

I. Objectifs

L'objectif visé par l'enseignement de la probabilité et de la statistique est de répondre aux besoins créés par le développement des nouvelles technologies de l'information et de la communication. En d'autres termes, l'objectif principal est la gestion et l'analyse de l'information.

Cet enseignement intervient à la fois par les activités traditionnelles de modélisation mathématique pour le traitement des données et l'aide à la décision dans un environnement aléatoire (ou incertain), auxquelles les nouvelles technologies apportent de nouveaux moyens de calcul et de traitement, et des problématiques liées directement à ces technologies (cryptographie, codage, compression,...)

Il s'appuie sur les connaissances acquises en statistiques, modélisation des phénomènes aléatoires, plans d'expérience, calcul scientifique..., ainsi que sur la maîtrise parfaite de l'outil informatique. Il ne suffira plus seulement de synthétiser sous forme chiffrée l'ensemble des données, mais d'estimer les relations entre différentes variables, l'intérêt étant de parvenir à établir des relations causales afin de pouvoir interpréter les résultats en termes d'impact.

A l'issue de leur formation, les ingénieurs seront aptes de mesurer, d'analyser, de modéliser, de planifier et de concevoir des projets.

II. Débouchés

Les métiers visés par cette formation " Gestion et analyse de l'information " dans le cadre des Nouvelles Technologies de l'Information et de la Communication sont les secteurs industriels, mais aussi les secteurs publics :

1. Gestion de l'information et de la connaissance dans l'entreprise : systèmes d'information, suivi de projet : enquêtes et sondages, documentation technique, télécommunication, automatisation des procédures bureautiques, veille technologique, intelligence économique.
2. Modélisation mathématique et traitement des données, avec des applications en aide à la décision ou en commande de processus.
3. Secteurs tertiaires : organismes publics de gestion et de prospective, collectivités territoriales, banques, bourse, assurances, administrations, bureaux d'études, grandes organisations, multinationales, offshoring ...
4. Télé-activités : télé-enseignement, et télé-conception en entreprise, télétravail, télé-services, commerce électronique.
5. Toutes autres activités liées au document électronique et multimédia : édition électronique, production de produits culturels multimédia.

III. Programme

Chapitre I Espaces de probabilités et vecteurs aléatoires

1. Espaces probabilisables et espaces probabilisés
 - Notion de σ -algèbre (ou tribu), définition d'un espace probabilisable
 - Axiomatique de Kolmogorov, définition d'un espace probabilisé, Lemme de Borel-Cantelli
 - Espace probabilisé conditionné par un événement (notion de probabilité conditionnelle), indépendance d'événements, théorème de Bayes.
2. Propriétés fondamentales sur les vecteurs aléatoires
 - Définitions : vecteur aléatoire, fonction de répartition, densité de probabilité
 - Moments d'un couple de variables aléatoires
 - Matrice des variances et covariances d'un vecteur aléatoire
 - Exemple d'application : loi normale multidimensionnelle $\mathcal{N}^n(\mu, \Sigma)$
3. Lois de probabilités usuelles : Normale, Exponentielle, Uniforme, Poisson, Binomiale, géométrique, Beta, Gamma, Khi-Deux, Student, Fisher-Snedecor,...
4. Fonctions génératrices et fonctions caractéristiques des moments d'un vecteur aléatoire

Chapitre II Introduction aux Processus Stochastiques et aux Chaînes de Markov

1. Généralités sur les processus stochastiques
 - Introduction
 - Notations et Définitions
 - Classification des processus stochastiques
2. Chaînes de Markov dans un espace d'états au plus dénombrable
 - Définitions et Propriétés de Markov
 - Chaînes de Markov Homogènes
 - Matrice stochastique (matrice de transition, noyau de transition)
 - Représentation graphique d'une chaîne de Markov
 - Théorème de Chapman-Kolmogorov
 - Etats possibles d'une chaîne de Markov (irréductible, , périodique,...)
 - Comportement asymptotique d'une chaîne de Markov (Ergodicité, stationnarité,...
 - Exemples d'application

Chapitre III Modèles statistiques paramétriques

1. Généralités sur l'inférence statistique
 - Définitions des concepts
 - Éléments de théorie de l'échantillonnage
2. Modèles statistiques
 - Définition et propriétés d'un modèle statistique

- Modèles statistiques classiques : Uniforme, Gaussien, Poisson, Binomial, Cauchy, Beta, Gamma, Exponentiel,...
- 3. Estimation ponctuelle des paramètres d'un modèle statistique paramétrique
 - Définitions : Estimateur, biais, convergence, efficacité
 - Estimation ponctuelle d'une moyenne et d'une variance
 - Méthodologie de construction de l'Estimateur du Maximum de Vraisemblance (EMV)
 - Règles de décision : fonction de perte, fonction de risque (cas du risque quadratique), Information de Fisher, test d'efficacité, inégalité de Rao-Cramer-Fréchet.
- 4. Estimation ensembliste : Intervalle de confiance
 - Généralités sur l'estimation par intervalle de confiance
 - Intervalle de confiance d'une moyenne
 - Intervalle de confiance d'une variance
 - Intervalle de confiance d'une proportion
 - Intervalle de confiance de la différence de : Deux moyennes, deux variances, deux proportions

Chapitre IV Géométrie des moindres carrés ordinaires

1. Introduction à la régression par la méthode des moindres carrés
 - Problématique
 - Modèles en économétrie
 - Théorie de la corrélation
2. Régression linéaire par la méthode des moindres carrés
 - Préviation par régression linéaire simple
 - Cas général : Préviation par régression polynomiale
3. Modèle de régression linéaire multiple par la méthode des moindres carrés
 - Position du problème
 - Estimation du vecteur $a = (a_0, a_1, a_2, \dots, a_k)$

IV. Quelques références bibliographiques

1. Y. Velenik, *Probabilités et Statistique*. Mars 2012 ;
[http ://www.unige.ch/math/folks/velenik/cours.html](http://www.unige.ch/math/folks/velenik/cours.html)
2. Jean Bertoin, *Probabilités : cours de licence en Mathématiques Appliquées* ; version 2000 – 2001 ; Université Paris 6.
3. Laurent Saloff-Coste, *Lectures on finite Markov chains*. In Lectures on probability theory and statistics (Saint-Flour, 1996), volume 1665 of Lecture Notes in Math., pages 301 – 413. Springer, Berlin, 1997

Chapitre 1

ESPACES DE PROBABILITES ET VECTEURS ALEATOIRES

1.1 Espaces Probabilisables et espaces probabilisés

1.1.1 Notion de σ -algèbre (ou tribu) de parties d'un ensemble

Soit Ω un ensemble non vide et $\mathcal{P}(\Omega)$ l'ensemble de ses parties.

Définition 1. Une tribu sur Ω (ou tribu de parties de Ω) est un sous ensemble $\mathcal{T}(\Omega)$ de $\mathcal{P}(\Omega)$ vérifiant :

- a) $\Omega \in \mathcal{T}(\Omega)$;
- b) Si $\forall A, \in \mathcal{T}(\Omega)$, alors $A^c \in \mathcal{T}(\Omega)$;
- c) $\mathcal{T}(\Omega)$ est stable pour la réunion dénombrable (si $(A_n, n \in \mathbb{N})$ est une suite d'éléments de $\mathcal{T}(\Omega)$, alors $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{T}(\Omega)$).

Remarque 1. – $\emptyset \in \mathcal{T}(\Omega)$;

- Par passage au complémentaire et sous les hypothèses de c), on a $\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{T}(\Omega)$.

Définition 2. Soit T un sous ensemble de $\mathcal{P}(\Omega)$. Soient \mathcal{T}_1 et \mathcal{T}_2 , deux tribus sur Ω et contenant T .

- 1) $\mathcal{T}_1 \cap \mathcal{T}_2$ est une tribu sur Ω , contenant T .
- 2) La tribu sur Ω engendré par T est la plus petite tribu sur Ω et contenant T .

Exemple 1. – $\mathcal{P}(\Omega)$ est une tribu sur Ω .

- Si $\Omega = \mathbb{R}^n$, la plus petite tribu contenant la topologie usuelle de \mathbb{R}^n (i.e. engendré par l'ensemble des ouverts de \mathbb{R}^n) est appelée tribu borelienne.
- Si $\Omega = \mathbb{R}$, on choisit en général $\mathcal{T}(\Omega) = \mathbb{B}_{\mathbb{R}}$ la tribu des boreliens de \mathbb{R} .
- Si Ω est fini ou dénombrable, on choisit en général $\mathcal{T}(\Omega) = \mathcal{P}(\Omega)$. Mais cela est impossible si Ω n'est pas dénombrable.

Définition 3. Soit Ω un ensemble non vide.

Le couple $(\Omega, \mathcal{T}(\Omega))$ est un espace probabilisable si $\mathcal{T}(\Omega)$ est une tribu sur Ω . Dans ce cas, tout élément A de $\mathcal{T}(\Omega)$ est un évènement.

1.1.2 Espaces Probabilisés

On utilisera désormais le langage probabiliste au lieu du langage ensembliste, à savoir qu'on appellera Ω l'univers, les parties de Ω des évènements, un élément $\omega \in \Omega$ un évènement élémentaire, et $\mathcal{T}(\Omega)$ une tribu (famille d'évènements observables sur Ω).

Définition 4. (*Axiomatique de Kolmogorov*)

Soit $(\Omega, \mathcal{T}(\Omega))$ un espace probabilisable.

1) On appelle probabilité (ou mesure de probabilité) sur $(\Omega, \mathcal{T}(\Omega))$, toute application $P : \mathcal{T}(\Omega) \rightarrow [0, 1]$ qui vérifie les deux axiomes suivants :

$A_1 : P(\Omega) = 1.$

$A_2 : \forall (A_n)_{n \geq 1}$ suite d'éléments de $\mathcal{T}(\Omega)$ telle que $A_i \cap A_j = \emptyset \forall i \neq j$, on a

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

2) Si P est une probabilité sur $(\Omega, \mathcal{T}(\Omega))$, alors le triplet $(\Omega, \mathcal{T}(\Omega), P)$ est appelé espace probabilisé.

Remarque 2. :

- L'axiome A_2 ci-dessus est la propriété de σ -additivité de P .
- Si Ω est fini ou dénombrable, on peut alors définir sur $(\Omega, \mathcal{P}(\Omega))$ une probabilité à l'aide d'une application $p : \Omega \rightarrow \mathbb{R}$ vérifiant : $p(\omega) \geq 0 \forall \omega \in \Omega$ et $\sum_{\omega \in \Omega} p(\omega) = 1$.

Il suffit de considérer pour $A \in \mathcal{P}(\Omega)$, $P(A) = \sum_{\omega \in A} p(\omega)$.

- Si Ω n'est pas dénombrable, on démontre qu'il n'existe pas une application $P : \mathcal{P}(\Omega) \rightarrow [0, 1]$ vérifiant les deux axiomes A_1 et A_2 .

Lemme 1. (*Borel-Cantelli*) :

Soit $(\Omega, \mathcal{T}(\Omega), P)$ un espace probabilisé, et $(A_n)_{n \geq 1}$ une suite d'évènements.

1) **Partie directe** : Si la série $\sum_{n \geq 1} P(A_n)$ converge, alors

$$P\left(\lim_{n \rightarrow +\infty} \sup A_n\right) = P\left(\lim_{n \rightarrow +\infty} \bigcup_{k \geq n} A_k\right) = P\left(\bigcap_{n \geq 1} \bigcup_{k \geq n} A_k\right) = 0 ; \text{ C'est à dire au plus}$$

un nombre fini d'évènements A_n se réalise.

2) **Partie réciproque** : Si les évènements $(A_n)_{n \geq 1}$ sont mutuellement indépendants et la série $\sum_{n \geq 1} P(A_n)$ diverge, alors $P\left(\lim_{n \rightarrow +\infty} \sup A_n\right) = 1$; C'est à dire une infinité d'évènements A_n se réalise avec une probabilité égale à 1.

1.1.3 Probabilités conditionnelles et indépendance d'évènements

Soit $(\Omega, \mathcal{T}(\Omega))$ un espace probabilisable et A un élément de $\mathcal{T}(\Omega)$.

On pose $\mathcal{T}_A(\Omega) = \{B \cap A, B \in \mathcal{T}(\Omega)\} = \{C / \exists B \in \mathcal{T}(\Omega), C = B \cap A\}$.

Définition 5. Si $\mathcal{T}_A(\Omega)$ est une tribu sur A , le couple $(A, \mathcal{T}_A(\Omega))$ est appelé "espace probabilisable $(\Omega, \mathcal{T}(\Omega))$ conditionné par A ".

Proposition 1. Soit $(\Omega, \mathcal{T}(\Omega), P)$ un espace probabilisé et $A \in \mathcal{T}(\Omega)$ tel que $P(A) \neq 0$.

- L'application $P' : \mathcal{T}_A(\Omega) \rightarrow \mathbb{R}$ définie par $P'(A \cap B) = P(B/A) = \frac{P(A \cap B)}{P(A)}$ est une probabilité sur $(A, \mathcal{T}_A(\Omega))$. Donc $(A, \mathcal{T}_A(\Omega), P')$ et $(\Omega, \mathcal{T}(\Omega), P(\cdot/A))$ sont des espaces probabilisés.
- $\forall B \in \mathcal{T}(\Omega)$, on a $P(A \cap B) = P(A)P(B/A)$.

Définition 6. Soit $(\Omega, \mathcal{T}(\Omega), P)$ un espace probabilisé.

1) Deux évènements A et B sont dits P -stochastiquement indépendants (ou tout simplement indépendants) si $P(A \cap B) = P(A)P(B)$.

2) Les évènements A_1, A_2, \dots, A_n sont dits mutuellement indépendants si pour toute sous famille $A_{i_1}, A_{i_2}, \dots, A_{i_k}$ avec $1 \leq i_1 < i_2 < \dots < i_k \leq n$, on a :

$$P\left(\bigcap_{l=1}^{l=k} A_{i_l}\right) = \prod_{l=1}^{l=k} P(A_{i_l})$$

Remarque 3. :

- L'indépendance dépend de la probabilisation de l'espace.
- Si A et B sont indépendants, alors A et \overline{B} , \overline{A} et B , \overline{A} et \overline{B} sont aussi indépendants.
- Si les évènements A_1, A_2, \dots, A_n sont mutuellement indépendants, alors ils sont indépendants 2 à 2. Mais la réciproque est fausse.

1.1.4 Théorème de BAYES

Soit $(\Omega, \mathcal{T}(\Omega), P)$ un espace probabilisé.

Définition 7. On appelle système de constituants (ou système complet d'évènements), toute famille finie ou dénombrable $(B_n)_{n \geq 1}$ d'éléments de $\mathcal{T}(\Omega)$, et qui forme une partition de Ω .

Théorème 1. On considère $B_1, B_2, \dots, B_n, \dots$ un système de constituants tel que $P(B_n) \neq 0 \forall n \geq 1$. Alors pour tout évènement A de probabilité non nulle, on a :

$$P(B_n/A) = \frac{P(B_n)P(A/B_n)}{\sum_{k \geq 1} P(B_k)P(A/B_k)} \quad \forall n \geq 1$$

$P(B_n)$ est la probabilité à priori de B_n et $P(B_n/A)$ est la probabilité à posteriori de B_n .

1.2 Propriétés fondamentales sur les vecteurs aléatoires

1.2.1 Généralités

Soient $(\Omega_1, \mathcal{T}(\Omega_1))$ et $(\Omega_2, \mathcal{T}(\Omega_2))$ deux espaces probabilisables.

Définition 8. 1) Une application $X : \Omega_1 \rightarrow \Omega_2$ est dite mesurable si : $\forall A \in \mathcal{T}(\Omega_2)$, $X^{-1}(A) \in \mathcal{T}(\Omega_1)$. On dit alors que X est une variable aléatoire à valeurs dans Ω_2 .
2) Si $X(\Omega_1)$ est un intervalle de \mathbb{R} et $\mathcal{T}(\Omega_2) \subset \mathbb{B}_{\mathbb{R}}$, on dit que X est une variable aléatoire réelle ou continue.
3) Si $X(\Omega_1)$ est finie ou dénombrable alors X est une variable aléatoire discrète.
4) Si $\Omega_2 = \mathbb{R}^n$ et $\mathcal{T}(\Omega_2) = \mathbb{B}_{\mathbb{R}^n}$, on dit que X est un vecteur aléatoire ou variable aléatoire vectorielle.

Proposition 2. $X = (X_1, X_2, \dots, X_n)$ est une variable aléatoire vectorielle de \mathbb{R}^n si et seulement si X_i est une variable aléatoire réelle $\forall i = 1, \dots, n$.

Définition 9. Soit $(\Omega_1, \mathcal{T}(\Omega_1), P)$ un espace probabilisé ; $(\Omega_2, \mathcal{T}(\Omega_2))$ et $(\Omega_3, \mathcal{T}(\Omega_3))$ deux espaces probabilisables.

1) Si $X : \Omega_1 \rightarrow \Omega_2$ est une variable aléatoire, l'application $P_X : \mathcal{T}(\Omega_2) \rightarrow [0, 1]$, définie par : $\forall A' \in \mathcal{T}(\Omega_2)$, $P_X(A') = P(X^{-1}(A')) = P(\{\omega \in \Omega_1 / X(\omega) \in A'\})$ est une probabilité sur $\mathcal{T}(\Omega_2)$. On l'appelle probabilité image de P par X ou loi de probabilité de X .
2) Si X_1, X_2, \dots, X_n sont n variables aléatoires réelles, la probabilité conjointe de (X_1, X_2, \dots, X_n) est la loi de probabilité du vecteur aléatoire $X = (X_1, X_2, \dots, X_n)$ à valeurs dans \mathbb{R}^n .
3) Deux variables aléatoires X et Y (à valeurs respectives dans Ω_2 et Ω_3) sont dites indépendantes si : $\forall A \in \mathcal{T}(\Omega_2)$ et $\forall B \in \mathcal{T}(\Omega_3)$, $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$.

1.2.2 Densité de probabilité et Fonction de répartition

Soit $(\Omega, \mathcal{T}(\Omega), P)$ un espace probabilisé, et $(\mathbb{R}^n, \mathbb{B}_{\mathbb{R}^n})$ l'espace borelien. On considère $X = (X_1, X_2, \dots, X_n)$ une variable aléatoire définie de Ω vers \mathbb{R}^n .

Définition 10. 1) La fonction de répartition de X (ou fonction de repartition de la probabilité P_X) est la fonction $F : \mathbb{R}^n \rightarrow [0, 1]$ définie par :

$$F(t_1, t_2, \dots, t_n) = P(X_1 < t_1, X_2 < t_2, \dots, X_n < t_n)$$

2) Si $n = 1$, la fonction F devient : $F(t) = P(X < t) = \int_{-\infty}^t dP_X(x) = P_X(]-\infty, t])$. C'est la fonction de répartition de la variable aléatoire réelle X .

Propriétés 1. Si F est une fonction de repartition, alors :

- F est croissante.
- F est continue à gauche par rapport à chaque variable.
- F tend vers 0 lorsque l'un des t_i tend vers $-\infty$.
- F tend vers 1 lorsque tous les t_i tendent vers $+\infty$.
- F détermine de manière unique P_X . En effet, on montre qu'il existe une bijection entre l'ensemble des fonctions de répartitions sur \mathbb{R}^n et l'ensemble des probabilités définies sur $(\mathbb{R}^n, \mathbb{B}_{\mathbb{R}^n})$.

Remarque 4. – La fonction de répartition F_{X_i} de la v.a X_i est donnée par :

$$F_{X_i}(t_i) = F(+\infty, \dots, +\infty, t_i, +\infty, \dots, +\infty)$$

- Si F est n fois différentiable, P_X admet pour densité la fonction $f_X : \mathbb{R}^n \rightarrow [0, 1]$ définie par :

$$f_X(t_1, t_2, \dots, t_n) = \frac{\partial^n F(t_1, t_2, \dots, t_n)}{\partial t_1 \partial t_2 \dots \partial t_n}$$

par rapport à la mesure de Borel sur \mathbb{R}^n .

- En particulier pour $n = 1$, f_X est la densité de probabilité de X par rapport à la mesure de Lebesgue sur \mathbb{R} .

1.2.3 Moments d'un couple de variables aléatoires réelles

Soient (X, Y) un couple de variables aléatoires réelles, de fonction de répartition F et de densité de probabilité f .

Définition 11. Soient h et k deux nombres réels. On appelle moment d'ordre h en X et d'ordre k en Y la quantité suivante quand elle existe :

$$m_{h,k} = E(X^h Y^k) = \int \int_{\mathbb{R}^2} x^h y^k dF(x, y) = \int \int_{\mathbb{R}^2} x^h y^k f(x, y) dx dy$$

Exemple 2. $m_{1,0} = E(X)$; $m_{0,1} = E(Y)$ et $m_{1,1} = E(XY)$

Définition 12. Si on pose $X_1 = X - E(X)$ et $Y_1 = Y - E(Y)$, alors les moments du couple (X_1, Y_1) , d'ordre h en X_1 et d'ordre k en Y_1 sont appelés **moments centrés** du couple (X, Y) et que l'on note souvent $\mu_{h,k}$.

