Cours de Mathématiques

UE : Maths 4

L3 génie civil

Probabilités et statistiques

4 septembre 2025

Alexandre MIZRAHI

Table des matières

Présentation de l'organisation de l'unité d'enseignement

Cette unité d'enseignement est organisée sous forme de classe inversée, elle est constituée de 10 séances de cours et de 20 séances de TD :

Pour chacune des semaines d'enseignement contenant un cours en amphi :

1. Elle commence par un travail personnel (environ 45 minutes)
   — Visionner 1 ou 2 vidéos sur la vidéothèque
   — Lire le chapitre du poly correspondant
   — Répondre à un QCM sur la plateforme pédagogique (Avant la veille du CM 23h00)

2. Elle se poursuit par un travail en CM
   — 30 minutes : Résumé de cours, questions (les retards ne sont pas acceptés).
   — 45 minutes : Travail en petit groupe sur un thème du cours, avec une production à rendre en fin de séance.
   — 15 minutes : Correction du travail de groupe.

3. Elle se termine par la recherche d'exercices en TD.

L'évaluation de cette UE est constituée de trois éléments :

— Résultats aux QCM.
— Productions de groupe en CM.
— Contrôle de fin de semestre. Durant le contrôle les documents sont interdits, toutefois une feuille manuscrite A4 est tolérée. Une calculatrice basique (4 opérations) ou une calculatrice scientifique en mode examen (diode clignotante) est autorisée. Les smartphones et tous les objets connectés sont strictement interdit.

Cours de la semaine 1

Chapitre 1

Espace de probabilité

1.1 Généralités

Une des difficultés de l'apprentissage des probabilités est la permanente confusion entre les mathématiques et les problèmes concrets que cette théorie permet d'étudier. Dans ce cours on ne précisera pas si on se trouve dans le monde des mathématiques ou dans le monde physique en étant conscient que l'on fait un grave abus de langage. Un des buts des Probabilités est d'étudier les phénomènes aléatoires, on a des informations sur eux mais pas assez pour prévoir les résultats. On va construire un modèle, c'est à dire un objet mathématique qu'il faudra interpréter concrètement. Par exemple, lors d'expériences, différents résultats sont possibles, on ne peut pas prévoir le résultat mais on peut créer un modèle, chaque résultat possible est représenté par un élément d'un ensemble Ω, et on lui associe un poids qui va correspondre à 'la chance' d'obtenir ce résultat.

Il y a deux façons naturelles de construire un modèle probabiliste : Soit on s'appuie sur des considérations physiques (symétrie, théorie physique,...) par exemple un dé a la même chance de tomber sur chacune des faces. Soit on s'appuie sur des expériences, on observe le phénomène puis on le modélise à l'aide de statistiques réalisées sur la partie expérimentale.

Exemple 1.1
1. On jette un dé. $\Omega = \{1, 2, 3, 4, 5, 6\}$.
2. Formation de binôme dans une classe de 50 étudiants $\Omega = \{(i, j) | 1 \leq i \leq j \leq 50\}$. Ω possède $\binom{5}{2} = \frac{50.49}{2}$ éléments, c'est le cardinal de Ω.

Définition 1.1 L'ensemble qui représente tous les possibles est souvent appelé univers, on le note souvent Ω. On appelle événement une partie de Ω, cela peut s'interpréter comme une famille de résultats possibles, on veut pouvoir calculer la probabilités d'une telle partie, c'est à dire 'la chance' que le résultat fasse partie de cette famille. on note $\mathcal{P}(\Omega)$ l'ensemble des parties de E.

Exemple 1.2 Si $\Omega = \{1; a\}$, alors $P(\Omega) = \big\{\emptyset, \{a\}; \{1\}; \{1; a\}\big\}$.

Définition 1.2 On appelle mesure de probabilité (ou probabilité) toute fonction P de $\mathcal{P}(\Omega)$ dans $\mathbb{R}^+$ qui a les propriétés suivantes :
1) $P(\Omega) = 1$
2) Si $A \cap B = \emptyset$ alors $P(A \cup B) = P(A) + P(B)$.
3) Si $(A_i)_{i \in \mathbb{N}}$ est une suite de partie deux à deux disjointes ($i \neq j \Rightarrow A_i \cap A_j = \emptyset$), alors :

$$P\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} P(A_i)$$

Remarque 1.1
- Les propriétés 1) et 2) sont des propriétés de mesures (longueurs, volumes, masses, etc...), la propriété 3) est une généralisation de la propriété 2.
- Si Ω est un ensemble fini (ou dénombrable) alors une probabilité P est définie dès que l'on connaît ses valeurs sur les singletons de Ω.
- Lorsque Ω est grand (par exemple pas dénombrable) on ne peut définir P que sur une partie de Ω, c'est à dire que l'on ne peut "mesurer" que certaines parties, on appelle tribu de Ω un ensemble de parties que l'on peut mesurer.

Vidéo 1.1 - Mesure de probabilités

Exemple 1.3 $\Omega = \{a; b; c\}, P(\{a\}) = P(\{b\}) = \frac{1}{4}$ et $P(\{c\}) = \frac{1}{2}$ alors $P(\{a, c\}) = \frac{3}{4}$.

Proposition 1.1 1. $P(\emptyset) = 0$
2. $P(\Omega \setminus A) = 1 - P(A)$
3. $A \subset B \Rightarrow P(B) = P(A) + P(B \setminus A) \geq P(A)$
4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Proposition 1.2 Si $(A_i)_i$ est une partition de Ω (c'est à dire $\cup A_i = \Omega$ et $i \neq j \Rightarrow A_i \cap A_j = \varnothing$), alors pour tout événement B on a :

$$P(B) = \sum_{i \in \mathbb{N}} P(A_i \cap B) = \lim_{n \to \infty} \sum_{i=1}^{n} P(A_i \cap B)$$

Proposition 1.3

$$P(A \cup B \cup C) = P(A) + P(B) + P(B) - P(A \cap B) - P(A \cap C) - P(C \cap B) + P(A \cap B \cap C)$$

1.2 Probabilité conditionnelle

On dispose d'une mesure de probabilité P sur un ensemble Ω, on cherche à en déduire une mesure de probabilité sur une partie A de Ω, on fait l'hypothèse que la probabilité reste proportionnelle à P mais que seule les valeurs dans A ont une chance d'arriver, on est donc amené à définir une nouvelle probabilité P_A telle que $P_A(B)$ soit proportionnelle à $P(A \cap B)$, car on se trouve obligatoirement dans A, pour être dans B il faut donc être dans $A \cap B$. $P_A(B) = \alpha P(A \cap B)$, mais la probabilité de se trouver dans A doit être égale à 1. On obtient donc $P_A(A) = \alpha P(A \cap A) = \alpha P(A) = 1$

Définition 1.3 Si A est un événement de probabilité non nulle, et B un événement quelconque, on définit la probabilité conditionnelle de B sachant A la probabilité :

$$P_A(B) = P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Proposition 1.4 P_A est une nouvelle mesure de probabilité sur Ω.

Exemple 1.4 On jette deux dés, on sait que la somme des deux dés est 10, quelle est la probabilité d'avoir un double.

Différentes modélisations sont possibles on peut par exemple prendre pour Ω l'ensemble des couples $\{(d_1; d_2) | d_1, d_2 \in \{1; 2; 3; 4; 5; 6\}\}$ cela revient à distinguer les deux dés. Pour P la probabilité uniforme (on parle d'équiprobabilité) sur Ω c'est à dire que l'on donne le même poids à chaque élément de Ω, comme il y a $6 \times 6 = 36$ éléments dans Ω, chaque élément a une probabilité de $\frac{1}{36}$. Obtenir 10 correspond à l'événement $A = \{(4; 6); (5; 5); (6; 4)\}$, $A \cap B$ correspond a obtenir un double dont la somme vaut 10, $A \cap B = \{(5; 5)\}$, on obtient donc

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{36}}{\frac{3}{36}} = \frac{1}{3}$$

On peut aussi prendre pour Ω l'ensemble $\{(4, 6), (5, 5), (6, 4)\}$ avec équiprobabilité, on obtient encore $\frac{1}{3}$, mais on n'utilise pas ici les probabilités conditionnelles.

Proposition 1.5 On a en outre la formule $P(A \cap B) = P(B|A)P(A)$

Preuve : trivial.

1.3 Indépendance

Définition 1.4 Deux événements A et B sont indépendants si $P(A \cap B) = P(A)P(B)$.

Remarque 1.2 Cela revient à dire que la probabilité de B sachant A est égale à la probabilité de B, le fait de savoir que l'on se trouve dans A ne nous apprend rien sur le fait de se trouver ou non dans B. Cela correspond bien à l'idée intuitive que A ne dépend pas de B. L'idée intuitive d'indépendance entre deux éléments se modélise par l'indépendance en probabilité, en revanche il peut y avoir indépendance en probabilité sans qu'il y ait cette notion d'indépendance intuitive.

Exemple 1.5 Par exemple on veut modéliser le fait de tirer une carte au hasard, A correspond à l'événement tirer un coeur et B à tirer un roi, on voit bien que A et B doivent être indépendants dans le modèle que l'on va proposer. Par contre si le jeu n'est pas complet, par exemple si il manque le 8 de pique alors on n'a plus indépendance.

Définition 1.5 n événements $(A_i)_{i\in I}$ sont indépendants dans leur ensemble si pour toute partie finie $J \subset I$, on a

$$P(\cap_{j\in J} A_j) = \prod_{j\in J} P(A_j)$$

Remarque 1.3 Si n événements sont indépendants dans leur ensemble alors ils sont indépendants deux à deux. La réciproque est fausse.

Vidéo 1.2 - Probabilité conditionnelle - évènements indépendants

1.4 Rappel de combinatoire

La combinatoire est particulièrement importante lorsqu'il y a équiprobabilité, dans ce cas le calcul de probabilités se ramène à des calculs de cardinaux d'ensembles (c'est de la combinatoire).

1. On peut former n^k mots de k lettres constitués avec n signes différents, ou encore il y a n^k façons de ranger k boules différentes dans n urnes différentes. Il y a répétition et ordre. Si E est un ensemble fini de cardinal n, il y a n^p, p-listes d'élément de E : c'est à dire avec ordre et répétition. Il y a aussi n^p application d'un ensemble de cardinal p vers un ensemble de cardinal n. On note souvent F^E l'ensemble des applications de E dans F. On a alors $\text{card}(F^E) = \text{card}(F)^{\text{card}(E)}$.

2. Il y a A_n^k façons de ranger k individus choisis parmi n, c'est aussi le nombre de mots de k lettres que l'on peut écrire avec n lettres en bois différentes. Il y a A_n^p p-liste sans répétition possible avec un ensemble à n éléments. On utilise les arrangements lorsqu'il n'y a pas répétition mais lorsqu'il y a un ordre.

$$A_n^k = n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$$

3. Il y a $\binom{n}{k}$ parties différentes à k éléments dans un ensemble à n éléments, c'est aussi le nombre de groupes de k personnes différents que l'on peut constituer à partir de n individus. On utilise les combinaisons lorsqu'il n'y a ni répétition ni ordre

$$\binom{n}{k} = \frac{n(n-1)\dots(n-k+1)}{k!} = \frac{A_n^k}{k!} = \frac{n!}{(n-k)!\,k!}$$

Proposition 1.6 Quelques propriétés des combinaisons :
- $\binom{n}{k} = \binom{n}{n-k}$.
- $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$. *Formule permettant de construire le triangle de Pascal.*
- Formule du binôme de Newton :

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Preuve : Pour la seconde on peut considérer un élément et regarder les parties qui le contiennent et celle qui ne le contiennent pas. Pour la troisième on développe un produit de n termes.

Remarque 1.4 Un ensemble de n éléments possède 2^n parties.

Vidéo 1.3 Cardinal de l'ensemble des parties et nombre de combinaisons

fin de la semaine 1

Cours de la semaine 2

Chapitre 2

Variables aléatoires

2.1 Introduction

Une façon très agréable de modéliser une expérience aléatoire est l'utilisation de fonctions particulières appelées variables aléatoires, pour lesquelles on définit des objets de la théorie des probabilités, permettant de les décrire.

2.2 Définitions

Définition 2.1 Soit Ω un ensemble et P une probabilité sur Ω, on appelle variable aléatoire sur Ω toute fonction de Ω dans $\mathbb{R}$.

Remarque 2.1 Une variable aléatoire est donc un objet mathématique qui peut prendre différentes valeurs.

Exemple 2.1 On peut modéliser la taille des individus d'une population donnée, à l'aide d'une variable aléatoire.

Exemple 2.2 On jette trois dés et on veut faire une étude probabiliste de la somme des trois résultats. On peut par exemple prendre $\Omega = \{1;2;3;4;5;6\}^3$ munie d'une probabilité uniforme et considérer la fonction

$$\begin{array}{rccc} X: & \Omega & \to & \mathbb{R} \\ & (\omega_1,\omega_2,\omega_3) & \mapsto & \omega_1+\omega_2+\omega_3 \end{array}$$

Il est intéressant de savoir par exemple quelle est la probabilité que la somme soit égale à 9, ce que l'on a envie de noter $P(X=9)$, en fait on veut mesurer la partie de Ω pour laquelle X prend la valeur 9, c'est à dire :

$$\{\{1,2,6\},\{1,3,5\},\{1,4,4\},\{1,5,3\},\{1,6,2\},\{2,1,6\},\{2,2,5\},\{2,3,4\},\{2,4,3\},\{2,5,1\},\{3,1,5\},\{3,2,4\},\dots\}$$

Définition 2.2 Soit X une variable aléatoire, A une partie de $\mathbb{R}$, x_0 un réel, on pose

$$\begin{array}{l} (X=x_0) = X^{-1}(x_0) = \{\omega\in\Omega | X(\omega)=x_0\} \\ (X\leq x_0) = X^{-1}(]-\infty;x_0[) = \{\omega\in\Omega | X(\omega)\leq x_0\} \\ (X\in A) = X^{-1}(A) = \{\omega\in\Omega | X(\omega)\in A\} \end{array}$$

Définition 2.3 Pour une variable aléatoire X, on peut définir une nouvelle mesure de probabilité P_X sur $\Omega=\mathbb{R}$ en posant :

$$\forall A\subset\mathbb{R},\ P_X(A) = P(X\in A)$$

On appelle cette mesure de probabilité, la loi de la variable aléatoire X.

