k \in E\}$.

Si de plus $E \subset \mathbb{R}$, la loi de X est entièrement déterminée par les familles suivantes :
– $\{\mathbb{P}(X \leq k) \ : \ k \in E\}$,
– $\{\mathbb{P}(X \geq k) \ : \ k \in E\}$.

Cas continu : Soient X une v.a. réelle continue.

La loi de X est entièrement déterminée par l'une des familles suivantes :
– $\{\mathbb{P}(X \in]-\infty, x]) \ : \ x \in \mathbb{R}\}$,
– $\{\mathbb{P}(X \in [x, +\infty[) \ : \ x \in \mathbb{R}\}$,
– $\{\mathbb{P}(X \in [a,b]) \ : \ a < b \in \mathbb{R}\}$.

Le cas usuel est celui des **variables aléatoires réelle à densité**, pour lesquelles il existe une fonction $f_X : \mathbb{R} \to \mathbb{R}_+$, intégrable sur $\mathbb{R}$ telle que

$$\forall a \leq b \in \mathbb{R}^2, \ \mathbb{P}(X \in [a,b]) = \int_a^b f_X(x)dx.$$

Dans ce cas-là, la **fonction de densité** f_X suffit à décrire entièrement la loi de X.
Attention : toutes les variables aléatoires continues ne sont pas à densité même si c'est souvent le cas...

Précisions : Certes, j'avais promis de plus faire de théorie de la mesure mais je ne peux pas m'empêcher... En fait, la loi de d'une v.a. $X : (\Omega, \mathcal{T}, \mathbb{P}) \to (\Omega', \mathcal{T}')$ est la mesure image $\mathbb{P}_X$ de $\mathbb{P}$ par X i.e. la mesure définie sur $(\Omega', \mathcal{T}')$ par :

$$\forall A \in \mathcal{T}', \mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A)).$$

Méthode : Pour montrer l'équidistribution...

Puisqu'on peut définir une probabilité uniquement sur une partie génératrice de la tribu de l'espace d'arrivée, pour vérifier l'équidistribution de se v.a., on peut se contenté de vérifier que leur lois coïncident sur une partie génératrice de la tribu de l'espace d'arrivée. D'où les méthodes suivantes :

1. Soient X et Y deux v.a. discrètes à valeurs dans E dénombrable. Si pour tout $k \in E$, $\mathbb{P}_X(\{k\}) = \mathbb{P}_Y(\{k\})$, alors $\mathbb{P}_X = \mathbb{P}_Y$.
2. Soient X et Y deux v.a.r.
   - Si pour tout $x \in \mathbb{R}$, $\mathbb{P}_X(]-\infty, x]) = \mathbb{P}_Y(]-\infty, x])$, alors $\mathbb{P}_X = \mathbb{P}_Y$,
   - Si pour tout $x \in \mathbb{R}$, $\mathbb{P}_X([x, +\infty[) = \mathbb{P}_Y([x, +\infty[)$, alors $\mathbb{P}_X = \mathbb{P}_Y$,
   - Si pour tout $a < b \in \mathbb{R}$, $\mathbb{P}_X([a, b]) = \mathbb{P}_Y([a, b])$, alors $\mathbb{P}_X = \mathbb{P}_Y$,

   Cela est valable si X et Y sont à valeurs réelles dans des ensemble finis, dénombrables ou continus.

Remarque 2.13. En fait, on va rarement travailler avec Ω. Ce qui intéresse, ce n'est pas tant l'espace des réalisations (Ω) mais plutôt l'espace des résultats Ω'... $\mathbb{P}_X$ représente la manière dont X "déforme la mesure $\mathbb{P}$". deux variables aléatoires peuvent être définies sur des espaces Ω très différents mais avoir au final le même comportement :

Il est tout a fait possible d'avoir $X \neq Y$ et $\mathbb{P}_X = \mathbb{P}_Y$...

Pour se convaincre, un petit exemple...

Exemple 2.14. Soit X la variable aléatoire définie sur $\Omega = \{a, b, c, d\}$ muni de sa tribu discrète et de la mesure uniforme et à valeurs dans $\{0, 1\}$ comme suit : $X(a) = X(b) = 1$, $X(c) = X(d) = 0$.

Soit Y la variable aléatoire définie sur $\mathbb{R}$ muni de sa tribu borélienne usuelle $\mathcal{B}(\mathbb{R})$ et de la mesure de Lebesgue et à valeurs dans $\{0, 1\}$ comme suit : $Y(x) = 1$ si la partie fractionnaire de x est plus petite que $\frac{1}{2}$, c'est-à-dire si $x - \lfloor x \rfloor \leq \frac{1}{2}$ et $Y(x) = 0$ sinon.

Ces 2 variables aléatoires sont a priori bien différentes et pourtant, elle sont équidistribuées... : $\mathbb{P}_X(\{1\}) = \mathbb{P}_Y(\{1\}) = \frac{1}{2}$ et $\mathbb{P}_X(\{0\}) = \mathbb{P}_Y(\{0\}) = \frac{1}{2}$. Elles suivent une loi $\mathcal{B}\left(\frac{1}{2}\right)$.

★ **Exercice 2.15.** On dispose d'une pièce dont les deux faces sont désignées par P et F. On jette cette pièce autant de fois qu'il est nécessaire pour obtenir le résultat P deux fois consécutivement. On désigne par N la variable aléatoire qui prend la valeur n ($n \geq 2$) si ces deux résultats sont obtenus (pour la première fois) aux jets de rangs $n-1$ et n. On note $p_n = \mathbb{P}(N = n)$.

1. Calculer p_n pour $n = 2, 3, 4$.
2. Soit $n \geq 4$. Exprimer p_n en fonction de p_{n-1} et p_{n-2}.
3. Calculer p_n en fonction de n.

2.2.2 Fonction de répartition d'une variable aléatoire

Définition 2.16. On appelle **fonction de répartition d'une v.a.r.** X la fonction F_X définie ci-dessous :

$$\begin{array}{rccl} F_X: & \mathbb{R} & \to & [0,1] \\ & t & \mapsto & \mathbb{P}(X \leq t) = \mathbb{P}_X\left(]-\infty, t]\right). \end{array}$$

$F_X(t)$ est la probabilité que X soit plus petit ou égal à t. De cette définition découle plusieurs propriétés évidentes :

Propriétés 2.17. *Soit X une v.a.r. de fonction de répartition F_X.*

F_X est toujours croissante, continue à droite et vérifie $\lim_{t\to-\infty} F_X(t) = 0$ et $\lim_{t\to+\infty} F_X(t) = 1$.

De plus :

- *si F_X est en escaliers, alors X est une v.a.r. discrète,*
- *si F_X est continue, on dit que X est une* **v.a.r. diffuse***,*
- *si F_X est dérivable presque partout, de dérivée f_X, alors X est une v.a.r. à densité, de densité f_X.*

Remarque : Pour toute fonction F croissante, continue à droite et vérifiant $\lim_{t\to-\infty} F(t) = 0$ et $\lim_{t\to+\infty} F(t) = 1$, il existe une variable aléatoire réelle X telle que $F = F_X$.

Formes usuelles des fonctions de répartition :

• **Fonction de répartition d'une v.a.r. discrète**

Si X est est une v.a.r. discrète, alors :

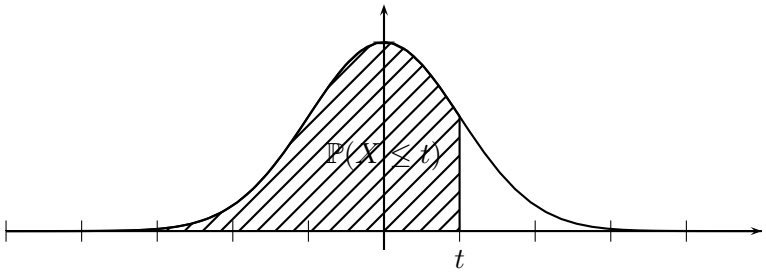
$$\forall t \in \mathbb{R},\ F_X(t) = \mathbb{P}(X \leq t) = \sum_{k \in X(\Omega) \cap]-\infty, t]} \mathbb{P}(X = k).$$

• **Fonction de répartition d'une v.a.r. à densité**

Si X est est une v.a.r. à densité f_X, alors :

$$\forall t \in \mathbb{R},\ F_X(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^{t} f_X(x)dx.$$

$F_X(t)$ est la valeur de l'aire comprise entre $y = 0$ et $y = f_X(x)$ jusqu'à $x = t$:



★ **Exercice 2.18.** Soit X une variable aléatoire de fonction de répartition F et a et b deux réels.

1. Exprimer la fonction de répartition de la variable aléatoire $Y = aX + b$ en fonction de F, a et b.
2. On suppose de plus que X admet une densité f. Déterminer la densité de la variable aléatoire Y.

Chapitre 3

Lois de probabilités usuelles

3.1 Lois de probabilité finies

La loi d'une v.a. finie X à valeurs dans $A = \{a_i\,; i \in I\}$ (avec I fini) a pour forme générale :

$$\forall a_i \in A,\ \mathbb{P}(X = a_i) = \alpha_i$$

où les $\alpha_i > 0$ vérifient $\sum_{i=1}^{N} \alpha_i = 1$.

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

C'est la loi de probabilité la plus simple : A a 2 éléments, souvent 0 et 1 et

$$\mathbb{P}(X = 1) = p = 1 - \mathbb{P}(X = 0).$$

Exemple : résultat d'un Pile ou face ; Présence ou absence d'un gène chez un individu ; sexe féminin ou masculin d'un individu ; etc...

Principales caractéristiques (avec $q = 1 - p$) :

- Fonction de répartition : $F_X(t) = \begin{cases} 0 & \text{pour } t < 0, \\ q & \text{pour } 0 < t < 1, \\ 1 & \text{pour } t > 1. \end{cases}$
- Espérance : $\mathbb{E}(X) = p$,
- Mode : $Mo = \begin{cases} 0 & \text{si } p < q, \\ 0,1 & \text{si } p = q, \\ 1 & \text{si } p > q. \end{cases}$
- Variance : $\mathbb{V}\mathrm{ar}(X) = pq$.

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

Une variable aléatoire X suivant une loi binomiale $\mathcal{B}(n,p)$, $n \in \mathbb{N}^*$ et $p \in [0,1]$ est à valeurs dans un ensemble à $n+1$ éléments A, souvent $\{0,1,\ldots,n\}$ et

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p^{n-k}).$$

Exemple : Nombre de Face obtenus à la suite de n lancers d'une même pièce ; Parmi n personne, nombres de porteurs d'un gène qui apparait avec fréquence p pour chaque individu ; etc...

Principales caractéristiques (avec $q = 1 - p$) :
- Espérance : $\mathbb{E}(X) = np$,
- Mode : $Mo = \lfloor (n+1)p \rfloor$
- Variance : $\mathbb{V}\text{ar}(X) = npq$.
- Génèse : La somme de n variables de Bernoulli $\mathcal{B}(p)$ suit une loi binomiale $\mathcal{B}(n, p)$.

3.1.3 Loi hypergéométrique $\mathcal{H}(n, p, N)$, $n \leq N \in \mathbb{N}^*$, $p \in [0, 1]$

Une variable aléatoire X suivant une loi hypergéométrique $\mathcal{H}(n, p, N)$, $n \leq N \in \mathbb{N}^*$ et $p \in [0, 1]$. On pose $q = 1 - p$. X est à valeurs dans $\{\max(0, n - qN), \ldots, \min(pN, n)\}$

$$\mathbb{P}(X = k) = \frac{\binom{pN}{k}\binom{qN}{n-k}}{\binom{N}{n}}.$$

Exemple : On a N boules dans une urne, p représente la proportion de boules blanches dans l'urne : on a pN boules blanches et qN boules noires. On tire au hasard n boules de l'urne. $\mathbb{P}(X = k)$ représente la probabilité que sur les n boules tirées, exactement k soient blanches.

Principales caractéristiques (avec $q = 1 - p$) :
- Espérance : $\mathbb{E}(X) = np$,
- Mode : $Mo = \left\lfloor (n+1)\dfrac{pN+1}{N+2} \right\rfloor$
- Variance : $\mathbb{V}\text{ar}(X) = npq\frac{N-n}{N-1}$.

Loi hypergéométrique et loi binomiale : Lorsque N tend vers l'infini et que p reste constant, la loi hypergéométrique $\mathcal{H}(n, p, N)$ converge en loi vers une loi binomiale $\mathcal{B}(n, p)$:

$$\mathcal{H}(n, p, N) \xrightarrow{\mathcal{L}} \mathcal{B}(n, p)$$

Autrement dit, si N est très grand devant n, on peut utiliser plutot une loi binomiale, qui est plus simple pour modéliser une loi hypergéométrique.

3.1.4 Loi multinomiale $\mathcal{M}(N, p_1, \ldots, p_n)$

la loi multinomiale généralise la loi binomiale : La loi binomiale compte le nombre de succès/échecs pour N tirages d'une expérience de Bernoulli et une loi multinomiale compte simultanément le nombre d'apparition des différents résultats pour N tirages d'une expérience à n résultats possibles :

Une variable $(X_1, X_2, \ldots, X_n)$ qui suit une loi multinomiale $\mathcal{M}(N, p_1, \ldots, p_n)$, avec $p_i \in]0, 1[$ et $\sum_{i=1}^n p_i = 1$ est telle que :

$\forall (N_1, N_2, \ldots, N_n)$ tels que $\sum_{i=1}^n N_i = N$,

$$\mathbb{P}\left(X_1 = N_1, X_2 = N_2, \ldots, X_n = N_n\right) = \frac{N!}{N_1!N_2!\ldots N_n!} p_1^{N_1} p_2^{N_2} \ldots p_n^{N_n}.$$

Principales caractéristiques :
- Espérances : $\mathbb{E}(X_i) = Np_i$,
- Variances : $\mathbb{V}\text{ar}(X_i) = Np_i(1-p_i)$,
- Covariances : $\mathbb{C}\text{ov}(X_i, X_j) = -Np_ip_j$ pour $i \neq j$.

3.2 Lois de probabilités discrètes dénombrables

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

Une variable aléatoire X qui suit une loi de Poisson $\mathcal{P}(\lambda)$, $\lambda > 0$ prend ses valeurs dans $\mathbb{N}$ et vérifie :

$$\forall k \in \mathbb{N},\ \mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Principales caractéristiques :
- Espérance : $\mathbb{E}(X) = \lambda$
- Mode : $Mo = \lceil \lambda \rceil - 1$
- Variance : $\mathbb{V}\text{ar}(X) = \lambda$.
- Génèse : La limite de la loi binomiale $\mathcal{B}(n, p_n)$ lorsque $n \to +\infty$ et $np_n \to \lambda > 0$ suit une loi de Poisson $\mathcal{P}(\lambda)$.

La loi de Poisson est soivent utilisée pour modéliser des caractères dont la variance est environ égale à la moyenne.

Loi de Poisson et évènements rares

La loi de Poisson est aussi appelée loi des évènements rares à cause du fait suivant : Si X représente le nombre de fois qu'un évènement de très faible probabilité apparait, alors on peut approcher X par une loi de Poisson. Par exemple, on suppose que sur la population mondiale, une très petite proportion p est porteur d'un gène G (c'est notre évènement rare). On tire des échantillons de N personnes et on s'intéresse la probabilité pour que l'echantillon contienne k porteurs du gène G, c'est-à-dire la probabilité $P(G = k) = \binom{N}{k}p^k(1-p)^k$. Le calcul effectif d'une telle probabilité est difficile, surtout à cause du facteur $\binom{N}{k}$, que replace par une quantité proche, plus facile à calculer : $P(G = k) = \binom{N}{k}p^k(1-p)^k \approx \frac{e^{-Np}}{k!}(Np)^k$.

Plus généralement, si X suit une loi $\mathcal{B}(N, p)$ avec N grand, p petit et Np raisonnable, on peut approcher la loi $\mathcal{B}(N, p)$ par une loi de Poisson de paramètre $\lambda = Np$...Les conditions d'une telle approximation sont un peu vagues, certes, mais en général, on considère que pour avoir une approximation raisonnable, il faut supposer $N > 30$ et $Np > 5$.

3.2.2 Loi géométrique $\mathcal{G}(p)$, $p \in]0, 1[$

Une variable aléatoire X qui suit une loi géométrique $\mathcal{G}(p)$ avec $p \in]0, 1[$ prend ses valeurs dans $\mathbb{N}^*$ et vérifie :

$$\mathbb{P}(X = k) = p(1-p)^{k-1}.$$

$\mathbb{P}(X = k)$ donne la probabilité que le premier succès d'une expérience de Bernoulli de parametre p arrive après $k-1$ échecs.

Principales caractéristiques :
- Espérance : $\mathbb{E}(X) = \frac{1}{p}$
- Mode : $Mo = 1$
- Variance : $\mathbb{V}\text{ar}(X) = \frac{1-p}{p^2}$.

On rencontre parfois pour la loi géométrique, la définition alternative suivante : $\mathbb{P}(Y = k)$ est la probabilité d'obtenir k échecs suivi d'un succès lors d'une succession d'épreuves de Bernoulli indépendantes. Elle modélise la durée de vie d'une entité qui aurait, à tout instant la probabilité p de mourir. On obtient alors, pour $k \in \mathbb{N}$, $\mathbb{P}(Y = k) = pq^k$. Ce n'est en fait qu'un décalage de la précédente définition : $Y = X + 1$.

3.2.3 Loi binomiale négative ou Loi de Pascal $\mathcal{N}eg\mathcal{B}(N, p)$, $N \in \mathbb{N}^*$, $p \in]0, 1[$

Une variable aléatoire X qui suit une loi binomiale négative ou Loi de Pascal $\mathcal{N}eg\mathcal{B}(N, p)$, avec $N \in \mathbb{N}^*$ et $p \in]0, 1[$ $(q = 1 - p)$ prend ses valeurs dans $\mathbb{N}$ et vérifie :

$$\forall k \in \mathbb{N},\ \mathbb{P}(X = k) = \binom{k + N + 1}{k} p^N q^k.$$

Cette variable aléatoire compte le nombre d'échecs nécessaires avant l'obtention de k succès d'une variable de Bernoulli $\mathcal{B}(p)$. On l'appelle Binomiale négative car on peut écrire $\mathbb{P}(X = k) = \binom{-N}{k} p^N (-q)^k$. avec la convention $\binom{-n}{k} = \frac{(-n)(-n-1)\ldots(-n-k+1)}{k!}$.

Principales caractéristiques ($q = 1 - p$) :
- Espérance : $\mathbb{E}(X) = \frac{q}{p}$
- Mode : $Mo = 0$
- Variance : $\mathbb{V}\text{ar}(X) = \frac{q}{p^2}$.
- Génèse : La somme de n varianles de loi géométrique $\mathcal{G}(p)$ indépendantes suit une loi $\mathcal{N}eg\mathcal{B}(N, p)$.

On trouve également une définition alternative de la loi binomiale négative, qui donne le nombre Y d'essais (au lieu des échecs) nécessaires à l'obtention de N succès. Dans ce cas, on a :$\mathbb{P}(Y = k) = \binom{k-1}{k-N} p^N q^{k-N}$. On a de toutes façons $Y = X + N$.

3.3 Lois de probabilité continues

3.3.1 Loi uniforme $\mathcal{U}[a,b]$, $a < b \in \mathbb{R}^2$

Une variable aléatoire X qui suit une loi uniforme $\mathcal{U}[a,b]$ avec $a < b \in \mathbb{R}^2$ est une variable aléaoire de densité f définie sur $\mathbb{R}$ par :

$$\forall x \in \mathbb{R},\ f(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a,b], \\ 0 & \text{si } x \notin [a,b]. \end{cases}$$

Principales caractéristiques :

- Fonction de répartition : $F_X(t) = \dfrac{t-a}{b-a}$.
- Espérance : $\mathbb{E}(X) = \dfrac{a+b}{2}$
- Médiane : $Me = \dfrac{a+b}{2}$
- Variance : $\mathbb{V}\text{ar}(X) = \frac{(b-a)^2}{6}$.

3.3.2 Loi exponentielle $E(\lambda)$, $\lambda \in \mathbb{R}_+^*$

Une variable aléatoire X qui suit une loi exponentielle $E(\lambda)$ avec $\lambda \in \mathbb{R}_+^*$ est une variable aléaoire de densité f définie sur $\mathbb{R}$ par :

$$\forall x \in \mathbb{R},\ f(x) = \begin{cases} 0 & \text{si } x > 0, \\ \lambda e^{-\lambda x} & \text{si } x \geq 0. \end{cases}$$

Principales caractéristiques :

- Fonction de répartition : $F_X(t) = 1 - e^{-\lambda x}$.
- Espérance : $\mathbb{E}(X) = \dfrac{1}{\lambda}$
- Médiane : $Me = \dfrac{\ln 2}{\lambda}$
- Mode : $Mo = 0$
- Variance : $\mathbb{V}\text{ar}(X) = \frac{1}{\lambda^2}$.

Propriété caractéristique des lois exponentielles :

$$\forall (s,t) \in \mathbb{R}^{+\,2},\ \mathbb{P}(X > s+t | X > t) = \mathbb{P}(X > s)$$

Les lois exponentielles modélisent la durée de vie d'un phénomène sans mémoire : la probabilité que le phénomène dure au moins s+t heures sachant qu'il a déjà duré t heures sera la même que la probabilité de durer s heures à partir de sa mise en fonction initiale. En d'autres termes, le fait que le phénomène ait duré pendant t heures ne change rien à son espérance de vie à partir du temps t.

3.3.3 Loi Normale ou Loi de Laplace-Gauss $\mathcal{N}(m, \sigma^2)$, $m \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}_+^*$

Une variable aléatoire X qui suit une loi Normale $\mathcal{N}(m, \sigma^2)$ avec $m \in \mathbb{R}$ et $\sigma^2 \in \mathbb{R}_+^*$ est une variable aléaoire de densité f définie sur $\mathbb{R}$ par :

$$\forall x \in \mathbb{R},\ f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_m)^2}{2\sigma^2}}.$$

L'importance de la loi normale est avant tout issue du théorème central limite : Si on somme un nombre suffisament grand de variables aléatoires indépendantes (ou faiblement liées) qui suivent des lois quelconques (ou presque...) alors cette somme se comporte comme une loi normale. Les hypothèses de ce théorème central limite sont assez faibles et se retrouvent dans beaucoup de cas pratiques, notamment pour le traitement des erreurs (pour s'assurer par exemple que la somme des erreurs n'influence pas trop les résultats de la somme des mesures, etc...)

Principales caractéristiques :

- Espérance : $\mathbb{E}(X) = m$
- Médiane : $Me = m$
- Mode : $Mo = m$
- Variance : $\mathbb{V}\text{ar}(X) = \sigma^2$.

La fonction de répartition de la loi normale n'ayant pas d'expression "sympathique", on se sert en général de tables des valeurs de F_X pour déterminer des quantités du type $\mathbb{P}(x \in [a, b])$ via l'égalité $\mathbb{P}(x \in [a, b]) = F_X(b) - F_X(a)$.

3.3.4 Loi Gamma $\Gamma(k, \theta)$, $(k, \theta) \in {\mathbb{R}_+^*}^2$

Une variable aléatoire X qui suit une loi Gamma $\Gamma(k, \theta)$, avec $k \in \mathbb{R}_+^*$ et $\theta \in \mathbb{R}_+^*$ est une variable aléaoire de densité f définie sur $\mathbb{R}$ par :

$$\forall x \in \mathbb{R},\ f(x) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\Gamma(k)\, \theta^k}.$$

où la fonction Γ est la fonction définie sur le demi-plan complexe de partie réelle positive par $\Gamma(z) = \displaystyle\int_0^{+\infty} t^{z-1}\, e^{-t}\, \mathrm{d}t$

3.3.5 Loi du χ_n^2, $n \in \mathbb{N}^*$

La loi du **Chi-2 (ou Khi-2, ou Khi-carré) à n degrés de liberté**, notée χ_n^2 est la loi obtenue en sommant n carrés de loi normales $\mathcal{N}(0, 1)$ indépendantes :

Une variable aléatoire K suit une loi χ_n^2 s'il existe n variables aléatoires indépendantes X_1, X_2,..., X_n de loi $\mathcal{N}(0, 1)$ telles que

$$K = X_1^2 + X_2^2 + \ldots X_n^2 = \sum_{i=1}^{n} X_i^2.$$

Ces lois sont à valeurs positives, et leurs fonctions de répartition ne sont pas très sympathique donc on utilise également des tables de sa fonction de répartition pour connaitre les valeurs de $\mathbb{P}(K \in [a, b])$.

Principales caractéristiques :

- Espérance : $\mathbb{E}(K) = n$
- Médiane : $Me = m$
- Mode : $Mo = k - 2$ si $k \geq 2$.
- Variance : $\mathbb{V}\text{ar}(K) = 2n$.

La principale utilisation de cette loi consiste à apprécier l'adéquation d'une loi de probabilité à une distribution empirique en utilisant le test du χ^2 basé sur la loi multinomiale (voir paragraphe 3.1.4).

3.3.6 Lois de Student $\mathcal{T}_n$, $n \in \mathbb{N}^*$

Une variable aléatoire T suit une **loi de Student à n degré de liberté** $\mathcal{T}_n$ s'il existe une variable aléatoire X de loi $\mathcal{N}(0, 1)$ et une variable aléatoire K de loi χ^2_n telles que :

$$T = \frac{X}{\sqrt{\frac{K}{n}}}.$$

Principales caractéristiques :

- Espérance : $\mathbb{E}(T) = \begin{cases} 0 & \text{si } n > 1, \\ \text{indet.} & \text{sinon.} \end{cases}$
- Médiane : $Me = 0$
- Mode : $Mo = 0$ si $k \geq 2$.
- Variance : $\mathbb{V}\text{ar}(X) = \begin{cases} \text{indet.} & \text{si } n \leq 1, \\ +\infty & \text{si } 1 < n \leq 2, \\ \frac{n}{n-2} & \text{si } n \geq 2. \end{cases}$.

3.3.7 Loi de Fisher $\mathcal{F}^n_p$, $(n, p) \in N^2$

Une variable aléatoire F suit une loi de Fisher $\mathcal{F}^n_p$ (ou Fisher-Snedecor) de paramètres (n, p) avec $(n, p) \in N^2$ s'il existe une variable aléatoire K_1 de loi χ^2_n et une variable aléatoire K_2 de loi χ^2_p telles que :

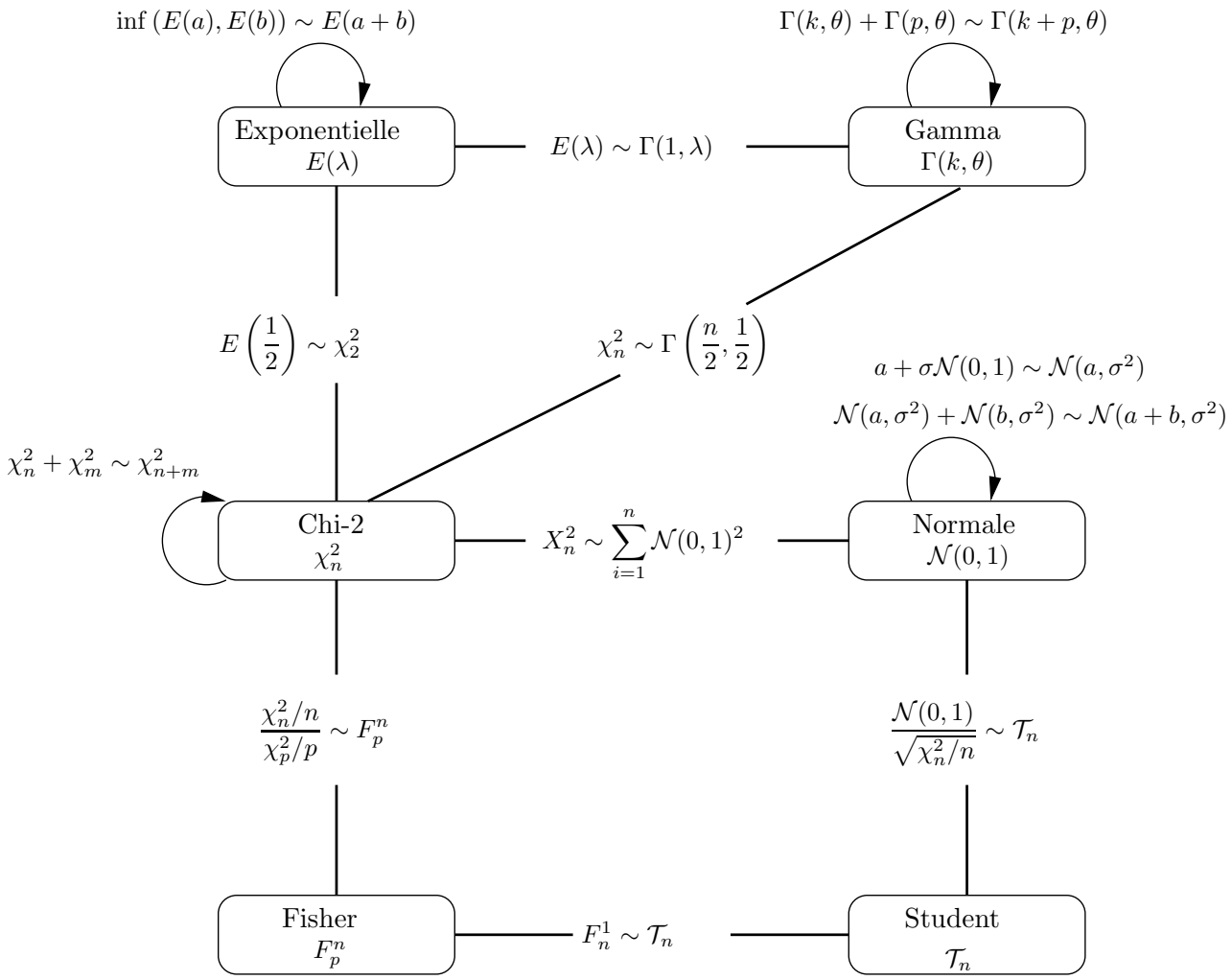
$$F = \frac{\frac{K_1}{n}}{\frac{K_2}{p}}.$$

Attention à l'ordre des degrés...

Principales caractéristiques :

- Espérance : $\mathbb{E}(T) = \dfrac{p}{p-2}$ si $p > 2$.
- Mode : $Mo = \dfrac{n-2}{n}\dfrac{p}{p+2}$ si $n \geq 2$.
- Variance : $\mathbb{V}\text{ar}(X) = \dfrac{2p^2(n+p-2)}{n(p-2)(p-4)}$ si $p > 4$.

3.3.8 Récapitulatif des relations entre les lois continues



Chapitre 4

Indépendance et probabilités conditionnelles

4.1 Probabilités conditionnelles

Définition 4.1. Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace de probabilités et soit $B \in \mathcal{T}$ tel que $\mathbb{P}(B) > 0$.
Pour tout $A \in \mathcal{A}$, on pose

$$\mathbb{P}(A \,|\, B) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(B)}.$$

On appelle $\mathbb{P}(A \,|\, B)$ la **probabilité de A sachant l'événement B** ou **probabilité de A conditionnée par l'événement B**.

Propriétés 4.2. *Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace de probabilités et soient $B \in \mathcal{T}$ tel que $\mathbb{P}(B) > 0$ et $A \in \mathcal{T}$.*

1. *L'application définie sur $\mathcal{T}$ par $A \mapsto \mathbb{P}(A \,|\, B)$ est une mesure de probabilité sur $\mathcal{T}$,*
2. $\mathbb{P}(A \,|\, B) = 0$ *si* $A \cap B = \emptyset$,
3. $\mathbb{P}(A \,|\, B) = \dfrac{\mathbb{P}(A)}{\mathbb{P}(B)}$ *si* $A \subset B$,
4. $\mathbb{P}(A \,|\, B) = 1$ *si* $B \subset A$,
5. $\mathbb{P}(A \,|\, B) = \mathbb{P}(A)$ *si* $A \perp B$.

★ **Exercice 4.3.** Démontrer les propriétés décrites ci-dessus.

Ci-dessous, trois résultats fondamentaux concernant les probabilités conditionnelles :

Proposition 4.4. Formule des probabilités composées
Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace de probabilité et soient $(B_i)_{i=1}^n$ des éléments de $\mathcal{T}$ tels que $\mathbb{P}(\cap_{i=1}^n B_i) > 0$.

$$\mathbb{P}(B_1 \cap B_2 \cap \ldots \cap B_n) = \mathbb{P}(B_1) \times \mathbb{P}(B_2 \,|\, B_1) \times \ldots \times \mathbb{P}(B_n \,|\, B_{n-1} \cap \ldots \cap B_1).$$

Cette reste vraie pour une quantité dénombrable d'évènements :

$$\mathbb{P}\left(\bigcap_{i=1}^{+\infty} B_i\right) = \mathbb{P}(B_1) \times \prod_{i=2}^{+\infty} \mathbb{P}\left(B_i \mid \cap_{k=1}^{i-1} B_k\right).$$

Proposition 4.5. Formule des probabilités totales

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace de probabilité et soient A et $(B_i)_{i\in\mathbb{N}}$ des éléments de $\mathcal{T}$ tels que, pour tout $i \in \mathbb{N}$, $\mathbb{P}(B_i) > 0$.

Si les $(B_i)_{i\in\mathbb{N}}$ forment une partition de Ω alors

$$\mathbb{P}(A) = \sum_{i\in\mathbb{N}} \mathbb{P}(B_i) \times \mathbb{P}\left(A \mid B_i\right).$$

La formule des probabilités totales est aussi vraie pour une partition finie. Il faut dans tous les cas qu'ils forment une partition de Ω ou du moins que les B_i soient disjoints et que $A \subset \cup_{i\in\mathbb{N}} B_i$.

Proposition 4.6. Formule de Bayes

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace de probabilité et Soient A et $(B_i)_{i\in\mathbb{N}}$ des éléments de $\mathcal{T}$ tels que $\mathbb{P}(A) \neq 0$ et pour tout $i \in \mathbb{N}$, $\mathbb{P}(B_i) > 0$.

Si les $(B_i)_{i\in\mathbb{N}}$ forment une partition de Ω, alors, pour tout $i \in \mathbb{N}$, on a :

$$\mathbb{P}\left(B_i \mid A\right) = \frac{\mathbb{P}(B_i) \times \mathbb{P}\left(A \mid B_i\right)}{\sum_{i\in\mathbb{N}} \mathbb{P}(B_i) \times \mathbb{P}\left(A \mid B_i\right)}.$$

La formule de Bayes permet en quelque sorte "d'inverser" le conditionnement : connaissant la probabilité de A sachant tous les B_i, on peut retrouver la probabilité de tous les B_i sachant A.

4.2 Evènements indépendants et variables aléatoires indépendantes

4.2.1 Evènements indépendants

La notion d'indépendance de v.a. vient en fait directement de la notion d'indépendance d'ensembles au sein de la tribu d'un espace de probabilités

Définition 4.7. Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace de probabilités.

Deux éléments A et B de $\mathcal{T}$ sont des **évènements indépendants** si

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

On note souvent $A \perp B$ lorsque A et B sont indépendants.

Remarque : Comme conséquences directes de la définition, on a :

1. Si $\mathbb{P}(B) = 0$, alors, pour tout évènement A, $A \perp B$.
2. Si $A \perp B$ et $\mathbb{P}(B) > 0$, alors $\mathbb{P}\left(A \mid B\right) = \mathbb{P}(A)$.

Proposition 4.8. *Soient A, B, C des évènements d'un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$.*

On suppose $A \perp B$ et $A \perp C$. On a :

1. $A \perp {}^cB$ *et* $A \perp {}^cC$,
2. *si* $B \subset C$, *alors* $A \perp C \setminus B$,
3. *Si les* $(A_i)_{i\in\{1,2\ldots,n\}}$ *sont des évènements disjoints,*

$$\forall i \in \{1, 2 \ldots, n\},\ A_i \perp B \Longrightarrow B \perp \bigcup_{i=1}^{n} A_i.$$

★ **Exercice 4.9.** Montrer les propriétés de la Proposition 4.8.

Remarque 4.10. Attention : pour généraliser à plusieurs évènements : L'indépendance 2 à 2 est différente de l'indépendance globale comme le montre l'exemple ci-dessous...

★ **Exercice 4.11.** On jette deux dés de façon indépendante et équiprobable. Soient A l'évenemnt "le chiffre du premier dé est impair", B l'évenemnt "le chiffre du deuxieme dé est pair" et C l'évenement "les chiffres des deux dés ont même parité".

1. Montrer que $A \perp B$, $B \perp C$ et $A \perp C$.
2. Calculer $\mathbb{P}(A \cap B \cap C)$ et $\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$.

A, B et C sont des évènements indépendants deux à deux mais pas dans leur ensemble...