Exemple 3. $\mu_{2,0} = V(X)$; $\mu_{0,2} = V(Y)$ et $\mu_{1,1} = COV(X, Y)$

1.2.4 Matrice des variances et covariances d'un vecteur aléatoire

Soit $X = (X_1, X_2, \dots, X_n)$ un vecteur aléatoire. Considérons $(e_i)_{i=1}^n$ une base de \mathbb{R}^n et $(e_i^*)_{i=1}^n$ sa base duale. Posons $x^* = \sum_{i=1}^n \alpha_i e_i^*$ et $X = \sum_{i=1}^n X_i e_i$.

Définition 13. On appelle variance de X l'application $Var(X)$ définie de $(\mathbb{R}^n)^*$ vers \mathbb{R} qui à x^* associe $Var(X)(x^*)$ telle que :

$$Var(X)(x^*) = E[(\sum_{i=1}^n \alpha_i X_i - \sum_{i=1}^n \alpha_i E(X_i))^2] = E[(\sum_{i=1}^n \alpha_i (X_i - E(X_i)))^2]$$

Après calcul, on obtient finalement,

$$Var(X)(x^*) = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j E[(X_i - E(X_i))(X_j - E(X_j))]$$

Remarque 5. – (i) La variance de X est donc une forme quadratique positive sur $(\mathbb{R}^n)^*$ (espace dual) dont la matrice dans la base duale $(e_i^*)_{i=1}^n$ est la matrice C d'éléments

$$C_{ij} = E[(X_i - E(X_i))(X_j - E(X_j))] \text{ avec } 1 \leq i, j \leq n.$$

- (ii) $C_{ii} = V(X_i)$; $C_{ij} = Cov(X_i, X_j) = Cov(X_j, X_i) = C_{ji}$
- (iii) C est une matrice symétrique et positive.
- (iv) C est appelée matrice des variances et covariances des variables aléatoires réelles X_1, X_2, \dots, X_n .

1.2.5 Exemple d'application : Loi normale multidimensionnelle $\mathcal{N}^n(\mu, \Sigma)$

Définition 14. Soient $\mu \in \mathbb{R}^n$, Σ une matrice symétrique définie positive. on considère le vecteur aléatoire $X = (X_1, X_1, \dots, X_n)$ qui suit la loi normale à n dimensions de paramètres μ et Σ . Naturellement, chaque X_i suit la loi normale de paramètres $E(X_i)$ et $\sigma^2(X_i)$.

On note alors $X \hookrightarrow \mathcal{N}^n(\mu, \Sigma)$

Proposition 3. On suppose que $X = (X_1, X_1, \dots, X_n)$ suit la loi $\mathcal{N}^n(\mu, \Sigma)$. Alors on a :

- (i) $\mu = E(X) = (E(X_1), E(X_2), \dots, E(X_n))$
- (ii) Σ est la variance de X , c'est à dire la matrice des variances et covariances des v.a.r X_1, X_2, \dots, X_n
- (iii) La densité de probabilité de X est donnée par :

$$f(x_1, x_2, \dots, x_n) = \frac{\sqrt{\det(\Sigma^{-1})}}{(2\pi)^{\frac{n}{2}}} \exp\left[-\frac{1}{2}Q(x_1, x_2, \dots, x_n)\right]$$

où $Q(x_1, x_2, \dots, x_n) = {}^t(x - \mu)\Sigma^{-1}(x - \mu)$ et $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$.

Remarque 6. Etant donné X qui suit la loi $\mathcal{N}^n(\mu, \Sigma)$, on souhaite déterminer μ et Σ connaissant l'expression quadratique de la densité de X .

- (i) μ est déterminé en résolvant le système d'équation :

$$\frac{\partial Q(x_1, x_2, \dots, x_n)}{\partial x_i} = 0, \quad \forall i = 1, 2, \dots, n$$

- (ii) La matrice $\Sigma^{-1} = (C'_{ij})_{1 \leq i, j \leq n}$; où $C'_{ij} = \frac{1}{2} \frac{\partial^2 Q(x_1, x_2, \dots, x_n)}{\partial x_i \partial x_j}$

Théorème 2. On suppose que le vecteur aléatoire X suit la loi $\mathcal{N}^n(\mu, \Sigma)$ et que Σ est une matrice diagonale. Alors, les v.a.r X_1, X_2, \dots, X_n sont Gaussiennes (lois normales) et indépendantes et la densité de X est :

$$f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right] = \prod_{i=1}^n f(x_i)$$

où $f(x_i)$ désigne la fonction densité de probabilité de la v.a.r X_i .

1.3 Quelques lois de probabilités usuelles

1.3.1 Loi normale

C'est la loi la plus importante en statistique car elle permet d'approcher plusieurs autres lois. Sa forme la plus utilisée est la forme centrée réduite qui s'obtient par un changement de variables. En effet, si X suit la loi $\mathcal{N}(\mu, \sigma)$, on pose alors $U = \frac{X - \mu}{\sigma}$. On montre alors facilement que U suit la loi normale centrée réduite $\mathcal{N}(0, 1)$.

1.3.2 Loi exponentielle

Si X suit la loi exponentielle, sa densité de probabilité est définie par :

$$f_q(x) = \begin{cases} qe^{-qx} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases} ; \text{ avec } q \text{ un réel strictement positif.}$$

$$\text{Ainsi, } E(X) = \frac{1}{q} ; \text{ et } V(X) = \frac{1}{q^2}$$

1.3.3 Loi uniforme

Si X suit la loi uniforme sur l'intervalle $[a, b]$, on pose $\theta = b - a$. Sa densité de probabilité est alors définie par :

$$f_X(x) = f_\theta(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{sinon} \end{cases} = \begin{cases} \frac{1}{\theta} & \text{si } x \in [0, \theta] \\ 0 & \text{sinon} \end{cases} ; \text{ avec } a \text{ et } b \text{ deux réels ;}$$

et la loi $U = X - a$.

$$\text{Ainsi, } E(X) = \frac{a+b}{2} ; \text{ et } V(X) = \frac{(b-a)^2}{12} = \frac{\theta^2}{12}.$$

1.3.4 Loi de Poisson

C'est une loi discrète de X et définie par : $p_\lambda(k) = \frac{\lambda^k e^{-\lambda}}{k!}$; où $k \in \{0, 1, 2, \dots\}$ et λ est un réel strictement positif. On note symboliquement $X \hookrightarrow \mathbb{P}_\lambda$.

Propriétés 2. – (i) $E(X) = \lambda$ et $V(X) = \lambda$.

– (ii) On montre que lorsque $\lambda \geq 10$, la loi de X (i.e \mathbb{P}_λ) peut être approchée par la loi normale $\mathcal{N}(\lambda, \sqrt{\lambda})$.

1.3.5 Loi Binomiale

C'est une loi discrète de X et définie par : $p(X = k) = C_n^k p^k (1-p)^{n-k}$; où $k \in \{0, 1, 2, \dots, n\}$ et $p \in [0, 1]$. On note symboliquement $X \hookrightarrow \mathcal{B}(n, p)$.

Propriétés 3. – (i) $E(X) = np$ et $V(X) = np(1-p)$.

– (ii) On montre que lorsque n est grand ou encore $np \geq 10$, la loi de X (i.e $\mathcal{B}(n, p)$) peut être approchée par la loi normale $\mathcal{N}(np, \sqrt{np(1-p)})$.

1.3.6 Loi Beta

$$a > 0, b > 0 \text{ avec } f_{a,b}(x) = \begin{cases} \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)} & \text{si } x \in [0, 1] \\ 0 & \text{sinon} \end{cases} ; \text{ où}$$

$$B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt$$

$$\text{Ainsi, } E(X) = \frac{a+1}{a+b+2} \text{ et } V(X) = \frac{(a+1)(b+1)}{(a+b+3)(a+b+2)^2}$$

1.3.7 Loi Gamma

$$p > 0, q > 0 \text{ avec } f_{p,q}(x) = \begin{cases} \frac{q^p e^{-qx} x^{p-1}}{\Gamma(p)} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases} ; \text{ où}$$

$$\Gamma(p) = \int_0^{+\infty} t^{p-1} e^{-t} dt. \text{ Si } p = 1, \text{ on retrouve le modèle exponentiel élémentaire.}$$

Ainsi, $E(X) = \frac{p}{q}$ et $V(X) = \frac{p}{q^2}$

NB : $X = \sum_{i=1}^p X_i$, et $X_i \hookrightarrow \mathcal{E}(q)$.

1.3.8 Loi Géométrique

C'est une loi discrète de X et définie par : $p(X = k) = p(1 - p)^{k-1}$; où $k \in \{1, 2, 3, \dots\}$ et $p \in]0, 1[$. p étant la probabilité de succès. On note symboliquement $X \hookrightarrow \mathcal{G}(p)$.

Ainsi, $E(X) = \frac{1}{p}$ et $V(X) = \frac{1-p}{p^2}$

1.3.9 Loi du Khi-deux

Soient X_1, X_2, \dots, X_n ; n v.a indépendantes telles que chaque X_i suit la loi normale centrée réduite. Alors, on dit que $X = X_1^2 + X_2^2 + \dots + X_n^2$ suit la loi du Khi-deux à n degrés de liberté et est symbolisée par χ_n^2 .

On constate qu'elle ne peut prendre que des valeurs positives et sa densité de probabilité est donc : $f(t) = C_n t^{\frac{n}{2}-1} e^{-\frac{t}{2}}$; où la constante C_n est obtenue en résolvant l'équation $\int_{-\infty}^{+\infty} f(t)dt = 1$ (C'est la propriété de probabilité totale).

Propriétés 4. – (i) $E(X) = n$ et $V(X) = 2n$

- (ii) On montre que lorsque n tend vers l'infini (en pratique $n > 30$), la loi $\frac{X-n}{\sqrt{2n}}$ suit la loi normale centrée réduite $\mathcal{N}(0, 1)$ (car dans ces conditions, X suit la loi normale $\mathcal{N}(n, \sqrt{2n})$ d'après le théorème central limite).
- (iii) Soient X_1, X_2, \dots, X_k ; k v.a indépendantes telles que chaque X_i suit la loi $\chi_{n_i}^2$. Alors, $Z = X_1 + X_2 + \dots + X_k$ suit la loi χ_n^2 où $n = n_1 + n_2 + \dots + n_k$. C'est la propriété d'additivité de la loi du Khi-deux.

1.3.10 Loi de Student

Elle joue un rôle très important dans l'estimation par intervalle de confiance. Elle est symétrique, de moyenne nulle et dépend d'un paramètre n appelé nombre de degrés de liberté. Soient X et Y deux v.a indépendantes telles que $X \hookrightarrow \mathcal{N}(0, 1)$ et $Y \hookrightarrow \chi_n^2$.

Alors, on dit que $Z = \frac{X}{\sqrt{\frac{Y}{n}}}$ suit la loi de Student à n degrés de liberté et est symbolisée par $Z \hookrightarrow St_n$. Sa densité de probabilité est définie par : $f(t) = C_n (1 + \frac{t^2}{n})^{-\frac{n+1}{2}}$; où la constante C_n est obtenue en résolvant l'équation $\int_{-\infty}^{+\infty} f(t)dt = 1$.

Propriétés 5. – (i) $E(Z) = 0$ si $n > 1$

- (ii) $V(Z) = \frac{n}{n-2}$ si $n > 2$
- (iii) On montre que pour des échantillons très élevés (en pratique $n > 30$), la loi $Z = St_n$ suit la loi normale centrée réduite $\mathcal{N}(0, 1)$.

1.3.11 Loi de Fischer-Snedecor

Elle joue un rôle très important dans l'estimation des proportions ou des fréquences empiriques. Soient Y_1 et Y_2 deux v.a indépendantes telles que $Y_1 \hookrightarrow \chi_{n_1}^2$ et $Y_2 \hookrightarrow \chi_{n_2}^2$.

Alors, on dit que $F = \frac{\frac{Y_1}{n_1}}{\frac{Y_2}{n_2}}$ suit la loi de Fischer-Snedecor à n_1 et n_2 degrés de liberté et est symbolisée par $F \hookrightarrow F(n_1, n_2)$. Sa densité de probabilité est donnée par :

$f(t) = C_{n_1, n_2} t^{\frac{n_1}{2}-1} (n_1 t + n_2)^{-\frac{n_1+n_2}{2}}$ si $t > 0$ et $f(t) = 0$ si $t \leq 0$; où la constante C_{n_1, n_2} est obtenue en résolvant l'équation $\int_{-\infty}^{+\infty} f(t) dt = 1$.

Propriétés 6. – (i) $E(F) = \frac{n_1}{n_2-2}$ si $n_2 > 2$

– (ii) $V(F) = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$ si $n_2 > 4$

1.4 Fonction génératrice et fonction caractéristique d'un vecteur aléatoire

La fonction génératrice et la fonction caractéristique d'une v.a permettent de calculer très facilement ses moments, pourvu qu'ils existent.

1.4.1 Fonction génératrice des moments d'un vecteur aléatoire

Soit $X = (X_1, X_2, \dots, X_n)$ une variable aléatoire à valeurs dans \mathbb{R}^n .

Définition 15. On définit la fonction m_X de \mathbb{R}^n vers \mathbb{R} par : $\forall t = (t_1, t_2, \dots, t_n) \in \mathbb{R}^n$, $m_X(t) = E(e^{tX}) = E(e^{t_1 X_1 + \dots + t_n X_n})$. Si $m_X(t)$ existe au voisinage de 0 dans \mathbb{R}^n , alors m_X est appelée fonction génératrice des moments (f.g.m) de la variable aléatoire vectorielle X . Elle caractérise la loi de probabilité de X (i.e m_X permet de déterminer entièrement la densité ou la fonction de répartition de X).

Remarque 7. On suppose que $m_X(t)$ existe au voisinage de 0. Alors,

- $E(X_1^{h_1} X_2^{h_2} \dots X_n^{h_n}) = \frac{\partial^{h_1+h_2+\dots+h_n} m_X}{\partial t_1^{h_1} \partial t_2^{h_2} \dots \partial t_n^{h_n}}(0, 0, \dots, 0)$
- Si $n = 1$, $m_X(t)$ est la f.g.m de la v.a.r X et est indéfiniment dérivable au voisinage de 0. Le moment d'ordre k de X est alors $E(X^k) = m_X^{(k)}(0)$.

Exemple 4. Si $X = (X_1, X_2, \dots, X_n)$ suit la loi normale $N^n(\mu, \Sigma)$, alors $m_X(t) = \exp(t\mu + \frac{1}{2}t\Sigma t)$ où $t = (t_1, t_2, \dots, t_n)$.

Il est alors évident que pour $n = 1$, on a $m_X(t) = e^{t\mu + \frac{1}{2}\sigma^2 t^2}$.

1.4.2 Fonction caractéristique des moments d'un vecteur aléatoire

Soit $X = (X_1, X_2, \dots, X_n)$ une variable aléatoire à valeurs dans \mathbb{R}^n , de loi de probabilité conjointe P_X .

Définition 16. On appelle fonction caractéristique de X de densité de probabilité f , l'application φ_X de \mathbb{R}^n vers \mathbb{C} définie par :

$$\varphi_X(t) = \varphi_X(t_1, t_2, \dots, t_n) = E(e^{itX}) = E(e^{i(t_1 X_1 + \dots + t_n X_n)}) = \int_{\mathbb{R}^n} e^{i(t_1 x_1 + \dots + t_n x_n)} f(x) dx$$

Donc φ_X est la transformée de FOURRIER de la loi de X .

En particulier pour $n = 1$, $\varphi_X(t) = \int_{\mathbb{R}} e^{itx} f_X(x) dx$ dépend uniquement de la distribution de X .

Propriétés 7. Soit φ_X la fonction caractéristique de la v.a à n dimensions X ; alors :

(i) $\forall t \in \mathbb{R}^n, |\varphi_X(t)| \leq 1$. En d'autres termes, φ_X prend ses valeurs dans le disque unité. En effet, $|E(e^{itX})| = E(|e^{itX}|) = 1$

(ii) $\varphi_X(0) = 1$.

(iii) φ_X est une fonction continue. C'est une conséquence immédiate de la continuité de l'application $t \mapsto e^{itx}$ pour chaque $x \in \mathbb{R}^n$.

(iv) Si $E(X_1^{h_1} X_2^{h_2} \dots X_n^{h_n}) < +\infty$, on a

$$\frac{\partial^{h_1+h_2+\dots+h_n}}{\partial t_1^{h_1} \partial t_2^{h_2} \dots \partial t_n^{h_n}} \varphi_X(0, 0, \dots, 0) = i^{h_1+h_2+\dots+h_n} E(X_1^{h_1} X_2^{h_2} \dots X_n^{h_n})$$

(v) Les v.a.r X_1, X_2, \dots, X_n sont indépendantes si et seulement si

$$\varphi_X(t_1, t_2, \dots, t_n) = \varphi_{X_1}(t_1) \varphi_{X_2}(t_2) \dots \varphi_{X_n}(t_n)$$

(vi) $n = 1$, $\varphi_X(-t) = \overline{\varphi_X(t)}$.

(vii) Deux variables aléatoires réelles X et Y ont même loi si et seulement si $\varphi_X = \varphi_Y$.

On supposera dans la suite que X est une v.a.r.

Exemple 5. Si X suit la loi uniforme U sur $[a, b]$, on a $\varphi_X(t) = \frac{e^{itb} - e^{ita}}{i(b-a)t}$, $t \in \mathbb{R}$.

Si X suit la loi exponentielle de paramètre $q > 0$, on a $\varphi_X(t) = \frac{q}{q-it}$, $t \in \mathbb{R}$.

Si X suit la loi de Cauchy, on a $\varphi_X(t) = e^{-|t|}$, $t \in \mathbb{R}$.

Si X suit la loi normale $N(\mu, \sigma)$, on a $\varphi_X(t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}$, $t \in \mathbb{R}$.

Si X suit la loi de poisson $\mathcal{P}(\lambda)$, on a $\varphi_X(t) = e^{\lambda(e^{it} - 1)}$, $t \in \mathbb{R}$.

Si X suit la loi binomiale $\mathcal{B}(n, p)$, on a $\varphi_X(t) = (q + pe^{it})^n$ avec $q = 1 - p$, $t \in \mathbb{R}$.

Théorème 3. Soit X une v.a.r, de loi de probabilité P_X et φ_X sa fonction caractéristique.

Si $E(X^n) < +\infty$ pour un entier $n \geq 1$, alors φ_X est continûment dérivable jusqu'à l'ordre n . Sa dérivée $k^{\text{ième}}$ est alors $\varphi_X^{(k)}(t) = i^k \int x^k e^{itx} dP_X(x)$. En particulier : $\varphi_X^{(k)}(0) = i^k E(X^k)$, donc $E(X^k) = \frac{1}{i^k} \varphi_X^{(k)}(0)$.

Pour Démontrer le théorème, nous avons besoin du lemme suivant :

Lemme 2. Soit $(\Omega, \mathcal{T}(\Omega), P)$ un espace probabilisé et $]a, b[$ un intervalle ouvert de \mathbb{R} . Soit une fonction $g : \Omega \times]a, b[\rightarrow \mathbb{R}$ telle que :

1) $\forall t \in]a, b[, g(., t)$ est une v.a.r sur $(\Omega, \mathcal{T}(\Omega))$ et $E(g(., t))$ existe.

2) $\forall \omega \in \Omega$, $\frac{\partial}{\partial t}g(\omega, t)$ existe sur $]a, b[$.

3) $|\frac{\partial}{\partial t}g(., t)| \leq Y$ où Y est une v.a.r telle que $E(Y)$ existe.

Alors $E(g(., t))$ est dérivable et $\frac{d}{dt}E(g(., t)) = E(\frac{\partial}{\partial t}g(., t))$; c'est à dire

$$\frac{d}{dt}\left(\int g(\omega, t)dP(\omega)\right) = \int \frac{\partial}{\partial t}g(\omega, t)dP(\omega)$$

.

Preuve du théorème 3 : Supposons que $E(X^n) < +\infty$ pour un certain $n \geq 1$. Ce qui implique $E(X^k) < +\infty \forall k \leq n$. Soit $g(X, t) = e^{itX}$. On a $\varphi_X(t) = E(g(X, t))$ et

$|\frac{\partial}{\partial t}g(., t)| = |iXe^{itX}| \leq |X|$ et $E(|X|)$ existe. En appliquant le lemme 2, on a :

$$\varphi_X'(t) = i \int xe^{itx}dP_X(x).$$

Soit ensuite $g(X, t) = iXe^{itX}$, on a $\varphi_X'(t) = E(g(X, t))$ et $|\frac{\partial}{\partial t}g(., t)| = |(iX)^2e^{itX}| \leq$

$|X|^2$ et $E(X^2)$ existe. En appliquant encore le lemme 2, on a : $\varphi_X^{(2)}(t) = i^2 \int x^2e^{itx}dP_X(x)$.

