



Ecole Nationale Supérieure d'Arts et Métiers

Aix-en-Provence

Probabilités et Statistiques

Version 2.1
Mai 2003

Jean-Louis POSS

Table des matières

1	Notion de probabilité	7
1.1	Événements. Espace probabilisable	7
1.2	Probabilité	8
1.3	Probabilités conditionnelles. Événements indépendants	9
1.4	Formule de BAYES	11
2	Variables aléatoires discrètes	13
2.1	Définition	13
2.2	Loi d'une variable aléatoire	14
2.3	Couple aléatoire. Variables indépendantes	15
2.3.1	Loi d'un couple aléatoire	15
2.3.2	Variables aléatoires indépendantes	16
2.4	Espérance mathématique. Variance. Coefficient de corrélation	17
2.4.1	Variable aléatoire à une dimension	17
2.4.2	Vecteur aléatoire	21
2.5	Inégalité de BIENAYMÉ-TCHEBYCHEV	21
2.6	Fonction génératrice	22
2.7	Lois de probabilité discrètes classiques	24
2.7.1	Loi binomiale	24
2.7.2	Loi multinomiale	25
2.7.3	Loi hypergéométrique	27
2.7.4	Loi géométrique (ou loi de PASCAL)	28
2.7.5	Loi de POISSON	29
3	Variables aléatoires absolument continues	33
3.1	Variable et vecteur aléatoires	33
3.2	Fonction de répartition. Densité	33
3.2.1	Variable aléatoire à une dimension	33
3.2.2	Vecteurs aléatoires	35
3.3	Espérance, variance, coefficient de corrélation	38
3.4	Détermination de la densité	38
3.5	Fonction caractéristique	40
3.6	Lois de probabilités continues classiques	44
3.6.1	Loi uniforme	44
3.6.2	Loi exponentielle	45

3.6.3	Loi normale (ou loi de LAPLACE-GAUSS)	47
3.6.4	Loi gamma	52
3.6.5	Loi bêta	56
3.6.6	Loi du khi-deux	57
3.6.7	Loi de STUDENT	60
3.6.8	Loi de FISHER-SNEDECOR	62
4	Convergences stochastiques	65
4.1	Différents types de convergence	65
4.1.1	Convergence en loi	65
4.1.2	Convergence en probabilité	65
4.2	Loi faible des grands nombres	67
4.3	Théorème central limit	67
4.4	Application : approximations de la loi binomiale	68
4.4.1	Approximation normale	68
4.4.2	Approximation de POISSON	68
5	Chaînes de MARKOV	71
5.1	Introduction	71
5.2	Définition	71
5.3	Méthodes algébriques	74
5.3.1	Relation fondamentale	74
5.3.2	Matrices stochastiques	74
5.3.3	Comportement asymptotique	76
5.4	Classification des états	78
5.4.1	Définitions	78
5.4.2	Périodicité	79
5.4.3	Récurrence	79
6	Estimation	87
6.1	Introduction	87
6.2	Estimation ponctuelle	87
6.2.1	Définitions	87
6.2.2	Exemples fondamentaux	88
6.2.3	Maximum de vraisemblance	92
6.3	Estimation par intervalle	93
6.3.1	Définitions	93
6.3.2	Exemples fondamentaux	94
6.4	Estimation d'une proportion	97
6.4.1	Estimation ponctuelle	97
6.4.2	Estimation par intervalle	97
6.5	Comparaison de moyennes et de variances	99
6.5.1	Intervalle de confiance de la différence de deux moyennes	99
6.5.2	Intervalle de confiance du rapport de deux variances	102
6.6	Méthode du « Bootstrap »	103

7	Tests	105
7.1	Introduction	105
7.2	Tests paramétriques	105
7.2.1	Test de la moyenne pour une loi normale (σ connu)	106
7.2.2	Test de la moyenne pour une loi normale (σ inconnu)	109
7.2.3	Test de l'écart-type pour une loi normale (m inconnue)	109
7.2.4	Comparaison de proportions	110
7.2.5	Comparaison de deux moyennes (observations appariées)	111
7.2.6	Comparaison de plusieurs moyennes (observations non appariées) – analyse de variance pour un facteur	112
7.2.7	Probabilité critique	113
7.3	Test du khi-deux	114
8	Régression linéaire	119
8.1	Introduction : lignes de régression	119
8.1.1	Cas discret	119
8.1.2	Cas continu	119
8.2	Régression linéaire simple	120
8.2.1	Estimation des paramètres par les moindres carrés	121
8.2.2	Estimation des paramètres par le maximum de vraisemblance	126
8.3	Régression linéaire multiple	134
8.3.1	Estimation des paramètres par les moindres carrés	135
8.3.2	Estimation des paramètres par le maximum de vraisemblance	139
A	Dénombrements	143
A.1	Suites. Arrangements sans répétition	143
A.2	Combinaisons sans répétition	144
A.3	Arrangements avec répétition	144
B	Fonctions eulériennes	145
C	Tables numériques	149
C.1	Loi normale	150
C.2	Loi du khi-deux	152
C.3	Loi de STUDENT	154
C.4	Loi de FISHER-SNEDECOR	155
	Bibliographie	159
	Index	161

Chapitre 1

Notion de probabilité

1.1 Événements. Espace probabilisable

Le résultat d'une expérience aléatoire (*épreuve*) est un événement... imprévisible : nous supposons qu'il appartient à un ensemble Ω appelé *ensemble des issues* ou *univers*. Dans le cas où les résultats possibles sont en nombre fini nous noterons $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ où les ω_i représentent les différentes issues.

Exemple élémentaire : on jette une pièce de monnaie ; on choisit $\Omega = \{\omega_1, \omega_2\}$ où ω_1 représente l'événement « le côté visible est pile » et ω_2 représente l'événement « le côté visible est face ». En fait ce choix présente une part d'arbitraire : nous aurions pu considérer le cas où la pièce reste en équilibre sur la tranche ou bien nous intéresser au temps mis par la pièce pour s'immobiliser. La même épreuve peut donner lieu à des modélisations diverses suivant le phénomène étudié.

Un événement A est une partie de Ω ($A \subset \Omega$ ou $A \in \mathcal{P}(\Omega)$) ; on dira que l'événement A est réalisé si l'issue ω_i de l'épreuve appartient à A ($\omega_i \in A$).

Afin de pouvoir attribuer une probabilité aux événements il est nécessaire de munir leur ensemble d'une structure appropriée.

Définition 1. Une tribu \mathcal{E} d'événements est une partie de $\mathcal{P}(\Omega)$ qui vérifie :

- Ω appartient à \mathcal{E} .
- $\forall A \in \mathcal{E}, \overline{A} = \Omega \setminus A \in \mathcal{E}$.
- $\forall (A, B) \in \mathcal{E}^2, A \cup B \in \mathcal{E}$ et $A \cap B \in \mathcal{E}$.
- $\forall I \subset \mathbb{N}, (\forall i \in I, A_i \in \mathcal{E}) \Rightarrow \bigcup_{i \in I} A_i \in \mathcal{E}$ et $\bigcap_{i \in I} A_i \in \mathcal{E}$.

Dans le cas où Ω est fini ou dénombrable on pourra toujours choisir $\mathcal{E} = \mathcal{P}(\Omega)$; lorsque Ω n'est pas dénombrable on choisira pour \mathcal{E} une partie stricte de $\mathcal{P}(\Omega)$.

Un peu de vocabulaire :

- l'univers Ω est également appelé *événement certain* puisqu'il contient toujours, par définition, l'issue de l'épreuve ;
- la partie vide $\emptyset = \overline{\Omega}$ est l'événement *impossible* ;
- \overline{A} est dit *événement contraire* de A ;
- lorsque $A \subset B$ on dit que A *implique* B ;
- $A \cup B$ se lit « A ou B » et $A \cap B$ se lit « A et B » ;
- lorsque $A \cap B = \emptyset$ on dit que les événements A et B sont *incompatibles*.

Définition 2. Le couple (Ω, \mathcal{E}) est un espace probabilisable.

1.2 Probabilité

Définition 3. Soit (Ω, \mathcal{E}) un espace probabilisable. Une probabilité est une application \mathbb{P} de \mathcal{E} dans l'intervalle $[0, 1]$ telle que :

- $\mathbb{P}(\Omega) = 1$,
- $\forall (A, B) \in \mathcal{E}^2, A \cap B = \emptyset \implies \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$,
- $\forall i \in I \subset \mathbb{N}, A_i \in \mathcal{E} \text{ et } (i \neq j \implies A_i \cap A_j = \emptyset) \implies \mathbb{P}\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mathbb{P}(A_i)$.

Définition 4. $(\Omega, \mathcal{E}, \mathbb{P})$ est un espace probabilisé.

Théorème 1. On a :

1. Probabilité du contraire

$$\forall A \in \mathcal{E}, \mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A).$$

2. Monotonie

$$\forall (A, B) \in \mathcal{E}^2, A \subset B \implies \mathbb{P}(A) \leq \mathbb{P}(B).$$

3. Théorème des « probabilités totales »

$$\forall (A, B) \in \mathcal{E}^2, \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

4. Généralisation

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{i=1}^n \sum_{j=i+1}^n \mathbb{P}(A_i \cap A_j) \\ &\quad + \sum_{i=1}^n \sum_{j=i+1}^n \sum_{k=j+1}^n \mathbb{P}(A_i \cap A_j \cap A_k) - \dots + (-1)^{n-1} \mathbb{P}\left(\bigcap_{i=1}^n A_i\right). \end{aligned}$$

(Les démonstrations sont rédigées en couleur bleue.)

1. $(A \cup \overline{A} = \Omega \text{ et } A \cap \overline{A} = \emptyset) \implies \mathbb{P}(A) + \mathbb{P}(\overline{A}) = 1$.

2. $(B = A \cup (B \setminus A) \text{ et } A \cap (B \setminus A) = \emptyset) \implies \mathbb{P}(B) = \mathbb{P}(A) + \underbrace{\mathbb{P}(B \setminus A)}_{\geq 0}$.

3. On écrit les réunions disjointes : $A \cup B = (A \setminus B) \cup (A \cap B) \cup (B \setminus A)$, $A = (A \setminus B) \cup (A \cap B)$ et $B = (B \setminus A) \cup (A \cap B)$.

4. Récurrence (pénible) sur n .

Cas particulier : définition d'une probabilité sur un ensemble dénombrable.

Soit $\Omega = \{\omega_i\}_{i \in \mathbb{N}}$ un univers dénombrable ; on pose $\mathcal{E} = \mathcal{P}(\Omega) : (\Omega, \mathcal{E})$ est un espace probabilisable.

Pour définir une probabilité il suffit de se donner les probabilités $\mathbb{P}(\{\omega_i\})$ telles que $\sum_{i=1}^{\infty} \mathbb{P}(\{\omega_i\}) = 1$;

on a alors pour tout événement $A : \mathbb{P}(A) = \sum_{\omega_i \in A} \mathbb{P}(\{\omega_i\})$.

La vérification est immédiate.

Cette méthode s'applique en particulier lorsque Ω est fini et les issues ω_i *équiprobables* :

$$\mathbb{P}(\{\omega_1\}) = \mathbb{P}(\{\omega_2\}) = \dots = \mathbb{P}(\{\omega_n\}).$$

On obtient alors :

$$\text{ÉQUIPROBABILITÉ} \implies \mathbb{P}(A) = \frac{\text{nombre d'issues favorables à } A}{\text{nombre d'issues possibles}}.$$

Les dénombrements des issues intervenant au numérateur et au dénominateur peuvent nécessiter des théorèmes rappelés en annexe A (page 143).

Théorème 2. Soit $(A_n)_{n \in \mathbb{N}}$ une suite monotone d'événements ; on a :

$$\mathbb{P}(\lim A_n) = \lim \mathbb{P}(A_n)$$

1. Supposons que la suite $(A_n)_{n \in \mathbb{N}}$ est croissante et que $A_0 = \emptyset$; on a : $\lim A_n = \bigcup_{i=0}^{\infty} A_i$.
Alors :

$$\begin{aligned} \mathbb{P}(\lim A_n) &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} (A_i \setminus A_{i-1})\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i \setminus A_{i-1}) \\ &= \lim \sum_{i=1}^n \mathbb{P}(A_i \setminus A_{i-1}) = \lim \sum_{i=1}^n (\mathbb{P}(A_i) - \mathbb{P}(A_{i-1})) = \lim \mathbb{P}(A_n). \end{aligned}$$

2. Supposons que la suite $(A_n)_{n \in \mathbb{N}}$ est décroissante ; on a : $\lim A_n = \bigcap_{i=0}^{\infty} A_i$.
La suite $(A_0 \setminus A_n)_{n \in \mathbb{N}}$ est croissante.

$$\begin{aligned} \mathbb{P}(A_0) - \mathbb{P}(\lim A_n) &= \mathbb{P}(A_0 \setminus \lim A_n) = \mathbb{P}(\lim (A_0 \setminus A_n)) = \lim \mathbb{P}(A_0 \setminus A_n) \\ &= \lim (\mathbb{P}(A_0) - \mathbb{P}(A_n)) = \mathbb{P}(A_0) - \lim \mathbb{P}(A_n). \end{aligned}$$

Donc :

$$\mathbb{P}(\lim A_n) = \lim \mathbb{P}(A_n).$$

1.3 Probabilités conditionnelles. Événements indépendants

Le fait de savoir qu'un événement s'est produit peut modifier la probabilité que nous attachons à un autre événement.

Exemple : jet de deux dés réguliers. On choisit comme univers l'ensemble des couples de résultats : $\Omega = \{(i, j) / 1 \leq i, j \leq 6\}$ où i est le nombre de points du premier dé et j le nombre de points du deuxième ; il y a donc 36 issues (« événements élémentaires ») possibles. Considérons les deux événements :

- A = « la somme des points est impaire »,
- B = « la somme des points est supérieure ou égale à 7 ».

On a, de façon immédiate : $\mathbb{P}(A) = \frac{18}{36}$ et $\mathbb{P}(B) = \frac{21}{36}$.

Quelle est la probabilité de l'événement : « la somme des points est supérieure ou égale à 7 sachant qu'elle est impaire » ?

Le nombre de cas possibles est égal au nombre de couples donnant une somme impaire, c'est-à-dire 18. Parmi ces cas possibles, qui sont équiprobables, il y en a 12 qui sont favorables à B ; la probabilité cherchée est donc : $\mathbb{P}(B/A) = \frac{12}{18} = \frac{24}{36}$. La probabilité attachée à B est plus forte si l'on sait que A s'est produit.

On remarque sur cet exemple que : $\mathbb{P}(B/A) = \frac{12/36}{18/36} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$.

Théorème 3. Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace probabilisé et soit A un événement de probabilité non nulle.

L'application \mathbb{P}_A définie sur \mathcal{E} par : $\mathbb{P}_A(B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$ est une probabilité sur l'espace probabilisable (Ω, \mathcal{E}) .

On vérifie les axiomes d'une probabilité :

$$1. \mathbb{P}_A(\Omega) = \frac{\mathbb{P}(A \cap \Omega)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A)}{\mathbb{P}(A)} = 1.$$

$$2. \forall (B, C) \in \mathcal{E}^2, B \cap C = \emptyset \implies \mathbb{P}_A(B \cup C) = \frac{\mathbb{P}(A \cap (B \cup C))}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \cap B) \cup (A \cap C))}{\mathbb{P}(A)}.$$

Or : $B \cap C = \emptyset \implies (A \cap B) \cap (A \cap C) = A \cap (B \cap C) = \emptyset$; donc :

$$\mathbb{P}_A(B \cup C) = \frac{\mathbb{P}(A \cap B) + \mathbb{P}(A \cap C)}{\mathbb{P}(A)} = \mathbb{P}_A(B) + \mathbb{P}_A(C).$$

3. On généralise sans peine la démonstration précédente.

Définition 5. La probabilité \mathbb{P}_A est appelée probabilité conditionnelle sachant que A s'est produit.

On peut noter $\mathbb{P}_A(B)$ ou bien $\mathbb{P}(B/A)$ la probabilité conditionnelle de B sachant que A s'est produit.

Théorème 4.

$$\forall (A, B) \in \mathcal{E}^2, \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B/A) = \mathbb{P}(B)\mathbb{P}(A/B).$$

La première égalité vient immédiatement de la définition de la probabilité conditionnelle, la deuxième résulte de la symétrie de l'intersection.

Ce résultat peut être généralisé :

Théorème 5.

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) = \mathbb{P}(A_1)\mathbb{P}_{A_1}(A_2)\mathbb{P}_{A_1 \cap A_2}(A_3) \dots \mathbb{P}_{A_1 \cap \dots \cap A_{n-1}}(A_n).$$

Récurrence immédiate sur n .

La probabilité de réalisation de l'événement B ne dépend pas de la réalisation de l'événement A lorsque $\mathbb{P}_A(B) = \mathbb{P}(B)$. La formule du théorème 4 se simplifie et on obtient : $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ et donc $\mathbb{P}_B(A) = \mathbb{P}(A)$. La probabilité de A ne dépend pas de la réalisation de B .

On est amené à poser :

Définition 6. Les événements A et B sont dits (stochastiquement) indépendants si

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

Exemple : les résultats de deux jets successifs d'un dé sont indépendants.

La généralisation de la notion d'indépendance à plus de deux événements présente une difficulté.

Exemple : jet de deux pièces équilibrées.

On peut choisir : $\Omega = \{(p, p), (f, f), (p, f), (f, p)\}$, où p et f représentent respectivement « pile » et « face », et supposer les issues équiprobables. Les événements :

- A = « la première pièce tombe sur pile »,
- B = « la deuxième pièce tombe sur face »,
- C = « les deux pièces tombent du même côté »

sont indépendants deux à deux, mais on n'a pas, par exemple :

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

Définition 7. Les événements A_1, A_2, \dots, A_n sont mutuellement indépendants si :

$$\forall k \in \{1, 2, \dots, n\}, 1 \leq i_1 < i_2 < \dots < i_k \leq n \implies \mathbb{P}\left(\bigcap_{j=1}^k A_{i_j}\right) = \prod_{j=1}^k \mathbb{P}(A_{i_j}).$$

La notion d'indépendance mutuelle est (beaucoup) plus forte que celle d'indépendance deux à deux ; c'est celle qui intervient le plus souvent.

Exemple : les résultats de n jets successifs d'un dé sont mutuellement indépendants.

1.4 Formule de BAYES

Définition 8. Un système complet d'événements $(B_i)_{1 \leq i \leq n}$ est une partition de Ω :

- $i \neq j \implies B_i \cap B_j = \emptyset$,
- $\bigcup_{i=1}^n B_i = \Omega$.

Théorème 6. Soit $(B_i)_{1 \leq i \leq n}$ un système complet d'événements et soit $A \in \mathcal{E}$. Alors :

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(B_i) \mathbb{P}_{B_i}(A).$$

$$\mathbb{P}(A) = \mathbb{P}(A \cap \Omega) = \mathbb{P}\left(A \cap \bigcup_{i=1}^n B_i\right) = \mathbb{P}\left(\bigcup_{i=1}^n (A \cap B_i)\right) = \sum_{i=1}^n \mathbb{P}(A \cap B_i) = \sum_{i=1}^n \mathbb{P}(B_i) \mathbb{P}_{B_i}(A).$$

On a utilisé : $i \neq j \implies B_i \cap B_j = \emptyset \implies (A \cap B_i) \cap (A \cap B_j) = \emptyset$.

Théorème 7 (BAYES). Soit $(B_i)_{1 \leq i \leq n}$ un système complet d'événements et soit $A \in \mathcal{E}$ tel que $\mathbb{P}(A) \neq 0$.

$$\mathbb{P}_A(B_i) = \frac{\mathbb{P}(B_i) \mathbb{P}_{B_i}(A)}{\sum_{j=1}^n \mathbb{P}(B_j) \mathbb{P}_{B_j}(A)}.$$

De $\mathbb{P}(A \cap B_i) = \mathbb{P}(A)\mathbb{P}_A(B_i) = \mathbb{P}(B_i)\mathbb{P}_{B_i}(A)$ on déduit $\mathbb{P}_A(B_i) = \frac{\mathbb{P}(B_i)\mathbb{P}_{B_i}(A)}{\mathbb{P}(A)}$ et on remplace $\mathbb{P}(A)$ par l'expression obtenue au théorème précédent.

Exemple :

Les pourcentages de production et de déchet de quatre machines sont fournis par le tableau suivant :

	Production	Déchets
Machine 1	15%	5%
Machine 2	20%	4%
Machine 3	30%	3%
Machine 4	35%	2%

Au contrôle on constate qu'une pièce est défectueuse ; quelle est la probabilité pour qu'elle vienne d'une machine donnée ?

Les « causes » sont les événements $B_i = \ll \text{la pièce vient de la machine } i \gg$; A est l'événement « la pièce est défectueuse ». On cherche les probabilités $\mathbb{P}_A(B_i)$ pour $i = 1, 2, 3, 4$.

On a :

$$\mathbb{P}(B_1) = 0.15, \quad \mathbb{P}(B_2) = 0.20, \quad \mathbb{P}(B_3) = 0.30, \quad \mathbb{P}(B_4) = 0.35,$$

$$\mathbb{P}_{B_1}(A) = 0.05, \quad \mathbb{P}_{B_2}(A) = 0.04, \quad \mathbb{P}_{B_3}(A) = 0.03, \quad \mathbb{P}_{B_4}(A) = 0.02.$$

D'où :

$$\mathbb{P}(A) = \sum_{i=1}^4 \mathbb{P}(B_i)\mathbb{P}_{B_i}(A) = (75 + 80 + 90 + 70)10^{-4} = 3.15\%$$

et

$$\mathbb{P}_A(B_1) = \frac{75}{315} = 23.8\%, \quad \mathbb{P}_A(B_2) = \frac{80}{315} = 25.4\%,$$

$$\mathbb{P}_A(B_3) = \frac{90}{315} = 28.6\%, \quad \mathbb{P}_A(B_4) = \frac{70}{315} = 22.2\%.$$

Chapitre 2

Variables aléatoires discrètes

2.1 Définition

Soit Ω un univers fini ou dénombrable ; on choisit pour tribu d'événements $\mathcal{E} = \mathcal{P}(\Omega)$.