Proposition 4.12. Critère d'indépendance pour une famille d'évènements I
Des événements $(A_j)_{j\in J}$ *sont indépendants dans leur ensemble si et seulement si on a :*

$$\forall I \subset J,\ \mathbb{P}\left(\bigcap_{i\in I} A_i\right) = \prod_{i\in I} \mathbb{P}(A_i).$$

Proposition 4.13. Critère d'indépendance pour une famille d'évènements II
Des événements $(A_j)_{j\in J}$ *sont indépendants dans leur ensemble si et seulement si on a :*

$$\forall \epsilon \in \{0,1\}^n,\ \mathbb{P}\left(\bigcap_{j\in J} A_j^{\epsilon_j}\right) = \prod_{j\in J} \mathbb{P}(A_j^{\epsilon_j}),$$

avec la convention $A_j^0 = {}^cA_j$ *et* $A_j^1 = A_j$.

Remarques 4.14. Il est évident que la définition de l'indépendance pour des familles d'évènements entraine les 2 propositions précédentes, car pour tout $j \in J$, A_j, Ω et ${}^cA_j$ appartiennent à $\sigma(A_j)$. En revanche, les réciproques ne sont pas immédiates.
Attention : L'indépendance des évènements $A_1, A_2, \ldots, A_n$ entraîne en particulier que :

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \prod_{i=1}^{n} \mathbb{P}(A_i)$$

ce qui correspond au choix particulier $I = \{1, 2, \ldots, n\}$, mais est une propriété beaucoup plus faible que l'indépendance pour les évènements (c.f. Exercice 4.11).

Ci-dessous, une application usuelle de la notion d'indépendance d'une famille d'évenements, le théorème de Borel-Cantelli. Avant son énoncé, on rappelle les définitions suivantes :

Définition 4.15. Soit $(A_n)_{n\geq 0}$ une suite d'évènements d'un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$.

On définit l'ensemble

$$\limsup_{n\geq 0} A_n = \bigcap_{n\geq 0} \bigcup_{k\geq n} A_k,$$

c'est l'ensemble des éléments de Ω qui appartiennent à une infinité d'evènements A_n.

On définit aussi

$$\liminf_{n\geq 0} A_n = \bigcup_{n\geq 0} \bigcap_{k\geq n} A_k$$

qui est l'ensemble des éléments de Ω qui appartiennent à tous les d'evènements A_n à partir d'un certain rang.

Lemme 4.16. Lemme de Borel-Cantelli

Soit $(A_n)_{n\geq 0}$ une suite d'évènements d'un espace de probabilité $(\Omega, \mathcal{T}, \mathbb{P})$.

1. *si $\sum_{n\geq 0} \mathbb{P}(A_n) < +\infty$, alors $\mathbb{P}(\limsup_{n\geq 0} A_n) = 0$.*
2. *si on suppose de plus que les $(A_n)_{n\geq 0}$ sont indépendants, on a :*
   si $\sum_{n\geq 0} \mathbb{P}(A_n) = +\infty$, alors $\mathbb{P}(\limsup_{n\geq 0} A_n) = 1$.

Remarque 4.17. Dans le premier point du lemme, l'hypothèse n'est pas affaiblissable. Il existe des familles telles que $\lim \mathbb{P}(A_n) = 0$ et $\mathbb{P}(\limsup_{n\geq 0} A_n) = 1$. Par exemple, dans l'espace de probabilités $([0,1], \mathcal{B}([0,1]), \lambda)$ où λ désigne la mesure de Lebesgue sur $[0,1]$, si on choisit $A_n =]0, \frac{1}{n}]$, on a $\limsup A_n = \emptyset$ et donc $\mathbb{P}(\limsup A_n) = 0$ et pourtant $\sum_{n\geq 1} \mathbb{P}(A_n) = +\infty$.

Démonstration.

1. Puisque $\limsup_{n\geq 0} A_n = \cap_{n\geq 0} \cup_{k\geq n} A_k,$, on a, pour tout $n \geq 0$, $\limsup_{n\geq 0} A_n \subset \cup_{k\geq n} A_k$. Par conséquent :

$$\mathbb{P}\left(\limsup_{n\geq 0} A_n\right) \leq \mathbb{P}\left(\bigcup_{k\geq n} A_k\right) \leq \sum_{k\geq n} \mathbb{P}(A_k).$$

Puisque la série $\sum_{n\geq 0} \mathbb{P}(A_n)$ converge, on obtient $\mathbb{P}\left(\limsup_{n\geq 0} A_n\right) = 0$.

2. On commence par remarquer que : $\mathbb{P}\left(\limsup_{n\geq 0} A_n\right) = 1 - \mathbb{P}\left(\liminf_{n\geq 0} {}^cA_n\right)$. Les évènements ${}^cA_n$ étant également indépendants, on a (puisque la famille des évènements $\cap_{k\geq n} {}^cA_n$ est une famille croissante) :

$$\mathbb{P}\left(\liminf_{n\geq 0} {}^cA_n\right) = \lim_{n\to+\infty} \lim_{q\to+\infty} \mathbb{P}\left(\bigcap_{k=n}^{q} {}^cA_n\right) = \lim_{n\to+\infty} \lim_{q\to+\infty} \prod_{k=n}^{q} \mathbb{P}({}^cA_n).$$

D'où :

$$\mathbb{P}\left(\liminf_{n\geq 0} {}^cA_n\right) = \lim_{n\to+\infty} \lim_{q\to+\infty} \prod_{k=n}^{q} (1 - \mathbb{P}(A_n)).$$

Or, quelque soit $x \in [0,1]$, on a $0 \leq 1-x \leq exp(-x)$, donc, pour tous entiers $0 \leq n \leq q$, on obtient :

$$0 \leq \prod_{k=n}^{q}(1-\mathbb{P}(A_n)) \leq \prod_{k=n}^{q} exp\left(\mathbb{P}(A_n)\right) = exp\left(-\sum_{k=n}^{q}\mathbb{P}(A_n)\right).$$

Puisque le terme de droite tends vers 0 lorsque $q \to +\infty$, on a :

$$\lim_{q\to+\infty}\prod_{k=n}^{q}(1-\mathbb{P}(A_n)) = 0,$$

d'où $\mathbb{P}\left(\liminf_{n\geq 0} {}^cA_n\right) = 0$ et donc $\mathbb{P}\left(\limsup_{n\geq 0} A_n\right) = 1$. □

4.2.2 Variables aléatoires indépendantes

En fait, le "passage" des évènements aux variables aléatoires se fait par l'intermédiaire des tribus d'évènements générées par les variables aléatoires :

Définition 4.18. On dit que deux v.a. $X : (\Omega, \mathcal{T}, \mathbb{P}) \to (\Omega', \mathcal{T}')$ et $Y : (\Omega, \mathcal{T}, \mathbb{P}) \to (\Omega'', \mathcal{T}'')$ forme un **couple de variables aléatoires indépendantes** si :

$$\forall (A,B) \in \mathcal{T}' \times \mathcal{T}'',\ \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

Remarque : Comme d'habitude, pour montrer l'indépendance de 2 v.a. X et Y, on pourrait toujours s'assurer que l'égalité $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ est vrai sur des parties génératrices des tribus $\sigma(X)$ et $\sigma(Y)$.

Ces résultats s'etendent bien sûr à n v.a. indépendantes dans leur ensemble via la définition ci-dessous.

Définition 4.19. On dit qu'une famille quelconque de v.a. $(X_i)_{i\in I}$, $X_i : (\Omega, \mathcal{T}, \mathbb{P}) \to (\Omega_i, \mathcal{T}_i)$ forme une **famille de variables aléatoires indépendantes** si :

$$\forall (A_i) \in \mathcal{T}_i,\ \mathbb{P}\left(\bigcap_{i\in I}\{X_i \in A_i\}\right) = \prod_{i\in I}\mathbb{P}(X_i \in A_i).$$

Cette reformulation permet notamment d'avoir des critères d'indépendances dans des cas particuliers usuels.

Proposition 4.20. Critères d'indépendance de v.a.
On fixe $X : (\Omega, \mathcal{T}, \mathbb{P}) \to (\mathbb{R}^{d_X}, \mathcal{B}(\mathbb{R}^{d_X}))$ et $Y : (\Omega, \mathcal{T}, \mathbb{P}) \to (\mathbb{R}^{d_Y}, \mathcal{B}(\mathbb{R}^{d_Y}))$ deux $\vec{v}.a.r.$

1. **Si X et Y sont discrètes.**
   X et Y sont indépendantes ssi :

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

2. **Si X et Y sont de densité respectives f_X et f_Y.**
Si X et Y sont indépendantes, alors la v.a. (X,Y) admet une densité $f_{(X,Y)}$, produit direct de f_X et f_Y :

$$\forall x \in \mathbb{R}^{d_X}, \forall y \in \mathbb{R}^{d_Y},\ f_{(X,Y)}(x,y) = f_X(x) f_Y(y).$$

Inversement, si la v.a. (X,Y) admet une densité $f_{(X,Y)}$, produit direct dedeux fonctions intégrables g_X et g_Y comme suit :

$$\forall x \in \mathbb{R}^{d_X}, \forall y \in \mathbb{R}^{d_Y},\ f_{(X,Y)}(x,y) = g_X(x) g_Y(y).$$

alors g_X et g_Y sont, à un facteur positif près, près les densités respectives de X et Y et ces deux v.a. sont indépendantes.

★ **Exercice 4.21.** Soient X et Y deux v.a.r. indépendantes. Exprimer les fonctions de répartitions des v.a.r. $\inf(X,Y)$ et $\sup(X,Y)$ en fonction de celles de X et Y.

★ **Exercice 4.22.** Montrer que la fonction caractéristique de la somme de 2 v.a.r. indépendantes est égal au produit des fonctions caractéristiques de chacune d'elle.

Si on admet que l'application qui à la loi de X associe sa fonction caractéristique est bijective, en déduire la loi suivie par la somme de 2 variables indépendantes de lois de Poisson $\mathcal{P}(\lambda_1)$ et $\mathcal{P}(\lambda_2)$.

Chapitre 5

Espérance, variance et moments

5.1 Espérance et variance

5.1.1 Espérance d'une variable aléatoire réelle

Définition 5.1. L'**espérance mathématique d'une v.a.** X est sa valeur moyenne.

- Si X une **v.a.r. discrète**, $\mathbb{E}(X) = \sum_{x \in X(\Omega)} x\mathbb{P}(X = x)$,
- Si X est une **v.a.r. de densité** f_X, $\mathbb{E}(X) = \displaystyle\int_{-\infty}^{+\infty} x f_X(x)\mathrm{d}x$.

Une v.a. admettant une espérance mathématique finie est dite **intégrable**.

Attention : Toutes les variables aléatoires n'ont pas forçément une esperance finie...

Lorsque $\mathbb{E}(X) = 0$, on dit que X est **centrée**.

Proposition 5.2. *L'opérateur $\mathbb{E}$ est linéaire :*

Pour toutes v.a.r. X et Y et tout réels a, b,

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b \qquad et \qquad \mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

D'autre part, on a le théorème de transfert suivant :

Proposition 5.3. *si g est une fonction continue par morceaux (en fait, mesurable suffit), alors, sous condition d'existence :*

- *Si X une* **v.a.r. discrète**, $\mathbb{E}(g(X)) = \sum_{x \in X(\Omega)} g(x)\mathbb{P}(X = x)$,
- *Si X est une* **v.a.r. de densité** f_X, $\mathbb{E}(g(X)) = \displaystyle\int_{-\infty}^{+\infty} g(x) f_X(x)\mathrm{d}x$.

5.1.2 Moments et moments centrés d'ordre p

Définition 5.4. Pour tout $p \geq 2$, Le **moment d'ordre** p d'une v.a.r. X est la moyenne de X^p :

- Si X une **v.a.r. discrète**, $\mathbb{E}(X^p) = \displaystyle\sum_{x \in X(\Omega)} x^p\mathbb{P}(X = x)$,
- Si X est une **v.a.r. de densité** f_X, $\mathbb{E}(X^p) = \displaystyle\int_{-\infty}^{+\infty} x^p f_X(x)\mathrm{d}x$.

Les moments d'ordre p ne sont pas forçément finis. par contre, si X admet un moment d'ordre p fini, alors tous ses moments d'ordre $q \leq p$ seront eux-aussi finis.

Définitions 5.5. On définit également le **moment centré d'ordre** p de la manière suivante :

- Si X une **v.a.r. discrète**, $m_p = \mathbb{E}((X - \mathbb{E}(X))^p) = \sum_{x \in X(\Omega)} x^p \mathbb{P}(X = x)$,
- Si X est une **v.a.r. de densité** f_X, $\mathbb{E}((X - \mathbb{E}(X))^p) = \int_{-\infty}^{+\infty} x^p f_X(x) \mathrm{d}x$.

La **variance** d'une v.a.r. X est la moyenne des carrés des écarts de ses valeurs à sa moyenne. La variance n'est autre que le moment centré d'ordre 2 de la variable aléatoire X :

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

- Si X une **v.a.r. discrète**, $\mathbb{V}\mathrm{ar}(X) = \sum_{x \in X(\Omega)} x^2 \mathbb{P}(X = x) - \mathbb{E}(X)^2$,
- Si X est une **v.a.r. de densité** f_X, $\mathbb{V}\mathrm{ar}(X) = \int_{-\infty}^{+\infty} x^2 f_X(x) \mathrm{d}x - \mathbb{E}(X)^2$.

Une v.a. admettant une variance finie est dite **de carré intégrable**. Une variable de variance égale à 1 est dite **réduite**.

La racine positive $\sigma(X) = \sqrt{\mathbb{V}\mathrm{ar}(X)}$ de la variance est appelée **écart-type**.

Proposition 5.6. Propriétés de la variance

1. *La variance* $\mathbb{V}\mathrm{ar}(X)$ *est toujours positive,*
2. *Si la v.a.* X *est constante* $\mathbb{P}$*-presque surement, alors* $\mathbb{V}\mathrm{ar}(X) = 0$.
3. *Pour tous réels* a *et* b, $\mathbb{V}\mathrm{ar}(aX + b) = a^2 \mathbb{V}\mathrm{ar}(X)$.

Attention : Toutes les variables aléatoires n'ont pas forçément une variance finie...

5.1.3 Indépendance et espérance

Proposition 5.7. *Soient* X *et* Y *deux v.a.r. indépendantes définies sur* $(\Omega, \mathcal{T}, \mathbb{P})$,

1. *Si* X *et* Y *admettent une moyenne, alors* XY *aussi et on a :*

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y),$$

2. *Si* $f(X)$ *et* $g(Y)$ *admettent une moyenne, alors* $f(X)g(Y)$ *aussi et on a :*

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

★ **Exercice 5.8.** Soit $(X_n)_{n \geq 1}$ une suite de v.a.r. indépendantes et de même loi. On note $m = \mathbb{E}(X_1)$ et $\sigma^2 = \mathbb{V}\mathrm{ar}(X_1)$. Soit N une v.a. entière, indépendante des X_n. On note $S_N = \sum_{k=1}^{N} X_k$, avec la convention $S_0 = 0$. Calculer $\mathbb{E}(S_N)$ et $\mathbb{V}\mathrm{ar}(S_N)$.

5.1.4 Intérêt de l'espérance, la variance et les moments

Outre le fait que l'espérance et la variance, ainsi que les moments d'ordre supérieurs permettent d'évaluer certains caractères descriptifs et dispersifs des variables aléatoires, il sont souvent pratiques pour définir une loi. En effet, lorsqu'on veux décrire la loi d'une variable aléatoire, on doit a priori donner soit sa fonction de densité pour les v.a.r. à densité, soit la valeurs de de toutes les probabilité $\mathbb{P}(X = k)$ pour les variables discrètes. Or il se trouve que beaucoup de "lois classiques" peuvent-être paramétrées par leur espérance et leur variance. On décrira souvent les lois avec des formulee du type : X suit une loi Machin d'espérance μ et de variance σ^2.

★ **Exercice 5.9.** Calcul (si existence) d'espérances et moments de v.a. usuelles (voir Chapitre 3).

★ **Exercice 5.10.** Au casino, une machine à sous fonctionne de la manière suivante : on introduit une pièce de 2 euros et trois roues tournent ; chaque roue s'arrête sur un chiffre de 0à 9 aléatoirement. Si les trois chiffres sont différents, le joueur perd sa mise, s'il y a un "double", celui-ci gagne 4 euros et enfin s'il obtient un "triple", il gagne x euros.

Quelle est la valeur limite de x pour laquelle le jeu est favorable à la banque ?

★ **Exercice 5.11.** Soit une variable aléatoire X de densité $f_X(x) = \frac{1}{2}\exp(-|x|)$ sur $\mathbb{R}$. Soit z un nombre réel et soit $g(x) = \exp(zx)$. Pour quelles valeurs de z la v.a. $Y = g(X)$ a-t-elle une espérance ? La calculer quand elle existe.

★ **Exercice 5.12.**

1. Soient X, Y deux v.a.r. Montrer que :
$$\mathbb{V}\mathrm{ar}(XY) = \mathbb{V}\mathrm{ar}(X)\mathbb{V}\mathrm{ar}(Y) + \mathbb{V}\mathrm{ar}(X)(\mathbb{E}(Y))^2 + \mathbb{V}\mathrm{ar}(Y)(\mathbb{E}(X))^2.$$
2. Soit une suite $(X_n)_{n\geq 0}$ de v.a.r. toutes de même variance σ^2. On note $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Montrer que : $\mathbb{V}\mathrm{ar}(\overline{X}) = \frac{\sigma^2}{n}$.

★ **Exercice 5.13. Autres expressions pour l'espérance**

1. Soit X une v.a. à valeurs dans $\mathbb{N}$. Montrer que son espérance est donnée par la formule suivante :
$$\mathbb{E}(X) = \sum_{k\geq 0} \mathbb{P}(X > k).$$
2. Soit X une v.a.r. d'espérance finie. Montrer que :
$$\mathbb{E}(X) = \int_0^{+\infty} \mathbb{P}(X > t)dt - \int_{-\infty}^0 \mathbb{P}(X < t)dt.$$
3. Soit X une v.a.r. positive, de fonction de répartition F et de densité f. Montrer que, pour tout $n \geq 1$,
$$\mathbb{E}(X^n) = \int_0^{+\infty} nx^{n-1}(1 - F(x))dx.$$
Montrer par un exemple que l'hypothèse de positivité de X est nécessaire.

4. Soit X une v.a.r. intégrable, de fonction de répartition F. Etablir que, pour tout réel a,

$$\mathbb{E}(|X-a|) = \int_a^{+\infty} (1-F(t))dt + \int_{-\infty}^{a} F(t)dt.$$

5. Soit X une v.a.r. intégrable, de fonction de répartition F, et de densité f. Montrer que le minimun de la fonction $a \mapsto \mathbb{E}(|X-a|)$ est atteint pour toute valeur de a pour laquelle $F(a) = \frac{1}{2}$. Comment s'interprète ces valeurs ?

L'exercice suivant est à la base de nombreux raisonnements de preuves, notamment dans les problèmes de convergence de suites de variables aléatoires.

★ **Exercice 5.14.** Soit X une v.a.r. positive de variance finie. Montrer que quelque soit $t \in \mathbb{R}$, on a :

$$\mathbb{E}(X) \leq t + \mathbb{E}(X \cdot 1_{\{X \geq t\}}).$$

5.1.5 Fonction caractéristique d'une variable aléatoire

Définition 5.15. Soit $X : (\Omega, \mathcal{T}, \mathbb{P}) \to \mathbb{R}^d$ une v.a. On appelle **fonction caractéristique** de X la fonction ϕ_X définie comme suit :

$$\forall t \in \mathbb{R}^d,\ \phi_X(t) := \mathbb{E}(e^{i<X,t>}) = \int_{\mathbb{R}^d} e^{i<x,t>} \mathrm{d}\mathbb{P}_X(x) = \int_\Omega e^{i<X(w),t>} \mathrm{d}\mathbb{P}(w).$$

La fonction caractéristique est un outil issu de l'analyse harmonique. On va voir qu'il y a, dans le cas des v.a. à densité, un lien étroit entre fonction caractéristique et transformée de Fourier de la fonction de densité.

La fonction caractéristique sert à résoudre de nombreux problèmes relatifs aux lois de probabilité, notamment des problème de convergence. C'est en particulier l'outil principal du théorème central limite...

Remarques 5.16.

1. Si X est une variable aléatoire réelle,

$$\phi_X(t) = \mathbb{E}\left(e^{itX}\right) = \mathbb{E}(\cos(tX)) + i\mathbb{E}(\sin(tX)).$$

2. Si X est une variable aléatoire entière, alors :

$$\phi_X(t) = \sum_{n \geq 0} \mathbb{P}(X=n))e^{int}.$$

3. Si X est une variable aléatoire réelle à densité f_X, alors :

$$\phi_X(t) = \int_{\mathbb{R}} f_X(x) e^{itx} \mathrm{d}x.$$

Proposition 5.17. Propriétés des la fonction caractéristique d'une variable aléatoire
Soit $X : (\Omega, \mathcal{T}, \mathbb{P}) \to \mathbb{R}^d$ une v.a. et soit ϕ_X sa fonction caractéristique.

1. $\phi_X(0) = 1$ *et* $\forall u \in \mathbb{R}^d$, $|\phi_X(u)| \leq 1$.
2. *La fonction* ϕ_X *uniformément continue,*
3. $\forall u \in \mathbb{R}^d$, $\phi_{-X}(u) = \phi_X(-u) = \overline{\phi_X(u)}$.
   Ainsi, ϕ_X *est à valeur réelles si* X *et symétrique et dans ce cas-là,* ϕ_X *est paire.*
4. $\forall a \in \mathbb{R}$, $\forall b \in \mathbb{R}^d$, $\forall u \in \mathbb{R}^d$, $\phi_{aX+b}(u) = e^{i<u,b>}\phi_X(au)$.
5. *si* $X \in L^1(\Omega)$, ϕ_X *tends vers* 0 *à l'infini.*

★ **Exercice 5.18.** Calculer la fonction caractéristique d'une v.a.r. X lorsqu'elle suit les lois suivantes :

1. Loi binomiale $\mathcal{B}(n, p)$,
2. Loi de Poisson $\mathcal{P}(\lambda)$,
3. Loi uniforme sur $[a, b]$,
4. Loi de Cauchy $\mathcal{C}(a)$ (de densité $x \mapsto \frac{a}{\pi}\frac{1}{x^2+a^2}$ sur $\mathbb{R}$),
5. Loi de Laplace $\mathcal{L}(a)$ (de densité $x \to \frac{a}{2}\exp(-a|x|)$ sur $\mathbb{R}$).

★ **Exercice 5.19.** Calcul de la fonction caractéristique d'une v.a.r. de loi $\mathcal{N}(m, \sigma^2)$.
On note ϕ_{m,σ^2} la f.c. d'une d'une v.a. de loi $\mathcal{N}(m, \sigma^2)$ et ϕ la f.c. d'une d'une v.a. de loi $\mathcal{N}(0, 1)$.

1. Exprimer ϕ_{m,σ^2} en fonction de ϕ.
2. Calculer ϕ'.
3. En intégrant par partie l'expression de ϕ', trouver une équation différentielle vérifiée par ϕ,
4. Conclure.

Remarque 5.20. v.a. à densité et fonction caractéristique.

Dans le cas d'une variable aléatoire à densité, la fonction ϕ_X est en fait la **transformée de Fourier** de la mesure $\mathbb{P}_X$. La transformé de Fourier $\mu \mapsto \hat{\mu}$ agit sur l'ensemble des mesures de $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ et est définie par :

$$\hat{\mu}(u) = \int_{\mathbb{R}^d} e^{i<x,u>} \mathrm{d}\mu(x).$$

Selon la convention qui est choisie pour la transformée de Fourier, il y a un facteur 2π ou encore un facteur -1 en plus qui apparait. Probablement pour cette raison, il arrive que l'on choisisse une convention différente, à savoir $\phi_X(u) = \mathbb{E}\left(e^{2i\pi<u,X>}\right)$ ou $\phi_X(u) = \mathbb{E}\left(e^{-<u,X>}\right)$. Ce changement de convention n'influence pas les résultats qui vont nous interesser, notamment ceux d'injectivité et d'inversibilité.

L'avantage, c'est surtout que la transformée de Fourier est un opérateur injectif, ainsi, 2 mesures finies ayant la même transformée de Fourier sont égales, d'où le résultat suivant :

Proposition 5.21. *Soient* $X : (\Omega, \mathcal{T}, \mathbb{P}_1) \to \mathbb{R}^d$ *et* $X : (\Omega', \mathcal{T}', \mathbb{P}_2) \to \mathbb{R}^d$ *deux v.a. et* ϕ_X *et* ϕ_Y *leurs fonctions caractéristiques.*

$$\phi_X = \phi_Y \iff X \text{ et } Y \text{ suivent la même loi de probabilité.}$$

Une autre conséquence due au lien avec la transformée de Fourier concerne les v.a. à densité. En effet, on peut (son condition d'intégrabilité) inverser la transformée de Fourier, d'où le résultat suivant :

Proposition 5.22. *Soit $X : (\Omega, \mathcal{T}, \mathbb{P}) \to \mathbb{R}$ une v.a. et ϕ_X sa fonction caractéristique.*

Si ϕ_X est intégrable, alors X est une v.a. à densité et sa densité f_X est presque surement donnée par la formule :

$$f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} \phi_X(u) \mathrm{d}u.$$

5.2 Des inégalités bien pratiques...

Proposition 5.23. Inégalité de Markov

Soit X une variable aléatoire admettant un moment d'ordre p.

$$\forall t \geq 0,\ \mathbb{P}(|X| \geq t) \leq \frac{E(|X|^p}{t^p}.$$

Proposition 5.24. Inégalité de Bienaymé-Tchebichev
Soit X une variable aléatoire d'espérance et de variance finies.

$$\forall t \geq 0,\ \mathbb{P}(|X - E(X)| \geq t) \leq \frac{Var(X)}{t^2}.$$

Cette dernière inégalité est très utilisée (voir Exercice 5.27 et Exercice 5.28).

Remarques 5.25. Lors de certaines démonstration, on utilise souvent (pour obtenir des majorations) le résultat suivant qui est une conséquence directe de l'inégalité de Markov :

Si X a une espérance finie, alors :

$$\forall \epsilon > 0, \exists M > 0,\ \mathbb{P}(X \geq M) \leq \epsilon.$$

En fait, d'autres inégalités issues de l'intégration peuvent être utilisées en probablités :

- **Inégalité de Jensen**
  Soient X une v.a.r. intégrable et $\phi : \mathbb{R} \to \mathbb{R}$ convexe et mesurable. On a :

$$\phi\left(\mathbb{E}(X)\right) \leq \mathbb{E}\left(\phi(X)\right).$$

- **Inégalité de Cauchy-Schwarz**

  Soient X et Y deux v.a.r. de carré intégrable. On a :

$$XY \text{ est une v.a. intégrable et } \mathbb{E}(XY) \leq \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}.$$

- **Inégalité de Minkovski**
  Soient X et Y deux v.a.r. admettant des moments d'ordre p. On a :

$$X + Y \text{ admet un moment d'ordre } p \text{ et } \mathbb{E}\left((X+Y)^p\right) \leq \left(\mathbb{E}(X^p)^{\frac{1}{p}} + \mathbb{E}(Y^p)^{\frac{1}{p}}\right)^p.$$

★ **Exercice 5.26.**

1. Montrer l'inégalité de Markov en utilisant la partition suivant de Ω :
$$\Omega = \{|X| \geq t\} \cup \{|X| \leq t\}.$$
2. En utilisant l'inégalité de Markov, montrer l'inégalité de Bienaymé-Tchebichev.
3. En utilisant l'inégalité de Markox, montrer l'**inégalité de Bernstein** décrite ci-dessous : Soient X une v.a.r. et $f : \mathbb{R} \to \mathbb{R}_+$ croissante telle que $f(X)$ admet une espérance finie.
$$\forall t \geq 0,\ \mathbb{P}(X \geq t) \leq \frac{E(f(X))}{f(t)}.$$

★ **Exercice 5.27. Application de l'inégalité de Bienaymé-Tchebichev I**

Soit $(X_n)_{n\geq 0}$ une suite de v.a. r. indépendantes et de même loi, admettant une esperance μ et une variance σ^2. On pose, pour tout $n \geq 0$
$$Y_n = \frac{(\sum_{k=1}^{n} X_k) - n\mu}{\sigma\sqrt{n}}.$$

1. Majorer, à l'aide de l'inégalité de Bienaymé-Chebichev, la quantité $\mathbb{P}(|Y_n| \geq k)$ pour $k \geq 0$,
2. Utiliser ce résultat dans le problème suivant :
Une montre subit des écarts quotidiens (positifs ou négatifs) que l'on suppose indépendants d'un jour à l'autre et qui suivent tous la même loi de moyenne nulle et de variance 16 secondes. En supposant la montre bien réglée au départ, déterminer un minorant la probabilité que l'écart sur une année (365 jours) soit inférieur à deux minutes. Conclusion ?

★ **Exercice 5.28. Application de l'inégalité de Bienaymé-Tchebichev II**

On dispose d'une pièce équilibrée.

1. On lance 100 fois la pièce. Majorer, à l'aide de l'inégalité de Bienaymé-Chebichev, la probabilité qu'avoir plus de 70 fois Face ou moins de 30 fois Face à l'issue des tirages, puis calculer cette probabilité explicitement. Commentez.
2. Estimer le nombre de lancers necessaires pour que la fréquence de Pile observée au Pile ou Face soit comprise entre 0.4 et 0.5, avec une probabilité au moins égale à 0.9.

★ **Exercice 5.29.** La **fonction génératrice d'une v.a. entière** est définie par
$$g_X(t) := E\left(t^X\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(X = n)t^n.$$

1. Donner la fonction génératrice d'une v.a.d. de loi $\mathcal{P}(\lambda)$ et d'une v.a.d. de loi $\mathcal{G}(p)$.
2. Justifier que g_X est une série entière de rayon de convergence au moins égal à 1 vérifiant $g_X(0) = \mathbb{P}(X = 0)$, $g_X(1) = 1$ et $\forall t \in [0, 1],\ g_X(t) \in [0, 1]$.
3. Montrer que, si X admet une espérance finie, alors g_X est dérivable en 1 et $g'_X(1) = \mathbb{E}(X)$.
4. Montrer que, si X admet un moment d'ordre $n \geq 2$ alors g_X est n fois dérivable en 1 et on a l'égalité :
$$g_X^{(n)}(1) = E\left(X(X-1)\cdots(X-(n-1))\right),$$
et en particulier, on a $var(X) = g'_X(1)\left(1 - g'_X(1)\right) + g''_X(1)$.

Chapitre 6

Couples de variables aléatoires

6.1 Lois conjointes et lois marginales

Ce paragraphe regroupe le vocabulaire usuel utilisé pour les couples de variables aléatoires et les vecterus aléatoires.

Définition 6.1. Soient $X, Y : (\Omega, \mathcal{T}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ deux v.a.r.

La **loi conjointe de X et Y** : $\mathbb{P}_{(X,Y)}$ est la probabilité image dans $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ de $\mathbb{P}$ par le couple (X, Y) :

$$\forall A \in \mathcal{B}(\mathbb{R}^2)),\ \mathbb{P}_{(X,Y)}(A) = \mathbb{P}((X, Y) \in A).$$

De même, si $X = (X_1, X_2, \ldots, X_n)$ est un vecteur aléatoire ($\vec{\text{v}}$.a.r.), la **loi conjointe des X_i** est la loi de X : $\mathbb{P}_X$, probabilité image dans $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ de $\mathbb{P}$ par X.

Cas des v.a. discrètes :

La loi conjointe de X et Y deux v.a.d. peut être donnée par :

$$\{\mathbb{P}(X = x, Y = y) \mid (x, y) \in X(\Omega) \times Y(\Omega)\},$$

Cas des v.a.r. à densité

La loi conjointe de X et Y deux v.a. peut être donnée par leur **densité conjointe**, c'est-à-dire une $f_{(X,Y)} : \mathbb{R}^2 \to \mathbb{R}_+$ telle que :

$$\int_{\mathbb{R}^2} f_{(X,Y)}(x, y)\mathrm{d}x\mathrm{d}y = 1.$$

Comme d'habitude, on peut se restreindre à donner les valeurs de la loi conjointe d'un couple ou d'un vecteur uniquement sur une classe qui genère la tribu d'arrivée.

Définition 6.2. Soient $X, Y : (\Omega, \mathcal{T}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ deux v.a.r.

On définit la **fonction de répartition d'un couple de v.a.r** comme étant la fonction suivante :

$$\begin{array}{rccl} F_{(X,Y)} : & \mathbb{R}^2 & \to & [0,1] \\ & (x, y) & \mapsto & \mathbb{P}(X \leq x, Y \leq y), \end{array}$$

autrement dit, $F_{(X,Y)}(x,y) = \mathbb{P}_{(X,Y)}\left(]-\infty,x]\times]-\infty,y]\right)$.

On peut de la même manière définir la **fonction de répartition d'un $\vec{\text{v}}$ar** $X = (X_1, X_2, \ldots, X_n)$ par la formule :

$$\begin{array}{rccl} F_X : & \mathbb{R}^n & \to & [0,1] \\ & (x_1, x_2, \ldots, x_n) & \mapsto & \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) = \mathbb{P}_X\left(\prod_{i=1}^n]-\infty, x_i]\right), \end{array}$$

autrement dit, $F_X(x_1, x_2, \ldots, x_n)\mathbb{P}_X\left(\prod_{i=1}^n]-\infty, x_i]\right)$.

Remarque : si un couple de v.a.r. (X,Y) admet pour densité conjointe $f_{(X,Y)} : \mathbb{R}^2 \to \mathbb{R}_+$, alors, en tout point de continuité de $f_{(X,Y)}$, on a :

$$f_{(X,Y)}(x,y) = \frac{\partial^2 F_{(X,Y)}}{\partial x \partial y}(x,y).$$

Définition 6.3. Soit $X = (X_1, X_2, \ldots, X_n)$ un $\vec{\text{v}}$.a.r. de $(\Omega, \mathcal{T}, \mathbb{P})$ vers $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.
La **loi marginale de** X_i est la probabilité $\mathbb{P}_{X_i}$ définie sur $\mathcal{B}(\mathbb{R})$ par :

$$\forall A \in \mathcal{B}(\mathbb{R}),\ \mathbb{P}_{X_i}(A) = \mathbb{P}(X_i \in A) = \mathbb{P}(X \in \mathbb{R}^{i-1} \times A \times \mathbb{R}^{n-i}).$$

Exemples 6.4. Par exemple, pour un couple de v.a. X et Y de loi conjointe donnée :

- **Cas des v.a. discrètes :**
  La loi marginale de X peut être donnée par :

$$\mathbb{P}(X = x) = \sum_{y \in \mathbb{R}} \mathbb{P}(X = x, Y = y) = \lim_{y \to +\infty} \left(F_{(X,Y)}(x,y) - \lim_{t \to x^-} F_{(X,Y)}(t,y) \right).$$

- **Cas des v.a.r. à densité** si un couple de v.a.r. (X,Y) admet pour densité conjointe $f_{(X,Y)} : \mathbb{R}^2 \to \mathbb{R}_+$, alors X est une v.a.r. à densité et sa densité f_X est donnée par :

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x,y)\mathrm{d}y.$$

Rappelons les critères d'indépendance pour les couples de variables aléatoires :

Proposition 6.5. Critères d'indépendance de v.a.
On fixe $X : (\Omega, \mathcal{T}, \mathbb{P}) \to (\mathbb{R}^{d_X}, \mathcal{B}(\mathbb{R}^{d_X}))$ et $Y : (\Omega, \mathcal{T}, \mathbb{P}) \to (\mathbb{R}^{d_Y}, \mathcal{B}(\mathbb{R}^{d_Y}))$ deux $\vec{\text{v}}$.a.r.

1. **Si X et Y sont discrètes**.
   X et Y sont indépendantes ssi :

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

2. **Si X et Y sont de densité respectives f_X et f_Y.**
   Si X et Y sont indépendantes, alors la v.a. (X,Y) admet une densité $f_{(X,Y)}$, produit direct de f_X et f_Y :

$$\forall x \in \mathbb{R}^{d_X}, \forall y \in \mathbb{R}^{d_Y},\ f_{(X,Y)}(x,y) = f_X(x) f_Y(y).$$

Inversement, si la v.a. (X,Y) admet une densité $f_{(X,Y)}$, produit direct dedeux fonctions intégrables g_X et g_Y comme suit :

$$\forall x \in \mathbb{R}^{d_X}, \forall y \in \mathbb{R}^{d_Y},\ f_{(X,Y)}(x,y) = g_X(x)g_Y(y).$$

alors g_X et g_Y sont, à un facteur positif près, près les densités respectives de X et Y et ces deux v.a. sont indépendantes.

★ **Exercice 6.6.** Soit (X,Y) un vecteur aléatoire de densité $f(x,y) = \frac{1}{4}(1+xy)\cdot 1_{[-1,1]^2}$ et soit (U,V) un vecteur aléatoire de densité $g(x,y) = \frac{1}{4}\cdot 1_{[-1,1]^2}$.