Ainsi en appliquant le lemme 2 à e^{itX} et à ses dérivées successives, on obtient :

$$\varphi_X^{(k)}(t) = i^k \int x^k e^{itx}dP_X(x). \quad \square$$

Proposition 4. Si la fonction caractéristique $\varphi_X(t)$ admet une dérivée d'ordre pair $2n$ à l'origine, alors les moments de X existent jusqu'à l'ordre $2n$. Autrement dit : si $\varphi_X^{(2n)}(0)$ existe, $n \in \mathbb{N}$, alors $E(X^k) < +\infty$ pour un entier $k \leq 2n$.

Remarque 8. Une conséquence immédiate de la proposition 4 est que si φ_X est indéfiniment dérivable en 0, X admet des moments de tous les ordres.

1.5 Travaux Dirigés

Exercice 1. Soient (Ω, \mathcal{U}) et (Ω', \mathcal{U}') deux espaces probabilisables, et $X : (\Omega, \mathcal{U}) \rightarrow (\Omega', \mathcal{U}')$ une application.

1. Montrer que $X^{-1}(\mathcal{U}')$ est une tribu des parties de Ω .
2. Soit C une famille des parties de Ω' . Montrer que $T(X^{-1}(C)) \subset X^{-1}(T(C))$, où $T(C)$ et $T(X^{-1}(C))$ désignent respectivement les tribus engendrées par C et $X^{-1}(C)$.

Exercice 2. Soit $\mathcal{T}(\Omega)$ une tribu sur Ω .

1. Montrer que $\emptyset \in \mathcal{T}(\Omega)$
2. Soit $(A_n, n \in \mathbb{N})$ une suite d'éléments de $\mathcal{T}(\Omega)$. Montrer que $\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{T}(\Omega)$.
3. Montrer que l'intersection de deux tribus est encore une tribu.
4. Soit $A \in \mathcal{T}(\Omega)$. On pose $\mathcal{T}_A(\Omega) = \{B \cap A, B \in \mathcal{T}(\Omega)\}$. Montrer que $\mathcal{T}_A(\Omega)$ est une tribu sur A .

Exercice 3. :

1. On considère un ensemble $E = \{1, 2, 3, 4\}$. La classe $C = \{\emptyset, E, \{1\}, \{2\}, \{3, 4\}, \{1, 2\}\}$ est-elle une tribu sur E ? Justifier votre réponse.
2. Soit E un espace et \mathbb{F} une tribu de parties de E . On suppose que $E = \mathbb{R}$ et on pose $\mathbb{A} = \{A \subset E, A \text{ est fini ou dénombrable ou } A^c \text{ est fini ou dénombrable}\}$ (\mathbb{A} est l'ensemble des parties qui sont finies ou dénombrables ou dont le complémentaire est fini ou dénombrable.).
Montrer que \mathbb{A} est une tribu et que $\mathbb{A} \neq P(\mathbb{R})$, où $P(\mathbb{R})$ est l'ensemble des parties de \mathbb{R} .

Exercice 4. Soient E et F deux espaces, f une application de E dans F et $T(F)$ une tribu sur F .

1. Montrer que $\Upsilon_f = \{f^{-1}(A), A \in T(F)\}$ est une tribu sur E .
2. La fonction $f : (E; \Upsilon_f) \rightarrow (F; T(F))$ est-elle une variable aléatoire ?
3. On pose $E = \{1; 2; 3; 4; 5\}$ et $C = \{\{1; 2\}; \{2; 3\}; \emptyset\}$.
Déterminer la tribu engendrée par la classe C de parties de E .

Exercice 5. :

1. Soit $(E, \mathcal{T}(E), P)$ un espace probabilisé et $(A_n)_{n \in \mathbb{N}}$ une suite de parties de E . On définit une nouvelle suite $(B_n)_{n \in \mathbb{N}}$ par $B_0 = A_0$ et $\forall n \in \mathbb{N}^*, B_n = A_n \cap (\bigcup_{k=0}^{n-1} A_k)^c$.

- i) Montrer que les B_n appartiennent à $\mathcal{T}(E)$, sont deux à deux disjoints, et vérifient :

$$\bigcup_{n \in \mathbb{N}} B_n = \bigcup_{n \in \mathbb{N}} A_n$$

- ii) En déduire l'inégalité : $P(\bigcup_{n \in \mathbb{N}} A_n) \leq \sum_{n \in \mathbb{N}} P(A_n)$
- iii) On suppose $\sum_{n \in \mathbb{N}} P(A_n) < +\infty$. En appliquant le résultat de la question 2 à la suite $(A_n)_{n \geq k}$ (pour $k \in \mathbb{N}$), et en faisant ensuite tendre k vers l'infini, montrer que l'on a la partie directe du lemme de Borel Cantelli :

$$P\left(\lim_{n \rightarrow +\infty} \sup A_n\right) = 0$$

On rappelle que $\lim_{n \rightarrow +\infty} \sup A_n = \bigcap_{k \in \mathbb{N}} \bigcup_{n \geq k} A_n$.

2. Enoncer et démontrer alors la partie indirecte du lemme de Borel-Cantelli.

Exercice 6. Soit $(\Omega, \mathcal{T}(\Omega), P)$ un espace probabilisé.

1. Montrer que si A et B sont indépendants, alors A et \bar{B} , \bar{A} et B , \bar{A} et \bar{B} sont aussi indépendants.
2. Montrer que deux événements $A, B \in \mathcal{T}(\Omega)$ sont indépendants si et seulement si

$$P(A \cap B)P(\bar{A} \cap \bar{B}) = P(\bar{A} \cap B)P(A \cap \bar{B})$$

Exercice 7. :

1. Soient p et q deux entiers naturels tels que $0 < q < p$.
 - i) Montrer que $\forall a > 0, a^q \leq 1 + a^p$.
 - ii) En déduire que si X est une v.a, $E(|X|^p) < +\infty \Rightarrow E(|X|^q) < +\infty$.
2. Si X et Y sont deux variables aléatoires,
 - i) montrer que $(E(XY))^2 \leq E(X^2)E(Y^2)$.
 - ii) Montrer que $|E(X)| \leq E(|X|)$.
 - iii) Montrer que si $E(X) = E(X^2) = 0$, alors $P(X = 0) = 1$.

Exercice 8. La consommation journalière d'une agglomération au cours du mois de septembre est une variable aléatoire X dont la densité a la forme : $f(t) = c(t-a)(b-t)1_{[a,b]}(t)$, où $t \in \mathbb{R}$; a, b et c sont des constantes strictement positives ($a < b$), avec la fonction indicatrice $1_{[a,b]}(t) = 1$ si $t \in [a, b]$ et 0 sinon.

1. Vérifier que l'on a pour tout $n \in \mathbb{N}$:

$$\int_a^b (t-a)^n (b-t) dt = \frac{(b-a)^{n+2}}{(n+1)(n+2)}.$$

2. Exprimer la constante c en fonction de a et b .
3. Calculer $E(X-a)$ et $E[(X-a)^2]$. En déduire $E(X)$ et $V(X)$.

4. Donner la fonction de répartition $F(x)$ de la variable aléatoire X (on distinguera les cas $x < a$, $a \leq x \leq b$ et $x > b$).
5. En notant X_i la consommation du i -ème jour et en supposant que les X_i sont indépendantes et de même loi que X , exprimer à l'aide de F et de n la fonction de répartition de la variable aléatoire $M_n = \max_{1 \leq i \leq n} X_i$.

Exercice 9. Soit X_1, X_1, \dots, X_n , n variables aléatoires indépendantes qui suivent chacune la loi uniforme sur $[\alpha, \beta]$ avec $0 < \alpha < \beta$.

1. Soit $i \in \{1, 2, \dots, n\}$. Calculer l'espérance $E(X_i)$ et la variance $\text{Var}(X_i)$.
2. Déterminer F_i la fonction de répartition de X_i , $i \in \{1, 2, \dots, n\}$.
3. On pose $Y = \min(X_1, X_1, \dots, X_n)$. Déterminer la fonction de répartition G de Y et en déduire sa densité de probabilité g .

Exercice 10. Calcul des moments de la loi gaussienne standard.

Soit X une variable aléatoire de loi normale $N(0, 1)$.

1. Que valent les $E(X^{2n+1})$ pour $n \in \mathbb{N}$?
2. On pose $c_n = E(X^{2n})$. Montrer en intégrant par parties que $\forall n \in \mathbb{N}, c_n = \frac{1}{2n+1} c_{n+1}$
3. En déduire une formule explicite pour $E(X^{2n})$.

Exercice 11. Soit $X = (X_1, X_2)$ un vecteur aléatoire à valeurs dans \mathbb{R}^2 de densité : $\forall (x_1, x_2) \in \mathbb{R}^2, f(x_1, x_2) = k \times \exp(-\frac{3}{16}x_1^2 + \frac{1}{4}x_1x_2 - \frac{1}{4}x_2^2)$

1. Déterminer $\mu = E(X)$ et \sum_X la matrice des variances et covariances de X . En déduire k .
2. Calculer le coefficient de corrélation linéaire des variables aléatoires réelles X_1 et X_2 .

Exercice 12. Soit X une variable aléatoire.

1. Montrer que X suit la loi normale $N(\mu, \Sigma)$ de dimension n si et seulement si $\forall a \in \mathbb{R}^n, a \neq 0, t_a X$ suit une loi normale unidimensionnelle.
2. Soit $X = (X_1, X_2)$ un vecteur aléatoire à valeurs dans \mathbb{R}^2 de densité : $\forall (x_1, x_2) \in \mathbb{R}^2, f(x_1, x_2) = k \times \exp(-\frac{1}{18}(5x_1^2 + 5x_2^2 - 8x_1x_2 - 14x_1 + 4x_2 + 17))$
 - i) Déterminer $\mu = E(X)$ et \sum_X la matrice des variances et covariances de X .
 - ii) En déduire k ; ainsi que le coefficient de corrélation linéaire entre X_1 et X_2 .

Exercice 13. Soit (X_1, X_2, \dots, X_n) un vecteur aléatoire de dimension n et désignant la transposée de X . On pose μ_X le vecteur moyenne de X et \sum_X la matrice des variances et covariances de X . On pose $Y = AX$ où A est une matrice $m \times n$.

1. Exprimer μ_Y et \sum_Y en fonction de μ_X et \sum_X .
2. On suppose que $n = 2$ et que X suit la loi normale de densité :
 $\forall (x_1, x_2) \in \mathbb{R}^2, f(x_1, x_2) = k \times \exp(-x_1^2 - x_2^2 + \frac{8}{5}x_1x_2 - \frac{26}{5}x_1 + \frac{28}{5}x_2 - \frac{41}{5})$
 - i) Déterminer k , $\mu_X = E(X)$ et \sum_X .
 - ii) On pose $U = 3X_1 + 2X_2$, $V = X_1 + 4X_2$ et $W = \begin{pmatrix} U \\ V \end{pmatrix}$.
Déterminer $\mu_W = E(W)$ et \sum_W .

Exercice 14. Soit $X = (X_1, X_2)$ un vecteur aléatoire à valeurs dans \mathbb{R}^2 de densité :
 $\forall (x_1, x_2) \in \mathbb{R}^2, f(x_1, x_2) = k \times \exp(-\frac{1}{2}x_1^2 - x_2^2 + \frac{1}{2}x_1x_2 + \frac{3}{2}x_1 + x_2 - 2)$

1. Déterminer k .
2. Déterminer $\mu = E(X)$ et \sum_X la matrice des variances et covariances de X .
3. En déduire la fonction génératrice des moments de X .

Exercice 15. Soit $X = (X_1, X_2, X_3)$ un vecteur aléatoire de \mathbb{R}^3 de densité :

$$\forall (x_1, x_2, x_3) \in \mathbb{R}^3,$$

$$f(x_1, x_2, x_3) = k \times \exp(-\frac{1}{4}(3x_1^2 + x_2^2 + 5x_3^2 - 2x_1x_2 - 6x_1x_3 + 2x_2x_3 + 2x_1 - 2x_2 - 10x_3 + 9))$$

1. Déterminer k , $\mu = E(X)$ et \sum_X la matrice des variances et covariances de X .
2. En déduire les coefficients de corrélation linéaire de (X_1, X_2) , (X_1, X_3) et (X_2, X_3) .
3. Calculer la fonction génératrice des moments de X .

Exercice 16. .

Soit X une variable aléatoire réelle de loi exponentielle de paramètre $q > 0$ ($f_X(t) = qe^{-qt}$ si $t \geq 0$, $f_X(t) = 0$ sinon).

1. Montrer que la fonction caractéristique de X est $\varphi_X(t) = \frac{q}{q - it}$, $t \in \mathbb{R}$.
2. En déduire le moment d'ordre 2 de X .
3. On note $Y = [X]$ sa partie entière. Montrer que Y suit la loi géométrique de paramètre e^{-q} .

Exercice 17. Soit X une variable aléatoire réelle.

1. Si X suit la loi uniforme sur $[a, b]$,
 - a) Calculer son espérance mathématique.
 - b) Montrer que sa variance vaut $V(X) = \frac{(b-a)^2}{12}$.
2. Si X suit la loi normale $N(\mu, \sigma)$, montrer (en utilisant les intégrales) que sa fonction génératrice des moments est $m_X(t) = e^{t\mu + \frac{1}{2}\sigma^2 t^2}$.

Exercice 18. Soit $X = (X_1, X_1, \dots, X_n)$ un vecteur aléatoire normal à valeurs dans \mathbb{R}^n , donc qui suit la loi normale $N^n(\mu, \sum_X)$ où $\sum_X = (c_{ij})_{1 \leq i, j \leq n}$ est sa matrice des variances et covariances. On rappelle que la fonction caractéristique de X est alors :

$$\forall t = (t_1, t_2, \dots, t_n) \in \mathbb{R}^n, \varphi_X(t) = \exp(it\mu - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n c_{ij} t_i t_j)$$

1. Montrer que si $n = 2$, la loi normale centrée réduite à deux dimensions X a pour fonction caractéristique $\varphi_X(t_1, t_2) = \exp(-\frac{1}{2}(t_1^2 + 2\rho t_1 t_2 + t_2^2))$ où ρ est le coefficient de corrélation linéaire des deux variables.
2. Montrer que $E(X_1) = 0$, $E(X_2) = 0$ et $E(X_1 X_2) = \rho$.

Chapitre 2

INTRODUCTION AUX PROCESSUS STOCHASTIQUES ET AUX CHAINES DE MARKOV

2.1 Généralités sur les processus stochastiques

2.1.1 Introduction

Les processus permettent de modéliser des systèmes dont le comportement n'est que partiellement prévisible. La théorie est fondée sur le calcul des probabilités et les statistiques. Les domaines d'application sont très divers surtout dans la modélisation et la gestion du trafic dans les réseaux de transport (carrefours à feux) et plus généralement la gestion des systèmes techniques complexes soumis à des perturbations aléatoires.

L'étude mathématique d'un système d'attente, se fait le plus souvent par l'introduction d'un processus stochastique approprié. Généralement, on qualifie de processus stochastique tout phénomène d'évolution temporelle dont l'analyse peut être soumise au calcul des probabilités. Du point de vue de l'observation, un processus stochastique est constitué par l'ensemble de ses réalisations. Une réalisation est obtenue par l'expérience qui consiste à enregistrer une suite d'événements au cours du temps. Le caractère aléatoire de l'évolution se montre par le fait que la répétition de l'expérience conduit à une autre séquence temporelle.

2.1.2 Notations et Définitions

Définition 1. 1) *Un processus stochastique est un ensemble des phénomènes produit par le hasard dans le temps, et dont l'évolution peut être décrite à l'aide de variable aléatoire.*

2) *Mathématiquement, un processus stochastique $(X_t)_{t \in \mathbb{T}}$ est une collection de variables aléatoires indexées par un paramètre t et définies sur un même espace de probabilités $(\Omega, \mathcal{T}(\Omega), P)$ et à valeurs dans un espace probabilisable $(E, \mathcal{T}(E))$.*

La variable X_t représente l'état du processus au temps t et l'ensemble de toutes les valeurs possibles E (partie de \mathbb{R}) est appelée espace des états du processus.

Définition 2. 1) Une suite de v.a.r $(X_n)_{n \in \mathbb{N}}$ sur $(\Omega, \mathcal{T}(\Omega), P)$ converge en loi vers une v.a.r X si $\lim_{n \rightarrow +\infty} F_{X_n}(x) = F_X(x)$, où F_{X_n} désigne la fonction de répartition de X_n .

2) Une suite de v.a.r $(X_n)_{n \in \mathbb{N}}$ sur $(\Omega, \mathcal{T}(\Omega), P)$ converge en probabilité vers une v.a.r X si $\forall \varepsilon > 0, \lim_{n \rightarrow +\infty} P(|X_n - X| \geq \varepsilon) = 0$.

On montre que la convergence en probabilité implique la convergence en loi.

2.1.3 Classification des processus stochastiques

La classification d'un processus stochastique dépend de trois quantités :

- l'espace d'état E
- l'indice t appartenant à l'espace temporel \mathbb{T} qui est soit une partie de \mathbb{N} , soit une partie de \mathbb{R}
- la dépendance statistique entre les variables aléatoires pour différentes valeurs de l'indice .

Lorsque la variable X_t peut prendre un nombre fini ou dénombrable de valeurs, le processus est dit à espace d'état discret. Dans le cas contraire, il est dit à espace d'état continu.

NB : On travaillera en temps discret (les indices sont des entiers) et on supposera de plus que l'espace d'état E dans lequel le processus prend ses valeurs est au plus dénombrable.

Relativement à la notion de probabilité conditionnelle $P(B|A)$, nous nous intéresserons au cas où $A = \{X_0 = x_0, \dots, X_n = x_n\}$ et $B = \{X_{n+1} = x_{n+1}\}$.

2.1.4 Exemples

Les processus stochastiques sont appliqués dans plusieurs domaines courants de la vie socio-économique et politique :

1. **Météorologie** : X_n = désigne la température en un lieu donné au jour n , ou encore la hauteur des précipitations au courant de l'année n .
2. **Bourse de valeurs** : X_n = désigne le cours d'une action le jour n , ou le chiffre d'affaire d'une société l'année n .
3. **Assurance** : X_n = désigne le montant total des indemnités versées par une compagnie d'assurance pour des sinistres survenus le mois n .
4. **Epidémiologie** : X_n = désigne le nombre d'individus infectés par une maladie contagieuse au bout de n jours.
5. **Jeux** : X_n = désigne l'ordre des cartes d'un jeu qui a été battu n fois.

2.2 Chaines de Markov dans un espace d'états au plus dénombrable

La catégorie des processus stochastiques est très large car elle permet de modéliser l'ensemble des phénomènes aléatoires. Cependant, on ne peut pas obtenir mathématiquement beaucoup de résultats sur un processus stochastique trop général. Il faut alors ignorer des hypothèses supplémentaires pour faciliter leur résolution.

En 1906, Andrei Andreyevich Markov a commencé l'étude d'un nouveau type de processus stochastique où le résultat d'une expérience donnée peut affecter le résultat de l'expérience suivante.

Dans ce cours, on propose de se restreindre à la modélisation des phénomènes aléatoires par des chaînes de Markov. Il est donc nécessaire de vérifier deux hypothèses mathématiques que sont la propriété de Markov et l'homogénéité.

2.2.1 Définitions et propriétés de Markov

Définition 3. *Intuitivement, un processus stochastique $(X_n)_{n \in \mathbb{N}}$ vérifie la propriété de Markov lorsque son évolution après une date n , ne dépend du passé $X_0, X_1, X_2, \dots, X_n$ qu'à travers sa position au temps n (et non pas du trajet qu'il a suivi pour atteindre cet état).*

Définition 4. *Formellement, on dit que $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov (donc vérifie la propriété de Markov) s'il existe des nombres réels positifs $(P_{ij})_{i,j \in E}$ avec $\sum_{j \in E} P_{ij} = 1$ pour tout $i \in E$, tels que pour tout entier n , tous $i, j, x_0, \dots, x_n \in E$:*

$$P(X_{n+1} = j | X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = i) = P(X_{n+1} = j | X_n = i) = P_{ij}$$

$(P_{ij})_{j \in E}$ désigne alors la probabilité de transition pour l'état i . Si l'espace d'état E est au plus dénombrable, $(P_{ij})_{i,j \in E}$ est alors appelée matrice de transition de la chaîne.