Définition 9. Une variable aléatoire sur (Ω, \mathcal{E}) est une application de Ω dans \mathbb{R} .

On notera par la suite les variables aléatoires par des lettres majuscules : $X, Y, Z, \dots, U, V, \dots$

Exemples :

1. Soit l'épreuve : on jette un dé ; on choisit pour univers $\Omega = \{\omega_1, \omega_2, \dots, \omega_6\}$ où ω_i représente l'événement « on obtient la face i ». On définit une variable aléatoire X en posant : $X(\omega_i) = i$; on peut aussi bien considérer les variables aléatoires définies par $Y(\omega_i) = i^2$ ou bien même $Z(\omega_i) = \log i$, bien que ce soit moins naturel.
2. On tire au hasard une personne dans une population donnée ; l'événement élémentaire ω_i est : « on a tiré la personne numéro i ». On définit les variables aléatoires U, V, W par :
 - (a) $U(\omega_i) =$ taille de la personne en centimètres,
 - (b) $V(\omega_i) =$ poids de la personne en kilogrammes,
 - (c) $W(\omega_i) =$ année de naissance.

On peut évidemment associer bien d'autres variables aléatoires à ce tirage.

3. Si X est une application constante on dit que c'est une *variable certaine*.

Théorème 8. L'ensemble des variables aléatoires définies sur (Ω, \mathcal{E}) a une structure d'espace vectoriel réel et d'anneau commutatif et unitaire pour les opérations définies par :

$$\forall \omega \in \Omega, (X + Y)(\omega) = X(\omega) + Y(\omega),$$

$$\forall \omega \in \Omega, \forall \lambda \in \mathbb{R}, (\lambda X)(\omega) = \lambda(X(\omega)),$$

$$\forall \omega \in \Omega, (XY)(\omega) = X(\omega)Y(\omega).$$

$$\text{On a de plus la propriété : } \lambda(XY) = (\lambda X)Y = X(\lambda Y).$$

C'est un résultat classique pour les fonctions à valeurs réelles.

Soit X une variable aléatoire et f une application de \mathbb{R} dans \mathbb{R} : l'application composée $f \circ X$ est définie sur Ω et à valeurs dans \mathbb{R} , donc c'est une variable aléatoire que l'on note usuellement $f(X)$ (Attention : ceci n'est plus tout à fait exact lorsque \mathcal{E} n'est pas égal à $\mathcal{P}(\Omega)$; voir le chapitre 3).

2.2 Loi d'une variable aléatoire

On suppose (Ω, \mathcal{E}) muni de la probabilité \mathbb{P} .

L'ensemble $X(\Omega)$ est fini ou dénombrable ; on pose : $X(\Omega) = (x_i)_{i \in I}$ où $I = \{1, 2, \dots, m\}$ dans le cas fini ou bien $I = \mathbb{N}^*$ dans le cas dénombrable en supposant $(i < j \Rightarrow x_i < x_j)$.

Définition 10. On appelle loi de la variable aléatoire X l'application p de $X(\Omega)$ dans $[0, 1]$ définie par :

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

On rappelle : $X^{-1}(x_i) = \{\omega \in \Omega \mid X(\omega) = x_i\}$. Cet événement peut s'énoncer : « X prend la valeur x_i » : on le notera donc en abrégé « $X = x_i$ » en commettant un *abus de notation* commode ; $p(x_i)$ est donc la probabilité pour que la variable aléatoire X prenne la valeur x_i .

Exemple : on considère la variable aléatoire U définie ci-dessus. Soit $u_i = 173$ cm : $p(u_i)$ est la probabilité de tirer au hasard une personne dont la taille est égale à 1,73 m. Cette probabilité dépend évidemment de la population étudiée, donc de \mathbb{P} .

Théorème 9. La loi de X définit une probabilité sur l'espace probabilisable $(X(\Omega), \mathcal{P}(X(\Omega)))$ par :

$$p(A) = \sum_{x_i \in A} p(x_i).$$

C'est un cas particulier de probabilité définie sur un ensemble fini ou dénombrable.

Définition 11. La fonction de répartition de la variable aléatoire X est l'application F définie sur \mathbb{R} par :

$$F(x) = \mathbb{P}(X^{-1}([-\infty, x])).$$

On notera plus simplement, *mais c'est un abus*, $F(x) = \mathbb{P}(X < x)$. Lorsque c'est nécessaire on note, plus précisément, F_X la fonction de répartition de la variable aléatoire X .

Théorème 10. La fonction de répartition F vérifie les propriétés suivantes :

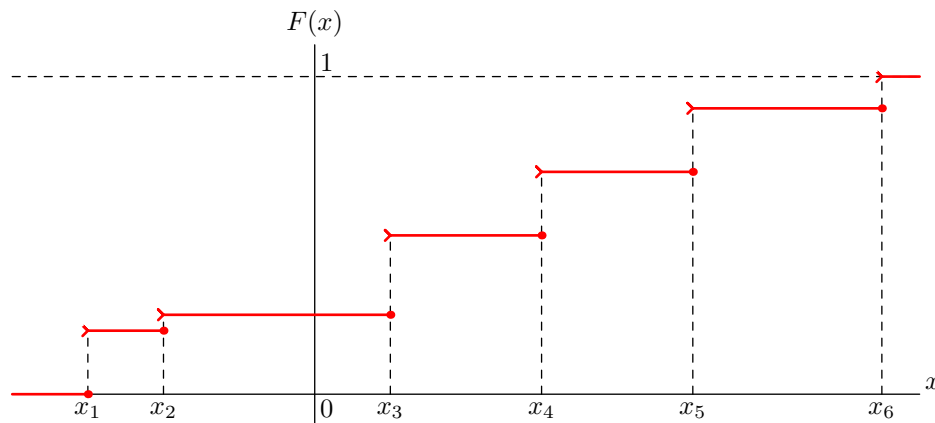
1. $F(x) = \sum_{x_i < x} p(x_i)$,
2. F est croissante, constante par intervalle (fonction en escalier),
3. F est continue à gauche en tout point,
4. $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow +\infty} F(x) = 1$,
5. $\forall i \in I$, $p(x_i) = F(x_i + 0) - F(x_i)$ où $F(x_i + 0) = \lim_{\substack{x \rightarrow x_i \\ x > x_i}} F(x)$.

1. L'événement $X^{-1}([-\infty, x])$ est la réunion disjointe des images réciproques $X^{-1}(x_i)$ pour tous les x_i strictement inférieurs à x .

2. Si x et x' appartiennent à un même intervalle $]x_i, x_{i+1}]$ les événements $X^{-1}([-\infty, x])$ et $X^{-1}([-\infty, x'])$ sont tous les deux égaux à la réunion disjointe $\bigcup_{j=1}^i X^{-1}(x_j)$; donc : $F(x) = F(x') = \sum_{j=1}^i p(x_j)$.

Si $x \leq x_i < x'$ on a $X^{-1}([-\infty, x']) = X^{-1}([-\infty, x]) \cup X^{-1}(x_i)$ et donc : $F(x') = F(x) + p(x_i) \geq F(x)$.

3. Si $x = x_i$ on a $F(x) = \sum_{j=1}^{i-1} p(x_j)$ qui est la valeur de $F(x')$ pour tout $x' \in]x_{i-1}, x_i]$.
4. Si $X(\Omega)$ est fini et égal à $\{x_1, x_2, \dots, x_n\}$ on a $F(x) = 0$ si $x \leq x_1$ et $F(x) = 1$ si $x > x_n$.
Dans le cas où $X(\Omega)$ n'est pas fini voir la démonstration du théorème 39.
5. On a vu ci-dessus que $F(x) = F(x_i) + p(x_i)$ si $x \in]x_i, x_{i+1}]$; d'où le résultat.



2.3 Couple aléatoire. Variables indépendantes

2.3.1 Loi d'un couple aléatoire

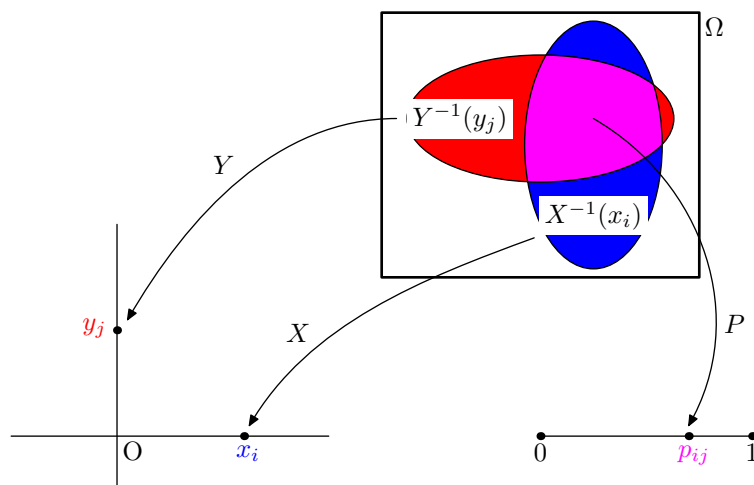
Soient X et Y deux variables aléatoires sur (Ω, \mathcal{E}) .

On a posé au paragraphe 2.2 : $X(\Omega) = (x_i)_{i \in I}$ avec $I = \{1, 2, \dots, m\}$ ou $I = \mathbb{N}^*$. De la même façon : $Y(\Omega) = (y_j)_{j \in J}$ où $J = \{1, 2, \dots, n\}$ dans le cas fini ou bien $J = \mathbb{N}^*$ dans le cas dénombrable.

Définition 12. La loi du couple (X, Y) , ou loi conjointe de X et Y , est l'application p définie sur $X(\Omega) \times Y(\Omega)$ par :

$$p(x_i, y_j) = \mathbb{P}(X^{-1}(x_i) \cap Y^{-1}(y_j)).$$

Ceci s'écrit plus simplement $\mathbb{P}((X = x_i) \cap (Y = y_j))$ et pourra se noter p_{ij} lorsqu'aucune confusion n'est à craindre.



Cette définition se généralise sans peine à un n -uplet (X_1, X_2, \dots, X_n) de variables aléatoires qui est un *vecteur aléatoire*.

Théorème 11. Sur l'espace probabilisable $(X(\Omega) \times Y(\Omega), \mathcal{P}(X(\Omega) \times Y(\Omega)))$ la loi conjointe définit une probabilité par :

$$\forall A \subset X(\Omega) \times Y(\Omega), p(A) = \sum_{(x_i, y_j) \in A} p(x_i, y_j).$$

C'est un cas particulier de définition de probabilité sur un espace dénombrable.

Les lois de X et Y s'appellent *lois marginales* (du couple). Si on connaît la loi du couple on peut en déduire les lois marginales, mais la réciproque est fausse !

Théorème 12. Les lois de X et Y sont données par :

$$\mathbb{P}(X = x_i) = \sum_{j \in J} \mathbb{P}((X = x_i) \cap (Y = y_j)), \quad \mathbb{P}(Y = y_j) = \sum_{i \in I} \mathbb{P}((X = x_i) \cap (Y = y_j)).$$

$$\begin{aligned} \mathbb{P}(X = x_i) &= \mathbb{P}((X = x_i) \cap \Omega) = \mathbb{P}\left((X = x_i) \cap \left(\bigcup_{j \in J} (Y = y_j)\right)\right) = \mathbb{P}\left(\bigcup_{j \in J} ((X = x_i) \cap (Y = y_j))\right) \\ &= \sum_{j \in J} \mathbb{P}((X = x_i) \cap (Y = y_j)) \quad \text{car la réunion est disjointe.} \end{aligned}$$

On notera en abrégé, lorsqu'aucune confusion n'est à craindre :

$$p_{i.} = \sum_{j \in J} p_{ij}, \quad p_{.j} = \sum_{i \in I} p_{ij}.$$

Lorsque I et J sont finis les lois peuvent se représenter par un tableau :

	x_1	x_2	\dots	x_i	\dots	x_m	$p_{.j}$
y_1	p_{11}	p_{21}	\dots	p_{i1}	\dots	p_{m1}	$p_{.1}$
y_2	p_{12}	p_{22}	\dots	p_{i2}	\dots	p_{m2}	$p_{.2}$
\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots
y_j	p_{1j}	p_{2j}	\dots	p_{ij}	\dots	p_{mj}	$p_{.j}$
\dots	\dots	\dots	\dots	\dots	\dots	\dots	\dots
y_n	p_{1n}	p_{2n}	\dots	p_{in}	\dots	p_{mn}	$p_{.n}$
$p_{i.}$	$p_{1.}$	$p_{2.}$	\dots	$p_{i.}$	\dots	$p_{m.}$	1

Dans les marges, en bas et à droite du tableau, on note les probabilités ... marginales, obtenues en additionnant les éléments de chaque colonne ou bien de chaque ligne.

2.3.2 Variables aléatoires indépendantes

Définition 13. Les variables aléatoires X et Y sont indépendantes si les événements $(X = x_i)$ et $(Y = y_j)$ sont indépendants pour toutes les valeurs x_i et y_j , c'est-à-dire :

$$\forall (i, j) \in I \times J, p_{ij} = p_{i.} p_{.j}.$$

On généralise sans difficulté la notion d'indépendance à n variables aléatoires : les variables pourront alors être deux à deux indépendantes ou, ce qui est plus intéressant et plus fréquent, mutuellement indépendantes.

2.4 Espérance mathématique. Variance. Coefficient de corrélation

2.4.1 Variable aléatoire à une dimension

Définition 14. Le moment d'ordre $k \in \mathbb{N}^*$ de la variable aléatoire X est défini par :

$$m_k(X) = \sum_{i \in I} x_i^k p(x_i)$$

si la série est absolument convergente (c'est-à-dire si la série $\sum_{i \in I} |x_i|^k p(x_i)$ converge).

Théorème 13. Si le moment d'ordre k existe tous les moments d'ordre inférieur à k existent également.

Supposons $k' \leq k$ et montrons :

$$\sum_{i \in I} |x_i|^k p(x_i) \text{ converge} \Rightarrow \sum_{i \in I} |x_i|^{k'} p(x_i) \text{ converge.}$$

On a : $\forall a > 0, 0 \leq k' \leq k \Rightarrow a^{k'} \leq 1 + a^k$; en effet :

- $a \geq 1 \Rightarrow a^k = a^{k'} a^{k-k'} \geq a^{k'}$ et donc $a^{k'} < 1 + a^k$;
- $a < 1 \Rightarrow a^{k'} < 1 \leq 1 + a^k$.

Donc :

$$\sum_{i \in I} |x_i|^{k'} p(x_i) \leq \sum_{i \in I} (1 + |x_i|^k) p(x_i) = 1 + \sum_{i \in I} |x_i|^k p(x_i).$$

Espérance mathématique

C'est le moment d'ordre 1 ; on le note :

$$\mathbf{E}(X) = \sum_{i \in I} x_i p(x_i).$$

Remarques

1. Si l'on considère que, pour tout i , $p(x_i)$ représente une masse ponctuelle placée à l'abscisse x_i , $\mathbf{E}(X)$ est l'abscisse du centre d'inertie.
2. Soient $f : \mathbb{R} \mapsto \mathbb{R}$ et la variable aléatoire $Z = f \circ X$, que l'on notera plus simplement $f(X)$.
On a ($Z = f(x_i)$) lorsque ($X = x_i$), donc avec la probabilité $p(x_i)$.

$$\mathbf{E}(Z) = \mathbf{E}(f(X)) = \sum_{i \in I} f(x_i) p(x_i).$$

Cas particulier : $f(x) = x^k$. On obtient : $\mathbf{E}(X^k) = \sum_{i \in I} x_i^k p(x_i) = m_k(X)$.

On n'utilisera plus la notation $m_k(X)$ par la suite.

Théorème 14. \mathbf{E} est une forme linéaire positive sur l'espace vectoriel des variables aléatoires définies sur (Ω, \mathcal{E}) .

1. Montrons d'abord le lemme : $\mathbf{E}(X) = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\})$.

On a : $p(x_i) = \mathbb{P}(X^{-1}(x_i)) = \sum_{\omega \in X^{-1}(x_i)} \mathbb{P}(\{\omega\})$. Donc :

$$\mathbf{E}(X) = \sum_{i \in I} x_i p(x_i) = \sum_{i \in I} x_i \left(\sum_{\omega \in X^{-1}(x_i)} \mathbb{P}(\{\omega\}) \right) = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\})$$

car chaque ω appartient à l'une des images réciproques $X^{-1}(x_i)$.

2. D'après le lemme :

$$\begin{aligned} \mathbf{E}(\alpha X + \beta Y) &= \sum_{\omega \in \Omega} (\alpha X + \beta Y)(\omega) \mathbb{P}(\{\omega\}) = \alpha \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\}) + \beta \sum_{\omega \in \Omega} Y(\omega) \mathbb{P}(\{\omega\}) \\ &= \alpha \mathbf{E}(X) + \beta \mathbf{E}(Y). \end{aligned}$$

3. Il est évident par ailleurs que \mathbf{E} est positive.

Variance

C'est le moment centré d'ordre 2 :

$$\mathbf{Var}(X) = \mathbf{E}\left((X - \mathbf{E}(X))^2\right).$$

Remarques

1. La variance de X représente le moment d'inertie des masses ponctuelles $p(x_i)$ placées aux abscisses x_i par rapport au centre d'inertie.

2. Lorsque $\mathbf{Var}(X) = 0$, X est une variable certaine.

$\mathbf{Var}(X) = \sum_{i \in I} (x_i - \mathbf{E}(X))^2 p(x_i)$: si la variance est nulle tous les x_i (pour lesquels $p(x_i)$ est non nul) sont égaux à une même constante, qui est $\mathbf{E}(X)$.

Pour des raisons d'homogénéité on introduit l'écart-type : $\sigma(X) = \sqrt{\mathbf{Var}(X)}$.

Théorème 15. *Propriétés de la variance :*

1. $\mathbf{Var}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2$.
2. $\forall a \in \mathbb{R}, \mathbf{E}((X - a)^2) = \mathbf{Var}(X) + (\mathbf{E}(X) - a)^2$ (théorème de HUYGENS).
3. $\forall (a, b) \in \mathbb{R}^2, \mathbf{Var}(aX + b) = a^2 \mathbf{Var}(X)$ et $\sigma(aX + b) = |a| \sigma(X)$.

$$1. \mathbf{E}\left((X - \mathbf{E}(X))^2\right) = \mathbf{E}(X^2 - 2X\mathbf{E}(X) + \mathbf{E}(X)^2) = \mathbf{E}(X^2) - 2\mathbf{E}(X)^2 + \mathbf{E}(X)^2.$$

$$2. \mathbf{E}((X - a)^2) = \mathbf{E}\left(\left((X - \mathbf{E}(X)) + (\mathbf{E}(X) - a)\right)^2\right) = \dots$$

$$3. \mathbf{Var}(aX + b) = \mathbf{E}\left(\left((aX + b) - \mathbf{E}(aX + b)\right)^2\right) = a^2 \mathbf{E}\left((X - \mathbf{E}(X))^2\right)$$

Définition 15. On appelle variable réduite associée à X la variable aléatoire :

$$X^* = \frac{X - E(X)}{\sigma(X)}.$$

On a immédiatement :

Théorème 16.

$$E(X^*) = 0, \quad \sigma(X^*) = 1.$$

Coefficient de corrélation

Définition 16. La covariance des variables aléatoires X et Y est définie par :

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))].$$

Remarques

1. $\text{Cov}(X, X) = \text{Var}(X)$.
2. $\text{Cov}(X, Y) = E(XY) - E(X) E(Y)$.

$\text{Cov}(X, Y) = E(XY - XE(Y) - YE(X) + E(X)E(Y))$, puis on utilise la linéarité de E .

Théorème 17. Si X et Y sont deux variables aléatoires indépendantes possédant une espérance mathématique :

$$E(XY) = E(X) E(Y).$$

$$E(XY) = \sum_{i,j} x_i y_j \underbrace{p_{ij}}_{=p_{i.}p_{.j}} = \left(\sum_i x_i p_{i.} \right) \left(\sum_j y_j p_{.j} \right) = E(X) E(Y).$$

Théorème 18. Si les variables aléatoires X et Y sont indépendantes leur covariance est nulle.

La réciproque est fausse.

La démonstration est immédiate.

Un contre-exemple pour la réciproque est donné ci-contre.

	$X = -1$	$X = 0$	$X = 1$
$Y = 0$	0	$\frac{1}{2}$	0
$Y = 1$	$\frac{1}{4}$	0	$\frac{1}{4}$

Théorème 19. Variance de la somme :

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j).$$

$$\begin{aligned} \text{Var}\left(\sum_{i=1}^n X_i\right) &= E\left(\left(\sum_{i=1}^n X_i - E\left(\sum_{i=1}^n X_i\right)\right)^2\right) = E\left(\left(\sum_{i=1}^n (X_i - E(X_i))\right)^2\right) \\ &= E\left(\sum_{i=1}^n (X_i - E(X_i))^2 + 2 \sum_{1 \leq i < j \leq n} (X_i - E(X_i))(X_j - E(X_j))\right) \\ &= \sum_{i=1}^n \underbrace{E(X_i - E(X_i))^2}_{=\text{Var}(X_i)} + 2 \sum_{1 \leq i < j \leq n} \underbrace{E(X_i - E(X_i))(X_j - E(X_j))}_{=\text{Cov}(X_i, X_j)} \end{aligned}$$

Cas particulier important : si les variables aléatoires sont deux à deux indépendantes la variance de leur somme est égale à la somme de leurs variances.