1. Vérifier que f et g sont bien des densités de probabilités,
2. trouver les densités des lois marginales f_X, f_Y, g_U et g_V.

6.2 Covariance

Définition 6.7. Soient $X, Y \in L^2(\Omega)$.

On appelle **covariance** de X et Y la quantité suivante :

$$\mathbb{C}\mathrm{ov}(X,Y) = E\left((X-E(X))(Y-E(Y))\right)$$

Remarque : D'après l'inégalité de Schwarz, XY et $(X-E(X))(Y-E(Y))$ sontbien dans $L^1(\Omega)$...

Proposition 6.8. *$X, Y \in L^2(\Omega)$. On a :*

$$\mathbb{C}\mathrm{ov}(X,Y) = E(XY) - E(X)E(Y) \quad \text{et} \quad Var(X+Y) = Var(X) + Var(Y) + 2\mathbb{C}\mathrm{ov}(X,Y).$$

Ces formules proviennent juste de la linéarité de l'espérance. La deuxième formule est l'analogue de $(x+y)^2 = x^2+y^2+2xy$... En fait, la plupart des propriétés de la covariance sont analogues à celles du produit de deux réels ou du produit scalaire de deux vecteurs :

★ **Exercice 6.9.** Montrer que la covariance est une forme bilinéaire symétrique positive sur l'espace vectoriel $L^2(\Omega)$ des v.a. et que la forme quadratique associée est la variance.

Définition 6.10. Soient $X, Y \in L^2(\Omega)$.

Si $\mathbb{C}\mathrm{ov}(X,Y) = 0$, on dit que X et Y sont des v.a.r. **non corrélées**.

Proposition 6.11. *Si X et Y sont 2 v.a. indépendantes qui admettent un moment d'ordre 2, alors XY aussi et on a :*

$$\mathbb{C}\mathrm{ov}(X,Y) = 0 \qquad \text{et } \mathbb{V}\mathrm{ar}(X+Y) = \mathbb{V}\mathrm{ar}(X) + \mathbb{V}\mathrm{ar}(Y),$$

Attention ! si $\mathbb{C}\mathrm{ov}(X,Y) = 0$ cela ne veut PAS dire que X et Y sont indépendantes !

★ **Exercice 6.12.** Soient X une v.a. de loi $\mathcal{N}(0,1)$ et Y uniformément distribuée sur $[-1,1]$. On note $Z = YX$.

1. Montrer que Z suit la loi $\mathcal{N}(0,1)$.
2. Calculer la covariance de X et Z.
3. Calculer $\mathbb{P}(X+Z=0)$ et en déduire que les variables X et Z ne sont pas indépendantes.

6.3 Vecteurs aléatoires

Définition 6.13. On définit la notion de **matrice aléatoire réelle** par la donnée de $M = (X_{i,j})$ où les $X_{i,j}$ sont des v.a.r. Un **vecteur aléatoire réel** ($\vec{\text{v}}$.a.r.) de dimension d est une matrice aléatoire de taille $d \times 1$.

Définition 6.14. Une matrice aléatoire réelle $M = (X_{i,j})$ où les $X_{i,j}$ sont des v.a.r. est dit **intégrable** (resp. de **carré intégrable**) si et seulement toutes ses composantes $X_{i,j}$ sont dans $L^1(\Omega)$.

Si M est intégrable, on appelle **espérance de la matrice** la matrice $\mathbb{E}(M) = (\mathbb{E}(X_{i,j}))$.

Proposition 6.15. *Soit $X : (\Omega, \mathcal{T}, \mathbb{P}) \to \mathbb{R}^d$ un $\vec{\text{v}}$.a.r. intégrable.*

1. *$\|X\| \in L^1(\Omega)$, quelque soit la norme choisie sur $\mathbb{R}^d$,*
2. *Si A est une matrice à coefficients réels ou complexes, AX est également intégrable et pour tout vecteur B de $\mathbb{R}^d$,*
$$\mathbb{E}(AX + B) = A\mathbb{E}(X) + B$$

.

Définition 6.16. Matrice de co-variance

Soient $X : (\Omega, \mathcal{T}, \mathbb{P}) \to \mathbb{R}^d$ et $Y : (\Omega, \mathcal{T}, \mathbb{P}) \to \mathbb{R}^d$, deux $\vec{\text{v}}$.a.r. de carré intégable. On définit la matrice de covariance $(d \times d)$ comme suit :

$$Cov(X, Y) = E\left((X - E(X))\ {}^t(Y - E(Y))\right).$$

La **matrice d'auto-covariance** du $\vec{\text{v}}$.a.r. X est donnée par :

$$K_X = Cov(X, {}^tX)$$

C'est une matrice symétrique et semi-définie positive. Elle est définie positives si les composantes de X sont linéairement indépendantes.

6.4 Changement de variables aléatoires

Proposition 6.17. *Soit X une v. a. à valeurs dans $\mathbb{R}^d$ de densité f_X, et $\phi : \mathbb{R}^d \to \mathbb{R}^d$ un difféomorphisme. La v.a. $Y = \phi \circ X$ admet une densité f_Y donnée par :*

$$\forall y \in \mathbb{R}^d,\ f_Y(y) = |det\left(J_{\phi^{-1}}(y)\right)| f_x(\phi^{-1}(y)).$$

où $det\left(J_{\phi^{-1}}\right)$ est le jacobien de ϕ^{-1}. La matrice $J_{\phi^{-1}}(y) = \left(\frac{\partial(\phi^{-1})_i}{\partial x_j}(y)\right)_{i,j}$ est la matrice des dérivées partielles de ϕ^{-1}.

Ce thèorème est très utilisé dans le cas ou les composantes de X sont des v.a. indépendantes. Ci-dessous, plusieurs exercices d'application.

On retrouve souvent ce théorème dans des applications du type : Soit (X, Y) deux v.a.r. de lois données. On pose $(U, V) = \phi(X, Y)$. Donner la loi conjointe de U et V. A-t-on $U \perp V$?

Ce théorème permet notamment de montrer la proposition suivante :

Proposition 6.18. Densité de la somme de 2 v.a.r. indépendantes

Soient X et Y deux v.a.r. indépendantes admettant des densités notées respectivement f_X et f_Y.

La v.a.r. $X+Y$ admet pour densité la fonction $f_{X+Y} = f_X * f_Y$, où $*$ est le produit de convolution :

$$\forall s \in \mathbb{R}, f_{X+Y}(s) = \int_{\mathbb{R}} f_X(t) f_Y(s-t) dt.$$

★ **Exercice 6.19.** Un couple de v.a.r (X,Y) admet pour densité la fonction $f(x,y) = 2 \cdot 1_\Delta(x,y)$ par rapport à la mesure de Lebesgue, où Δ est le triangle de sommets $(0,0)$, $(0,1)$ et $(1,0)$.

1. Donner les lois des v.a. suivantes : X, Y, $X+Y$, $X-Y$ $\frac{Y}{1-X}$.
2. Montrer que X et $\frac{Y}{1-X}$ sont indépendantes et donner la loi de $\frac{Y}{1-X}$.

★ **Exercice 6.20.** Autour de la loi exponentielle...

1. Soient X et Y 2 v.a.r. indépendantes de loi $Exp(\lambda)$. Calculer la loi du couple $(\frac{X}{Y}, X^2+Y^2)$ et la loi de $X+Y$.
2. Soient X_i, pour $i = 1, 2, \ldots, n$ une séquence de v.a.r. indépendantes de lois respectives $Exp(\lambda_i)$. Montrer que $\inf_{i=1}^n X_i$ suit aussi une loi exponentielle dont on déterminera le paramètre.

★ **Exercice 6.21. Coordonnées aléatoires cartésiennes et polaires**

Soient X et Y deux v.a. i.i.d. de loi normale $\mathcal{N}(0,1)$.
On pose $(X,Y) = (R\cos\theta, R\sin\theta)$ pour $R \geq 0$ et $\theta \in [0, 2\pi]$.
Determiner la loi du couple (R, θ).

★ **Exercice 6.22. Transformations de la loi normale**.

Soit X une v.a. de loi normale $\mathcal{N}(m, \sigma^2)$.

1. On definit la v.a. Y par $\log Y = X$ (on dit que Y suit une loi log-normale de parametres m et σ^2). Donner la densité de probabilité de Y, son espérance et sa variance.
2. Si X est centrée et réduite ($m = 0$ et $\sigma = 1$), on définit $Z = X^2$ sur $\mathbb{R}^+$. Donner la densité de Z. On dit que Z suit une Chi-deux à un degré de liberté $\chi_2(1)$. Déterminer $\mathbb{E}(Z)$ et $\mathbb{V}\text{ar}(Z)$.

En pratique, souvent, on se sert de ce théorème pour simuler des v.a. à densité à partir d'autres.

★ **Exercice 6.23.** Soient X et Y deux v.a.r. indépendantes. On suppose que
- la v.a. X a pour densité $f_X(x) = \frac{1}{\pi}(1-x^2)^{-\frac{1}{2}} \cdot 1_{]-1,1[}$
- la v.a. Y a pour densité $f_Y(y) = y \exp\left(\frac{-y^2}{2}\right) \cdot 1_{\mathbb{R}_+^*}$.

Montrer que XY suit une loi normale $\mathcal{N}(0,1)$.

★ **Exercice 6.24. Coordonnées aléatoires cartésiennes et polaires**

Soient X et Y deux v.a. i.i.d. de loi normale $\mathcal{N}(0,1)$.
On pose $(X,Y) = (R\cos\theta, R\sin\theta)$ pour $R \geq 0$ et $\theta \in [0, 2\pi]$.
Determiner la loi du couple (R, θ).

★ **Exercice 6.25. Simulation de la loi normale**

Soient U_1 et U_2 deux v.a. i.i.d. de loi uniforme sur $[0,1]$.

1. Calculer la loi de $V = -2 \ln U_1$,
2. Calculer la loi de $R = \sqrt{V}$,
3. Calculer la loi de $\theta = 2\pi U_2$,
4. Calculer la loi du couple de v.a. (X, Y) défini par :$X = R\cos\theta$ et $Y = R\sin\theta$.
5. Conclure sur les lois (et l'indépendance) de X et Y.

★ **Exercice 6.26.** On note U une v.a. de loi uniforme sur $[0, 1]$.

1. Soient x_0, x_1,... x_n des nombres réels quelconques et p_0 , p_1 ,... p_n des nombres réels de $[0, 1]$ tels que $\sum_{i=1}^{n} p_i = 1$. On définit $q_0 = 0$ et pour tout $i = 1, \ldots, n$, $q_i = \sum_{j=1}^{i} p_j$.
   Déterminer la loi de la v.a. Z, définie de la manière suivante :
$$Z = \sum_{i=1}^{n} x_i 1_{]q_i, q_{i+1}]}(U).$$
2. Soit $\lambda > 0$. Déterminer la loi de la v.a. Y définie par $Y = -\frac{1}{\lambda}\ln(1 - U)$.
3. Sur un ordinateur, on dispose d'une fonction *Random* qui tire au hasard un nombre dans l'intervalle $[0, 1]$. Ecrire une procedure fournissant en sortie la réalisation d'une v.a. de binomiale. Même question pour une loi exponentielle.

Chapitre 7

Suites de variables aléatoires

7.1 Les différents types de convergence

Les variables aléatoires sont des fonctions mesurables. Lorsqu'on considère des suites de variables aléatoires, ce sont donc des suites de fonctions... et comme pour les suites de fonctions, nous avons donc plusieurs mode de convergence : convergence presque sûre, convergence dans L^p, en mesure (en probabilité)... mais les variables aléatoires étant caractérisées par leurs lois de probabilité, c'est-à-dire des mesures, on peut envisager la convergence des mesures elles-même : c'est la convergence en loi.

Ce paragraphe présente ces différents modes de convergence ainsi que les liens entre celles-ci et donne quelques résultats classiques bien utiles pour montrer les convergences de suites de variables aléatoires.

7.1.1 Convergence $\mathbb{P}$-presque sûre ($\mathbb{P}$-p.s.)

Définition 7.1. Soit $(X_n)_{n\in\mathbb{N}}$ et X des $\vec{\text{v}}$.a.r. de dimension d définis sur $(\Omega, \mathcal{T}, \mathbb{P})$.

On dit que $(X_n)_{n\in\mathbb{N}}$ **converge $\mathbb{P}$-presque sûrement (ou converge $\mathbb{P}$-p.s.) vers** X si la suite $(X_n(\omega))_{n\in\mathbb{N}}$ converge vers $X(\omega)$ pour $\mathbb{P}$-presque tout $\omega \in \Omega$:

$$\boxed{X_n \xrightarrow[n\to\infty]{\mathbb{P}\text{-p.s.}} X \quad \Longleftrightarrow \quad \mathbb{P}\left(\left\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1.}$$

Remarques 7.2. Dans la deuxième partie de l'équivalence, la convergence de $X_n(\omega)$ vers $X(\omega)$ est la convergence simple dans $\mathbb{R}^d$: Dire qu'un v.a. converge $\mathbb{P}$-p.s. c'est la dire qu'on a la convergence simple de la suite de fonctions $(X_n)_{n\in\mathbb{N}}$ vers X sur un ensemble de mesure totale.

D'autre part, la convergence $\mathbb{P}$-p.s. est seul type de convergence qui fait intervenir "explicitement" Ω. Ce sera de ce fait le type de convergence le plus difficile à montrer (en général).

7.1.2 Convergence en probabilité ($\mathbb{P}$)

Définition 7.3. Soit $(X_n)_{n\in\mathbb{N}}$ et X des $\vec{\text{v}}$.a.r. de dimension d définis sur $(\Omega, \mathcal{T}, \mathbb{P})$.

On dit que $(X_n)_{n\in\mathbb{N}}$ **converge en probabilité vers** X si pour tout $\epsilon > 0$, la probabilité de l'évènement $\{|X_n - X| > \epsilon\}$ tend vers 0 :

$$\boxed{X_n \xrightarrow[n\to\infty]{\mathbb{P}} X \quad \Longleftrightarrow \quad \lim_{n\to\infty} \mathbb{P}\left(|X_n - X| > \epsilon\right) = 0.}$$

Remarque 7.4. Cette convergence est en fait la convergence dite "en mesure" qui peut être définie pour n'importe quelle mesure sur $(\Omega, \mathcal{T})$.

Méthode : Pour montrer une convergence en probabilité, on utilisera souvent des inégalités du type inégalité de Markov, de Bienaymé-Chebichev ou de Bernstein.

7.1.3 Convergence en moyenne d'ordre p (L^p)

Définition 7.5. Soit $(X_n)_{n\in\mathbb{N}}$ et X des $\vec{\text{v}}$.a.r. de dimension d définis sur $(\Omega, \mathcal{T}, \mathbb{P})$.

On dit que $(X_n)_{n\in\mathbb{N}}$ **converge en moyenne d'ordre p (ou converge dans L^p) vers** X si la v.a. $|X_n - X|$ converge dans $L^p(\Omega, \mathcal{T}, \mathbb{P})$ vers 0 :

$$\boxed{X_n \xrightarrow[n\to\infty]{L^p} X \quad \Longleftrightarrow \quad \lim_{n\to\infty} \mathbb{E}\left(|X_n - X|^p\right) = 0.}$$

Lorsque $p = 1$, on parle simplement de **converge en moyenne** et lorsque $p = 2$, on parle souvent de **convergence en moyenne quadratique**.

Propriété 7.6. *Si* $X_n \xrightarrow[n\to\infty]{L^p} X$, *alors* $X_n \xrightarrow[n\to\infty]{L^q} X$ *pour tout* $q \leq p$.

7.1.4 Convergence en loi ($\mathcal{L}$)

Définition 7.7. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de $\vec{\text{v}}$.a.r. de dimension d et X un $\vec{\text{v}}$.a.r. de dimension d.

On dit que $(X_n)_{n\in\mathbb{N}}$ **converge en loi vers** X si la suite des fonctions de répartition $(F_{X_n})_{n\in\mathbb{N}}$ converge simplement vers la fonction de répartition F_X de X en tout point où cette dernière est continue :

$$\boxed{X_n \xrightarrow[n\to\infty]{\mathcal{L}} X \quad \Longleftrightarrow \quad \Big(F_X \text{ continue en } t \in \mathbb{R}^d \Longrightarrow \lim_{n\to\infty} F_{X_n}(t) = F_X(t)\Big).}$$

Remarque 7.8. Dire que X_n converge en loi vers X est équivalent à dire que X_n converge vers X au sens des distributions ou de manière équivalente, en terme de mesure, dire que la mesure $\mathbb{P}_{X_n}$ converge étroitement vers $\mathbb{P}_X$.

Remarque 7.9. Lorsqu'une suite de v.a.r. converge en loi, sa v.a. limite n'est pas unique... par contre la loi de la v.a. limite est unique. En effet, si $X_n \xrightarrow{\mathcal{L}} X$ et qu'une v.a. Y est de même loi que X, alors on peut écrire $X_n \xrightarrow{\mathcal{L}} Y$.

En ce fait, en pratique, on écrira souvent la limite sous forme de loi. Par exemple : $X_n \xrightarrow{\mathcal{L}} \mathcal{B}(p)$ ou $X_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$...

7.1.5 Outils pour montrer la convergence

Comme on l'a déjà mentionné, la convergence $\mathbb{P}$-p.s. est le seul mode de convergence qui fait apparaitre explicitement Ω mais cela peut être ennuyeux, surtout qu'on a rarement des informations très précises sur cet espace. Pour montrer la convergence $\mathbb{P}$-p.s. d'une suite de $\vec{v}$.a.r. , on utilise donc le lemme suivant, qui ramène l'étude de la convergence $\mathbb{P}$-p.s. à une étude de convergence en probabilité :

Lemme 7.10. Lemme fondamental de la convergence $\mathbb{P}$-p.s.
Soit $(X_n)_{n\in\mathbb{N}}$ des $\vec{v}$.a.r. de dimension d définis sur $(\Omega, \mathcal{T}, \mathbb{P})$.

$$X_n \xrightarrow[n\to\infty]{\mathbb{P}\text{-}p.s.} 0 \quad \Longleftrightarrow \quad \sup_{k\geq n} |X_k| \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Proposition 7.11. Critère pour la convergence $\mathbb{P}$-p.s.
Soit $(X_n)_{n\in\mathbb{N}}$ des $\vec{v}$.a.r. de dimension d définis sur $(\Omega, \mathcal{T}, \mathbb{P})$.

$$\forall \epsilon > 0, \sum_{n\in\mathbb{N}} \mathbb{P}(|X_n| > \epsilon) < +\infty \quad \Longrightarrow \quad X_n \xrightarrow[n\to\infty]{\mathbb{P}\text{-}p.s.} 0.$$

Remarque : Pour avoir la convergence en probabilité, il suffit que la suite des $\mathbb{P}(|X_n| > \epsilon)$ tende vers 0 : la convergence $\mathbb{P}$-p.s. est donc plus forte que la convergence en probabilité.

La démonstration de cette proposition repose sur le lemme fondamental et le lemme de Borel-Cantelli.

Les résultats d'intégration classiques fournissent des outils permettant de montrer la convergence en moyenne et les convergence L^p.

Proposition 7.12. Outils d'intégration pour la convergence L^p

1. **Théorème de la convergence monotone**
   Soit $(X_n)_{n\geq 0}$ une suite de v.a.r. positives.

   $$(X_n)_{n\geq 0} \nearrow, (X_n)_{n\geq 0} \xrightarrow{\mathbb{P}\text{-}p.s.} X \implies X \text{ est une v.a.r., et } \lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X).$$

   Les espérances mises en jeu ici peuvent être infinies...

2. **Théorème de la convergence dominée**
   Soit $(X_n)_{n\geq 0}$ une suite de v.a.r.

   $$\left.\begin{array}{l} (X_n)_{n\geq 0} \xrightarrow{\mathbb{P}\text{-}p.s.} X, \\ |X_n| \leq Y,\ Y \text{ v.a.r. intégrable} \end{array}\right\} \implies \lim_{n\to\infty} \mathbb{E}(X_n) = E(X) < +\infty.$$

3. **Lemme de Fatou**
   Soit $(X_n)_{n\geq 0}$ une suite de v.a.r. positives On a :

   $$\mathbb{E}\left(\liminf_{n\to\infty} X_n\right) \leq \liminf_{n\to\infty} \mathbb{E}(X_n).$$

4. **Lemme de Scheffé**
Soit $(X_n)_{n\geq 0}$ une suite de v.a.r. positives qui converge $\mathbb{P}$-p.s. vers une v.a. X. On a :

$$\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X) \iff \lim_{n\to\infty} \mathbb{E}(|X_n - X|) = 0.$$

Proposition 7.13. *Soit $(X_n)_{n\in\mathbb{N}}$ et X des $\vec{\text{v}}$.a.r. de dimension d définis sur $(\Omega, \mathcal{T}, \mathbb{P})$.*

1. **Critère de Cauchy pour la convergence $\mathbb{P}$-p.s.** *:*

$$X_n \xrightarrow[n\to\infty]{\mathbb{P}\text{-}p.s.} X \quad \iff \quad \forall \epsilon > 0,\ \mathbb{P}\left(\sup_{k\geq 0} |X_{n+k} - X_n| > \epsilon\right) \xrightarrow[n\to\infty]{} 0.$$

2. **Critère de Cauchy pour la convergence en probabilité** *:*

$$X_n \xrightarrow[n\to\infty]{\mathbb{P}} X \quad \iff \quad \forall \epsilon > 0,\ \sup_{k\geq 0} \left(\mathbb{P}\left(|X_{n+k} - X_n| > \epsilon\right)\right) \xrightarrow[n\to\infty]{} 0.$$

3. **Critère de Cauchy pour la convergence L^p** *:*

$$X_n \xrightarrow[n\to\infty]{L^p} X \quad \iff \quad \sup_{k\geq 0} \mathbb{E}\left(|X_{n+k} - X_n|^p\right) \xrightarrow[n\to\infty]{} 0.$$

Remarque : Le critère de Cauchy pour la convergence $\mathbb{P}$-p.s. est calqué sur le lemme fondamental de la convergence $\mathbb{P}$-p.s. Le critère de Cauchy pour la convergence L^p est calqué sur le lemme de Scheffé.

7.1.6 Opérations sur les v.a. et convergences

Proposition 7.14. *Soient $(X_n)_{n\in\mathbb{N}}$ une suite de $\vec{\text{v}}$.a.r. indépendants de dimension d définis sur $(\Omega, \mathcal{T}, \mathbb{P})$ et soit $f : \mathbb{R}^d \to \mathbb{R}^n$ une fonction continue.*

$$X_n \xrightarrow[n\to\infty]{\mathbb{P}} X \quad \Longrightarrow \quad f(X_n) \xrightarrow[n\to\infty]{\mathbb{P}} f(X).$$

$$X_n \xrightarrow[n\to\infty]{\mathbb{P}\text{-}p.s.} X \quad \Longrightarrow \quad f(X_n) \xrightarrow[n\to\infty]{\mathbb{P}\text{-}p.s.} f(X).$$

$$\left.\begin{array}{l} X_n \xrightarrow[n\to\infty]{L^p} X \\ f(X_n) \in L^P(\Omega) \\ f(X) \in L^P(\Omega) \end{array}\right\} \Longrightarrow f(X_n) \xrightarrow[n\to\infty]{L^p} f(X).$$

Une conséquence importante de théorème est la suivante :

Corollaire 7.15. *Soient $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ des suites de $\vec{\text{v}}$.a.r. indépendants de dimensions d définis sur $(\Omega, \mathcal{T}, \mathbb{P})$.*

Si $X_n \to X$ et $Y_n \to Y$ $\mathbb{P}$-p.s., $\mathbb{P}$ ou L^1, alors $X_n + Y_n \to X + Y$ avec le même mode de convergence.

Attention : Ce corollaire n'est pas valable pour la convergence en loi : Si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} -X$ où X suit une loi symétrique par rapport à 0, on n'aura PAS $X_n + Y_n \xrightarrow{\mathcal{L}} 0$... Pour les convergences $\mathbb{P}$ et $\mathbb{P}$-p.s., un autre corollaire de la Proposition 7.14 est aussi fréquemment utilisé :

Corollaire 7.16. *Soient $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ des suites de $\vec{\text{v}}$.a.r. indépendants de dimensions respectives d_X et d_Y définis sur $(\Omega, \mathcal{T}, \mathbb{P})$ et soit $f : \mathbb{R}^{d_X} \times \mathbb{R}^{d_Y} \to \mathbb{R}^n$ une fonction continue.*

$$\left.\begin{array}{l} X_n \xrightarrow[n\to\infty]{\mathbb{P}} X \\ Y_n \xrightarrow[n\to\infty]{\mathbb{P}} Y \end{array}\right\} \quad \Longrightarrow \quad f(X_n, Y_n) \xrightarrow[n\to\infty]{\mathbb{P}} f(X, Y).$$

$$\left.\begin{array}{l} X_n \xrightarrow[n\to\infty]{\mathbb{P}\text{-}p.s.} X \\ Y_n \xrightarrow[n\to\infty]{\mathbb{P}\text{-}p.s.} Y \end{array}\right\} \quad \Longrightarrow \quad f(X_n, Y_n) \xrightarrow[n\to\infty]{\mathbb{P}\text{-}p.s.} f(X, Y).$$

Remarque : Ce corollaire permet par exemple d'écrire que si $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$ et $Y_n \xrightarrow[n\to\infty]{\mathbb{P}} Y$ alors :

$X_nY_n \xrightarrow[n\to\infty]{\mathbb{P}} XY$, $\max(X_n, Y_n) \xrightarrow[n\to\infty]{\mathbb{P}} \max(X, Y)$, $\min(X_n, Y_n) \xrightarrow[n\to\infty]{\mathbb{P}} \max(X, Y)$, etc...

D'autre part, pour obtenir un énoncé similaire à celui-là pour la convergence L^p, il faudra s'assurer de l'intégrabilité de toutes les v.a. mises en jeu.

7.1.7 Quelques précisions sur la convergence en loi

Remarque 7.17. sur la continuité de la fonction de répartition limite...

On a $X_n \xrightarrow[n\to\infty]{\mathcal{L}} X$ si et seulement si $F_{X_n}(t) \longrightarrow F_X(t)$ en tout point t où F_X est continue. Cette restriction de la convergence simple de F_{X_n} vers F_X à l'ensemble de continuité de F_X est une condition nécessaire. Pour illustrer cette nécessité, voici un exemple. Si X_n suit une loi de Dirac $\delta_{\frac{1}{n}}$, alors la suite $(X_n)_{n\geq 0}$ converge en loi vers une v.a. X qui suit une loi de Dirac δ_0. Or on a, pour tout $n \geq 0$, $F_{X_n}(0) = \mathbb{P}(X_n \leq 0) = 0$ et $F_X(0) = \mathbb{P}(X \leq 0) = 1$...

D'une manière générale, $X_n \xrightarrow[n\to\infty]{\mathcal{L}} X$ n'implique pas que $\mathbb{P}(X_n \in A)$ converge vers $\mathbb{P}(X \in A)$ pour tout évènement A.

Comme précisé dans la remarque 7.8, dire que X_n converge en loi vers X est équivalent à dire que la mesure $\mathbb{P}_{X_n}$ converge étroitement vers $\mathbb{P}_X$. Ce qui se traduit par la proposition suivante :

Proposition 7.18.

X_n converge en loi vers X ssi X_n converge vers X au sens des distributions :

$$X_n \xrightarrow[n\to\infty]{\mathcal{L}} X \quad \Longleftrightarrow \quad \forall f \in \mathcal{K}(\mathbb{R}^d, \mathbb{R}),\ \lim_{n\to\infty} \mathbb{E}\left(f(X_n)\right) = \mathbb{E}\left(f(X)\right).$$

où $\mathcal{K}(\mathbb{R}^d, \mathbb{R})$ est l'ensemble des fonctions continues à support compact de $\mathbb{R}^d$ dans $\mathbb{R}$.

Complément à la Remarque 7.9 Notez bien que la convergence en loi est le seul mode de convergence qui "permet" que les différents $\vec{\text{v}}$.a.r. de la suite et la limite ne soit pas définis forçément sur le même espace de probabilité puisque c'est un type de convergence qui ne dépend que des

mesures images, c'est-à-dire de l'espace où arrive les $\vec{\text{v}}$.a.r. ... Cela permet d'avoir des résultats du type : $X_n \xrightarrow{\mathcal{L}} X$, $X_n \xrightarrow{\mathcal{L}} Y$ et $X \neq Y$. Pour cela, il suffit de choisir X et Y ayant la même loi, mais définies sur des espaces de probabilité différents comme par exemple dans l'Exemple 2.14 où dans l'exercice suivant...

★ **Exercice 7.19.**

1. Donner deux v.a. aléatoires X et Y définies sur des espaces de probabilité différents et qui suivent une loi $\mathcal{U}([0,1])$.
2. Soit $(X_n)_{n\geq 0}$ une suite de v.a. définies sur des espaces $(\Omega_n, \mathcal{T}_n, \mathbb{P}_n)$, à valeurs dans $[0,1]$, telles que, pour tout n, X_n suit la loi $\mathbb{P}_{X_n} = \frac{1}{n+1}\sum_{k=0}^{n} \delta_{\frac{k}{n}}$.
   Montrer que X_n converge en loi à la fois vers X et vers Y.

Proposition 7.20. CNS de convergence en loi pour les v.a.r. discrètes *Soient $(X_n)_{n\geq 0}$ une suite de v.a.r. discrètes et X une v.a.r. discrète.*

$$X_n \xrightarrow[n\to\infty]{\mathcal{L}} X \quad \Longleftrightarrow \quad \forall k \in \mathbb{Z},\ \lim_{n\to+\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k).$$

★ **Exercice 7.21.** Démontrer la Proposition 7.20.

Proposition 7.22. CS de convergence en loi pour les v.a.r. à densité (Scheffé) *Soient $(X_n)_{n\geq 0}$ une suite de v.a.r. à densité. On note f_{X_n} la densité de X_n.*

$$\left.\begin{array}{l} f_{X_n} \xrightarrow{\lambda\text{-p.p.}} f \\ \int_{\mathbb{R}} f \mathrm{d}\lambda = 1 \end{array}\right\} \quad \Longrightarrow \quad X_n \xrightarrow{\mathcal{L}} X \text{ où } X \text{ est une v.a.r. de densité } f.$$

Dans ce cas, on a de plus :

$$\lim_{n\to+\infty} \sup_{A\in\mathcal{B}(\mathbb{R})} |\mathbb{P}(X_n \in A) - \mathbb{P}(X \in A)| = 0.$$

Attention : la réciproque est fausse ! Pour s'en convaincre, voici un exemple :

★ **Exercice 7.23.** On considère la suite de v.a.r à densité $(X_n)_{n\geq 1}$, telle que, pour tout $n \leq 1$, $f_{X_n}(t) = (1 - \cos(2\pi n t))\mathbf{1}_{]0,1]}(t)$.

1. Justifier le fait que, pour tout $t \in]0,1]$, la fonction $f_{X_n}(t)$ diverge.
2. Montrer que la suite $(X_n)_{n\geq 1}$ converge en loi vers une loi que l'on précisera.

La proposition qui suit est une proposition fondamentale, dont la partie "réciproque" sert notamment de pivot à la démonstration du Théorème central limite (voir le Paragraphe 7.3).

Proposition 7.24. Convergence en loi et fonctions caractéristiques (Lévy)
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de $\vec{\text{v}}$.a.r. de dimension d définis sur $(\Omega, \mathcal{T}, \mathbb{P})$ et X un $\vec{\text{v}}$.a.r. de dimension d défini sur $(\Omega', \mathcal{T}', \mathbb{P}')$.

$$X_n \xrightarrow{\mathcal{L}} X \quad \Longrightarrow \quad \forall t \in \mathbb{R},\ \lim_{n\to\infty} \phi_{X_n}(t) = \phi_X(t).$$

Le réciproque est vraie si ϕ_X est continue en 0 :

$$\left.\begin{array}{l} \forall t \in \mathbb{R},\ \lim_{n\to\infty} \phi_{X_n}(t) = \phi_X(t) \\ \phi_X \text{ continue en } 0 \end{array}\right\} \quad \Longrightarrow \quad X_n \xrightarrow{\mathcal{L}} X.$$

7.1.8 Liens entre les différents types de convergence

Théorème 7.25. *Soient $(X_n)_{n\geq 0}$ une suite de v.a.r. et X une v.a.r. définies sur $(\Omega, \mathcal{T}, \mathbb{P})$.*

- $X_n \xrightarrow{\mathbb{P}\text{-}p.s.} X \quad \Longrightarrow \quad X_n \xrightarrow{\mathbb{P}} X$,
- $X_n \xrightarrow{\mathbb{P}} X \quad \Longrightarrow \quad$ *Il existe une suite extraite* $(X_{n_k})_{n\geq 0}$, $X_{n_k} \xrightarrow{\mathbb{P}\text{-}p.s.} X$.

La convergence $\mathbb{P}$-p.s. est donc plus forte que la convergence en probabilité. Cela était en fait déjà "visible" à la Proposition 7.11 : avoir la convergence en probabilité, c'est dire que, pour tout $\epsilon > 0$, la suite $(\mathbb{P}(|X_n - X| > \epsilon))_{n\geq 0}$ temps vers zéro. On la condition nécessaire mise en évidence à la Proposition 7.11 pour avoir la convergence $\mathbb{P}$-p.s. porte sur la convergence non pas de la suite $(\mathbb{P}(|X_n - X| > \epsilon))_{n\geq 0}$, mais de la série de terme général $(\mathbb{P}(|X_n - X| > \epsilon))_{n\geq 0}$, ce qui est une condition beaucoup plus forte...

Ci-dessous, un exemple de suites de v.a.r. qui converge en probabilité mais $\mathbb{P}$-p.s. nulle part.

★ **Exercice 7.26.** Soit la suite d'intervalles $(I_n)_{n\geq 0}$ de $\mathbb{R}$ définis par $I_n = \left[\frac{\ell_n}{2^{k_n}}, \frac{\ell_n+1}{2^{k_n}}\right]$ où ℓ_n et k_n sont les entiers (uniques) tels que $2^{k_n} \leq n < 2^{k_n+1}$ et $n = 2^{k_n} + \ell_n$.

On définit la suite v.a.r suivante : pour tout $n \geq 0$, X_n est définie sur $[0, 1]$ (muni de sa tribu borélienne et de la mesure de Lebesgue) par : $X_n = \mathbf{1}_{I_n}$.

1. Montrer que la suite $(X_n)_{n\geq 0}$ converge en probabilité vers 0,
2. Lorsque k est fixé, pour tout $\epsilon > 0$, et tout $n \in [2^k, 2^{k+1}[$, calculer $\mathbb{P}(X_n > \epsilon)$. Conclure que la suite $(X_n)_{n\geq 0}$ ne converge pas $\mathbb{P}$-p.s. vers 0.

Théorème 7.27. *Soient $(X_n)_{n\geq 0}$ une suite de v.a.r. et X une v.a.r. définies sur $(\Omega, \mathcal{T}, \mathbb{P})$.*

- $X_n \xrightarrow{L^p} X \quad \Longrightarrow \quad X_n \xrightarrow{\mathbb{P}} X$,
- *Si la suite $(|X_n|^p)_{n\geq 0}$ est équi-intégrable et que $X \in L^p$, alors :*

$$X_n \xrightarrow{\mathbb{P}} X \quad \Longleftrightarrow \quad X_n \xrightarrow{L^p} X.$$

Définitions 7.28. Une famille de v.a.r. $(Y_n)_{n\geq 0}$ définies sur $(\Omega, \mathcal{T}, \mathbb{P})$ est dite **équi-intégrable** si

$$\lim_{a\to+\infty} \sup_{n\geq 0} \mathbb{E}\left(|Y_n|\mathbf{1}_{\{|Y_n|>a\}}\right) = 0.$$

Elle est dite **bornée dans** L^1 si $\sup_{n\geq 0} \mathbb{E}(|Y_n|) < \infty$.

Remarque 7.29. Si une famille est équi-intégrable, alors elle est bornée dans L^1 mais la réciproque est fausse, comme le montre l'exercice suivant.

★ **Exercice 7.30.** On considère une suite de v.a.r. $(Y_n)_{n\geq 0}$ telles que la loi de Y_n soit $\frac{1}{n}\delta_n + (1-\frac{1}{n})\delta_0$.

1. Montrer que la famille $(Y_n)_{n\geq 0}$ est bornée dans L^1.
2. Calculer $\sup_{n\geq 0} \mathbb{E}\left(|Y_n|\mathbf{1}_{\{|Y_n|>a\}}\right)$ pour tout $a > 0$ et en déduire que la famille $(Y_n)_{n\geq 0}$ n'est pas équi-intégrable.

Pour qu'une famille bornée dans L^1 soit équi-intégrable, il faut hypothèse supplémentaire :

Proposition 7.31. CNS pour l'équi-intégrabilité d'une famille de v.a.r.