Proposition 1. *Si $x_0, x_1, \dots, x_n \in E$, alors on a*

$$P(X_1 = x_1, \dots, X_n = x_n | X_0 = x_0) = P_{x_0, x_1} P_{x_1, x_2} \dots P_{x_{n-1}, x_n}$$

En particulier pour un espace d'état $E = \{1, 2, \dots, n\}$ au plus dénombrable, $P(X_n = j | X_0 = i) = P_{ij}^n$

2.2.2 Chaînes de Markov homogènes et matrice stochastique

Définition 5. *La matrice stochastique (ou de transition) d'une chaîne de Markov est une matrice carrée telle que :*

- (i) *Son ordre est égal au cardinal de l'espace des états E*
- (ii) *Chaque élément de la matrice est une probabilité*
- (iii) *Chaque ligne représente une loi de probabilité (la somme des éléments de chaque ligne vaut 1).*

Définition 6. *La matrice stochastique d'une chaîne de Markov est la matrice d'une forme quadratique \mathcal{P} définie de $E \times E$ vers $[0, 1]$ et encore appelée noyau de transition. par : $\mathcal{P}(x, y) = p(X_{n+1} = y | X_n = x)$.*

On montre alors que $\sum_{y \in E} \mathcal{P}(x, y) = 1$ pour tout $x, y \in E$.

Théorème 1. Une chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ est caractérisée par :

- (i) L'espace d'états $E = \{x_1, \dots, x_m\}$;
- (ii) A chaque instant n , la matrice de transition \mathbb{P}_n modélisant le passage à l'état x_{n+1} sachant l'état x_n est donnée par : $\mathbb{P}_n = (P(X_{n+1} = j / X_n = i))_{i,j \in E} = (P_{ij})_{i,j \in E}$
- (iii) La distribution initiale π_0 (c'est une loi de probabilités sur E à l'instant $n = 0$).

Définition 7. 1) Intuitivement, une chaîne de Markov homogène est une chaîne telle que la probabilité qu'elle a pour passer dans un certain état au même état soit indépendante du temps. En d'autres termes, la loi de probabilité caractérisant la prochaine étape ne dépend pas du temps (de l'étape précédente), et en tout temps la loi de probabilité à tout moment de la chaîne est toujours la même pour caractériser la transition à l'étape en cours.

2) Formellement, une chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ à dimension finie est dite homogène lorsque : $\forall n \in \mathbb{N}, \mathbb{P}_n = \mathbb{P}$ (i.e la matrice de transition \mathbb{P}_n reste invariante quelque soit l'instant de transition n).

Théorème 2. Une chaîne de Markov homogène $(X_n)_{n \in \mathbb{N}}$ est caractérisée par :

- (i) L'espace d'états $E = \{x_1, \dots, x_m\}$;
- (ii) La matrice de transition \mathbb{P} ;
- (iii) La distribution initiale π_0

La forme analytique de la matrice de transition \mathbb{P} est alors :

$$\mathbb{P} = \begin{pmatrix} P(X_{n+1} = x_1 | X_n = x_1) & P(X_{n+1} = x_2 | X_n = x_1) & \dots & P(X_{n+1} = x_m | X_n = x_1) \\ P(X_{n+1} = x_1 | X_n = x_2) & P(X_{n+1} = x_2 | X_n = x_2) & \dots & P(X_{n+1} = x_m | X_n = x_2) \\ \vdots & \vdots & \ddots & \vdots \\ P(X_{n+1} = x_1 | X_n = x_m) & P(X_{n+1} = x_2 | X_n = x_m) & \dots & P(X_{n+1} = x_m | X_n = x_m) \end{pmatrix}$$

$$= \begin{pmatrix} P(x_1|x_1) & P(x_2|x_1) & \dots & P(x_m|x_1) \\ P(x_1|x_2) & P(x_2|x_2) & \dots & P(x_m|x_2) \\ \vdots & \vdots & \ddots & \vdots \\ P(x_1|x_m) & P(x_2|x_m) & \dots & P(x_m|x_m) \end{pmatrix}$$

où encore

$$\mathbb{P} = \begin{pmatrix} P(x_1, x_1) & P(x_1, x_2) & \dots & P(x_1, x_m) \\ P(x_2, x_1) & P(x_2, x_2) & \dots & P(x_2, x_m) \\ \vdots & \vdots & \ddots & \vdots \\ P(x_m, x_1) & P(x_m, x_2) & \dots & P(x_m, x_m) \end{pmatrix}$$

NB : La première écriture est appelée **écriture conditionnelle** tandis que la deuxième est l'**écriture conventionnelle**.

2.2.3 Représentation graphique d'une chaîne de Markov

L'évolution d'une chaîne de Markov peut être représentée graphiquement sous la forme d'un graphe orienté G ayant pour sommet les points i et pour arêtes les

couples orientés (i, j) . Le graphe est donc défini à partir d'un ensemble d'états et d'une matrice de transition. Les noeuds du graphe représentent les états possibles dans une chaîne de Markov, et les flèches représentent les transitions d'un état à un autre.

2.2.4 Théorème de Chapman-Kolmogorov

Exemple 1. (Chaîne à deux états) : **Etat d'une ligne de téléphone**

Considérons l'état d'une ligne de téléphone $X_n = 0$ si la ligne est libre à l'instant n , et $X_n = 1$ si la ligne est occupée. Supposons que sur chaque intervalle de temps, il y a une probabilité p qu'un appel arrive (un appel au plus). Si la ligne est déjà occupée, l'appel est perdu. Supposons également que si la ligne est occupée au temps n , il y a une probabilité q qu'elle se libère au temps $n + 1$. Cherchons la matrice de transition de ce processus stochastique.

Solution : On peut alors modéliser une chaîne de Markov à valeurs dans $E = \{0, 1\}$, et ayant pour matrice de transition

$$\mathbb{P} = \begin{pmatrix} P(0|0) & P(1|0) \\ P(0|1) & P(1|1) \end{pmatrix} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

où $P(0|0)$ désigne la probabilité qu'à l'instant $n + 1$, la ligne soit libre sachant qu'elle était également libre à l'instant n ;

$P(0|1)$ est la probabilité qu'à l'instant $n + 1$, la ligne soit libre sachant qu'elle était occupée à l'instant n ;

$P(1|0)$ est la probabilité qu'à l'instant $n + 1$, la ligne soit occupée sachant qu'elle était libre à l'instant n ;

$P(1|1)$ est la probabilité qu'à l'instant $n + 1$, la ligne soit occupée sachant qu'elle était également occupée à l'instant n .

Exemple 2. (Chaîne à trois états) : **File d'attente simple**

On modifie l'exemple précédent en supposant qu'on peut mettre un appel en attente. Les appels arrivent et la ligne se libère comme avant. Si un appel arrive pendant que la ligne est occupée et si le système n'est pas saturé, l'appel est mis en attente. Si un appel arrive alors qu'il y a déjà un en attente, il est perdu. Déterminons la matrice de transition de ce processus stochastique.

Solution : L'espace d'état est $E = \{0, 1, 2\}$ et $X_n = 2$ si on peut mettre un appel qui arrive en attente. Le cas où il y a exactement un appel retenu au temps n est un peu plus délicat. En effet, la situation "l'appel se termine et pas d'appel nouveau arrive" correspond à $P(0|1) = q(1 - p)$; et $P(2|1) = p(1 - q)$ est lorsqu'un appel nouveau arrive et celui en cours continue. Comme $P(1|0) + P(1|1) + P(1|2) = 1$, on trouve finalement

$$\mathbb{P} = \begin{pmatrix} P(0|0) & P(1|0) & P(2|0) \\ P(0|1) & P(1|1) & P(2|1) \\ P(0|2) & P(1|2) & P(2|2) \end{pmatrix} = \begin{pmatrix} 1-p & p & 0 \\ q(1-p) & 1-q(1-p)-p(1-q) & p(1-q) \\ 0 & q & 1-q \end{pmatrix}$$

Comme exemple d'interprétation, $P(2|1)$ est la probabilité qu'un appel soit en attente à l'instant $n + 1$ sachant que la ligne était occupée à l'instant n .

Soit π_n la loi (distribution) de probabilité de la variable aléatoire X_n . π_n est encore la distribution de probabilités d'états pour une chaîne de Markov.

Théorème 3. : (Chapman-Kolmogorov)

Soit $X_n[E, \mathbb{P}, \pi_0]$, une chaîne de Markov homogène. Alors la distribution à l'instant n est donnée par la formule : $\pi_n = \pi_{n-1}\mathbb{P} = \pi_{n-2}\mathbb{P}^2 = \dots = \pi_0\mathbb{P}^n$

2.2.5 Etats possibles d'une chaîne de Markov

Définition 8. On appelle état accessible à partir de l'état $x_i \in E$ tout état $x_j \in E$ d'une chaîne de Markov tel que il existe un entier n , $(P(X_{n+1} = x_j | X_n = x_i))^n > 0$. Deux états mutuellement accessibles l'un à partir de l'autre sont dits communicants.

Définition 9. Désignons par T_{ij} le temps de premier passage de l'état x_i à l'état x_j (ou temps de premier retour en dans une chaîne de Markov discrète : $T_{ij} = \min\{k, X_k = x_j | X_0 = x_i\}$

Si $i = j$ alors T_i est le temps de premier retour en x_i .

1) L'état x_i est dit état récurrent si $p(T_{ij} < \infty) = 1$

Les états récurrent sont dits aussi persistants. Un état récurrent est donc visité un nombre infini de fois.

2) x_i est dit transitoire si $p(T_{ij} < \infty) < 1$

Une chaîne dont tous les états sont récurrents est dite récurrente ; une chaîne dont tous les états sont transitoires est dite transiente.

Définition 10. Un état $x_j \in E$ est absorbant si la probabilité de rester dans cet état lorsqu'on y parvient est 1 ($P(X_{n+1} = x_j | X_n = x_j) = 1$)

Définition 11. Une chaîne est dite irréductible si tout état est accessible à partir de n'importe quel autre état, avec une probabilité non nulle.

Remarque 1. – Toute chaîne possédant un état absorbant n'est pas irréductible.
 – Les états périodiques sont des états qui ne sont atteints que périodiquement.
 – Si tous les états d'une chaîne sont communicants mutuellement, alors elle est irréductible.

2.2.6 Calcul du temps moyen de retour en $x_j \in E$

Définition 12. La probabilité de première transition en n unités de temps, de l'état x_i à l'état x_j , notée $f_{ij}^{(n)}$, est définie par :

$$f_{ij}^{(n)} = P(T_j = n | X_0 = x_i) = P(X_n = x_j, X_{n-1} \neq x_j, \dots, X_1 \neq x_j | X_0 = x_i)$$

$f_{ii}^{(n)}$ est la probabilité de premier retour à l'état x_i en n unités de temps.

Théorème 4. x_i est récurrent si et seulement si $\sum_{n=1}^{\infty} f_{ii}^{(n)} = 1$
 x_i est transitoire si et seulement si $\sum_{n=1}^{\infty} f_{ii}^{(n)} < 1$

Définition 13. 1) On appelle temps moyen de retour en x_j , la quantité :

$$\mu_j = E(T_j | X_0 = x_j) = \sum_n n p_{jj}^{(n)}$$

2) Un état x_j est récurrent positif (ou non nul) s'il est récurrent et si le temps moyen μ_j de retour est fini. Dans le cas où le temps moyen de retour est infini, l'état est dit récurrent nul.

2.2.7 Comportement asymptotique des chaînes de Markov homogènes

Définition 14. Un vecteur de probabilités $v = (v_1, \dots, v_m)$ sur E (les coefficients de v sont positifs et de somme 1) est dit vecteur de probabilités invariant (distribution invariante ou vecteur de probabilités d'équilibre), pour la chaîne de matrice de transition \mathbb{P} , si v est un vecteur propre à gauche de valeur propre 1 ($v = v\mathbb{P}$).

Remarque 2. On voit que si v est un vecteur de probabilités invariant et si l'état initial de la chaîne X_0 suit la loi v , alors X_n suit également la loi v pour tout n (puisque la loi de X_n est $v\mathbb{P}^n = v$).

Proposition 2. Il existe toujours un vecteur de probabilités invariant dans un processus stochastique markovien à dimension fini.

Les résultats suivants donnent les conditions d'unicité de ce vecteur de probabilités invariant.

Lemme 1. Soit $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov sur un espace à m éléments et ayant pour matrice de transition \mathbb{P} . On suppose que les coefficients de \mathbb{P} sont tous strictement positifs. Alors il existe un unique vecteur de probabilités invariant $v = (v_1, \dots, v_m)$ telle que quelque soit la distribution initiale de la chaîne π_0 , X_n converge en loi vers v quand n tend vers $+\infty$.

Théorème 5. Soit $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov sur un espace d'état fini et ayant pour matrice de transition \mathbb{P} . On suppose qu'il existe un entier k pour lequel les coefficients de \mathbb{P}^k sont tous strictement positifs (\mathbb{P} est dite régulière). Alors il existe un unique vecteur de probabilités invariant v tel que quelque soit la distribution initiale de la chaîne π_0 , X_n converge en loi vers v quand n tend vers $+\infty$ (i.e, $\forall i, j \in E$, $\lim_{n \rightarrow +\infty} P(X_n = j | X_0 = i) = v_j$).

La matrice \mathbb{P} est alors dite régulière.

Preuve : Considérons une loi initiale π_0 , fixons $r \in \{0, 1, \dots, k\}$ et notons $\pi_0 u = \pi_0 \mathbb{P}^r$. On a $\pi_0 \mathbb{P}^{km+r} = \pi_0 \mathbb{P}^{km}$, et d'après le lemme ci-dessus, $\pi_0 \mathbb{P}^{km+r}$ converge vers v quand m tend vers $+\infty$. Puisque cette limite ne dépend pas de r (quelque soit la distribution initiale), X_n converge en loi vers le vecteur propre probabilité (à gauche) associé à la valeur propre 1 de \mathbb{P}^k .

Définition 15. Une chaîne de Markov homogène $X_n[E, \mathbb{P}, \pi_0]$ est dite stationnaire si la distribution n'évolue pas dans le temps. Dans ce cas, π_0 est une distribution invariante (ou stationnaire) de cette chaîne.

Théorème 6. 1) On pose $\pi = (\pi_1, \pi_1, \dots, \pi_i, \dots)$ une distribution stationnaire, avec $\pi_i = p(X_\infty = x_i)$.

Alors pour tout état récurrent $x_i : \pi_i = \frac{1}{\mu_i}$,
où μ_i est le temps moyen de retour à x_i .

2) Une chaîne transiente infinie n'admet pas de distribution stationnaire.

3) Si la chaîne est irréductible, elle possède une distribution stationnaire unique $\pi = (\frac{1}{\mu_i})_{x_i \in E}$ si et seulement si tous les états sont récurrents positifs.

4) Toute distribution limite π^* d'une chaîne, si elle existe, est l'unique distribution stationnaire, nécessairement portée par les états récurrents.

Définition 16. : (Propriétés d'Ergodicité d'une chaîne de Markov)

1) Si $\lim_{n \rightarrow +\infty} \mathbb{P}^n = \mathbb{P}^\infty$ où \mathbb{P}^∞ est une matrice stochastique ne possédant aucun élément nul, on dit que le système est ergodique, ou qu'il est stable en probabilité, ou encore qu'il possède un régime permanent, et est dite matrice ergodique.

2) Un état récurrent positif et apériodique est dit ergodique.

3) Une chaîne irréductible, apériodique et récurrente positive est dite ergodique.

Théorème 7. (Chacon-Ornstein)

Toute chaîne ergodique possède une distribution limite unique $\pi^* = (\pi_i^*)_i$ qui s'identifie à l'unique distribution stationnaire $\pi = (\pi_i)_i$ de la chaîne ;

$\forall x_i \in E, \lim_{n \rightarrow +\infty} p_i^{(n)} = \pi_i^* = \pi_i = \frac{1}{\mu_i}$, où $\pi = \pi \mathbb{P}$.

Remarque 3. – Il existe plusieurs situations où les chaînes de Markov tendent vers un équilibre du système ; c'est à dire une chaîne dont la distribution évolue et se stabilise à partir d'un certain rang.

– Etant donnée une chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ de matrice de transition \mathbb{P} , on cherche à étudier le comportement de la distribution de X_n quand $n \rightarrow +\infty$. Ce qui revient à étudier la suite de matrice $(\mathbb{P}^n)_{n \in \mathbb{N}}$.

Exemple 3. Reprenons le phénomène aléatoire précédent sur la ligne téléphonique de matrice de transition

$$\mathbb{P} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

avec $0 < p$ et $q < 1$. On cherche alors une expression simplifiée de \mathbb{P}^n pour calculer facilement sa limite.

On peut diagonaliser \mathbb{P} et on montre que ses valeurs propres sont 1 et $1-p-q$.

On a alors $D = Q^{-1} \mathbb{P} Q$ avec

$$Q = \begin{pmatrix} 1 & -p \\ 1 & q \end{pmatrix}; Q^{-1} = \begin{pmatrix} q/(p+q) & p/(p+q) \\ -1/(p+q) & 1/(p+q) \end{pmatrix}; D = \begin{pmatrix} 1 & 0 \\ 0 & 1-p-q \end{pmatrix}.$$

Par ailleurs $\mathbb{P}^n = (QDQ^{-1})^n = QD^nQ^{-1}$; et un calcul simple donne

$$\mathbb{P}^n = \begin{pmatrix} (q + p(1 - p - q)^n)/(p + q) & (p - p(1 - p - q)^n)/(p + q) \\ (q - q(1 - p - q)^n)/(p + q) & (p + q(1 - p - q)^n)/(p + q) \end{pmatrix}$$

Comme $|1 - p - q| < 1$, on voit que

$$\lim_{n \rightarrow +\infty} \mathbb{P}^n = \begin{pmatrix} q/(p + q) & p/(p + q) \\ q/(p + q) & p/(p + q) \end{pmatrix} = \mathbb{P}^\infty$$

2.2.8 Exemple d'application

On dispose d'une parcelle de terrain dont les végétations potentielles sont : herbe (h), arbuste (a) et forêt (f). On observe régulièrement la végétation après chaque trois ans ; et on obtient alors les informations suivantes sur la parcelle de terrain :

- S'il y a de l'herbe, 3 ans plus tard, on aura une végétation composée de 50 pour cent d'herbe, 45 pour cent arbuste et 5 pour cent de forêt.
- S'il y a des arbustes, 3 ans plus tard, on aura 10 pour cent d'herbe, 50 pour cent d'arbuste et 40 pour cent de forêt.
- S'il y a de la forêt, 3 ans plus tard, on aura 10 pour cent d'arbuste et 90 pour cent de forêt.

1. Donner l'ensemble E des états de ce processus Markovien.
2. Montrer que la matrice de passage de l'instant initial à l'instant suivant est

$$\mathbb{P}_0 = \begin{pmatrix} 0.5 & 0.45 & 0.05 \\ 0.1 & 0.5 & 0.4 \\ 0 & 0.1 & 0.9 \end{pmatrix}$$

3. On suppose que l'évolution de cette végétation peut être modélisée par une chaîne de Markov homogène de distribution initiale $\pi_0(h, a, f) = (1, 0, 0)$. Calculer π_1, π_2, π_3 .
4. On suppose que $\forall n \geq 25$,

$$\mathbb{P}^n = \begin{pmatrix} 0.0377 & 0.1887 & 0.7736 \\ 0.0377 & 0.1887 & 0.7736 \\ 0.0377 & 0.1887 & 0.7736 \end{pmatrix}$$

Montrer alors qu'au bout de 75 ans, la distribution se stabilise autour de $\pi_\infty = (0.0377, 0.1887, 0.7736)$.

2.3 Travaux Dirigés

Exercice 1. Une boulangerie dispose d'un seul vendeur. On suppose que l'arrivée de sa clientèle décrit l'évolution d'un phénomène aléatoire à travers une suite de variables aléatoires réelles $(X_n)_{n \in \mathbb{N}}$. Ainsi, $X_n = 0$ signifie que le vendeur est libre à l'instant n , et $X_n = 1$ signifie que le vendeur est occupé à l'instant n . Supposons que sur chaque intervalle de temps, la probabilité qu'un client arrive est q (un patient au plus), et doit rentrer si le vendeur est déjà occupé. Supposons également que si le vendeur est occupé à l'instant n , il y a une probabilité p qu'il se libère à l'instant $n + 1$.

- 1) Evaluer la matrice de transition \mathbb{P} de cette chaîne de Markov ainsi modélisée.
- 2) Etudier son comportement asymptotique, c'est à dire, montrer que :

$$\lim_{n \rightarrow +\infty} \mathbb{P}^n = \begin{pmatrix} p/(p+q) & q/(p+q) \\ p/(p+q) & q/(p+q) \end{pmatrix}; \text{ avec } 0 < q \text{ et } p < 1.$$

Exercice 2. 1. On considère une chaîne de Markov homogène de matrice de transition

$$\mathbb{P} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Montrer que \mathbb{P} admet une infinité de vecteurs de distribution invariante.

$$2. \text{ On suppose maintenant que } \mathbb{P} = \begin{pmatrix} \frac{1}{5} & \frac{1}{5} & \frac{3}{5} \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \end{pmatrix}$$

Montrer que $\pi = (\frac{5}{11}, \frac{2}{11}, \frac{4}{11})$ est un vecteur de distribution invariante de \mathbb{P} .