Définition 2.4 On définit pour une variable aléatoire X sa fonction de répartition par

$$F_X(t) = P(X\leq t)$$

Proposition 2.1 On démontre facilement que F_X est croissante, tend vers 0 en $-\infty$ et vers 1 en $+\infty$. Moins facilement que la fonction de répartition F_X caractérise la loi de X : c'est à dire que si $F_X=F_Y$ alors $P_X=P_Y$.

Définition 2.5 Deux variables aléatoires X,Y sont indépendantes si pour toutes parties A,B de $\mathbb{R}$ les parties de Ω, $(X\in A)$ et $(Y\in B)$ sont indépendantes. C'est à dire pour toutes parties A et B de $\mathbb{R}$
$P\Big((X\in A)\cap(X\in B)\Big) = P(X\in A).P(X\in B)$.

Remarque 2.2 La notion intuitive d'indépendance, le fait de "ne pas dépendre de", se traduit pour des variables aléatoires par l'indépendance, par exemple le résultat d'un lancé de dé et le suivant, par contre le poids et la taille d'un individu choisi au hasard dans une population, ne peuvent raisonnablement pas être modélisés par des variables aléatoires indépendantes.

Définition 2.6 n variables aléatoires sont indépendantes si pour toutes parties $A_1, A_2, ..., A_n$ de $\mathbb{R}$ les événements $(X_1 \in A_1), \ldots, (X_n \in A_n)$ sont indépendants

Proposition 2.2 (admis) Si X et Y sont indépendantes alors pour toutes fonctions Φ et Ψ, $\Phi \circ X$ et $\Psi \circ Y$ sont indépendantes, on note usuellement ces nouvelles variables aléatoires $\Phi(X)$ et $\Psi(X)$.

Exemple 2.3 On jette deux dés, on modélise la somme par une variable aléatoire X et la valeur absolue de la différence par une variable aléatoire Y, est-il raisonnable de supposer X et Y indépendants ?

2.3 Variables aléatoires discrètes

Définition 2.7 On dit qu'une variable aléatoire X est discrète si $X(\Omega)$ est fini ou dénombrable, on note alors : $X(\Omega) = \{x_1, x_2, \ldots, x_n, \ldots\}$.

Proposition 2.3 La loi d'une variable aléatoire discrète est totalement déterminée par la donnée des $P(X = x_i)$, en effet pour toute partie A de $\mathbb{R}$:

$$P(X \in A) = \sum_{\{i/x_i \in A\}} P(X = x_i)$$

Vidéo 2.1 Variables aléatoires discrètes

2.3.1 L'espérance

Définition 2.8 Lorsque $X(\Omega) = \{x_1, x_2, \ldots, x_n\}$ est fini, on définit l'espérance de X par

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

Lorsque $X(\Omega) = \{x_1, x_2, \ldots, x_n, \ldots\}$ est dénombrable (non fini) si la série suivante est absolument convergente on définit l'espérance de X par

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} x_i P(X = x_i)$$

Vidéo 2.2 - Espérance d'une variable aléatoire discrète

Proposition 2.4 (admis) soit ϕ une fonction de $\mathbb{R}$ dans $\mathbb{R}$, si la série suivante est absolument convergente alors la variable aléatoire $\phi(X)$, possède une espérance et on a

$$\mathbb{E}(\phi(X)) = \sum_{i} \phi(x_i) P(X = x_i)$$

Remarque 2.3 Si on interprète $P(X = x_i)$ comme une fréquence théorique pour que le caractère $\mathcal{C}$ prenne la valeur x_i, alors l'espérance s'interprète comme une moyenne théorique des valeurs prises par le caractère $\mathcal{C}$.

Proposition 2.5 Soient X et Y des variables aléatoires, possédant une espérance, a et b des réels alors :

1. $\mathbb{E}(1) = 1$
2. $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$.
3. Si X et Y sont indépendantes alors $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Définition 2.9 On définit, lorsqu'elles existent, la variance par $\mathbb{V}\mathrm{ar}\,(X) = \mathbb{E}\big((X - \mathbb{E}(X))^2\big)$, l'écart type comme la racine carrée de la variance, et la covariance par

$$\mathrm{covar}\,(X, Y) = \mathbb{E}\Big(\big(X - \mathbb{E}(X)\big)\big(Y - \mathbb{E}(Y)\big)\Big)$$

Remarque 2.4 La variance permet de mesurer l'éloignement de X avec l'espérance de X, Plus la variance est grande plus souvent X est éloignée de son espérance un peu comme dans une classe ou la moyenne en proba stat serait de 10/20, si tous les étudiants ont 10/20 la variance est nulle si la moitié de la classe à 0 et l'autre moitié à 20 la variance est égale à 100 et l'écart type à 10.

Proposition 2.6 Soient X et Y des variables aléatoires possédant une variance et a un réel.

1. $\mathbb{V}\text{ar}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.
2. $\mathbb{V}\text{ar}(aX) = a^2\mathbb{V}\text{ar}(X)$.
3. $\text{covar}(X) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.
4. Si X et Y sont indépendantes alors $\mathbb{V}\text{ar}(X+Y) = \mathbb{V}\text{ar}(X) + \mathbb{V}\text{ar}(Y)$.
5. Si X et Y sont indépendantes alors $\text{covar}(X,Y) = 0$.

Remarque 2.5 On peut remarquer que la réciproque de la dernière propriété est fausse, la covariance peut être nulle et les variables aléatoires non indépendantes.

Vidéo 2.3 - Variance dune variable aléatoire discrète discrète

2.4 Variables aléatoires discrètes : lois classiques

2.4.1 Loi de Bernoulli

Définition 2.10 X suit une loi de Bernoulli de paramètre $p \in [0;1]$, si X ne prend que deux valeurs 0 et 1.

$$P(X=1) = p \quad \text{et} \quad P(X=0) = 1-p = q$$

On note $X \sim \mathcal{B}(p)$.

Proposition 2.7 Si X suit une loi de Bernoulli de paramètre p alors $\mathbb{E}(X) = p$ et $\mathbb{V}\text{ar}(X) = pq$.

Preuve : $\mathbb{E}(X) = 1P(X=1) + 0P(X=0) = p + 0 = p$
$\mathbb{V}\text{ar}(X) = (1-p)^2P(X=1) + (0-p)^2P(X=0) = (1-p)^2p + p^2(1-p) = p(1-p)(1-p+p) = p(1-p)$

2.4.2 Loi Binômiale

Définition 2.11 X suit une loi Binômiale de paramètre $(n,p) \in \mathbb{N} \times [0;1]$, si X prend les valeurs entre 0 et n et :

$$\forall k \in \{0;1;\ldots;n\},\ P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

On note $X \sim \mathcal{B}(n,p)$.

Proposition 2.8 Si X suit une loi de Binômiale de paramètre (n,p) alors

$$\mathbb{E}(X) = np\ , \mathbb{V}\text{ar}(X) = npq$$

Proposition 2.9 Si $X_1, \ldots, X_n$ sont n variables aléatoires indépendantes de même loi de Bernoulli de paramètre p alors la variable aléatoire $X_1 + X_2 + \ldots + X_n$ suit une loi binômiale de paramètre (n,p).

Remarque 2.6 C'est le cas typique du pile ou face, du oui ou non, en général on fait des sommes de variables aléatoires indépendantes de Bernoulli.

2.4.3 Loi de Poisson

Définition 2.12 X suit une loi de Poisson de paramètre $\lambda > 0$, si

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

On note $X \sim \mathcal{P}(\lambda)$.

Proposition 2.10 Si X suit une loi de Poisson de paramètre λ, alors

$$\mathbb{E}(X) = \lambda,\ \mathbb{V}\text{ar}(X) = \lambda$$

Remarque 2.7 On utilise ces lois pour modéliser un décompte : nombre d'événements arrivant durant une période donnée, ou pour approcher une loi binômiale lorsque n est très grand et p très petit(loi des événements rares).

2.4.4 Loi géométrique

Définition 2.13 X suit une loi géométrique de paramètre $a \in]0;1[$, si

$$\forall k \in \mathbb{N}^*,\ P(X=k) = a(1-a)^k$$

Proposition 2.11 Si X suit une loi géométrique de paramètre a alors

$$\mathbb{E}(X) = \frac{1}{a}\ , \mathbb{V}\mathrm{ar}\,(X) = \frac{1-a}{a^2}$$

Remarque 2.8 Typiquement si on veut modéliser, lors d'une suite de lancés d'une pièce, la première fois ou un pile sort.

2.4.5 Loi hypergéométrique

Définition 2.14 Soient $n \leq N < M$, X suit une loi hypergéométrique de paramètre (n, N, M) si

$$\forall k \in \big[\max(0, n-(N-M)); \min(n, M)\big],\ P(X=k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$$

Remarque 2.9 Typiquement si on veut modéliser le tirage sans remise de n boules le nombre de boules blanches prises, dans une urne contenant N boules dont M sont blanches.

⚐ *fin de la semaine 2*

▶ *Cours de la semaine 3*

Chapitre 3

Variable aléatoire à densité

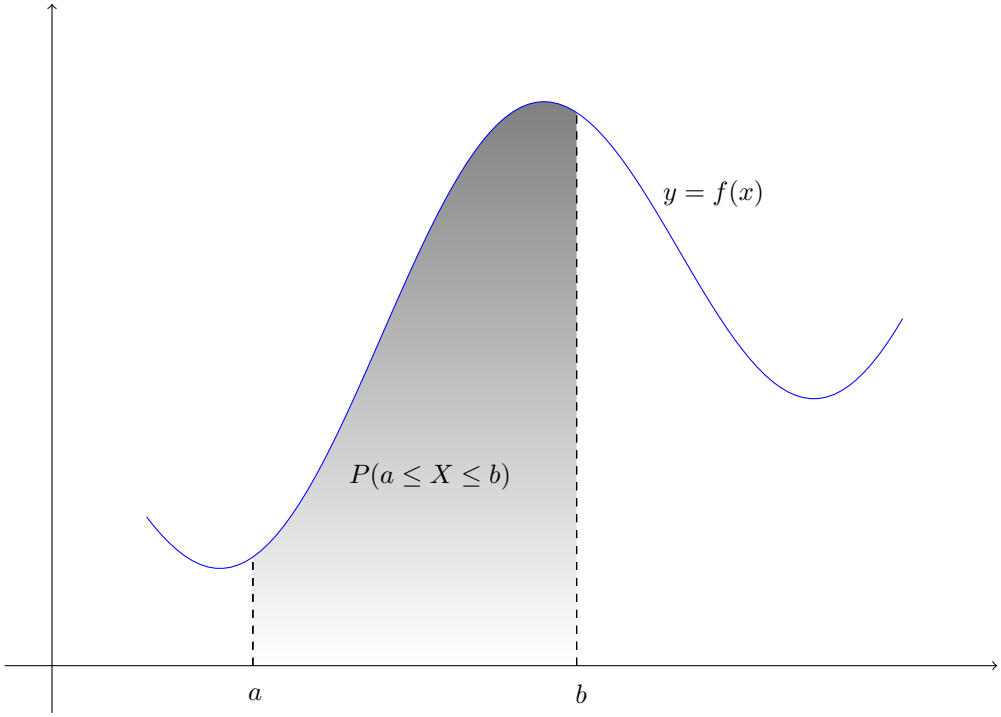
Définition 3.1 On dit qu'une variable aléatoire X est à densité si il existe une fonction $f : \mathbb{R} \to \mathbb{R}^+$ telle que pour tout réel $a \leqslant b$:

$$P(a \leq X \leq b) = \int_a^b f(t)\,\mathrm{d}t$$

f est appelée une densité de f.

Proposition 3.1 (admis) On a l'équivalence entre

1. f est une densité de X.
2. Pour tout ensemble A de réels on a $P(X \in A) = \int_A f(x)\,\mathrm{d}x$.
3. Pour tout réel t, $F_X(t) = \displaystyle\int_{-\infty}^{t} f(x)\,\mathrm{d}x$



Vidéo 3.1 - Variables aléatoires à densité

Remarque 3.1 La loi d'une variable aléatoire à densité est totalement déterminée par la donnée d'une densité f. Si X est une variable aléatoire à densité $P(X = a) = 0$ pour n'importe quel réel a. Cela peut sembler un peu étrange, il faut avoir en tête que pour δ petit, la probabilté $P(a \leq X \leq a + \delta) \approx f(a)\delta$.

Proposition 3.2 La fonction de répartition se déduit de la densité par $F_X(t) = P(X \leq t) = \int_{-\infty}^{t} f(x)\,\mathrm{d}x$. Réciproquement si F_X la fonction de répartition de X est continue sur $\mathbb{R}$ et dérivable sauf peut être en certains points alors X possède une densité donnée par $f_X = F'_X$, la dérivée de la fonction de répartition.

3.0.1 L'espérance

Vidéo 3.2 - Comment définir l'espérance d'une variable aléatoire à densité

Définition 3.2 Pour une variable aléatoire admettant la fonction f pour densité Lorsque l'intégrale suivante est convergente en valeur absolue ($\int_{-\infty}^{+\infty} |tf(t)|\,\mathrm{d}t$) on définit l'espérance de X par

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} tf(t)\,\mathrm{d}t$$

Remarque 3.2 L'espérance est une sorte de moyenne théorique des valeurs possibles de la variable aléatoire X.

Proposition 3.3 (admis) soit ϕ une fonction de $\mathbb{R}$ dans $\mathbb{R}$, si l'intégrale ($\int_{-\infty}^{+\infty} \phi(t)f(t)|\,\mathrm{d}t$ est convergente alors la variable aléatoire $\phi(X)$, possède une espérance et on a

$$\mathbb{E}(\Phi(X)) = \int_{-\infty}^{+\infty} \phi(t)f(t)\,\mathrm{d}t$$

En particulier $\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} t^2 f(t)\,\mathrm{d}t$

Proposition 3.4 soient X et Y des variables aléatoires, possédant une espérance, a et b des réels alors :

1. $\mathbb{E}(1) = 1$
2. $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$.
3. Si X et Y sont indépendantes et possède une espérance alors XY possède une espérance et $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Définition 3.3 On définit, lorsqu'elles existent, la variance par $\mathbb{V}\text{ar}(X) = \mathbb{E}\big((X-\mathbb{E}(X))^2\big)$, l'écart type comme pour les variables aléatoires discrètes est la racine carrée de la variance, et la covariance par

$$\text{covar}(X,Y) = \mathbb{E}\Big(\big(X-\mathbb{E}(X)\big)\big(Y-\mathbb{E}(Y)\big)\Big)$$

Proposition 3.5 Soient X et Y des variable aléatoires possédant une variance et a un réel.