Soient $(Y_n)_{n\geq 0}$ une suite de v.a.r. définies sur $(\Omega, \mathcal{T}, \mathbb{P})$.

$(Y_n)_{n\geq 0}$ est équi-intégrable si et seulement si :

1. *$(Y_n)_{n\geq 0}$ est bornée dans L^1 : $\sup_{n\geq 0} \mathbb{E}(|Y_n|) < \infty$,*
2. *$\forall \epsilon > 0, \exists \eta > 0,\ \forall A \in \mathcal{B}(\mathbb{R}),\ \mathbb{P}(A) < \eta \implies \sup_{n\geq 0} \mathbb{E}(|Y_n|\mathbf{1}_A) < \epsilon$.*

On rappelle que $\mathbb{E}(|Y_n|\mathbf{1}_A) = \int_A |Y_n| \mathrm{d}\mathbb{P}$.

En pratique, pour vérifier qu'une suite de v.a.r. est équi-intégrable, on utilise une des deux conditions suffisantes présentées ci-dessous :

Proposition 7.32. CS pour l'équi-intégrabilité d'une famille de v.a.r. I

Soient $(X_n)_{n\geq 0}$ une suite de v.a.r. définies sur $(\Omega, \mathcal{T}, \mathbb{P})$.

Si il existe Y dans L^1 telle que $|X_n| \leq Y$ pour tout $n \geq 0$, alors la suite $(X_n)_{n\geq 0}$ est équi-intégrable.

Proposition 7.33. CS pour l'équi-intégrabilité d'une famille de v.a.r. II

Soient $(X_n)_{n\geq 0}$ une suite de v.a.r. définies sur $(\Omega, \mathcal{T}, \mathbb{P})$.

Si la suite $(|X_n|^p)_{n\geq 0}$ est bornée pour un certain $p > 1$, alors elle est équi-intégrable.

Cela se démontre en utilisant l'inégalité de Hölder. **Attention !** l'entier p qui intervient dans cette CS doit vraiment être STRICTEMENT PLUS GRAND QUE 1, sinon, c'est faux.

Théorème 7.34. *Soient $(X_n)_{n\geq 0}$ une suite de v.a.r. et X une v.a.r. définies sur $(\Omega, \mathcal{T}, \mathbb{P})$.*

- $X_n \xrightarrow{L^p} X \implies$ *Il existe une suite extraite* $(X_{n_k})_{n\geq 0},\ X_{n_k} \xrightarrow{\mathbb{P}\text{-}p.s.} X$,
- $\left.\begin{array}{l} X_n \xrightarrow{\mathbb{P}\text{-}p.s.} X, \\ \exists Y \in L^p, \forall n \geq 0,\ |X_n| < Y \end{array}\right\} \implies X_n \xrightarrow{L^p} X$.

Théorème 7.35. *Soient $(X_n)_{n\geq 0}$ une suite de v.a.r. et X une v.a.r.*

- $X_n \xrightarrow{\mathbb{P}} X \implies X_n \xrightarrow{\mathcal{L}} X$,
- $X_n \xrightarrow{\mathcal{L}} a$, *v.a. constante* $\implies X_n \xrightarrow{\mathbb{P}} a$.

De ce théorème, on déduit évidemment les implications suivantes :

- $X_n \xrightarrow{\mathbb{P}\text{-p.s.}} X \implies X_n \xrightarrow{\mathcal{L}} X$,
- $X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{\mathcal{L}} X$,
- $X_n \xrightarrow{\mathcal{L}} a$, v.a. constante $\implies$ Il existe une suite extraite $(X_{n_k})_{n\geq 0},\ X_{n_k} \xrightarrow{\mathbb{P}\text{-p.s.}} a$.

7.2 Lois des grands nombres

Le principe de la loi des grands nombres, ou plutôt DES loiS des grands nombres est le suivant :

Si on effectue une suite de réalisations d'une expérience aléatoire $(X_n)_{n\geq 1}$, en supposant qu'à chaque instant, la "moyenne spatiale" des observations est la même ($\mathbb{E}(X_n) = m$ quelque soit

$n \geq 1$), alors la " moyenne temporelle" $\frac{1}{n}\sum_{k=1}^{n} X_k$ des observations converge vers la moyenne spatiale commune :

$$\text{Si } \forall n \geq 1,\ \mathbb{E}(X_n) = m,\ \text{alors } \overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to+\infty]{} m.$$

Le qualificatif de loi faible ou loi forte va dépendre du type de convergence de la suite $\frac{1}{n}\sum_{k=1}^{n} X_k$:
- On parle de loi faible si la convergence est en probabilité ou dans L^p, avec en général des hypothèses faibles d'indépendance : non-corrélation ou indépendance 2 à 2.
- On parle de loi forte si la convergence est la convergence $\mathbb{P}$-p.s., avec des hypothèses fortes d'indépendance : indépendance globale.

Les lois fortes sont donc plus difficiles à obtenir car la convergence $\mathbb{P}$-p.s. est plus contraignante que les convergence en probabilité ou L^p. Cela en fait des résultats très forts mais en contrepartie, les hypothèses doivent être beaucoup plus restrictives que pour les lois faibles (indépendance globale VS indépendance 2 à 2 ou non corrélation). La souplesse des hypothèses requises pour les lois faibles rend ces théorèmes bien utiles dans de nombreuses situations.

7.2.1 Lois faibles de grands nombres

Théorème 7.36. (Tchebichev) Loi faible des grands nombres dans L^2

Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. de L^2, centrées et non corrélées.

$$\frac{1}{n^2}\sum_{k=1}^{n}\mathbb{E}(X_k^2) \xrightarrow[n\to+\infty]{} 0 \qquad \Longrightarrow \qquad \overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to+\infty]{L^2,\,\mathbb{P}} 0.$$

Ce résultat est la plus facile à démontrer parmis les lois des grand nombres.

Démonstration. Comme les v.a. X_n sont centrées et non corrélées, on a, pour tout $j \neq k$, $\mathbb{E}(X_iX_j) = \mathbb{E}(X_i)\mathbb{E}(X_j) = 0$. Ainsi,

$$\left\|\overline{X}_n\right\|_2^2 = \int_\Omega \overline{X}_n^2 \mathrm{d}\mathbb{P} = \frac{1}{n^2}\sum_{k=1}^{n}\|X_k\|_2^2 = \frac{1}{n^2}\sum_{k=1}^{n}\mathbb{E}(X_k^2).$$

On obtient donc $\left\|\overline{X}_n\right\|_2^2 \xrightarrow[n\to+\infty]{} 0$, autrement dit $\overline{X}_n \xrightarrow[n\to+\infty]{L^2} 0$. Puisque la convergence L^2 entraine la convergence en probabilité, on obtient le résultat. □

Corollaire 7.37. Loi faible des grands nombres dans L^2, cas i.i.d.

Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. de L^2 i.i.d.,

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to+\infty]{L^2,\,\mathbb{P}} \mathbb{E}(X_1).$$

Démonstration. Les $(X_n)_{n\geq 1}$ étant i.i.d., elles sont en particulier non corrélées. Sion note $m = \mathbb{E}(X_1)$ et $\sigma^2 = \mathbb{E}(X_1^2)$. alors la suites de v.a.r. $(X_n - m)_{n\geq 1}$ vérifie les hypothèses du théorème précédent. Comme $\frac{1}{n^2}\sum_{n\geq 1}\mathbb{E}(X_k^2) = \frac{1}{n}\sigma^2$, on a donc : $\overline{X}_n - m = \frac{1}{n}\sum_{k=1}^n (X_k - m) \xrightarrow[n\to+\infty]{L^2,\,\mathbb{P}} 0$. On en déduit alors que $\overline{X}_n \xrightarrow[n\to+\infty]{L^2,\,\mathbb{P}} m$. □

On peut étendre la loi faible des grands nombres au suites de v.a.r. de L^1 (qui admettent une espérance mais pas forcément une variance), mais on doit pour cela renforcer les hypothèses en faisant apparaitre l'équi-intégrabilité de la suite $(X_n)_{\geq 1}$. On rappelle qu'une famille de v.a.r. $(X_n)_{n\geq 0}$ définies sur $(\Omega, \mathcal{T}, \mathbb{P})$ est dite équi-intégrable si $\lim\limits_{a\to+\infty} \sup\limits_{n\geq 0} \mathbb{E}\left(|X_n|\mathbf{1}_{\{|X_n|>a\}}\right) = 0$.

Théorème 7.38. Loi faible des grands nombres dans L^1

Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. de L^1, indépendantes et de même moyenne m.

La famille $(X_n)_{n\geq 1}$ est équi-intégrable $\quad\Longrightarrow\quad$ $\overline{X}_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^n X_k \xrightarrow[n\to+\infty]{L^1,\,\mathbb{P}} m.$

La preuve de ce théorème est une application des inégalités de Bienaymé-Tchebichev et de Markov.

Démonstration. Sous la condition d'équi-intégrabilité de la famille $(X_n)_{n\geq 1}$, montrer la convergence L^1 de $\overline{X}_n$ vers 0 est équivalent à montrer la convergence en probabilité (voir Théorème 7.27).

On va montrer la convergence en probabilité.

Soit $\epsilon > 0$. Puisque la suite $(X_n)_{n\geq 1}$ est équi-intégrable, il existe $M > 0$ tel que :

$$\forall n \geq 1,\ \mathbb{E}\left(|X_n|\mathbf{1}_{\{|X_n|>M\}}\right) \leq \epsilon^2.$$

Pour tout $n \geq 1$, on définit :

- $Y_n = X_n^{[M]} = X_n\mathbf{1}_{\{|X_n|\leq M\}}$ la v.a. X_n tronquée au dessus de M.
- $Z_n = X_n - Y_n = X_n\mathbf{1}_{\{|X_n|>M\}}$.

Puisque $\mathbb{E}(X_n) = m$ pour tout $n \geq 1$, on a $\mathbb{E}(\overline{X}_n) = m$ et donc :

$$\overline{X}_n - m = \overline{X}_n - \mathbb{E}(\overline{X}_n) = \left(\overline{Y}_n - \mathbb{E}(\overline{Y}_n)\right) + \left(\overline{Z}_n - \mathbb{E}(\overline{Z}_n)\right).$$

Ainsi,

$$\mathbb{P}\left(\left|\,\overline{X}_n - m\,\right| \geq 2\epsilon\right) \leq \mathbb{P}\left(\left|\,\overline{Y}_n - \mathbb{E}(\overline{Y}_n)\,\right| \geq \epsilon\right) + \mathbb{P}\left(\left|\,\overline{Z}_n - \mathbb{E}(\overline{Z}_n)\,\right| \geq \epsilon\right).$$

On va majorer les 2 quantités à gauche pour monter que $\mathbb{P}\left(\left|\,\overline{X}_n - m\,\right| \geq 2\epsilon\right)$ tend vers 0 avec n.

1. Puisque Y_n est une v.a. bornée, elle est dans L^2 et il en va de même pour $\overline{Y}_n$. L'inégalité de Bienaymé-Tchebichev donne alors :

$$\mathbb{P}\left(\left|\,\overline{Y}_n - \mathbb{E}(\overline{Y}_n)\,\right| \geq \epsilon\right) \leq \frac{1}{\epsilon^2}\mathbb{V}\mathrm{ar}\left(\overline{Y}_n\right).$$

Les Y_n étant des v.a. indépendantes, car les X_n le sont, on a $\mathbb{V}\mathrm{ar}\left(\overline{Y}_n\right) = \dfrac{1}{n^2}\displaystyle\sum_{k=1}^{n} \mathbb{V}\mathrm{ar}\left(Y_k\right)$. D'autre part, pour tout $n \geq 1$, on a $\mathbb{V}\mathrm{ar}(Y_n) = \mathbb{E}(Y_n^2) - \mathbb{E}(Y_n)^2 \leq \mathbb{E}(Y_n^2) \leq M^2$ et on obtient ainsi :

$$\mathbb{P}\left(\left|\,\overline{Y}_n - \mathbb{E}(\overline{Y}_n)\,\right| \geq \epsilon\right) \leq \frac{1}{n^2\epsilon^2} M^2.$$

Pour n assez grand, on a alors $\mathbb{P}\left(\left|\,\overline{Y}_n - \mathbb{E}(\overline{Y}_n)\,\right| \geq \epsilon\right) \leq \epsilon$.

2. Les v.a. Z_n sont dans L^1. En effet, on a choisit M de sorte que

$$\forall n \geq 1,\ \mathbb{E}\left(|Z_n|\right) = \mathbb{E}\left(|X_n|\mathbf{1}_{\{|X_n|>M\}}\right) \leq \epsilon^2.$$

Cela implique notamment que $\overline{Z_n}$ est aussi dans L^1. Appliquons l'inégalité de Markov à $\overline{Z_n}$:

$$\mathbb{P}\left(\left|\,\overline{Z}_n - \mathbb{E}(\overline{Z}_n)\,\right| \geq \epsilon\right) \leq \frac{1}{\epsilon}\mathbb{E}\left(\left|\,\overline{Z}_n - \mathbb{E}(\overline{Z}_n)\,\right|\right) \leq \frac{2}{\epsilon}\mathbb{E}\left(\left|\,\overline{Z}_n\,\right|\right).$$

Or $\mathbb{E}\left(\left|\,\overline{Z}_n\,\right|\right) \leq \displaystyle\sum_{k=1}^{n} \mathbb{E}\left(|Z_k|\right)$ et donc :

$$\mathbb{P}\left(\left|\,\overline{Z}_n - \mathbb{E}(\overline{Z}_n)\,\right| \geq \epsilon\right) \leq \frac{2}{n\epsilon} n\epsilon^2 = 2\epsilon.$$

Ainsi, on déduit des majorations de et $\mathbb{P}\left(\left|\,\overline{Y}_n - \mathbb{E}(\overline{Y}_n)\,\right| \geq \epsilon\right)$ de $\mathbb{P}\left(\left|\,\overline{Z}_n - \mathbb{E}(\overline{Z}_n)\,\right| \geq \epsilon\right)$ que, pour tout $\epsilon > 0$, lorsque n assez grand on a :

$$\mathbb{P}\left(\left|\,\overline{X}_n - m\,\right| \geq \epsilon\right) \leq 3\epsilon.$$

ce qui implique la convergence en probabilité de la suite $\overline{X}_n$ vers m. □

Corollaire 7.39. Loi faible des grands nombres dans L^1, cas i.i.d.

Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. de L^1, i.i.d.

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to+\infty]{L^1,\,\mathbb{P}} 0.$$

Démonstration. Il suffit de voir que lorsque les v.a. sont i.i.d. alors $\mathbb{E}\left(|X_n|\mathbf{1}_{\{|X_n|>M\}}\right)$ ne dépend pas de n et $\lim_{M\to+\infty} \mathbb{E}\left(|X_1|\mathbf{1}_{\{|X_1|>M\}}\right) = 0$ car $X_1 \in L^1$. une famille de v.a. i.i.d. de L^1 est donc toujours équi-intégrable. On peut donc lui appliquer le Théorème 7.38. □

7.2.2 Lois fortes de grands nombres

On donne ici juste les enoncés des lois fortes de grand nombres.

Théorème 7.40. Loi forte des grands nombres dans L^2

Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. indépendantes de L^2.

$$\left.\begin{array}{l} \mathbb{E}(X_n) \xrightarrow[n\to+\infty]{} m, \\ \displaystyle\sum_{n\geq 1} \frac{1}{n^2}\mathbb{V}\mathrm{ar}(X_n) < +\infty. \end{array}\right\} \quad \Longrightarrow \quad \overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to+\infty]{\mathbb{P}\text{-}p.s.} m.$$

Théorème 7.41. (Rajchman) Loi forte des grands nombres dans L^2, cas centré

Soient $(X_n)_{n\geq 1}$ une suite de v.a.r. indépendantes et centrées de L^2 et $(u_n)_{n\geq 1}$ une suite croissante de nombres réels positifs qui tends vers $+\infty$.

$$\sum_{n\geq 1} \frac{1}{u_n^2}\mathbb{E}(X_n^2) < +\infty \quad \Longrightarrow \quad \overline{X}_n = \frac{1}{u_n}\sum_{k=1}^{n} X_k \xrightarrow[n\to+\infty]{\mathbb{P}\text{-}p.s.} 0.$$

Corollaire 7.42. Loi forte des grands nombres dans L^2, cas i.i.d.

Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. i.i.d. de L^2.

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to+\infty]{\mathbb{P}\text{-}p.s., L^2} \mathbb{E}(X_1).$$

Dans le cadre L^1, la convergence $\mathbb{P}$-p.s. de $\overline{X}_n$ n'est garantie que dans le cas le plus restrictif, à savoir le cas des suites v.a.r. qui sont i.i.d. :

Théorème 7.43. (Kolmogorov-Khintchine) Loi forte des grands nombres dans L^1

Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. i.i.d. de L^1.

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to+\infty]{\mathbb{P}\text{-}p.s., L^1} \mathbb{E}(X_1).$$

7.3 Théorème central limite

Théorème 7.44. (Lévy) Le théorème central limite (TCL)

Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. i.i.d de L^2, définies sur un même espace de probabilité $(\Omega, \tau, \mathbb{P})$.
On suppose $\mathbb{E}(X_1) = m$ et $\mathbb{V}\mathrm{ar}(X_1) = \sigma^2$.

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{X_k - m}{\sigma} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Remarque 7.45. Le coefficient $\sqrt{n}$ est en fait assez intuitif car :
- L'esperance de X_k est m, donc

$$\mathbb{E}\left(\sum_{k=1}^{n} \frac{X_k - m}{\sigma}\right) = 0.$$

- La variance de $X_k - m$ est σ^2, donc :

$$\mathbb{V}\mathrm{ar}\left(\sum_{k=1}^{n} \frac{X_k - m}{\sigma}\right) = \sum_{k=1}^{n} \mathbb{V}\mathrm{ar}\left(\frac{X_k - m}{\sigma}\right) = n\frac{1}{\sigma^2}\mathbb{V}\mathrm{ar}(X_1) = n.$$

On en déduit donc que $\mathbb{V}\mathrm{ar}\left(\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{X_k - m}{\sigma^2}\right) = 1$.

Ainsi, la variable aléatoire $\dfrac{1}{\sqrt{n}} \displaystyle\sum_{k=1}^{n} \frac{X_k - m}{\sigma^2}$ suit toujours une loi d'espérance 0 et de variance 1... la Théorème central limite indique que lorsque n est grand, cette loi devient "normale".

Une façon plus pertinente d'interpréter ce théorème est la suivante : Sous les hypothèses du TCL, on est dans le cadre d'application de la loi forte des grands nombres, donc :

$$\overline{X}_n - m = \frac{1}{n} \sum_{k=1}^{n} (X_k - m) \xrightarrow[n\to+\infty]{\mathbb{P}\text{-p.s.}} 0.$$

Ce qui nous assure qu'en substance, on a $\mathbb{P}$-presque sûrement $\overline{X}_n - m = o(1)$.

Puisque toutes les variables aléatoires sont dans L^2 et de variance σ^2, le TCL permet d'obtenir un ordre de grandeur de $\overline{X}_n - m$: en effet, il indique qu'on a

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{X_k - m}{\sigma} = \sqrt{n}\frac{(\overline{X}_n - m)}{\sigma} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

ce qui permet d'évaluer la vitesse de convergence de $\overline{X}_n$ vers m :
Quand n devient grand, $\overline{X}_n - m$ se comporte presque comme une loi normale $\mathcal{N}(0, \frac{\sigma^2}{n})$.

Cet angle permet retouver l'énoncé du TCL à l'aide des constatations suivantes : $\overline{X_n}$ est une v.a. d'espérance m et de variance $\frac{\sigma^2}{n}$ (le calcul est rapide !) donc la variable aléatoire $\sqrt{n}\frac{\overline{X}_n - m}{\sigma}$ est toujours une variable aléatoire dont la loi a pour espérance 0 et pour variance 1. Le TCL indique que lorsque n devient grand, cette loi devient "normale".

Remarque 7.46. La preuve du Théorème central limite repose sur la caractérisation de la convergence en loi à l'aide des fonctions caractéristiques : Une suite de v.a. $(X_n)_{n\geq 0}$ converge en loi vers une v.a. X si et seulement si la suite $(\phi_{X_n})_{n\geq 0}$ converge simplement vers ϕ_X.

Deuxième partie

STATISTIQUES

Introduction

Aussi loin que l'on remonte dans le temps et dans l'espace (en Chine et en Egypte, par exemple), les Etats ont toujours senti le besoin de disposer d'informations sur leurs sujets ou sur les biens qu'ils possedent et produisent. Mais les recensements de population et de ressources, les statistiques (du latin *status* : état) sont restées purement descriptives jusqu'au 17ème siècle. Puis s'est développé le calcul des probabilités et des méthodes statistiques sont apparues en Allemagne, en Angleterre et en France. Beaucoup de scientifiques ont apporté leur contribution au développement de cette science : Pascal, Huygens, La famille Bernouilli, Moivre, Laplace, Gauss, Mendel, Pearson, Ficher, etc....

Actuellement, beaucoup de domaines utilisent les méthodes statistiques (médecine, agronomie, sociologie, industrie etc....).

La statistique englobe les méthodes de recueil, de traitement, de présentation, d'organisation, de synthèse, d'analyse et d'interprétation de données (numériques ou non), afin de décrire et de prévoir des phénomènes issus de domaines très variés. C'est évidemment un branche des mathématiques de référence pour la recherche scientifique basée sur des expérimentations : On observe, on note les observations, on synthétise et on traite les données, on propose un modèle, on teste le modèle, on valide (ou pas).

Les données sont collectées soit de manière exhaustive soit par sondage. On trie les données que l'on organise en tableaux, diagrammes, etc... Puis on interprète les résultats : on les compare avec ceux déduits de la théorie des probabilités. On pourra donc évaluer des grandeurs statistiques comme la moyenne ou la variance (estimateurs, intervalles de confiance) afin de savoir si deux populations sont comparables (tests d'hypothèses)ou de déterminer si des grandeurs sont liées et de quelle façon (corrélation, ajustement analytique). Les conclusions, toujours entachées d'un certain pourcentage d'incertitude, nous permettent alors de prendre une décision.

Ce cours ouvre sur deux parties du monde statistique :

La statistique déscriptive, qui a pour objectif la synthèse des données expérimentales reccueillies : sous forme d'indices numériques, de tableaux et graphiques divers.

La statistique inférentielle, qui tente d'extrapoler à la population entière le comportement d'un ou plusieurs caractèreS à partir des données reccueillies sur une partie de la population (sondages, prélèvements de poissons dans un lac, essais cliniques...). A partir de ces informations partielles, on essaye d'imaginer le comportement des caractères sur toute la population. Evidemment, c'est ici qu'il sera question d'estimation de parametres, de risque d'erreur et de confiance en les résultats obtenus...

Chapitre 1

Statistique descriptive : Séries univariées

1.1 Les bases du vocabulaire statistique

Avant de plonger dans le joyeux monde des statistiques, un peu de vocabulaire s'impose car comme pour les probabilités, la statistique a son propre dialecte... Notez que le langage probabiliste et son bestiaire de lois usuelles doit être maitrisé pour faire des statistiques inférentielles... Un bref rappel à ce propos sera fait en début des chapitres concernés.

• Population, individu, écnatillon et caractères

Une **population** Ω est l'ensemble des éléments sur lesquels portent l'étude statistique (l'ensemble des élèves d'une classe, l'ensemble des assurés d'une compagnie, l'ensemble des ~~cobayes~~ patients lors de l'étude des effets d'un médicament, l'ensemble des emplacements géographiques dans une zone à risque, etc...). Il est le plus souvent discrèt, mais pas forcément. Cela dit, si la population est "continue", il n'est pas rare de la discrétiser en utilisant un petit maillage.

Chaque élément de la population Ω est appelé **individu**.

Un **échantillon** est une partie (la plupart du temps finie puisqu'il s'agit en général des données expérimentales...) de la population Ω.

La statistique est l'étude d'un ou plusieurs **caractère** (ou **variable**) $\xi : \Omega \to C$ qui n'est rien d'autre qu'une fonction définie sur Ω à valeurs dans un ensemble C appelé ensemble des **valeurs** (ou des **modalités**).

• Caractères quantitatifs

Lorsque ξ ne prend que des valeurs numériques, il est **quantitatif** :

- **discret** s'il ne peut prendre que des valeurs isolées (notes, âge, nombre de dents ...)
- **continu** dans le cas contraire (poids, taille, temps de reaction à un blague ...). Dans ce cas on effectue souvent un regroupement des valeurs par intervalles appellés **classes**. La longueur d'une classe est appelée **amplitude**.

  Pour construire ces intervalles, on respecte en général les règles suivantes :
  - Le nombre de classes est compris entre 5 et 20 (de préférence entre 6 et 12)

– Chaque fois que cela est possible, les amplitudes des classes sont égales.
– Chaque classe (sauf la dernière) contient sa borne inférieure mais pas sa borne supérieure.
Dans les calculs, une classe sera représentée par son centre, qui est le milieu de l'intervalle.

• Caractères qualitatifs, caractères catégoriels

Lorsque ξ n'est pas à valeurs numériques, on dit qu'il est **qualitatif** (couleur des yeux, sport pratiqué, espèce ...).

Lorsque l'ensemble des modalités est discrèt (caractère quantitatif discrèt ou caractère qualitatif), on parle de **caractère catégoriel** ou de **variable catégorielle**.

A chaque caractère catégoriel ξ, on peut associer une **partition** $(P_i)_{i\in I}$ **de la population** Ω **associée au caractère** ξ, indexée par l'ensemble des modalités $(C_i)_{i\in I}$ du caractère, définie par $P_i = \xi^{-1}(\{C_i\}) = [\xi = X_i]$.

• Effectifs et fréquences

Quand Ω est un ensemble discrèt, chaque valeur (ou classe) c est associée un **effectif** n : c'est le nombre d'individus associés à cette valeur : $n = \text{Card}(\xi^{-1}(c))$. Quelle que soit la nature de Ω, on peut associer à chaque valeur (ou classe) c sa **fréquence d'apparition**.

Dans la suite de ce chapitre, on considèrera pour les exemples les 3 séries statistiques suivantes :

Série A : Notes obtenues à un examen dans une classe de 30 élèves :

$-3-11-9-8-9-4-7-9-7-16-7-8-9-9-5-6-9-10-11-9-11-2-6-3-11-9-13-10-13-8-8-15-8$

Population : Classe de 30 élèves, Caractère quantitatif discrèt : note de l'examen.

Série B : Salaires en euros des employés d'une entreprise :

Salaires	$[900, 1200]$	$[1200, 1400]$	$[1400, 1600]$	$[1600, 1800]$	$[1800, 2000]$	$[2000, 2400]$	TOTAL
Effectif	30	30	60	40	20	20	200

Population : Entreprise de 280 employés, Caractère quantitatif continu, regroupé en classes d'amplitudes variées : salair).

Série C : Proportion d'adhérents à un club sportif dans différentes sections :

– 17% jouent au handball,
– 25% houent au rugby,
– 58% jouent au tennis.

Population : Club sportif, effectif inconnu, Caractère qualitatif : sport.

1.2 Représentations des données brutes

Il existe plusieurs manière de représenter une série statistique, plus ou moins lisibles selon la nature du caractère étudié.

1.2.1 Variables quantitatives

• **Tri plat d'une variable quantitative**

Pour les séries quantitatives, la première représentation est le **tri plat** : Le caractère est présenté en tableau par modalités ou classes (ordonnées croissantes ou décroissantes), associé aux effectifs (on parle alors de **tableau d'effectifs**) et éventuellement complété par les fréquences d'apparition.

Exemple 1.1. La série **B** est déja présentée par son tableau d'effectifs par classes.

Exemple 1.2. Ci-dessous, le tableau d'effectifs de la série **A**, completé par la distribution des fréquences :

Notes	2	3	4	5	6	7	8	9	10	11	13	15	16
Effectif	1	2	1	1	2	3	5	6	2	3	2	1	1
Fréq. en %	3	7	3	3	7	10	17	20	7	10	7	3	3

On peut éventuellement regrouper les modalités en classe, selon les besoins. Ci-dessous, le tableau d'effectifs de la série **A** regroupées par classes d'amplitude 5 :

Notes	$[0,5[$	$[5,10[$	$[10,15[$	$[15,20[$
Effectif	4	17	7	2
Fréq. en %	13	57	23	7

• **Effectifs et fréquences cumulés**

On peut aussi choisir d'ajouter au tableau des effectifs la ligne des

- **effectifs cumulés croissants** : L'effectif cumulé croissant d'une modalité (ECC) est le nombre d'individus dont la modalité est plus petite ou égale à la modalité en cours.
- **fréquences cumulées croissantes** : la fréquence cumulée croissante d'une modalité (FCC) est la fréquence d'apparition d'une modalité est plus petite ou égale à la modalité en cours. C'est le pendant statitistique de la fonction de répartition d'une loi de probabilité.

Exemple 1.3. Ci-dessous, le tableau d'effectifs de la série **A** completé par les effectifs cumulés croissants et les fréquences cumulées croissantes :

Notes	2	3	4	5	6	7	8	9	10	11	13	15	16
Effectif	1	2	1	1	2	3	5	6	2	3	2	1	1
ECC	1	3	4	5	7	10	15	21	23	26	28	29	30
FCC en %	3	10	13	16	23	33	50	70	77	87	94	97	100

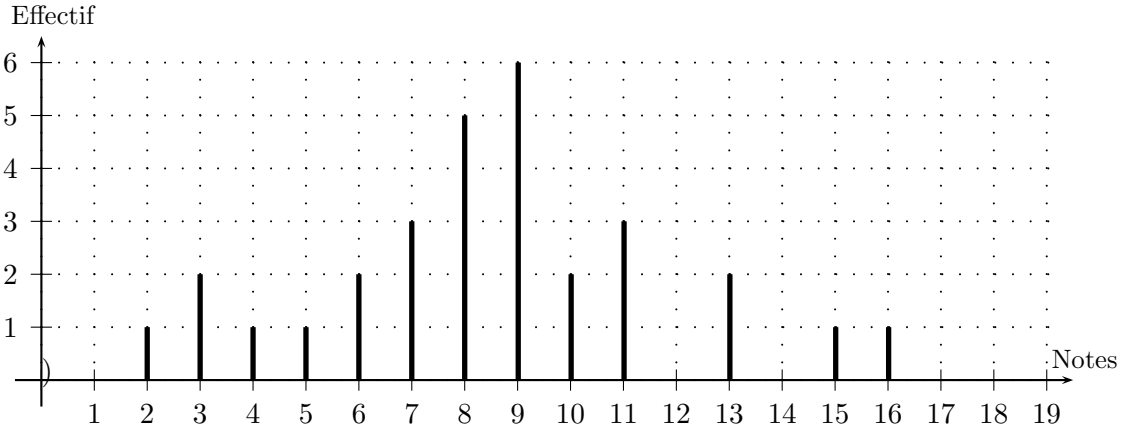
Exemple 1.4. Ci-dessous, le tableau d'effectifs de la série **B** completé par les effectifs cumulés croissants et les fréquences cumulées croissantes :

Salaires	$[900, 1200]$	$[1200, 1400]$	$[1400, 1600]$	$[1600, 1800]$	$[1800, 2000]$	$[2000, 2400]$
ECC	30	60	120	160	180	200
FCC en %	15	30	60	80	90	100

• **Diagramme en bâtons**

Pour les caractères quantitatifs discrèts, on peut également représenter la série statistique avec un **diagramme en bâtons**, où chaque bâton est proportionnel à l'effectif (ou à la fréquence, forçément) associé à chaque modalité.

Exemple 1.5. Voici le diagramme en bâtons représentant la série des notes de la **série A** :

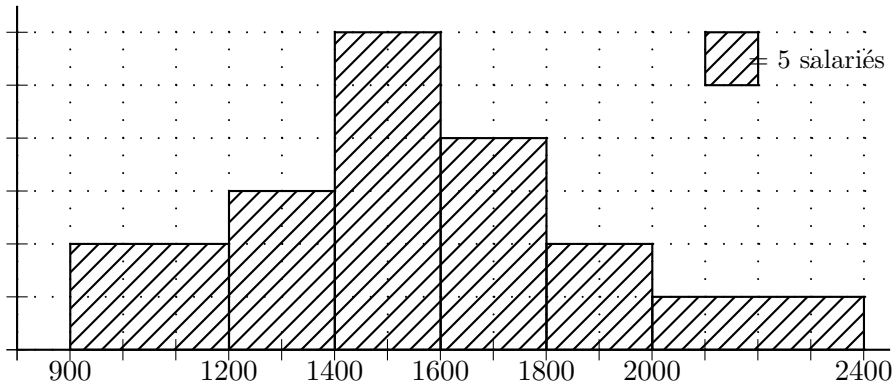


• **Histogrammes**

Pour les caractères quantitatifs continus avec modalités regroupées par classes, on représente la série statistique par un **histogramme** : chaque classe est représentée par un rectangle dont la base correspond à l'intervalle de classe et son aire est proportionnelle à l'effectif (ou à la fréquence d'apparition) de la classe.

Evidemment, lorsque toutes les classes ont même amplitude, il suffit que la hauteur soit directement proportionnelle directement à l'éffectif mais ce n'est pas le cas si les classes sont d'amplitude variable...

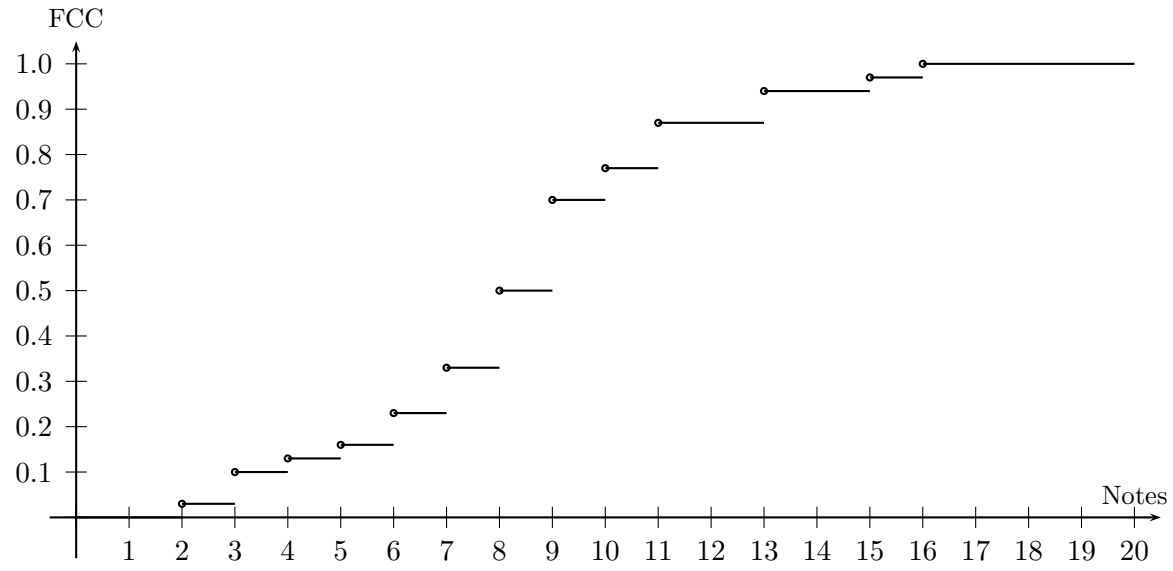
Exemple 1.6. Pour la série **B**, on obtient par exemple l'histogramme suivant :



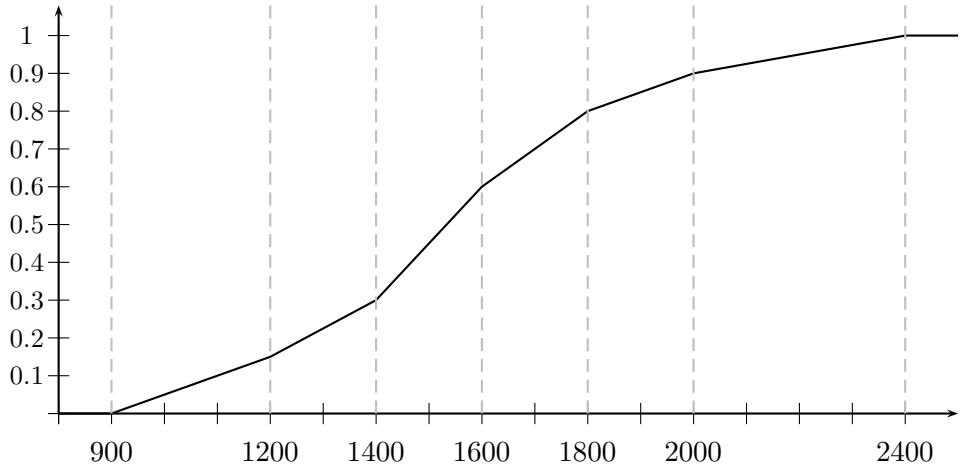
• Graphes des fréquences cumulées croissantes

Pour les caractères quantitatifs, on peut représenter la graphe des fréquences cumulées. C'est une fonction croissante, de 0 à 1. Lorsque le caractère est discrèt, on obtient une fonction en escalier. Pour les caractère continus regroupés par classe, on obtient une fonction linéaire par morceaux.