Définition 17. Le coefficient de corrélation des variables aléatoires X et Y est défini par :

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X) \sigma(Y)}$$

si X et Y possèdent des écart-types non nuls.

Théorème 20. Si X et Y sont indépendantes leur coefficient de corrélation est nul.

La réciproque est fausse ; le coefficient de corrélation donne donc une indication plus ou moins précise de l'indépendance de deux variables aléatoires. Cependant on verra plus loin que la réciproque est exacte pour la loi normale que l'on utilise fréquemment : ceci explique l'intérêt de ce coefficient.

Définition 18. Si $\rho(X, Y)$ est nul les deux variables sont dites non-corrélées.

Théorème 21. 1. On a : $\rho(X, Y) \in [-1, 1]$.

2. Si $Y = \alpha X + \beta$ (avec $\alpha \neq 0$) on a $|\rho(X, Y)| = 1$.

3. Si $|\rho(X, Y)| = 1$ les variables aléatoires X et Y sont liées par la relation affine :

$$\frac{X - \mathbf{E}(X)}{\sigma(X)} - \rho(X, Y) \frac{Y - \mathbf{E}(Y)}{\sigma(Y)} = 0 \quad (\text{avec } \rho(X, Y) = \pm 1!).$$

1. L'espérance mathématique étant une forme linéaire positive :

$$\forall \lambda \in \mathbb{R}, \mathbf{E} \left(\left((X - \mathbf{E}(X)) + \lambda(Y - \mathbf{E}(Y)) \right)^2 \right) \geq 0.$$

D'où :

$$\forall \lambda \in \mathbb{R}, \lambda^2 \text{Var}(Y) + 2\lambda \text{Cov}(X, Y) + \text{Var}(X) \geq 0.$$

Le discriminant du trinôme est négatif ou nul c'est-à-dire : $\text{Cov}(X, Y)^2 \leq \text{Var}(X) \text{Var}(Y)$; donc $\rho(X, Y)^2 \leq 1$.

2. On a :

$$\mathbf{E}(Y) = \alpha \mathbf{E}(X) + \beta, \sigma(Y) = |\alpha| \sigma(X), \mathbf{E}(XY) = \alpha \mathbf{E}(X^2) + \beta \mathbf{E}(X).$$

Donc :

$$\text{Cov}(X, Y) = \alpha \mathbf{E}(X^2) + \beta \mathbf{E}(X) - \mathbf{E}(X)(\alpha \mathbf{E}(X) + \beta) = \alpha \text{Var}(X)$$

et

$$\rho(X, Y) = \frac{\alpha \text{Var}(X)}{\sigma(X)(|\alpha| \sigma(X))} = \begin{cases} 1 & \text{si } \alpha > 0; \\ -1 & \text{si } \alpha < 0. \end{cases}$$

3. Si $|\rho(X, Y)| = 1$ le discriminant du trinôme est nul : le trinôme a alors une racine double $\lambda_0 = -\text{Cov}(X, Y)/\text{Var}(Y)$ et donc $\mathbf{E} \left(\left((X - \mathbf{E}(X)) + \lambda_0(Y - \mathbf{E}(Y)) \right)^2 \right) = 0$.

Ceci entraîne $(X - \mathbf{E}(X)) + \lambda_0(Y - \mathbf{E}(Y)) = 0$, soit, en reportant et en simplifiant :

$$(X - \mathbf{E}(X)) - \rho(X, Y) \frac{\sigma(X)}{\sigma(Y)} (Y - \mathbf{E}(Y)) = 0.$$

2.4.2 Vecteur aléatoire

Lorsqu'on utilise le calcul matriciel on suppose que les vecteurs sont représentés par des matrices colonnes ; on note m^T la transposée de la matrice m .

Définition 19. Soit $X = (X_1 \ X_2 \ \dots \ X_n)^T$ un vecteur aléatoire. On pose, en notant :

$$\begin{aligned} \mathbf{E}(X) &= (\mathbf{E}(X_1) \ \mathbf{E}(X_2) \ \dots \ \mathbf{E}(X_n))^T, \\ \mathbf{Var}(X) &= \mathbf{E}\left((X - \mathbf{E}(X))(X - \mathbf{E}(X))^T\right) \\ &= \begin{pmatrix} \mathbf{Var}(X_1) & \mathbf{Cov}(X_1, X_2) & \dots & \mathbf{Cov}(X_1, X_n) \\ \mathbf{Cov}(X_2, X_1) & \mathbf{Var}(X_2) & \dots & \mathbf{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Cov}(X_n, X_1) & \mathbf{Cov}(X_n, X_2) & \dots & \mathbf{Var}(X_n) \end{pmatrix} \quad (\text{matrice de covariance}). \end{aligned}$$

Théorème 22. Posons $Y = (Y_1, Y_2, \dots, Y_m) = aX + b$ où a est une matrice ayant m lignes et n colonnes et $b \in \mathbb{R}^m$. On a :

$$\mathbf{E}(Y) = a\mathbf{E}(X) + b \quad \text{et} \quad \mathbf{Var}(Y) = a\mathbf{Var}(X)a^T.$$

Ce théorème généralise le théorème 14, la dernière propriété du théorème 15 et le théorème 19.

Pour l'espérance mathématique le résultat découle immédiatement de la linéarité.

$$\begin{aligned} \mathbf{Var}(Y) &= \mathbf{E}\left((Y - \mathbf{E}(Y))(Y - \mathbf{E}(Y))^T\right) = \mathbf{E}\left((aX - a\mathbf{E}(X))(aX - a\mathbf{E}(X))^T\right) \\ &= a\mathbf{E}\left((X - \mathbf{E}(X))(X - \mathbf{E}(X))^T\right)a^T = a\mathbf{Var}(X)a^T. \end{aligned}$$

Théorème 23. La matrice de covariance est symétrique, définie positive.

La symétrie est évidente.

Il existe une matrice orthogonale u telle que $u^{-1}\mathbf{Var}(X)u$ est la matrice diagonale des valeurs propres de $\mathbf{Var}(X)$; d'après le théorème précédent cette matrice est égale à $\mathbf{Var}(Y)$ avec $Y = u^T X$. Les valeurs propres de $\mathbf{Var}(X)$ sont donc positives ; elles sont non nulles si $\forall i, \mathbf{Var}(Y_i) \neq 0$, c'est-à-dire si la distribution est non dégénérée.

2.5 Inégalité de BIENAYMÉ-TCHEBYCHEV

Théorème 24. Pour toute variable aléatoire X ayant une espérance mathématique et une variance :

$$\forall \varepsilon > 0, \mathbb{P}\left(|X - \mathbf{E}(X)| \geq \varepsilon\right) \leq \frac{\mathbf{Var}(X)}{\varepsilon^2}.$$

$$\mathbf{Var}(X) = \sum_{i \in I} (x_i - \mathbf{E}(X))^2 p(x_i) \geq \sum_{i \in I'} (x_i - \mathbf{E}(X))^2 p(x_i)$$

où $I' = \{i \in I \mid |x_i - \mathbf{E}(X)| \geq \varepsilon\}$, car tous les termes sont positifs.

Minorant alors chaque $|x_i - \mathbf{E}(X)|$ par ε on obtient :

$$\mathbf{Var}(X) \geq \varepsilon^2 \sum_{i \in I'} p(x_i) = \varepsilon^2 \mathbb{P}(|X - \mathbf{E}(X)| \geq \varepsilon).$$

Remarque : le théorème ci-dessus est un cas particulier (pour $Z = X - \mathbf{E}(X)$ et $r = 2$) de l'inégalité

$$\forall r > 0, \forall \varepsilon > 0, \mathbb{P}(|Z| \geq \varepsilon) \leq \frac{\mathbf{E}(|Z|^r)}{\varepsilon^r}.$$

La démonstration est identique à celle ci-dessus :

$$\begin{aligned} \mathbf{E}(|Z|^r) &= \sum_{i \in I} |z_i|^r p(z_i) \geq \sum_{i \in I'} |z_i|^r p(z_i) \quad \text{où} \quad I' = \{i \in I \mid |z_i| \geq \varepsilon\} \\ &\geq \varepsilon^r \sum_{i \in I'} p(z_i) = \varepsilon^r \mathbb{P}(|Z| \geq \varepsilon). \end{aligned}$$

Variante : en posant $\varepsilon = \lambda \sigma(X)$ on obtient :

$$\mathbb{P}(\mathbf{E}(X) - \lambda \sigma(X) < X < \mathbf{E}(X) + \lambda \sigma(X)) > 1 - \frac{1}{\lambda^2}.$$

Sous cette forme on voit clairement que l'écart-type caractérise la dispersion de la variable aléatoire autour de son espérance mathématique.

2.6 Fonction génératrice

Définition 20. Soit X une variable aléatoire à valeurs entières positives ($X(\Omega) \subset \mathbb{N}$). La fonction génératrice de X est définie par :

$$\forall s \in \mathbb{C}, |s| \leq 1, g_X(s) = \sum_{k \in \mathbb{N}} \mathbb{P}(X = k) s^k.$$

Le rayon de convergence de la série entière de définition est au moins égal à 1 puisque la série converge pour $s = 1$:

$$g_X(1) = \sum_{k \in \mathbb{N}} \mathbb{P}(X = k) = 1.$$

La connaissance de la fonction génératrice de X entraîne celle de la loi de X : $\mathbb{P}(X = k)$ est le coefficient de s^k dans le développement en série entière de $g_X(s)$.

Théorème 25. Fonction génératrice d'une somme de variables aléatoires indépendantes.

$$X \text{ et } Y \text{ indépendantes} \Rightarrow g_{X+Y}(s) = g_X(s) g_Y(s).$$

La réciproque est fausse.

$\mathbb{P}(X + Y = k) = \mathbb{P}\left(\bigcup_{i+j=k} ((X = i) \cap (Y = j))\right) = \sum_{i=0}^k \mathbb{P}(X = i) \mathbb{P}(Y = k - i)$, car les variables aléatoires X et Y sont indépendantes et prennent des valeurs entières positives. Les séries utilisées étant sommables :

$$\begin{aligned} g_{X+Y}(s) &= \sum_{k=0}^{\infty} \left(\sum_{i=0}^k \mathbb{P}(X = i) \mathbb{P}(Y = k - i) \right) s^k = \sum_{i=0}^{\infty} \left(\sum_{k=i}^{\infty} \mathbb{P}(X = i) \mathbb{P}(Y = k - i) s^k \right) \\ &= \sum_{i=0}^{\infty} \mathbb{P}(X = i) s^i \sum_{k=i}^{\infty} \mathbb{P}(Y = k - i) s^{k-i} = \sum_{i=0}^{\infty} \mathbb{P}(X = i) s^i \sum_{j=0}^{\infty} \mathbb{P}(Y = j) s^j \\ &= g_X(s) g_Y(s). \end{aligned}$$

Théorème 26. *Généralisation :*

$$X_1, X_2, \dots, X_n \text{ mutuellement indépendantes} \Rightarrow g_{\sum_{i=1}^n X_i}(s) = \prod_{i=1}^n g_{X_i}(s).$$

Théorème 27. Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires à valeurs dans \mathbb{N} , ayant toutes la même loi qu'une variable X et mutuellement indépendantes. Soit N une variable aléatoire à valeurs dans \mathbb{N}^* indépendante de chaque X_i . On pose : $R = \sum_{i=1}^N X_i$. La fonction génératrice de R est donnée par :

$$g_R = g_N \circ g_X.$$

On a :

$$\mathbb{P}(R = k) = \sum_{n=1}^{\infty} \mathbb{P}(N = n) \mathbb{P}(R = k / N = n) = \sum_{n=1}^{\infty} \mathbb{P}(N = n) \mathbb{P}\left(\sum_{i=1}^n X_i = k\right).$$

Utilisant la sommabilité des séries on obtient :

$$\begin{aligned} g_R(s) &= \sum_{k=0}^{\infty} \sum_{n=1}^{\infty} \mathbb{P}(N = n) \mathbb{P}\left(\sum_{i=1}^n X_i = k\right) s^k = \sum_{n=1}^{\infty} \mathbb{P}(N = n) \sum_{k=0}^{\infty} \mathbb{P}\left(\sum_{i=1}^n X_i = k\right) s^k \\ &= \sum_{n=1}^{\infty} \mathbb{P}(N = n) g_{\sum_{i=1}^n X_i}(s) = \sum_{n=1}^{\infty} \mathbb{P}(N = n) (g_X(s))^n \text{ d'après la généralisation ci-dessus,} \\ &= g_N(g_X(s)). \end{aligned}$$

Théorème 28. Si la série est convergente on a :

$$g_X^{(m)}(1) = \sum_{k=m}^{\infty} k(k-1) \dots (k-m+1) \mathbb{P}(X = k) = \mathbf{E}(X(X-1) \dots (X-m+1)).$$

En particulier

$$\mathbf{E}(X) = g'_X(1) \text{ et } \mathbf{Var}(X) = g''_X(1) + g'_X(1) - g'_X(1)^2.$$

Le rayon de convergence de la série de définition de $g_X(s) = \sum_{k=0}^{\infty} \mathbb{P}(X = k)s^k$ est supérieur ou égal à 1 : on peut donc dériver terme à terme sur l'intervalle $] -1, 1[$ et l'égalité obtenue sera encore valable pour $s = 1$ si la série converge. Or :

$$g_X^{(m)}(s) = \sum_{k=m}^{\infty} \mathbb{P}(X = k)k(k-1)\dots(k-m+1)s^{k-m}.$$

Pour $m = 1$ on obtient : $g_X'(1) = \sum_{k=1}^{\infty} k\mathbb{P}(X = k) = \mathbf{E}(X)$.

Pour $m = 2$ on obtient : $g_X''(1) = \sum_{k=2}^{\infty} k(k-1)\mathbb{P}(X = k) = \mathbf{E}(X(X-1)) = \mathbf{E}(X^2) - \mathbf{E}(X)$.

2.7 Lois de probabilité discrètes classiques

2.7.1 Loi binomiale

Une urne contient des boules blanches et des boules noires. La probabilité de tirer une boule blanche au hasard est égale à p ; $q = 1 - p$ est la probabilité de tirer une boule noire. On tire n boules au hasard et *avec remise* ; on étudie la loi de probabilité de la variable aléatoire X égale au nombre de boules blanches tirées.

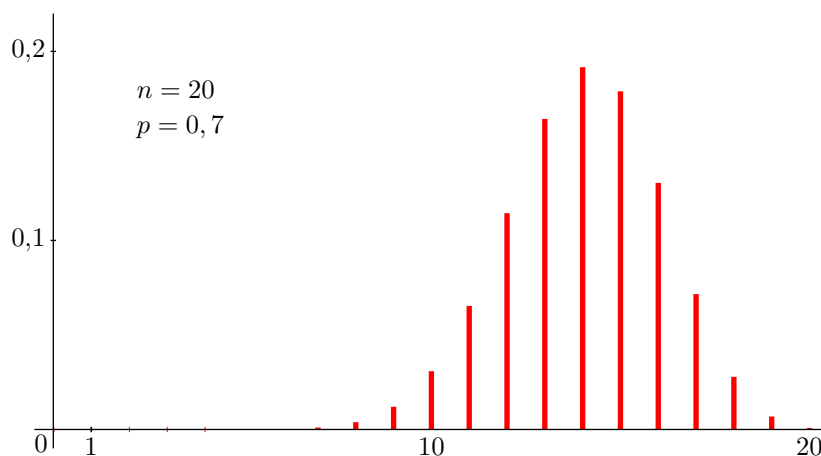
La variable aléatoire X peut prendre des valeurs entières comprises entre 0 et n . Déterminons la probabilité pour que X soit égale à k . L'univers Ω est constitué des suites des résultats des n tirages ; la probabilité d'une suite correspondant au tirage de k boules blanches et $(n - k)$ boules noires est égale à $p^k q^{n-k}$, car les tirages sont indépendants (il y a remise de la boule tirée après chaque tirage). Le nombre de suites incompatibles correspondant au tirage de k boules blanches est égal au nombre de façons de placer k boules parmi n : il y en a $\binom{n}{k}$ et la probabilité cherchée est donc :

$$\forall k \in \mathbb{N}, 0 \leq k \leq n, \mathbb{P}(X = k) = \binom{n}{k} p^k q^{n-k}.$$

On vérifie immédiatement qu'il s'agit d'une loi de probabilité :

$$\sum_{k=0}^n \mathbb{P}(X = k) = (p + q)^n = 1.$$

On dit que X est une *variable binomiale* de loi $\mathcal{B}(n, p)$: une variable binomiale représente le *nombre de succès* dans une suite d'épreuves *indépendantes*.



Théorème 29. La fonction génératrice de la loi binomiale $\mathcal{B}(n, p)$ est : $g_X(s) = (ps + q)^n$.

$$g_X(s) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} s^k = (ps + q)^n.$$

Théorème 30. Espérance mathématique et variance :

$$\mathbf{E}(X) = np, \quad \mathbf{Var}(X) = npq.$$

Pour tout s réel on a

$$g'_X(s) = np(ps + q)^{n-1} \text{ et } g''_X(s) = n(n-1)p^2(ps + q)^{n-2},$$

et donc (voir théorème 28)

$$\begin{aligned} \mathbf{E}(X) &= g'_X(1) = np, \\ \mathbf{Var}(X) &= g''_X(1) + g'_X(1) - (g'_X(1))^2 = n(n-1)p^2 + np - (np)^2 = np(1-p). \end{aligned}$$

Théorème 31 (stabilité). Si X et Y sont des variables binomiales indépendantes de lois $\mathcal{B}(n_1, p)$ et $\mathcal{B}(n_2, p)$ leur somme $(X + Y)$ est une variable binomiale de loi $\mathcal{B}(n_1 + n_2, p)$.

D'après le théorème 25 :

$$g_{X+Y}(s) = g_X(s)g_Y(s) = (ps + q)^{n_1}(ps + q)^{n_2} = (ps + q)^{n_1+n_2}.$$

On peut aussi remarquer que si l'on tire successivement n_1 boules puis n_2 boules on a tiré $(n_1 + n_2)$ boules...

Remarque : Lorsque n est grand le calcul de $b(k, n, p) = \binom{n}{k} p^k q^{n-k}$ est délicat ; on peut utiliser des expressions approchées : voir le paragraphe 4.4.

2.7.2 Loi multinomiale

Une population est composée d'éléments de m types, la proportion d'éléments de type j étant égale à p_j ($\sum_{j=1}^m p_j = 1$). On tire avec remise dans cette population un échantillon de taille n . Soit N_j le nombre d'éléments de type j tirés et $\mathbf{N} = (N_1, N_2, \dots, N_m)$; on a $N_1 + N_2 + \dots + N_m = n$.

Soient n_1, n_2, \dots, n_m des entiers naturels tels que $n_1 + n_2 + \dots + n_m = n$. La probabilité d'un tirage comportant n_j éléments de type j est égale à $p_1^{n_1} p_2^{n_2} \dots p_m^{n_m}$ et le nombre de tirages (incompatibles)

de cette sorte est $\frac{n!}{n_1! n_2! \dots n_m!}$ (cf. § A.3). Donc :

$$\mathbb{P}\left(\bigcap_{j=1}^m (N_j = n_j)\right) = \frac{n!}{n_1! n_2! \dots n_m!} p_1^{n_1} p_2^{n_2} \dots p_m^{n_m}.$$

Remarque : pour tout j la variable aléatoire N_j suit la loi binomiale $\mathcal{B}(n, p_j)$.

$$\mathbf{E}(N_j) = np_j \quad \mathbf{Var}(N_j) = np_j(1 - p_j)$$

Théorème 32. On a, pour $j \neq k$:

$$\text{Cov}(\mathbf{N}_j, \mathbf{N}_k) = -n p_j p_k.$$

Les variables $(\mathbf{N}_j)_{1 \leq j \leq m}$ ne sont donc pas indépendantes, ce qui est d'ailleurs assez évident puisque $\sum_{j=1}^m \mathbf{N}_j = n$.