1. $\mathbb{V}\text{ar}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.
2. $\mathbb{V}\text{ar}(aX) = a^2\mathbb{V}\text{ar}(X)$.
3. $\text{covar}(X) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.
4. Si X et Y sont indépendantes alors $\mathbb{V}\text{ar}(X+Y) = \mathbb{V}\text{ar}(X) + \mathbb{V}\text{ar}(Y)$.
5. Si X et Y sont indépendantes alors $\text{covar}(X,Y) = 0$.

Remarque 3.3 On peut remarquer que la réciproque de la dernière propriété est fausse, la covariance peut être nulle et les variables aléatoires non indépendantes.

3.1 Lois de probabilités à densité classiques

Vidéo 3.3 - Exemples de variables aléatoires à densité

3.1.1 Loi uniforme sur $[a,b]$

C'est la plus simple des lois, mais elle sert relativement peu. Une densité est donnée par :

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

Proposition 3.6 Son espérance et sa variance sont :

$$\mathbb{E}(X) = \frac{a+b}{2}; \quad \mathbb{V}\text{ar}(X) = \frac{(b-a)^2}{12}$$

3.1.2 Loi exponentielle

X suit une loi exponentielle de paramètre $\lambda > 0$ si elle admet pour densité :

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

Proposition 3.7 Son espérance et sa variance sont :

$$\mathbb{E}(X) = \frac{1}{\lambda}; \quad \mathbb{V}\text{ar}(X) = \frac{1}{\lambda^2}$$

Remarque 3.4 On utilise ces variables aléatoires pour modéliser des temps d'attentes, cela vient d'une caractérisation importante des VA exponentielles

$$P(X \geq t+t' | X \geq t') = P(X \geq t)$$

La probabilité que le phénomène se passe après l'instant $t+t'$ sachant qu'il ne s'est pas passé avant l'instant t' est égale à la probabilité qu'il se passe après l'instant t. En quelque sorte il n'y a pas de mémoire des phénomènes, le fait de savoir que le phénomène ne soit pas apparu avant l'instant t' remet tout à zéro.

3.1.3 Loi normale

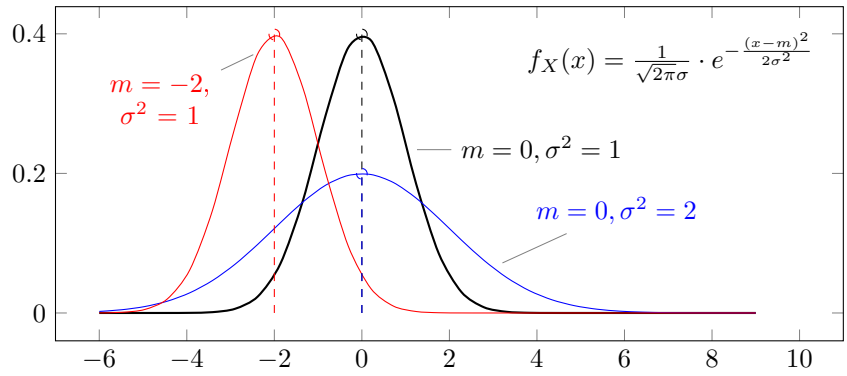
Appelée loi normale, loi de Gauss, loi de Laplace-Gauss, loi gaussienne, ou loi en cloche. On utilise la valeur de l'intégrale suivante assez difficile à calculer :

$$\int_{-\infty}^{+\infty} e^{-\frac{1}{2}t^2}\,\mathrm{d}t = \sqrt{2\pi}$$

Définition 3.4 X suit une loi normale de paramètre $(m, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*$, si elle admet pour densité :

$$f_X(t) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(t-m)^2}{2\sigma^2}}$$

On note souvent $X \sim \mathcal{N}(m, \sigma^2)$.



Proposition 3.8 (admis) Si $X \sim \mathcal{N}(m, \sigma^2)$ et λ un réel non nul, alors $X + \lambda$ suit une loi normale $\mathcal{N}(m + \lambda, \sigma^2)$ et la variable aléatoire λX suit une loi normale $\mathcal{N}(\lambda m, \lambda^2\sigma^2)$

Proposition 3.9 Si $X \sim \mathcal{N}(m, \sigma^2)$ alors la variable aléatoire

$$X^* = \frac{X - m}{\sigma}$$

s'appelle la variable aléatoire centrée (son espérance est nulle) réduite (sa variance est égale à 1) associée à X. L'espérance et la variance de $X \sim \mathcal{N}(m, \sigma^2)$ sont donc :

$$\mathbb{E}(X) = m; \quad \mathbb{V}\text{ar}\,(X) = \sigma^2$$

Théorème 3.1 (admis) *Soient $X_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ et $X_2 \sim \mathcal{N}(m_2, \sigma_2^2)$ deux variables aléatoires normales indépendantes alors*

$$X_1 + X_2 \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$$

⚐ *fin de la semaine 3*

▶ *Cours de la semaine 4*

Chapitre 4

Couples de variables aléatoires

🎞 *Vidéo 4.1 : Couples de variables aléatoires discrètes*
🎞 *Vidéo 4.2 : Couple de variables aléatoires discrètes et espérance*

4.1 Densité d'un couple de variables aléatoires.

Le fait de connaître la loi de deux variables aléatoires X et Y ne permet pas de calculer des probabilités du genre $P(X \in A, Y \in B)$ car les lois de X et de Y ne nous apprennent rien sur le lien qui existe entre X et Y.

Définition 4.1 On dit qu'un couple de variables aléatoires $X = (X_1, X_2)$ a pour densité $f : \mathbb{R}^2 \to \mathbb{R}^+$ lorsque pour tous réels a, b, c, d tels que $a \leqslant b$ et $c \leqslant d$, on a

$$P(a \leq X_1 \leq b, c \leq X_2 \leq d) = \int_a^b \left(\int_c^d f(t_1, t_2)\, \mathrm{d}t_2 \right) \mathrm{d}t_1$$

Proposition 4.1 f est une densité de X si et seulement si pour tout ensemble A du plan :

$$P(X \in A) = \iint_A f(t_1, t_2)\, \mathrm{d}t_1\, \mathrm{d}t_2$$

f est une densité de X si et seulement si pour tous réels t_1, t_2 :

$$P(X_1 \leqslant t_1; X_2 \leqslant t_2) = \int_{-\infty}^{t_2} \left(\int_{-\infty}^{t_1} f(x_1, x_2)\, \mathrm{d}x_1 \right) \mathrm{d}x_2$$

Proposition 4.2 Si le couple de variables aléatoires (X_1, X_2) possède f pour densité alors $\int_{-\infty}^{+\infty} f(t_1, t_2)\, \mathrm{d}t_1$ est une densité de X_2 et $\int_{-\infty}^{+\infty} f(t_1, t_2)\, \mathrm{d}t_2$ est une densité de X_1

Preuve : Il suffit de regarder la fonction de répartition de X_i.

Proposition 4.3 Si les variables aléatoires X_1 et X_2 sont indépendantes de densité f_{X_1} et f_{X_2} alors le couple de variables aléatoires (X_1, X_2) possède une densité définie par $f(t_1, t_2) = f_{X_1}(t_1) f_{X_2}(t_2)$.

Proposition 4.4 Si les variables aléatoires X_1 et X_2 sont indépendantes alors $\mathbb{V}\mathrm{ar}\,(X_1 + X_2) = \mathbb{V}\mathrm{ar}\,(X_1) + \mathbb{V}\mathrm{ar}\,(X_2)$, et $\mathrm{cov}(X_1; X_2) = 0$

Proposition 4.5 (admis) Si le couple de variables aléatoires (X_1, X_2) possède f pour densité alors pour toute fonction telle que l'intégrale est absolument convergente on a

$$\mathbb{E}(\phi(X_1, X_2)) = \iint_{\mathbb{R}^2} \phi(t_1, t_2) f(t_1, t_2)\, \mathrm{d}t_1\, \mathrm{d}t_2$$

4.2 Autres lois à densité

Définition 4.2 La somme des carrées de n variables aléatoires indépendantes de même loi $\mathcal{N}(0, 1)$, suit une loi dite de chi deux (χ^2) à n degrés de liberté (noté $Z = X_1^2 + X_2^2 + \cdots + X_n^2 \sim \chi^2(n)$).

Définition 4.3 Soient N une variable aléatoire de loi normale centrée réduite ($\mathcal{N}(0, 1)$) et χ une variable aléatoire de loi de χ^2 à n degrés de liberté telles que N et χ soient indépendantes alors $T = \frac{N}{\sqrt{\frac{\chi}{n}}}$ suit une loi de Student à n degrés de liberté (noté $T \sim \mathcal{S}(n)$).

Définition 4.4 Soient deux variables aléatoires indépendantes U_1 et U_2 de loi de chi deux à d_1 et d_2 degrés de libertés alors $S = \frac{\frac{U_1}{d_1}}{\frac{U_2}{d_2}}$ suit une loi de Fisher de paramètre (d_1, d_2) (noté $S \sim \mathcal{F}(d_1, d_2)$).

4.3 Propagation des erreurs

4.3.1 Avec une seule variable

Un problème classique en sciences expérimentales, on mesure une grandeur x avec une certaine incertitude, une certaine erreur, et on a une seconde grandeur y qui s'écrit comme une fonction de la grandeur x, par exemple $y = g(x)$ quelle erreur a-t-on sur y. En probabilité on modélise ces grandeurs par des variables aléatoires X et Y, on suppose que l'on connaît la loi de X, et on essaie de trouver des informations sur $Y = g(X)$, trouver la loi de Y peut s'avérer compliqué. Si l'on suppose que les valeurs prises par X sont resserrées autours de sa moyenne μ_x, on a $X = \mu_x + \varepsilon$, ou ε est une variable aléatoire centrée de variance faible, on peut alors effectuer un DL_1 de g en μ_x.

$$Y = g(X) = g(\mu_x + \varepsilon) = g(\mu_x) + g'(\mu_x)\varepsilon + o(\varepsilon) \approx g(\mu_x) + g'(\mu_x)\varepsilon$$

Si l'on prend l'espérance de ceci on obtient : $\mathbb{E}(Y) \approx \mathbb{E}(g(\mu_x)) + \mathbb{E}(g'(\mu_x)\varepsilon) = g(\mu_x) + g'(\mu_x)\mathbb{E}(\varepsilon) = g(\mu_x)$. Ce n'est pas une égalité en général mais une approximation correcte si les erreurs ε restent petites. Faisons de même avec la variance de Y :

$$\mathbb{V}\mathrm{ar}\,(Y) = \mathbb{E}\Big(\big(g(X) - \mathbb{E}(g(X))\big)^2\Big) \approx \mathbb{E}\Big(\big(g(\mu_x + \varepsilon) - g(\mu_x)\big)^2\Big) = \mathbb{E}\Big(\big(g(\mu_x) + g'(\mu_x)\varepsilon - g(\mu_x)\big)^2\Big)$$

Donc $\mathbb{V}\mathrm{ar}\,(Y) \approx \mathbb{E}\Big(\big(g'(\mu_x)\varepsilon\big)^2\Big) = g'(\mu_x)^2\mathbb{E}\big(\varepsilon^2\big) = g'(\mu_x)^2\mathbb{V}\mathrm{ar}\,(X)$.
Que l'on peut encore écrire $\sigma_Y \approx |g'(\mu_x)|\sigma_X$.
$|g'(\mu_x)|$ est un coefficient multiplicateurs des erreurs.

4.3.2 Avec deux variables ou plus

On mesure un vecteur $\overrightarrow{x} = (x_1; x_2)$ avec une certaine incertitude, une certaine erreur, et on a une seconde grandeur y qui s'écrit comme une fonction de la grandeur $\overrightarrow{x}$, par exemple $y = g(\overrightarrow{x})$ quelle erreur a-t-on sur y. En probabilité on modélise ces grandeurs par des variables aléatoires X_1, X_2 et Y, on suppose que l'on connaît les lois de X_1 et X_2, et on essaie de trouver des informations sur $Y = g(X_1, X_2)$, trouver la loi de Y peut s'avérer compliqué. Si l'on suppose que les valeurs prises par X_1 et X_2 sont resserrées autours de leur moyenne $\mu_{x_1} = \mathbb{E}(X_1), \mu_{x_2} = \mathbb{E}(X_2)$, on a $X_1 = \mu_{x_1} + \varepsilon_1$, où ε_1 est une variable aléatoire centrée de variance faible, de même $X_2 = \mu_{x_2} + \varepsilon_2$ on peut alors effectuer un DL_1 de g en (μ_{x_1}, μ_{x_2}).

$$Y = g(X_1, X_2) = g(\mu_{x_1} + \varepsilon_1, \mu_{x_2} + \varepsilon_2) \approx g(\mu_x) + \frac{\partial g}{\partial x_1}(\mu_{x_1}, \mu_{x_2})\varepsilon_1 + \frac{\partial g}{\partial x_2}(\mu_{x_1}, \mu_{x_2})\varepsilon_2$$

Si l'on prend l'espérance de ceci on obtient :
$\mathbb{E}(Y) \approx \mathbb{E}\big(g(\mu_{x_1}, \mu_{x_2})\big) + \mathbb{E}\big(\frac{\partial g}{\partial x_1}(\mu_{x_1}, \mu_{x_2})\varepsilon_1\big) + \mathbb{E}\big(\frac{\partial g}{\partial x_2}(\mu_{x_1}, \mu_{x_2})\varepsilon_2\big) = g(\mu_{x_1}, \mu_{x_2})$. Ce n'est pas une égalité en général mais une approximation correcte si les erreurs ε_i restent petites. Faisons de même avec la variance de Y :