Exercice 3. On considère une marche aléatoire dont on peut modéliser par une chaîne de Markov homogène $(X_n)_{n \in \mathbb{N}}$ à trois états $E = \{a, b, c\}$ de matrice de transition :

$$\mathbb{P} = \begin{pmatrix} \frac{1}{4} & 0 & \frac{3}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix} \text{ telle que } \lim_{n \rightarrow +\infty} \mathbb{P}^n = \begin{pmatrix} \frac{1}{4} & \frac{3}{10} & \frac{9}{20} \\ \frac{1}{4} & \frac{3}{10} & \frac{9}{20} \\ \frac{1}{4} & \frac{3}{10} & \frac{9}{20} \end{pmatrix}$$

On suppose que la distribution initiale à l'instant $n = 0$ est $\pi_0 = (1, 0, 0)$.

1. Montrer que $X_n[E, \mathbb{P}, \pi_0]$ n'est pas une chaîne stationnaire.
2. Montrer alors que pour $\pi_0^* = (\frac{1}{4}, \frac{3}{10}, \frac{9}{20})$, la chaîne de Markov homogène $X_n[E, \mathbb{P}, \pi_0^*]$ est stationnaire
3. En déduire une distribution invariante de \mathbb{P} .

Exercice 4. On considère une marche aléatoire dont on peut modéliser par une chaîne de Markov homogène $(X_n)_{n \in \mathbb{N}}$ à trois états $E = \{a, b, c\}$ de matrice de transition :

$$\mathbb{P} = \begin{pmatrix} \frac{1}{4} & 0 & \frac{3}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$$

1. Trouver le vecteur probabilité invariante V (encore appelé distribution invariante ou d'équilibre) pour cette chaîne de Markov homogène.
2. Etudier alors le comportement asymptotique de cette chaîne de Markov homogène, c'est à dire calculer $\lim_{n \rightarrow +\infty} \mathbb{P}^n$.

Exercice 5. Doudou le paresseux.

Doudou, le paresseux, ne connaît que trois endroits dans sa cage : (i) les copeaux où il dort ; (ii) la mangeoire où il mange ; (iii) la roue où il s'amuse. Ses journées sont assez semblables les une aux autres, et son activité se modélise aisément par une chaîne de Markov homogène. Toutes les minutes, il peut soit changer d'activité, soit continuer celle qu'il était entrain de faire.

- Quand il dort, il a 9 chances sur 10 de ne pas se réveiller la minute suivante. Et lorsqu'il se réveille, il y a 1 chance sur 2 qu'il aille manger et 1 chance sur 2 qu'il parte s'amuser.
- Le repas ne dure qu'une minute, après il fait autre chose. Ainsi, après avoir manger, il y a 3 chances sur 10 qu'il parte s'amuser dans la roue, mais surtout 7 chances sur 10 qu'il retourne dormir.
- Quand il s'amuse, il y a 8 chances sur 10 qu'il retourne dormir au bout d'une minute. Sinon il continue à s'amuser.

On suppose qu'à l'instant $n = 0$, l'état initial de la distribution est $\pi_0 = (a, b, c)$ (c'est à dire qu'au début des observations, on avait a pour cent de chance de trouver Doudou entrain de dormir, b pour cent de chance de trouver Doudou entrain de manger et c pour cent de chance de trouver Doudou entrain de s'amuser).

1. Donner l'ensemble E des états de ce processus Markovien .
2. Quelle est la matrice de transition de ce processus ?
3. Quelles sont les chances de trouver Doudou entrain de dormir, manger ou s'amuser, respectivement après une et deux minutes dès le début de l'observation de ce dernier ?
4. Retrouver de deux manière différentes la distribution initiale $\pi_0^* = (0.884, 0.0442, 0.0718)$ pour que la chaîne de Markov homogène $X_n[E, \mathbb{P}, \pi_0^*]$ soit stationnaire.

Exercice 6. Modélisation d'un brin d'ADN

La modélisation incomplète la plus simple d'un brin d'ADN, enchainement "désordonné" de nucleotides de l'un des 3 types adenine (a), cytosine (c) et guanine (g), est de le considérer comme une trajectoire d'une chaîne de Markov homogène $(X_n)_{n \in \mathbb{N}}$ à trois états $E = \{a, c, g\}$ dont la matrice de transition \mathbb{P} fournit les probabilités que l'un de ces états succède à un autre. Si le pas de temps entre chaque observation est 3 minutes, on obtient alors les informations suivantes :

- Si c'est de l'adenine, 3 minutes plus tard notre brin d'ADN sera constitué de 50 pour cent d'adenine, 45 pour cent de cytosine et 5 pour cent de guanine.
- Si c'est de la cytosine, 3 minutes plus tard notre brin d'ADN sera constitué de 10 pour cent d'adenine, 50 pour cent de cytosine et 40 pour cent de guanine.
- Si c'est de la guanine, 3 minutes plus tard notre brin d'ADN sera constitué de 10 pour cent de cytosine et 90 pour cent de guanine.

1. Donner alors la matrice de transition \mathbb{P} de passage d'un instant à l'autre.
2. On suppose qu'à l'instant $n = 0$, l'état initial de la distribution de probabilité est $\pi_0 = (1, 0, 0)$ (c'est à dire qu'au début des observations, notre brin d'ADN sera constitué de 100 pour cent d'adénine, zero pour cent de cytosine et zero pour cent de guanine). La distribution à l'instant n étant donnée par la formule $\pi_n = \pi_{n-1}\mathbb{P} = \pi_0\mathbb{P}^n$ pour $n \geq 1$,
 - i) Calculer les distributions au bout de $3mn$, $6mn$, $75mn$, $78mn$ et $81mn$, correspondant respectivement à π_1 , π_2 , π_{25} , π_{26} , π_{27} ; sachant que

$$\mathbb{P}^{25} = \begin{pmatrix} 0.0377 & 0.1887 & 0.7736 \\ 0.0377 & 0.1887 & 0.7736 \\ 0.0377 & 0.1887 & 0.7736 \end{pmatrix}$$

Que constate-t-on ?

- ii) En déduire alors la matrice $\lim_{n \rightarrow +\infty} \mathbb{P}^n$
3. Une distribution λ est dite invariante si on a $\lambda\mathbb{P} = \lambda$. La chaîne de Markov homogène $(X_n)_{n \in \mathbb{N}}$ de distribution initiale λ (encore notée $X_n[E, \mathbb{P}, \lambda]$) est alors dite stationnaire.
 - i) Montrer que $X_n[E, \mathbb{P}, \pi_0]$ n'est pas une chaîne stationnaire.
 - ii) Montrer que pour la distribution initiale $\pi_0^* = (0.0377, 0.1887, 0.7736)$, la chaîne de Markov homogène $X_n[E, \mathbb{P}, \pi_0^*]$ est stationnaire.

Exercice 7. L'évolution de la végétation sur une parcelle de terrain présente quatre types de végétation : sol nu (sn), herbe (h) arbuste (a) et forêt (f) ; et est considérée comme une trajectoire d'une chaîne de Markov homogène $(X_n)_{n \in \mathbb{N}}$ à quatre états $E = \{sn, h, a, f\}$. On suppose que le pas de temps entre deux observations est de 3 ans et que les probabilités que l'un de ces états succède à un autre sont évaluées grâce à des analyses biologiques et conduisent à la matrice de transition \mathbb{P} .

$$\mathbb{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

On suppose qu'à l'instant $n = 0$, l'état initial de la distribution de probabilité de la végétation est $\pi_0 = (a, b, c, d)$ (c'est à dire qu'au début des observations, la parcelle de terrain était constitué d'une proportion a de sol nu, b d'herbe, c d'arbuste et d de forêt).

1. Calculer les distributions de la végétation sur la parcelle au bout de 3 ans, 6 ans.
2. On suppose que $\lim_{n \rightarrow +\infty} \mathbb{P}^n = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{2}{3} & 0 & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ 0 & 0 & 0 & 1 \end{pmatrix}$; Calculer $\lim_{n \rightarrow +\infty} \pi_n$.
3. Montrer que $X_n[E, \mathbb{P}, \pi_0]$ n'est pas une chaîne stationnaire.

4. A l'aide de la deuxième question, trouver les constantes a, b, c, d pour que la chaîne de Markov homogène $X_n[E, \mathbb{P}, \pi_0]$ soit stationnaire. Que constate-t-on ?

Exercice 8. Un patient qui arrive dans un cabinet médical est dirigé vers une salle d'attente. S'il y a déjà 3 personnes qui attendent, le patient découragé repart immédiatement. Un médecin est présent dans ce cabinet. Le médecin vient toutes les 20 mn dans la salle d'attente pour voir s'il y a des patients en attente. Si c'est le cas, il prend en consultation l'un des patients, sinon il revient 20 mn plus tard. On suppose qu'une consultation ne dure pas plus de 20 mn. On discrétise le temps en intervalles de temps de durée 20 mn. On note X_0 le nombre de personnes dans la salle d'attente à l'instant t_0 et X_n le nombre de personnes qui sont dans la salle d'attente lorsque les médecins arrivent à l'instant t_n dans la salle d'attente. On peut alors montrer que la suite $(X_n)_n$ est une chaîne de Markov homogène de matrice de

$$\text{transition } \mathbb{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

1. Déterminer l'ensemble des états E de cette chaîne de Markov.
2. Calculer les distributions π_1, π_2 et π_3 lorsque $\pi_0 = (1, 0, 0, 0)$.
En déduire que la chaîne $X_n[E, \mathbb{P}, \pi_0]$ n'est pas stationnaire.
3. Trouver une distribution initiale invariante π_0^* laquelle la chaîne $X_n[E, \mathbb{P}, \pi_0^*]$ est stationnaire.
4. Montrer que la chaîne admet un état absorbant que l'on précisera.
5. On suppose maintenant que $\pi_0 = (0, 0, 0, 1)$. Calculer $\lim_{n \rightarrow +\infty} \pi_n$.
6. En déduire $\lim_{n \rightarrow +\infty} \mathbb{P}^n$

Exercice 9. Dans un certain pays, il ne fait jamais beau deux jours de suite. Si un jour il fait beau (BT), le lendemain il peut neiger (N) ou pleuvoir (PL) avec autant de chances. Si un jour il pleut ou il neige, il y a une chance sur deux qu'il y ait changement de temps le lendemain, et s'il y a changement, il y a une chance sur deux que ce soit pour du beau temps.

1. Former, à partir de celà, une chaîne de Markov (en précisant l'espace des états) et en déterminer sa matrice de transition.
2. Si un jour il fait beau, quel est le temps le plus probable pour le surlendemain ?
3. On suppose que l'on a que deux états (beau temps et mauvais temps).
 - a) Déterminer la matrice de transition de la nouvelle chaîne ainsi obtenue.
 - b) Montrer que la chaîne admet un état absorbant que l'on précisera.
 - c) Montrer que la chaîne peut tendre vers un système d'équilibre (On pourra si possible calculer $\lim_{n \rightarrow +\infty} \mathbb{P}^n$).

Exercice 10. Un individu vit dans un environnement où il est susceptible d'être contaminé par une maladie. Son état de santé est suivi mensuellement. Pour un mois donné m , trois états sont possibles : Il peut être immunisé (état I), malade (état M), non malade et non immunisé (état S).

D'un mois m au mois suivant $m+1$, son état peut évoluer selon les règles épidémiologiques suivantes :

- étant immunisé au mois m , il peut, au mois $m+1$, être encore immunisé avec une probabilité 0,9 ou passer à l'état S avec une probabilité 0,1
- étant malade au mois m , il peut, au mois $m+1$, être encore malade avec une probabilité 0,2 ou passer à l'état immunisé avec une probabilité 0,8
- étant en l'état S au mois m , il peut, au mois $m+1$, être encore en l'état S avec une probabilité 0,5 ou passer à l'état malade M avec une probabilité 0,5.

1. Ecrire la matrice \mathbb{A} qui résume les probabilités de transition entre le mois m et le mois $m+1$.
2. En supposant le processus homogènement Markovien, calculer les probabilités pour qu'un individu soit dans l'état e au mois $m+2$, $e = I$ ou M ou S , sachant avec certitude qu'il était immunisé au mois m .
3. Quelle distribution initiale rend le processus stationnaire ?

Exercice 11. Une application économique des chaînes de Markov

La plupart des entreprises qui émettent des obligations sont notées par des agences de notations (Moody's ou Standard and Poors). Les notes peuvent être sous la forme AAA, BBB, CCC; du plus sûr au plus risqué. Après plusieurs évaluations mensuelles, on a constaté que cela obéissait à un processus Markovien homogène de matrice de transition (en pourcentage) :

$$\mathbb{P} = \begin{pmatrix} 90.81 & 8.33 & 0.86 \\ 0.70 & 90.65 & 8.65 \\ 0.09 & 2.27 & 97.64 \end{pmatrix}$$

On rappelle que c'est l'agrégation des capacités (ou crédibilités) de solvabilité des entreprises d'un pays vis à vis de leurs bailleurs de fonds (Institutions financières qui prêtent de l'argent aux entreprises).

On s'intéresse à la situation financière actuelle de l'Allemagne dont le processus de mesure de crédibilité obéit à un algorithme généré par la matrice \mathbb{P} .

On suppose qu'à l'instant $n = 0$ (janvier 2013), l'état initial des notes attribuées à l'Allemagne était $\pi_0 = (1, 0, 0)$ (c'est à dire qu'en janvier, ce pays avait une probabilité certaine d'avoir le triple A (AAA); indicateur d'une très bonne santé économique du pays).

1. Donner l'ensemble E des états possibles de ce processus Markovien de notations.
2. à quelle période (préciser le mois et l'année) l'Allemagne perdra-t-elle son triple A ? (cette situation se produit lorsqu'il existe une distribution où un état est plus probable que l'état AAA).

3. Trouver une distribution initiale π_0^* pour que l'Allemagne garde son triple A quelque soit l'ampleur de la crise financière mondiale qui sévit actuellement.
4. Reprendre les questions précédentes (avec $\pi_0 = (1, 0, 0, 0, 0, 0, 0, 0)$) si les notes sont de la forme AAA, AA, A, BBB, BB, B, CCC, Défaut; du plus sûr au plus risqué, sachant que cela obéissait à un processus Markovien homogène de matrice de transition (en pourcentage) :

$$\mathbb{P} = \begin{pmatrix} 90.81 & 8.33 & 0.68 & 0.06 & 0.12 & 0.00 & 0.00 & 0.00 \\ 0.70 & 90.65 & 7.79 & 0.64 & 0.06 & 0.14 & 0.02 & 0.00 \\ 0.09 & 2.27 & 91.05 & 5.52 & 0.74 & 0.26 & 0.01 & 0.06 \\ 0.02 & 0.33 & 5.95 & 86.93 & 5.30 & 1.17 & 0.12 & 0.18 \\ 0.02 & 0.14 & 0.67 & 7.73 & 80.53 & 8.84 & 1.00 & 1.06 \\ 0.00 & 0.11 & 0.24 & 0.43 & 6.48 & 83.46 & 4.08 & 5.20 \\ 0.22 & 0.00 & 0.22 & 1.30 & 2.38 & 5.00 & 64.85 & 19.79 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 100 \end{pmatrix}$$

- Exercice 12.**
1. Montrer que le produit de deux matrices stochastiques (de même taille) est encore une matrice stochastique.
 2. Montrer que toutes les valeurs propres d'une matrice stochastique ont un module inférieur ou égal à 1.
 3. Une matrice \mathbb{P} est bistochastique si la somme de chaque ligne et de chaque colonne est égale à 1. Montrer que dans ce cas, une telle chaîne de Markov ne possède que des états persistants.
 4. Montrer que la matrice stochastique d'une chaîne de Markov ayant un état absorbant, ne peut pas être régulière.
 5. Montrer que toute chaîne de Markov irréductible et apériodique admet une unique distribution invariante et qui est indépendante de la distribution initiale.

Exercice 13. On considère une chaîne de Markov de matrice stochastique

$$\mathbb{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix};$$

1. Montrer que cette chaîne est irréductible mais périodique de période 2.
2. Montrer alors que $\lim_{n \rightarrow +\infty} \mathbb{P}^n$ n'existe pas.
3. Montrer que cette chaîne de Markov ne possède qu'une distribution invariante $(\frac{1}{2}; \frac{1}{2})$, et qui rend le processus stationnaire dès le départ.

Chapitre 3

MODELES STATISTIQUES PARAMETRIQUES

3.1 Généralités sur l'inférence statistique

3.1.1 Définitions

Définition 1. *La statistique est une discipline mathématique qui élabore des modèles probabilistes pour les données réelles.*

L'une des applications intrinsèques consiste à faire des prévisions. Par exemples :

- Contrôle de qualité en industrie : cartes en temps réels des bouchons sur le réseau RATP.
- Culture expérimentale en agronomie : mécanisme de dispersion spatiale des pollens.
- Enquêtes par des instituts de sondage sur les différents acteurs ou secteurs de la société tels que l'économie, la démographie, ou les opinions publiques.
- Trafic d'internet : requêtes reçues par un serveur.

En statistique, on décrit un groupe d'unité (population ou échantillon) à l'aide de mesures ou caractéristiques (effectif, moyenne, écart-type, variance, proportion, ...).

Définition 2. *1) Ces caractéristiques sont appelées **paramètres**, lorsqu'il s'agit d'étudier ou de décrire une population.*

*2) Ces caractéristiques sont appelées **observations** ou **réalisations statistiques** lorsqu'il s'agit de décrire un échantillon.*

Définition 3. *L'inférence statistique est alors l'ensemble des méthodes permettant de tirer des conclusions sur un groupe déterminé à partir des données provenant d'un échantillon choisi dans cette population.*

La méthodologie consiste donc à :

- Déterminer ces caractéristiques à étudier ;
- Estimer ces caractéristiques à partir de la population totale ;
- Procéder au test d'hypothèse pour étudier la précision de ces estimations.

3.1.2 Éléments de théorie de l'échantillonnage

L'échantillonnage n'est autre que le processus de prélèvement d'un échantillon sur une population. Il existe plusieurs types :

a) Echantillonnages sur la base des méthodes empiriques : Une illustration parfaite est la **méthode des quotas**. Elle est plus utilisée dans les contextes sociaux, et politico-économiques.

b) Echantillonnages aléatoires : On distingue plusieurs types.

- **L'échantillonnage aléatoire simple** : On tire au hasard et avec remise les unités dans la population concernée.
- **L'échantillonnage stratifié** : On subdivise d'abord la population en sous ensembles (ou strates) relativement homogènes ; ensuite on extrait de chaque strate un échantillon aléatoire simple ; enfin on regroupe tous les petits échantillons pour ainsi constituer l'échantillon souhaité.
- **L'échantillonnage par grappes** : Par exemple diviser la ville en quartiers, et choisir un certain nombre de quartiers et faire l'enquête auprès des populations y résidents.

3.2 Modèle statistique

3.2.1 Définition du concept

Définition 4. On appelle modèle statistique pour l'observation X , une famille de mesure de probabilités \mathcal{P} qui contient la loi inconnue suivie par X à partir d'un échantillon.

Définition 5. On appelle modèle statistique pour l'observation X , le triplet $(\chi, \mathcal{A}, \mathcal{P})$ où χ est l'ensemble des résultats, \mathcal{A} une tribu de parties de χ et \mathcal{P} une famille de loi de probabilités sur l'espace mesurable (χ, \mathcal{A}) .

Définition 6. Lorsque le modèle statistique est indicé par un paramètre θ appartenant à un espace des paramètres Θ (i.e $\mathcal{P} = \{P_\theta / \theta \in \Theta\}$), alors l'application $\theta \mapsto P_\theta$ s'appelle **une paramétrisation** du modèle statistique.

Parmi les modèles statistiques, on distingue les modèles paramétriques et les modèles non paramétriques.

Définition 7. Un modèle statistique est dit paramétrique lorsque l'espace des paramètres Θ est un sous ensemble de l'espace euclidien \mathbb{R}^p . p est alors appelé dimension paramétrique du modèle. Dans le cas contraire, le modèle est dit non paramétrique ou encore de dimension infinie.

Remarque 1. 1) Il peut avoir plusieurs, voire une infinité de paramétrisations pour un même modèle statistique.

2) Les caractéristiques de la loi recherchée en modèle paramétrique sont généralement : l'espérance mathématique ou moyenne $E_\theta(X)$; la variance $V_\theta(X)$; la loi ou densité de probabilité $P_\theta(X \in A)$.

3.2.2 Quelques propriétés

- (i) Un modèle statistique paramétrique est dit identifiable s'il admet une paramétrisation injective.
- (ii) Un modèle statistique paramétrique $\{P_\theta/\theta \in \Theta\}$ est dit homogène si le support de sa loi P_θ ne dépend pas du paramètre θ (rappel : $\text{supp}(P_\theta) = \{x \in \mathbb{R}/P_\theta(x) \neq 0\}$)

3.2.3 Modèles statistiques classiques

1. **Modèle gaussien à paramètre bidimensionnel** : $\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2)/\mu \in \mathbb{R}, \sigma^2 > 0\}$

Lorsque l'un des deux paramètres est connu à l'avance, le modèle est alors à un paramètre.