- Dans le cas où $m = 2$ le calcul direct est aisé ; puisque $\mathbf{N}_1 + \mathbf{N}_2 = n$ on a :

$$\mathbf{E}(\mathbf{N}_1 \mathbf{N}_2) = \mathbf{E}(\mathbf{N}_1(n - \mathbf{N}_1)) = n\mathbf{E}(\mathbf{N}_1) - \mathbf{E}(\mathbf{N}_1^2) = n(np_1) - (np_1(1 - p_1) + (np_1)^2).$$

En reportant (et en rappelant que $p_1 + p_2 = 1$) :

$$\text{Cov}(\mathbf{N}_1, \mathbf{N}_2) = \mathbf{E}(\mathbf{N}_1 \mathbf{N}_2) - \mathbf{E}(\mathbf{N}_1)\mathbf{E}(\mathbf{N}_2) = n^2 p_1 - np_1 p_2 - n^2 p_1^2 - n^2 p_1 p_2 = -np_1 p_2.$$

- Pour le cas général la démonstration utilise les fonctions caractéristiques (voir paragraphe 3.5). On introduit les matrices :

$$p = \begin{pmatrix} p_1 & p_2 & \cdots & p_m \end{pmatrix}^T \quad \text{et} \quad d = \begin{pmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & p_m \end{pmatrix}.$$

On a :

$$\begin{aligned} \widehat{f}_{\mathbf{N}}(t) &= \mathbf{E}(\exp(it^T \mathbf{N})) = \sum_{\sum n_j = n} \frac{n!}{n_1! n_2! \cdots n_m!} p_1^{n_1} p_2^{n_2} \cdots p_m^{n_m} \exp\left(\sum_{j=1}^m t_j n_j\right) \\ &= \left(\sum_{j=1}^m p_j \exp(it_j)\right)^n \\ &= \left(\sum_{j=1}^m p_j \left(1 + it_j - \frac{1}{2} t_j^2 + o(t_j^2)\right)\right)^n \\ &= \left(1 + it^T p - \frac{1}{2} t^T d t + o(\|t\|^2)\right)^n \\ &= 1 + it^T (np) - \frac{n}{2} t^T d t - \frac{n(n-1)}{2} (t^T p)^2 + o(\|t\|^2). \end{aligned}$$

D'après le théorème 50 on obtient :

$$\mathbf{E}(\mathbf{N}) = n p,$$

ce qui n'est pas une découverte, et

$$\mathbf{E}(\mathbf{N} \mathbf{N}^T) = nd + n(n-1)p p^T \quad (\text{car } (t^T p)^2 = t^T p p^T t).$$

D'où :

$$\begin{aligned} \text{Var}(\mathbf{N}) &= \mathbf{E}(\mathbf{N} \mathbf{N}^T) - \mathbf{E}(\mathbf{N}) \mathbf{E}(\mathbf{N})^T = nd + n(n-1)p p^T - n^2 p p^T = n(d - p p^T) \\ &= n \begin{pmatrix} p_1(1-p_1) & -p_1 p_2 & \cdots & -p_1 p_m \\ -p_1 p_2 & p_2(1-p_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & -p_{m-1} p_m \\ -p_1 p_m & \cdots & -p_{m-1} p_m & p_m(1-p_m) \end{pmatrix}. \end{aligned}$$

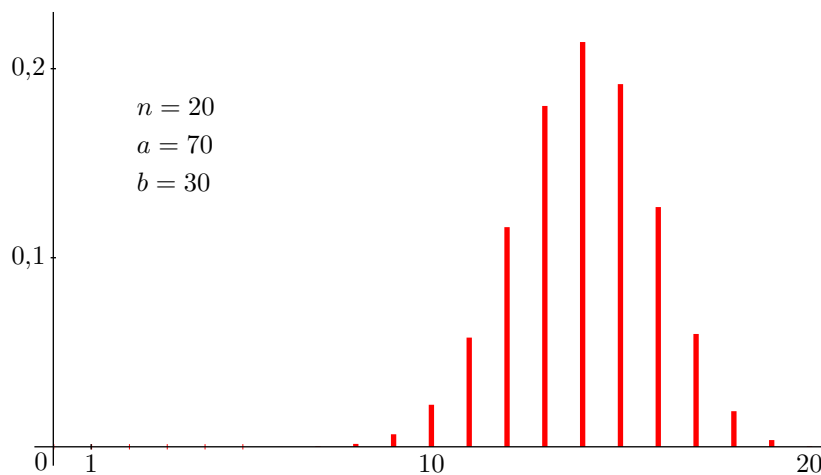
2.7.3 Loi hypergéométrique

Une urne contient a boules blanches et b boules noires ; on tire n boules de l'urne au hasard et *sans remise* (tirage exhaustif). Déterminons la loi de la variable aléatoire X , nombre de boules blanches tirées.

Le tirage ayant lieu sans remise on peut tirer les n boules en une seule fois : le nombre de tirages est égal au nombre de sous-ensembles de n boules parmi $(a + b)$, soit $\binom{a+b}{n}$. Ces tirages sont équiprobables si les boules ne diffèrent que par la couleur.

Le nombre de tirages donnant k boules blanches est égal au produit du nombre de façons de choisir k boules blanches parmi a et du nombre de façons de choisir $(n - k)$ boules noires parmi b , soit $\binom{a}{k} \binom{b}{n-k}$. D'où :

$$\mathbb{P}(X = k) = \frac{\binom{a}{k} \binom{b}{n-k}}{\binom{a+b}{n}}.$$



Théorème 33. *Espérance et variance :*

$$\mathbf{E}(X) = n \frac{a}{a+b} \quad \mathbf{Var}(X) = n \frac{a}{a+b} \frac{b}{a+b} \frac{a+b-n}{a+b-1}.$$

Il n'est pas interdit de faire le calcul direct à partir de la définition, mais c'est assez pénible ; il est plus intéressant d'utiliser une méthode indirecte.

On note que $X = \sum_{i=1}^n X_i$ où X_i représente le nombre de boule blanche tirée au i -ième tirage ; pour tout i la variable aléatoire X_i suit la loi de BERNOULLI, c'est-à-dire la loi binomiale $\mathcal{B}(1, p)$ avec $p = \frac{a}{a+b}$. On a

$$\forall i \in \{1, \dots, n\}, \mathbf{E}(X_i) = p \text{ et } \mathbf{Var}(X_i) = pq = \frac{ab}{(a+b)^2} \text{ où } q = 1 - p = \frac{b}{a+b},$$

et donc

$$\mathbf{E}(X) = \sum_{i=1}^n \mathbf{E}(X_i) = np = n \frac{a}{a+b}.$$

Les variables aléatoires X_i ne sont pas indépendantes puisque les tirages ont lieu sans remise. La variable aléatoire $X_j X_k$ suit la loi de BERNOULLI dont la probabilité de succès, et donc l'espérance mathématique, est

$$\mathbb{P}(X_j X_k = 1) = \mathbb{P}((X_j = 1) \cap (X_k = 1)) = \mathbb{P}(X_j = 1) \mathbb{P}_{(X_j=1)}(X_k = 1) = \frac{a}{a+b} \frac{a-1}{a+b-1}.$$

D'où

$$\text{Cov}(X_j, X_k) = \frac{a(a-1)}{(a+b)(a+b-1)} - \left(\frac{a}{a+b}\right)^2 = \frac{-ab}{(a+b)^2(a+b-1)}$$

et

$$\begin{aligned} \text{Var}(X) &= \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq j < k \leq n} \text{Cov}(X_j, X_k) \\ &= n \frac{ab}{(a+b)^2} + 2 \frac{n(n-1)}{2} \frac{-ab}{(a+b)^2(a+b-1)} = \frac{nab(a+b-n)}{(a+b)^2(a+b-1)} \\ &= n \frac{a}{a+b} \frac{b}{a+b} \frac{a+b-n}{a+b-1} = npq \frac{a+b-n}{a+b-1}. \end{aligned}$$

Remarque : il n'est peut-être pas évident que chaque variable X_i suit la loi binomiale $\mathcal{B}(1, p)$; vérifions-le pour X_2 :

$$\begin{aligned} \mathbb{P}(X_2 = 1) &= \mathbb{P}((X_2 = 1) \cap ((X_1 = 0) \cup (X_1 = 1))) \\ &= \mathbb{P}(X_1 = 0) \mathbb{P}_{(X_1=0)}(X_2 = 1) + \mathbb{P}(X_1 = 1) \mathbb{P}_{(X_1=1)}(X_2 = 1) \\ &= \frac{b}{a+b} \frac{a}{a+b-1} + \frac{a}{a+b} \frac{a-1}{a+b-1} = \frac{a}{a+b}. \end{aligned}$$

Remarque : la loi hypergéométrique est voisine de la loi binomiale ; la seule différence est qu'il y a tirage avec ou sans remise. On constate que le fait de ne pas remettre les boules ne modifie pas l'espérance mathématique mais diminue la variance ($\frac{a+b-n}{a+b-1} \leq 1$).

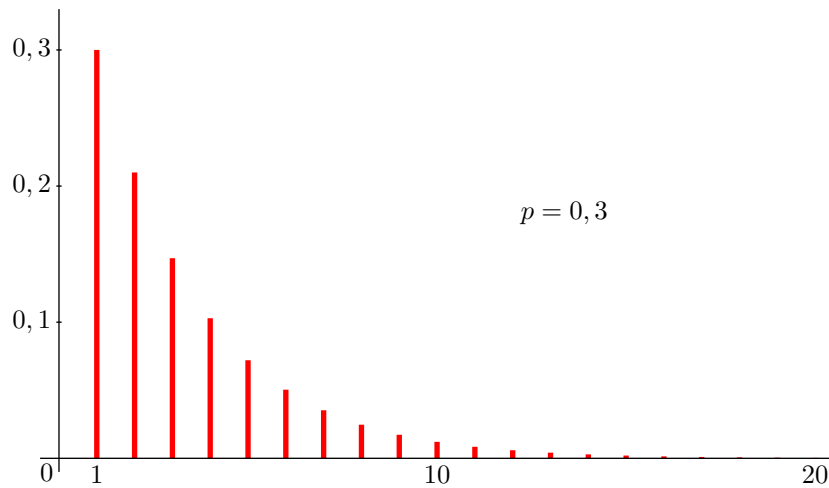
2.7.4 Loi géométrique (ou loi de PASCAL)

Une urne contient des boules blanches et des boules noires. La probabilité de tirer une boule blanche au hasard est égale à p ($q = 1 - p$ est la probabilité de tirer une boule noire). On tire des boules au hasard et *avec remise* jusqu'à ce qu'on obtienne la première boule blanche. On étudie la loi de probabilité de la variable aléatoire X égale au nombre de boules tirées.

La variable aléatoire X peut prendre toutes les valeurs entières strictement positives. La probabilité pour que X soit égale à k est celle de la suite correspondant au tirage de $(k-1)$ boules noires suivies d'une boule blanche : elle est égale à $q^{k-1}p$, car les tirages sont indépendants.

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

On dit que X est une variable géométrique de loi $\mathcal{G}(p)$: une variable géométrique représente le *nombre d'épreuves indépendantes* nécessaires pour obtenir le *premier succès*.



Théorème 34. La fonction génératrice de la loi géométrique $\mathcal{G}(p)$ est :

$$g_X(s) = \frac{ps}{1 - qs}.$$

$$g_X(s) = \sum_{k=1}^{\infty} (q^{k-1}p)s^k = \frac{ps}{1 - qs} \text{ avec } |qs| < 1.$$

Théorème 35. Espérance mathématique et variance :

$$\mathbf{E}(X) = \frac{1}{p}, \quad \mathbf{Var}(X) = \frac{q}{p^2}.$$

On a

$$g'_X(s) = \frac{p}{(1 - qs)^2} \quad \text{et} \quad g''_X(s) = \frac{2pq}{(1 - qs)^3}.$$

D'où (théorème 28)

$$\mathbf{E}(X) = g'_X(1) = \frac{p}{(1 - q)^2} = \frac{1}{p}$$

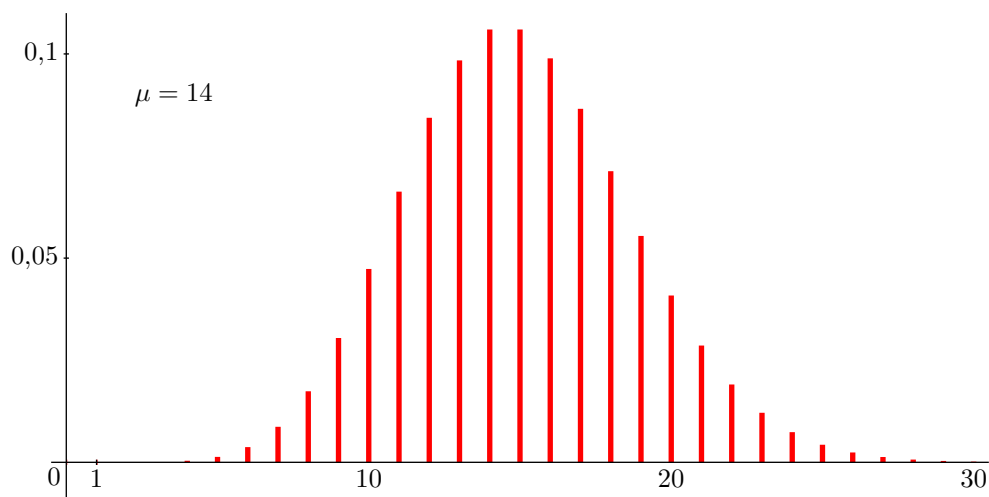
et

$$\mathbf{Var}(X) = g''_X(1) + g'_X(1) - (g'_X(1))^2 = \frac{2pq}{(1 - q)^3} + \frac{1}{p} - \frac{1}{p^2} = \frac{q}{p^2}.$$

2.7.5 Loi de POISSON

Définition 21. La variable aléatoire X suit la loi de POISSON $\mathcal{P}(\mu)$ si :

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



Théorème 36. La fonction génératrice de la loi de POISSON $\mathcal{P}(\mu)$ est :

$$g_X(s) = e^{\mu(s-1)}.$$

$$g_X(s) = \sum_{k=0}^{\infty} e^{-\mu} \frac{(\mu s)^k}{k!} = e^{-\mu} e^{\mu s} = e^{\mu(s-1)}.$$

Théorème 37. Espérance mathématique et variance :

$$\mathbf{E}(X) = \mu, \quad \mathbf{Var}(X) = \mu.$$

On a :

$$g'_X(s) = \mu e^{\mu(s-1)} \quad \text{et} \quad g''_X(s) = \mu^2 e^{\mu(s-1)}$$

D'où (théorème 28)

$$\mathbf{E}(X) = g'_X(1) = \mu \text{ et } \mathbf{Var}(X) = g''_X(1) + g'_X(1) - (g'_X(1))^2 = \mu^2 + \mu - \mu^2 = \mu.$$

Théorème 38 (stabilité). Si X et Y sont des variables de POISSON indépendantes de lois $\mathcal{P}(\mu_1)$ et $\mathcal{P}(\mu_2)$ leur somme $(X + Y)$ est une variable de POISSON de loi $\mathcal{P}(\mu_1 + \mu_2)$.

D'après le théorème 25 :

$$g_{X+Y}(s) = g_X(s)g_Y(s) = e^{\mu_1(s-1)}e^{\mu_2(s-1)} = e^{(\mu_1+\mu_2)(s-1)}.$$

Processus de POISSON

Un événement A (appel téléphonique arrivant à un central, par exemple) peut se produire à n'importe quel instant $t > 0$. On fait les hypothèses suivantes :

1. La probabilité pour que A se produise entre les instants t et $t + \tau$ est égale à $\lambda\tau$ où λ est une constante indépendante de t , de τ et de la réalisation antérieure de A .
2. L'événement A ne peut se produire plus d'une fois entre les instants t et $t + \tau$; plus précisément la probabilité pour que A se produise deux fois ou plus entre t et $t + \tau$ est un infiniment petit par rapport à τ .

La variable aléatoire X_t qui représente le nombre de réalisations de A entre les instants 0 et t suit la loi de POISSON $\mathcal{P}(\lambda t)$.

On pose $p_n(t) = \mathbb{P}(X_t = n)$ et on note $\left[\frac{nA}{t_1, t_2} \right]$ l'événement « A se produit n fois entre les instants t_1 et t_2 ».

Calculons $p_n(t)$; on commence par le cas $n = 0$ puis on effectue une récurrence.

$$\begin{aligned} p_0(t + \tau) &= \mathbb{P}\left(\left[\frac{0A}{0, t + \tau}\right]\right) = \mathbb{P}\left(\left[\frac{0A}{0, t}\right] \cap \left[\frac{0A}{t, t + \tau}\right]\right) = \mathbb{P}\left(\left[\frac{0A}{0, t}\right]\right) \mathbb{P}\left(\left[\frac{0A}{t, t + \tau}\right]\right) \\ &= p_0(t) \mathbb{P}\left(\left[\frac{0A}{0, \tau}\right]\right) = p_0(t)(1 - \lambda\tau - o(\tau)). \end{aligned}$$

D'où :

$$\frac{p_0(t + \tau) - p_0(t)}{\tau} = -(\lambda + o(1))p_0(t) \xrightarrow{\tau \rightarrow 0} p'_0(t) = -\lambda p_0(t).$$

Tenant compte de la condition initiale $p_0(0) = 1$ on obtient immédiatement :

$$p_0(t) = e^{-\lambda t}.$$

Récurrence sur n :

$$\left[\frac{nA}{0, t + \tau} \right] = \left(\left[\frac{nA}{0, t} \right] \cap \left[\frac{0A}{t, t + \tau} \right] \right) \cup \left(\left[\frac{(n-1)A}{0, t} \right] \cap \left[\frac{1A}{t, t + \tau} \right] \right) \cup \left(\bigcup_{k=2}^n \left(\left[\frac{(n-k)A}{0, t} \right] \cap \left[\frac{kA}{t, t + \tau} \right] \right) \right).$$

D'où

$$p_n(t + \tau) = p_n(t)(1 - \lambda\tau - o(\tau)) + \lambda\tau p_{n-1}(t) + \sum_{k=2}^n p_{n-k}(t)o(\tau).$$

et

$$\frac{p_n(t + \tau) - p_n(t)}{\tau} = -(\lambda + o(1))p_n(t) + \lambda p_{n-1}(t) + \sum_{k=2}^n p_{n-k}(t)o(1) \xrightarrow{\tau \rightarrow 0} p'_n(t) = -\lambda p_n(t) + \lambda p_{n-1}(t).$$

La solution générale de l'équation homogène est : $p_n(t) = C_n e^{-\lambda t}$. On détermine la solution générale de l'équation complète par la « méthode de la variation de la constante » :

$$p'_n(t) = (C'_n - \lambda C_n) e^{-\lambda t} = -\lambda C_n e^{-\lambda t} + \lambda p_{n-1}(t).$$

D'où : $C'_n = \lambda p_{n-1}(t) e^{\lambda t}$. Pour $n = 1$:

$$C'_1 = \lambda p_0(t) e^{\lambda t} = \lambda \Rightarrow C_1 = \lambda t + K_1 \Rightarrow p_1(t) = (\lambda t + K_1) e^{-\lambda t}.$$

Or, d'après la condition initiale : $p_1(0) = 0 = K_1$; donc :

$$p_1(t) = \lambda t e^{-\lambda t}.$$

On suppose (hypothèse de récurrence vérifiée pour $n = 1$ et $n = 2$) : $p_{n-1}(t) = \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t}$.

Alors :

$$C'_n = \lambda^n \frac{t^{n-1}}{(n-1)!} \Rightarrow C_n = \frac{(\lambda t)^n}{n!} + K_n \Rightarrow p_n(t) = \left(\frac{(\lambda t)^n}{n!} + K_n \right) e^{-\lambda t}.$$

D'après la condition initiale $p_n(0) = 0 = K_n$ on obtient finalement :

$$p_n(t) = \mathbb{P}(X_t = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}.$$

Le processus de POISSON est le cas le plus simple d'apparition « naturelle » de la loi de POISSON.

Chapitre 3

Variables aléatoires absolument continues

Dans ce chapitre l'univers Ω est non dénombrable ; la tribu \mathcal{E} des événements doit être strictement incluse dans $\mathcal{P}(\Omega)$. On suppose définie une probabilité \mathbb{P} sur (Ω, \mathcal{E}) .

3.1 Variable et vecteur aléatoires

Définition 22. La tribu de BOREL sur \mathbb{R}^n est la plus petite tribu contenant les pavés ouverts.

Elle contient également les pavés fermés et, plus généralement, toutes les parties « usuelles » de \mathbb{R}^n . Ses éléments s'appellent les boréliens de \mathbb{R}^n .

Définition 23. Une variable (resp. un vecteur) aléatoire sur l'espace probabilisable (Ω, \mathcal{E}) est une application X de Ω dans \mathbb{R} (resp. dans \mathbb{R}^n) qui vérifie : pour tout borélien B de \mathbb{R} (resp. de \mathbb{R}^n) on a $X^{-1}(B) \in \mathcal{E}$.

Il suffit de vérifier que l'image réciproque de toute demi-droite $] - \infty, x[$ (resp. de toute partie de \mathbb{R}^n de la forme $\prod_{i=1}^n] - \infty, x_i[$) est un événement.