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\,(Y) &= \mathbb{E}\Big(\big[g(X_1, X_2) - \mathbb{E}(g(X_1, X_2))\big]^2\Big) \\
&\approx \mathbb{E}\Big(\big[g(\mu_{x_1} + \varepsilon_1, \mu_{x_2} + \varepsilon_2) - g(\mu_{x_1}, \mu_{x_2})\big]^2\Big) \\
&\approx \mathbb{E}\Big(\big[g(\mu_{x_1}, \mu_{x_2}) + \frac{\partial g}{\partial x_1}(\mu_{x_1}, \mu_{x_2})\varepsilon_1 + \frac{\partial g}{\partial x_2}(\mu_{x_1}, \mu_{x_2})\varepsilon_2 - g(\mu_{x_1}, \mu_{x_2})\big]^2\Big) \\
&\approx \mathbb{E}\Big(\big[\frac{\partial g}{\partial x_1}(\mu_{x_1}, \mu_{x_2})\varepsilon_1 + \frac{\partial g}{\partial x_2}(\mu_{x_1}, \mu_{x_2})\varepsilon_2\big]^2\Big) \\
&\approx \Big(\frac{\partial g}{\partial x_1}(\mu_{x_1}, \mu_{x_2})\Big)^2\mathbb{E}(\varepsilon_1^2) + \Big(\frac{\partial g}{\partial x_2}(\mu_{x_1}, \mu_{x_2})\Big)^2\mathbb{E}(\varepsilon_2^2) + 2\Big(\frac{\partial g}{\partial x_1}(\mu_{x_1}, \mu_{x_2})\Big)\Big(\frac{\partial g}{\partial x_2}(\mu_{x_1}, \mu_{x_2})\Big)\mathbb{E}(\varepsilon_1\varepsilon_2) \\
&\approx \Big(\frac{\partial g}{\partial x_1}(\mu_{x_1}, \mu_{x_2})\Big)^2\mathbb{V}\mathrm{ar}\,(X_1) + \Big(\frac{\partial g}{\partial x_2}(\mu_{x_1}, \mu_{x_2})\Big)^2\mathbb{V}\mathrm{ar}\,(X_2) + 2\Big(\frac{\partial g}{\partial x_1}(\mu_{x_1}, \mu_{x_2})\Big)\Big(\frac{\partial g}{\partial x_2}(\mu_{x_1}, \mu_{x_2})\Big)\mathrm{covar}\,(X_1, X_2)
\end{aligned}$$

Dans la cas ou $\mathrm{covar}\,(X_1, X_2) = 0$, par exemple si X_1 et X_2 sont indépendantes on a alors $\mathbb{V}\mathrm{ar}\,(Y) = \Big(\frac{\partial g}{\partial x_1}(\mu_{x_1}, \mu_{x_2})\Big)^2\mathbb{V}\mathrm{ar}\,(X_1) + \Big(\frac{\partial g}{\partial x_2}(\mu_{x_1}, \mu_{x_2})\Big)^2\mathbb{V}\mathrm{ar}\,(X_2)$. et donc :

$$\sigma_Y = \sqrt{\Big(\frac{\partial g}{\partial x_1}(\mu_{x_1}, \mu_{x_2})\Big)^2\sigma_{X_1}^2 + \Big(\frac{\partial g}{\partial x_2}(\mu_{x_1}, \mu_{x_2})\Big)^2\sigma_{X_2}^2}$$

⚐ *fin de la semaine 4*

▶ *Cours de la semaine 5*

Chapitre 5

Convergences des suites de variables aléatoires

5.1 Généralités

On fait n mesures (ou n expériences) et on modélise chaque résultat à l'aide d'une variable aléatoire X_i, on peut supposer que ces variables aléatoires ont la même loi et sous certaines hypothèses qu'elles sont indépendantes. On note alors $S_n = \sum_{k=1}^n X_k$ et $\overline{X}_{[n]} = \frac{1}{n}S_n = \frac{1}{n}(X_1 + \ldots + X_n)$, ce sont deux nouvelles variables aléatoires. Pour simplifier on note souvent la moyenne empirique sans l'indice n, $\overline{X}$ pour $\overline{X}_{[n]}$ même si cette variable aléatoire dépend de n.

Définition 5.1 Soit $X_1, X_2, ..., X_n, ..$, une suite de variables aléatoires, on pose $\overline{X}_n$ la moyenne des n premières variables aléatoires, et on appelle moyenne empirique cette nouvelle variable aléatoire.

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + ... + X_n)$$

Proposition 5.1 Comme les X_i ont la même loi elles ont même espérance et même variance, de plus $\mathbb{E}(S_n) = n\mathbb{E}(X_1)$, $\mathbb{E}(\overline{X}) = \mathbb{E}(X_1)$ et si les X_i sont indépendantes $\mathbb{V}\text{ar}\,(\overline{X}_{[n]}) = \frac{1}{n}\mathbb{V}\text{ar}\,X_1$.

Exemple 5.1 Dans une population ou une proportion p vote pour A et $q = 1-p$ vote pour B, si on tire un échantillon ω de 1000 personnes, on modélise le fait que la ième personne vote pour A par $X_i(\omega) = 1$. Si l'échantillon est choisi de façon aléatoire avec remise alors les X_i sont indépendants de loi $\mathcal{B}(p)$, on a alors $\mathbb{V}\text{ar}\,\overline{X} = \frac{1}{n}pq$, plus n est grand plus $\overline{X}$ est "proche" de $\mathbb{E}(\overline{X}_{[n]}) = p$.

5.2 Loi des grands nombres

On suppose dans cette partie que toutes les variables aléatoires possèdent espérance et variance.

Théorème 5.1 *Inégalité de Bien Aymé Tchebichev*

$$P(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\mathbb{V}\text{ar}\,X}{\varepsilon^2}$$

Vidéo 5.1 : Inégalité de Bienaymé Tchebychev (BAT)

Théorème 5.2 *Loi faible des grands nombres :*
Soit (X_n) une suite de variables aléatoires indépendantes de même loi, alors :

$$\forall \varepsilon > 0, \ \lim_{n\to\infty} P(|\overline{X}_{[n]} - \mathbb{E}(X_1)| \geq \varepsilon) = 0$$

Vidéo 5.2 : loi des grands nombres

Définition 5.2 Il existe différents types de convergence pour les suites de variables aléatoires.
On dit que la suite de variables aléatoires (X_n) converge vers la variable aléatoire X

$$\begin{cases} \text{en probabilité lorsque} & \forall \varepsilon > 0 \ \lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0 \\ \text{presque sûrement lorsque} & P\big(\lim_{n\to\infty} X_n = X\big) = 1 \\ \text{en moyenne quadratique lorsque} & \lim_{n\to\infty} \mathbb{E}(|X_n - X|^2) = 0 \\ \text{en loi lorsque} & \lim_{n\to\infty} F_{X_n}(t_0) = F_X(t_0) \text{ en tout } t_0 \text{ où } F_Y \text{ est continue} \end{cases}$$

Théorème 5.3 (admis) *Loi forte des grands nombres :*
Soit (X_n) une suite de variables aléatoires indépendantes de même loi possédant une espérance, alors la suite des moyennes empiriques $(\overline{X_{[n]}})$ converge presque sûrement vers l'espérance des variables aléatoires X_i :

$$\overline{X_{[n]}} \xrightarrow{ps} \mathbb{E}(X_1)$$

5.3 Théorème de la limite centrée

Proposition 5.2 Soient $Y_1, Y_2, \ldots Y_n, \ldots$ et Y des variables aléatoires dont les fonctions de répartitions sont $F_1, F_2, \ldots F_n, \ldots$ et F_Y, la suite de variables aléatoire (Y_n) converge en loi vers la variable aléatoire Y si et seulement si pour tous points a, b tel que F_Y est continue en a et b on a :

$$\lim_{n\to\infty} P(a \leq Y_n \leq b) = P(a \leq Y \leq b)$$

Théorème 5.4 (admis) *dit de la limite centrée (TCL) :*
Soit (X_i) une suite de variables aléatoires indépendantes de même loi, ayant m pour espérance et σ pour écart type. La suite de variable aléatoire (Z_n) :

$$Z_n = \frac{X_1 + \ldots + X_n - nm}{\sqrt{n}\,\sigma} = \sqrt{n}\frac{\overline{X_n} - m}{\sigma} = \frac{\overline{X_n} - m}{\sqrt{\frac{\sigma^2}{n}}}$$

converge en loi vers une loi $\mathcal{N}(0;1)$.

Vidéo 5.3 : Théorème central limite

Remarque 5.1 Dans la pratique pour n assez grand nous supposerons que Z_n suit une loi $\mathcal{N}(0;1)$ alors qu'en réalité elle suit une loi proche d'une loi normale centrée réduite.
On remarque que $\mathbb{E}(\overline{X}) = m$ et $\mathbb{V}\text{ar}(\overline{X}) = \frac{\sigma^2}{n}$, la variable aléatoire Z_n est donc centrée et réduite, ce qui est extraordinaire c'est qu'elle suit une loi proche d'une loi normale.

Exemple 5.2 Dans un pays où 52% des gens vont voter pour A et 48% pour B, on choisit un échantillon de 900 personnes. Quelle est la probabilité que dans l'échantillon B soit vainqueur ?
On peut modéliser ceci de la fonction suivante on considère des variables aléatoires X_i qui suivent des lois de Bernoulli de paramètre .52. On interpréte les X_i de la façon suivante, dans un échantillon ω, si le ième personne interrogée vote pour A, alors $X_i(\omega) = 1$, si elle vote pour B alors $X_i(\omega) = 0$. $S_{900} = X_1 + \ldots + X_{900}$ représente donc le nombre de personnes qui votent pour A dans l'échantillon. Ce qui nous intéresse c'est la probabilité que le sondage donne B gagnant c'est à dire $P(S_{900} < 450)$. On peut utiliser la loi de S_{900} qui est une binomiale, mais ça demande beaucoup de calcul, on peut appliquer l'approximation centrale c'est à dire supposer que $Z_{900} = \frac{X_1+\ldots+X_{900}-900m}{\sqrt{900}\,\sigma}$ suit une loi normale centrée réduite. Or on sait que pour une loi de Bernoulli de paramètre p la variance vaut $p(1-p)$ on peut calculer $\sigma = \sqrt{p(1-p)}$, on a alors

$$P(S_{900} < 450) = P\Big(\frac{S_{900} - 900m}{\sqrt{900}\sigma} < \frac{450 - 900 \times 0,52}{30\sqrt{0.52 \times 0.48}}\Big) = P(Z_{900} < -1.20) \simeq 12\%$$

Il y a 12% de chance que l'institut donne B gagnant.

fin de la semaine 5

Cours de la semaine 6

Chapitre 6

Statistiques : échantillonnage et estimation

6.1 Introduction

Le but est d'étudier un caractère $\mathcal{C}$, d'une population au vu d'un échantillon de cette population. Une généralisation en quelque sorte de la méthode du sondage. On cherche un modèle probabiliste qui va décrire le phénomène, on suppose que $\mathcal{C}$ peut être représenté par une variable aléatoire on cherche alors à estimer sa loi, son espérance, ses paramètres..., pour cela il y a d'abord le choix de l'échantillon (échantillonnage), puis les calculs qui permettent d'estimer les paramètres.

Lors du choix du modèle on fait l'hypothèse à priori d'un type de lois (Exponentielle, normale, etc...) puis on cherche à estimer les paramètres de cette loi.

Vidéo 6.1 : Modèle probabiliste en statistiques

6.2 Statistique d'échantillonnage

Soit $\mathcal{P}$ notre population, choisissons un échantillon ω de n individus de cette population, et déterminons pour chaque individu le caractère $\mathcal{C}$, on note $X(\omega) = (x_1, \ldots, x_n)$, ou x_i est la valeur de $\mathcal{C}$ pour le ième individu de l'échantillon ω. On obtient donc un vecteur aléatoire :

$$\begin{array}{rccl} X & : & \mathcal{P}^n & \to & \mathbb{R}^n \\ & & \omega & \mapsto & X(\omega) = (X_1(\omega), X_2(\omega), \ldots, X_n(\omega)) \end{array}$$

Remarque 6.1 Si ω est tiré au hasard alors les X_i suivent toutes la même loi, si de plus ω est tirée avec remise, c'est à dire que l'on peut tirer plusieurs fois le même individu alors les X_i sont indépendantes, on parle d'échantillonnage non exhaustif.

Définition 6.1 On appelle estimateur (ou statistique d'échantillonnage) toute variable aléatoire de la forme $g(X_1, \dots, X_n)$ qui permet d'estimer un paramètre de la population, on appelle loi d'échantillonnage sa loi.

Définition 6.2 Soient $X_1, X_2, \dots, X_n$ des variables aléatoires de même loi indépendantes (iid : indépendantes, identiquement distribuées), et θ un paramètre de leur loi.

1. Un estimateur $K_n = g(X_1, \dots, X_n)$ du paramètre θ est sans biais si
$$\mathbb{E}(K_n) = \theta$$
2. Un estimateur K_n du paramètre θ est convergeant si la suite de variable aléatoire $(K_n)_n$ converge en probabilité vers θ c'est à dire :
$$\forall \varepsilon > 0, \lim_{n\to\infty} P(|K_n - \theta| > \varepsilon) = 0$$
3. K_n et H_n deux estimateurs du paramètre θ ; K_n est plus efficace que H_n si
$$\mathbb{E}\big((K_n - \theta)^2\big) \leq \mathbb{E}\big((H_n - \theta)^2\big)$$

6.2.1 Moyenne

Vidéo 6.2 : Moyenne empirique

Remarque 6.2 La moyenne de l'échantillon $\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$ est une statistique d'échantillonnage qui permet d'estimer la moyenne de la population.
$\mathbb{E}(\overline{X}) = \mathbb{E}(X_1)$, l'espérance de $\overline{X}$ est celle de l'ensemble de la population.
$\mathbb{V}\text{ar}\,\overline{X} = \frac{1}{n}\mathbb{V}\text{ar}\,X_1$ la variance de $\overline{X}$ est d'autant plus petite que n est grand, l'écart type varie inverse-proportionnellement à la racine carrée de n.

Remarque 6.3 La loi faible des grands nombres nous apprend donc que $\overline{X}$ est un estimateur convergeant de la moyenne, et on vient de voir que $\overline{X}$ est un estimateur sans biais de la moyenne.

Proposition 6.1 Si les X_i sont indépendantes de même loi normale (m, σ^2) alors $\overline{X}$ suit une loi normale de paramètre $\mathcal{N}(m, \frac{\sigma^2}{n})$.