Exemple 1.7. Ci-dessous, le graphe des fréquences cumulées de la série **A** :



Exemple 1.8. Ci-dessous, le graphe des fréquences cumulées de la série **B** :



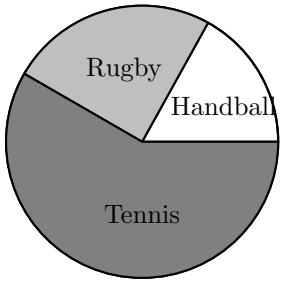
1.2.2 Variables qualitatives

Il existe plusieurs manière de représenter une variable qualitative, selon son humeur...

• **Diagrammes circulaires : les fameux camemberts...**

Chaque modalité est représentée par un secteur angulaire proportionnel à l'effetif (et donc la fréquence) associée.

Exemple 1.9. Diagramme circulaire de la **série C**



• **Diagrammes en tuyaux d'orgue et Diagrammes en bande**

Dans ces 2 représentations, chaque modalité est représentée par un rectangle de même base et de hauteur proportionnelle à l'effetif (et donc la fréquence) associée.

La différence réside dans la présentation : Dans les diagrammes en tuyaux d'orgues, les rectangles associés aux modalités sont mis côte à côte. Dans les diagramme en bande, les rectangles associés aux modalités sont mis bout à bout.

Exemple 1.10. Diagrammes de la **série C**

Diagramme en tuyaux d'orgue

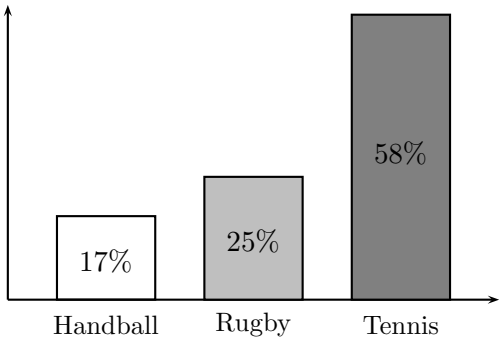
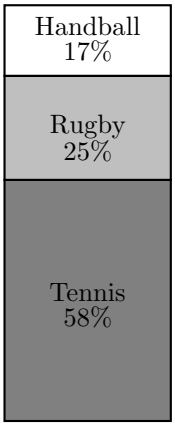


Diagramme en bande



1.3 Indicateurs de position d'une série statistique quantitative

• Mode d'un caractère

Le **mode** d'une variable statistique Mo (quantitavie ou qualitative) est la modalité ou la classe qui apparait avec le plus grand effectif (ou la plus grande fréquence si on préfère...). Pour des séries quantitatives continues, on parle plutôt de **classe modale** pour désigner l'intervalle de plus grande fréquence.

Exemples 1.11. Sur nos exemples, le mode de la série **A** est 9, celui de la série **B** est la classe $[1400, 1600]$ et celui de la série **C** est "Tennis".

• Moyenne

Pour un caratère quantitatif X, la **moyenne** ou **caractère moyen**, notée $\overline{X}$ est la moyenne des modalités pondérées avec leur fréquences d'apparition (c'est la moyenne au sens usuel du terme...).

Si les modalités de X sont les valeurs $C_1, C_2, \ldots, C_k$, qui apparaissant avec fréquences respectives $f_1, f_2, \ldots f_k$, alors

$$\overline{X} = \sum_{i=1}^{k} f_i C_i.$$

Lorsque le caractère est donnée par des classes, on calculer la moyenne en utilisant le centre de chaque classe (moyenne des 2 extrémités...) : si les modalités de X sont les intervalles $C_1 = [a_1, b_1]$, $C_2 = [a_2, b_2]$, ..., $C_k = [a_k, b_k]$, qui apparaissant avec fréquences respectives $f_1, f_2, \ldots f_k$, alors

$$\overline{X} = \sum_{i=1}^{k} f_i \frac{b_i - a_i}{2}.$$

Un caractère dont la moyenne est nulle est dit **centré**. Pour tout caractère X, le caractère $X - \overline{X}$ est dite **caractère centré** ou **variable centrée**.

Exemples 1.12.
Moyenne de la série **A** :

$$\overline{X}_{\mathrm{A}} = \frac{1}{30}(2\times1+3\times2+4\times1+5\times1+6\times2+7\times3+8\times5+9\times6+10\times2+11\times3+13\times2+15\times1+16\times1) = \frac{254}{30}$$

La note moyenne des étudiants à l'examen est donc environ : $8,47$.

Moyenne de la série **B** :

$$\overline{X}_{\mathrm{B}} = \frac{1}{200}(1050\times30+1300\times30+1500\times60+1700\times40+1900\times20+2200\times20) = \frac{310500}{200} = 1552,5$$

Le salaire moyen des employés de l'entreprise est donc de 1552,5 €.

Le choix de privilégier la moyenne arithmétique par rapport aux autres moyennes (géométrique ou harmonique) sera justifié au paragraphe 1.4.1

• **Intervalle médian et valeur médiane pour les caractères quantitatifs discrèts**

L'**intervalle médian** I_M d'un caractère quantitatif discrèt est l'ensemble des valeurs Me qui permettent de scinder la population en 2 groupes de même effectif, de sorte que la moitié de la population a un caractère plus petit ou égal à Me et l'autre partie a un caractère plus grand que Me.

Si I_m est réduit à un singleton, alors l'unique valeur possbile est appellée **médiane**, sinon, on appèle médiane le centre de l'intervalle I_M.

Attention : La parité du nombre n d'individus de la population joue un vrai rôle dans la médiane. Si on ordonne de manière croissante les valeurs de la série statistique (avec multiplicité d'apparition), alors on a :

- **si n est impair**, l'intervalle médian est réduit à un singleton et la médiane est la valeur du milieu de la série ordonnée : le $\frac{n+1}{2}$-ième terme.
- **si n est pair**, l'intervalle médian est l'intervalle compris entre le $\frac{n}{2}$-ième et le $\left(\frac{n}{2}+1\right)$-ième terme. La valeur médiane est le centre de cet intervalle.

Exemples 1.13.

- La série statistique $2-4-4-\boxed{5}-5-6-7$ a pour valeur médiane 5.
- La série statistique $2-4-4-\boxed{4-5}-5-6-7$ a pour intervalle médian $[4,5[$ et sa médiane est donc $4,5$.
- La série statistique $2-4-4-\boxed{5-5}-5-6-7$. a pour intervalle médian 5 et sa médiane est donc 5.

L'intervalle médian et/ou la médiane se lisent très bien sur les graphes des fréquences cumulées.

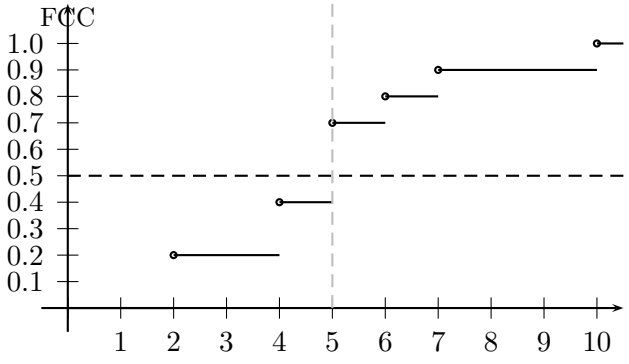
Exemple 1.14. Si on considère la (petite) série statistique des notes à un examen suivante : $2-2-4-4-5-5-5-6-7-10$.

La valeur médiane de cette série est 5.

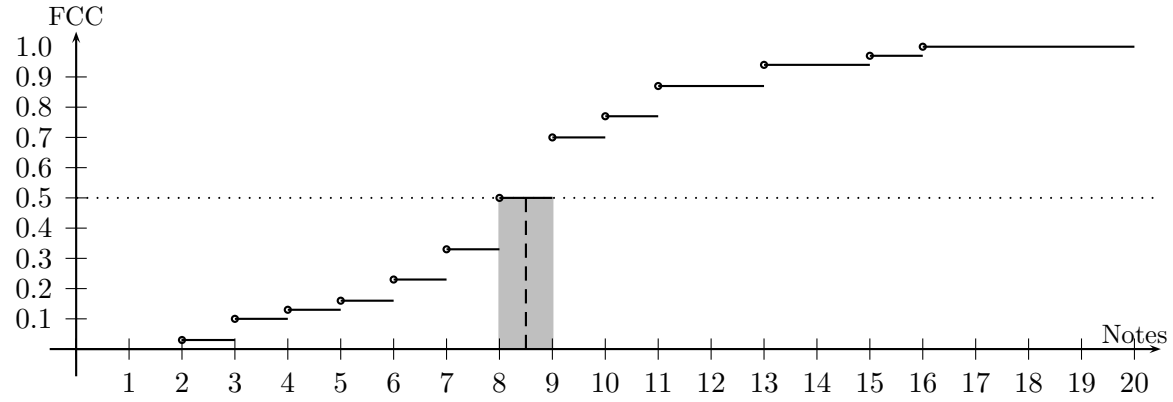
Le tableau des fréquences cumulées est donné par :

Notes	2	4	5	6	7	10
FCC en %	20	40	70	80	90	100

ce qui donne graphiquement :



Exemple 1.15. Pour la série **A** :



Ici, l'intervalle médian est $[8,9[$. La valeur médiane est $8,5$. 15 personnes ont eu une note inférieure ou égale à $8,5$.

• **Valeur médiane pour les caractères quantitatifs continus**

Lorsque le caratère est quantitatif continu, on détermine la **valeur médiane** grâce au graphe des fréquences cumulées :

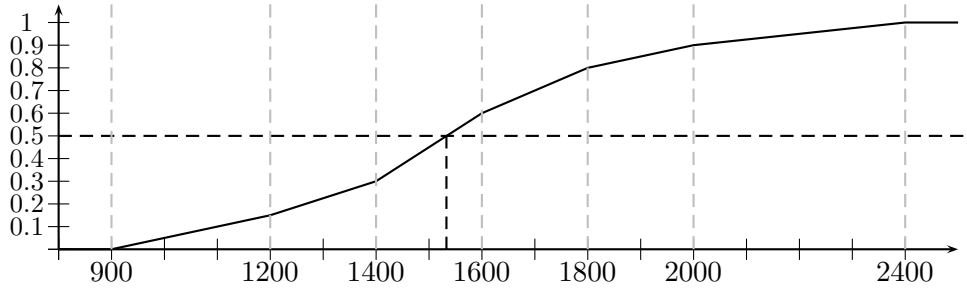
- On ordonne les classes de manière croissantes : C_1, C_2,..., C_r.
- On détermine la classe médiane $C_i = [a_i, b_i]$: la première (dans l'ordre croissant des valeurs) pour laquelle la fréquence cumulée F_i est supérieure à $0,5$ ($0,5 \in [F_{i-1}, F_i[$). cela correspond à la première classe pour laquelle les effectifs cumulés sont supérieurs à la moitié du nombre d'individus de la population.
- La valeur médiane est donnée par la formule :$Me = a_i + \dfrac{b_i - a_i}{F_i - F_{i-1}}\left(F_{i-1} - 0,5\right)$.

Il n'est ABSOLUMENT PAS NÉCESSAIRE de connnaitre cette formule par coeur ! En fait, le moyen le plus simple est encore de "voir" la valeur médiane sur le graphe des fréquences cumulées et résoudre une simplé équation de croisement de droites comme dans l'exemple ci-dessous...

Exemple 1.16. Pour la série **B**, on peut observer sur le graphe des fréquences cumulées que la classe médiane est $[1400, 1600]$, sur laquelle la courbe des fréquences cumulées est d'équation : $y = \frac{0,6-0,3}{1600-1400}(x - 1400) + 0,3$. On cherche l'antécédent de $0,5$ par cette équation :

$$0,5 = \frac{0,6 - 0,3}{1600 - 1400}(x - 1400) + 0,3 \Longleftrightarrow x = 1533,3.$$

La valeur du salaire médian est donc 1533,3€.



1.4 Indicateurs de dispersion d'une série statistique quantitative

Les indicateurs de position présenté au paragraphe précédent peuvent être complétés par des indicateurs de dispersion qui nous renseignent sur la variabilité du caractère. En effet, les deux séries statistiques 10-10-10-10-10 et 2-2-10-18-18 ont même effectif, même moyenne et même médiane mais on voit bien que ces seules données sont insuffissantes pour distinguer ces deux séries car elles ne rendent pas compte de la répatition globale des valeurs et de leur dispersion...

• Étendue d'une série statistique quantitative

C'est le premier indice de dispersion d'une série statistique : c'est l'écart entre les 2 valeurs extrémales max et min de la série statistique.

Si la série statistique est quantitative continue présentée par classe, on prend la différence entre le max de la plus grande classe et le min de la plus petite classe.

1.4.1 associés à la moyenne : variance et écart-type

• Variance

Pour un caractère quantitatif discrèt X de modalités $C_1, C_2, \ldots, C_k$ apparaissant avec fréquences respectives $f_1,\ f_2,\ \ldots f_k$ et de moyenne $\overline{X}$, la **variance** est le réel :

$$\mathbb{V}\mathrm{ar}(X) = \sum_{i=1}^{k} f_i(C_i - \overline{X})^2 = \sum_{i=1}^{k} f_i C_i^2 \ - \ \overline{X}^2.$$

Lorsque le caractère est donnée par des classes, les C_i sont les centres de chaque classe (moyenne des 2 extrémités...).

La variance est la moyenne des carrés des distances entre les valeurs et la moyenne. C'est un nombre toujours positif, nul si et seulement si le caractère est constant, égal à la moyenne.

Exemple 1.17. Pour la série **B** :

$V_{\mathbf{B}} = \frac{3}{20}(1050 - 1552.5)^2 + \frac{3}{20}(1100 - 1552.5)^2 + \frac{3}{10}(1500 - 1552.5)^2 + \frac{1}{5}(1700 - 1552.5)^2 \ldots$
$\cdots + \frac{1}{10}(1900 - 1552.5)^2 + \frac{1}{10}(2200 - 1552.5)^2 = 127768,75$

La variance n'est pas en soit très parlante même si plus elle est grande, plus le caractère sera dispersé autour de sa moyenne, ce qui est plus signficatif, c'est sa racine...

• Écart-type

Pour un caractère quantitatif X, l'écart-type $\sigma(X)$ est la racine de la variance $\mathbb{V}\mathrm{ar}(X)$:

$$\sigma(X) = \sqrt{\mathbb{V}\mathrm{ar}(X)}.$$

Comme son nom l'indique, l'écart-type donne une idée de l'écart moyen entre la moyenne et la valeur du caractère sur un individu pris au hasard.

Un caractère dont la variance vaut 1 (et donc l'écart-type vaut 1 aussi) est dit **réduit**. Pour tout caractère X, le caractère $\frac{1}{\sigma(X)}X$ est dit **caractère réduit** ou **variable réduite**.

Exemple 1.18. Pour la série **B** : $\sigma_{\mathbf{B}} = 357,44€$.

Remarque 1.19. Lorsqu'on se livre à des observations scientifiques, les mesures ne sont pas toujours exactement identiques d'une fois sur l'autre, même lorsque les conditions semblent être similaires. Il se produit ce que l'on appelle une erreur d'observation . On a la relation suivante : valeur observée = valeur exacte + erreur d'observation avec : x_o = valeur observée x_e = valeur exacte x_o - x_e = erreur d'observation On décide alors de prendre pour x_e la valeur qui minimise les erreurs d'observation, matérialisé par la moyenne des carrés de ces erreurs (critère des moindres carrés). Les calculs prouve que la meilleure valeur estimant x_e suivant ce critère est $\overline{X}$ et qu'en choisissant cette valeur pour x_e, alors le minimum de la moyenne des carrés de ces erreurs est la variance $\mathbb{V}\text{ar}(X)$. Dans ce cas-là, la moyenne des erreurs d'observations est $\sigma(X)$.

Cela ne signifie pas que $\overline{X}$ soit la valeur exacte x_e de la grandeur observée mais que c'est la meilleure évaluation possible que l'on puisse en faire selon le critère des moindres carrés.

1.4.2 associés à la médiane : Quartiles, fractiles et boîtes à moustaches !

• Quartiles, déciles, centiles et autres fractiles...

Une fois qu'on a compris l'histoire des médianes, pas de problèmes pour comprendre les quartiles, centiles et fractiles...

Les **quartiles** sont les valeurs Q_1, Q_2 et Q_3 de la série ordonnée qui permettent de scinder la population en 4 groupes de même effectif. Ils correspondent aux moment où les fréquences cumulées égales à $0,25$, $0,5$ et $0,75$.

Les **déciles** sont les valeurs D_1, D_2, ...et D_9 de la série ordonnée qui permettent de scinder la population en 10 groupes de même effectif. Ils correspondent aux moment où les fréquences cumulées égales à $0,1$, $0,2$, ...et $0,9$.

... et (je vous le donne en mille... non, en cent !), les **centiles** sont les valeurs C_1, C_2, ..., C_{99} de la série ordonnée qui permettent de scinder la population en 100 groupes de même effectif. Ils correspondent aux moment où les fréquences cumulées égales à $0,01$, $0,02$, ...et $0,99$.

Bref, la médiane, c'était en quelque sorte le "demitile" de la série caractéristique (petite private joke au passage : "Au demitile, tout est tranquille... au sommet de l'ile, tout paraît facile"... bon, c'est pourri, je vous l'accorde).

Evidemment, le deuxième quartile, le cinquième décile et le cinquantième centile sont tous égaux à la médiane :

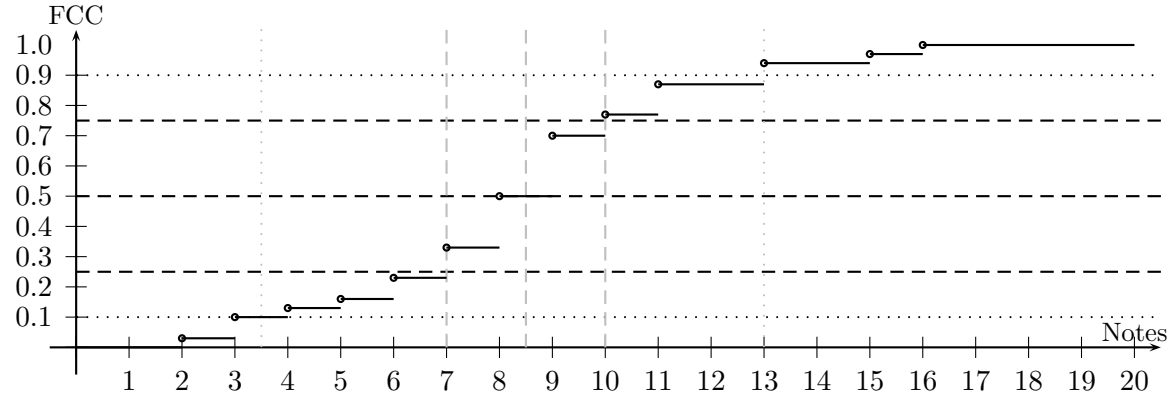
$$Me = Q_2 = D_5 = C_{50}.$$

On les calcule en utilisant les mêmes procédés que pour la médiane : en utilisant les considérations de cardinalité pour les caractères quantitatifs discrets ou à partir du graphe des fréquences cumulés directement par lecture pour les caractères discrèts et par interpolation linéaire pour les caractères continus présentés par classes.

Les centiles ne sont vraiments significatifs que pour les séries statistiques assez longues.

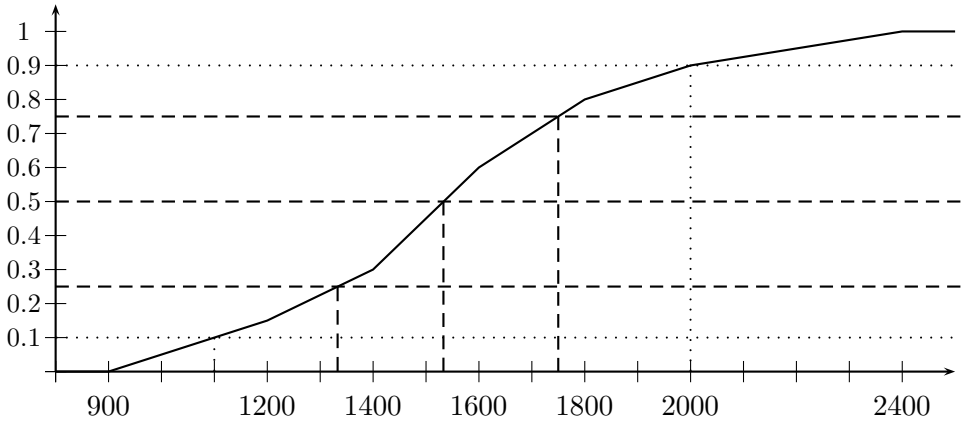
Exemple 1.20. Quartiles de la série **A** : $Q_1 = 7$, $Me = 8,5$, $Q_3 = 10$.

Premier et neuvième déciles de la série **A** : $D_1 = 3.5$ $D_9 = 13$



Exemple 1.21. Quartiles de la série **B** : On trouve graphiquement environ $Q_1 = 1330$, $Me = 1533$ et $Q_3 = 1750$.

Premier et neuvième déciles : $D_1 = 1100$ et $D_9 = 2000$.



• **Écart inter-quartiles et Écart inter-déciles**

Les quartiles et les déciles (en particulier le premier et le neuvième) sont des outils très utilisé du fait des propriétés ci-dessous :

- 50% de la population a un caractère compris entre Q_1 et Q_3,
- 25% de la population a un caractère plus petit que Q_1,
- 25% de la population a un caractère plus grand que Q_3,
- 80% de la population a un caractère compris entre D_1 et D_9.

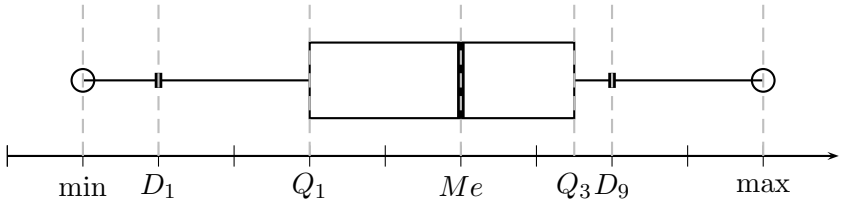
L'**écart inter-quartiles** est la quantité $Q_3 - Q_1$. Il donne l'étendue de l'intervalle où se situent 50% des notes.

L'**écart inter-déciles** est la quantité $D_9 - D_1$. Il donne l'étendue de l'intervalle où se situent 80% des notes.

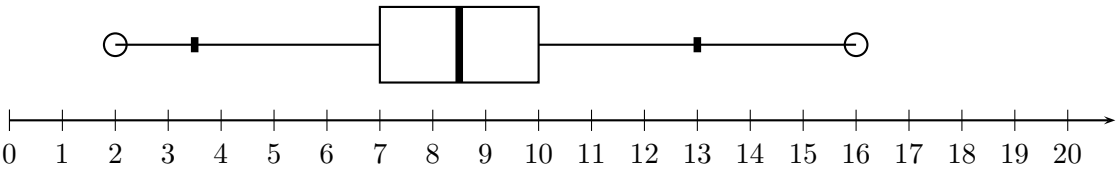
• Diagrammes en boîtes

Pour les caractères quantitatifs, les **diagrammes en boîtes**, ou **diagrammes à pattes** ou **boites à moustaches** (mon appelation préférée...) ou encore **Whiskers plots** synthétisent les différents indicateurs de positions et de dispersion associés à la médiane que sont les valeurs extrémales du caractère : minimum et maximum des modalités du caractère min et max, les quartiles (dont la médiane) Q_1, Me et Q_3 et les 1ier et 9ième déciles D_1 et D_9 (qui permettent de matérialiser les écart inter-déciles et inter-quartiles).

Ce diagramme est une première "ébauche" de la distribution de la série statistique.

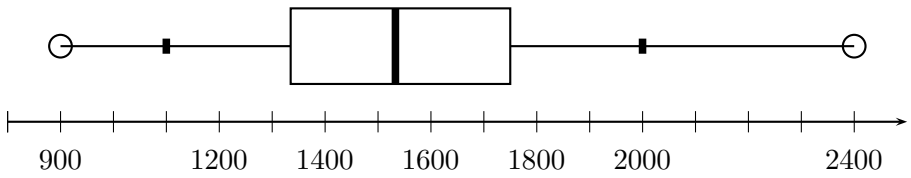


Exemple 1.22. Diagramme en boîte de la série **A** :



Sur cet exemple, si les déciles ne sont pas indiqués, rien ne laisse supposer que la distribution est relativement symétrique autour de la moyenne, ce qui n'est pas le cas... Ils donnent des informations supplémentaires sur la répartition des notes

Exemple 1.23. Diagramme en boîte de la série **B** :



Chapitre 2

Statistique descriptive : Séries bi-variées

L'étude de séries statistiques bivariées est souvent motivée par la recherche de phénomènes de corrélation entre des caractères.

L'approche est un peu différente selon que les variables étudiées sont toutes les deux catégorielles, toutes les deux continues, oul'une catégorielle et l'autre continue.

2.1 Cas de deux variables catégorielles

On considère deux caractères catégoriels A et B définis sur une population Ω, d'ensembles de modalités respectifs $(a_i)_{i\in I}$ et $(b_j)_{j\in J}$. La partition de Ω associé au couple de variables catégorielles (A, B) sera la **partition croisée** $((a_i, b_j))_{(i,j)\in I\times J}$.

Dans ce paragraphe, on considèrera pour les exemple la série statistique suivante :

Série D : Propriétés des 60 plantes d'une roseraie : couleur (Blanc, Rose, Jaune) et taille des fleurs (petites - P, moyennes - M, grandes - G, et très gandes - TG) sur chaque plant.

Couleur /Taille fleurs	P	M	G	TG
Blanc	1	9	9	5
Rose	4	6	3	3
Jaune	2	5	4	9

2.1.1 Représentation : Distribution conjointe Tableau de contingence

• Distributions conjointes et marginales des effectifs

Pour tout couple de modalité (a_i, b_j) de $I \times J$, on note n_{ij} le nombre d'individus ayant la modalité a_i du caractère A et la modalité b_j du caractère B.

La donnée des $(n_{ij})_{(i,j)\in I\times J}$ est la **distritubtion conjointe des effectifs**.

On pose $n_{i\cdot} = \sum_{j\in J} n_{ij}$ et $n_{\cdot j} = \sum_{i\in I} n_{ij}$.

Les distributions **distributions marginales des effectifs** sont données par $(n_{i\cdot})_{i\in I}$ pour le caractère A et par et $(n_{\cdot j})_{j\in J}$ pour le caractère B.

Le nombre total d'individus est bien sûr $n = \sum_{(i,j)\in I\times J} n_{ij} = \sum_{j\in J} n_{i\cdot} = \sum_{j\in J} n_{\cdot j}$.

De même, on note les fréquences associées à chaques couples de modalités : $f_{ij} = \frac{n_{ij}}{n}$ et on pose $f_{i\cdot} = \sum_{j\in J} f_{ij}$ et $f_{\cdot j} = \sum_{i\in I} f_{ij}$, de sorte que $\sum_{(i,j)\in I\times J} n_{ij} = \sum_{j\in J} n_{i\cdot} = \sum_{j\in J} n_{\cdot j} = 1$.

On nomme de la même manière $(f_{ij})_{(i,j)\in I\times J}$ la **distribution conjointe des fréquences** ainsi que $(f_{i\cdot})_{i\in I}$ et $(f_{\cdot j})_{j\in J}$ les deux **distributions marginales des fréquences** de A et de B.

• **Tableau de contingence**

Le **tableau de contingence des caractères A et B** regroupe d'une part la distritubtion conjointe des effectifs : $(n_{ij})_{(i,j)\in I\times J}$ ainsi que les deux distributions marginales des effectifs.

Usuellement, le premier caractère A indice sur les lignes et le deuxième caractère B indice les colonnes.

Exemple 2.1. Tableau de contingence de la série **D** :

	P	M	G	TG	Population
Blanc	0	9	9	6	24
Rose	4	6	3	3	16
Jaune	2	5	4	9	20
Population	6	20	16	18	60

Distribution conjointe des fréquences de la série **D** donnée en pourcentages :

	P	M	G	TG	Population
Blanc	0	15	15	10	40
Rose	6,7	10	5	5	26,7
Jaune	3,3	8,3	6,7	15	33,3
Population	10	33,3	26,7	30	100

2.1.2 Distributions conditionnelles (profils lignes et profils colonnes)

La **distribution du caractère B conditionnellement au sous-ensemble** $\{A = a_i\}$ est la donnée des fréquences du caractère B parmi les individus ayant a_i pour modalité du caractère A. C'est donc la donnée des $\left(\frac{n_{ij}}{n_{i\cdot}}\right)_{j\in J}$.

On a autant de distribution conditionnelle de B que de modalité pour A (nombre de lignes du tableau de contingence).

Ces **profil lignes** sont représentés dans un tableau regroupant ces lignes, ainsi qu'une ligne supplémentaire contenant la distribution marginale de B : $\left(\frac{n_{\cdot j}}{n}\right)_{j\in J}$.

Exemple 2.2. Profils lignes de la série **D** en pourcentages :

$B\|A$	P	M	G	TG	Population
Blanc	0	37,5	37,5	25	100
Rose	25	37,5	18,75	18,75	100
Jaune	10	25	20	45	100
Freq. marginale de B	10	33,3	26,7	30	100

De la même manière, la **distribution du caractère A conditionnellement au sous-ensemble** $\{B = b_i\}$ est la donnée des fréquences du caractère A parmi les individus ayant b_i pour modalité du caractère B. C'est donc la donnée des $\left(\frac{n_{ij}}{n_{\cdot j}}\right)_{i\in I}$.

On a autant de distribution conditionnelle de A que de modalité pour B (nombre de colonnes du tableau de contingence).

Ces **profil colonnes** sont représentés dans un tableau regroupant ces colonnes, ainsi qu'une colonne supplémentaire contenant la distribution marginale de A : $\left(\frac{n_{i\cdot}}{n}\right)_{i\in I}$.

Exemple 2.3. Profils colonnes de la série **D** en pourcentages :

$A\|B$	P	M	G	TG	Freq. marginale de A
Blanc	0	45	56,75	16,7	40
Rose	66,7	30	18,75	33,3	26,7
Jaune	33,3	25	25	50	33,3
Population	100	100	100	100	100

2.1.3 Ecart à l'indépendance : Indice du χ^2

L'intérêt de l'étude conjointe de 2 variables catégorielle est avant tout de savoir si les 2 caractères étudiés sont liés ou non... L'absence de liaisons peut se traduire par une des 3 propriétés équivalentes suivantes :

- Les profils lignes sont identiques : Pour tout $(i,j) \in I \times J$, $\frac{n_{ij}}{n_{i\cdot}} = \frac{n_{\cdot j}}{n}$,
- Les profils colonnes sont identiques : Pour tout $(i,j) \in I \times J$, $\frac{n_{ij}}{n_{\cdot j}} = \frac{n_{i\cdot}}{n}$,
- Pour tout $(i,j) \in I \times J$, $\frac{n_{ij}}{n} = \frac{n_{i\cdot} n_{\cdot j}}{n}$,

On peut dresser le **tableau théorique d'indépendance** construit à partir des distributions marginales des effectifs de A et de B, dont les éléments sont les $\left(\frac{n_{i\cdot} n_{\cdot j}}{n}\right)_{(i,j)\in I\times J}$, qui serait le tableau obtenu exactement si les 2 caractères étaient indépendants.

Exemple 2.4. Tableau théorique d'indépendance de la série **D** :

$A\|B$	P	M	G	TG	Distrib. marginale de A
Blanc	2,4	8	6,4	7,2	24
Rose	1,6	5,3	4,3	4,8	16
Jaune	2	6,7	5,3	6	20
Distrib. marginale de B	6	20	16	18	60

Ce tableau permet de contruire le **tableau des écarts entre les effectifs observés et les effectifs théoriques d'indépendance**, formé des écarts $\left(n_{ij} - \frac{n_{i\cdot}n_{\cdot j}}{n}\right)_{(i,j)\in I\times J}$.

La somme de chaque ligne et de chaque colonne doit être nulle. Certains écarts sont positifs, indiquant une **sur-représentation du couple de modalités** correspondant par rapport à l'indépendance, les écart négatifs indiquent quant à eux une **sous-représentation du couple de modalités** correspondant par rapport à l'indépendance.

Exemple 2.5. Tableau des écarts entre les effectifs observés et les effectifs théoriques d'indépendance de la série **D** :

$A\|B$	P	M	G	TG
Blanc	-2,4	1	2,6	-1,2
Rose	2,4	-0,3	-1,3	-1,8
Jaune	0	-1,7	-1,3	3

• **Indice du χ^2**

En fait, les écarts entre les effectifs observés et théorique d'indépendance ne sont en fait pas très représentatifs, notamment car il dépendent grandement de la taille de la population : plus il y a d'individus, plus les écarts peuvent être grands (même si en proportion, ils sont infimes...). Ce qui est important, c'est donc plus le ratio $\frac{n_{ij} - \frac{n_{i\cdot}n_{\cdot j}}{n}}{\frac{n_{i\cdot}n_{\cdot j}}{n}}$ qui donne une idée de la vrai différence entre effectifs réels et effectifs théoriques.

L'indice du χ^2 permet de quantifier la distance entre la distribution observée et la distribution théorique qu'on aurait du observer si les deux caractères avaient étés indépendants. (le ^{2} dans χ^2 indique qu'on quantifie en fait le carré d'une distance...), avec l'idée que plus le χ^2 d'une série statistique est grande plus les caractères sont dépendants l'un de l'autre,

Le χ^2 de la série statistique est donné par la formule suivante :

$$\chi^2 = \sum_{(i,j)\in I\times J} \frac{\left(n_{ij} - \frac{n_{i\cdot}n_{\cdot j}}{n}\right)^2}{\frac{n_{i\cdot}n_{\cdot j}}{n}} = n \sum_{(i,j)\in I\times J} \frac{f_{i,j} - f_{i\cdot}f_{\cdot j}}{f_{i\cdot}f_{\cdot j}}.$$

Exemple 2.6. Indice du χ^2 de la série statistique **D** :

$$\chi^2 = \frac{2,4^2}{2,4} + \frac{1^2}{8} + \frac{2,6^2}{6,4} + \frac{1,2^2}{7,2} + \frac{2,4^2}{1,6} + \frac{0,3^2}{5,3} + \frac{1,3^2}{4,3} + \frac{1,8^2}{4,8} + 0 + \frac{1,7^2}{6,7} + \frac{1,3^2}{5,3} + \frac{3^2}{6} \approx 10,72.$$

Pour raffiner la connaissance des liens de dépendance qui unissent les caractères A et B, on peut dresser le **tableau des contributions au χ^2** qui indique quels couples de modalités (a_i, b_j) contribuent le plus à l'indépendance. Pour tout couple $(i,j) \in I \times J$, on définit la **contribution du couple de modalités (a_i, b_j) au χ^2** par :

$$ctr_{ij} = \frac{1}{\chi^2} \cdot \frac{\left(n_{ij} - \frac{n_{i\cdot}n_{\cdot j}}{n}\right)^2}{\frac{n_{i\cdot}n_{\cdot j}}{n}} = \frac{n}{\chi^2} \cdot \frac{f_{i,j} - f_{i\cdot}f_{\cdot j}}{f_{i\cdot}f_{\cdot j}}.$$

La somme des éléments du tableau des contributions au χ^2 est égale à 1 (ou environ égal à 1 si les valeurs sont approchées).

Ce tableau est en fait plus représentatif du défaut d'indépendance que celui des écarts d'effectifs.

Exemple 2.7. Tableau des contributions au χ^2 de la série **D** (en valeurs approchées) :

$A\|B$	P	M	G	TG
Blanc	0,22	0.01	0,10	0,02
Rose	0,33	0,001	0,03	0,06
Jaune	0	0,04	0,03	0,14

2.2 Cas d'une variable catégorielle et d'une variable réelle

On considère un caractère catégoriels A d'ensemble de modalités $(a_i)_{i\in I}$ et un caractère réel X définis sur une population Ω.

La partition de Ω associée à la variable catégorielle A est notée $(A_i)_{i\in I}$, avec $A_i = X^{-1}(\{a_i\}) = [X = a_i]$. Pour tout $i \in I$, on note n_i l'effectif de A_i.

Remarque : On peut utiliser bien sûr les résultats et objets de ce paragraphe lorsque la variable X est continue ou discrète.

Dans ce paragraphe, on considèrera pour les exemples la **Série E** :

45 poissons d'une même espèce (Ti jaune ou *Lutjanus Casmira*) sont prélevés dans le lagon. On note leur taille et après dissection, le contenu de leur estomac : Mollusques, Crustacés ou Poissons (données grâcieusement fournies par Cécile Mablouké, Doctorante en Biologie Marine).