2. **Modèle uniforme** : $\mathcal{P} = \{\mathcal{U}[0, \theta]/\theta > 0\}$

3. **Modèle de Poisson** : $\mathcal{P} = \{P_\lambda/\lambda > 0\}$

4. **Modèle de Cauchy** : $\mathcal{P} = \{f_\theta/\theta \in \mathcal{R}\}$, avec $f_\theta(x) = \frac{1}{\pi(1+(x-\theta)^2)}$

5. **Modèle Binomial** : $\mathcal{P} = \{\mathcal{B}(n, p)/n \in \mathcal{N}^*, p \in [0, 1]\}$

6. **Modèle Beta** : $\mathcal{P} = \{f_{a,b}/a > 0, b > 0\}$ avec $f_{a,b}(x) = \begin{cases} \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)} & \text{si } x \in [0, 1] \\ 0 & \text{sinon} \end{cases}$;

où $B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt$

7. **Modèle Gamma** : $\mathcal{P} = \{f_{p,q}/p > 0, q > 0\}$ avec $f_{p,q}(x) = \begin{cases} \frac{q^p e^{-qx} x^{p-1}}{\Gamma(p)} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases}$;

où $\Gamma(p) = \int_0^{+\infty} t^{p-1} e^{-t} dt$. Si $p = 1$, on retrouve le modèle exponentiel élémentaire.

8. **Modèle Exponentiel général** : $\mathcal{P} = \{f_\theta/\theta \in \Theta\}$, où f_θ est la densité de probabilité susceptible d'être celle de la variable aléatoire X et telle que $f_\theta(x) = C(\theta)h(x)e^{\eta(\theta)T(x)}$

Remarque 2. 1) Si T est la fonction identité alors la famille est dite naturelle.

2) Si $\eta(\theta) = \theta$ alors θ est appelé paramètre canonique.

3) Si on pose $\theta = q$, $C(\theta) = q$, $h(x) = 1$, $\eta(\theta) = -q$ et $T(x) = x$, on retrouve le modèle exponentiel élémentaire.

4) Les lois Normale, Binomiale, Poisson, Gamma, ... sont de la famille exponentielle.

Exercice d'application : Préciser dans chaque cas ci-dessus l'espace de paramétrisation Θ ainsi que sa dimension.

3.3 Estimation ponctuelle des paramètres d'un modèle statistique paramétrique

3.3.1 Définitions : Estimateur du Maximum de Vraisemblance (EMV), biais, convergence, efficacité

Définition 8. Soit X une variable aléatoire dont la loi dépend d'un paramètre θ . Un échantillon de taille n est la donnée de n variables aléatoires X_1, \dots, X_n de même loi que X . Une réalisation de l'échantillon est la donnée d'un n -uplets (x_1, \dots, x_n) où $x_i \in X_i(\Omega)$.

Problème : Il s'agit d'exprimer le paramètre inconnu θ .

Définition 9. Un estimateur de θ est une fonction mesurable souvent notée $\hat{\theta}$ où T encore appelé **statistique**. Elle est définie par $T : W_i^n \rightarrow D \subset \mathbb{R}$ qui à x associe $T(x)$ avec $x = (x_1, \dots, x_n)$ et $W_i = X_i(\Omega)$. $T(x)$ est une estimation de θ .

NB : Pour un même paramètre, il peut avoir plusieurs estimateurs possibles. Pour pouvoir choisir, il faut définir les qualités qui font qu'un estimateur soit meilleur.

Définition 10. 1) On appelle erreur d'estimation la quantité :

$T - \theta = T - E(T) + E(T) - \theta = (T - E(T)) + (E(T) - \theta)$. On pose alors $B(T) = E(T) - \theta$ qui représente l'erreur systématique et s'appelle **biais de l'estimateur** T .

2) Un estimateur T de θ est dit sans biais lorsque $E(T) = \theta$ c'est-à-dire $B(T) = 0$.

3) Un estimateur T de θ est dit asymptotiquement sans biais $E(T) \rightarrow \theta$ lorsque $n \rightarrow +\infty$.

4) Un estimateur sans biais ou asymptotiquement sans biais est dit convergent si sa variance $V(T) \rightarrow 0$ lorsque $n \rightarrow +\infty$.

5) Un estimateur $T(\theta)$ est dit efficace s'il n'est pas biaisé et si $V(T) \leq V(T')$ où T' est un autre estimateur quelconque de θ .

NB : Pour comparer deux estimateurs, on regarde d'abord leurs biais et ensuite leurs variances.

3.3.2 Méthodologie de construction de l'estimateur du maximum de vraisemblance (EMV)

Fonction de vraisemblance

Considérons un modèle statistique $\mathcal{P} = \{P_\theta, \theta \in \Theta\} = \{f_\theta = f(\bullet, \theta), \theta \in \Theta\}$ Etant donné l'observation X , la probabilité de son apparition peut être mesurée par une fonction $f(X, \theta)$. Puisque θ est inconnue, il semble alors naturel de favoriser les valeurs de θ pour lesquelles $f(X, \theta)$ est élevée : **c'est la notion de vraisemblance de θ pour une observation X .**

Définition 11. On appelle fonction de vraisemblance la fonction $L : \Theta \rightarrow \mathbb{R}$ qui à θ associe $f(x, \theta)$ et on note $L(\theta) = L(X, \theta) = f(x, \theta)$.

Lorsque $X = (X_1, X_2, \dots, X_n)$ avec les X_i indépendantes; a pour densité marginale $f(x_i, \theta)$ alors la fonction de vraisemblance de X est $L(X, \theta) = \prod_{i=1}^n f(x_i, \theta)$.

Définition 12. On appelle estimateur du maximum de vraisemblance de θ , tout point maximal de la fonction de vraisemblance noté $\hat{\theta}_{MV} = \sup_{\theta \in \Theta} L(X, \theta)$ il s'agit donc de résoudre l'équation de vraisemblance :

$$\overrightarrow{\text{grad}}(\ln(L(X, \theta))) = \overrightarrow{0}$$

L'estimateur du maximum de vraisemblance sera une solution candidate. On retiendra alors les valeurs pour lesquelles la dérivée seconde est négative.

Propriétés de l'EMV

1. $\hat{\theta}_{MV}$ n'est pas nécessairement sans biais.
2. Si la méthode du maximum de vraisemblance admet une solution unique, alors cette solution converge et est asymptotiquement efficace.
3. $g(\hat{\theta}_{MV})$ est l'estimateur du maximum de vraisemblance de $g(\hat{\theta})$
4. $\hat{\theta}_{MV}$ converge s'il converge en probabilité vers θ . C'est-à-dire :

$$\forall \varepsilon > 0, \lim_{n \rightarrow +\infty} P(|\hat{\theta}_{MV} - \theta| > \varepsilon) = 0$$

3.3.3 Règles de décision : Fonction de perte, Fonction de risque (cas du risque quadratique), Information de Fisher, Test d'efficacité, Inégalité de Rao-Cramer-Frechet

Fonction de perte d'un estimateur

Soit T un estimateur ponctuel du paramètre θ . On sait définir $T : \mathbb{R}^n \rightarrow \Theta \subseteq \mathbb{R}^m$ qui à (x_1, \dots, x_n) associe $(\theta_1, \dots, \theta_m)$.

Définition 13. Toute fonction non négative $l : \Theta \times \Theta \rightarrow \mathbb{R}$ qui à (t, θ) associe $l(t, \theta)$ et convexe en θ est appelé fonction de perte de l'estimateur T .

Remarque 3. La fonction de perte sert à mesurer la qualité d'un estimateur. Il est naturel de supposer que $l(\theta, \theta) = 0$.

Lorsque $l(t, \theta) = \|t - \theta\|^2$, la fonction de perte est dite quadratique.

Exemple 1. On peut avoir :

$$l(t, \theta) = |t - \theta|; \quad l(t, \theta) = \left(1 - \frac{t}{\theta}\right)^2; \quad l(t, \theta) = \frac{t}{\theta} - \log\left(\frac{t}{\theta}\right) - 1; \dots$$

Fonction de Risque d'un estimateur

Définition 14. C'est la fonction définie par $R_l : \Theta \times \Theta \rightarrow \mathbb{R}$ qui à (t, θ) associe $\mathbb{E}[l(t, \theta)]$. C'est donc la perte moyenne lorsqu'on utilise l'estimateur T quand la vraie valeur du paramètre est θ .

Remarque 4. 1) Lorsque la fonction de perte choisie est quadratique, le risque associé est quadratique.

2) Lorsque l'estimateur est sans biais, le risque quadratique correspond à la variance de cet estimateur.

3) Dans le cas où $\theta = (\theta_1, \dots, \theta_m)$, la variance de T n'est autre que la matrice des variances-covariances de (X_1, X_2, \dots, X_n) .

Information de Fisher

Définition 15. On pose $S(X, \bullet) : \Theta \rightarrow \mathbb{R}^m$ qui à θ associe $\overrightarrow{\text{grad}}(\log L(X, \theta))$ c'est-à-dire

$$S(X, \theta) = \left(\frac{\partial \log L(X, \theta)}{\partial \theta_1}, \dots, \frac{\partial \log L(X, \theta)}{\partial \theta_m} \right)$$

C'est la fonction score d'un échantillon $X = (X_1, X_2, \dots, X_n)$.

Définition 16. On appelle information de Fischer au point $\theta = (\theta_1, \dots, \theta_m)$ d'un modèle statistique, la matrice

$$I(\theta) = \mathbb{E} \left[S(X, \theta) \cdot S(X, \theta)^T \right] = (I_{ij}(\theta))_{1 \leq i, j \leq n}; \text{ avec } I_{ij}(\theta) = \mathbb{E} \left[\frac{\partial \log L(X, \theta)}{\partial \theta_i} \cdot \frac{\partial \log L(X, \theta)}{\partial \theta_j} \right]$$

Théorème 1. Pour un modèle statistique régulier, l'information de Fisher est donnée par : $I(\theta) = (I_{ij}(\theta))$; avec

$$I_{ij}(\theta) = -\mathbb{E} \left(\frac{\partial^2 \log L(X, \theta)}{\partial \theta_i \partial \theta_j} \right)$$

Définition 17. Un modèle statistique est dit régulier si :

1. Θ est un ouvert de \mathbb{R}^m .
2. L'ensemble des x pour lesquelles $f(x, \theta) = 0$ ne dépend pas du paramètre θ .
3. La fonction $f(x, \theta)$ est de classe C^2 par rapport à θ .

Propriétés 1. 1. L'information de Fischer d'un modèle régulier est symétrique et définie positive.

2. On montre que $I(\theta) = V(S(x, \theta))$.
3. On rappelle qu'une matrice est positive lorsque sa forme quadratique associée est positive c'est-à-dire que toutes ses valeurs propres sont positives.
4. L'information de Fisher est additive; c'est à dire $I_{X+Y}(\theta) = I_X(\theta) + I_Y(\theta)$, $\forall \theta \in \Theta$, où X et Y sont deux variables aléatoires indépendantes.

Théorème 2. Inégalité de Rao-Cramer-Frechet

1. Soit T un estimateur sans biais de paramètre θ . Si notre modèle statistique est régulier alors $V(T) \geq \frac{1}{I(\theta)}$ avec $\frac{1}{I(\theta)} = \left(-\frac{1}{I_{ij}(\theta)} \right)_{1 \leq i, j \leq n}$

2. Un estimateur sans biais $T(\theta)$ sera dit **efficace** si $V(T) = \frac{1}{I(\theta)}$
3. Un estimateur du maximum de vraisemblance n'est pas toujours efficace mais est au moins **asymptotiquement sans biais**.

3.3.4 Estimateur ponctuelle de la moyenne et de la variance

Estimateur de la moyenne

Soit X une variable aléatoire dont on veut estimer la moyenne $\mu = \mathbb{E}(X)$ à partir d'un n -échantillon X_1, X_2, \dots, X_n de X .

Définition 18. On pose alors $\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n) = \frac{1}{n} \sum_{i=1}^n X_i$. C'est la moyenne empirique. Sa réalisation est : $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ qui est la moyenne de l'échantillon, ou moyenne observée.

Propriétés 2. 1. $\mathbb{E}(\bar{X}) = \mu$ donc \bar{X} est sans biais.

2. Si on pose $\mathbb{V}(X) = \sigma^2$. On montre que $\mathbb{V}(\bar{X}) = \frac{\mathbb{V}(X)}{n} = \frac{\sigma^2}{n}$. D'où $\mathbb{V}(\bar{X}) \rightarrow 0$ lorsque $n \rightarrow +\infty$; et donc \bar{X} converge.

3. D'après 1) et 2), \bar{X} est un estimateur efficace de μ .

Exercice d'application : On suppose que $\mathbb{E}(X) = \mu$, on pose $Y = X_1 + 2X_2 + \dots + nX_n$.

1. Montrer que $\mathbb{E}(Y) = \frac{n(n+1)}{2}\mu$

2. Sachant que $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, montrer que $\mathbb{E}(\bar{X}) = \mu$ et que Y est un estimateur biaisé de μ .

Estimateur de la variance

Etant donné une variable aléatoire X telle que $\mu = \mathbb{E}(X)$ et $\sigma^2 = \mathbb{V}(X)$, on veut estimer sa variance σ^2 . On distingue alors deux cas :

i) Si μ est connu

Définition 19. On appelle *Variance empirique*, la statistique $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2$

Elle est encore appelée **variance observée** car c'est la variance de l'échantillon.

Théorème 3. Si μ est connu, un estimateur efficace de σ^2 est $T^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ et $\mathbb{E}(T^2) = \sigma^2$.

ii) Si μ est inconnue

Théorème 4. $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$. La variance empirique est alors un estimateur biaisé de σ^2 , mais asymptotiquement sans biais.

Propriétés 3. 1. On montre que $\mathbb{E}(S^2) = \frac{n-1}{n}\sigma^2$ donc $\mathbb{E}(S^2) \rightarrow \sigma^2$ lorsque $n \rightarrow +\infty$.

$$2. B(S^2) = \mathbb{E}(S^2) - S^2 = \frac{(n-1)}{n}\sigma^2 - \sigma^2 = -\frac{1}{n}\sigma^2.$$

Théorème 5. $(S')^2 = \frac{n}{n-1}S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ est un estimateur sans biais de σ^2 . En effet,

$$\mathbb{E}(S'^2) = \mathbb{E}\left(\frac{n}{n-1}S^2\right) = \frac{n}{n-1}\mathbb{E}(S^2) = \frac{n}{n-1} \times \frac{n-1}{n}\sigma^2 = \sigma^2$$

Remarque 5. 1) Si n est grand, $\mathbb{E}(S^2) \cong \mathbb{E}(S'^2)$ on préfère S^2 , sinon on utilise S'^2

2) D'après le théorème Centrale limite $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \rightarrow \mathcal{N}(0,1)$ lorsque $n \rightarrow +\infty$ ($n > 30$).

3) On montre que : $\frac{nS^2}{\sigma^2} \rightsquigarrow X_n^2$ et que la version S' est $\frac{(n-1)S'^2}{\sigma^2} \rightsquigarrow X_{n-1}^2$

4) $\frac{nT^2}{\sigma^2} \rightsquigarrow X_n^2$.

3.4 Estimateur par Intervalle de confiance

3.4.1 Généralités

Il est plus réaliste de fournir une estimation du type, $t_1 < \theta < t_2$ plutôt que d'écrire sèchement $\theta = t$, car on sait que la valeur estimée t diffère toujours de la valeur exacte du paramètre recherché θ .

Il est donc souhaitable de donner la **précision de l'estimation**, en acceptant de commettre une erreur epsilon (ε) sur celle-ci.

Définition 20. Soit X une variable aléatoire dont la loi dépend d'un paramètre d'inconnue θ . On appelle **intervalle de confiance de θ** , de niveau de confiance $1 - \varepsilon$ ou encore de seuil ε , un intervalle qui a la probabilité $1 - \varepsilon$ de contenir la vraie valeur de θ . C'est-à-dire $P(t_1 < \theta < t_2) = 1 - \varepsilon$ c'est-à-dire $P(\theta \notin [t_1, t_2]) = \varepsilon$.

Remarque 6. 1) Les niveaux de confiance les plus fréquemment utilisés sont : 90% ($\varepsilon = 10\%$), 95% ($\varepsilon = 5\%$) et 99% ($\varepsilon = 1\%$).

2) ε est encore appelée le risque. On choisira dans la plupart des cas un intervalle à risque symétrique c'est-à-dire : $P(\theta < t_1) = \frac{\varepsilon}{2}$ et $P(\theta > t_2) = \frac{\varepsilon}{2}$.

3) Si on augmente le niveau de confiance $1 - \varepsilon$, on augmente la largeur de l'intervalle de confiance. On prendra généralement $I_\varepsilon = [\mu - c, \mu + c]$ et chercher l'intervalle de confiance I_ε tel que $P(\theta \in I_\varepsilon) = 1 - \varepsilon$ revient à chercher c . μ représente θ dans l'échantillon.

3.4.2 Intervalle de confiance pour une moyenne

i) **Si σ est connue et $n < 30$**

On sait que $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \rightsquigarrow \mathcal{N}(0,1)$. On se fixe le risque ε et on cherche c dans la table de la loi normale satisfaisant l'équation probabiliste :

$$P(\mu - \varepsilon < \bar{X} < \mu + \varepsilon) = 1 - \varepsilon$$

$$\begin{aligned}
&\Leftrightarrow P(\bar{X} \in [\mu - c, \mu + c]) = 1 - \varepsilon \\
&\Leftrightarrow P(-c < X - \mu < c) = 1 - \varepsilon \\
&\Leftrightarrow P\left(-\frac{c}{\frac{\sigma}{\sqrt{n}}} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{c}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - \varepsilon \\
&\Leftrightarrow \phi\left(\frac{c}{\frac{\sigma}{\sqrt{n}}}\right) - \phi\left(-\frac{c}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - \varepsilon
\end{aligned}$$

où ϕ est la fonction de répartition de la loi normale centrée réduite. Puisque $\mathcal{N}(0, 1)$ a une densité de probabilité paire, on a :

$$\phi\left(-\frac{c}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} > \frac{c}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{c}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - \phi\left(\frac{c}{\frac{\sigma}{\sqrt{n}}}\right)$$

D'où

$$\begin{aligned}
&\phi\left(\frac{c}{\frac{\sigma}{\sqrt{n}}}\right) - \left(1 - \phi\left(\frac{c}{\frac{\sigma}{\sqrt{n}}}\right)\right) = 1 - \varepsilon \\
&\Leftrightarrow \phi\left(\frac{c}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - \frac{\varepsilon}{2} \\
&\Leftrightarrow \phi^{-1} \circ \phi\left(\frac{c}{\frac{\sigma}{\sqrt{n}}}\right) = \phi^{-1}\left(1 - \frac{\varepsilon}{2}\right) \\
&\Leftrightarrow \frac{c}{\frac{\sigma}{\sqrt{n}}} = \phi^{-1}\left(1 - \frac{\varepsilon}{2}\right) \\
&\Leftrightarrow \boxed{c = \frac{\sigma}{\sqrt{n}} \phi^{-1}\left(1 - \frac{\varepsilon}{2}\right)}
\end{aligned}$$

On a donc : $I_\varepsilon = [\mu - c, \mu + c]$ où μ est la moyenne de l'échantillon

ii) **Si σ est inconnu et $n < 30$**

Si σ est inconnu, on sait que la variable décisionnelle $\frac{\bar{X} - \mu}{S/\sqrt{n-1}} \rightsquigarrow St_{n-1}$ où S est l'écart type de l'échantillon. On se sert de la table de la loi de Student après avoir isolé c , comme dans le paragraphe précédent :

$$\begin{aligned}
P(\mu - c < \bar{X} < \mu + c) = 1 - \varepsilon &\Rightarrow P\left(-\frac{c}{\frac{S}{\sqrt{n-1}}} < \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n-1}}} < \frac{c}{\frac{S}{\sqrt{n-1}}}\right) = 1 - \varepsilon \\
&\Leftrightarrow \phi_{St_{n-1}}\left(\frac{c}{\frac{S}{\sqrt{n-1}}}\right) - \phi_{St_{n-1}}\left(-\frac{c}{\frac{S}{\sqrt{n-1}}}\right) = 1 - \varepsilon
\end{aligned}$$

On continue le processus comme ci-dessus pour trouver c puisque la densité de la loi de student est aussi paire.

On obtient

$$\boxed{c = \frac{S}{\sqrt{n-1}} \phi_{St_{n-1}}^{-1} \left(1 - \frac{\varepsilon}{2}\right)}$$

iii) **Si $n > 30$ et σ inconnue**

On sait que

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \rightsquigarrow \mathcal{N}(0, 1)$$

d'après le théorème central limite.

On fixe alors l'erreur ε et on utilise la table de la loi normale centrée réduite pour déterminer la constante c comme ci-dessus.