6.2.2 Variance

De même il est naturel de déterminer un estimateur sans biais de la variance, et comme pour la moyenne on peut essayer avec la variance de l'échantillon. On pose alors

$$S^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X})^2$$

Notons m l'espérance de X_1 et σ son écart type, et déterminons l'espérance de S^2.

$$\begin{aligned}
\sum_{k=1}^{n}(X_k - \overline{X})^2 &= \sum_{k=1}^{n}(X_k - m + m - \overline{X})^2 && (6.1)\\
&= \sum_{k=1}^{n}\big((X_k - m)^2 + (m - \overline{X})^2 + 2(X_k - m)(m - \overline{X})\big) && (6.2)\\
&= \sum_{k=1}^{n}(X_k - m)^2 + \sum_{k}(m - \overline{X})^2 + 2(m - \overline{X})\sum_{k}(X_k - m) && (6.3)\\
&= \sum_{k=1}^{n}(X_k - m)^2 + n(m - \overline{X})^2 + 2(m - \overline{X})n(\overline{X} - m) && (6.4)\\
&= \sum_{k=1}^{n}(X_k - m)^2 - n(m - \overline{X})^2 && (6.5)\\
\mathbb{E}\left(\sum_{k=1}^{n}(X_k - \overline{X})^2\right) &= \mathbb{E}\left(\sum_{k=1}^{n}(X_k - m)^2 - n(m - \overline{X})^2\right) && (6.6)\\
&= \sum_{k=1}^{n}\mathbb{E}\left((X_k - m)^2\right) - n\mathbb{E}\left((m - \overline{X})^2\right) && (6.7)\\
&= \sum_{k=1}^{n}\sigma^2 - n\frac{1}{n}\sigma^2 && (6.8)\\
&= (n-1)\sigma^2 && (6.9)
\end{aligned}$$

On remarque donc que S^2 est un estimateur biaisé de la variance car $\mathbb{E}(S^2) = \frac{n-1}{n}\sigma^2$. on définit alors un estimateur non biaisé de la variance $\widehat{S}^2$ par :

$$\widehat{S}^2 = \frac{n}{n-1}S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2$$

Théorème 6.1 (admis) *Si les X_i sont indépendantes de même loi normale $\mathcal{N}(m, \sigma^2)$ alors*

1. *$\widehat{S}^2$ et $\overline{X}$ sont indépendantes.*
2. *$\frac{n-1}{\sigma^2}\widehat{S}^2$ suit une loi de chi deux à $n-1$ degrés de libertés, on note $\frac{n-1}{\sigma^2}\widehat{S}^2 \sim \chi^2(n-1)$.*
3. *La variable aléatoire Z_n suit une loi de Student à $n-1$ degrés de liberté, on note $Z_n \sim \mathcal{S}(n-1)$ avec*

$$Z_n = \sqrt{n}\frac{\overline{X} - m}{\widehat{S}}$$

Remarque 6.4 On remarque que $\frac{n-1}{\sigma^2}\widehat{S}^2 = \sum_{k=1}^{n}\left(\frac{X_k - \overline{X}}{\sigma}\right)^2$
Si les X_i sont indépendantes de même loi, alors pour n grand Z_n suit une loi proche d'une $\mathcal{N}(0;1)$.

6.3 Intervalle de confiance

Définition 6.3 On appelle intervalle de confiance d'un paramètre θ, au seuil α, un intervalle $[A;B]$, ou A et B sont des variables aléatoires telles que

$$P(m \in [A,B]) = \alpha \quad \Big(\text{ou plus généralement } P(m \in [A,B]) \geqslant \alpha\Big)$$

Remarque 6.5 On parle aussi d'un intervalle de confiance au risque $1-\alpha$.
Il est important de remarquer que contrairement à l'intuition ce n'est pas m qui est une variable aléatoire, mais A et B les bornes de l'intervalle de confiance. La probabilité a un sens avant l'expérience, après l'expérience on a un intervalle $I(\omega)$, si l'expérience fait partie des 95% (ou α) les plus "moyennes" notre paramètre appartient à I, si notre expérience est "particulière" notre paramètre ne lui appartient pas, il n'y a plus à proprement parlé de probabilité.
La notion plus simple d'intervalle de fluctuation donne un intervalle fixe I telle que la probabilité qu'une variable aléatoire appartienne à cet intervalle soit égal à α.

Exemple 6.1 On suppose que la mesure de la taille d'une pièce usinée suit une loi normale, on veut estimer la taille moyenne m à l'aide d'un intervalle de confiance, on obtient 13 résultats, on calcule alors leur moyenne $\overline{X}(\omega) = 12$ et leur variance $S^2(\omega) = 0,5$, on sait d'après le théorème 6.1 que $Z = \sqrt{13}\frac{\overline{X}-m}{\widehat{S}}$ suit une loi de Student à 12 degrés de liberté donc d'après la table

$$\begin{aligned} P\left(\left|\sqrt{13}\frac{\overline{X}-m}{\widehat{S}}\right| < 2,18\right) &= 95\% \\ P\left(\overline{X} - \frac{2,18}{\sqrt{13}}\widehat{S} < m < \overline{X} + \frac{2,18}{\sqrt{13}}\widehat{S}\right) &= 95\% \end{aligned}$$

Donc $I = [\overline{X} - \frac{2,18}{\sqrt{13}}\widehat{S}; \overline{X} + \frac{2,18}{\sqrt{13}}\widehat{S}]$ est un intervalle de confiance de m au niveau de 95 %. avec nos valeurs numériques on obtient $I(\omega) = [12 - \frac{2,18}{\sqrt{12}}0,5; 12 + \frac{2,18}{\sqrt{12}}0,5] = [11,68; 12,32]$.

Remarque 6.6 Intervalle de confiance d'une proportion : On prélève un échantillon ω de taille n d'une population dont une proportion p d'individus possèdent un caractère qualitatif. On modélise par X_i une variable de Bernoulli le fait que le ième individu de l'échantillon ait le caractère $(X_i(\omega) = 1)$ ou ne l'ait pas $(X_i(\omega) = 0)$, la proportion du caractère dans l'échantillon est égale $\overline{X}(\omega)$. L'approximation centrale permet d'affirmer que

$$Z_1 = \frac{\overline{X} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

suit une loi proche d'une $\mathcal{N}(0;1)$, on pourra appliquer cette approximation des que $n\overline{X}(\omega) \geq 5$ et $n\Big(1 - \overline{X}(\omega)\Big) \geq 5$, on sait d'après la loi des grand nombre que $\overline{X}$ converge en probabilité vers p, dans la pratique on approxime souvent p par $\overline{X}$, dans la formule précédente ce qui simplifie les calculs :

$$Z_2 = \frac{\overline{X} - p}{\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}}$$

suit une loi proche d'une $\mathcal{N}(0;1)$

6.4 Estimateur du maximum de vraisemblance

Un des problème important de la statistique est de trouver de bons estimateurs, on en a donné un pour la moyenne et un pour la variance de façon empirique mais comment trouver une méthode générale. Nous donnons ici l'idée générale d'une méthode très efficace. On donne juste l'exemple sur des variable aléatoires à valeurs dans $\mathbb{N}$. On suppose que les X_i suivent une loi entièrement déterminée par un paramètre θ inconnu, on fait le tirage de notre échantillon et on obtient des valeurs $x_1, x_2, \ldots, x_n$, la probabilité d'obtenir ce résultat est donné par la vraisemblance

$$L(x_1, x_2, \ldots, x_n, \theta) = P_\theta(X_1 = x_1, \ldots, X_n = x_n) = P_\theta(X_1 = x_1)P_\theta(X_2 = x_2)\ldots P_\theta(X_n = x_n)$$

car les X_i sont indépendants, on peut se dire que la probabilité d'obtenir ce que l'on a obtenu est assez importante et même puisqu'on l'a obtenu que c'était ce qu'il y a de plus probable, on va donc estimer θ à l'aide de $\widehat{\theta}$ qui maximise la fonction L, on va donc obtenir un $\widehat{\theta} = f(x_1, \ldots, x_n)$ telle que $\forall\theta : L(x_1, x_2, \ldots, x_n, \widehat{\theta}) \geqslant L(x_1, x_2, \ldots, x_n, \theta)$

Exemple 6.2 Supposons que les X_i suivent une loi de Bernoulli de paramètre p, alors

$$L(x_1, x_2, \ldots, x_n, p) = p^{(\sum_{i=1}^n x_i)}(1-p)^{(n-\sum_{i=1}^n x_i)}$$

pour obtenir le p qui maximise cette quantité, étudions la fonction

$$\begin{array}{rcl} f(p) & = & p^u(1-p)^{n-u} \\ f'(p) & = & up^{u-1}(1-p)^{n-u} - (n-u)p^u(1-p)^{n-u-1} \\ & = & \big(u(1-p) - (n-u)p\big)p^{u-1}(1-p)^{n-u-1} \\ & = & (u-np)p^{u-1}(1-p)^{n-u-1} \end{array}$$

La fonction f est donc croissante entre 0 et $\frac{u}{n}$ puis décroissante entre $\frac{u}{n}$ et 1. On en déduit donc que le maximum est atteint en $\frac{u}{n}$. Le maximum de vraisemblance est donc atteint en $\frac{1}{n}\sum_{i=1}^n x_i$, on obtient l'estimateur $\widehat{p} = \frac{1}{n}\sum_{i=1}^n X_i$, on retrouve bien $\overline{X}$.

On peut utiliser la même méthode pour les variables aléatoires à densité on remplace juste $P(X_i = x_i)$ par $f_{X_i}(x_i)$.

⚐ *fin de la semaine 6*

▶ *Cours de la semaine 7*

Chapitre 7

Tests statistiques

7.1 Introduction

L'objet des tests statistiques dit tests d'aide à la décision, est de prendre une décision pour l'ensemble d'une population à la vue d'un échantillon de cette population. Dans ces tests il y a toujours une dissymétrie qu'il est impératif de comprendre.

Exemple 7.1 Un juge doit juger un homme qu'on accuse d'utiliser une pièce truquée, pour laquelle la face pile apparaît moins souvent que la face face. On voit bien ici la dissymétrie : si la pièce est truquée il vaut mieux condamner le joueur, si la pièce n'est pas truquée il ne faut pas condamner le joueur :

	Pièce truquée	Pièce non truquée
Condamnation du joueur	Bien	Très mauvais
Acquittement du joueur	Mauvais	Bien

🎞 *Vidéo 7.1 : Objectif d'un test statistique*

7.2 Formalisme d'un test statistique

Dans le modèle que l'on se fixe au départ, il y a deux possibilités, l'une des deux est vrai, l'autre est fausse, mais on ne sait pas laquelle est vraie. On note H_0 celle que nous accepterons en cas de doute et H_1 l'autre, le test va être construit de telle sorte que nous refuserons H_0 (c'est à dire nous accepterons H_1) seulement si H_0 est très vraisemblablement faux.

Il y a quatre possibilités pour l'issu du test :

	H_0 est vrai	H_1 est vrai
On conclut H_0	Bon	Erreur de deuxième espèce
On conclut H_1	Erreur de première espèce	Bon

Exemple 7.2 Si l'on reprend notre exemple, on va choisir pour H_0 la pièce n'est pas truquée, et pour H_1 la pièce est truquée. Le juge lance 1000 fois la pièce si il obtient 516 piles il ne va pas condamner le joueur,si il obtient 487 piles non plus, si il obtient 137 piles il va le condamner, mais ou est la limite , à partir de quel seuil le résultat ne peut-il plus venir d'une pièce normale.

Définition 7.1 On appelle risque (d'erreur) d'un test le réel α tel que si H_0 est vrai la probabilité de conclure H_1 est égale (ou inférieure) à α. On appelle seuil de confiance du test la grandeur $1-\alpha$. α correspond lorsque H_0 est vrai à la probabilité de faire une erreur de première espèce.

Exemple 7.3 Le juge décide de faire un test au risque 0,2%. Le juge va lancer 1000 fois le dé, il note X_i la variable aléatoire qui vaut 1 si le ième lancé donne pile et 0 si le ième lancé donne face.

$$S = \sum_{k=1}^{1000} X_k = X_1 + X_2 + \ldots + X_{1000}$$

S est une variable aléatoire qui correspond donc au nombre de pile obtenu lors des 1000 lancés. On décide de choisir H_1 si $S(\omega)$ est inférieur à un certain seuil ζ, c'est à dire si le nombre de piles obtenu est 'trop' faible.

$$\begin{cases} \text{Si } S(\omega) < \zeta \text{ alors on choisit } H_1 \\ \text{Si } S(\omega) > \zeta \text{ alors on choisit } H_0 \end{cases}$$

Pour construire le test, il faut donc déterminer ζ tel que si H_0 est vrai (on dit souvent sous H_0) alors $P(S < \zeta) = 0,2\%$. Supposons la pièce non truquée, les X_i sont indépendants de même loi de Bernoulli de paramètre $\frac{1}{2}$. On va utiliser l'approximation centrale

$$Z = \frac{S - 500}{\sqrt{1000}\sqrt{\frac{1}{2}\frac{1}{2}}}$$

suit une loi proche d'une $\mathcal{N}(0;1)$. Donc $P(S < \zeta) = P(Z < 2\frac{\zeta - 500}{\sqrt{1000}}) = 0,002$ donc d'après la table de la loi normale on a

$$2\frac{\zeta - 500}{\sqrt{1000}} = -2,88 \text{ donc } \zeta = 500 - 1,44\sqrt{1000} \simeq 454,46$$

finalement le juge lance 1000 fois la pièce, si il obtient plus de 454 piles il acquitte le joueur si il obtient moins de 454 piles il condamne le joueur.

Vidéo 7.2 : Formalisme d'un test statistique

Déroulement d'un test

1. Choix des 2 hypothèses : H_0 et H_1.
2. Choix du risque du test.
3. Choix de la forme du test.
4. Construction du test : calcul du seuil.
5. Expérimentation et confrontation des résultats avec le seuil.
6. Décision : H_0 ou H_1.

Application à l'exemple précédent

1. Choix des 2 hypothèses : H_0 et H_1.
   H_0 la pièce n'est pas truquée, H_1 la pièce est truquée
2. Choix du risque du test.
   0,2%
3. Choix de la forme du test.
   On choisit H_1 si $S(\omega)$ est inférieur à un certain seuil ζ

4. Construction du test : calcul du seuil.
   Calcul permettant de trouver $\zeta \simeq 454,46$.
5. Expérimentation et confrontation des résultats avec le seuil.
   $S(\omega) = 432$ *suite à l'expérience de lancé.*
6. Décision : H_0 ou H_1.
   Comme 432<454 on conclut H_1.