Taille en mm	109	92	78	105	131	92	76	91	96	87	85	98	136	130	94
Régime	M	M	M	M	M	M	M	M	C	C	C	C	C	C	C
Taille en mm	144	87	78	114	112	98	116	140	92	203	112	144	94	118	159
Régime	C	C	C	C	C	C	C	C	C	C	P	P	P	P	P
Taille en mm	113	131	133	103	88	111	134	95	98	92	87	76	78	109	112
Régime	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P

2.2.1 Représentation : Boites parallèles

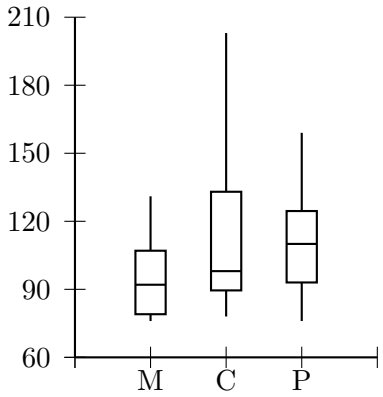
Sur chaque ensemble de la partition $(A_i)_{i\in I}$, on calcule la médiane, les quartiles, le maximum et le minimum de B. Cela permet de représenter les diagrammes en boites associé au caractère B sur l'ensemble des individus de la population ayant a_i comme modalité du caractère A.

Les diagrammes en boites de B selon leur modialité de A sont alignés verticalement sur un unique graphique.

Un sens (A selon les modalité de B ou B selon les modalité de A) est souvent plus pertinent que l'autre si les deux caractères sont quantitatif, mais cela dépend de l'analyse qui veut être faite.

On utilise la représentation de B en fonction de A lorsque A est qualitative et B quantitative.

Exemple 2.8. Pour la série **E** :
Pour les poissons mangeurs de mollusques : min $= 76$, max $= 131$, $Me = 92$, $Q_1 = 79,5$ $Q_3 = 107$.
Pour les poissons mangeurs de crustacés : min $= 78$, max $= 203$, $Me = 98$, $Q_1 = 89,5$ $Q_3 = 133$.
Pour les poissons mangeurs de poissons : min $= 76$, max $= 159$, $Me = 110$, $Q_1 = 93$ $Q_3 = 124,5$.



2.2.2 Décomposition de la moyenne et de la variance sur une partition

Pour tout $i \in I$, on note $\overline{X}_i$ et $V_i(X)$ **la moyenne et la variance du caractère X conditionnées à la modalité a_i de A** : ce sont la moyenne et la varianc calculées pour les seuls individus de A_i, c'est-à-dire :

$$\overline{X}_i = \frac{1}{n_i} \sum_{\omega \in A_i} X(\omega) \quad \text{et} \quad V_i(X) = \frac{1}{n_i} \sum_{\omega \in A_i} \left(X(\omega) - \overline{X}_i \right)^2 .$$

Remarque :
La moyenne $\overline{X}$ du caractère X est la moyenne des moyennes conditionnées $\overline{X}_i$ pondérées par les effectifs des modalités de A :

$$\overline{X} = \frac{1}{n} \sum_{i \in I} n_i \overline{X}_i = \sum_{i \in I} f_i \overline{X}_i,$$

si f_i est la fréquence d'apparition de la modalité a_i du caractère A.

Ces moyennes et variances de X conditionnées au caractère A permettent de définir la **variance expliquée** ou **variance inter-classe** (ou inter-groupes ou encore inter-catégories) comme la variance des moyennes.

$$V_{inter}(X) = \frac{1}{n} \sum_{i \in I} n_i \left(\overline{X}_i - \overline{X} \right)^2 = \sum_{i \in I} f_i \left(\overline{X}_i - \overline{X} \right)^2 .$$

De même on définit la **variance résiduelle** ou **variance intra-classe** (ou intra-groupes ou encore intra-catégories) qui résume la variabilité à l'intérieur des classes : c'est la moyenne des variances V_i pondérées par les effectifs des modalités de A :

$$V_{intra}(X) = \frac{1}{n} \sum_{i \in I} n_i V_i(X) = \sum_{i \in I} f_i V_i(X).$$

On peut montrer facilement que :

$$\mathbb{V}\mathrm{ar}(X) = V_{inter}(X) + V_{intra}(X).$$

2.2.3 Rapport de corrélation

La variable catégorielle A et la variable réelle X sont liées si, dans chaque catégorie de la partition associée à A, les individus sont homogènes vis-à-vis du caractère X, c'est-à-dire si la variance intra-catégorielle ne contribue que faiblement à la variance. Cette propriété est décrite par le **coefficient de corrélation entre A et X** :

$$\eta = \sqrt{\frac{V_{inter}(X)}{\mathbb{V}\mathrm{ar}(X)}}.$$

Cet indice est compris entre 0 et 1. Il est égal à 1 si $V_{inter}(X) = \mathbb{V}\mathrm{ar}(X)$, c'est-à-diresi $V_{intra}(X) = 0$ et donc si pour tout $i \in I$, $V_i(X) = 0$: Pour chaque modalité de A les valeurs prises par X sont toutes les mêmes.

Plus le coefficient est proche de 1, plus la donnée de la modalité de A "va influencer" la valeur de X.

La coefficient de corrélation est nul lorsque $V_{inter}(X) = 0$, c'est-à-dire si pour tout $i \in I$, $\overline{X}_i = \overline{X}$: Dans chaque modalité de A, les valeurs prises par X sont équitablement réparties autour de la moyenne de X : il y a une vraie hétérogénéïté des valeurs de X au sein de chaque population A_i.

Plus le coefficient de corrélation est proche de 0, moins la donnée de la modalité de A "va influencer" la valeur de X.

Exemple 2.9. Pour les données de la série **E** :
Moyenne et variance des tailles des Ti jaunes mangueurs de mollusques :

$$\overline{X}_M = 96,75 \quad \text{et} \quad V_M(X) = 281,48 \text{ (écart-type } \sigma_M(X) = 16,78)$$

Moyenne et variance des tailles des Ti jaunes mangueurs de crustacés :

$$\overline{X}_C = 112,35 \quad \text{et} \quad V_C(X) = 912,58 \text{ (écart-type } \sigma_C(X) = 30,02)$$

Moyenne et variance des tailles des Ti jaunes mangueurs de poissons :

$$\overline{X}_P = 108,24 \quad \text{et} \quad V_P(X) = 468,438 \text{ (écart-type } \sigma_M(X) = 21,64)$$

Moyenne et variance de la population :

$$\overline{X} = 108,24 \quad \text{et} \quad \mathbb{V}\mathrm{ar}(X) = 632,98$$

Variance inter-classes et intra-classes :

$$V_{inter}(X) = 30,41 \quad \text{et} \quad V_{intra}(X) = 602,98$$

Il y a une très forte variabilité à l'interieur des classes (V_{intra} grand) mais les moyennes varient très peu d'une classe à l'autre (V_{inter} petit). Cela est confirmé par le coefficient de corrélation Taille/alimentation : $\eta = 0,22$ qui est relativement petit : La taille des poissons semble être assez très peu corrélée à la taille des poissons...

2.3 Cas de deux variables réelles

On considère deux caractères réels X et Y définis sur une population Ω de n individus.
On note $(X,Y)(\Omega)$ l'image de Ω par (X,Y), c'est-à-dire :

$$(X,Y)(\Omega) = \{(X(\omega), Y(\omega)) \ : \ \omega \in \Omega\}\,.$$

On note $\{(x_i, y_i) \ : \ i \in I\}$ l'ensemble des couples distincts de Ω.

Pour ce paragraphe, on considère la série statistique suivante :

Série F

Sur 30 poissons Ti-jaunes récupérés dans le lagon (encore merci Cécile !), on mesure le pourcentage d'azote (N) et le pourcentage de carbone (C) dans leurs muscles. Les données sont résumées dans ce tableau :

% N	%C	% N	%C	% N	%C	% N	%C	% N	%C	% N	%C
13,32	46,48	13,99	45,25	13,92	45,06	13,68	44,63	13,93	44,65	12,89	42,22
13,83	44,72	13,94	44,15	13,96	45,16	14,14	44,89	13,63	43,61	14,10	45,44
13,66	44,52	14,71	46,49	14,82	47,82	13,47	45,05	13,65	45,58	13,09	45,42
14,72	47,33	15,05	47,70	14,20	45,23	14,37	47,30	12,73	39,90	14,91	46,06
14,93	45,84	15,12	47,75	13,66	44,13	14,47	45,88	13,95	45,23	14,35	46,83

2.3.1 Distribution d'effectifs, distribution de fréquences et nuage de points

• Distributions d'éffectifs et de fréquence

Pour tout $(x_i, y_i) \in (X,Y)(\Omega)$, on définit

$$n_i = \text{Card}\,\{\omega \in \Omega \ : \ X(\omega) = x,\, Y(\omega) = y.\} \quad \text{et} \quad f_i = \frac{n_i}{n}.$$

Le cas qui nous intéresse ici est celui où les 2 variables sont continues, et même si cela est peu probable, il se peut que certains individus aient exactement les même caractère X et Y simultanéement. On doit doit alors tout de même définir la **distribution d'effectifs** (*resp.* **distribution des fréquences**), donnée par l'ensemble :

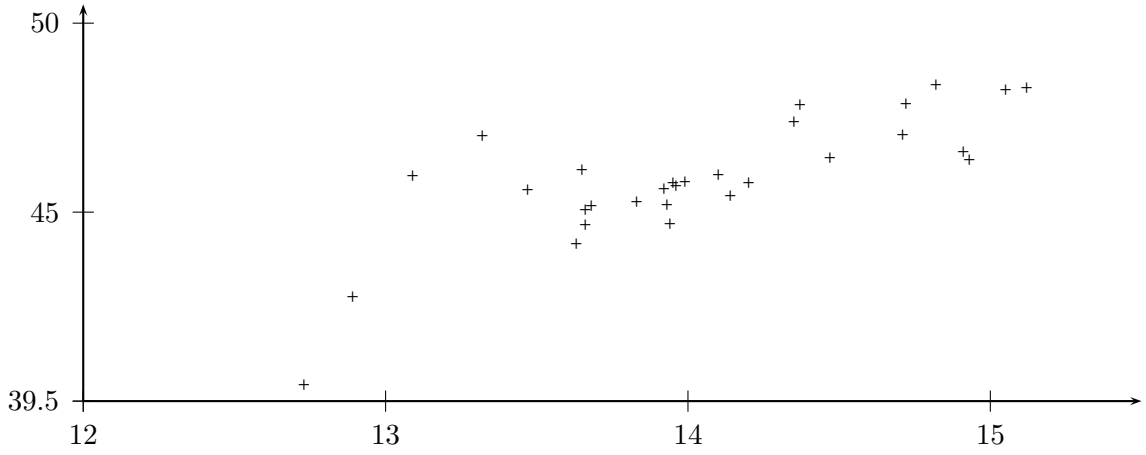
$$\{((x_i, y_i), n_i) \ : \ i \in I\} \quad (\text{resp. } \{((x_i, y_i), f_i) \ : \ i \in I\})$$

• Nuage de points

On peut représenter le **nuage de points** associé au couple de caractères (X,Y) en représentant dans un repère les couples (x_i, y_i) par le point de coordonnées (x_i, y_i). On peut également adapter cette représentation au cas où certains effectifs sont plus grand que 1 en représentant chaque couple (x_i, y_i) par un disque centré au point (x_i, y_i) et de rayon proportionnel à l'effectif.

On appelle **point moyen** le point de coordonnées $(\overline{X}, \overline{Y})$, c'est le centre de gravité du nuage de points.

Exemple 2.10. Nuage de points associé à la série **F** :



2.3.2 Covariance et coefficient de corrélation linéaire

• **Covariance**

La **covariance des caractères** X **et** Y est définie comme la moyenne des produits des différences à la moyenne :

$$\mathbb{C}\text{ov}(X,Y) = \frac{1}{n}\sum_{\omega\in\Omega}(X(\omega)-\overline{X})(Y(\omega)-\overline{Y}) = \frac{1}{n}\sum_{i\in I} n_i(x_i-\overline{X})(y_i-\overline{Y}).$$

L'unité de la covariance est l'unité produit des unités des 2 variables.

On a bien sur $\mathbb{C}\text{ov}(Y,X) = \mathbb{C}\text{ov}(X,Y)$, $\mathbb{C}\text{ov}(X,X) = \mathbb{V}\text{ar}(X)$ et si l'une des deux variable$X$ ou Y est constante, alors $\mathbb{C}\text{ov}(X,Y) = 0$.

Exemple 2.11. Pour la série **F**, on peut à l'aide d'un tableau excel par exemple, rapidement calculer que $\mathbb{C}\text{ov}(N,C) = 0,78$. Son unité est le un pour un dix-millième.

• **Coefficient de corrélation linéaire** Dans le cas où aucun des deux caractères n'est constant ($\mathbb{V}\text{ar}(X)$ et $\mathbb{V}\text{ar}(Y)$ non nuls), on définit le **Coefficient de corrélation linéaire** par le rapport de la covariance de X et Y sur le produit des écart-types $\sigma(X)$ et $\sigma(Y)$:

$$r(X,Y) = \frac{\mathbb{C}\text{ov}(X,Y)}{\sigma(X)\sigma(Y)} = Cov\left(\frac{X-\overline{X}}{\sigma(X)}, \frac{Y-\overline{Y}}{\sigma(Y)}\right).$$

cet indice n'a aucune unité de mesure.

Quelques propriétés du coefficient de corrélation linéaire :

- $r(X,Y) = r(Y,X)$,
- $r(X,Y) \in [-1,1]$,
- $|r(X,Y)| = 1$ si et seulement s'il existe $(a,b) \in \mathbb{R}^* \times \mathbb{R}$ tels que $Y = aX + b$.
  Le signe de a est donné par le signe de $r(X,Y)$.

Le signe de $r(X,Y)$ nous indique le sens de liaison : s'il est positif, les 2 caractères on tendance à varier dans le même sens (nuage plutot ascendant). Sinon, les 2 caractères varient en sens inverse (nuage plutot descendant).

La dernière propriété nous indique que plus $|r(X,Y)|$ est proche de 1, plus la variable Y a de chance d'être proche d'une combinaison linéaire de X et donc fortement corrélée à X...

Exemple 2.12. Pour la série **F**, les variances de N et C sont respectivement : $\mathbb{V}\text{ar}(N) = 0,37$ et $\mathbb{V}\text{ar}(C) = 2,63$, d'où

$$r(N,C) = \frac{0,78}{\sqrt{0,37}\sqrt{2,63}} = 0,78.$$

2.4 Régression linéaire

Lorsque deux variables quantitatives sont assez corrélées ($r(X,Y)$ proche de 1), et que l'une (ici X) est a priori la cause de l'autre (ici Y), il est assez naturel de chercher une fonction f de X qui approche Y "au mieux". La méthode statistique permattant de trouver cette fonction f s'appelle **régression de Y sur X**.

Pour mettre en oeuvre une régression, deux choses préalables sont nécessaires :

1. choisir l'ensemble de fonctions parmi lesquelles on va chercher f.
2. définir mathématiquement la notion de "au mieux" évoquée au paragraphe ci-dessus.

Si l'ensemble des fonctions parmi lesquelles on cherche f est l'ensemble des applications affines ($x \mapsto ax + b$), on parle de **régression linéaire**.

Ce paragraphe présente 3 méthodes de régression linéaires dont le plus usuel et éfficace est le dernier : la méthode des moindres carrés.

2.4.1 Méthode de Mayer

La **méthode de Mayer** consiste à scinder le nuage de points en 2 parties (qui diffèrent d'au plus un individu) selon les valeurs de X, puis à calculer les points moyens respectifs de ses 2 sous-nuages de points. La **droite de Mayer** est la la droite qui joint ces 2 points.

Les points de $\{(x_i, y_i) : i \in, I\}$ sont ordonnés selon les x_i croissants. Pour une même valeur de x_i, lescouples (x_i, y) sont ordonnés par y croissants. Soit G_1 l'ensemble des $\lfloor \frac{n}{2} \rfloor$ premiers couples du nuage ordonné et $G_2 = (X,Y)(\Omega) \setminus G_1$.

On note $(\overline{X}_1, \overline{Y}_1)$ le point moyen des points de G_1 et $(\overline{X}_2, \overline{Y}_2)$ le point moyen des points de G_2.

La droite de Mayer est la droite que relie $(\overline{X}_1, \overline{Y}_1)$ à $(\overline{X}_2, \overline{Y}_2)$.

Exemple 2.13. Pour la série **F** :

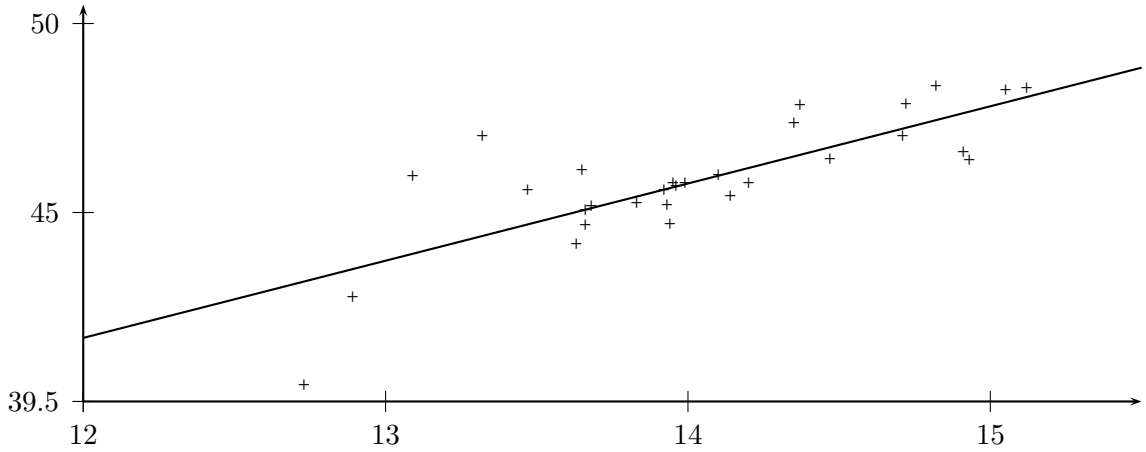
Données réordonnées (par lignes) :

% N	%C	% N	%C	% N	%C	% N	%C	% N	%C	% N	%C
12,73	39,90	12,89	42,22	13,09	45,42	13,32	46,48	13,47	45,05	13,63	43,61
13,65	45,58	13,66	44,52	13,66	44,13	13,68	44,63	13,83	44,72	13,92	45,06
13,93	44,65	13,94	44,15	13,95	45,23	13,96	45,16	13,99	45,25	14,10	45,44
14,14	44,89	14,20	45,23	14,35	46,83	14,37	47,30	14,47	45,88	14,71	46,49
14,72	47,33	14,82	47,82	14,91	46,06	14,93	45,84	15,05	47,70	15,12	47,75

Point moyen de la première partie du tableau : $M_1 = (13.46, 44.36)$ (2 premières lignes et première moitié de la troisième ligne).

Point moyen de la première partie du tableau : $M_2 = (14.52, 46.33)$ (deuxième moitié de la troisième ligne et 2 dernières ligne).

Equation de la droite passant par M_1 et M_2 : $y = 2,04x + 16,68$.



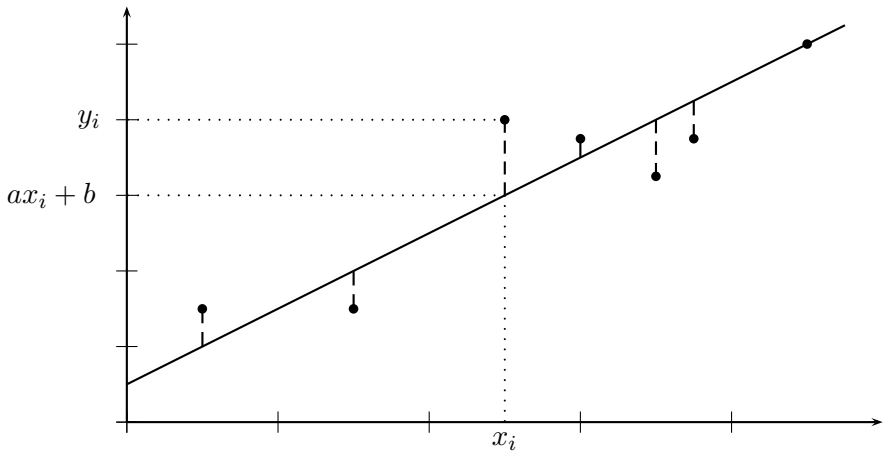
2.4.2 Méthode des moindres carrés

La méthode des moindres carré consiste à trouver l'équation de la droite pour laquelle la somme des carrés des distances (verticales) entre les points du nuage et les points de la droite est minimale.

On cherche dont a et b tels que la quantité :

$$F(a,b) = \sum_{\omega \in \Omega} \left(Y(\omega) - (aX(\omega) + b)\right)^2 .$$

Graphiquement, on cherche à minimiser la somme des carrés des distances entre y_i et $ax_i + b$.



La quantité $F(a,b)$ permet de quantitifer l'erreur globale commise en approchant Y par $aX + b$. L'intérêt de considérer les carrés de ces valeurs permet d'éviter les compensations qu'il pourrait y avoir entre les erreurs positives et négatives.

La fonction F attend son minimum au point $(\hat{a}, \hat{b})$, définis par :

$$\hat{a} = \frac{\mathbb{C}\text{ov}(X,Y)}{\mathbb{V}\text{ar}(X)} \quad \text{et} \quad \hat{b} = \overline{Y} - a\overline{X}.$$

Remarque : X est la variable de base, privilégiée comme cause de Y : on ne trouvera pas la même droite si on cherche plutôt à expliquer X en fonction de Y : En effectuant la régression $X = \hat{c}Y + \hat{d}$ on aura pas forcément $\hat{a} = \frac{1}{\hat{c}}$ et $\hat{b} = -\frac{\hat{d}}{\hat{c}}$.

Pour démontrer les valeurs de $\hat{a}$ et $\hat{b}$, on peut considérer la fonction $(a,b) \mapsto F(a,b)$, dont l'unique point critique (point où les dérivées partielles s'annulent) est $(\hat{a}, \hat{b})$ et vérifier que c'est effectivement un minimum (matrice Hessienne à valeurs propres positives). Sinon, on peut le faire "à la main" comme décrit ci-dessous :

$$F(a,b) = \sum_{\omega\in\Omega} \left(Y(\omega) - (aX(\omega) + b)\right)^2 = \sum_{i\in I} n_i \left(y_i - \overline{Y} - a(x_i - \overline{X}) + \overline{Y} - a\overline{X} - b\right)^2$$

$$F(a,b) = \sum_{i\in I} n_i \left(y_i - \overline{Y} - a(x_i - \overline{X})\right)^2 + 2\left(\overline{Y} - a\overline{X} - b\right) \sum_{i\in I} \left(y_i - \overline{Y} - a(x_i - \overline{X})\right) + \sum_{i\in I} \left(\overline{Y} - a\overline{X} - b\right)^2,$$

or $n\overline{X} = \sum_{i\in I} x_i$ et $n\overline{Y} = \sum_{i\in I} y_i$ donc le second terme de la somme est nul. et ainsi :

$$F(a,b) = \sum_{i\in I} n_i \left(y_i - \overline{Y} - a(x_i - \overline{X})\right)^2 + n\left(\overline{Y} - a\overline{X} - b\right)^2.$$

De plus en réarrangeant le premier terme, on obtient :

$$\sum_{i\in I} n_i \left(y_i - \overline{Y} - a(x_i - \overline{X})\right)^2 = \sum_{i\in I} n_i \left(y_i - \overline{Y}\right)^2 - 2a \sum_{i\in I} n_i \left(y_i - \overline{Y}\right)\left(x_i - \overline{X}\right) + a^2 \sum_{i\in I} n_i \left(x_i - \overline{X}\right)^2$$

$$\sum_{i\in I} n_i \left(y_i - \overline{Y} - a(x_i - \overline{X})\right)^2 = \mathbb{V}\text{ar}(Y) - 2a\mathbb{C}\text{ov}(X,Y) + a^2\mathbb{V}\text{ar}(X).$$

Le minimum du polynome en a est atteint lors du changement de signe de sa dérivée (car on a $\mathbb{V}\text{ar}(X) > 0$), c'est-à-dire en $\hat{a} = \dfrac{\mathbb{C}\text{ov}(X,Y)}{\mathbb{V}\text{ar}(X)}$.

Le second terme $n\left(\overline{Y} - a\overline{X} - b\right)^2$ est toujours positif ou nul, et lorsque a est fixé, il est nul si et seulement en $\hat{b} = \overline{Y} - a\overline{X}$. Ainsi, le minimum de $F(a,b)$ est donc obtenu en $(\hat{a}, \hat{b})$.

On en déduit donc l'équation de la **droite de régression linéaire des moindres carrés** :

$$y = \overline{Y} + \frac{\mathbb{C}\text{ov}(X,Y)}{\mathbb{V}\text{ar}(X)}(x - \overline{X}).$$

Le **modèle ajusté de** Y, noté $\hat{Y}$ est le caractère :

$$\hat{Y} = \frac{\mathbb{C}\text{ov}(X,Y)}{\mathbb{V}\text{ar}(X)}\, X + \overline{Y} - a\overline{X}.$$

Ce modèle ajusté peut permettre de prévoir la valeur de Y correspondant à une valeur x de X donnée. On parle d'**interpolation** si x est contenu dans l'intervalle des valeurs déjà prises par X sur la population. Lorsque x n'est pas dans l'intervalle des valeurs prises par f dans la population, on parle d'**extrapolation**.

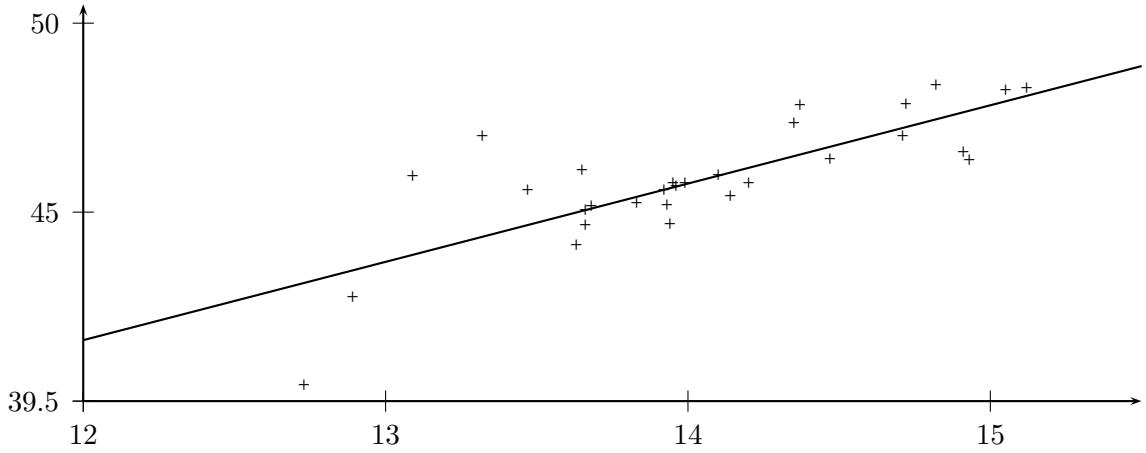
Evidemment, cette "prédiction" de la valeur prise par Y peut à priori faire apparaitre des erreurs, quantifiés par la **variable des résidus** ou **erreur d'ajustement** $\hat{E}$, qui est le caractère définit par :

$$\hat{E} = Y - \hat{Y}.$$

C'est une variable centrée dont la variance est $F(\hat{a}, \hat{b})$ (qui est minimum de F).

On a les relations suivantes :
- $\mathbb{V}\mathrm{ar}(\hat{Y}) = r(X,Y)^2 \mathbb{V}\mathrm{ar}(Y)$,
- $\mathbb{V}\mathrm{ar}(\hat{E}) = \mathbb{V}\mathrm{ar}(Y)(1 - r(X,Y)^2)$,
- $\mathbb{V}\mathrm{ar}(Y) = \mathbb{V}\mathrm{ar}(\hat{Y}) + \mathbb{V}\mathrm{ar}(\hat{E})$ et $\mathbb{C}\mathrm{ov}(\hat{Y}, \hat{E}) = 0$.

Exemple 2.14. Pour la série **F** : les calculs donnent $\hat{a} = 2,07$ et $\hat{b} = 16,24$



Sur cet exemple, les droites obtenues par la méthode de Mayer et par la méthode des moindres carrés sont quasiment confondues mais ce n'est pas pas toujours le cas...

Cela dit, ces droites de regression dépendent fortement de la population : dès qu'on rajoute un point (surtout s'il a des coordonnées loin des moyennes !), on peut fortement modifier la droite de regression.

2.4.3 Régression orthogonale

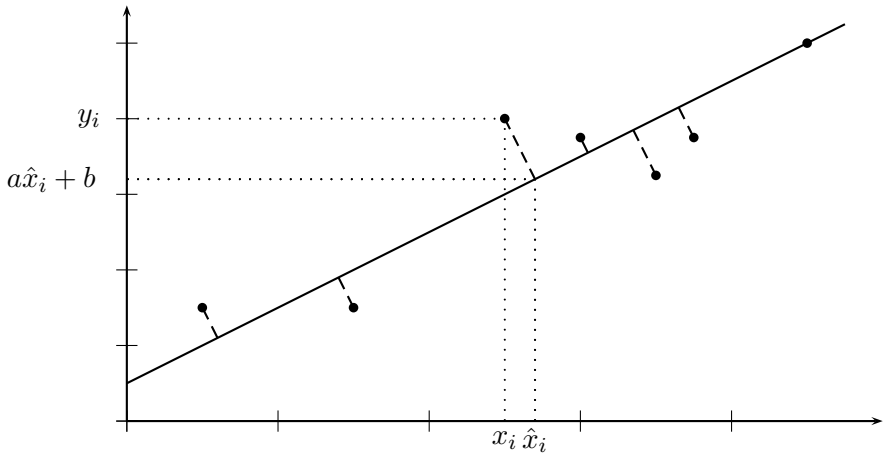
La régression orthogonale consiste à trouver l'équation de la droite pour laquelle la somme des carrés des distances entre les points du nuage et la droite est minimale. Si $\Delta_{a,b}$ désigne la droite d'équation $y = ax + b$, on cherche dont a et b tels que la quantité :

$$R(a,b) = \sum_{\omega \in \Omega} d\left((X,Y)(\omega); \Delta_{a,b}\right)^2,$$

C'est-à-dire :

$$F(a,b) = \sum_{\omega\in\Omega} \left(X(\omega)^2 + (Y(\omega)-b)^2 - \frac{1}{1+a^2}(X(\omega)+aY(\omega)-ab)^2 \right).$$

Graphiquement, on cherche à minimiser la somme des carrés des distances entre y_i et $\Delta_{a,b}$.



Remarque : Ce mode de régression ne privilégie pas X ou Y et donne une droite identique selon qu'on cherche à expliquer Y en fonction de X ou X en fonction de Y : En effectuant la régression orthogonale $X = \tilde{c}Y + \tilde{d}$ on aura exactement $\tilde{a} = \frac{1}{\tilde{c}}$ et $\tilde{b} = -\frac{\tilde{d}}{\tilde{c}}$. On l'utilise lorsque les erreurs sur X et Y sont du me ordre de grandeur.

La fonction R attend son minimum au point $(\tilde{a},\tilde{b})$, définis par :

$$\tilde{a} = C + \sqrt{C^2+1} \quad \text{et} \quad \tilde{b} = \overline{Y} - \tilde{a}\overline{X},$$

où $C = \dfrac{\sigma(Y)^2 - \sigma(X)^2}{2\mathbb{C}\mathrm{ov}(X,Y)}$.

Pour démontrer les valeurs de $\tilde{a}$ et $\tilde{b}$, on considère la fonction $(a,b) \mapsto R(a,b)$, dont l'unique point critique (point où les dérivées partielles s'annulent) est $(\tilde{a},\tilde{b})$ et on vérifie que c'est effectivement un minimum (matrice Hessienne à valeurs propres positives).

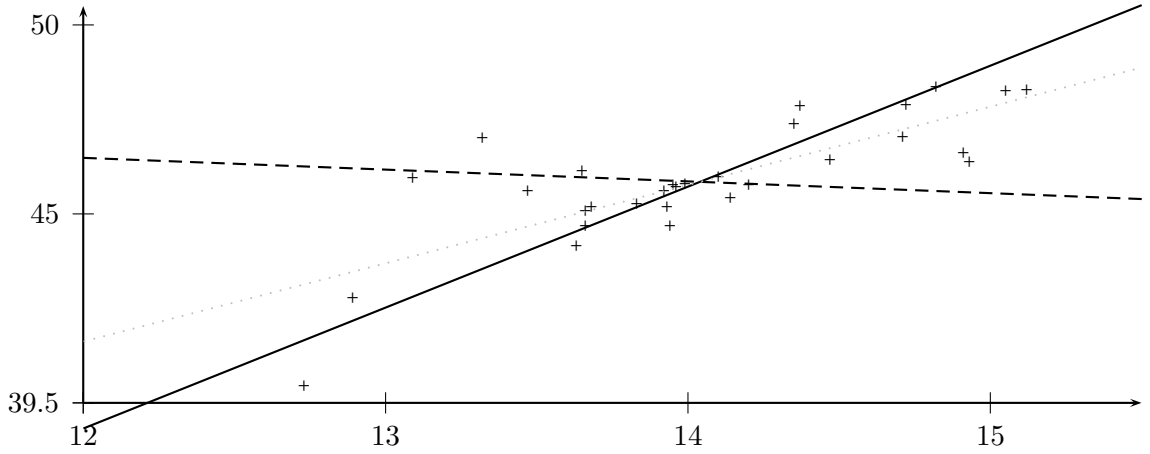
On en déduit donc l'équation de la **droite de régression orthogonale** ou **axe principal du nuage de points** :

$$y = \overline{Y} + \left(C + \sqrt{C^2+1}\right)(x - \overline{X}), \text{ avec } C = \frac{\sigma(Y)^2 - \sigma(X)^2}{\mathbb{C}\mathrm{ov}(X,Y)}.$$

La droite orthognonale à la droite principale, passant par son point $(\overline{X},\overline{Y})$ est appellée **axe secondaire du nuage de points**. SOn equation est donnée par :

$$y = \overline{Y} + \left(C - \sqrt{C^2+1}\right)(x - \overline{X}), \text{ avec } C = \frac{\sigma(Y)^2 - \sigma(X)^2}{\mathbb{C}\mathrm{ov}(X,Y)}.$$

Exemple 2.15. Pour la série **F** : les calculs donnent $\tilde{a} = 3.20$ et $\tilde{b} = 0.43$. Sur le dessin qui suit, l'axe principal est en noir, l'axe secondaire est en pointillé (orthogonal à l'axe principal même si la différence entre les echelles d'axes ne le montrer pas visuellement). La droite grise est la droite de régression des moindres carrés.



2.4.4 Application à la détermination de tendances dans les séries temporelles

On peut, éventuellement graphiquement, déceller des tendances plutot logarithmique ou exponentielles des courbes, bref, un type simple de courbe non linéaire. Dans ce cas-là, on serait tentés d'exprimer Y par $Y = b\ln(aX)$ ou $Y = be^{aX}$. Dans ces cas-là, on se sert de la régression linéaire, mais un peu adaptée :

• **Pour la régression logarithmique :**

Dire que $Y = b\ln(aX) = b\ln X + b\ln a$, c'est exactement dire que Y est linéaire en $\ln X$. On pose alors $Z = \ln X$ et on étudie la corrélation linéaire entre Z et Y, afin d'obtenir une régression du type $Y = \hat{A}Z + \hat{B}$. On en déduit les estimations $\hat{a}$ et $\hat{b}$ de a et b par identification $\hat{b} = \hat{A}$ et $\hat{B} = \hat{A}\ln\hat{a}$.

• **Pour la régression exponentielle :**

Dire que $Y = be^{aX}$, c'est exactement dire que $\ln Y$ est linéaire en X. On pose alors $Z = \ln Y$ et on étudie la corrélation linéaire entre X et Z, afin d'obtenir une régression du type $Z = \hat{A}X + \hat{B}$. On en déduit les estimations $\hat{a}$ et $\hat{b}$ de a et b par identification $\hat{a} = \hat{A}$ et $\hat{b} = e^{\hat{B}}$.

Une **série temporelle** ou **chronologique** $(Y_t)_{t=1,2,\ldots,n}$ est une variable réelle dont les unités statistiques sont les instants. Ces instants peuvent être considérés comme les valeurs d'un caractère T, en posant $T_t = t$ pour tout $t = 1, 2, \ldots, n$.