Exemple (trivial) : toute application constante est une variable aléatoire. Soit $X = k$ cette variable :

$$X^{-1}(] - \infty, x[) = \begin{cases} \emptyset & \text{si } x \leq k, \\ \Omega & \text{si } x > k. \end{cases}$$

3.2 Fonction de répartition. Densité

3.2.1 Variable aléatoire à une dimension

Définition 24. La fonction de répartition de la variable aléatoire X est l'application F définie sur \mathbb{R} par :

$$F(x) = \mathbb{P}(X^{-1}(] - \infty, x[)).$$

Théorème 39. La fonction de répartition vérifie les propriétés suivantes :

1. F est croissante ;

2. F est continue à gauche en tout point ;
 3. $\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1.$

1. Supposons $x < y$:

$$\begin{aligned}
 F(y) &= \mathbb{P}\left(X^{-1}(] - \infty, y[)\right) = \mathbb{P}\left(X^{-1}(] - \infty, x[\cup [x, y[)\right) \\
 &= \mathbb{P}\left(X^{-1}(] - \infty, x[) \cup X^{-1}([x, y[)\right) \quad (\text{propriété de l'image réciproque}) \\
 &= \underbrace{\mathbb{P}\left(X^{-1}(] - \infty, x[)\right)}_{=F(x)} + \underbrace{\mathbb{P}\left(X^{-1}([x, y[)\right)}_{\geq 0} \quad (\text{événements incompatibles}) \\
 &\geq F(x).
 \end{aligned}$$

2. Soit $a \in \mathbb{R}$:

$$X^{-1}(] - \infty, a[) = \lim_{n \rightarrow \infty} X^{-1}(] - \infty, a - \frac{1}{n}[)$$

la suite étant croissante. D'après le théorème 2 :

$$\begin{aligned}
 F(a) &= \mathbb{P}\left(X^{-1}(] - \infty, a[)\right) = \mathbb{P}\left(\lim_{n \rightarrow \infty} X^{-1}(] - \infty, a - \frac{1}{n}[)\right) \\
 &= \lim_{n \rightarrow \infty} \mathbb{P}\left(X^{-1}(] - \infty, a - \frac{1}{n}[)\right) = \lim_{n \rightarrow \infty} F(a - \frac{1}{n}) \\
 &= F(a - 0).
 \end{aligned}$$

On peut montrer par ailleurs que F admet une limite à droite en tout point. On a

$$X^{-1}(] - \infty, a]) = \lim_{n \rightarrow \infty} X^{-1}(] - \infty, a + \frac{1}{n}[)$$

la suite étant décroissante. D'après le théorème 2 :

$$\begin{aligned}
 \mathbb{P}(X \leq a) &= \mathbb{P}\left(X^{-1}(] - \infty, a])\right) = \mathbb{P}\left(\lim_{n \rightarrow \infty} X^{-1}(] - \infty, a + \frac{1}{n}[)\right) \\
 &= \lim_{n \rightarrow \infty} \mathbb{P}\left(X^{-1}(] - \infty, a + \frac{1}{n}[)\right) = \lim_{n \rightarrow \infty} F(a + \frac{1}{n}) = F(a + 0).
 \end{aligned}$$

Remarque :

$$\mathbb{P}(X = a) = \mathbb{P}(X \leq a) - \mathbb{P}(X < a) = \lim_{x \rightarrow a^+} F(x) - F(a).$$

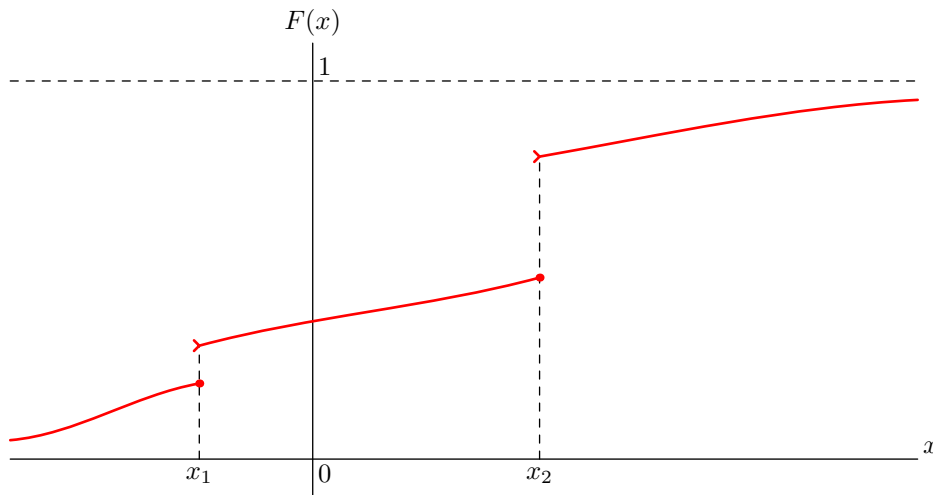
Si F est continue en a alors $\mathbb{P}(X = a) = 0$.

3. La suite d'événements $\left(X^{-1}(] - \infty, -n[)\right)_{n \in \mathbb{N}}$ est décroissante :

$$\begin{aligned}
 \lim_{x \rightarrow -\infty} F(x) &= \lim_{n \rightarrow \infty} \mathbb{P}\left(X^{-1}(] - \infty, -n[)\right) \\
 &= \mathbb{P}\left(\lim_{n \rightarrow \infty} X^{-1}(] - \infty, -n[)\right) \quad (\text{voir théorème 2}) \\
 &= \mathbb{P}(X^{-1}(\emptyset)) = \mathbb{P}(\emptyset) = 0.
 \end{aligned}$$

La suite d'événements $\left(X^{-1}(]-\infty, n[))\right)_{n \in \mathbb{N}}$ est croissante :

$$\begin{aligned} \lim_{x \rightarrow \infty} F(x) &= \lim_{n \rightarrow \infty} \mathbb{P}\left(X^{-1}(]-\infty, n[))\right) \\ &= \mathbb{P}\left(\lim_{n \rightarrow \infty} X^{-1}(]-\infty, n[))\right) \quad (\text{voir théorème 2}) \\ &= \mathbb{P}(X^{-1}(\mathbb{R})) = \mathbb{P}(\Omega) = 1. \end{aligned}$$



La fonction de répartition représentée sur le graphe ci-dessus possède deux points de discontinuité d'abscisses x_1 et x_2 ; nous supposons par la suite que toute fonction de répartition est continue sur \mathbb{R} et qu'il existe une application $f : \mathbb{R} \mapsto \mathbb{R}$, continue sauf en un nombre fini de points, telle que :

$$\forall x \in \mathbb{R}, F(x) = \int_{-\infty}^x f(t) dt.$$

La variable aléatoire X est alors dite absolument continue ; la fonction f est sa *densité de probabilité*. En faisant l'analogie entre une probabilité et une masse cela correspond au cas où la masse est répartie et où il n'y a pas de masse ponctuelle.

On a alors :

$$\mathbb{P}(X^{-1}[a, b[)) = \mathbb{P}(a \leq X < b) = F(b) - F(a) = \int_a^b f(x) dx.$$

3.2.2 Vecteurs aléatoires

Définition 25. La fonction de répartition du vecteur aléatoire $X = (X_1, X_2, \dots, X_n)$ est l'application F définie sur \mathbb{R}^n par :

$$F(x_1, x_2, \dots, x_n) = \mathbb{P}\left(\bigcap_{i=1}^n (X_i^{-1}(]-\infty, x_i[))\right).$$

On supposera qu'il existe une densité de probabilité $f : \mathbb{R}^n \mapsto \mathbb{R}$, continue sauf en un nombre fini de points, telle que :

$$F(x_1, x_2, \dots, x_n) = \int \int \dots \int_{K_n} f(t_1, t_2, \dots, t_n) dt_1 dt_2 \dots dt_n \quad \text{où } K_n = \prod_{i=1}^n]-\infty, x_i[.$$

On a alors, pour tout borélien B de \mathbb{R}^n :

$$\mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B) = \int \int \dots \int_B f(t_1, t_2, \dots, t_n) dt_1 dt_2 \dots dt_n.$$

Si l'on connaît la densité d'un vecteur aléatoire on peut en déduire la densité de ses composantes ; on donne le résultat dans le cas le plus simple :

Théorème 40. Soit $V = (X, Y)$ un couple aléatoire ayant f pour densité ; les densités de probabilité des variables marginales X et Y sont données par :

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

On a :

$$F_X(x) = \mathbb{P}(X < x) = \mathbb{P}((X < x) \cap (Y \in \mathbb{R})) = \int \int_{D_x} f(t, y) dt dy$$

où $D_x = \{(t, y) \mid t < x\}$. D'où :

$$F_X(x) = \int_{-\infty}^x \left(\int_{-\infty}^{\infty} f(t, y) dy \right) dt.$$

En dérivant par rapport à x on obtient :

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy.$$

Remarque : en général la connaissance des densités de X et Y est insuffisante pour déterminer la densité du couple (X, Y) .

Définition 26. Les variables aléatoires X et Y sont indépendantes si, quels que soient les boréliens B_1 et B_2 , les événements $X^{-1}(B_1)$ et $Y^{-1}(B_2)$ sont indépendants.

Théorème 41. Les propositions suivantes sont équivalentes.

- Les variables aléatoires X et Y sont indépendantes.
- $\forall (x, y) \in \mathbb{R}^2, F_{(X, Y)}(x, y) = F_X(x)F_Y(y)$.
- $\forall (x, y) \in \mathbb{R}^2, f_{(X, Y)}(x, y) = f_X(x)f_Y(y)$.

1. Si X et Y sont indépendantes, alors, pour tous x et y réels, les événements $(X < x)$ et $(Y < y)$ sont indépendants et

$$F(x, y) = \mathbb{P}((X < x) \cap (Y < y)) = \mathbb{P}(X < x)\mathbb{P}(Y < y) = F_X(x)F_Y(y).$$

En dérivant par rapport à x et y on obtient :

$$f(x, y) = f_X(x)f_Y(y).$$

2. On suppose : $\forall (x, y) \in \mathbb{R}^2$, $f(x, y) = f_X(x)f_Y(y)$. Alors, pour tous réels a, b, c et d :

$$\begin{aligned} \mathbb{P}((a \leq X < b) \cap (c \leq Y < d)) &= \iint_K f(x, y) dx dy, \text{ où } K = [a, b[\times [c, d[, \\ &= \int_a^b f_X(x) dx \int_c^d f_Y(y) dy \\ &= \mathbb{P}(a \leq X < b) \mathbb{P}(c \leq Y < d). \end{aligned}$$

On conclut que X et Y sont indépendantes en remarquant que les boréliens sont obtenus à partir des pavés par passage au complémentaire, intersections et réunions dénombrables (voir [16]).

Théorème 42. Si X et Y sont deux variables aléatoires indépendantes et si φ et ψ sont deux applications de \mathbb{R} dans \mathbb{R} les variables aléatoires $\varphi(X)$ et $\psi(Y)$ sont indépendantes.

Pour tous boréliens B_1 et B_2 on a :

$$\begin{aligned} \mathbb{P}((\varphi \circ X)^{-1}(B_1) \cap (\psi \circ Y)^{-1}(B_2)) &= \mathbb{P}((X^{-1}(\varphi^{-1}(B_1))) \cap (Y^{-1}(\psi^{-1}(B_2)))) \\ &= \mathbb{P}(X^{-1}(\varphi^{-1}(B_1))) \mathbb{P}(Y^{-1}(\psi^{-1}(B_2))) \\ &= \mathbb{P}((\varphi \circ X)^{-1}(B_1)) \mathbb{P}((\psi \circ Y)^{-1}(B_2)) \end{aligned}$$

Théorème 43 (Somme de deux variables). Soit un couple aléatoire $V = (X, Y)$ de densité de probabilité f . La variable aléatoire $Z = X + Y$ a pour densité de probabilité la fonction f_Z définie par :

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z - x) dx.$$

Soit F_Z la fonction de répartition de la variable aléatoire $Z = X + Y$:

$$F_Z(z) = \mathbb{P}(Z < z) = \mathbb{P}(X + Y < z) = \iint_{\Delta_z} f(x, y) dx dy \quad \text{où} \quad \Delta_z = \{(x, y) \mid x + y < z\}.$$

On effectue le changement de variables

$$\begin{cases} t &= x, \\ u &= x + y, \end{cases} \quad \text{soit} \quad \begin{cases} x &= t, \\ y &= u - t. \end{cases} \quad \text{de jacobien} \quad \frac{D(x, y)}{D(t, u)} = 1.$$

Alors :

$$\begin{aligned} F_Z(z) &= \iint_{D_z} f(t, u - t) dt du, \quad \text{où } D_z = \{(t, u) \mid u < z\}, \\ &= \int_{-\infty}^z \left(\int_{-\infty}^{\infty} f(t, u - t) dt \right) du. \end{aligned}$$

En dérivant par rapport à z :

$$f_Z(z) = \int_{-\infty}^{\infty} f(t, z - t) dt.$$

3.3 Espérance, variance, coefficient de corrélation

Définition 27. Le moment d'ordre k , k entier supérieur ou égal à 1, de la variable aléatoire X est défini par :

$$m_k(X) = \int_{-\infty}^{\infty} x^k f(x) dx$$

où f est la densité de probabilité de X , si l'intégrale est absolument convergente.

On définit l'espérance mathématique, la variance, l'écart-type, la covariance et le coefficient de corrélation comme au paragraphe 2.4 (page 17) ; les propriétés sont identiques.

L'inégalité de BIENAYMÉ-TCHEBYCHEV est encore valable.

3.4 Détermination de la densité

On peut déterminer la densité en dérivant la fonction de répartition, mais il est souvent possible, et plus commode, d'utiliser le théorème suivant.

Théorème 44. Soit \mathcal{D}_k l'algèbre des fonctions de classe \mathcal{C}^∞ de \mathbb{R}^k dans \mathbb{R} à support borné. On pose, pour tout $f : \mathbb{R}^k \rightarrow \mathbb{R}$ localement intégrable :

$$\forall \varphi \in \mathcal{D}_k, f^*(\varphi) = \int_{\mathbb{R}^k} f \varphi \quad (\text{distribution régulière associée à } f).$$

Alors, si f et g sont localement intégrables :

$$f^* = g^* \iff f = g \text{ presque partout.}$$

Remarques :

- La densité de probabilité étant essentiellement utilisée dans des intégrations on peut modifier arbitrairement sa valeur sur un ensemble de mesure nulle.
- Si f est la densité de probabilité d'une variable aléatoire X on a : $f^*(\varphi) = \mathbf{E}(\varphi(X))$.

Pour la démonstration on consultera un cours sur les distributions de SCHWARTZ ([25], [4], [28] ou, mieux, [29]).

Exemples d'applications :

1. Loi de $Z = X + Y$.

$$\forall \varphi \in \mathcal{D}_1, f_Z^*(\varphi) = \mathbf{E}(\varphi(Z)) = \mathbf{E}(\varphi(X + Y)) = \iint_{\mathbb{R}^2} \varphi(x + y) f_{(X,Y)}(x, y) dx dy.$$

On effectue le changement de variables

$$\begin{cases} t &= x, \\ z &= x + y, \end{cases} \quad \text{soit} \quad \begin{cases} x &= t, \\ y &= z - t, \end{cases} \quad \text{de jacobien} \quad \frac{D(x, y)}{D(t, z)} = 1.$$

Alors :

$$\begin{aligned} f_Z^*(\varphi) &= \iint_{\mathbb{R}^2} \varphi(z) f_{(X,Y)}(t, z - t) dt dz = \int_{\mathbb{R}} \varphi(z) \left(\int_{\mathbb{R}} f_{(X,Y)}(t, z - t) dt \right) dz \\ &= g^*(\varphi) \quad \text{où } g(z) = \int_{\mathbb{R}} f_{(X,Y)}(t, z - t) dt. \end{aligned}$$

D'après le théorème 44 f_Z est égale à g presque partout ; on peut choisir $f_Z = g$, soit :

$$f_Z(z) = \int_{\mathbb{R}} f_{(X,Y)}(t, z-t) dt.$$

On a montré à nouveau le théorème 43.

2. Loi du couple $(U, V) = (X + Y, X - Y)$.

$$\begin{aligned} \forall \varphi \in \mathcal{D}_2, f_{(U,V)}^*(\varphi) &= \mathbf{E}(\varphi(U, V)) = \mathbf{E}(\varphi(X + Y, X - Y)) \\ &= \int \int_{\mathbb{R}^2} \varphi(x + y, x - y) f_{(X,Y)}(x, y) dx dy. \end{aligned}$$

On effectue le changement de variables

$$\begin{cases} u &= x + y, \\ v &= x - y, \end{cases} \quad \text{soit} \quad \begin{cases} x &= \frac{u+v}{2}, \\ y &= \frac{u-v}{2}, \end{cases} \quad \text{de jacobien} \quad \frac{D(x, y)}{D(u, v)} = -\frac{1}{2}.$$

Alors :

$$f_{(U,V)}^*(\varphi) = \int \int_{\mathbb{R}^2} \varphi(u, v) f_{(X,Y)}\left(\frac{u+v}{2}, \frac{u-v}{2}\right) \frac{1}{2} du dv.$$

On peut donc choisir :

$$f_{(U,V)}(u, v) = \frac{1}{2} f_{(X,Y)}\left(\frac{u+v}{2}, \frac{u-v}{2}\right).$$

Théorème 45. Soit $X = (X_1 \ X_2 \ \dots \ X_n)^T$ un vecteur aléatoire de densité f_X ; définissons le vecteur aléatoire $Y = (Y_1 \ Y_2 \ \dots \ Y_n)^T$ par un changement de variables affine :

$$Y = aX + b \quad \text{où } a \in \mathcal{M}_n(\mathbb{R}) \text{ est inversible et } b \in \mathbb{R}^n.$$

La densité de probabilité de Y est définie par :

$$\forall y \in \mathbb{R}^n, f_Y(y) = \frac{1}{|\det a|} f_X(a^{-1}(y - b)).$$

$$\forall \varphi \in \mathcal{D}_n, \mathbf{E}(\varphi(Y)) = \mathbf{E}(\varphi(aX + b)) = \int_{\mathbb{R}^n} \varphi(ax + b) f_X(x) dx.$$

Effectuons le changement de variable :

$$y = ax + b, \quad x = a^{-1}(y - b), \quad \text{dont le jacobien est } \det(a^{-1}) = \frac{1}{\det a}.$$

D'où :

$$\mathbf{E}(\varphi(Y)) = \int_{\mathbb{R}^n} \varphi(y) f_X(a^{-1}(y - b)) \frac{1}{|\det a|} dy.$$

On conclut grâce au théorème 44.

Exemple : connaissant la densité de probabilité du couple aléatoire $X = (X_1, X_2)$, déterminer la densité de probabilité du couple $Y = (Y_1, Y_2) = (X_1 + X_2, X_1 - X_2)$.

On a $Y = aX$ avec

$$a = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \text{ D'où } a^{-1} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix} \text{ et } \det a = -2.$$

La densité de probabilité de Y est donc définie par :

$$f_{(Y_1, Y_2)}(y_1, y_2) = \frac{1}{2} f_{(X_1, X_2)}\left(\frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2}\right).$$

3.5 Fonction caractéristique

Définition 28. 1. La fonction caractéristique de la variable aléatoire X est définie par :

$$\widehat{f_X}(t) = \mathbf{E}(e^{itX}).$$

2. La fonction caractéristique du vecteur aléatoire $X = (X_1 \ X_2 \ \dots \ X_n)^T$ est définie par :

$$\widehat{f_X}(t) = \mathbf{E}(e^{it^T X}) \quad \text{où} \quad t^T X = \sum_{j=1}^n t_j X_j.$$

Avec les notations habituelles la fonction caractéristique d'une variable aléatoire est donnée par

$$\widehat{f_X}(t) = \sum_{k \in I} p(x_k) e^{itx_k} \quad \left| \quad \widehat{f_X}(t) = \int_{-\infty}^{\infty} e^{itx} f_X dx \right.$$

dans le cas discret, dans le cas continu.

On généralise aisément au cas d'un vecteur aléatoire.

Dans le cas continu la fonction caractéristique est la transformée de FOURIER de la densité de probabilité.

Exemples

1. Cas discret

Si $X(\Omega) \subset \mathbb{N}^*$ on a $\widehat{f_X}(t) = g_X(e^{it})$ où g_X est la fonction génératrice de X (voir § 2.6).

En particulier :

Loi binomiale $\mathcal{B}(n, p)$ $\widehat{f_X}(t) = (pe^{it} + q)^n,$	Loi géométrique $\mathcal{G}(p)$ $\widehat{f_X}(t) = \frac{pe^{it}}{1 - qe^{it}},$	Loi de POISSON $\mathcal{P}(\mu)$ $\widehat{f_X}(t) = \exp(\mu(e^{it} - 1)).$
--	---	--

2. Cas continu : loi de CAUCHY

La variable aléatoire X de densité $f_X(x) = \frac{1}{\pi} \frac{1}{1 + x^2}$ a pour fonction caractéristique

$$\widehat{f_X}(t) = e^{-|t|}.$$

La fonction caractéristique $\widehat{f_X}$ est définie par

$$\widehat{f_X}(t) = \int_{-\infty}^{\infty} e^{itx} \frac{dx}{\pi(1+x^2)}.$$

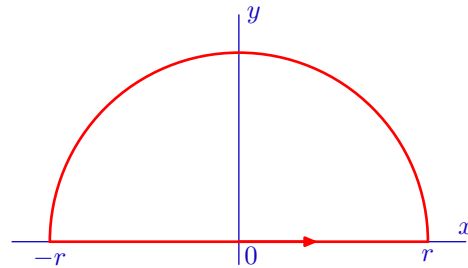
Elle est paire : on peut supposer $t > 0$.

On applique le théorème des résidus (voir [3], par exemple) en intégrant sur le lacet γ ci-contre la

fonction $g(z) = \frac{e^{itz}}{\pi(1+z^2)}$.

On obtient lorsque r tend vers l'infini :

$$\int_{\gamma} g(z) dz = \underbrace{\int_{-r}^r \frac{e^{itx}}{\pi(1+x^2)} dx}_{\rightarrow \widehat{f_X}(t)} + \underbrace{\int_0^{\pi} \frac{\exp(itre^{i\theta})}{1+r^2e^{2i\theta}} ire^{i\theta} d\theta}_{\rightarrow 0} = 2i\pi \underbrace{\text{Res}(g, i)}_{= \frac{e^{-t}}{2i\pi}}$$



D'où :

$$\widehat{f_X}(t) = e^{-|t|}.$$

Théorème 46. La fonction caractéristique d'une variable aléatoire (resp. d'un vecteur aléatoire) est une fonction continue de \mathbb{R} (resp. \mathbb{R}^n) dans \mathbb{C} et $|\widehat{f_X}| \leq 1$.

Variable à une dimension.

1. Cas discret :

$$|p(x_k)e^{itx_k}| \leq p(x_k) \text{ et } \sum_{k \in I} p(x_k) = 1.$$

2. Cas continu :

$$|f_X(x)e^{itx}| \leq f_X(x) \text{ et } \int_{\mathbb{R}} f_X(x) dx = 1.$$

La démonstration est identique pour un vecteur aléatoire.

Théorème 47. La fonction caractéristique d'une variable (resp. d'un vecteur) aléatoire X vérifie les propriétés suivantes :

1. $\widehat{f_X}(0) = 1$;
2. $\forall t \in \mathbb{R}, \widehat{f_X}(-t) = \overline{\widehat{f_X}(t)}$;
3. $\forall (a, b) \in \mathbb{R}^2, \forall t \in \mathbb{R}, \widehat{f_{aX+b}}(t) = e^{itb} \widehat{f_X}(at)$
(resp. $\forall a \in \mathcal{M}_{\mathbb{R}}(m, n), \forall b \in \mathbb{R}^m, \forall t \in \mathbb{R}^m, \widehat{f_{aX+b}}(t) = e^{it^T b} \widehat{f_X}(a^T t)$).

Ces propriétés résultent immédiatement de la définition.

Théorème 48. La fonction caractéristique d'une variable (resp. d'un vecteur) aléatoire X détermine sa loi.