Définition 7.2 Lors d'un test on peut calculer la p-valeur (ou p-value) cela correspond à la probabilité sous H_0 d'avoir un résultat plus extrême que celui observé. Cette quantité est souvent calculée par les programmes de statistiques.

Vidéo 7.3 : Exemple d'un test statistique avec une loi normale

fin de la semaine 7

Cours de la semaine 8

7.3 Différents tests statistiques

7.3.1 Comparaison d'une moyenne à une valeur donnée

Les X_i sont des variables aléatoires indépendantes de même loi d'espérance μ, $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ et $\widehat{S}^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$, si on pose

$$Z = \sqrt{n}\frac{\overline{X} - \mu}{\widehat{S}}$$

1. Si les X_i suivent une loi normale alors Z suit une loi de Student à $n-1$ degrés de liberté.
2. Si les X_i suivent une loi proche d'une loi normale alors Z suit une loi proche d'une loi de Student à $n-1$ degrés de liberté.
3. Si n est grand Z suit une loi proche d'une loi normale centrée réduite.

7.3.2 Comparaison d'une proportion à une valeur donnée

On utilise pour n assez grand la partie précédente avec les X_i des variables aléatoires de Bernoulli.

7.3.3 Comparaison de deux moyennes

On a deux populations normalement distribuées de même moyenne et de même écart type. Pour la première population on a un échantillon de n_1 individus de moyenne $\overline{X}^1$ et de variance estimée ($\widehat{S}_1^2 = \frac{1}{n_1-1}\sum_{i=1}^n (X_i - \overline{X}^1)^2$), pour la seconde population on a un échantillon de n_2 individus de moyenne $\overline{X}^2$ et de variance estimée ($\widehat{S}_2^2 = \frac{1}{n_2-1}\sum_{i=1}^n (X_i - \overline{X}^2)^2$) alors Z suit une loi de Student à $n_1 + n_2 - 2$ degrés de liberté avec :

$$Z = \frac{\overline{X}^1 - \overline{X}^2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\sqrt{\frac{(n_1-1)\widehat{S}_1^2 + (n_2-1)\widehat{S}_2^2}{n_1+n_2-2}}}$$

7.3.4 Comparaison de deux proportions

Voir le test du chi carrée.

7.3.5 Test du chi carrée d'ajustement

Il s'agit d'un test assez différent des exemples précédents, il permet de comparer des résultats expérimentaux à une distribution donnée, par exemple on mesure la résistance à la compression d'un béton et on se demande si les résultats suivent une loi normale. Pour cela on répartit les n résultats en différentes classes : $A_1, \ldots, A_k$, sous l'hypothèse H_0 que l'on teste, la probabilité qu'un résultat se trouve dans la classe A_i est p_i. L'espérance du nombre d'individus qui se trouve dans la classe A_i est donc np_i, on parle souvent du nombre d'observations espérées. On va comparer ce nombre np_i au nombre de résultats expérimentaux n_i se trouvant dans la classe A_i. On peut représenter ceci sous la forme d'un tableau :

	A_1	A_2	...	A_k
résultat espéré sous H_0	np_1	np_2	...	np_k
résultat observé	n_1	n_2	...	n_k

Proposition 7.1 (admis) Sous l'hypothèse H_0 et si le nombre d'individus espérés par classe (np_i) est supérieur à 5, alors la variable aléatoire Z suit une loi de probabilité proche d'un χ^2 à $k-1$ degrés de liberté avec :

$$Z = \sum_{i=1}^{k} \frac{(np_i - n_i)^2}{np_i}$$

Sous l'hypothèse H_0 si le calcul des p_i nécessite l'estimation de t paramètres, alors la loi de Z est proche d'une loi de χ^2 à $(k-1-t)$ degrés de liberté.

Vidéo 8.1 : Test du Khi deux d'ajustement à une loi

7.3.6 Test du chi carrée d'homogénéité ou d'indépendance

Deux utilisations sont possibles pour ce test :

1. On considère deux caractères différents sur une même population, et on teste le fait que ces caractères sont indépendants.
2. On a deux populations différentes et on teste leur homogénéité pour un certain caractère.

Notons $A_1, A_2, \ldots, A_k$ les classes du premier caractère et $B_1, B_2, \ldots, B_l$ les classes du second caractère. On présente les résultats expérimentaux à l'aide d'un tableau de contingence :

	A_1	A_2	$\ldots$	A_k	total
B_1	$n_{(1;1)}$	$n_{(1;2)}$	$\ldots$	$n_{(1;k)}$	$\sum_j n_{(1;j)}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
B_l	$n_{(l;1)}$	$n_{(l;2)}$	$\ldots$	$n_{(l;k)}$	$\sum_j n_{(l;j)}$
total	$\sum_i n_{(i;1)}$	$\sum_i n_{(i;2)}$	$\ldots$	$\sum_i n_{(i;k)}$	n

Les $n_{(i,j)}$ représentent le nombre d'individus de l'échantillon se trouvant dans la classe A_i pour le premier caractère et dans la classe B_j pour le second, sous H_0 c'est à dire si les caractères sont indépendants, la probabilité d'appartenir à l'intersection des deux classes est égale au produit des probabilités d'appartenir à chacune des classes, or pour estimer la probabilité d'appartenir à A_i il est naturel de considérer $\frac{1}{n}\sum_i n_{(i;j)}$, et pour estimer la probabilité d'appartenir à B_j, $\frac{1}{n}\sum_i n_{i;j}$, ce qui nous donne comme probabilité $p_{i;j}$ de se trouver dans la classe A_i pour le premier caractère et dans la classe B_j pour le second :

$$p_{(i;j)} = \left(\frac{1}{n}\sum_j n_{j;i}\right)\left(\frac{1}{n}\sum_i n_{j;i}\right) = \frac{(\sum_j n_{j;i})(\sum_i n_{j;i})}{n^2}$$

Le nombre d'individus espérés dans $A_i \cap B_j$:

$$np_{(i;j)} = n\left(\frac{1}{n}\sum_j n_{j;i}\right)\left(\frac{1}{n}\sum_i n_{j;i}\right) = \frac{(\sum_j n_{j;i})(\sum_i n_{j;i})}{n}$$

Proposition 7.2 (admis) Sous l'hypothèse H_0 et si le nombre d'individus espérés par case $(np_{(i;j)})$ est supérieur à 5, alors la variable aléatoire Z suit une loi de probabilité proche d'un χ^2 à $(k-1)(l-1)$ degrés de liberté avec :

$$Z = \sum_{\substack{1 \le i \le l \\ 1 \le j \le k}} \frac{(np_{(i;j)} - n_{(i;j)})^2}{np_{(i;j)}}$$

7.4 Formalisme d'un test statistique

Lorsque l'on effectue un test statistique, dans un exercice par exemple, on précise toujours :

1. Le type de test
2. Le niveau du test
3. Les hypothèses H_0 et H_1
4. La forme du test
5. Les hypothèses sur les données (taille de l'échantillon, loi normale, homoscédasticité,...)
6. La statistique du test, puis sa loi sous H_0

7. Enfin le calcul correspondant
8. Une conclusion

⚐ *fin de la semaine 8*

▶ *Cours de la semaine 9*

Chapitre 8

Régression linéaire

8.1 Introduction

L'objet de la régression linéaire est de modéliser une variable Y à expliquer par une fonction affine d'une autre variable x dite explicative : dans la pratique on est en possession de couple de valeurs $(x_i; y_i)$ correspondant à un même individu et l'on cherche une relation du genre $y_i \approx \beta_1 x_i + \beta_0$, mais un tel couple (β_0, β_1) n'existe que si les points $(x_i; y_i)$ sont alignés ce qui n'est en général pas le cas. On suppose alors que $Y_i = \beta_1 x_i + \beta_0 + \varepsilon_i$, ce ε_i correspondant à une variation par rapport à une relation idéal. Nous supposerons que

1. les x_i sont des points déterministes
2. les Y_i sont des variables aléatoires
3. les ε_i sont des variables aléatoires indépendantes et suivent toutes une même loi normale centrée ($\mathbb{E}(\varepsilon_i) = 0$) nous supposons de plus qu'elles ont toutes la même variance σ^2(homoscédasticité).

β_0, β_1 et σ^2 sont des constantes que l'on ne connaît pas et que l'on cherche à estimer.
On a donc $\mathbb{E}(Y_i) = \beta_1 x_i + \beta_0$ et $\mathbb{V}\text{ar}(Y_i) = \sigma^2$.

Exemple 8.1 Par exemple Y pourrait représenter la longueur d'une poutre d'acier soumise à une certaine traction x. Dans la phase élastique l'élongation est proportionnelle à la traction, toutefois expérimentalement les points ne sont pas complètement aligné : erreur de mesure, inhomogénéité de la barre, etc ... A partir de quelques points on cherche à estimer la relation entre la traction et l'allongement de la poutre.

🎞 *Vidéo : 9.1 : Modèle de régression linéaire*

8.2 Moindre carrés

Se pose la question de l'estimation de β_0 et β_1. On peut par exemple maximiser la vraisemblance, ici $V(b_0, b_1) = \prod_i f(y_i - (b_0 + b_1 x_i)) = \frac{1}{\sqrt{2\pi}^n} \exp\left[-\frac{1}{2\sigma^2} \left(\sum_i \big(y_i - (b_0 + b_1 x_i)\big)^2 \right) \right]$ qui est maximum ssi $\sum_i \big(y_i - (b_0 + b_1 x_i)\big)^2 \Big)$ est minimum, cela s'interprète comme la somme des carrés des écarts verticaux des points (x_i, y_i) à la droite d'équation $y = b_0 + b_1 x$. Il s'agit de minimiser la fonction des deux variables b_0 et b_1

$$\varphi(b_0; b_1) = \sum_i \big(y_i - (b_0 + b_1 x_i)\big)^2$$

En calculant les dérivées partielles on obtient $\begin{cases} \frac{\partial \varphi}{\partial b_0}(b_0; b_1) = -2 \sum_i \big(y_i - (b_0 + b_1 x_i)\big) \\ \frac{\partial \varphi}{\partial b_1}(b_0; b_1) = -2 \sum_i x_i \big(y_i - (b_0 + b_1 x_i)\big) \end{cases}$ En un minimum les dérivées partielles s'annulent, il s'agit donc de résoudre un système de deux équations à deux inconnues.

$$\begin{cases} \overline{y} - b_0 - b_1 \overline{x} = 0 \\ \overline{xy} - (b_0 \overline{x} + b_1 \overline{x^2}) = 0 \end{cases}$$

ce qui s'écrit encore, avec les notations suivantes

$$\overline{x} = \frac{1}{n} \sum_{i=1}^n x_i; \quad \overline{y} = \frac{1}{n} \sum_{i=1}^n y_i; \quad \overline{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i; \quad \overline{x^2} = \frac{1}{n} \sum_{i=1}^n x_i^2; \quad \text{et} \quad \sigma_x^2 = \overline{x^2} - \overline{x}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \overline{x})^2$$

$$\begin{pmatrix} 1 & \overline{x} \\ \overline{x} & \overline{x^2} \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \begin{pmatrix} \overline{y} \\ \overline{xy} \end{pmatrix}$$

Ce qui donne

$$\begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \begin{pmatrix} 1 & \overline{x} \\ \overline{x} & \overline{x^2} \end{pmatrix}^{-1} \begin{pmatrix} \overline{y} \\ \overline{xy} \end{pmatrix} = \frac{1}{\overline{x^2} - \overline{x}^2} \begin{pmatrix} \overline{x^2} & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix} \begin{pmatrix} \overline{y} \\ \overline{xy} \end{pmatrix} = \frac{1}{\sigma_x^2} \begin{pmatrix} \overline{x^2}\overline{y} - \overline{x}\ \overline{xy} \\ -\overline{x}\ \overline{y} + \overline{xy} \end{pmatrix}$$

Ce qui avec des notations usuelles donne $\begin{cases} b_1 = \frac{\text{covar}\,(x,y)}{\mathbb{V}\text{ar}\,(x)} = \frac{\overline{xy} - \overline{x}\ \overline{y}}{\sigma_x^2} \\ b_0 = \overline{y} - b_1\overline{x} \end{cases}$

Pour étudier la nature de ce point critique on peut calculer la hessienne de φ, on a

$$\begin{cases} \frac{\partial^2 \varphi}{\partial b_0^2}(b_0; b_1) = -2\sum_i -1 = 2n \\ \frac{\partial^2 \varphi}{\partial b_1^2}(b_0; b_1) = -2\sum_i (-x_i^2) = 2n\overline{x^2} \\ \frac{\partial^2 \varphi}{\partial b_0 \partial b_1}(b_0; b_1) = -2\sum_i (-x_i) = 2n\overline{x} \end{cases}$$

La hessienne est donc égale à $\begin{pmatrix} 2n & 2n\overline{x} \\ 2n\overline{x} & 2n\overline{x^2} \end{pmatrix} = 2n \begin{pmatrix} 1 & \overline{x} \\ \overline{x} & \overline{x^2} \end{pmatrix}$ Cette matrice possède deux valeurs propres dont la somme vaut $1 + \overline{x^2}$ et le produit $\overline{x^2}$, les deux valeurs propres sont donc positives et φ possède en son unique point critique un minimum local strict. Ce minimum est en fait global.

La droite qui approxime "le mieux" le nuage de points $(x_i; y_i)_{i \leq n}$ est donc la droite d'équation $y = b_0 + b_1 x$ appelée droite de régression linéaire avec $\begin{cases} b_1 = \frac{\text{covar}\,(x,y)}{\mathbb{V}\text{ar}\,(x)} = \frac{\overline{xy} - \overline{x}\ \overline{y}}{\sigma_x^2} \\ b_0 = \overline{y} - b_1\overline{x} \end{cases}$

Vidéo 9.2 : Calcul des coefficients de régression linéaire

8.3 Estimateurs des paramètres

Pour étudier la qualité de notre modèle linéaire, on va étudier d'un point de vue probabiliste les quantités b_0 et b_1 que l'on vient de déterminer, pour cela on définit les estimateurs $\hat{\beta}_0$ et $\hat{\beta}_1$:

Définition 8.1 $\begin{cases} \hat{\beta}_1 = \frac{\sum_i^n x_i Y_i - \sum_i^n x_i \sum_i^n Y_i}{\sigma_x^2} = \frac{\overline{xY} - \overline{x}\,\overline{Y}}{\sigma_x^2} \\ \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{x} \end{cases}$

A chaque point (x_i, y_i) correspond un point sur la droite de régression de coordonnées $(x_i, \hat{y}_i)$ avec $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, qui est un estimateur de $\mathbb{E}(Y_i)$.