Pour ce paragraphe, on utilise la série temporelle suivante :

Série G

Nombres d'adhérent à un club de sport selon le mois (unité de temps) :

Mois	J	F	M	A	M	J	J	A	S	O	N	D
Effectifs	31	37	39	41	46	48	52	53	57	59	62	63

Lorsque les différences $Y_{t+1} - Y_t$ ou plutôt $\frac{Y_{t+1}-Y_t}{Y_t}$ sont relativement constantes, on est amenés à considérer un modèle linéaire de la forme $Y = at + b$. Les coefficients de la droite de régression de Y par rapport au temps T sont alors :

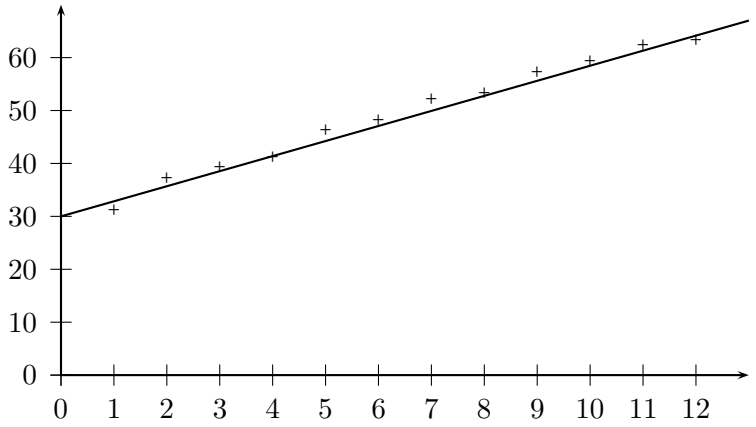
$$\hat{a} = \frac{\mathbb{C}\text{ov}(T, Y)}{\mathbb{V}\text{ar}(T)} \quad \text{et} \quad \hat{b} = \overline{Y} - \hat{a}\overline{T},$$

avec $\overline{T} = \sum_{t=1}^{n} t = \frac{n+1}{2}$ et $\mathbb{V}\text{ar}(T) = \sum_{t=1}^{n} t^2 - \left(\sum_{t=1}^{n} t\right)^2 = \frac{(n^2-1)}{12}$.

Exemple 2.16. Pour la série **G**, de moyenne $\overline{Y} = 49$, le calcul de la covariance avec le temps donne $\mathbb{C}\text{ov}(T, Y) = 34,25$. On en déduit :

$$\hat{a} = \frac{35,25}{11,9} = 2,87 \quad \text{et} \quad \hat{b} = 49 - 2,87 \times 6,5 = 30,31.$$

Ce qui donne graphiquement :



Lorsque Y est à valeurs strictement positives et lorsque les rapports $\frac{Y_{t+1}}{Y_t}$ sont relativement constants : $\frac{Y_{t+1}}{Y_t} \approx a$, on est aussi amené à considérer un modèle de la forme $Y_t = ba^t$, avec $b = Y_0$, appelé **modèle exponentiel**. On effectue alors le changement de variable $Z_y = \ln Y_t$ et on pose $A = \ln a$ et $B = \ln b$, puis on cherche la regression linéaire de $Z_t = At + B$.

Exemple 2.17. Pour la série **G**. Le passage au logarithmes donne la série suivante :

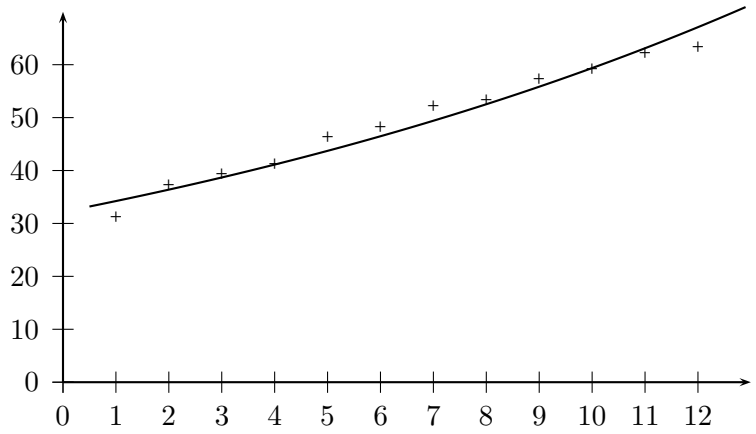
Mois	J	F	M	A	M	J	J	A	S	O	N	D
$Z = \ln Y$	3,43	3,61	3,66	3,71	3,83	3,87	3,95	3,97	4,04	4,08	4,13	4,14

Le caractère Z a pour moyenne $\overline{Z} = 3,87$. On a $\mathbb{C}\text{ov}(T, Z) = 0,73$ où T représente la variable aléatoire de temps.

On en déduit : $\hat{A} = \dfrac{0,73}{11,9} = 0,061$ et $\hat{B} = 3,87 - 0,061 \times 6,5 = 3,47$.

D'où $\hat{a} = e^{\hat{A}} = 1,063$ et $\hat{b} = e^{\hat{B}} = 32,20$.

Ce qui donne graphiquement :



On peut aussi être amené à considérer un modèle de la forme $Y_t = a \ln t + b$, appelé **modèle logarithmique**. On effectue alors le changement de variable $X = \ln T$, puis on cherche la regression linéaire de $Y_t = aX + b$.

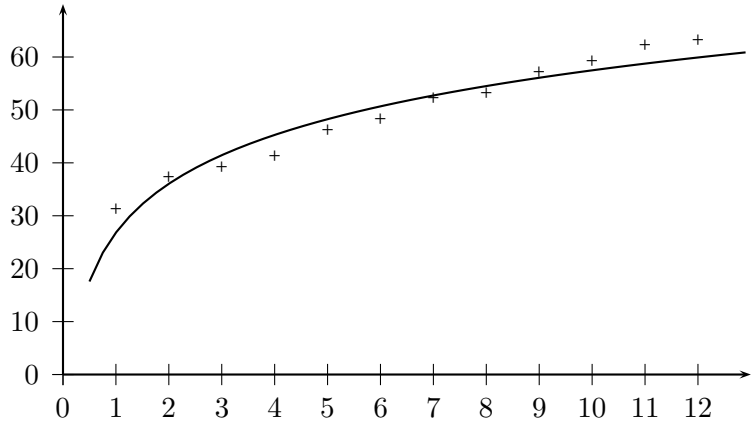
Exemple 2.18. Pour la série **G**. Le passage au logarithmes donne la série suivante :

Mois	J	F	M	A	M	J	J	A	S	O	N	D
$\ln T$	0	0,69	1,10	1,39	1,61	1,79	1,95	2,08	2,20	2,30	2,40	2,48
Effectif	31	37	39	41	46	48	52	53	57	59	62	63

Le caractère $X = \ln T$ a pour moyenne $\overline{X} = 1,66$ et de variance $\mathbb{V}\mathrm{ar}(X) = 0,52$, le calcul de la covariance avec le temps donne $\mathbb{C}\mathrm{ov}(X, Y) = 6,97$.

On en déduit $\hat{a} = \dfrac{6,97}{0,52} = 13,32$ et $\hat{b} = 49 - 13,32 \times 1,66 = 26,81$.

Ce qui donne graphiquement :



Chapitre 3

Introduction à la statistique inférentielle

3.1 Principes de la statistique inférentielle

La statitistique inférentielle regroupe les méthodes qui permettent de tirer des conclusion sur une population entière à partir des données recueillies sur un **échantillon** (sous-ensemble de la population).

Cela permet de repondre à plusieurs types de problèmes dont voici des exemples :

Problème A : ESTIMER DES PARAMETRES

Un petit exemple : Afin de mieux gérer son budget pub, le responsable de la diffusion d'un produit fait un sondage pour connaitre la dépense moyenne pour ce type d'articles par différentes catégories socio-professionnelles. Il fera ainsi une estimation de la dépense moyenne par catégories socio-professionnelles. Il peut vouloir également savoir la précision de ces estimations pour déterminer s'il peut faire confiance au resultats obtenus par le sondage.

Le premier problème qui se pose est donc de faire des **estimations des parametres souhaités**, ponctuelles ("la dépense moyenne est m") ou par intervalle de confiance ("il y a de très forte chances pour que la dépense moyenne soit comprise entre a et b").

Problème B : CONFORMITÉ À UNE HYPOTHÈSE

Un petit exemple : Pour un controle qualité, on veut, lors de la reception d'un echantillon de pièces mécaniques, comparer le taux de pièces non conformes observé par rapport à une norme fixée, de manière à refuser le lot de pièce si son taux de rebus dépasse la norme.

Dans la plupart des situations réelles, la valeur du paramêtre (pourcentage de pieces non conformes) est inconnue mais on peut avoir une idée de sa valeur et qu'on puisse formuler une **hypothèse** sur celui-ci (on suppose que l'usine qui a envoyé le lot est sérieuse et que son pourcentage de pieces non conformes est inférieur à la norme n). Les observations faites sur l'échantillon vont alors permettre de confirmer ou d'infirmer cette hypothèse, en se basant sur des critères obtenus par des **test de conformité**.

Problème C : COMPARAISON, INDÉPENDANCE, HOMOGÉNÉÏTÉ

Les décideurs sont souvent intéressé à determiner si 2 ou pluisuers populations données sont semblables ou nettement différentes par rapport à une caractéristique particulière, comme par exemple un médecin, qui souhaiterai savoir si un médicament est efficace (une population de test qui reçoit le traitement, une population qui reçoit un placebo), ou un acheteur qui s'intéresse à la durée de vie de machines semblables provenant de 2 fournisseurs différents. Pour cela, on utilise des **tests de comparaison** qui fournissent la réponse à ce type de question.

Dans le même esprit que les test de comparaisons, on peut vouloir savoir si deux caractères qualitatifs ou quantitatifs discrets sont indépendants ou non, c'est l'objets des **tests d'indépendance**. Par exemple, un médecin testant un médicament veut évidemment savoir si la guérison des patients depend ou non de la prise du médicament. Une autre façon de tourner le problème est de se demander si tous les groupes de tests ont les mêmes taux de guérison, i.e. si toutes les populations sont homogènes pour le caractère de guérison (ou pas...). On fait alors appel à des **tests d'homogénéïté**.

Problème D : AJUSTEMENT

Face à la répartition d'un paramètre sur un échantillon, une fois les principales caractéristiques estimées (moyenne, variance, etc...) on peut éventuellement vouloir savoir si un caractère d'une population suit une loi de probabilité particulière connue, comme par exemple la loi normale...

Des **tests d'ajustement analytique** ou **tests d'adéquation à une loi théorique** permettent de vérifier la qualité de l'ajustement du caractère étudié à une loi normale, binomiale, de Poisson ou encore à une loi uniforme. Il permette de savoir s'il est plausible que le caractère étudié suive effectivement la loi spécifiée avant le test.

3.2 Échantillonnage

3.2.1 Méthode d'échantillonage aléatoire simple

On parlera ici de l'**échantillonage aléatoire simple**. Ce type d'échantillonnage consiste à extraite un echantillon de taille n par **tirages aléatoires équiprobables et indépendants avec remise** parmi une population de N individus.

Le fait qu'on considère des tirages avec remise peut être un peu déroutant car cela va à l'encontre de l'idée du sondage classique (on ne sonde pas 2 fois la même personne au telephone, en général, on suit une liste) mais s'adapte bien à d'autre forme de prise d'échantillon (par exemple : on pêche un poisson, on relève sa taille et on le relache, puis on recommence...).

De plus lorsque l'échantillon est de taille faible devant la taille de la population ($\frac{n}{N} < 0,1$), on peut assimiler un tirage sans remise avec un tirage avec remise et cela simplifie beaucoup les calculs. Le modèle est donc le suivant :

Définition 3.1. échantillonage aléatoire simple

- On note X le caractère (variable aléatoire) que l'on veut étudier sur l'ensemble de la population.
- On note X_k (pour $k = 1,\ldots, n$) le résultat aléatoire du k-ième tirage. Ce sont des variables aléatoires indépendantes et identiquement distribuées de même loi que X.

Le n-uplet $(X_1, X_2, \ldots, X_n)$ est appellé **n-échantillon** ou **échantillon de taille** n de X.
- La réalisation unique $(x_1, x_2, \ldots, x_n)$ de l'échantillon $(X_1, X_2, \ldots, X_n)$ est l'ensemble des **valeurs observées**.
- Une **statistique** ou **variable d'échantillonnage** est une variable aléatoire $Y = f(X_1, X_2, \ldots, X_n)$ (où f est une fonction mesurable).
  Après réalisation la v.a. Y prend la valeur $f(x_1, x_2, \ldots, x_n)$.

Remarque 3.2. Il existe d'autre types plus élaborés d'échantillonnage adaptés à certaines situations :
- Echantillonnage sur la base de méthodes empiriques : **Méthode des quotats** qui tient compte de la répartition de la population selon certains critères (age, catégorie socio-professionnelle, etc...).
- **Echantillonnage stratifié** : On subsdivise la population en sous-ensembles (strates) relativement homogènes (ex : hommes/femmes, chaises/tables/armoires...), puis on extrait de chaque strate un échantillon aléatoire simple qui sont ensuite regorupés.
- **échantillonnage par groupe** : On choisit n échantillon aléatoire d'unités qui sont elles-mêmes des sous-ensembles de la population. Exemple : On divise une ville en quartiers, puis parmi ces quartiers on en choisit (aléatoirement) un certain nombre qui forment l'échantillon et on fait l'enquête sur tous les habitants des quartiers sélectionnés.

3.2.2 Variables d'échatillonnage

On suppose que X suit une loi d'esperance $\mathbb{E}(X) = \mu$ et de variance $\mathbb{V}\mathrm{ar}(X) = \sigma^2$ inconnues.

Définition 3.3. On appelle **moyenne empirique** ou **moyenne d'échantillonnage** de l'échantillon $(X_1, X_2, \ldots, X_n)$ la variable d'échantillonnage :

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k.$$

Sa réalisation, appellée **moyenne observée** sera noté $\overline{x}_n = \frac{1}{n}\sum_{k=1}^{n} x_k$; C'est la moyenne des valeurs de la réalisation de l'échantillon.

Dans le cas où X suit une loi de Bernoulli de parametre inconnu p, $\overline{X}_n$ est la fréquence de succès lors de n expériences. On appelle alors souvent cette variable **fréquence empirique** ou **fréquence d'échantillonnage**.

Dans le cas où X est une v.a. finie à valeurs dans l'ensemble $\{v_i\,;\, i = 1, 2, \ldots, h\}$, on peut définir la **distribution de fréquence d'échantillonnage** comme un vecteur $(F_{1,n}, F_{2,n}, \ldots, F_{h,n})$ dont les réalisations $(f_{1,n}, f_{2,n}, \ldots, f_{h,n})$ sont les fréquences des valeurs $\{v_i\,;\, i = 1, 2, \ldots, h\}$: f_i est la fréquence d'apparition de v_i dans l'échantillon $(x_1, x_2, \ldots, x_n)$.

Propriétés 3.4. $\mathbb{E}\left(\overline{X}_n\right) = \mu \qquad et \qquad \mathbb{V}\mathrm{ar}\left(\overline{X}_n\right) = \dfrac{1}{n}\sigma^2.$

★ **Exercice 3.5.** Montrer la propriété précédente.

Définition 3.6.

On appelle **variance empirique** ou **variance d'échantillonnage** de $(X_1, X_2, \ldots, X_n)$ la variable d'échantillonnage :

$$S_n^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X})^2.$$

Sa réalisation, appellée **variance observée** sera noté $s_n^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})^2$; c'est la variance des valeurs de la réalisation de l'échantillon. la racine positive de la variance observée est l'**écart-type observé**.

Propriété 3.7. $\mathbb{E}\left(V_n^2\right) = \dfrac{n-1}{n}\sigma^2$.

Démonstration :

$$\mathbb{E}\left(V_n^2\right) = \mathbb{E}\left(\frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X})^2\right) = \mathbb{E}\left(\frac{1}{n}\left(\sum_{k=1}^{n}X_k\right) - \overline{X}^2\right) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left(X_k^2\right) - \mathbb{E}\left(\overline{X}^2\right)$$

Or pour toute v.a. Y, $\mathbb{E}(Y^2) = \mathbb{V}\mathrm{ar}(Y) + \mathbb{E}(Y)^2$, d'où :

$$\mathbb{E}\left(V_n^2\right) = \frac{1}{n}\left(\sum_{k=1}^{n}\mathbb{V}\mathrm{ar}(X_k) + \mathbb{E}(X_k)^2\right) - \left(\mathbb{V}\mathrm{ar}(\overline{X}) + \mathbb{E}(\overline{X})^2\right).$$

De plus $\mathbb{E}(X_k) = \mathbb{E}(X)$ et $\mathbb{V}\mathrm{ar}(X_k) = \mathbb{V}\mathrm{ar}(X)$, d'où :

$$\mathbb{E}\left(V_n^2\right) = \mathbb{V}\mathrm{ar}(X) + \mathbb{E}(X)^2 - \left(\mathbb{V}\mathrm{ar}(\overline{X}) + \mathbb{E}(\overline{X})^2\right) = \sigma^2 - \mu^2 - \frac{1}{n}\sigma^2 - \mu^2 = \frac{n-1}{n}\sigma^2.$$

Définition 3.8.

On appelle **variance corrigée d'échantillonnage** de $(X_1, X_2, \ldots, X_n)$ la variable d'échantillonnage :

$$S_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2.$$

On montre facilement grâce au résultat précédent la propriété suivante :

Propriété 3.9. $\mathbb{E}\left(S_n^2\right) = \sigma^2$.

3.2.3 Approche des lois de probabilité des variables d'échantillonnage

Pour approcher les lois de probabilité des variables d'échantillonnage présentées ci-dessus, on fait appel au 2 théorèmes limites classiques suivants :

Théorème 3.10. Loi forte des grands nombres

Soit X une v.a. telle que $\mathbb{E}(X) = \mu$ et $\mathbb{V}\mathrm{ar}(X) = \sigma^2 \neq 0$. On considère une suite $(X_k)_{k\geq 1}$ de v.a. i.i.d de même loi que X.

La v.a. $\overline{X}_n$ converge en probabilité vers $\mathbb{E}(X) = \mu$. Ce qu'on peut reformuler ainsi :

La v.a. $\overline{X}_n$ est asymptotiquement constante égale à $\mathbb{E}(X) = \mu$.

En effet, pour tout $n \geq 2$, on a : $\mathbb{E}(\overline{X}_n) = \mu$ et $\mathbb{V}\text{ar}(\overline{X}_n) = \frac{\sigma^2}{n}$, donc $\lim_{n\to+\infty} \mathbb{V}\text{ar}(\overline{X}) = 0$. $\overline{X}_n$ est asymptotiquement constante et comme $\mathbb{E}(\overline{X}_n) = \mu$ quelque soit n, alors $\overline{X}_n$ est asymptotiquement constante égale à μ.

Théorème 3.11. Théorème central limite

Soit X une v.a. telle que $\mathbb{E}(X) = \mu$ et $\mathbb{V}\text{ar}(X) = \sigma^2 \neq 0$. On considère une suite $(X_k)_{k\geq 1}$ de v.a. i.i.d de même loi que X.

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1),$$

Résultat que l'on peut reformuler ainsi :

Lorsque n est grand ($n > 30$ en général), alors on a $\overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$

Remarque 3.12. La variable aléatoire $\overline{X}_n$ suit toujours une loi d'espérance $\mathbb{E}(\overline{X}_n) = \mu$ et de variance $\mathbb{V}\text{ar}(\overline{X}_n) = \frac{\sigma^2}{n}$. Le théorème central limite nous dit donc que lorsque n est grand, la loi de $\overline{X}_n$ peut être assimilée à une loi normale.

Exemple 3.13. On considère un lot de 500 bonbons artisanaux. Le poids d'un bonbon est une variable aléatoire X d'esperance $\mu = 5g$ et d"écart-type $\sigma = 0.5$. Quelle est la probabilité pour qu'un paquet de 50 bonbons issus de ce lot ait un poids supérieur à $260g$?

Le poids d'une boite de bonbons est donné par la v.a. $T = \sum_{k=1}^{50} X_k = 50\overline{X}_{50}$.

Comme l'échantillon est grand ($50 > 30$), on peut assimiler la loi de $\overline{X}_{50}$ à une loi $\mathcal{N}\left(5, \frac{0.5^2}{50}\right)$. Donc T suit approximativement une loi $\mathcal{N}\left(50 \times 5, 50^2 \times \frac{0.5^2}{50}\right)$, c'est-à-dire la loi $\mathcal{N}(250, 12.5)$.

Ainsi, la v.a. $\frac{T-250}{\sqrt{12.5}}$ suit une loi $\mathcal{N}(0,1)$. On va donc pouvoir se servir de la table de la loi normale $\mathcal{N}(0,1)$.

$$\mathbb{P}(T > 260) = \mathbb{P}\left(\frac{T-250}{\sqrt{12.5}} > \frac{260-250}{\sqrt{12.5}}\right) = \mathbb{P}\left(\frac{T-250}{\sqrt{12.5}} > 2.83\right) = 1 - \mathbb{P}\left(\frac{T-250}{\sqrt{12.5}} < 2.83\right).$$

En lisant la table de la loi $\mathcal{N}(0,1)$, on obtient :

$$\mathbb{P}(T > 260) = 1 - 0.9977 = 0.0023.$$

Il y a donc 2.3% de chances que le poids d'un paquet de bonbons dépasse 260g.

Remarque 3.14. Si X suit effectivement une loi normale $\mathcal{N}(\mu, \sigma^2)$, alors :

1. $\overline{X}_n$ suit une loi $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$
2. $\frac{n}{\sigma^2} S_n^2$ suit une loi χ^2_{n-1},
3. $\frac{\overline{X}_n - \mu}{\sqrt{S^2}/\sqrt{n-1}}$ suit une loi de Student $\mathcal{T}_{n-1}$.

3.3 Intervalles de fluctuation

Un intervalle de fluctuation (à ne pas confondre avec intervalle de confiance...) permet de détecter un écart important entre la réalisation d'une grandeur (espérance empirique, variance empirique, etc) et sa valeur théorique connue.

C'est un intervalle dans lequel la grandeur observée est censée se trouver avec une forte probabilité (souvent de l'ordre de 95 %). L'intervalle de flucturation n'est pas unique mais il est courant d'utiliser l'intervalle centré autour la grandeur connue.

Le fait d'obtenir une valeur en dehors de cet intervalle s'interprète alors en mettant en cause la représentativité de l'échantillon ou la valeur théorique. À l'inverse, le fait que la moyenne soit comprise dans l'intervalle n'est pas une garantie de la validité de l'échantillon ou du modèle.

Exemple 3.15. On préleve 25 pièces parmi la production journalière d'une usine. Une étude préalable a montré que la taille des pièces suit une loi normale de moyenne 10mm et d'ecart-type 2mm.

Déterminer un intervalle dans lequel ont a 85% d'avoir l'écart-type observé ?

On cherche α et β tels que :

$$\mathbb{P}(\alpha < S_n < \beta) = 0,85.$$

Pour cela, on va utiliser le fait que la v.a. $\frac{n}{\sigma^2}S_n^2$ ($n = 25$) suit une loi $\chi^2_{25-1} = \chi^2_{24}$ et la table de cette loi.

$$\mathbb{P}\left(a < \frac{n}{\sigma^2}S_n^2 < b\right) = \mathbb{P}\left(\frac{n}{\sigma^2}S^2 > a\right) - \mathbb{P}\left(\frac{n}{\sigma^2}S_n^2 > b\right) = 0,85 = 0,9 - 0,05.$$

La décomposition $0,85 = 0,9 - 0,05$ dépend des données disponibles dans la table... On aurait aussi pu choisir la décomposition $0,85 = 0,95 - 0,1$.

On cherche donc a et b tels que :

$$\mathbb{P}\left(\frac{n}{\sigma^2}S_n^2 > a\right) = 0,90 \qquad \text{et} \qquad \mathbb{P}\left(\frac{n}{\sigma^2}S_n^2 > b\right) = 0,05.$$

La lecture de la table donne $a = 15,659$ et $b = 36,415$. On en déduit donc

$$\mathbb{P}\left(15,659 < \frac{n}{\sigma^2}S_n^2 < 36,415\right) = 0,85$$

$$\mathbb{P}\left(2,5054 < S_n^2 < 5,8264\right) = 0,85$$

$$\mathbb{P}\left(1,58 < S_n < 2,41\right) = 0,85$$

Ainsi, l'écart-type observée sur un échantillon de pièces de la production a 85% de chance être compris entre $1,58$ et $2,41$.

Lorsque la loi du parametre étudié X n'est pas entièrement connue mais que l'on connait sa moyenne ou encore si elle est connue mais que les calculs sont très lourds, on peut établir des intervalles de fluctuation asymptotiques pour la moyenne :

Proposition et définition 3.16. Intervalle de fluctuation asymptotique pour la moyenne empirique

Soit X une v.a. telle que $\mathbb{E}(X) = \mu$ et $\mathbb{V}\text{ar}(X) = \sigma^2 \neq 0$. On considère une suite $(X_k)_{k\geq 1}$ de v.a. i.i.d de même loi que X.

Pour $\alpha \in]0,1[$ et Z une v.a. de loi normale $\mathcal{N}(0,1)$. On note u_α l'unique réel pour lequel $\mathbb{P}(|Z| < u_\alpha) = 1-\alpha$. On a :

$$\lim_{n\to\infty} \mathbb{P}\left(\overline{X}_n \in \left[\mu - u_\alpha \frac{\sigma}{\sqrt{n}}; \mu + u_\alpha \frac{\sigma}{\sqrt{n}}\right]\right) = 1-\alpha.$$

autrement dit : $\lim_{n\to\infty} \mathbb{P}\left(\,|\,\overline{X}_n - \mu\,|\, \leq u_\alpha \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$. L'intervalle $I_n = \left[\mu - u_\alpha \frac{\sigma}{\sqrt{n}}; \mu + u_\alpha \frac{\sigma}{\sqrt{n}}\right]$ est appellé **intervalle de fluctuation asymptotique au seuil** $1-\alpha$.

Remarques 3.17. Pour $\alpha = 0.05$, $u_\alpha \approx 1.96$. Pour $\alpha = 0.01$, $u_\alpha \approx 2.58$

Cette proposition est en fait une conséquence immédiate du théorème central limite (Démonstration en exercice).

On considère que pour $n > 30$, on peut identifier $\mathbb{P}(\overline{X}_n \in I_n)$ à $1-\alpha$.

Proposition 3.18. Intervalles de fluctuation asymptotiques au seuil de 95% et au seuil de 99% pour la moyenne empirique d'une variable de Bernoulli

Soit X une v.a. de loi $\mathcal{B}(p)$. On considère une suite $(X_k)_{k\geq 1}$ de v.a. i.i.d de même loi que X. On a :

$$\lim_{n\to\infty} \mathbb{P}\left(\overline{X}_n \in \left[\mu - 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}}; \mu + 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]\right) = 0.95.$$

L'intervalle $\left[\mu - 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}}; \mu + 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ *est appellé* **intervalle de fluctuation asymptotique au seuil de 95%**. *et*

$$\lim_{n\to\infty} \mathbb{P}\left(\overline{X}_n \in \left[\mu - 2.58\frac{\sqrt{p(1-p)}}{\sqrt{n}}; \mu + 2.58\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]\right) = 0.99.$$

L'intervalle $\left[\mu - 2.58\frac{\sqrt{p(1-p)}}{\sqrt{n}}; \mu + 2.58\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ *est appellé* **intervalle de fluctuation asymptotique au seuil de 99%**.

Remarque 3.19. On admet en général que

$$\mathbb{P}\left(\overline{X}_n \in \left[\mu - 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}}; \mu + 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]\right) = 1-\alpha$$

sous les conditions suivantes :

$$n > 30, \quad np \geq 5, \quad n(1-p) \geq 5.$$

Remarque 3.20. Intervalle de fluctuation simplifié au seuil de 95% pour la moyenne empirique d'une variable de Bernoulli

Soit X une v.a. de loi $\mathcal{B}(p)$. On considère une suite $(X_k)_{k\geq 1}$ de v.a. i.i.d de même loi que X. Pour n assez grand, on a :

$$\mathbb{P}\left(\overline{X}_n \in \left[\mu - \frac{1}{\sqrt{n}}; \mu + \frac{1}{\sqrt{n}}\right]\right) \leq 0.95.$$

cela vient simplement du fait que, quelque soit p, $1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \frac{1}{\sqrt{n}}$.

★ **Exercice 3.21.** Dans l'arrière-boutique d'un bar, il y 180 bouteilles de Dodo et 60 de Cot citron en vrac. Pendant une coupure de courant, le barman part à l'aveugle dans l'arrière-boutique et revient avec 60 bouteilles. On note la proportion p de bouteilles de bières dans le lot qu'il a ramené.

1. Déterminer un intervalle dans lequel p a 95% de chance de se trouver.
2. Parmi les 60 bouteilles ramenées, 40 sont des bouteilles de Cot. Cet échantillon est-il représentatif du stock ?

Remarque 3.22. Dans la littérature, les intervalles de fluctuation au seuil $1-\alpha$ sont aussi désignés par les expressions **intervalles de dispersion au seuil** $1-\alpha$ ou **intervalles de probabilité au seuil** $1-\alpha$.

Remarque 3.23. Intervalle de fluctuation et intervalle de confiance

Ile faut pas confondre Intervalle de fluctuation et intervalle de confiance. Pour illustrer cela, on reprend l'Exercice 3.21

Déterminer un intervalle dans lequel p a 95% de chance de se trouver revient à déterminer l'intervalle de fluctuation au seuil de 05% de la proportion de bouteilles de Dodo parmi les bouteilles ramenées.

Mais si on suppose maintenant que le barman ne connait pas l'état des stocks avant d'aller chercher les bouteilles dans le noir et qu'il doit évaluer la composition du stock uniquement à partir des 60 bouteilles ramenée (par Exemple : 25 Cots, et 35 Dodo), alors on va chercher à trouver un intervalle de confiance pour la proportion de Dodo dans le stock, qui lui permettra de dire : à la vue de mon échantillon, mon stock doit vraisemblablement contenir entre x% et y% de bouteilles de Dodo.

Chapitre 4

Estimation

4.1 Introduction

4.1.1 Principe et notations

La distribution exacte d'une variable X modélisant le caractère qui intéresse un statisticien est en général partiellement connue. On peut par exemple, à partir d'un échantillon et des différentes représentations issues des statistiques descriptives supposer d'un caractère suit une loi qui ressemble vaguement à une loi connue (loi normale, loi exponentielle, etc...). Reste que la loi de X dépend d'un (ou plusieurs) parametre(s) inconnus (esperance et variance pour une loi normale, espérance pour une exponentielle, etc...). On cherche donc a se faire une idée plus précise sur ce(s) paramètre(s) à partir des données observées sur l'échantillon : on cherche à estimer ce(s) parametre(s).

On peut tenter d'attribuer une valeur numérique unique au parametre : c'est une **estimation ponctuelle**. Pour ce faire, on utilise un **estimateur**, c'est-à-dire une v.a. dont la valeur observée donnera l'estimation du parametre.

Le problème est qu'il est fort peu probable que l'estimation poctuelle soit exacte... Ainsi, plutôt que d'estimer un parametre par une valeur unique, il est fréquent qu'on fasse une **estimation par intervalle de confiance**, c'est à dire que l'on donne un intervalle dans lequel notre paramêtre a de "fortes chances de se trouver". C'est ce qu'on appelle l'**intervalle d'estimation** ou **intervalle de confiance**. Il va sans dire qu'évidemment la notion de "fortes chances de se trouver dans un intervalle donné" doit être précisée dès le départ car il précise le **degré de confiance** que l'on peut accorder aux résultats obtenus.

Dans la suite, on s'intéresse à l'estimation des principales caractéristiques d'un caractère : l'espérance, la variance et les fréquences (pour des v.a. à valeurs dans un ensemble fini).

Notations

- Les parametres à estimer seront notés par des lettres greques minuscules :
  μ : moyenne/espérance
  σ^2 : variance
  σ : écart-type
  π : proportion

- Les réalisations d'échantillon seront notées par des lettres latines minuscules :
  $(x_1, x_2, \ldots, x_n)$: valeurs observées de l'échantillon
  $\overline{x}$: moyenne de l'échantillon
  σ^2 : variance observée de l'échantillon
  σ : écart-type observé de l'échantillon
  f : fréquence observée de l'échantillon
- Les estimateurs (v.a.) seront notés en majuscules :
  $\overline{X}_n$
  S_n^2
  F_n
  ou par des lettres greques minuscules "chapeautées" :
  $\widehat{\mu}_n$
  $\widehat{\sigma^2}_n$
  $\widehat{\sigma}_n$
  $\widehat{\pi}_n$

4.1.2 Notion d'estimateur et qualités d'un estimateur

Définition 4.1. On appelle **estimateur d'un paramêtre** θ d'une population une v.a. $T_n = f(X_1, X_2, \ldots, X_n)$ qui a de "bonnes propriétés" par rapport à θ (voir définition suivantes).
Sa réalisation sera notée $t_n = f(x_1, x_2, \ldots, x_n)$ est appelée **estimation**.

En pratique, on suppose en général que T_n tend vers $g(\theta)$ avec g simple (encore faut-il préciser à quelle type de limite fait référence l'expression " T_n tend vers $g(\theta)$").

Pour un même parametre, il peut exister plusieurs estimateurs intéressants (par exemple si X suit une loi de Poisson de parametre λ, alors $\overline{X}$ et S^2 sont deux estimateurs de λ ; Pour une loi symétrique $\overline{X}$ et Me la médiane de l'échantillon sont deux estimateurs de la moyenne), le choix de l'un ou l'autre va donc dépendre des qualités qu'on demande à l'estimateur...

L'une des propriétés presque évidente que l'on demande à un estimateur, c'est d'est être convergent, c'est-à-dire que lorsque l'échantillon de vient très grand, sa valeur tende vers la vrai valeur du parametre. Il y a évidemment plusieurs moyens de quantifier le "rapprochement" d'un estimateur vers le parametre...

On présente ici les 2 principaux types de convergence qui englobent les 2 exemples suivants :

1. convergence de la moyenne empirique $\overline{X}_n$ d'une suite de v.a. i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$ vers m.
2. convergence de la fréquence empirique $F_n = \frac{1}{n}\text{Card}\{k \in [1, n]\,;\, X_i = 1\}$ d'une suite de de v.a. i.i.d. de loi $\mathcal{B}(p)$ vers p.

Définition 4.2. Erreur d'estimation et biais
Soit un estimateur $T_n = f(X_1, X_2, \ldots, X_n)$ du parametre θ.
On appelle **erreur d'estimation** la quantité $T_n - \theta$.
L'espérance de l'erreur $B(T_n) = \mathbb{E}(T_n) - \theta$ est appellé **biais** de l'estimateur T_n.

- Si $\mathbb{E}(T_n) = \theta$, alors $B(T_n) = 0$ et on dit que T_n est un **estimateur sans biais** de θ, L'esperance de l'erreur d'estimation est alors nulle.
- Si $\mathbb{E}(T_n) \neq \theta$, alors $B(T_n) \neq 0$ et on dit que T_n est un **estimateur biaisé** de θ,
- Si $\lim_{n\to+\infty} \mathbb{E}(T_n) = \theta$, alors $\lim_{n\to+\infty} B(T_n) = 0$ et on dit que T_n est un **estimateur asymptotiquement sans biais**.

Remarque 4.3. L'erreur d'estimation $T_n - \theta$ peut être décomposée en $(T_n - \mathbb{E}(T_n)) + (\mathbb{E}(T_n) - \theta)$. La quantité $(T_n - \mathbb{E}(T_n))$ traduit la fluctuation de T_n autour de son espérance et le biais $(\mathbb{E}(T_n) - \theta)$ représente l'erreur systématique faite en estimation θ avec T_n.

Exemple 4.4. La moyenne empirique est un estimateur san biais du parametre λ d'une loi de Poisson. La variance empirique est un estimateur biaisé de même parametre . En effet, si X suit une loi de Poisson $\mathcal{P}(\lambda)$, alors :

$$\mathbb{E}(\overline{X}_n) = \lambda,\ \mathbb{E}(V_n) = \frac{n-1}{n}\lambda, \quad \text{et } \mathbb{E}(X) = \mathbb{V}\text{ar}(X) = \lambda.$$

Définition 4.5. Soit un estimateur $T_n = f(X_1, X_2, \ldots, X_n)$ du parametre θ.

On appelle **erreur quadratique moyenne** la quantité

$$EQM(T_n) = \mathbb{E}((T_n - \theta)^2) = \mathbb{V}\text{ar}(T_n) - (\mathbb{E}(T_n) - \theta)^2 = \mathbb{V}\text{ar}(T_n) - B(T_n).$$

Conséquence : De 2 estimateurs sans biais, on privilégie donc celui de variance minimale... C'est celui pour lequel $EQM(T_n)$ est le plus petit !