3.4.3 Estimation de la variance σ^2 par intervalle de confiance

i) **Si μ est connue**

On sait que $T^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \Rightarrow \frac{nT^2}{S_e^2} = \sum_{i=1}^n \left(\frac{X_i - \mu}{S_e}\right)^2$
où S_e désigne l'écart-type de l'échantillon. or $\frac{X_i - \mu}{S_e} \rightsquigarrow \mathcal{N}(0, 1)$ donc $\frac{nT^2}{S_e^2} \rightsquigarrow X_n^2$

On cherche I_ε tel que $P(T^2 \in I_\varepsilon) = 1 - \varepsilon$.

Posons $I_\varepsilon = [a, b]$. Alors $P(T^2 \in [a, b]) = 1 - \varepsilon$

$$\Leftrightarrow P(a \leq T^2 \leq b) = 1 - \varepsilon$$

$$\Leftrightarrow P\left(\frac{na}{S_e^2} \leq \frac{nT^2}{S_e^2} \leq \frac{nb}{S_e^2}\right) = 1 - \varepsilon$$

$$\Leftrightarrow P\left(\frac{nT^2}{S_e^2} \in \left]-\infty, \frac{na}{S_e^2}\right] \cup \left[\frac{nb}{S_e^2}, +\infty\right]\right) = 1 - (1 - \varepsilon) = \varepsilon$$

Pour que le risque d'estimation soit symétrique il faut que :

$$\begin{cases} P\left(\frac{nT^2}{S_e^2} > \frac{nb}{S_e^2}\right) = \frac{\varepsilon}{2} & (*) \\ P\left(\frac{nT^2}{S_e^2} < \frac{na}{S_e^2}\right) = \frac{\varepsilon}{2} & (**) \end{cases}$$

$$(**) \Rightarrow \phi_{X_n^2}\left(\frac{na}{S_e^2}\right) = \frac{\varepsilon}{2} \Rightarrow \boxed{a = \frac{S_e^2}{n} \phi_{X_n^2}^{-1}\left(\frac{\varepsilon}{2}\right)}$$

$$(*) \Rightarrow 1 - P\left(\frac{nT^2}{S_e^2} \leq \frac{nb}{S_e^2}\right) = \frac{\varepsilon}{2}$$

$$\Leftrightarrow P\left(\frac{nT^2}{S_e^2} \leq \frac{nb}{S_e^2}\right) = 1 - \frac{\varepsilon}{2}$$

$$\Leftrightarrow \boxed{b = \frac{S_e^2}{n} \phi_{X_n^2}^{-1}\left(1 - \frac{\varepsilon}{2}\right)}$$

ii) **Si μ est inconnue**

On sait que $\frac{nS^2}{S_e^2} \rightsquigarrow X_{n-1}^2$

S^2 est le nouveau estimateur. On cherche deux réels a et b tels que $P(a \leq S^2 \leq b) = 1 - \varepsilon$

$$\Leftrightarrow P\left(\frac{na}{S_e^2} \leq \frac{nS^2}{S_e^2} \leq \frac{nb}{S_e^2}\right) = 1 - \varepsilon$$

Pour que le risque soit symétrique il faut que : $P\left(\frac{nb}{S_e^2} \leq \frac{nS^2}{S_e^2}\right) = \frac{\varepsilon}{2}$ et $P\left(\frac{nS^2}{S_e^2} \leq \frac{na}{S_e^2}\right) = \frac{\varepsilon}{2}$. On trouve $a = \frac{S_e^2}{n} \phi_{X_{n-1}^2}^{-1}\left(\frac{\varepsilon}{2}\right)$ et $b = \frac{S_e^2}{n} \phi_{X_{n-1}^2}^{-1}\left(1 - \frac{\varepsilon}{2}\right)$

3.4.4 Estimation d'une proportion par intervalle de confiance

Estimation ponctuelle

Soit une population ayant des individus possédant une caractéristique A , on veut estimer à partir d'un échantillon de taille n , la proportion d'individu Π possédant cette caractéristique A .

Soit K une variable aléatoire qui représente le nombre d'individus dans l'échantillon possédant la caractéristique A . Alors $K \rightsquigarrow \mathcal{B}(n, \Pi_e)$ où Π_e est la proportion d'individus ayant cette caractéristique dans l'échantillon.

On montre que : $\mathbb{E}(K) = n\Pi_e$, et $\mathbb{V}(K) = n\Pi_e(1 - \Pi_e)$.

Théorème 6. La fréquence empirique $F = \frac{K}{n}$ est l'estimateur efficace de Π .

Intervalle de confiance d'une proportion

Puisque $F = \frac{K}{n}$, on montre que $\frac{F - \Pi_e}{\sqrt{\frac{\Pi_e(1 - \Pi_e)}{n}}} \rightsquigarrow \mathcal{N}(0, 1)$ lorsque $\begin{cases} n\Pi_e > 5 \\ \text{ou} \\ n(1 - \Pi_e) > 5 \end{cases}$

On cherche alors $I_\varepsilon = [\Pi_e - r, \Pi_e + r]$ tel que $P(F \in I_\varepsilon) = 1 - \varepsilon$

$$\Leftrightarrow P(\Pi_e - r \leq F \leq \Pi_e + r) = 1 - \varepsilon \Leftrightarrow P(-r \leq F - \Pi_e \leq r) = 1 - \varepsilon$$

$$\Leftrightarrow P\left(-\frac{r}{\sqrt{\frac{\Pi_e(1 - \Pi_e)}{n}}} \leq \frac{F - \Pi_e}{\sqrt{\frac{\Pi_e(1 - \Pi_e)}{n}}} \leq \frac{r}{\sqrt{\frac{\Pi_e(1 - \Pi_e)}{n}}}\right) = 1 - \varepsilon$$

$$\text{D'où } \phi_{\mathcal{N}(0,1)}\left(\frac{r}{\sqrt{\frac{\Pi_e(1 - \Pi_e)}{n}}}\right) - \phi_{\mathcal{N}(0,1)}\left(-\frac{r}{\sqrt{\frac{\Pi_e(1 - \Pi_e)}{n}}}\right) = 1 - \varepsilon.$$

Comme la densité de probabilité de $\mathcal{N}(0, 1)$ est paire, on a donc :

$$\phi_{\mathcal{N}(0,1)}\left(\frac{r}{\sqrt{\frac{\Pi_e(1 - \Pi_e)}{n}}}\right) - \left(1 - \phi_{\mathcal{N}(0,1)}\left(\frac{r}{\sqrt{\frac{\Pi_e(1 - \Pi_e)}{n}}}\right)\right) = 1 - \varepsilon$$

$$\Leftrightarrow 2\phi_{\mathcal{N}(0,1)}\left(\frac{r}{\sqrt{\frac{\Pi_e(1 - \Pi_e)}{n}}}\right) = 2 - \varepsilon$$

$$\text{D'où } \boxed{r = \sqrt{\frac{\Pi_e(1 - \Pi_e)}{n}} \phi_{\mathcal{N}(0,1)}^{-1}\left(1 - \frac{\varepsilon}{2}\right)}$$

3.5 Travaux Dirigés

Exercice 1. Une enquête a montré que des automobilistes qui résident à Melen, faisaient un excès de vitesse lors d'un passage devant l'Ecole Nationale Supérieure Polytechnique de Yaoundé(ENSPY). Face à de tels excès, le Maire de l'Arrondissement de Yaoundé 3^{ème} décide de lancer une campagne de prévention . Suite à cette campagne, on a observé 15 excès de vitesse sur 150 voitures passant devant l'ENSPY.

1. Donner une estimation ponctuelle du pourcentage p de voitures en excès de vitesse devant l'ENSPY pendant la campagne de prévention.
2. Soit F la variable aléatoire qui à tout échantillon de 150 voitures, on associe le pourcentage de voitures en excès de vitesse.
 - a) Déterminer la moyenne $E(F)$ et la variance $V(F)$ de F .
 - b) On suppose que F suit la loi normale. Estimer par intervalle de confiance la proportion π de voitures de toute la ville de Yaoundé de passage en excès de vitesse devant l'ENSP au seuil $\varepsilon = 0.05$.

Exercice 2. Soit X une variable aléatoire réelle suivant la loi normale de moyenne θ et d'écart-type θ . On considère (X_1, X_2, \dots, X_n) un n -échantillon de X et on pose : $\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$ et $(S')^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.

1. Calculer $E(\bar{X}^2)$ et $E((S')^2)$.
2. Parmi $T_1 = \frac{n\bar{X}^2}{n+1}$ et $T_2 = (S')^2$, quel est le meilleur estimateur de la variance de X ?

Exercice 3. Soit X une variable aléatoire de densité $f_\theta(x) = \frac{\sqrt{\theta}}{\sqrt{2\pi}} e^{-\frac{\theta x^2}{2}}$, avec $\theta > 0$. Soit $(X_1; X_2; \dots; X_n)$ un n -échantillon aléatoire de X .

1. Montrer que f_θ est bien la densité de probabilité d'une loi normale dont on précisera la moyenne et l'écart-type.
2. Déterminer l'estimateur du maximum de vraisemblance $\tilde{\theta}$ du paramètre θ .
3. Montrer que $E(\tilde{\theta}) = \frac{n}{n-2}\theta$.
4. En déduire un estimateur $\hat{\theta}_1$ de θ qui n'est pas biaisé.
5. Montrer que $V(\tilde{\theta}_1) = \frac{2}{n-4}\theta^2$. En déduire que l'estimateur $\tilde{\theta}_1$ de θ converge.

Exercice 4. 1. Qu'appelle t-on modèle statistique pour une observation $X = (X_1, X_2, \dots, X_n)$?

Donner deux exemples de modèles statistiques.

2. Quand dit-on qu'un modèle statistique est paramétrique ?
3. Quand dit-on qu'un modèle statistique est identifiable ?
4. Définir la fonction de perte quadratique d'un estimateur T de θ .

5. Montrer que son risque quadratique vaut : $R(T, \theta) = V(T) + (b(T))^2$ où V est la fonction variance et b la fonction biais.
6. Donner la forme générale de la densité de probabilité d'un modèle statistique exponentiel de paramètre inconnu θ .

Exercice 5. On considère deux modèles statistiques : (a) celui de la loi de Poisson $(\mathbb{N}; \mathbf{P}(\mathbb{N}); P(\lambda) : \lambda > 0)$, où $P(\lambda)$ désigne la loi de poisson de paramètre λ ; (b) celui de la loi exponentielle $(\mathbb{R}_+; \mathbf{B}_{\mathbb{R}_+}; \epsilon(q) : q > 0)$, où $\epsilon(q)$ désigne la loi exponentielle de paramètre q .

On considérera à chaque fois l'observation d'un échantillon $X = (X_1; X_2; \dots; X_n)$

1. Rappeler l'expression de la loi (fonction densité de probabilité) pour chaque modèle statistique.
2. Calculer la fonction de vraisemblance de X pour chaque modèle statistique.
3. Montrer que l'estimateur du maximum de vraisemblance du paramètre de chaque modèle par rapport à la variable aléatoire X vaut respectivement $\tilde{\lambda} = \frac{1}{n} \sum_{i=1}^n X_i$ et $\tilde{q} = \frac{1}{\frac{1}{n} \sum_{i=1}^n X_i}$
4. Montrer que $\tilde{\lambda}$ est sans biais tandis que \tilde{q} est biaisé.
5. On rappelle que le risque quadratique d'un estimateur quelconque $\tilde{\theta}$ de θ est défini par : $R(\tilde{\theta}, \theta) = V(\tilde{\theta}) + (b(\tilde{\theta}))^2$ où V est la fonction variance et b la fonction biais.
Montrer que $R(\tilde{\lambda}, \lambda) = \frac{\lambda}{n}$ et que $R(\tilde{q}, q) = \frac{q^2(n+2)}{(n-1)(n-2)}$
6. On rappelle également que l'information de Fisher I d'un modèle régulier à un paramètre est l'opposé de l'espérance de la dérivée seconde du logarithme de la fonction de vraisemblance de la variable X .
Montrer que $I(\lambda) = \frac{n}{\lambda}$ et que $I(q) = \frac{n}{q^2}$
7. En déduire que $\tilde{\lambda}$ est efficace tandis que \tilde{q} est asymptotiquement efficace.

Exercice 6. Un organisme indépendant a effectué les dosages d'un certain polluant contenue dans les déchets d'une usine. Sur 10 prélèvements provenant de cette usine, on a obtenu une moyenne de 6.8 mg/kg avec un écart-type de 1.9 mg/kg. On admet que le taux de ce polluant contenu dans de tels déchets obéit à une loi $N(m, \sigma)$

1. Donner un intervalle de confiance de la moyenne m au seuil $\varepsilon = 0.05$ en supposant que la variance σ est inconnue.
2. Donner un intervalle de confiance de σ^2 au seuil $\varepsilon = 0.05$ en supposant que la moyenne m est inconnue.

Exercice 7. On introduit une modification sur une chaîne de production (la chaîne fabrique un produit) et l'on souhaite en mesurer l'impact. Pour cela, on utilise un indicateur de performance lié au nombre de clients servis dans les délais par mois et sur 10 mois consécutifs. On obtient les valeurs suivantes au cours de l'expérimentation :
Numéro du mois d'observation : 1 2 3 4 5 6 7 8 9 10

Situation 1(Avant modification) : 148 155 144 129 154 144 132 147 151 119

Situation 2(Après modification) : 165 155 132 152 133 145 151 145 144 143

On supposera que ces mesures sont des réalisations de variables aléatoires indépendantes de loi $X_1 \in N(m_1, \sigma_1)$ et $X_2 \in N(m_2, \sigma_2)$.

Donner une estimation ponctuelle des paramètres m_1 , σ_1^2 , m_2 et σ_2^2 .

Exercice 8. Soit X une variable aléatoire suivant une loi normale $\mathcal{N}(\mu, \sigma^2)$. Soit $(X_1; X_2; \dots; X_n)$ un n -échantillon aléatoire de X

1. Calculer l'information de Fisher $I(\mu)$ pour le paramètre μ (on supposera σ connu).
2. Calculer l'information de Fisher $I(\sigma^2)$ pour le paramètre σ^2 (on supposera μ connu).
3. La statistique X est-elle efficace pour μ ?

Exercice 9. Soit X une variable aléatoire de densité $f_\theta(x) = \frac{A}{x^{1+\frac{1}{\theta}}}$ si $x \geq 1$ et $f_\theta(x) = 0$ sinon, avec $\theta > 0$.

Nous disposons de $(X_1; X_2; \dots; X_n)$ un échantillon aléatoire de taille n de loi parente X .

1. Déterminer A .
2. Déterminer l'estimateur du Maximum de vraisemblance $\tilde{\theta}$ du paramètre θ .
3. Calculer, après avoir justifié pourquoi ce calcul est possible, la borne de Fréchet.
4. L'estimateur $\tilde{\theta}$ est-il efficace ?

Exercice 10. Soit X une variable aléatoire suivant une loi uniforme $U[0, \theta]$. Nous disposons de $(X_1; X_2; \dots; X_n)$, un échantillon aléatoire de taille n de loi parente X .

Soit U et T les statistiques : $U(X_1; X_2; \dots; X_n) = \frac{2}{n} \sum_{i=1}^n X_i$ et $T(X_1; X_2; \dots; X_n) = \sup_{1 \leq i \leq n} X_i$

1. Montrer que U et T sont deux estimateurs de θ .
2. Comparer les deux estimateurs.

Exercice 11. La durée de fonctionnement d'un matériel électrique est représentée par une variable aléatoire réelle X suivant une loi de densité $f(x, \theta, \lambda) = \frac{\lambda}{\theta} x^{\lambda-1} \exp \left\{ -\frac{x^\lambda}{\theta} \right\}$ si $x > 0$, avec $\theta > 0$ et $\lambda > 0$. On suppose que λ est connu.

1. Déterminer la loi de $U = X^\lambda$ puis calculer $E(X^\lambda)$ et $V(X^\lambda)$.
2. On considère un échantillon $(X_1; X_2; \dots; X_n)$ de même loi que X . Calculer l'estimateur du Maximum de vraisemblance $\tilde{\theta}$ du paramètre θ . Cet estimateur est-il sans biais ?, convergent ?, efficace ?. Calculer alors son risque quadratique.

Exercice 12. Les éléments d'une population possèdent un caractère X qui suit une loi de densité $f_\theta(x) = \frac{2}{\sqrt{\pi}\theta^{\frac{3}{2}}} x^2 \exp \left\{ -\frac{x^2}{\theta} \right\}$, avec $\theta > 0$.

Une suite de n expériences indépendantes a donné les valeurs x_1, \dots, x_n .

1. Déterminer un estimateur $\tilde{\theta}$ du paramètre θ par la méthode du maximum de vraisemblance.
2. Examiner les qualités suivantes de $\tilde{\theta}$: efficacité, biais, consistance (convergence).

Exercice 13. Étudier l'efficacité de l'estimateur du maximum de vraisemblance de chacun des modèles statistiques classiques suivants :

1. X suit la loi de Bernoulli $B(1, p)$.
2. X suit la loi uniforme $U[0, \theta]$.
3. X suit la loi de poisson de paramètre $\lambda > 0$.
4. X suit la loi normale bidimensionnelle $\mathcal{N}(\mu, \sigma^2)$.
5. X suit la loi exponentielle de paramètre $q > 0$.

Chapitre 4

GEOMETRIE DES MOINDRES CARRES

4.1 Généralités sur la régression par la méthode des moindres carrés

4.1.1 Problématique

On considère un nuage de point $M_i(x_i, y_i)$ qu'on désire ajuster au mieux par une courbe mathématique parfois appelée tendance de type $x \mapsto f(x)$ dont on devra choisir le modèle de façon pertinente relativement au phénomène étudié.

On cherche alors les paramètres de f qui peut être à priori une fonction affine, polynomiale, exponentielle, puissance, logarithmique, ou périodique.

Rappelons que l'objectif principal de l'étude est la prévision des réalisations futures très souvent économiques (prévoir l'évolution de la vente d'un produit pour ajuster au mieux les moyens de production, savoir en général l'évolution d'un marché).

On cherche donc à estimer les paramètres de la tendance en minimisant (ou en régressant) la somme des carrés des distances entre y_i et $f(x_i)$.

C'est-à-dire on cherche à minimiser $\Delta = \sum_{i=1}^n (y_i - f(x_i))^2 = \sum_{i=1}^n \mathcal{E}_i^2$
 \mathcal{E}_i^2 est appelé erreur d'ajustement ou résidu associé à la valeur explicative x_i

Minimiser Δ consiste donc à résoudre un problème d'optimisation sans contrainte.

4.1.2 Modèles économétriques

Définition 1. *L'économétrie est une branche de l'économie qui traite de l'estimation pratique des relations économiques.*

La démarche économétrique comporte trois étapes :

1. Construire un modèle testable qui puisse être vérifié par la statistique
2. Estimer les paramètres du modèle
3. Vérifier que les écarts entre les observations et les valeurs théoriques du modèle ne sont pas systématiques

Définition 2. *Un modèle en économétrie est une représentation formalisée d'un phénomène sous forme d'équation dont les variables sont des grandeurs économiques.*

Le tableau suivant donne quelques modèles de base.

Modèle	Formule
Linéaire	$y = ax + b$
Log-linéaire	$y = bx^a$
Exponentiel ou Géométrique	$y = e^{ax+b}$
Logarithmique	$y = a \log x + b$
Polynomiale	$y = \sum_{i=0}^p a_i x^i$

4.1.3 Théorie de la corrélation

Définitions

1. Deux phénomènes sont corrélés lorsqu'ils ont une évolution commune.
2. La corrélation simple mesure le degré de liaison entre deux phénomènes (ou deux variables aléatoires).
3. La corrélation multiple cherche la relation entre trois variables ou plus.
4. La corrélation est linéaire si tous les points du couple (X,Y) des deux variables semblent alignés sur une droite.
5. La corrélation est dite non-linéaire si tous les points (X,Y) se trouvent sur une courbe d'allure quelconque.
6. La corrélation est dite positive si les valeurs des variables évoluent dans le même sens.
7. La corrélation est dite négative si les valeurs des variables évoluent en sens contraire.

Coefficient de corrélation linéaire

Soient X et Y deux variables aléatoires. On a par définition :

$$r_{XY} = \frac{cov(X,Y)}{\sigma(X)\sigma(Y)} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Propriétés

- i) $-1 \leq r_{XY} \leq 1$
- ii) Si r_{XY} est proche de zéro, alors il y a absence de corrélation.
- iii) La corrélation est dite forte si $|r_{XY}|$ est proche de 1.
- iv) Le coefficient de corrélation nous permet alors de mesurer la force et le sens de la corrélation entre X et Y .

4.2 Régression linéaire par la méthode des moindres carrés

4.2.1 Prédiction par régression linéaire simple

Position du problème

Soient X et Y deux variables sur lesquelles on veut étudier un modèle économique linéaire. X est appelée **variable explicative** ou **hexogène** ou **prédictive**.

Y est appelée **variable expliquée** ou **à expliquer** ou **endogène** ou **à prédire**.