1. Cas discret : on développe $\widehat{f_X}(t)$ sous la forme $\sum_{k \in I} p(x_k) e^{itx_k}$.

2. Cas continu : on utilise le théorème d'inversion de FOURIER (voir [4], par exemple).

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{f_X}(t) e^{-itx} dt.$$

Démonstration identique dans le cas d'un vecteur aléatoire.

Théorème 49. Si le moment d'ordre k de la variable aléatoire X existe, sa fonction caractéristique admet une dérivée d'ordre k au voisinage de 0 et

$$\widehat{f_X}^{(k)}(0) = i^k \mathbf{E}(X^k).$$

1. Cas discret

On a :

$$\frac{d^k}{dt^k}(e^{itx_n}) = (ix_n)^k e^{itx_n} \quad \text{et} \quad |(ix_n)^k e^{itx_n}| = |x_n|^k.$$

Puisque $\mathbf{E}(X^k)$ existe, la série de terme général $(|x_n|^k p(x_n))$ est convergente et, donc, la série de terme général $((ix_n)^k e^{itx_n} p(x_n))$ converge uniformément sur \mathbb{R} et

$$\widehat{f_X}^{(k)}(t) = \sum_{n=0}^{\infty} (ix_n)^k e^{itx_n} p(x_n).$$

Donc :

$$\widehat{f_X}^{(k)}(0) = \sum_{n=0}^{\infty} (ix_n)^k p(x_n) = i^k \mathbf{E}(X^k).$$

2. Cas continu

On a :

$$\frac{d^k}{dt^k}(e^{itx} f_X(x)) = (ix)^k e^{itx} f_X(x) \quad \text{et} \quad |(ix)^k e^{itx} f_X(x)| = |x|^k f_X(x).$$

Puisque $\mathbf{E}(X^k)$ existe, l'intégrale $\int_{\mathbb{R}} |x|^k f_X(x) dx$ est convergente :

$$\widehat{f_X}^{(k)}(t) = \int_{\mathbb{R}} (ix)^k e^{itx} f_X(x) dx.$$

Donc :

$$\widehat{f_X}^{(k)}(0) = \int_{\mathbb{R}} (ix)^k f_X(x) dx = i^k \mathbf{E}(X^k).$$

Théorème 50. Si le vecteur aléatoire X possède une matrice de covariance on a :

$$\widehat{f_X}(t) = 1 + it^T \mathbf{E}(X) - \frac{1}{2} t^T \mathbf{E}(X X^T) t + o(\|t\|^2).$$

$$\begin{aligned} \widehat{f_X}(t) &= \mathbf{E}(\exp(it^T X)) = \mathbf{E}\left(1 + it^T X + \frac{1}{2}(it^T X)^2 + o(\|t\|^2)\right) \\ &= \mathbf{E}\left(1 + it^T X - \frac{1}{2}t^T X X^T t + o(\|t\|^2)\right) \quad \text{car } t^T X = X^T t \in \mathbb{R}, \\ &= 1 + it^T \mathbf{E}(X) - \frac{1}{2}t^T \mathbf{E}(X X^T) t + o(\|t\|^2) \quad \text{par linéarité.} \end{aligned}$$

Théorème 51. Si le couple aléatoire (X, Y) a pour fonction caractéristique \widehat{f} les fonctions caractéristiques des variables marginales X et Y sont données par :

$$\widehat{f_X}(t) = \widehat{f}(t, 0) \quad \text{et} \quad \widehat{f_Y}(u) = \widehat{f}(0, u).$$

$$\widehat{f_X}(t) = \mathbf{E}(e^{itX}) = \mathbf{E}(e^{i(tX+0Y)}) = \widehat{f}(t, 0).$$

Théorème 52.

$$X \text{ et } Y \text{ sont indépendantes} \Leftrightarrow \forall (t, u) \in \mathbb{R}^2, \widehat{f_{(X,Y)}}(t, u) = \widehat{f_X}(t) \widehat{f_Y}(u).$$

Plus généralement les variables aléatoires X_1, X_2, \dots, X_n sont mutuellement indépendantes si et seulement si la fonction caractéristique du vecteur aléatoire $X = (X_1 \ X_2 \ \dots \ X_n)^T$ est égale au produit des fonctions caractéristique des X_i .

On applique la définition et la formule d'inversion.

Théorème 53.

$$\forall (\lambda, \mu) \in \mathbb{R}^2, \forall t \in \mathbb{R}, \widehat{f_{\lambda X + \mu Y}}(t) = \widehat{f_{(X,Y)}}(\lambda t, \mu t).$$

$$\widehat{f_{\lambda X + \mu Y}}(t) = \mathbf{E}(e^{it(\lambda X + \mu Y)}) = \mathbf{E}(e^{i(\lambda t X + \mu t Y)}) = \widehat{f_{(X,Y)}}(\lambda t, \mu t).$$

Ce théorème s'étend immédiatement à une combinaison linéaire de n variables aléatoires.

Corollaire 1. Si on connaît la fonction caractéristique $\widehat{f_{\lambda X + \mu Y}}$ pour tous les couples (λ, μ) , alors on connaît la loi du couple (X, Y) .

D'après le théorème précédent on connaît $\widehat{f_{(X,Y)}}$.

Théorème 54. Si n variables aléatoires sont mutuellement indépendantes la fonction caractéristique de leur somme est égale au produit de leurs fonctions caractéristiques.

On applique les deux théorèmes précédents.

Théorème 55. Soit $(X_k)_{k \in \mathbb{N}^*}$ une suite de variables aléatoires mutuellement indépendantes ayant toutes la même loi de probabilité qu'une variable X . Soit N une variable aléatoire à valeurs dans \mathbb{N}^* , indépendante de chaque X_k .

La fonction caractéristique de la variable aléatoire $R = \sum_{k=1}^N X_k$ est donnée par :

$$\widehat{f_R} = g_N \circ \widehat{f_X}$$

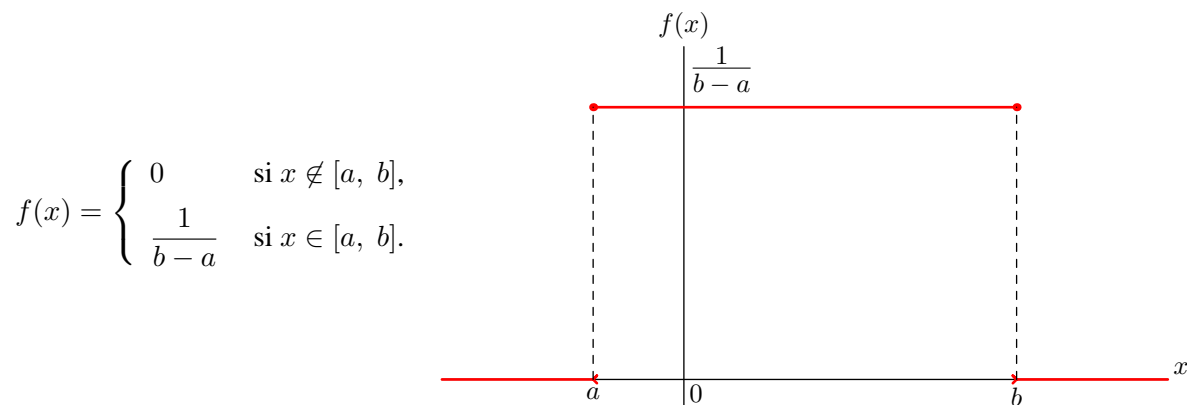
où g_N représente la fonction génératrice de N .

On admet ce théorème.

3.6 Lois de probabilités continues classiques

3.6.1 Loi uniforme

La densité de la loi uniforme $\mathcal{U}(a, b)$ est définie par :



Théorème 56. *Espérance mathématique et variance :*

$$\mathbf{E}(X) = \frac{a+b}{2}, \quad \mathbf{Var}(X) = \frac{(a-b)^2}{12}.$$

$$\mathbf{E}(X) = \int_{\mathbb{R}} x f(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{a+b}{2},$$

et

$$\mathbf{E}(X^2) = \int_{\mathbb{R}} x^2 f(x) dx = \int_a^b \frac{x^2}{b-a} dx = \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b = \frac{a^2 + ab + b^2}{3}.$$

D'où :

$$\mathbf{Var}(X) = \frac{a^2 + ab + b^2}{3} - \frac{a^2 + 2ab + b^2}{4} = \frac{(a-b)^2}{12}.$$

Théorème 57. *Fonction caractéristique :*

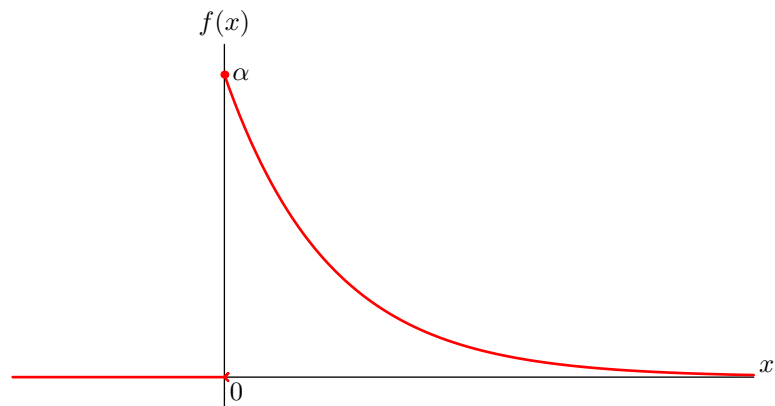
$$\widehat{f}(t) = \exp\left(it \frac{a+b}{2}\right) \frac{\sin \frac{b-a}{2} t}{\frac{b-a}{2} t}.$$

$$\widehat{f}(t) = \int_a^b \frac{e^{itx}}{b-a} dx = \frac{1}{b-a} \left[\frac{e^{itx}}{it} \right]_a^b = \frac{e^{itb} - e^{ita}}{it(b-a)}.$$

3.6.2 Loi exponentielle

La densité de la loi exponentielle $\mathcal{E}(\alpha)$ est définie par :

$$f(x) = \begin{cases} 0 & \text{si } x < 0, \\ \alpha e^{-\alpha x} & \text{si } x \geq 0. \end{cases}$$



Théorème 58. *Fonction caractéristique :*

$$\hat{f}(t) = \frac{\alpha}{\alpha - it}.$$

$$\hat{f}(t) = \int_0^{\infty} \alpha e^{(-\alpha + it)x} dx = \alpha \left[\frac{e^{(-\alpha + it)x}}{-\alpha + it} \right]_0^{\infty} = \frac{\alpha}{\alpha - it}.$$

Théorème 59. *Espérance mathématique et variance :*

$$\mathbf{E}(X) = \frac{1}{\alpha}, \quad \mathbf{Var}(X) = \frac{1}{\alpha^2}.$$

On applique le théorème 49.

$$\hat{f}'(t) = \frac{i\alpha}{(\alpha - it)^2}. \text{ D'où : } \hat{f}'(0) = \frac{i}{\alpha} = i\mathbf{E}(X) \text{ et donc } \mathbf{E}(X) = \frac{1}{\alpha}.$$

$$\hat{f}''(t) = \frac{-2\alpha}{(\alpha - it)^3}. \text{ D'où : } \hat{f}''(0) = \frac{-2}{\alpha^2} = -\mathbf{E}(X^2). \text{ Donc } \mathbf{E}(X^2) = \frac{2}{\alpha^2}$$

et

$$\mathbf{Var}(X) = \frac{2}{\alpha^2} - \left(\frac{1}{\alpha}\right)^2 = \frac{1}{\alpha^2}.$$

Théorème 60 (Propriété caractéristique). Soit X la variable aléatoire représentant la durée de vie d'un système et Y celle représentant la durée de survie du même système :

$$Y = X - \tau \text{ sachant que } X \geq \tau.$$

1. Si X suit la loi exponentielle $\mathcal{E}(\alpha)$ alors Y suit la même loi exponentielle.
2. Si X et Y suivent la même loi c'est une loi exponentielle.

1. On a :

$$F_Y(y) = \mathbb{P}(Y < y) = \mathbb{P}(X - \tau < y \mid X \geq \tau) = \mathbb{P}(X < y + \tau \mid X \geq \tau).$$

D'où :

- Si $y \leq 0$: $F_Y(y) = 0$.
- Si $y > 0$:

$$\begin{aligned} F_Y(y) &= \frac{\mathbb{P}(\tau \leq X < \tau + y)}{\mathbb{P}(X \geq \tau)} = \frac{F_X(y + \tau) - F_X(\tau)}{1 - F_X(\tau)} = \frac{(1 - e^{-\alpha(y+\tau)}) - (1 - e^{-\alpha\tau})}{1 - (1 - e^{-\alpha\tau})} \\ &= 1 - e^{-\alpha y}. \end{aligned}$$

2. On a vu :

$$F_Y(y) = \frac{\mathbb{P}(\tau \leq X < \tau + y)}{\mathbb{P}(X \geq \tau)}.$$

Donc :

- Si $y \leq 0$: $F_Y(y) = 0$.
- Si $y > 0$:

$$F_Y(y) = \frac{F_X(y + \tau) - F_X(\tau)}{1 - F_X(\tau)} = F_X(y) \quad \text{par hypothèse.}$$

La fonction G définie par $G(x) = 1 - F_X(x)$ vérifie donc :

$$\forall y \in \mathbb{R}_+, G(y + \tau) = G(y)G(\tau).$$

D'où $G(0) = 1$ et

$$\frac{G(y + \tau) - G(\tau)}{(y + \tau) - \tau} = \frac{G(\tau)(G(y) - G(0))}{y} \xrightarrow{y \rightarrow 0} G'(\tau) = G(\tau)G'(0).$$

On intègre immédiatement :

$$G(\tau) = e^{-\alpha\tau} \text{ où } \alpha = -G'(0).$$

D'où, en reportant :

$$F_X(x) = F_Y(x) = 1 - e^{-\alpha x} \text{ si } x \geq 0.$$

Théorème 61. Supposons que les variables aléatoires X_k sont mutuellement indépendantes et suivent la même loi exponentielle $\mathcal{E}(\alpha)$; supposons que la variable aléatoire N est indépendante de chaque X_k et suit la loi géométrique $\mathcal{G}(p)$. Alors la variable aléatoire

$$R = \sum_{k=1}^N X_k$$

suit la loi exponentielle $\mathcal{E}(\alpha p)$.

La fonction génératrice de la loi géométrique $\mathcal{G}(p)$ est $g_N(s) = \frac{ps}{1 - qs}$. D'après le théorème 55

$$\widehat{f_R}(t) = g_N(\widehat{f_{X_k}}(t)) = g_N\left(\frac{\alpha}{\alpha - it}\right) = \frac{p \frac{\alpha}{\alpha - it}}{1 - q \frac{\alpha}{\alpha - it}} = \frac{\alpha p}{\alpha p - it}$$

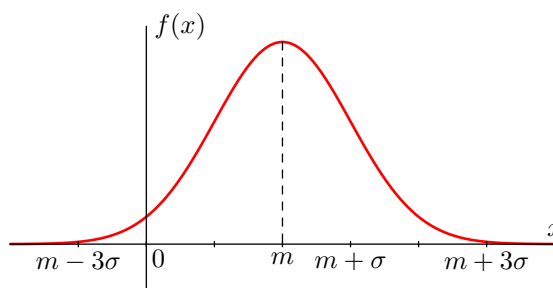
qui est la fonction caractéristique de la loi exponentielle $\mathcal{E}(\alpha p)$.

3.6.3 Loi normale (ou loi de LAPLACE-GAUSS)

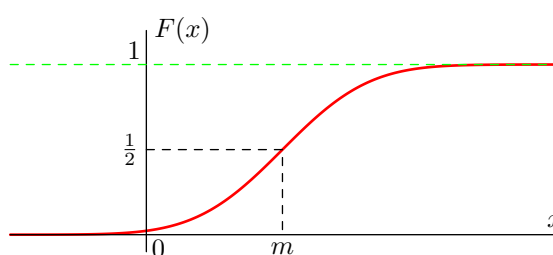
Loi normale à une dimension

La densité de la loi normale $\mathcal{N}(m, \sigma)$ est définie par :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$



La fonction de répartition de la loi normale ne peut s'exprimer à l'aide des fonctions élémentaires ; il faut utiliser une table spéciale (cf. page 150).



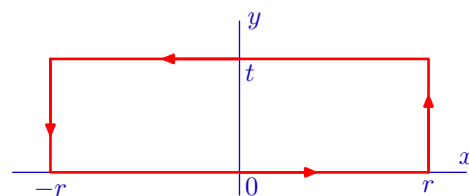
Théorème 62. La fonction caractéristique de la loi normale $\mathcal{N}(m, \sigma)$ est définie par :

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

On cherche d'abord la fonction caractéristique $\hat{\varphi}$ de la loi normale réduite $\mathcal{N}(0, 1)$:

$$\hat{\varphi}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) \exp(itx) dx.$$

On intègre la fonction $g(z) = \exp\left(-\frac{z^2}{2}\right)$, holomorphe sur \mathbb{C} , sur le lacet ci-contre. (Voir, par exemple [3].)



Lorsque r tend vers l'infini on obtient :

$$\begin{aligned} 0 &= \underbrace{\int_{-r}^r \exp\left(-\frac{x^2}{2}\right) dx}_{\rightarrow \sqrt{2\pi}} + i \underbrace{\int_0^t \exp\left(-\frac{1}{2}(r+iy)^2\right) dy}_{\rightarrow 0} \\ &\quad - \underbrace{\int_{-r}^r \exp\left(-\frac{1}{2}(-x+it)^2\right) dx}_{\rightarrow \exp\left(\frac{t^2}{2}\right)\sqrt{2\pi}\hat{\varphi}(t)} + i \underbrace{\int_t^0 \exp\left(-\frac{1}{2}(-r+iy)^2\right) dy}_{\rightarrow 0}. \end{aligned}$$

D'où : $\hat{\varphi}(t) = \exp\left(-\frac{t^2}{2}\right)$.

On obtient le résultat annoncé en appliquant la dernière propriété du théorème 47.

Théorème 63. *Espérance mathématique et variance :*

$$\mathbf{E}(X) = m, \quad \mathbf{Var}(X) = \sigma^2.$$

$$\widehat{f}'(t) = (im - \sigma^2 t) \exp\left(itm - \frac{\sigma^2 t^2}{2}\right). \text{ D'où : } \widehat{f}'(0) = im \text{ et } \mathbf{E}(X) = m.$$

$$\widehat{f}''(t) = ((im - \sigma^2 t)^2 - \sigma^2) \exp\left(itm - \frac{\sigma^2 t^2}{2}\right). \text{ D'où : } \widehat{f}''(0) = -m^2 - \sigma^2 \text{ et } \mathbf{E}(X^2) = m^2 + \sigma^2.$$

Donc :

$$\mathbf{Var}(X) = (m^2 + \sigma^2) - m^2 = \sigma^2.$$

Théorème 64. *Si X suit la loi normale $\mathcal{N}(m, \sigma)$ alors $Y = aX + b$, avec $a \neq 0$, suit la loi normale $\mathcal{N}(am + b, |a|\sigma)$.*

D'après la dernière propriété du théorème 47 :

$$\widehat{f}_Y(t) = e^{itb} \widehat{f}_X(at) = \exp(itb) \exp\left(i(at)m - \frac{\sigma^2}{2}(at)^2\right) = \exp\left(i(am + b)t - \frac{(a\sigma)^2}{2}t^2\right)$$

qui est la fonction caractéristique de la loi normale $\mathcal{N}(am + b, |a|\sigma)$.

Théorème 65 (stabilité). *La somme de deux variables normales indépendantes suit une loi normale. Plus précisément si X suit la loi normale $\mathcal{N}(m_1, \sigma_1)$, si Y suit la loi normale $\mathcal{N}(m_2, \sigma_2)$ et si X et Y sont indépendantes, alors $X + Y$ suit la loi normale $\mathcal{N}(m_1 + m_2, \sqrt{\sigma_1^2 + \sigma_2^2})$.*

D'après le théorème 54 :

$$\widehat{f}_{(X,Y)}(t) = \widehat{f}_X(t) \widehat{f}_Y(t) = \exp\left(itm_1 - \frac{\sigma_1^2}{2}t^2\right) \exp\left(itm_2 - \frac{\sigma_2^2}{2}t^2\right) = \exp\left(it(m_1 + m_2) - \frac{\sigma_1^2 + \sigma_2^2}{2}t^2\right).$$

Théorème 66. *Soient X_1, X_2, \dots, X_n n variables aléatoires mutuellement indépendantes suivant la même loi normale $\mathcal{N}(m, \sigma)$; leur moyenne $Y = \frac{1}{n} \sum_{i=1}^n X_i$ suit la loi normale $\mathcal{N}(m, \frac{\sigma}{\sqrt{n}})$.*

On montre comme pour le théorème précédent que la somme Z suit la loi normale $\mathcal{N}(nm, \sigma\sqrt{n})$. Puisque $Y = Z/n$ on applique la dernière propriété du théorème 47 :

$$\widehat{f}_Y(t) = \widehat{f}_Z\left(\frac{t}{n}\right) = \exp\left(itm - \frac{\sigma^2}{2n}t^2\right).$$

Vecteur gaussien

Définition 29. Un vecteur aléatoire $X = (X_1, X_2, \dots, X_n)$ est dit gaussien (non dégénéré) si sa densité est de la forme :

$$f(x) = \alpha \exp\left(-\frac{1}{2}(x - m)^T q (x - m)\right)$$

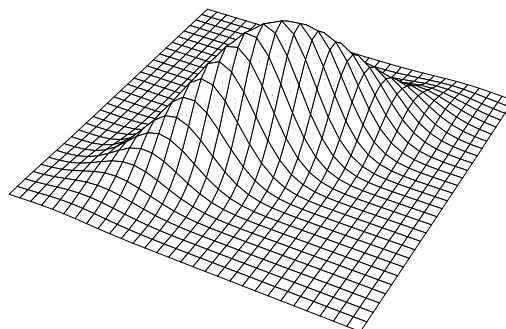
où $x \in \mathbb{R}^n$ et $m \in \mathbb{R}^n$ sont représentés par des matrices colonnes et $q \in \mathcal{M}_n(\mathbb{R})$ est symétrique, définie, positive.