En termes de variables aléatoires les x_i sont des constantes et les Y_i des variables aléatoires, les estimateurs $\hat{\beta}_0$ et $\hat{\beta}_1$ sont des variables aléatoires, on peut déterminer leur espérance et leur variance ainsi que leur loi :

Proposition 8.1 Les estimateurs $\hat{\beta}_0$ et $\hat{\beta}_1$ de β_0 et β_1 sont sans biais.

Preuve :
$\begin{cases} \hat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^n x_i Y_i - \overline{x}\overline{Y_i}}{\overline{x^2} - \overline{x}^2} \\ \hat{\beta}_0 = \frac{1}{n}\sum_{i=1}^n Y_i - \hat{\beta}_1\overline{x} \end{cases}$ donc $\begin{cases} \mathbb{E}(\hat{\beta}_1) = \frac{\frac{1}{n}\sum_{i=1}^n x_i \mathbb{E}(Y_i) - \overline{x}\mathbb{E}(Y_i)}{\overline{x^2} - \overline{x}^2} \\ \mathbb{E}(\hat{\beta}_0) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(Y_i) - \mathbb{E}(\hat{\beta}_1)\overline{x} \end{cases}$

donc $\begin{cases} \mathbb{E}(\hat{\beta}_1) = \frac{\frac{1}{n}\sum_{i=1}^n x_i(\beta_1 x_i + \beta_0) - \overline{x}(\beta_1 x_i + \beta_0)}{\overline{x^2} - \overline{x}^2} = \frac{(\beta_1\overline{x^2} + \beta_0\overline{x}) - \overline{x}(\beta_1\overline{x} + \beta_0)}{\overline{x^2} - \overline{x}^2} = \beta_1 \\ \mathbb{E}(\hat{\beta}_0) = \frac{1}{n}\sum_{i=1}^n (\beta_1 x_i + \beta_0) - \mathbb{E}(\hat{\beta}_1)\overline{x} = \beta_1\overline{x} + \beta_0 - \beta_1\overline{x} = \beta_0 \end{cases}$

Proposition 8.2 Les estimateurs $\hat{\beta}_0$ et $\hat{\beta}_1$ suivent des lois normales .
$\mathbb{E}(\hat{\beta}_0) = \beta_0 \quad \mathbb{E}(\hat{\beta}_1) = \beta_1$
$\mathbb{V}\text{ar}\,(\hat{\beta}_0) = \frac{\sigma^2\overline{x^2}}{n\sigma_x^2} \quad \mathbb{V}\text{ar}\,(\hat{\beta}_1) = \frac{\sigma^2}{n\sigma_x^2} \quad \mathcal{C}\text{ov}\,(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\overline{x}\sigma^2}{n\sigma_x^2}$ en notant $\sigma_x^2 = \overline{x^2} - \overline{x}^2$

On peut montrer que parmi tous les estimateurs sans biais de β_0 et β_1, $\hat{\beta}_0$ et $\hat{\beta}_1$ sont de variance minimale.
En outre, $\hat{\beta}_0$ converge en probabilité vers β_0 et $\hat{\beta}_1$ converge en probabilité vers β_1

Preuve : $\hat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^n x_i Y_i - \overline{x}\overline{Y_i}}{\overline{x^2} - \overline{x}^2} = \sum_{i=1}^n \frac{(x_i - \overline{x})Y_i}{n(\overline{x^2} - \overline{x}^2)}$ Donc β_1 est la somme de variables aléatoires indépendantes de lois normales car $Y_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$ donc

$$\frac{(x_i - \overline{x})Y_i}{n(\overline{x^2} - \overline{x}^2)} \sim \mathcal{N}\left(\frac{(x_i - \overline{x})(\beta_0 + \beta_1 x_i)}{n(\overline{x^2} - \overline{x}^2)}; \left(\frac{(x_i - \overline{x})\sigma}{n(\overline{x^2} - \overline{x}^2)} \right)^2 \right)$$

$\mathbb{V}\text{ar}\,(\hat{\beta}_1) = \sum_i^n \left(\frac{(x_i - \overline{x})\sigma}{n(\overline{x^2} - \overline{x}^2)} \right)^2 = \frac{1}{n^2\sigma_x^4}\left(\sum_{i=1}^n (x_i - \overline{x})^2\sigma^2 \right) = \frac{1}{n^2\sigma_x^4}\left(n\sigma_x^2\sigma^2 \right) = \frac{\sigma^2}{n\sigma_x^2}$

De même
$\hat{\beta}_0 = \frac{1}{n}\sum_{i=1}^n Y_i - \hat{\beta}_1\overline{x} = \frac{1}{n}\sum_{i=1}^n Y_i - \frac{\frac{1}{n}\sum_{i=1}^n x_iY_i - \overline{x}Y_i}{\sigma_x^2}\overline{x} = \frac{1}{n\sigma_x^2}\sum_{i=1}^n Y_i\sigma_x^2 - x_i\overline{x}Y_i + \overline{x}^2Y_i$
donc $\hat{\beta}_0 = \frac{1}{n\sigma_x^2}\sum_{i=1}^n(\sigma_x^2 - x_i\overline{x} + \overline{x}^2)Y_i = \sum_{i=1}^n(\frac{\overline{x^2} - x_i\overline{x}}{n\sigma_x^2})Y_i$

Donc $\hat{\beta}_0$ est la somme de variables aléatoires indépendantes de lois normales car $Y_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$ donc

$$\left(\frac{\overline{x^2} - x_i\overline{x}}{n\sigma_x^2}\right)Y_i \sim \mathcal{N}\left(\frac{\overline{x^2} - x_i\overline{x}}{n\sigma_x^2}(\beta_0 + \beta_1 x_i); \left(\frac{\overline{x^2} - x_i\overline{x}}{n\sigma_x^2}\right)^2\sigma^2\right)$$

$\mathbb{V}\text{ar}\,(\hat{\beta}_0) = \frac{1}{n^2\sigma_x^4}\sum_{i=1}^n \mathbb{V}\text{ar}\left((\overline{x^2} - x_i\overline{x})Y_i\right) = \frac{1}{n^2\sigma_x^4}\sum_{i=1}^n(\overline{x^2} - x_i\overline{x})^2\mathbb{V}\text{ar}\,(Y_i)$
donc $\mathbb{V}\text{ar}\,(\hat{\beta}_0) = \frac{1}{n^2\sigma_x^4}(n\overline{x^2}^2 - 2n(\overline{x})^2\overline{x^2} + n\overline{x^2}\overline{x}^2)\mathbb{V}\text{ar}\,(Y_i) = \frac{\sigma^2\overline{x^2}}{n\sigma_x^4}(\overline{x^2} - (\overline{x})^2) = \frac{\sigma^2\overline{x^2}}{n\sigma_x^2}$
Enfin $\mathcal{C}\text{ov}\,(\hat{\beta}_0, \hat{\beta}_1) = \mathcal{C}\text{ov}\left(\frac{\frac{1}{n}\sum_{i=1}^n(x_i - \overline{x})Y_i}{\sigma_x^2}; \frac{1}{n\sigma_x^2}\sum_{i=1}^n(\overline{x^2} - x_i\overline{x})Y_i\right) = \frac{1}{n^2\sigma_x^4}\sum_{i=1}^n\sum_{j=1}^n\mathcal{C}\text{ov}\left((x_i - \overline{x})Y_i; (\overline{x^2} - x_j\overline{x})Y_i\right)$
donc $\mathcal{C}\text{ov}\,(\hat{\beta}_0, \hat{\beta}_1) = \frac{\sigma^2}{n^2\sigma_x^4}\sum_{i=1}^n(x_i - \overline{x})(\overline{x^2} - x_i\overline{x}) = \frac{\sigma^2}{n^2\sigma_x^4}(-n\overline{x^2}\overline{x} + n\overline{x}^3) = \frac{\sigma^2\overline{x}}{n\sigma_x^4}(-\overline{x^2} + \overline{x}^2) = \frac{-\overline{x}\sigma^2}{n\sigma_x^2}$

On pose $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ c'est une estimation de l'espérance de Y_i, on peut donc estimer la variance des ε_i à partir de cette valeur par $\frac{1}{n}\sum_i(y_i - \hat{y}_i)^2$, cet estimateur de σ^2 est biaisé,

Définition 8.2 On note $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^n(Y_i - \hat{y}_i)^2 = \frac{1}{n-2}\sum_{i=1}^n(Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$.

Remarque 8.1 Le $n-2$ vient du fait que l'on estime deux paramètres : β_0 et β_1.

Proposition 8.3 (admis) $\hat{\sigma}^2$ est un estimateur sans biais de σ^2.
$K = \dfrac{(n-2)\hat{\sigma}^2}{\sigma^2} = \displaystyle\sum_{i=1}^n\left(\frac{Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)}{\sigma}\right)^2$ suit une loi de Chi deux à $n-2$ degrés de libertés.
$(\hat{\beta}_0; \hat{\beta}_1)$ est indépendant de $\hat{\sigma}^2$

Définition 8.3 On pose :

$$\begin{cases} \sigma_0^2 = \dfrac{\sigma^2\overline{x^2}}{n\sigma_x^2} & \text{La variance de } \hat{\beta}_0 \\ \sigma_1^2 = \dfrac{\sigma^2}{n\sigma_x^2} & \text{La variance de } \hat{\beta}_1 \\ \hat{\sigma}_0^2 = \dfrac{\hat{\sigma}^2\overline{x^2}}{n\sigma_x^2} & \text{La variance estimée de } \hat{\beta}_0 \\ \hat{\sigma}_1^2 = \dfrac{\hat{\sigma}^2}{n\sigma_x^2} & \text{La variance estimée de } \hat{\beta}_1 \end{cases}$$

Proposition 8.4 Une autre façon d'écrire la proposition 8.2 :

$$B_0 = \frac{\hat{\beta}_0 - \beta_0}{\sigma_0} \sim \mathcal{N}(0;1) \qquad \text{et} \qquad B_1 = \frac{\hat{\beta}_1 - \beta_1}{\sigma_1} \sim \mathcal{N}(0;1)$$

Qui donne si on la croise avec la proposition 8.3

$$T_0 = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_0} \sim \mathcal{S}(n-2) \qquad \text{et} \qquad T_1 = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_1} \sim \mathcal{S}(n-2)$$

Vidéo 9.3 : Régression linéaire et tests statistiques

8.4 Intervalle de confiance d'une prévision obtenue par régression linéaire

On aimerait pouvoir prévoir la valeur de Y_{n+1} en un point x_{n+1} pour lequel on n'a aucune données, on va estimer pour cela $y_{n+1} = \beta_0 + \beta_1 x_{n+1}$ à l'aide de $\hat{y}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{n+1}$, c'est tout l'intérêt de la régression linéaire. Mais on aimerait faire mieux et donner un intervalle de confiance pour y_{n+1}.

Proposition 8.5 $T_n = \sqrt{n}\dfrac{\hat{y}_{n+1} - y_{n+1}}{\hat{\sigma}\sqrt{1 + \frac{(x_{n+1} - \overline{x})^2}{\sigma_x^2}}}$ suit une loi de Student à $d = n-2$ degrés de libertés.

Preuve : $\hat{y}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{n+1} = \sum_{i=1}^n(\frac{\overline{x^2} - x_i\overline{x}}{n\sigma_x^2})Y_i + \sum_{i=1}^n\frac{(x_i - \overline{x})Y_i}{n\sigma_x^2}x_{n+1} = \sum_{i=1}^n\left(\frac{\overline{x^2} - x_i\overline{x}}{n\sigma_x^2} + \frac{(x_i - \overline{x})}{n\sigma_x^2}x_{n+1}\right)Y_i$

$$\hat{y}_{n+1} = \sum_{i=1}^n\left(\frac{\overline{x^2} + x_i(x_{n+1} - \overline{x}) - \overline{x}x_{n+1}}{n\sigma_x^2}\right)Y_i$$

Donc $\hat{y}_{n+1}$ est la somme de variables aléatoires indépendantes de lois normales car $Y_i \sim \mathcal{N}(\beta_0+\beta_1 x_i, \sigma^2)$ et $\hat{y}_{n+1}$ suit donc une loi normale.

$\mathbb{E}(\hat{y}_{n+1}) = \mathbb{E}(\hat{\beta}_0+\hat{\beta}_1 x_{n+1}) = \mathbb{E}(\hat{\beta}_0)+\mathbb{E}(\hat{\beta}_1)x_{n+1} = \beta_0+\beta_1 x_{n+1} = y_{n+1}$

$\mathbb{V}\text{ar}\,(\hat{y}_{n+1}) = \mathbb{V}\text{ar}\,(\hat{\beta}_0+\hat{\beta}_1 x_{n+1}) = \mathbb{V}\text{ar}\,(\hat{\beta}_0)+\mathbb{V}\text{ar}\,(\hat{\beta}_1)x_{n+1}^2+2\mathcal{C}\text{ov}\,(\hat{\beta}_0;\hat{\beta}_1)x_{n+1}$

en remplaçant les variances de $\hat{\beta}_0$ et $\hat{\beta}_1$ par leur valeur on obtient

$\mathbb{V}\text{ar}\,(\hat{y}_{n+1}) = \frac{\sigma^2\overline{x^2}}{n\sigma_x^2}+\frac{\sigma^2}{n\sigma_x^2}x_{n+1}^2+2\frac{-\overline{x}\sigma^2}{n\sigma_x^2}x_{n+1} = \frac{\sigma^2\overline{x^2}+\sigma^2x_{n+1}^2-2\overline{x}\sigma^2x_{n+1}}{n\sigma_x^2} = \frac{\sigma^2}{n\sigma_x^2}(\overline{x^2}+x_{n+1}^2-2\overline{x}x_{n+1})$

$\mathbb{V}\text{ar}\,(\hat{y}_{n+1}) = \frac{\sigma^2}{n\sigma_x^2}(\sigma_x^2+\overline{x}^2+x_{n+1}^2-2\overline{x}x_{n+1}) = \frac{\sigma^2}{n\sigma_x^2}\left(\sigma_x^2+(\overline{x}-x_{n+1})^2\right) = \frac{\sigma^2}{n}\left(1+\frac{(x_{n+1}-\overline{x})^2}{\sigma_x^2}\right)$

8.5 Résidus, coefficient de corrélation et coefficient de détermination

Définition 8.4 On appelle résidu l'écart entre y_i et $\hat{y}_i$.