On suppose que $y_t = a_0 + a_1x_t + \mathcal{E}_t$ où $t \in [1, n]$, a_0, a_1 sont les coefficients de la droite ou paramètres du modèle, \mathcal{E}_t est l'erreur de spécification (différence entre le modèle vrai et le modèle spécifié). Elle est généralement inconnue et mesure l'ensemble des phénomènes explicatifs non liés à X . On rappelle que n est le nombre d'observations.

Estimation ponctuelle de a_0 et a_1

Soient $(x_t, y_t)_{1 \leq t \leq n}$ un échantillon de n couples de valeurs observées.

On suppose que $y_t = a_0 + a_1x_t + \mathcal{E}_t = y(x_t) + \mathcal{E}_t$.

Déterminons a_0 et a_1 par la méthode des moindres carrés. Il s'agit alors de résoudre le problème d'optimisation

$$(P_m) \iff \begin{cases} \min \Delta(a, b) = \min(\sum_{t=1}^n (y_t - b - ax_t)^2) \\ \text{avec } (a, b) \in \mathbb{R}^2 \end{cases}$$

C'est à dire trouver les couples (a_1, a_0) tels que $\forall (a, b) \in \mathbb{R}^2, \Delta(a, b) \geq \Delta(a_1, a_0)$

Condition de 1^{er} ordre : Il s'agit de résoudre l'équation $\overrightarrow{\text{grad}} \Delta(a, b) = \overrightarrow{0}$

$$\text{Soit } \begin{cases} \frac{\partial \Delta(a, b)}{\partial a} = 0 \\ \frac{\partial \Delta(a, b)}{\partial b} = 0 \end{cases} \iff \begin{cases} \sum_{t=1}^n \frac{\partial}{\partial a} (y_t - b - ax_t)^2 = 0 \\ \sum_{t=1}^n \frac{\partial}{\partial b} (y_t - b - ax_t)^2 = 0 \end{cases}$$

$$a = -\frac{\frac{1}{n} \sum_{t=1}^n x_t y_t - \overline{XY}}{(\overline{X})^2 - \frac{1}{n} \sum_{t=1}^n x_t^2} = \frac{E(XY) - E(X)E(Y)}{E(X^2) - [E(X)]^2} \iff \boxed{a = \frac{\text{Cov}(X, Y)}{V(X)}}$$

On montre que $\boxed{b = \overline{Y} - a\overline{X}}$

On a un unique candidat $\boxed{A_1 = \begin{pmatrix} a_1 \\ a_0 \end{pmatrix} = \begin{pmatrix} \frac{\text{Cov}(X, Y)}{V(X)} \\ \overline{Y} - [\frac{\text{Cov}(X, Y)}{V(X)}]\overline{X} \end{pmatrix}}$

Etudions si le point $A_1 \begin{pmatrix} a_1 \\ a_0 \end{pmatrix}$ est un maximum ou un minimum de $\Delta(a, b)$.

On procèdera par la matrice Hessienne au point A_1 car Δ est de classe C^2 .

$$\text{Hess}(\Delta, A_1) = \begin{pmatrix} \frac{\partial^2 \Delta}{\partial a^2}(a_1, a_0) & \frac{\partial^2 \Delta}{\partial a \partial b}(a_1, a_0) \\ \frac{\partial^2 \Delta}{\partial a \partial b}(a_1, a_0) & \frac{\partial^2 \Delta}{\partial b^2}(a_1, a_0) \end{pmatrix} = \begin{pmatrix} 2 \sum_{t=1}^n x_t^2 & 2 \sum_{t=1}^n x_t \\ 2 \sum_{t=1}^n x_t & 2n \end{pmatrix}$$

$$\begin{aligned}
\det (Hess (\Delta, A_1)) &= 4 \left(n \sum_{t=1}^n x_t^2 - \left(\sum_{t=1}^n x_t \right)^2 \right) \\
&= 4 \left(n^2 \mathbb{E} (X^2) - n^2 (\mathbb{E} (X))^2 \right) \\
&= 4n^2 \mathbb{V} (X) \geq 0 \\
tr (Hess (\Delta, A_1)) &= 2 \left(\sum_{t=1}^n x_t^2 + n \right) = 2n (\mathbb{E} (X^2) + 1) \geq 0
\end{aligned}$$

Donc $Hess(\Delta, A_1)$ est définie positive. D'où le point A_1 est bien un minimum de $\Delta(a, b)$. On pose alors :

$$\tilde{a}_1 = \frac{cov(X, Y)}{V(X)} \text{ et } \tilde{a}_0 = \bar{Y} - \tilde{a}_1 \bar{X}.$$

NB : Nous montrerons ultérieurement que les estimateurs \tilde{a}_1 et \tilde{a}_0 de a_1 et a_0 sont efficaces.

4.2.2 Cas général : Prédiction par régression polynomiale

On admet qu'on cherche les paramètres d'une fonction polynomiale f qui minimise la somme des carrés des distances entre y_i et $f(x_i)$ avec $f(x) = c_0 + c_1x + c_2x^2 + \dots + c_px^p$.

On pose alors $\Delta(c_0, c_1, \dots, c_p) = \sum_{i=1}^n (y_i - \sum_{k=0}^p c_k x_i^k)^2$.

On peut obtenir efficacement un minimum de Δ soit en développant celui-ci selon Taylor soit en utilisant la méthode d'optimisation basée sur les fonctions numériques à plusieurs variables. Une condition nécessaire (non suffisante) d'extremum pour Δ est que toutes ses dérivées partielles soient nulles.

$$\forall k = 0, \dots, p, \frac{\partial \Delta}{\partial c_k} = 2 \sum_{i=1}^n [(y_i - f(x_i)) (-x_i^k)] = 0 \quad (4.1)$$

C'est-à-dire $\sum_{i=1}^n x_i^k f(x_i) = \sum_{i=1}^n y_i x_i^k$ avec $k \in \{0, 1, \dots, p\}$
Pour simplifier, on pose :

$$S_k = \sum_{i=1}^n x_i^k, \quad W_k = \sum_{i=1}^n y_i x_i^k \quad (4.2)$$

Ainsi pour $k = 0$, $\sum_{i=1}^n f(x_i) = \sum_{i=1}^n y_i$
C'est-à-dire $nc_0 + c_1 S_1 + c_2 S_2 + \dots + c_p S_p = W_0$.
Pour $k = 1$, $c_0 S_1 + c_1 S_2 + c_2 S_3 + \dots + c_p S_{p+1} = W_1$

⋮

$k = p - 1$, $c_0 S_{p-1} + c_1 S_p + c_2 S_{p+1} + \dots + c_p S_{p+p-1} = W_{p-1}$
Pour $k = p$, $c_0 S_p + c_1 S_{p+1} + c_2 S_{p+2} + \dots + c_p S_{2p} = W_p$

Les $p + 1$ équations précédentes donnent le système matriciel suivant :

$$\begin{bmatrix} n & S_1 & \cdots & S_p \\ S_1 & S_2 & \cdots & S_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ S_p & S_{p+1} & \cdots & S_{2p} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_p \end{bmatrix} = \begin{bmatrix} W_0 \\ W_1 \\ \vdots \\ W_p \end{bmatrix} \quad (4.3)$$

On constate que la matrice A_{ij} ($1 \leq i, j \leq p+1$) du système est symétrique et inversible. Par conséquent, le système admet une solution unique.

Remarque 1. si $p = 1$ on a $\begin{pmatrix} n & S_1 \\ S_1 & S_2 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \end{pmatrix} = \begin{pmatrix} W_0 \\ W_1 \end{pmatrix}$, $\det = nS_2 - S_1^2$.

On obtient $c_0 = \frac{W_0 S_2 - W_1 S_1}{nS_2 - S_1^2}$; $c_1 = \frac{nW_1 - S_1 W_0}{nS_2 - S_1^2}$

En remplaçant W_0, W_1, S_1 et S_2 par leurs valeurs, on retrouve les résultats sur la loi de régression linéaire simple. C'est-à-dire

$$c_1 = \frac{\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}{V(X)} ; c_0 = \mathbb{E}(Y) - c_1 \mathbb{E}(X).$$

Exemple 1. Ajuster par le modèle polynomial le nuage de point représenté dans le tableau ci-dessous.

x_i	-3	-2	1	2	3	4	5
y_i	2	0,5	-1	-1	0	2	4

Solution

On sait que lorsque le nuage de points présente i sommets, alors la tendance est un polynôme de degré $i + 1$.

Dans ce cas le nuage ayant 1 sommet, on l'ajuste alors avec un modèle polynômiale de degré $1 + 1 = 2$.

On pose ainsi $\forall x \in \mathbb{R}$, $f(x) = \sum_{i=0}^2 c_i x^i = c_0 + c_1 x + c_2 x^2$.

Le vecteur $(c_i)_{0 \leq i \leq 2}$ bâti sur les c_i , $\forall i \in \{0, 1, 2\}$ est solution de l'équation matricielle.

$$\begin{pmatrix} S_0 & S_1 & S_2 \\ S_1 & S_2 & S_3 \\ S_2 & S_3 & S_4 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} W_0 \\ W_1 \\ W_2 \end{pmatrix}$$

$$S_j = \sum_{i=1}^7 x_i^j, \quad \forall j \in [[0, 4]] \text{ et } W_j = \sum_{i=1}^7 y_i x_i^j, \quad \forall j \in [[0, 2]]$$

$$S_0 = \sum_{i=1}^7 1 = 7$$

$$S_1 = \sum_{i=1}^7 x_i^1 = -3 - 2 + 1 + 2 + 3 + 4 + 5 = 10$$

$$S_2 = \sum_{i=1}^7 x_i^2 = 9 + 4 + 1 + 4 + 9 + 16 + 25 = 65$$

$$S_3 = \sum_{i=1}^7 x_i^3 = -27 - 8 + 1 + 8 + 27 + 64 + 125 = 190$$

$$S_4 = \sum_{i=1}^7 x_i^4 = 81 + 16 + 1 + 16 + 81 + 256 + 625 = 1076$$

$$W_0 = \sum_{i=1}^7 y_i x_i^0 = 2 + 0,5 - 1 - 1 + 0 + 2 + 4 = 6,5$$

$$W_1 = \sum_{i=1}^7 y_i x_i^1 = -3 \times 2 - 2 \times \frac{1}{2} - 1 - 2 + 8 + 20 = 18$$

$$W_2 = \sum_{i=1}^7 y_i x_i^2 = 18 + 2 - 1 - 4 + 0 + 32 + 100 = 147$$

Ainsi on obtient
$$\begin{pmatrix} 7 & 10 & 68 \\ 10 & 68 & 190 \\ 68 & 190 & 1076 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} \frac{13}{2} \\ 18 \\ 147 \end{pmatrix}$$

Donc
$$\begin{cases} c_0 = -\frac{20101}{15974} = -1,258 \\ c_1 = -\frac{4861}{15974} = -0,304 \\ c_2 = \frac{4311}{15974} = 0,269 \end{cases}$$

D'où
$$\underline{|\forall x \in \mathbb{R}, f(x) = 0,269x^2 - 0,304x - 1,258|}$$

4.3 Modèle de régression linéaire multiple

4.3.1 Position du problème

Les phénomènes économiques et sociaux appréhendés par une seule variable sont très rares. D'où le modèle linéaire général encore appelé modèle de régression multiple. On a une variable endogène y_t expliquée par plusieurs variables exogènes x_{it} ($i = 1, \dots, k ; t = 1, \dots, n$). On a lors :

$$y_t = a_0 + \sum_{i=1}^k a_i x_{it} + \mathcal{E}_t$$

où a_0, a_1, \dots, a_k sont alors les paramètres du modèle, n est la taille de l'échantillon, k le nombre de variables explicatives ayant posé le phénomène observé.

Pour $t = 1$, on a $y_1 = a_0 + a_1 x_{11} + a_2 x_{21} + \dots + a_k x_{k1} + \mathcal{E}_1$

⋮

Pour $t = n$, on a $y_n = a_0 + a_1 x_{1n} + a_2 x_{2n} + \dots + a_k x_{kn} + \mathcal{E}_n$

On obtient alors l'équation matricielle

$$Y = Xa + \mathcal{E} \text{ où } a = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{pmatrix}; \mathcal{E} = \begin{pmatrix} \mathcal{E}_1 \\ \vdots \\ \mathcal{E}_n \end{pmatrix}; Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \text{ et}$$

$$X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{pmatrix}$$

Estimer les paramètres de départ revient donc à estimer le vecteur a qui est l'unique paramètre de l'équation matricielle.

4.3.2 Etude de l'estimateur \hat{a} du vecteur $a = (a_0, a_1, \dots, a_k)$

$Y = Xa + \mathcal{E}$ où $X \in \mathcal{M}(n, k+1)$, $Y, \mathcal{E} \in \mathbb{R}^n$, $a \in \mathbb{R}^{k+1}$.

Posons $S = S(a_0, a_1, \dots, a_k) = \sum_{t=1}^n \mathcal{E}_t^2$. Déterminons alors l'estimateur \hat{a} qui minimise S .

On sait que $\mathcal{E} = Y - Xa$.

Posons $\mathcal{E}' = \mathcal{E}^T$ (transposée de \mathcal{E}).

$\mathcal{E}' = (Y - Xa)' = Y' - a'X'$

Comme $S = \mathcal{E}'\mathcal{E}$ alors $S = (Y' - a'X')(Y - Xa)$. Ce qui implique $S = YY' - Y'Xa - a'XY + a'X'Xa$.

Cherchons alors les points candidats qui minimisent S . Il suffit d'appliquer la condition de premier ordre $\frac{\partial S}{\partial a} = \vec{0}$.

Or $\frac{\partial S}{\partial a} = 0 - 2YX' + \frac{\partial(a'X'Xa)}{\partial a} = -2Y'X + X'Xa + a'X'X$

Par ailleurs $X'Xa = a'X'X$; d'où $\frac{\partial S}{\partial a} = -2Y'X + 2X'Xa$

Ainsi, $\frac{\partial S}{\partial a} = \vec{0} \Leftrightarrow -2Y'X + 2X'Xa = \vec{0}$. Or $X'X$ est symétrique donc inversible.

Donc on peut écrire

$$(X'X)^{-1}(X'Xa) = (X'X)^{-1}(Y'X) \text{ d'où } \boxed{\hat{a} = (X'X)^{-1}(Y'X)}$$

Remarque 2. Comme $Y'Xa = a'X'Y$, on peut encore écrire :

$$S = Y'Y - 2a'X'Y + a'X'Xa \text{ et on obtient } \boxed{\hat{a} = (X'X)^{-1}(X'Y)}$$

Question : \hat{a} est-il un estimateur sans biais de a ?

Il suffit d'évaluer $\mathbb{E}(\hat{a})$. On sait que $\hat{a} = (X'X)^{-1}(X'Y) = (X'X)^{-1}X'(Xa + \mathcal{E})$

$$\hat{a} = (X'X)^{-1}(X'X)a + (X'X)^{-1}X'\mathcal{E}$$

$$\hat{a} = a + (X'X)^{-1}X'\mathcal{E}$$

Donc $\mathbb{E}(\hat{a}) = \mathbb{E}(a) + \mathbb{E}[(X'X)^{-1}X'\mathcal{E}] = a + (X'X)^{-1}X'\mathbb{E}(\mathcal{E})$.

Montrons que $\mathbb{E}(\mathcal{E}) = 0$.

On sait que $\hat{a}_0 = a_0 + \sum_{i=1}^n (\alpha_{0i} \mathcal{E}_i)$, \dots , $\hat{a}_k = a_k + \sum_{i=1}^n \alpha_{ki} \mathcal{E}_i$
 Or par hypothèse, on a supposé que $\mathbb{E}(\mathcal{E}_i) = 0$; D'où

$$\mathbb{E}(\mathcal{E}) = \begin{pmatrix} \mathbb{E}(\mathcal{E}_1) \\ \mathbb{E}(\mathcal{E}_2) \\ \vdots \\ \mathbb{E}(\mathcal{E}_n) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \vec{0}$$

D'où $\mathbb{E}(\hat{a}) = a + 0 = a$ donc \hat{a} est bien un estimateur sans biais de a .

4.4 Travaux Dirigés

Exercice 1. 1. Qu'appelle t-on modèle économétrique ?

2. Donner quelques modèles de base en économétrie et si possible leurs contextes d'applications.
3. Quelle est la différence entre la régression linéaire simple et la régression polynômiale ?
4. Pourquoi utilise t-on plus la méthode des moindres carrées pour faire de la régression ?

Exercice 2. : Régression linéaire avec un modèle logarithmique.

On considère le nuage de points ci-dessous dont on désire faire un ajustement :

x	2	5	9	14	20	27	35
y	-0.49	1.49	2.77	3.73	4.51	5.16	5.72

Déterminer l'équation de la tendance qui donne l'ajustement logarithmique entre x et y ?

Exercice 3. Lors du séisme survenu le 11 Mars 2011 au Japon, le marché financier de Tokyo à travers son indice appelé "NIKKEI", a aussitôt réagi par quelques fluctuations. Ainsi, ses valeurs de clôture journalière se résument dans le tableau ci-dessous :

	11 Mars (1 ^{er} jour)	14 Mars (2 ^{eme} jour)	15 Mars (3 ^{eme} jour)	16 Mars (4 ^{eme} jour)	17 Mars (5 ^{eme} jour)
Valeur de l'indice (en points)	10000	8000	8160	8005	8200

18 Mars (6 ^{eme} jour)	21 Mars (7 ^{eme} jour)	22 Mars (8 ^{eme} jour)
8310	8350	8325

On souhaite utiliser la méthode des moindres carrées pour prévoir les fluctuations futures.

1. Utiliser le modèle de régression linéaire simple pour simuler un meilleur comportement futur de ce marché financier (on déterminera l'équation de la tendance).
2. Quelle est véritablement l'équation de la tendance dans ce contexte financier ?

Exercice 4. La crise d'endettement qu'a connu la Grèce, a causé plusieurs perturbations dans les places boursières européennes. Ainsi, quelques valeurs de clôture journalière du CAC40 (Indice de la bourse de Paris) à compter du 07 Novembre 2011, se résument dans le tableau ci-dessous :

	7 Nov	8 Nov	9 Nov	10 Nov	11 Nov	14 Nov
Valeur de l'indice (en points)	3250	3200	3245	3195	3140	3180

On souhaite utiliser la méthode des moindres carrées pour prévoir les fluctuations futures.

1. Utiliser le modèle de régression linéaire simple pour simuler un meilleur comportement futur de ce marché financier (on déterminera l'équation de la tendance).
2. Quelle est véritablement l'équation de la tendance dans ce contexte économique ?

Exercice 5. : Gestion d'un portefeuille par la régression linéaire multiple.

Un fonds de placements internationaux gère son portefeuille d'actions et d'obligations en fonction des critères économiques, démographiques et sociaux des 4 pays dans lesquels une partie des souscriptions qui lui sont confiées sont investies ou en voie de l'être. Pour cela, il dispose des statistiques suivantes sur les 4 pays :

SR : taux moyen d'épargne par personne entre 1960 à 1970
 POP15 : pourcentage de la population de moins de 15 ans
 POP75 : pourcentage de la population de plus de 75 ans
 DDPI : taux moyen de croissance du revenu moyen par personne.

	Autriche	Espagne	Japon	Zimbabwe
SR	12.07	21.10	11.77	13.30
POP15	23.32	27.01	27.74	31.92
POP75	4.41	1.91	2.87	1.52
DDPI	3.93	8.21	4.35	2.00

On souhaite utiliser la méthode des moindres carrées pour prévoir les fluctuations futures du portefeuille.

1. Effectuer la régression linéaire multiple du taux moyen d'épargne (variable expliquée) par le taux moyen de croissance du revenu moyen par personne et les pourcentages de population de moins de 15 ans et de plus de 75 ans (trois variables explicatives) .
2. Que peut-on dire des coefficients de corrélation (ou de régression) ci-dessus obtenus entre le taux moyen d'épargne et les pourcentages de population de moins de 15 ans et de plus de 75 ans ? Comment interpréter les signes de ces coefficients de régression ?

Exercice 6. Le Laboratoire ERIC (Equipe de recherche en ingénierie des connaissances) souhaite étudier l'impact de fumer (Variables exogènes : TAR ; NICOTINE ; WEIGHT) sur l'hypertension (Variable endogène : CO). Pour cela, ce Laboratoire dispose des statistiques suivantes sur 4 marques de cigarettes :

<i>Cigarettes</i>	<i>TAR(mg)</i>	<i>NICOTINE(mg)</i>	<i>WEIGHT(g)</i>	<i>CO(mg)</i>
<i>Benson and Hedges</i>	16	1.06	1.0938	16.6
<i>Camelights</i>	8	0.67	0.928	10.2
<i>Carlton</i>	4.1	0.4	0.9462	5.4
<i>Marlboro</i>	15.1	0.9	0.9316	14.4

On souhaite utiliser la méthode des moindres carrées pour prévoir l'influence de la cigarette sur l'Hypertension.

1. Effectuer la régression linéaire multiple de la variable à expliquer (CO) par les trois variables explicatives (TAR ; NICOTINE ; WEIGHT).
2. Interpréter les signes des différents coefficients de régression obtenus ci-dessus.