Cas particulier important :

Définition 30. Un couple aléatoire (X, Y) est dit gaussien (non dégénéré) si sa densité est de la forme :

$$f(x, y) = \alpha \exp\left(-\frac{1}{2}Q(x, y)\right)$$

où $Q(x, y) = A(x - a)^2 + 2B(x - a)(y - b) + C(y - b)^2$ est une forme quadratique définie positive en $(x - a)$ et $(y - b)$: $A > 0$, $C > 0$ et $B^2 - AC < 0$.



Théorème 67. Soit X un vecteur gaussien :

$$\alpha = \sqrt{\frac{\det q}{(2\pi)^n}}, \quad \mathbf{E}(X) = m \quad \text{et} \quad \mathbf{Var}(X) = q^{-1}.$$

On notera $\mathcal{N}(m, c)$ la loi de X où $c = q^{-1}$ est la matrice de covariance et dont la densité est donc :

$$f(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det c}} \exp\left(-\frac{1}{2}(x - m)^T c^{-1} (x - m)\right).$$

Remarque : la notation utilisée pour la loi normale n'est pas exactement la même dans le cas d'une variable aléatoire ou dans celui d'un vecteur (écart-type remplacé par matrice de covariance).

D'après le théorème 45 la densité du vecteur aléatoire $Y = X - m$ est :

$$f_Y(y) = f_X(y + m) = \alpha \exp\left(-\frac{1}{2}y^T q y\right).$$

Par raison de symétrie on a $\mathbf{E}(Y) = 0$ et donc $\mathbf{E}(X) = \mathbf{E}(Y) + m = m$.

Puisque q est symétrique, définie et positive, il existe une matrice orthogonale u telle que

$$d = u^{-1} q u = u^T q u$$

est diagonale avec des éléments diagonaux (valeurs propres) λ_j strictement positifs.

Soit $Z = u^T X$; d'après le théorème 45 la densité de $Z = (Z_1, Z_2, \dots, Z_n)$ est :

$$\begin{aligned} f_Z(z) &= \underbrace{\frac{1}{|\det u|}}_{=1} f_X(uz) = \alpha \exp\left(-\frac{1}{2}(uz)^T q(uz)\right) = \alpha \exp\left(-\frac{1}{2}z^T \underbrace{(u^T q u)}_{=d} z\right) \\ &= \alpha \exp\left(-\frac{1}{2} \sum_{j=1}^n \lambda_j z_j^2\right) = \alpha \prod_{j=1}^n \exp\left(-\frac{1}{2} \lambda_j z_j^2\right) = \prod_{j=1}^n f_{Z_j}(z_j). \end{aligned}$$

Les variables aléatoires Z_j sont indépendantes et suivent les lois normales $\mathcal{N}(0, \sigma_j)$ avec $\lambda_j = \frac{1}{\sigma_j^2}$:

$$f_Z(z) = \prod_{j=1}^n \frac{1}{\sigma_j \sqrt{2\pi}} \exp\left(-\frac{z_j^2}{2\sigma_j^2}\right) = \frac{1}{(2\pi)^{n/2} \prod_{j=1}^n \sigma_j} \exp\left(-\frac{1}{2} \sum_{j=1}^n \frac{z_j^2}{\sigma_j^2}\right).$$

Donc

$$\frac{1}{\alpha} = (2\pi)^{n/2} \prod_{j=1}^n \sigma_j \quad \text{et} \quad \alpha^2 = \frac{1}{(2\pi)^n} \prod_{j=1}^n \lambda_j = \frac{\det q}{(2\pi)^n}.$$

Enfin :

$$\mathbf{Var}(X) = \mathbf{E}(XX^T) = \mathbf{E}((uZ)(uZ)^T) = u\mathbf{E}(ZZ^T)u^T = u d^{-1} u^T = q^{-1}.$$

Corollaire 2. Si le couple (X, Y) est gaussien les variables marginales X et Y sont gaussiennes. Dans ce cas si X suit la loi $\mathcal{N}(m_1, \sigma_1)$ et Y la loi $\mathcal{N}(m_2, \sigma_2)$ et si leur coefficient de corrélation est égal à ρ la densité du couple (X, Y) est :

$$f(x, y) = \frac{\exp\left[-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-m_1}{\sigma_1}\right)^2 - 2\rho\frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \left(\frac{y-m_2}{\sigma_2}\right)^2\right)\right]}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}.$$

On a :

$$q^{-1} = \mathbf{Var}((X, Y)) = \begin{pmatrix} \mathbf{Var}(X) & \mathbf{Cov}(X, Y) \\ \mathbf{Cov}(Y, X) & \mathbf{Var}(Y) \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

D'où :

$$\det q^{-1} = \sigma_1^2 \sigma_2^2 (1 - \rho^2), \quad \text{et} \quad q = \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}.$$

Enfin :

$$\alpha = \sqrt{\frac{1}{(2\pi)^2 \sigma_1^2 \sigma_2^2 (1 - \rho^2)}} = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

et

$$\begin{aligned} Q(x, y) &= \begin{pmatrix} x - m_1 & y - m_2 \end{pmatrix} q \begin{pmatrix} x - m_1 \\ y - m_2 \end{pmatrix} \\ &= \frac{1}{1 - \rho^2} \left(\left(\frac{x - m_1}{\sigma_1} \right)^2 - 2\rho \frac{(x - m_1)(y - m_2)}{\sigma_1\sigma_2} + \left(\frac{y - m_2}{\sigma_2} \right)^2 \right). \end{aligned}$$

Théorème 68. Si le couple (X, Y) est gaussien et si $\rho(X, Y) = 0$ les deux variables sont indépendantes.

En reportant $\rho = 0$ dans l'expression de la densité du couple on voit que $f(x, y) = f_X(x)f_Y(y)$.

Attention : il se peut que X et Y soient gaussiennes sans que le couple (X, Y) soit gaussien, auquel cas on ne peut conclure à l'indépendance lorsque $\rho(X, Y) = 0$.

Exemple

Considérons le couple aléatoire (non gaussien...) de densité

$$f(x, y) = \frac{1}{4\pi\sqrt{1-r^2}} \left(\exp\left(-\frac{x^2 - 2rxy + y^2}{2(1-r^2)}\right) + \exp\left(-\frac{x^2 + 2rxy + y^2}{2(1-r^2)}\right) \right) \text{ où } -1 < r < 1.$$

Les variables marginales X et Y suivent la loi normale réduite. En effet :

$$\begin{aligned} f_X(x) &= \frac{1}{4\pi\sqrt{1-r^2}} \left\{ \int_{-\infty}^{\infty} \exp\left(-\frac{(y-rx)^2}{2(1-r^2)} - \frac{x^2}{2}\right) dy + \int_{-\infty}^{\infty} \exp\left(-\frac{(y+rx)^2}{2(1-r^2)} - \frac{x^2}{2}\right) dy \right\} \\ &= \frac{\sqrt{2}}{\pi} \exp\left(-\frac{x^2}{2}\right) \int_0^{\infty} \exp(-u^2) du \text{ en posant } u = \frac{y \pm rx}{\sqrt{2(1-r^2)}}, \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \end{aligned}$$

Elles ne sont pas indépendantes car $f(x, y) \neq f_X(x)f_Y(y)$ bien que $\text{Cov}(X, Y) = \mathbf{E}(XY) = 0$ à cause de la parité de $f(x, y)$ par rapport à x et par rapport à y .

Fonction caractéristique

Théorème 69. Le vecteur $\mathbf{X} = (X_1, X_2, \dots, X_n)$ est gaussien de loi $\mathcal{N}(m, c)$, où c est la matrice de covariance, si et seulement si la fonction caractéristique de \mathbf{X} est définie par :

$$\forall t \in \mathbb{R}^n, \widehat{f_{\mathbf{X}}}(t) = \exp\left(it^T m - \frac{1}{2}t^T c t\right).$$

1. Supposons $m = 0$. Pour tout λ de \mathbb{R}^n on a, d'après le théorème 22 :

$$\mathbf{E}(\lambda^T \mathbf{X}) = 0 \quad \text{et} \quad \sigma^2 = \mathbf{Var}(\lambda^T \mathbf{X}) = \lambda^T c \lambda.$$

La fonction caractéristique de $\lambda^T \mathbf{X}$ est donc, d'après le théorème 62 :

$$\widehat{f_{\lambda^T \mathbf{X}}}(u) = \exp\left(-\frac{1}{2}\sigma^2 u^2\right) = \exp\left(-\frac{1}{2}\lambda^T c \lambda u^2\right)$$

En appliquant le corollaire 1 :

$$\widehat{f_{\mathbf{X}}}(u\lambda) = \widehat{f_{\lambda^T \mathbf{X}}}(u) = \exp\left(-\frac{1}{2}(u\lambda)^T c (u\lambda)\right).$$

En posant $t = u\lambda$ on obtient :

$$\widehat{f_{\mathbf{X}}}(t) = \exp\left(-\frac{1}{2}t^T c t\right).$$

2. Lorsque $m \neq 0$ on remarque que :

$$\widehat{f_{X+m}}(t) = \mathbf{E}\left(\exp(it^T(X+m))\right) = \exp(it^T m) \widehat{f_X}(t).$$

La réciproque résulte du théorème 48.

Corollaire 3. Soit $X = (X_1, X_2, \dots, X_n)$ un vecteur gaussien de loi $\mathcal{N}(m, c)$: $Y = aX + b$, où $a \in \mathcal{M}_{\mathbb{R}}(m, n)$ et $b \in \mathbb{R}^m$, est un vecteur gaussien de loi $\mathcal{N}(am + b, a c a^T)$.

D'après la dernière propriété du théorème 47 on a :

$$\begin{aligned} \widehat{f_Y}(t) &= \exp(it^T b) \widehat{f_X}(a^T t) = \exp(it^T b + i(a^T t)^T m - \frac{1}{2}(a^T t)^T c (a^T t)) \\ &= \exp(it^T (am + b) - \frac{1}{2}t^T (aca^T)t). \end{aligned}$$

On retrouve (voir théorème 22) :

$$\mathbf{E}(Y) = a\mathbf{E}(X) + b \quad \text{et} \quad \mathbf{Var}(Y) = a\mathbf{Var}(X)a^T.$$

Calculs numériques

On dispose de tables numériques donnant les valeurs de la densité φ et de la fonction de répartition Φ de la loi normale réduite $\mathcal{N}(0, 1)$ (voir page 150) ; les valeurs de la fonction de répartition F_X et de la densité f_X d'une variable X suivant la loi normale $\mathcal{N}(m, \sigma)$ s'en déduisent :

$$F_X(x) = \Phi\left(\frac{x-m}{\sigma}\right), \quad f_X(x) = \frac{1}{\sigma}\varphi\left(\frac{x-m}{\sigma}\right).$$

D'après le théorème 64 si X suit la loi normale $\mathcal{N}(m, \sigma)$ alors $X^* = \frac{X-m}{\sigma}$, variable réduite associée, suit la loi normale réduite.

$$F_X(x) = \mathbb{P}(X < x) = \mathbb{P}(X^* < \frac{x-m}{\sigma}) = \Phi\left(\frac{x-m}{\sigma}\right).$$

En dérivant :

$$f_X(x) = \frac{1}{\sigma}\varphi\left(\frac{x-m}{\sigma}\right)$$

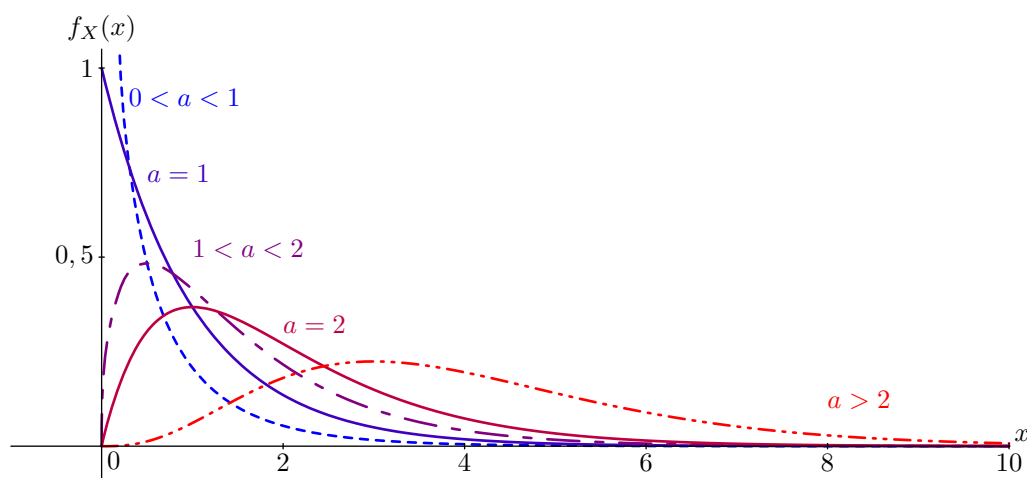
3.6.4 Loi gamma

Définition 31. La variable aléatoire X suit la loi $\gamma(a)$ si sa densité est définie par :

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(a)} x^{a-1} e^{-x} & \text{si } x > 0, \\ 0 & \text{sinon,} \end{cases}$$

où a est un paramètre réel strictement positif.

La fonction eulérienne de deuxième espèce Γ est définie à la page 145.



Théorème 70. La fonction caractéristique de la loi $\gamma(a)$ est :

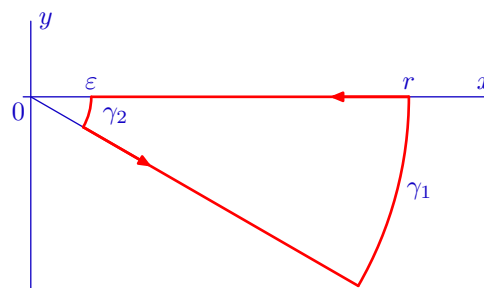
$$\widehat{f_X}(t) = \frac{1}{(1 - it)^a}.$$

La fonction caractéristique est définie par

$$\widehat{f_X}(t) = \int_0^\infty \frac{1}{\Gamma(a)} x^{a-1} e^{-x(1-it)} dx.$$

On applique le théorème des résidus en intégrant sur le lacet γ ci-contre la fonction

$$g(z) = \frac{1}{\Gamma(a)} z^{a-1} e^{-z}.$$



On obtient lorsque r tend vers l'infini et ϵ tend vers zéro :

$$\underbrace{\int_\epsilon^r \frac{(1-it)^a}{\Gamma(a)} x^{a-1} e^{-x(1-it)} dx}_{\rightarrow (1-it)^a \widehat{f_X}(t)} + \underbrace{\int_{\gamma_1} f(z) dz}_{\rightarrow 0} + \underbrace{\int_r^\epsilon \frac{1}{\Gamma(a)} x^{a-1} e^{-x} dx}_{\rightarrow -1} + \underbrace{\int_{\gamma_2} f(z) dz}_{\rightarrow 0} = 0.$$

D'où :

$$\widehat{f_X}(t) = \frac{1}{(1 - it)^a}.$$

Théorème 71. Espérance mathématique et variance :

$$\mathbf{E}(X) = \mathbf{Var}(X) = a.$$

Ce résultat est immédiat

– soit à partir de la définition ;

– soit en utilisant le théorème 49.

Théorème 72 (stabilité). Si les variables aléatoires X et Y suivent les lois $\gamma(a)$ et $\gamma(b)$ et sont indépendantes leur somme $Z = X + Y$ suit la loi $\gamma(a + b)$.

D'après le théorème 54

$$\widehat{f_Z}(t) = \widehat{f_X}(t)\widehat{f_Y}(t) = \frac{1}{(1 - it)^a} \frac{1}{(1 - it)^b} = \frac{1}{(1 - it)^{a+b}}.$$

On peut généraliser la définition.

Définition 32. Soient λ (facteur d'échelle) et a (facteur de forme) deux paramètres réels strictement positifs. La variable aléatoire T suit la loi $\gamma(a, \lambda)$ si sa densité est définie par :

$$f_T(t) = \begin{cases} \frac{\lambda}{\Gamma(a)} (\lambda t)^{a-1} e^{-\lambda t} & \text{si } t > 0, \\ 0 & \text{sinon.} \end{cases}$$

Remarques

1. La loi $\gamma(a, 1)$ est la loi $\gamma(a)$.
2. La loi $\gamma(1, \lambda)$ est la loi exponentielle $\mathcal{E}(\lambda)$.

Théorème 73. Soit $\alpha > 0$.

$$T \text{ suit la loi } \gamma(a, \lambda) \iff U = \alpha T \text{ suit la loi } \gamma(a, \frac{\lambda}{\alpha}).$$

En particulier : T suit la loi $\gamma(a, \lambda) \iff X = \lambda T$ suit la loi $\gamma(a)$.

1. Supposons que T suit la loi $\gamma(a, \lambda)$. La fonction de répartition de $U = \alpha T$ est :

$$F_U(u) = \mathbb{P}(U < u) = \mathbb{P}(T < \frac{u}{\alpha}) = F_T(\frac{u}{\alpha}).$$

On obtient la densité de probabilité en dérivant :

$$f_U(u) = \frac{1}{\alpha} f_T(\frac{u}{\alpha}) = \begin{cases} \frac{\lambda/\alpha}{\Gamma(a)} (\frac{\lambda}{\alpha} u)^{a-1} e^{-\frac{\lambda}{\alpha} u} & \text{si } u > 0, \\ 0 & \text{sinon.} \end{cases}$$

Donc U suit la loi $\gamma(a, \frac{\lambda}{\alpha})$.

2. Supposons que U suit la loi $\gamma(a, \frac{\lambda}{\alpha})$. On a, pour tout $\varphi \in \mathcal{D}_1$:

$$\begin{aligned} \mathbf{E}(\varphi(T)) &= \mathbf{E}\left(\varphi\left(\frac{U}{\alpha}\right)\right) = \int_{-\infty}^{\infty} \varphi\left(\frac{u}{\alpha}\right) f_U(u) du \\ &= \int_{-\infty}^{\infty} \varphi(t) f_U(\alpha t) \alpha dt, \quad \text{en posant } u = \alpha t. \end{aligned}$$

Donc, d'après le théorème 44 :

$$f_T(t) = \alpha f_U(\alpha t) = \frac{\lambda}{\Gamma(a)} (\lambda t)^{a-1} e^{-\lambda t} \text{ si } t > 0.$$

Remarque : nous avons utilisé deux méthodes différentes à titre d'exemple et de comparaison.

Théorème 74. Si T suit la loi $\gamma(a, \lambda)$ on a :

$$\mathbf{E}(T) = \frac{a}{\lambda} \quad \text{et} \quad \mathbf{Var}(T) = \frac{a}{\lambda^2}.$$

La variable aléatoire $X = \lambda T$ suit la loi $\gamma(a)$; donc :

$$\mathbf{E}(T) = \mathbf{E}\left(\frac{X}{\lambda}\right) = \frac{1}{\lambda} \mathbf{E}(X) = \frac{a}{\lambda} \quad \text{et} \quad \mathbf{Var}(T) = \mathbf{Var}\left(\frac{X}{\lambda}\right) = \frac{1}{\lambda^2} \mathbf{E}(X) = \frac{a}{\lambda^2}.$$

Théorème 75 (stabilité). Si les variables aléatoires T et U suivent les lois $\gamma(a, \lambda)$ et $\gamma(b, \lambda)$ et sont indépendantes leur somme suit la loi $\gamma(a + b, \lambda)$.

On pose $X = \lambda T$ et $Y = \lambda U$ et on applique le théorème 72.

Théorème 76. La fonction caractéristique de la loi $\gamma(a, \lambda)$ est :

$$\widehat{f}(t) = \frac{1}{(1 - it/\lambda)^a}.$$

Si T suit la loi $\gamma(a, \lambda)$ alors $X = \lambda T$ suit la loi $\gamma(a)$; on applique le théorème 47 :

$$\widehat{f}_T(t) = \widehat{f_{X/\lambda}}(t) = \widehat{f_X}\left(\frac{t}{\lambda}\right) = \frac{1}{(1 - it/\lambda)^a}.$$

Théorème 77. Si X suit la loi normale réduite alors $V = \frac{X^2}{2}$ suit la loi $\gamma(\frac{1}{2})$.

On applique le théorème 44. Pour toute fonction φ appartenant à \mathcal{D}_1 :

$$\begin{aligned} \mathbf{E}(\varphi(V)) &= \mathbf{E}\left(\varphi\left(\frac{X^2}{2}\right)\right) = \int_{-\infty}^{\infty} \varphi\left(\frac{x^2}{2}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\ &= \sqrt{\frac{2}{\pi}} \int_0^{\infty} \varphi\left(\frac{x^2}{2}\right) \exp\left(-\frac{x^2}{2}\right) dx = \sqrt{\frac{2}{\pi}} \int_0^{\infty} \varphi(v) e^{-v} \frac{dv}{\sqrt{2v}}, \quad \text{où } x = \sqrt{2v}. \end{aligned}$$

D'où :

$$f_V(v) = \begin{cases} \frac{1}{\sqrt{\pi}} v^{-1/2} e^{-v} = \frac{1}{\Gamma(1/2)} v^{-1/2} e^{-v} & \text{si } v > 0, \\ 0 & \text{sinon.} \end{cases}$$

Donc V suit la loi $\gamma(\frac{1}{2})$.

Théorème 78. Si X_1, X_2, \dots, X_n suivent la loi normale réduite et sont indépendantes alors

$$W = \frac{1}{2} \sum_{k=1}^n X_k^2$$

suit la loi $\gamma(\frac{n}{2})$.