S S E = S C E = $\sum_{i=1}^n(\hat{y}_i-\bar{y})^2$.

SSE est la variation expliquée par la régression (Sum of Squares Explained).

S S R = S C R = $\sum_{i=1}^n(y_i-\hat{y}_i)^2$

est la variation expliquée par les résidus (Sum of Squared Residuals).

S S T = S C T = $\sum_{i=1}^n(y_i-\bar{y})^2$ est la variation totale (Sum of Squares Total).

Le coefficient de corrélation $R = \frac{\mathcal{C}\text{ov}\,(x,Y)}{\sqrt{\mathbb{V}\text{ar}\,(x)\mathbb{V}\text{ar}\,(Y)}} = \frac{\overline{xY}-\overline{x}\overline{Y}}{\sqrt{(\overline{x^2}-\overline{x}^2)(\overline{y_i^2}-\overline{y}^2)}}$.

Le coefficient de détermination $R^2 = \frac{\text{S S E}}{\text{S S T}} = \frac{\sum_{i=1}^n(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^n(y_i-\bar{y})^2}$.

Remarque 8.2 On peut remarquer que S S T=S S E + S S R

Conformément aux notations le coefficient de détermination est le carré du coefficient de corrélation!!! $SSE = \sum_{i=1}^n(\hat{\beta}_0+\hat{\beta}_1x_i-(\hat{\beta}_0+\hat{\beta}_1\overline{x}))^2 = \hat{\beta}_1^2 n\sigma_x^2 = \frac{n}{\sigma_x^2}(\overline{xY}-\overline{x}\overline{Y})^2$

R^2 varie entre 0 et 1, lorsque R^2 est proche de 0, le pouvoir prédictif de la régression est plutôt faible et lorsque R^2 est proche de 1 son pouvoir prédictif est plutôt fort.

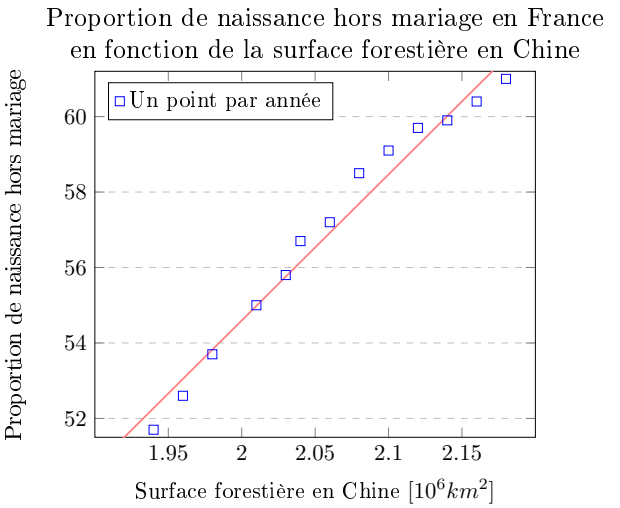
Proposition 8.6 (admis) Il est possible d'effectuer un test $H_0 : \beta_1 = 0$ contre $H_1 : \beta_1 \neq 0$.
Sous l'hypothèse H_0 les variables aléatoires $\frac{1}{\sigma^2}SSE$ et $\frac{1}{\sigma^2}SSR$ sont indépendantes de loi de chi deux à 1 et $n-2$ degrés de liberté. La variable aléatoire F défini par

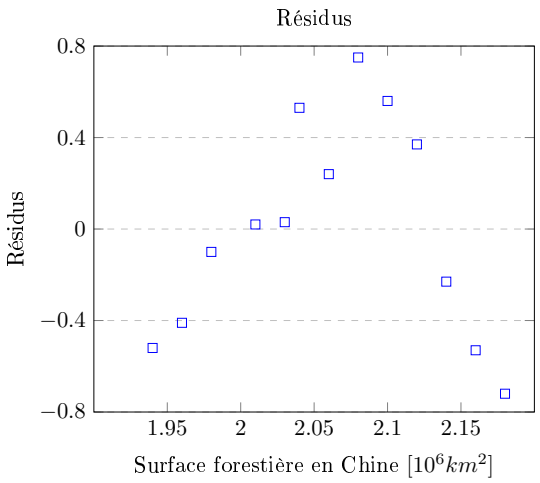
$$F = \frac{\frac{SSE}{1}}{\frac{SSR}{n-2}} = (n-2)\frac{SSE}{SSR} = (n-2)\frac{R^2}{1-R^2}$$

suit donc une loi de Fisher de paramètre $(1, n-2)$.

Remarque 8.3 La régression linéaire est un outil formidable, il ne faut pas oublier les hypothèses faites au début du chapitre, le fait que les "perturbations" (ε) suivent des lois normales, qu'elles sont indépendantes entre elles (souvent non vérifiée pour les séries temporelles par exemple) et surtout l'hypothèse d'homoscédasticité qui n'est pas toujours vérifiée.
Il faut aussi faire bien attention au fait qu'un lien statistiques même très bon n'est pas un lien de causalité.

Exemple 8.2 On étudie la corrélation entre la proportion de naissances hors mariage en France et la surface forestière en Chine : on trouve une très forte corrélation (~ 0.99) évidement il n'y a pas de causalité entre eux (ce n'est pas un des deux phénomènes qui entraîne l'autre), les deux phénomènes augmentent avec le temps, ils apparaissent donc fortement liés statistiquement. Un premier risque serait donc de conclure à un lien de cause à effet entre eux (causalité).
Ici on remarque que le modèle n'est pas très bon, si l'on regarde les résidus qui devraient être indépendants des x_i et de même loi, ce n'est clairement pas le cas. Et pourtant n'importe quel programme de statistiques va effectuer tous les calculs de régression mais les hypothèses n'étant pas vérifiées ce n'est pas vraiment interprétable.
Voyons les calculs effectués par *LibreOffice Calc* sur cet exemple.





Modèle de régression	Linéaire					
Statistiques de régression						
R^2 calculé	0,977					
Erreur type	0,49					
Nombre de variables X	1,00					
Observations	13,00					
R^2 estimé sans biais	0,975					
Analyse de la Variance (ANOVA)						
	df	SS	MS	F	Précision F	
Régression	1,00	111,22	111,22	470,79	2.10^{-10}	
Résidu	11,00	2,60	0,24			
Total	12,00	113,82				
Niveau de confiance	0,95					
	Coefficients	Erreur type	Student	Valeur P	Infér 95%	Supér 95%
Intercepter	-24,59	3,76	-6,53	4.10^{-5}	-32,87	-16,31
X1	39,59	1,82	21,70	2.10^{-10}	35,57	43,60

TABLE 8.1 – Rendu de la fonction Régression dans LibreOffice 7.3 Calc

Notations du cours	Notation LibreOffice	Résultats numériques
$\hat{\sigma}(\omega)$	Erreur type	0, 49
$\hat{\beta}_0(\omega)$	Coefficients Intercepter	-24,59
$\hat{\beta}_1(\omega)$	Coefficients X1	39,59
$\hat{\sigma}_0(\omega)$	Erreur type Intercepter	3,76
$\hat{\sigma}_1(\omega)$	Erreur type X1	1, 82
$SSE(\omega) = SS_{\text{expliqué}}(\omega)$	SS Régression	111,22
$SSR(\omega) = SS_{\text{résidu}}(\omega)$	SS Résidu	2,60
$F(\omega)$	F Régression	470,79
$R^2(\omega)$	R^2 calculé	0,977
Sous l'hypothèse $\beta_0 = 0$, $P\big(\lvert\hat{\beta}_0\rvert \geq \hat{\beta}_0(\omega)\big)$	Valeur P Intercepter	4.10^{-5}
Sous l'hypothèse $\beta_1 = 0$, $P\big(\lvert\hat{\beta}_1\rvert \geq \hat{\beta}_1(\omega)\big)$	Valeur P Intercepter	2.10^{-10}
Intervalle de confiance pour β_0	Intercepter [Infér 95% ; Supér 95%]	[-32,87;-16,31]
Intervalle de confiance pour β_1	X1 [Infér 95% ; Supér 95%]	[35,57;43,60]

TABLE 8.2 – On retrouve dans le Rendu de la fonction Régression dans *LibreOffice 7.3 Calc* un grand nombre des valeurs étudiées précédemment Pour notre échantillon ω

⚐ *fin de la semaine 9*

Cours de la semaine 10

Chapitre 9

Analyse de la variance

9.1 Introduction

L'objet de l'analyse de la variance est de tester si différentes populations possèdent la même moyenne pour un caractère donné. On suppose que l'on a I populations, et pour chaque population i, n_i est la taille de l'échantillon provenant de la population i. Le modèle s'écrit :

$$Y_{i,j} = A_i + \varepsilon_{i,j} \qquad \text{pour} \quad 1 \leq i \leq I \quad \text{et} \quad 1 \leq j \leq n_i$$

L'indice i correspond à la ième population, l'indice j correspond au jième individu de la ième population. On fait les hypothèses :

1. les $Y_{i,j}$ sont des variables aléatoires
2. A_i est une constante, égale à la moyenne de la ième population.
3. les $\varepsilon_{i,j}$ sont des variables aléatoires indépendantes et suivent toutes une même loi normale centrée ($\mathbb{E}(\varepsilon_{i,j}) = 0$) nous supposons de plus qu'elles ont toutes la même variance σ^2(homoscédasticité).

On prend comme estimateur

$$\begin{cases} \text{pour } A_i & : \quad \overline{Y_i} = \dfrac{1}{n_i} \displaystyle\sum_{j=1}^{n_i} Y_{i,j} \\ \text{pour } \sigma^2 & : \quad \hat{\sigma}^2 = \dfrac{1}{n-I} \displaystyle\sum_{i=1}^{I} \sum_{j=1}^{n_i} (Y_{i,j} - \overline{Y_i})^2 \end{cases}$$

On cherche à savoir si $A_1 = A_2 = ... = A_I$. On pose, en vue d'effectuer un test, l'hypothèse nulle

$$H_0 : A_1 = A_2 = ... = A_I \qquad\qquad H_1 : \exists i_1, i_2 \leq I; \quad A_{i_1} \neq A_{i_2}$$

Exemple 9.1 Par exemple on veut comparer la résistance à la compression de 5 formulations de béton qui différent par la teneur en un plastifiant, on casse une dizaine d'éprouvette de chacune des 5 formulations, comment savoir si les différences observées sont significatives ou pas ?

9.2 Test Anova

On peut appliquer le théorème 6.1 et l'on voit que $\frac{n_i-1}{\sigma^2}\hat{S}_i^2 = \sum_{j=1}^{n_i} \left(\frac{Y_{i,j}-\overline{Y_i}}{\sigma}\right)^2$ suit une loi de χ^2 à $n_i - 1$ degrés de libertés, donc en posant $SSR = \sum_{i=1}^{I}(n_i - 1)\hat{S}_i^2 = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{i,j} - \overline{Y_i})^2$, (variabilité intra-classe) la variable aléatoire $\frac{SSR}{\sigma^2}$ suit une loi de chi deux à $\sum_{i=1}^{I}(n_i - 1) = n - I$ degrés de libertés.

Sous l'hypothèse H_0 on peut appliquer le théorème 6.1 les $Y_{i,j}$ suivent tous la même loi normale de paramètre (A_1, σ^2), et en posant

$\overline{Y} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{I}\sum_{j=1}^{n_i} Y_{i,j}$ et $\hat{S}^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{i,j} - \overline{Y})^2$ la variable aléatoire $(n-1)\frac{\hat{S}^2}{\sigma^2}$ suit une loi de chi deux à $n-1$ degrés de liberté.

Posons $SSE = \sum_{i=1}^{I} n_i(\overline{Y}_i - \overline{Y})^2$ (variabilité inter-classe)

$$SST = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{i,j}-\overline{Y})^2 = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{i,j}-\overline{Y}_i+\overline{Y}_i-\overline{Y})^2 \quad (9.1)$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{n_i}\left((Y_{i,j}-\overline{Y}_i)^2+(\overline{Y}_i-\overline{Y})^2+2(Y_{i,j}-\overline{Y}_i)(\overline{Y}_i-\overline{Y})\right) \quad (9.2)$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{i,j}-\overline{Y}_i)^2+\sum_{i=1}^{I}\sum_{j=1}^{n_i}(\overline{Y}_i-\overline{Y})^2+2\sum_{i=1}^{I}\sum_{j=1}^{n_i}(\overline{Y}_i-\overline{Y})(Y_{i,j}-\overline{Y}_i) \quad (9.3)$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{i,j}-\overline{Y}_i)^2+\sum_{i=1}^{I}n_i(\overline{Y}_i-\overline{Y})^2+2\sum_{i=1}^{I}(\overline{Y}_i-\overline{Y})\sum_{j=1}^{n_i}(Y_{i,j}-\overline{Y}_i) \quad (9.4)$$

$$= \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{i,j}-\overline{Y}_i)^2+\sum_{i=1}^{I}n_i(\overline{Y}_i-\overline{Y})^2 \quad (9.5)$$

On a donc

$$(n-1)\hat{S}^2 = SSR + SSE$$

Théorème 9.1 (admis) *Sous l'hypothèse H_0 on peut montrer que SSR et SSE sont indépendantes, $(n-1)\frac{\hat{S}^2}{\sigma^2} \sim \chi^2(n-1)$ et $\frac{SSR}{\sigma^2} \sim \chi^2(n-I)$, et $\frac{SSE}{\sigma^2} \sim \chi^2(I-1)$ d'ou*

$$F = \frac{\frac{SSE}{I-1}}{\frac{SSR}{n-I}} \quad \textit{suit une loi de Fisher de paramètres } (I-1; n-I)$$

⚐ *fin de la semaine 10*

Index