Définition 4.6. Un estimateur $T_n = f(X_1, X_2, \ldots, X_n)$ est **convergent en moyenne quadratique** vers θ si :

$$\lim_{n\to+\infty} \mathbb{E}((T_n - \theta)^2) = 0.$$

Remarque 4.7. Dire que T_n converge en moyenne quadratique vers θ revient à dire que la distance moyenne entre les réalisations de T_n et θ tends vers 0.

Proposition 4.8. *Si* $\lim_{n\to+\infty} \mathbb{E}(T_n) = \theta$ *et* $\lim_{n\to+\infty} \mathbb{V}\text{ar}(T_n) = 0$ *alors* T_n *converge en moyenne quadratique vers* θ*.*

Cette proposition découle du fait que $\mathbb{E}((T_n - \theta)^2) = \mathbb{V}\text{ar}(T_n) + (\mathbb{E}(T_n) - \theta)^2$.

Définition 4.9. Efficacité

Soient T_n et T'_n deux estimateurs d'un parametre θ. T est **plus efficace que** T' si

$$EQM(T) \leq EQM(T').$$

Si T_n et T'_n sont sans biais, alors T est plus efficace que T' si $\mathbb{V}\text{ar}(T) \leq \mathbb{V}\text{ar}(T')$. L'estimateur sans biais de variance minimale est appelé **estimateur efficace**. La variance minimale est appellée borne de Cramer-Rao.

Remarque 4.10. Dans certains cas et malgré les apparences, il se peut qu'un estimateur avec biais soit plus efficace qu'un estimateur sans biais.

★ **Exercice 4.11.** Qui de la moyenne empirique ou de la variance empirique est le meilleur estimateur du parametre λ d'une loi de Poisson ?

Les estimateurs étant des variables aléatoires, il y a évidemment d'autres notions de convergence :

Définition 4.12. Un estimateur $T_n = f(X_1, X_2, \ldots, X_n)$ est **convergent en probabilité** vers la valeur θ si :

$$\text{Pour tout } \epsilon > 0 \text{ fixé, } \lim_{n \to +\infty} \mathbb{P}\left(|T_n - \theta| > \epsilon\right) = 0,$$

i.e. T_n converge en probabilité vers la v.a. constante égale à θ. On parle dans ce cas d'estimateur **faiblement consistant**.

On peut interpréter ce type de convergence comme suit : Si on considère l'intervalle $I_\epsilon = [\theta - \epsilon; \theta + \epsilon]$, alors en effectuant un grand nombre de réaliations de T_n, la proportion de réalisations de T_n qui "tombe" hors de l'intervalle I_ϵ tend vers 0.

Propriété 4.13. *Un estimateur convergent en moyenne quadratique vers θ est également convergent en probabilité.*

Définition 4.14. Un estimateur $T_n = f(X_1, X_2, \ldots, X_n)$ est **fortement consistant** si T_n converge presque sûrement vers la v.a. constante égale à θ, i.e. :

$$\mathbb{P}\left(\lim_{n \to +\infty} T_n = \theta\right) = 1.$$

C'est la convergence $\mathbb{P}$-presque sûre de T_n vers θ.

4.2 Estimation ponctuelle

4.2.1 Estimation de parametres usuels

4.2.1.1 Moyenne et proportion

Proposition 4.15. *Soit X une variable aléatoire d'esperance μ et de variance σ^2. On dispose d'un échantillon $(X_1, X_2, \ldots, X_n)$ de X.*

La moyenne empirique $\overline{X}_n = \sum_{k=1}^{n} X_k$ est un estimateur sans biais qui converge en moyenne quadratique vers $\mathbb{E}(X) = \mu$.

La moyenne empirique est un estimateur sans biais car $\mathbb{E}(\overline{X}_n) = \mu$. D'autre part $\mathbb{V}\text{ar}\left(\overline{X}_n\right) = \frac{\sigma^2}{n}$, donc $\lim_{n \to +\infty} \mathbb{V}\text{ar}\left(\overline{X}_n\right) = 0$.

Remarque 4.16. On peut montrer en outre que $\overline{X}_n$ est un estimateur efficace de θ.

Si on étudie une population d'individus possédant (ou pas) une certaine caractéristique A et que l'on a besoin d'estimer à partir d'un échantillon de taille n la proportion d'individus de cette population ayant effectivement la caractéristique A.

Cela revient à s'interesser à l'estimation de la moyenne d'une variable aléatoire Y qui suit une loi de Bernoulli $\mathcal{B}(p)$ ($Y(\omega) = 1$ si et seulement si l'individu ω a la caractéristique A et $Y(\omega) = 0$ sinon) et dont on cherche à estimer p à partir d'un échantillon de taille n.

La moyenne $\overline{Y}$ représente alors la **fréquence empirique du caractère** et est usuellement notée F_n :

$$F_n = \frac{\text{Card}\left(\{k \,:\, X_k \text{ a le caractère } A\}\right)}{n}$$

Comme cas particulier de la proposition précédente, on a donc :

La fréquence empirique F d'un caractère A, est un estimateur sans biais qui converge en moyenne quadratique vers la proportion exacte p d'individus de la population ayant le caractère A.

★ **Exercice 4.17.** Parmi une population d'étudiants de L1, on a prélevé indépendamment 2 échantillons d'étudiants :

1. Dans l'echantillon n°1, composé de 120 étudiants, on en note 48 qui vont s'orienter en L2 maths.
2. Dans l'echantillon n°2, composé de 150 étudiants, on en note 66 qui vont s'orienter en L2 maths.

On note p la proportion d'étudiants d'étudiants voulant intégrer le L2 maths. En donner 3 estimations ponctuelles.

4.2.1.2 Variance et écart-type

Proposition 4.18. *Soit X une variable aléatoire d'espérance μ et de variance σ^2. On dispose d'un échantillon $(X_1, X_2, \ldots, X_n)$ de X.*

1. *Si $\mathbb{E}(X) = \mu$ est connue. La variance corrigée $\overline{T}_n^2 = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n}(X_k - \mu)^2$ est un estimateur sans biais de σ^2.*

   L'écart-type σ sera estimé ponctuellement par l'estimateur T_n, estimateur biaisé en général ($\mathbb{E}(T_n) \neq \sqrt{\mathbb{E}(T_n^2)}$ en général).
2. *Si $\mathbb{E}(X) = \mu$ est inconnue :*

   La variance empirique $V_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n}(X_k - \overline{X})^2$ est un estimateur biaisé, asymptotiquement sans biais, qui converge en moyenne quadratique vers σ^2.

   La variance corrigée $\overline{S}_n^2 = \dfrac{1}{n-1}\displaystyle\sum_{k=1}^{n}(X_k - \overline{X})^2$ est un estimateur sans biais qui converge en moyenne quadratique vers σ^2.

   L'écart-type σ sera estimé ponctuellement par l'estimateur S_n, estimateur biaisé en général ($\mathbb{E}(S_n) \neq \sqrt{\mathbb{E}(S_n^2)}$ en général).

Démonstration.

1. On a $\mathbb{E}(T_n^2) = \mathbb{E}\left(\frac{1}{n}\sum_{k=1}^{n}(X_k-\mu)^2\right) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left(X_k^2\right) - \frac{2\mu}{n}\sum_{k=1}^{n}\mathbb{E}\left(X_k\right) + \mu^2 = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left(X_k^2\right) - \mu^2$

Or pour tout X_k, $\mathbb{E}(X_k^2) = \mathbb{V}\mathrm{ar}(X_k) + \mathbb{E}(X_k)^2 = \sigma^2 + \mu^2$. On en déduit donc

$$\mathbb{E}(T_n^2) = \sigma^2 + \mu^2 - \mu^2 = \sigma^2,$$

c'est-à-dire que T_n^2 est un estimateur sans biais.

2. On a $\mathbb{E}(V_n) = \frac{n-1}{n}\sigma^2$ donc V_n est un estimateur biaisé de σ^2 et $B(V_n) = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{1}{n}\sigma^2$, donc V_n est asymptotiquement sans biais.

$\mathbb{E}(S_n^2) = \frac{n}{n-1}\mathbb{E}(V_n) = \sigma^2$ donc S_n^2 est un estimateur sans biais de σ^2. □

Remarque 4.19. Lorsque n est grand, $\mathbb{E}(S_n^2) \approx \mathbb{E}(V_n)$ on privilégie V_n mais lorsque n est petit, il vaut mieux choisir l'estimateur sans biais S_n^2.

Remarque 4.20. Si X admet un moment d'ordre 4 et μ connue, l'estimateur T_n^2 converge en moyenne quadratique vers θ :

$\mathbb{V}\mathrm{ar}(T_n^2) = \mathbb{V}\mathrm{ar}\left(\frac{1}{n}\sum_{k=1}^{n}(X_k-\mu)^2\right) = \frac{1}{n^2}\sum_{k=1}^{n}\mathbb{V}\mathrm{ar}\left((X_k-\mu)^2\right)$ et $\mathbb{V}\mathrm{ar}\left((X_k-\mu)^2\right) = \mathbb{E}\left((X_k-\mu)^4\right) - \mathbb{E}\left((X_k-\mu)^2\right)^2 = \mathbb{E}\left((X-\mu)^4\right) - \mathbb{E}\left((X-\mu)^2\right)^2$.

Ainsi $\mathbb{V}\mathrm{ar}(T_n^2) = \frac{1}{n}\left(\mathbb{E}\left((X-\mu)^4\right) - \mathbb{E}\left((X-\mu)^2\right)^2\right)$ qui tend vers 0 lorsque n tend vers $+\infty$: T_n^2 converge donc en moyenne quadratique vers θ.

On peut également montrer que c'est un estimateur efficace.

★ **Exercice 4.21.** Dans une grosse entreprise, on note X le nombre d'appel de clients reçus au standard. On suppose que X suit une loi normale $\mathcal{N}(\mu, \sigma^2)$, alors μ et σ^2 inconnus. Sur 10 jours, on obtient les résultats suivants :

Jour	1	2	3	4	5	6	7	8	9	10
Nbd'appels	200	240	190	150	220	180	170	230	210	210

Donner des estimations ponctuelles de μ et de σ^2.

4.3 Estimation par intervalle de confiance

4.3.1 Principe

Lorsqu'on cherche à estimer une quantité θ, il est plus réaliste de fournir une réponse du type θ est très probablement dans l'intervalle $[t_1, t_2]$, en acceptant un risque (mesuré !) de faire une erreur plutot qu'une estimation ponctuelle brute $\theta = t$ qui n'a, de toutes façons, presque aucune chance pour s'avérer exacte.

Définition 4.22. Soit X une variable aléatoire dont la loi dépend d'un parametre inconnu θ. On appelle **intervalle de confiance pour θ de niveau de confiance** $1-\alpha$ ou **de seuil** α un intervalle $[t_1, t_2]$ pour lequel on a :

$$\mathbb{P}(\theta \in [t_1, t_2]) = 1 - \alpha.$$

Le parametre θ est compris entre t_1 et t_1 avec probabilité $1-\alpha$.

Le seuil α est également appellé **risque**

Evidemment, plus le niveau de confiance est élevé, plus la certitude que la méthode fournisse un intervalle contenant la vraie valeur de θ augmente.

Remarque 4.23. Les niveau de confiance les plus utilisé sont 90%, 95% et 99%. Plus le niveau de confiance augmente, plus l'intervalle (et donc l'imprecision sur θ) augmente. Il faut donc choisir un bon compromis, qui dépendra vraiment au cas par cas du problème étudié.

On choisira le plus souvent les intervalles à risque symétriques, c'est-à-dire pour lesquels on a

$$\mathbb{P}(\theta < t_1) = \mathbb{P}(\theta > t_2) = \frac{\alpha}{2}.$$

Remarque 4.24. Intervalles de fluctuation et intervalles de confiance

La construction des intervalles de confiance est en quelque sorte la construction inverse de celle des intervalles de fluctuation : On prend vraiment le problème inversé :

Par exemple, dans le cas d'un caractère qui suit une loi de Bernouilli de parametre p :

On utilise un intervalle de fluctuation lorsque la proportion p est connue ou si on fait en amont une hypothèse sur sa valeur (prise de décision à partir d'un échantillon). La fréquence f observée dans un échantillon DOIT appartenir À l'intervalle de fluctuation considérée.

On utilise un intervalle de confiance lorsqu'on veut estimer une proportion inconnue p dans une population à partir de la fréquence f observée dans un échantillon (estimation, par exemple dans le cadre d'un sondage).

4.3.2 Intervalle de confiance d'une moyenne

4.3.2.1 Cas Gaussien

On supporte que la loi du caractère étudié X est la loi normale $\mathcal{N}(\mu, \sigma^2)$. On distingue deux cas de figures, selon que σ^2 est connue ou non.

Si σ^2 est connue :

Puisque X suit une loi $\mathcal{N}(\mu, \sigma^2)$, $\overline{X}_n$ suit une loi $\mathcal{N}(\mu, \frac{\sigma^2}{n})$, ou encore $\dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$ suit une loi $\mathcal{N}(0,1)$.

On fixe le risque α et on cherche dans la table de la loi normale $\mathcal{N}(0,1)$ la valeur $u_{\frac{\alpha}{2}}$ pour laquelle on a :

$$\mathbb{P}\left(-u_{\frac{\alpha}{2}} < \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} < u_{\frac{\alpha}{2}}\right) = 1-\alpha.$$

C'est-à-dire que

$$\mathbb{P}\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} < u_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}.$$

($u_{\frac{\alpha}{2}}\alpha$ est le fractile d'ordre $1-\frac{\alpha}{2}$ de la loi normale centrée réduite) On en déduit que

$$\mathbb{P}\left(\overline{X}_n - u_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X}_n + u_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha.$$

Ainsi, si $\overline{x}_n$ est une réalisation de $\overline{X}_n$, alors l'intervalle de confiance au niveau $1-\alpha$ pour la valeur de μ est :

$$I_{1-\alpha} = \left[\overline{x}_n - u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \ , \ \overline{x}_n + u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right].$$

Si σ^2 est inconnue :

On a vu au chapitre précédent (Remarque 3.14) que $\dfrac{\overline{X}_n - \mu}{S_n/\sqrt{n}}$ suit une loi de Student $\mathcal{T}_{n-1}$.

On fixe le risque α et on cherche dans la table de la loi $\mathcal{T}_{n-1}$ la valeur $t_{\frac{\alpha}{2}}$ pour laquelle on a :

$$\mathbb{P}\left(-t_{\frac{\alpha}{2}} < \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} < t_{\frac{\alpha}{2}} \right) = 1-\alpha.$$

ou de manière équivalente, $t_{\frac{\alpha}{2}}$ tel que :

$$\mathbb{P}\left(\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} < t_{\frac{\alpha}{2}} \right) = 1 - \frac{\alpha}{2}.$$

Ainsi, on en déduit que

$$\mathbb{P}\left(\overline{X}_n - t_{\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}} < \mu < \overline{X}_n + t_{\frac{\alpha}{2}} \frac{\S_n}{\sqrt{n}} \right) = 1-\alpha.$$

et si $\overline{x}_n$ est une réalisation de $\overline{X}_n$ et s_n une réalisation de S_n, alors l'intervalle de confiance au niveau $1-\alpha$ pour la valeur de μ est :

$$I_{1-\alpha} = \left[\overline{x}_n - t_{\frac{\alpha}{2}} \frac{s_n}{\sqrt{n}} \ , \ \overline{x}_n + t_{\frac{\alpha}{2}} \frac{s_n}{\sqrt{n}} \right].$$

4.3.2.2 Cas non gaussien

On supporte que la loi du caractère étudié X inconnue, d'epérance μ et de variance σ^2. On distingue également deux cas de figures, selon que σ^2 est connue ou non mais tout repose sur le fait que lorsque n est assez grand ($n > 30$), va pouvoir se ramener au cas gaussien grâce au théor-m central limite. **On se limite donc à des grands échantillons : n¿30**.

Si σ^2 est connue :

La démarche est la même que dans le cas gaussien. Puisque X suit une loi d'epérance μ et de variance σ^2 et que l'echantillon est grand, $\overline{X}_n$ peut être assimilée à une loi $\mathcal{N}(\mu, \frac{\sigma^2}{n})$, ou encore $\dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$ est assimilée une loi $\mathcal{N}(0,1)$.

On fixe le risque α et on en déduit que si $\overline{x}_n$ est une réalisation de $\overline{X}_n$, alors l'intervalle de confiance au niveau $1-\alpha$ pour la valeur de μ est :

$$I_{1-\alpha} = \left[\overline{x}_n - u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \ , \ \overline{x}_n + u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right].$$

Si σ^2 est inconnue :

Puisque $\overline{X}_n$ est assimilée à une loi gaussienne lorsque $n > 30$, on peut également approcher $\dfrac{\overline{X}_n - \mu}{S_n/\sqrt{n}}$ par une loi de Student $\mathcal{T}_{n-1}$ lorsque $n > 30$.

On fixe le risque α et on peut prendre l'intervalle de confiance au niveau $1-\alpha$ pour la valeur de μ trouvé dans le cas gaussien :

Si $\overline{x}_n$ est une réalisation de $\overline{X}_n$ et s_n une réalisation de S_n, alors l'intervalle de confiance au niveau $1-\alpha$ pour la valeur de μ est

$$I_{1-\alpha} = \left[\overline{x}_n - t_{\frac{\alpha}{2}}\frac{s_n}{\sqrt{n}}\ ,\ \overline{x}_n + t_{\frac{\alpha}{2}}\frac{s_n}{\sqrt{n}}\right].$$

Remarque 4.25. D'une manière générale, plus n est grand, plus la taille de l'intervallede confiance à un seuil donné est petit et donc plus l'estimation est précise.

4.3.3 Intervalle de confiance de la variance d'une loi Normale

On suppose que le caractère étudié X suit une loi $\mathcal{N}(\mu, \sigma^2)$.

Si μ est connue (peu fréquent)

D'après la Proposition 4.18, $T_n^2 = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n}(X_k - \mu)^2$ est un estimateur sans biais de σ^2. On peut montrer également qu'il est efficace.

De plus, puisque X suit une loi $\mathcal{N}(\mu, \sigma^2)$, la variable aléatoire $\dfrac{nT_n^2}{\sigma^2} = \displaystyle\sum_{k=1}^{n}\left(\frac{X_k - \mu}{\sigma}\right)^2$ est la somme de n variables de loi $\mathcal{N}(0,1)$ donc suit une loi du Khi-2 χ_n^2.

Ainsi, l'erreur α étant fixée, on cherche dans la table de χ_n^2 les valeurs $k_{\frac{\alpha}{2}}^n$ et $k_{1-\frac{\alpha}{2}}^n$ telles que :

$$\mathbb{P}\left(k_{\frac{\alpha}{2}}^n < \frac{nT_n^2}{\sigma^2} < k_{1-\frac{\alpha}{2}}^n\right) = 1-\alpha,$$

autrement dit $k_{\frac{\alpha}{2}}^n$ et $k_{1-\frac{\alpha}{2}}^n$ telles que : $\begin{cases} \mathbb{P}\left(k_{\frac{\alpha}{2}}^n < \frac{nT_n^2}{\sigma^2}\right) = \frac{\alpha}{2}, \\ \mathbb{P}\left(\frac{nT_n^2}{\sigma^2} < k_{1-\frac{\alpha}{2}}^n\right) = 1-\frac{\alpha}{2}, \end{cases}$

De plus, on a : $\mathbb{P}\left(\dfrac{nT_n^2}{k_{1-\frac{\alpha}{2}}^n} < \sigma^2 < \dfrac{nT_n^2}{k_{\frac{\alpha}{2}}^n}\right) = 1-\alpha.$

Si t_n^2 est une réalisation de T_n^2, alors l'intervalle de confiance pour σ^2 de seuil $1-\alpha$ est :

$$I_{1-\alpha} = \left[\frac{nt_n^2}{k_{1-\frac{\alpha}{2}}^n}\ ,\ \frac{nt_n^2}{k_{\frac{\alpha}{2}}^n}\right]$$

Si μ est inconnue

D'après la Proposition 4.18, la variance corrigée $\overline{S}_n = \dfrac{1}{n-1}\displaystyle\sum_{k=1}^{n}(X_k - \overline{X})^2$ est un estimateur sans biais qui converge en moyenne quadratique vers σ^2.

De plus, puisque X suit une loi $\mathcal{N}(\mu, \sigma^2)$, la variable aléatoire $\dfrac{nS_n^2}{\sigma^2}$ suit une loi du Khi-2 χ^2_{n-1}.

Ainsi, l'erreur α étant fixée, on cherche dans la table de χ^2_{n-1} les valeurs $k^{n-1}_{\frac{\alpha}{2}}$ et $k^{n-1}_{1-\frac{\alpha}{2}}$ telles que :

$$\mathbb{P}\left(k^{n-1}_{\frac{\alpha}{2}} < \frac{nS_n^2}{\sigma^2} < k^{n-1}_{1-\frac{\alpha}{2}}\right) = 1-\alpha,$$

autrement dit $k^{n-1}_{\frac{\alpha}{2}}$ et $k^{n-1}_{1-\frac{\alpha}{2}}$ telles que : $\begin{cases} \mathbb{P}\left(k^{n-1}_{\frac{\alpha}{2}} < \frac{nS_n^2}{\sigma^2}\right) = \frac{\alpha}{2}, \\ \mathbb{P}\left(\frac{nS_n^2}{\sigma^2} < k^{n-1}_{1-\frac{\alpha}{2}}\right) = 1-\frac{\alpha}{2}, \end{cases}$

De plus, on a : $\mathbb{P}\left(\dfrac{nS_n^2}{k^{n-1}_{1-\frac{\alpha}{2}}} < \sigma^2 < \dfrac{nS_n^2}{k^{n-1}_{\frac{\alpha}{2}}}\right) = 1-\alpha.$

Si s_n^2 est une réalisation de S_n^2, alors l'intervalle de confiance pour σ^2 de seuil $1-\alpha$ est :

$$I_{1-\alpha} = \left[\frac{ns_n^2}{k^{n-1}_{1-\frac{\alpha}{2}}}\ ,\ \frac{ns_n^2}{k^{n-1}_{\frac{\alpha}{2}}}\right]$$

4.3.4 Intervalle de confiance d'une proportion

La fréquence empirique $F_n = \dfrac{\text{Card}\left(\{k\ :\ X_k \text{ a le caractère } A\}\right)}{n}$ est un estimateur de la proportion π d'apparition d'un caractère A.

D'autre part, lorsque $n\pi$ et $n(1-\pi)$ sont supérieurs à 5 (ou autres conditions, du type $n > 30$...), on peut assimiler la loi de F_n à une loi $\mathcal{N}(\pi, \pi(1-\pi))$, ou encore dire que $\dfrac{F_n - \pi}{\sqrt{\pi(1-\pi)/n}}$ suit une loi $\mathcal{N}(0,1)$.

On fixe le risque α et on cherche dans la table de la loi normale $\mathcal{N}(0,1)$ la valeur $u_{\frac{\alpha}{2}}$ pour laquelle on a :

$$\mathbb{P}\left(-u_{\frac{\alpha}{2}} < \frac{F_n - \pi}{\sqrt{\pi(1-\pi)/n}} < u_{\frac{\alpha}{2}}\right) = 1-\alpha.$$

C'est-à-dire que

$$\mathbb{P}\left(\frac{F_n - \pi}{\sqrt{\pi(1-\pi)/n}} < u_{\frac{\alpha}{2}}\right) = 1-\frac{\alpha}{2}.$$

($u_{\frac{\alpha}{2}}\alpha$ est le fractile d'ordre $1-\frac{\alpha}{2}$ de la loi normale centrée réduite) On en déduit que

$$\mathbb{P}\left(F_n - u_{\frac{\alpha}{2}}\sqrt{\frac{\pi(1-\pi)}{n}} < \pi < F_n + u_{\frac{\alpha}{2}}\sqrt{\frac{\pi(1-\pi)}{n}}\right) = 1-\alpha.$$

Le problème, c'est que $\pi(1-\pi)$ est inconnu... On a 2 méthodes :

1. Si f_n est une réalisation de F_n, on approche $\pi(1-\pi)$ par $f_n(1-f_n$.

Ainsi, l'intervalle de confiance au niveau $1-\alpha$ pour la valeur de μ est :

$$I_{1-\alpha} = \left[f_n - u_{\frac{\alpha}{2}} \sqrt{\frac{f_n(1-f_n)}{n}} \ , \ f_n + u_{\frac{\alpha}{2}} \sqrt{\frac{f_n(1-f_n)}{n}} \right].$$

2. Méthode de l'ellipse (moin classique, mais plus rigoureuse).
$\mathbb{P}\left(F_n - u_{\frac{\alpha}{2}} \sqrt{\frac{\pi(1-\pi)}{n}} < \pi < F_n - u_{\frac{\alpha}{2}} \sqrt{\frac{\pi(1-\pi)}{n}} \right) = \mathbb{P}\left(|F_n - \pi| < u_{\frac{\alpha}{2}} \sqrt{\frac{\pi(1-\pi)}{n}} \right).$
donc $\mathbb{P}\left(|F_n - \pi| < u_{\frac{\alpha}{2}} \sqrt{\frac{\pi(1-\pi)}{n}} \right) = \mathbb{P}\left((F_n - \pi)^2 < \left(u_{\frac{\alpha}{2}} \sqrt{\frac{\pi(1-\pi)}{n}} \right)^2 \right) = 1-\alpha$. La quantité $\left(u_{\frac{\alpha}{2}} \sqrt{\frac{\pi(1-\pi)}{n}} \right)^2 - (F_n - \pi)^2$ est un polynôme de degré 2 en π, dont on peut calculer ses racines π_1 et π_2 (ordonnées) et de ce fait connaitre l'intervalle $[\pi_1, \pi 2]$ pour lequel $\left(u_{\frac{\alpha}{2}} \sqrt{\frac{\pi(1-\pi)}{n}} \right)^2 - (F_n - \pi)^2 > 0$. Cet intervalle donne l'intervalle de confiance de π au seuil $1-\alpha$: $I_{1-\alpha} = [\pi_1, \pi 2]$

Chapitre 5

Tests statistiques

5.1 Généralités sur les tests

Les tests sont des méthodes statistiques d'aide à la décision. En général, on formule une hypothèse relative à une population (répartition, indépendance, liaison entre caractères, valeur d'un parametre, etc...) et on cherche à valider cette hypothèse à partir des données recuiellies sur un échantillon. En gros, on cherche à répondre à la question :

L'échantillon que l'on a est-il compatible avec l'hypothèse qu'on a faite ?

Le principe est donc le suivant : On émet une hypothèse H énoncée en langage courant. Exemples : Un lot de pièces produits par une usine est-il conforme à une norme ? deux caractères sont-ils indépendants ? y a-t-il un lien entre 2 caractères ? un caractère est-il reparti selon une loi fixée ?

Cette hypothèse H est traduite sous la forme de 2 hypothèses contradictoires :

- (H_0) **Hypothèse nulle**,
  (égalité d'un parametre à une valeur, indépendance, liaison lineaire,etc...)
- (H_1) **Hypothèse alternative**

L'hypothèse nulle sera acceptée ou refusée à l'aide d'une variable d'échantillon V (on se restreint ici au cas ou il n'y en a qu'une seule mais il pourrait eventuellement en avoir plusieurs).

La définition des règle de décision nécessite d'abord de définir un **risque** α petit (1%, 5%...) déterminant le **seuil de signification du test** : $1-\alpha$ (plus le risque est faible, plus le résultat est significatif).

On obtient sur l'échantillon une réalisation v de la statistique V qui sera comparée à une **valeur critique** v_α ou à un **intervalle critique** I_α. Ce test de comparaison permettra soit d'accepter ou de rejeter H_0.

Par exemple, si l'on définit v_α tel que $\mathbb{P}(V \geq v_\alpha) = \alpha$, H_0 sera acceptée si $v \leq v_\alpha$ et refusée sinon ; si l'on définit I_α tel que $\mathbb{P}(V \notin I_\alpha) = \alpha$, H_0 sera acceptée si $v \in I_\alpha$ et refusée sinon.

Quoi qu'il advienne, il y a toujours un **risque d'erreur**, mais qui peut être scindé en deux cas de figure :

- **Erreur de type 1** : H_0 est rejetée à tort.
  Cette erreur est controlée : c'est α.
- **Erreur de type 2** : H_0 est acceptée à tort.
  Cette erreur n'est ni controlée ni controlable. On la note en général β.

En fait, plus α diminue, plus β augmente, et inversement.Il faut donc trouver un compromis qui dépend en général du cadre d'application du test et les risques encourus. En général, on choisi $\alpha = 1\%$ ou $\alpha = 5\%$.

5.2 Tests du Chi-2

5.2.1 Chi-2 d'adéquation à une distribution - Test d'ajustement

Cas typique : Une étude antérieure a montré une répartition d'un caractère X, de modalités $(x_i)_{i=1...m}$, selon une répartition des fréquences $(f_i)_{i=1...m}$ (chaque modalité x_i apparait avec la fréquence f_i). On refait un sondage quelque temps après, auprès de N individus. Peut-on dire si la répartition a évolué ou si elle est toujours la même ?

- (H_0) : la répartition des fréquences du caractère X est $(f_i)_{i=1...m}$
- (H_1) : certaines modalités x_i de X n'apparaissent pas avec la fréquence f_i.

Les variables d'échantillon dont on se sert ici seront les fréquences observées F_i de chaque modalité x_i, ou les effectifs observés E_i selon les données. La variable de test est :

$$K^2 = N \sum_{i=1}^{m} \frac{(F_i - f_i)^2}{f_i} = \frac{1}{N} \sum_{i=1}^{m} \frac{(E_i - N f_i)^2}{E_i}.$$

Si la répartition des fréquences est bien $(f_i)_{i=1...m}$, alors pour N assez grand ($Nf_i > 5$ pour au moins 80% des modalités), les effectifs observés F_i sur répartissent autour de leur moyennes f_i selon une loi normale... Dans ce cas, K^2 est en fait une somme de carrés de variables aleatoires qui suivent toutes la loi normale $N(0,1)$ (la dernière étant en fait dépendante des $m-1$ autres).

K^2 suit donc une loi du chi-2 à $m-1$ degrés de liberté :

$$K^2 \sim \chi^2_{m-1}.$$

La valeur critique de ce test sera donc v_α telle que

$$\mathbb{P}\left(\chi^2_{m-1} > v_\alpha\right) = \alpha.$$

Ainsi, si k^2 est la réalisation de K^2 sur notre échantillon :

- Si $k^2 > v_\alpha$, (H_0) est rejetée,
- Si $k^2 \leq v_\alpha$, (H_0) est acceptée.

5.2.2 Chi-2 d'indépendance

Cas typique : Une étude porte sur deux caractères X et Y, de modalités $(x_i)_{i=1...m}$ et $(y_j)_{j=1...p}$, On prélève un échantillon de N individus. Les caractères X et Y sont-ils indépendants ?

- (H_0) : X et Y sont indépendants sur l'ensemble de la population,
- (H_1) : il existe un lien entre X et Y sur l'ensemble de la population.

Les variables d'échantillon dont on se sert ici seront les fréquences observées F_i de chaque modalité x_i, ou les effectifs observés E_i selon les données. La variable de test est :

$$K^2 = N \sum_{i=1}^{m} \sum_{j=1}^{p} \frac{(F_{ij} - f_{ij})^2}{f_{ij}}.$$

Si les deux caractères sont bien indépendants, alors pour N assez grand ($Nf_i > 5$ pour au moins 80% des modalités), les effectifs observés F_{ij} sur répartissent autour de leur moyennes f_{ij} selon une loi normale... Dans ce cas, K^2 est en fait une somme de carrés de variables aleatoires qui suivent toutes la loi normale $N(0,1)$ (la dernière ligne $i = m$ et la dernière colonne $j = p$ étant en fait dépendantes des $(m-1)(p-1)$ autres).

K^2 suit donc une loi du chi-2 à $(m-1)(p-1)$ degrés de liberté :

$$K^2 \sim \chi^2_{(m-1)(p-1)}.$$

La valeur critique de ce test sera donc v_α telle que

$$\mathbb{P}\left(\chi^2_{(m-1)(p-1)} > v_\alpha\right) = \alpha.$$

Ainsi, si k^2 est la réalisation de K^2 sur notre échantillon :

- Si $k^2 > v_\alpha$, (H_0) est rejetée,
- Si $k^2 \leq v_\alpha$, (H_0) est acceptée.

5.3 Tests de comparaison de moyennes

5.3.1 Comparaison d'une moyenne à une valeur fixée

Cas typique : Il existe une norme concernant le caractère X d'un produit (taille donné, poids fixé, etc...). Une usine fabrique ce produit et on veux savoir à partir d'un echantillon de taille N le produit fabriqué correspond aux normes.

- (H_0) : $\mathbb{E}(X) = m$
- (H_1) : $\mathbb{E}(X) \neq m$ (ou $\mathbb{E}(X) > m$ ou $\mathbb{E}(X) < m$ selon le problème...

Les variables d'échantillon dont on se sert ici seront la moyenne $\overline{X}_N$ et éventuellement la variance corrigée S^2_N si jamais la variance n'est pas connue.

La variable de test étant $\overline{X}_N$, ou plus précisement :

– Si la variance σ^2 de X est connue : $T = \dfrac{\overline{X}_N - \mu}{\sigma/\sqrt{N}}$.
– Si la variance σ^2 de X est inconnue : $T = \dfrac{\overline{X}_N - \mu}{S_N/\sqrt{N}}$.

Lorsque X est sensée suivre une loi normale ou si l'échantillon est grand ($N > 30$) la loi de X étant inconnue, on a :

– Si la variance σ^2 de X est connue : $T = \dfrac{\overline{X}_N - \mu}{\sigma/\sqrt{N}} \sim \mathcal{N}(0,1)$.
– Si la variance σ^2 de X est inconnue : $T = \dfrac{\overline{X}_N - \mu}{S_N/\sqrt{N}} \sim \mathcal{T}_{N-1}$.

L'intervalle critique pour T au risque α de ce test sera un intervalle I_α qui dépend en fait de la formulation de (H_1) :

– Si (H_1) est $\mathbb{E}(X) \neq m$:
L'intervalle est choisi symétrique par rapport à 0. C'est le cas le plus classique.

$$I_\alpha = [-u_\alpha, u_\alpha] \text{ avec } \mathbb{P}(-u_\alpha < T < u_\alpha) = 1 - \alpha$$

– Si (H_1) est $\mathbb{E}(X) > m$: $I_\alpha =]-\infty, u_\alpha]$ avec $\mathbb{P}(T < u_\alpha) = 1 - \alpha$
– Si (H_1) est $\mathbb{E}(X) < m$: $I_\alpha = [-u_\alpha, +\infty]$ avec $\mathbb{P}(-u_\alpha < T) = 1 - \alpha$

Ainsi, si t est la réalisation de T sur notre échantillon :
– Si $t \notin I_\alpha$, (H_0) est rejetée,
– Si $t \in I_\alpha$, (H_0) est acceptée.

5.3.2 Comparaison de deux moyennes

Cas typique : Un stock d'objets ayant un certain caractère provient de 2 fournisseurs différents et indépendants. On observe le même caractère sur chacune des 2 populations : $X^{(1)}$ et $X^{(2)}$. A partir de deux échantillons : N_1 individus du premier fournisseur et N_2 individus du deuxième fournisseur, peut-on considérer que les moyennes respectives m_1 et m_2 de $X^{(1)}$ et $X^{(2)}$ sont-elles les mêmes ?

En fait, on se ramène à un test de comparaison d'une moyenne à une valeur fixe en utilisant la variable de test suivante :

$$D = \overline{X^{(1)}_{N_1}} - \overline{X^{(2)}_{N_2}}.$$

En effet, si comme d'habitude, on se place dans le cas où $X^{(1)}$ et $X^{(2)}$ suivent une loi normale ou si les échantillons sont grands ($N_1 > 30$ et $N_2 > 30$) les lois de $X^{(1)}$ et $X^{(2)}$ sont inconnues, on a $\overline{X^{(1)}_{N_1}} \sim \mathcal{N}(m_1, \dfrac{\sigma_1^2}{N_1})$ et $\overline{X^{(2)}_{N_2}} \sim \mathcal{N}(m_2, \dfrac{\sigma_2^2}{N_2})$ et donc

$$D \sim \mathcal{N}\left(m_1 - m_2, \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}\right).$$

Dans le cas où les variances sont connues, on ramène donc le test de comparaison de 2 moyennes à un test de comparaison d'une moyenne à une valeur fixée :

– $(H_0) : \mathbb{E}(D) = 0$
– $(H_1) : \mathbb{E}(D) \neq 0$ (ou $\mathbb{E}(D) \leq 0$ ou $\mathbb{E}(D) \geq 0$ selon les données du problème)

Si les variances sont inconnues, il faut avoir recours à à variance corrigée pour chacune des 2 variables $X^{(1)}$ et $X^{(2)}$, afin d'obtenir un estimateur de la variance de D : $S^2_{N_1+N_2} = \frac{{S^{(1)}}^2}{N_1} + \frac{{S^{(2)}}^2}{N_2}$