Les variables aléatoires $V_k = \frac{X_k^2}{2}$ sont indépendantes et suivent toutes la loi $\gamma(\frac{1}{2})$; d'après le théorème 72 $W = \sum_{k=1}^n V_k$ suit la loi $\gamma(\frac{n}{2})$.

3.6.5 Loi bêta

Théorème 79. Soient X et Y deux variables aléatoires indépendantes suivant les lois $\gamma(a)$ et $\gamma(b)$ respectivement. La variable aléatoire $Z = \frac{X}{Y}$ suit la loi $\beta(a, b)$ (loi bêta de deuxième espèce) dont la densité est définie par :

$$f_Z(z) = \begin{cases} \frac{1}{B(a, b)} \frac{z^{a-1}}{(z+1)^{a+b}} & \text{si } z > 0, \\ 0 & \text{sinon.} \end{cases}$$

La fonction eulérienne de première espèce B est définie à la page 145.

La fonction de répartition de la variable aléatoire Z est nulle sur \mathbb{R}^- ; sur \mathbb{R}^+ on a :

$$F_Z(z) = \mathbb{P}\left(\frac{X}{Y} < z\right) = \iint_{D_z} f_{(X, Y)}(x, y) dx dy$$

où $D_z = \{(x, y) \in (\mathbb{R}^+)^2 \mid x < zy\}$ et, les variables aléatoires X et Y étant indépendantes,

$$f_{(X, Y)}(x, y) = f_X(x) f_Y(y) = \left(\frac{x^{a-1}e^{-x}}{\Gamma(a)}\right) \left(\frac{y^{b-1}e^{-y}}{\Gamma(b)}\right) \quad \text{si } (x, y) \in (\mathbb{R}^+)^2.$$

On effectue le changement de variables défini par :

$$\begin{cases} x = u, \\ y = \frac{u}{v}. \end{cases}$$

On obtient

$$F_Z(z) = \int_0^z \left(\int_0^\infty \frac{u^{a-1}e^{-u}}{\Gamma(a)} \frac{1}{\Gamma(b)} \left(\frac{u}{v}\right)^{b-1} e^{-\frac{u}{v}} \frac{u}{v^2} du \right) dv$$

puis la densité de probabilité de Z en dérivant par rapport à z

$$f_Z(z) = \frac{1}{\Gamma(a)\Gamma(b)z^{b+1}} \int_0^\infty u^{a+b-1} \exp\left(-\left(1 + \frac{1}{z}\right)u\right) du.$$

On pose : $t = \left(1 + \frac{1}{z}\right)u$.

$$\begin{aligned} f_Z(z) &= \frac{z^{a-1}}{\Gamma(a)\Gamma(b)(z+1)^{a+b}} \int_0^\infty t^{a+b-1} e^{-t} dt \\ &= \frac{z^{a-1}\Gamma(a+b)}{\Gamma(a)\Gamma(b)(z+1)^{a+b}} = \frac{1}{B(a, b)} \frac{z^{a-1}}{(z+1)^{a+b}}. \end{aligned}$$

Théorème 80. Espérance mathématique et variance.

$$\mathbb{E}(Z) = \frac{a}{b-1} \quad (b > 1) \quad \text{et} \quad \text{Var}(Z) = \frac{a(a+b-1)}{(b-1)^2(b-2)} \quad (b > 2).$$

$$\begin{aligned}
\mathbf{E}(Z^r) &= \frac{1}{B(a, b)} \int_0^\infty \frac{z^{a+r-1}}{(z+1)^{a+b}} dz. \text{ On pose : } u = \frac{z}{z+1}. \\
&= \frac{1}{B(a, b)} \int_0^1 \frac{\left(\frac{u}{1-u}\right)^{a+r-1}}{\left(\frac{1}{1-u}\right)^{a+b}} \frac{du}{(1-u)^2} = \frac{1}{B(a, b)} \underbrace{\int_0^1 u^{a+r-1} (1-u)^{b-r-1} du}_{=B(a+r, b-r)} \\
&= \frac{B(a+r, b-r)}{B(a, b)} \quad \text{si } r < b.
\end{aligned}$$

D'où

$$\begin{aligned}
\mathbf{E}(Z) &= \frac{B(a+1, b-1)}{B(a, b)} = \frac{\Gamma(a+1)\Gamma(b-1)}{\Gamma(a)\Gamma(b)} = \frac{a}{b-1} \quad \text{si } b > 1, \\
\mathbf{E}(Z^2) &= \frac{B(a+2, b-2)}{B(a, b)} = \frac{\Gamma(a+2)\Gamma(b-2)}{\Gamma(a)\Gamma(b)} = \frac{a(a+1)}{(b-1)(b-2)} \quad \text{si } b > 2, \\
\mathbf{Var}(Z) &= \frac{a(a+1)}{(b-1)(b-2)} - \left(\frac{a}{b-1}\right)^2 = \frac{a(a+b-1)}{(b-1)^2(b-2)} \quad \text{si } b > 2.
\end{aligned}$$

3.6.6 Loi du khi-deux

Définition 33. Soient X_1, X_2, \dots, X_n des variables normales réduites indépendantes. La loi du χ^2 à n degrés de liberté est la loi de la variable aléatoire

$$\chi_n^2 = \sum_{k=1}^n X_k^2.$$

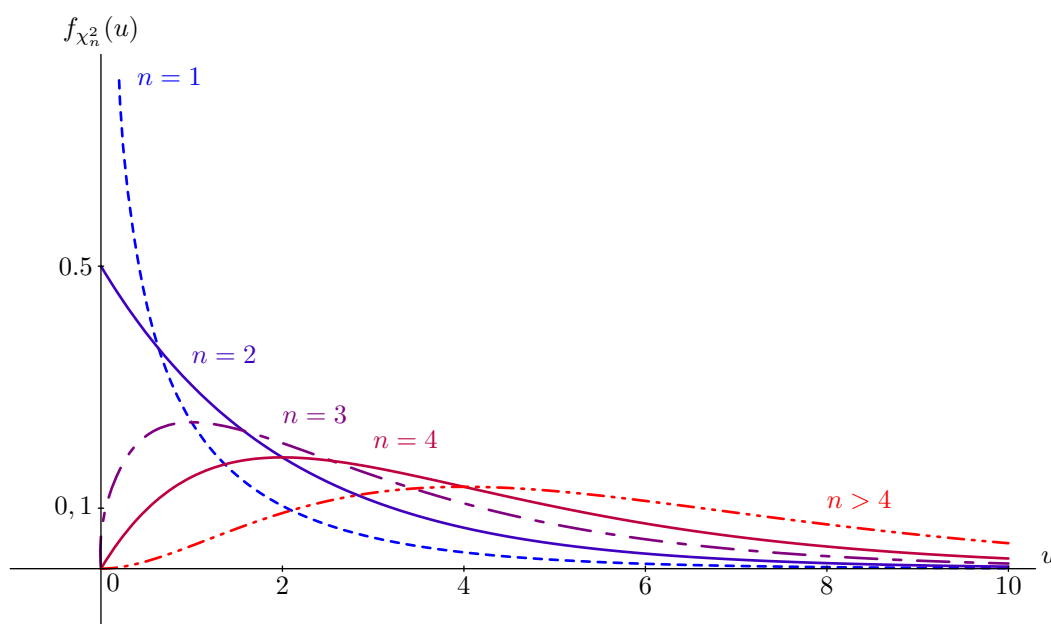
On sait (cf. théorème 78) que $W = \frac{1}{2} \sum_{k=1}^n X_k^2$ suit la loi $\gamma(\frac{n}{2})$. D'après le théorème 73 la variable aléatoire $\chi_n^2 = 2W$ suit donc la loi $\gamma(\frac{n}{2}, \frac{1}{2})$.

La densité de la loi du χ^2 à n degrés de liberté est donc définie par

$$f_{\chi_n^2}(u) = \begin{cases} \frac{1}{2\Gamma(\frac{n}{2})} \left(\frac{u}{2}\right)^{\frac{n}{2}-1} e^{-\frac{u}{2}} & \text{si } u > 0, \\ 0 & \text{sinon,} \end{cases}$$

et sa fonction caractéristique par

$$\widehat{f_{\chi_n^2}}(t) = \frac{1}{(1-2it)^{n/2}}.$$



D'après le théorème 74, on obtient

$$\mathbf{E}(\chi_n^2) = n \quad \text{et} \quad \mathbf{Var}(\chi_n^2) = 2n.$$

Théorème 81 (Stabilité). Si X et Y sont indépendantes et suivent respectivement les lois du khi-deux à n_1 et n_2 degrés de liberté alors $(X + Y)$ suit la loi du khi-deux à $(n_1 + n_2)$ degrés de liberté.

La loi du khi-deux étant une loi gamma il suffit d'appliquer le théorème 75. On peut également remarquer que $(X + Y)$ est la somme des carrés de $(n_1 + n_2)$ variables normales réduites indépendantes.

Théorème 82. Soit X un vecteur gaussien de dimension n de loi $\mathcal{N}(m, c)$; la variable aléatoire $Y = (X - m)^T c^{-1} (X - m)$ suit la loi du khi-deux à n degrés de liberté.

On peut supposer, sans restreindre la généralité, que $m = 0$.

La matrice c^{-1} est symétrique, définie positive : il existe une matrice orthogonale u telle que $d = u^T c^{-1} u$ soit diagonale, les éléments diagonaux étant strictement positifs. On a :

$$Y = X^T c^{-1} X = X^T u d u^T X = (u^T X)^T d (u^T X).$$

D'après le corollaire 3 $Z = u^T X$ est un vecteur gaussien d'espérance nulle et dont la matrice de covariance est $u^T c u = d^{-1}$. Donc

$$Y = \sum_{i=1}^n \frac{Z_i^2}{\mathbf{Var}(Z_i)}$$

suit la loi du khi-deux à n degrés de liberté.

Approximations pour n grand

Théorème 83. La variable aléatoire $\frac{\chi_n^2 - n}{\sqrt{2n}}$ converge en loi vers une variable normale réduite.

La convergence en loi est définie au paragraphe 4.1.

Il suffit d'appliquer le théorème 95 avec $X_n = Y_n^2$ où Y_n suit la loi normale réduite ; on trouve immédiatement $m = 1$ et $\sigma = \sqrt{2}$. On a $\chi_n^2 = S_n$; d'où le résultat.

Théorème 84 (FISHER). *La variable aléatoire $\sqrt{2\chi_n^2} - \sqrt{2n-1}$ converge en loi vers une variable normale réduite.*

Déterminons la densité de probabilité de $Z = \sqrt{2\chi_n^2} - \sqrt{2n-1}$. Pour toute fonction φ appartenant à \mathcal{D}_1 :

$$\begin{aligned} \mathbf{E}(\varphi(Z)) &= \mathbf{E}(\varphi(\sqrt{2\chi_n^2} - \sqrt{2n-1})) = \int_0^\infty \varphi(\sqrt{2u} - \sqrt{2n-1}) \frac{1}{2\Gamma(\frac{n}{2})} \left(\frac{u}{2}\right)^{\frac{n}{2}-1} e^{-\frac{u}{2}} du \\ &= \int_{-\sqrt{2n-1}}^\infty \varphi(z) \frac{1}{\Gamma(\frac{n}{2})} \left(\frac{1}{2}(z + \sqrt{2n-1})\right)^{n-1} \exp\left(-\frac{1}{4}(z + \sqrt{2n-1})^2\right) dz \end{aligned}$$

où l'on a posé : $z = \sqrt{2u} - \sqrt{2n-1}$. Donc :

$$f_Z(z) = \begin{cases} 0 & \text{si } z < -\sqrt{2n-1}, \\ \frac{1}{\Gamma(\frac{n}{2})} \left(\frac{1}{2}(z + \sqrt{2n-1})\right)^{n-1} \exp\left(-\frac{1}{4}(z + \sqrt{2n-1})^2\right) & \text{si } z > -\sqrt{2n-1}. \end{cases}$$

Soit $z \in \mathbb{R}$ fixé ; pour n assez grand on a $z > -\sqrt{2n-1}$ et donc :

$$\ln(f_Z(z)) = \underbrace{-\ln\left(\Gamma\left(\frac{n}{2}\right)\right)}_A + \underbrace{(n-1)(\ln(z + \sqrt{2n-1}) - \ln 2)}_B - \underbrace{\frac{1}{4}(z + \sqrt{2n-1})^2}_C$$

Or :

$$\begin{aligned} A &= -\frac{n}{2} \ln n + \frac{n}{2} \ln 2 + \frac{n}{2} + \frac{1}{2} \ln n - \frac{1}{2} \ln 2 - \frac{1}{2} \ln(2\pi) + o(1) \quad (\text{STIRLING}), \\ B &= (n-1) \left(-\frac{1}{2} \ln 2 + \frac{1}{2} \ln n + \frac{z}{\sqrt{2n}} - \frac{1}{4n} - \frac{z^2}{4n} + o\left(\frac{1}{n}\right) \right), \\ C &= -\frac{z^2}{4} + \frac{1}{4} - \frac{n}{2} - \frac{z}{2} \sqrt{2n-1}. \end{aligned}$$

En regroupant et en simplifiant :

$$\ln(f_Z(z)) = -\frac{1}{2} \ln(2\pi) - \frac{z^2}{2} + z \sqrt{\frac{n}{2}} \left(1 - \sqrt{1 - \frac{1}{2n}}\right) + o(1) \xrightarrow{n \rightarrow \infty} -\frac{1}{2} \ln(2\pi) - \frac{z^2}{2}.$$

Donc :

$$f_Z(z) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}.$$

Théorème 85 (WILSON-HILFERTY). *La variable aléatoire*

$$\sqrt{\frac{9n}{2}} \left(\sqrt[3]{\frac{\chi_n^2}{n}} - \left(1 - \frac{2}{9n}\right) \right)$$

converge en loi vers une variable normale réduite.

On pose :

$$Z = \sqrt{\frac{9n}{2}} \left(\sqrt[3]{\frac{\chi_n^2}{n}} - \left(1 - \frac{2}{9n}\right) \right).$$

Par la même méthode que ci-dessus on obtient pour n grand, lorsque z est fixé :

$$f_Z(z) = \frac{1}{\Gamma(\frac{n}{2})} \left(\frac{n}{2}\right)^{\frac{n-1}{2}} \left(z\sqrt{\frac{2}{9n}} + 1 - \frac{2}{9n}\right)^{\frac{3n}{2}-1} \exp\left(-\frac{n}{2}\left(z\sqrt{\frac{2}{9n}} + 1 - \frac{2}{9n}\right)^3\right).$$

On développe comme ci-dessus $\ln f_Z(z)$ lorsque n tend vers l'infini ; on obtient :

$$f_Z(z) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}.$$

Ces approximations donnent des résultats numériquement satisfaisants dès que n est supérieur à 30, l'approximation de WILSON-HILFERTY étant un peu meilleure que celle de FISHER.

3.6.7 Loi de STUDENT

Définition 34. Soient X et Y deux variables aléatoires indépendantes, X suivant la loi normale réduite et Y la loi du χ^2 à n degrés de liberté. La variable aléatoire

$$T = \frac{X}{\sqrt{Y/n}}$$

suit la loi de STUDENT à n degrés de liberté.

Théorème 86. La densité de probabilité de la loi de STUDENT à n degrés de liberté est définie par :

$$f_T(t) = \frac{1}{\sqrt{n} B\left(\frac{1}{2}, \frac{n}{2}\right) \left(1 + \frac{t^2}{n}\right)^{\frac{n+1}{2}}}.$$

On a : $Z = \frac{T^2}{n} = \frac{X^2/2}{Y/2}.$

Or, d'après le théorème 77 (page 55), $X^2/2$ suit la loi $\gamma(\frac{1}{2})$ et, d'après le théorème 78, $Y/2$ suit la loi $\gamma(\frac{n}{2})$; d'autre part ces variables sont indépendantes. D'après le théorème 79 la variable aléatoire Z suit la loi $\beta(\frac{1}{2}, \frac{n}{2})$.

Par ailleurs

$$\begin{aligned} F_T(-t) &= \mathbb{P}(T < -t) = \mathbb{P}(X < -t\sqrt{Y/n}) \\ &= \mathbb{P}(X > t\sqrt{Y/n}), \quad \text{car } X \text{ suit la loi normale réduite,} \\ &= \mathbb{P}(T > t) = 1 - F_T(t); \end{aligned}$$

donc la densité de probabilité f_T est paire et, pour $t > 0$:

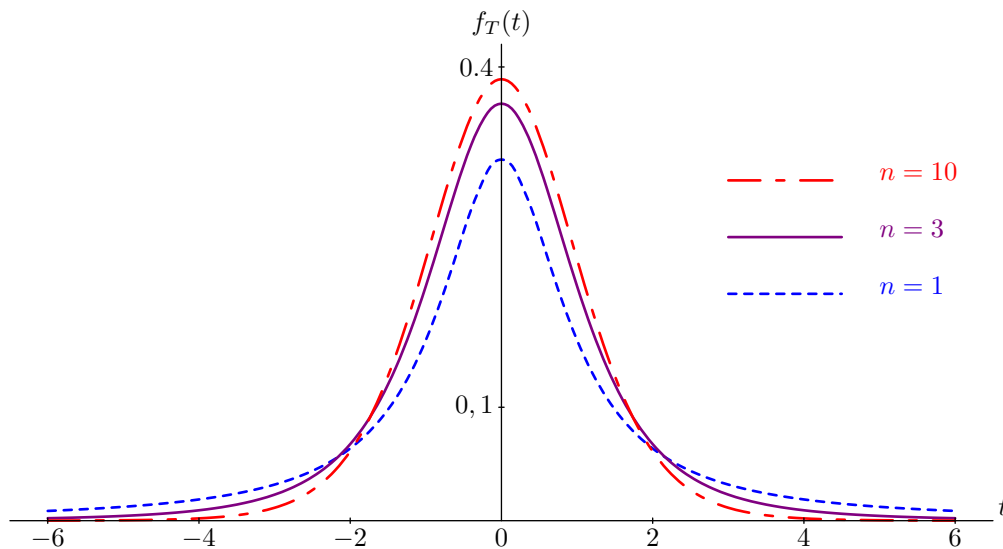
$$\begin{aligned} \mathbb{P}(-t < T < t) &= F_T(t) - F_T(-t) = 2F_T(t) - 1 \\ &= \mathbb{P}(T^2 < t^2) = \mathbb{P}(Z < t^2/n) = F_Z(t^2/n). \end{aligned}$$

En dérivant on obtient :

$$2f_T(t) = \frac{2t}{n} f_Z(t^2/n)$$

et la densité de probabilité de T :

$$f_T(t) = \frac{t}{n} \frac{1}{B(\frac{1}{2}, \frac{n}{2})} \frac{(t^2/n)^{-1/2}}{(1 + t^2/n)^{(n+1)/2}} = \frac{1}{\sqrt{n} B(\frac{1}{2}, \frac{n}{2}) (1 + t^2/n)^{(n+1)/2}}$$



L'espérance mathématique est définie si $n > 1$; elle est nulle puisque la densité de probabilité est paire. La variance est définie pour $n > 2$:

$$\text{Var}(T) = \mathbf{E}(T^2) = \mathbf{E}(nZ) = n \frac{1/2}{n/2 - 1} = \frac{n}{n - 2}.$$

Théorème 87. *La loi de STUDENT converge en loi vers une loi normale réduite.*

La convergence en loi est définie au paragraphe 4.1.

On a

$$B\left(\frac{1}{2}, \frac{n}{2}\right) = \frac{\Gamma(\frac{1}{2})\Gamma(\frac{n}{2})}{\Gamma(\frac{n+1}{2})} \sim \sqrt{\pi} \frac{\sqrt{2\pi}(\frac{n}{2})^{(n-1)/2} e^{-n/2}}{\sqrt{2\pi}(\frac{n+1}{2})^{n/2} e^{-(n+1)/2}} \sim \sqrt{\pi} e^{1/2} \left(\frac{n}{n+1}\right)^{n/2} \left(\frac{n}{2}\right)^{-1/2} \sim \sqrt{\frac{2\pi}{n}}$$

et

$$\left(1 + \frac{t^2}{n}\right)^{(n+1)/2} = \sqrt{\left(1 + \frac{t^2}{n}\right)^n} \left(1 + \frac{t^2}{n}\right)^{1/2} \xrightarrow{n \rightarrow \infty} e^{t^2/2}.$$

Donc :

$$f_T(t) \xrightarrow{n \rightarrow \infty} \frac{e^{-t^2/2}}{\sqrt{2\pi}}.$$

3.6.8 Loi de FISHER-SNEDECOR

Définition 35. Soient X et Y deux variables aléatoires indépendantes suivant des lois du χ^2 à m et n degrés de liberté respectivement. La variable aléatoire

$$Z = \frac{X/m}{Y/n}$$

suit la loi de FISHER-SNEDECOR à m et n degrés de liberté, notée $\mathcal{F}(m, n)$.

Théorème 88. La densité de probabilité de la loi $\mathcal{F}(m, n)$ est définie par :

$$f_{m,n}(z) = \begin{cases} \frac{(m/n)^{\frac{m}{2}}}{B(m/2, n/2)} \frac{z^{\frac{m}{2}-1}}{(\frac{m}{n}z + 1)^{\frac{m+n}{2}}} & \text{si } z > 0, \\ 0 & \text{sinon.} \end{cases}$$

X suit la loi $\gamma(m/2, 1/2)$, donc, d'après le théorème 73, $X/2$ suit la loi $\gamma(m/2)$; de même $Y/2$ suit la loi $\gamma(n/2)$.

D'après le théorème 79 la variable aléatoire $U = \frac{X}{Y} = \frac{X/2}{Y/2}$ suit la loi $\beta(m/2, n/2)$ dont la densité est définie par :

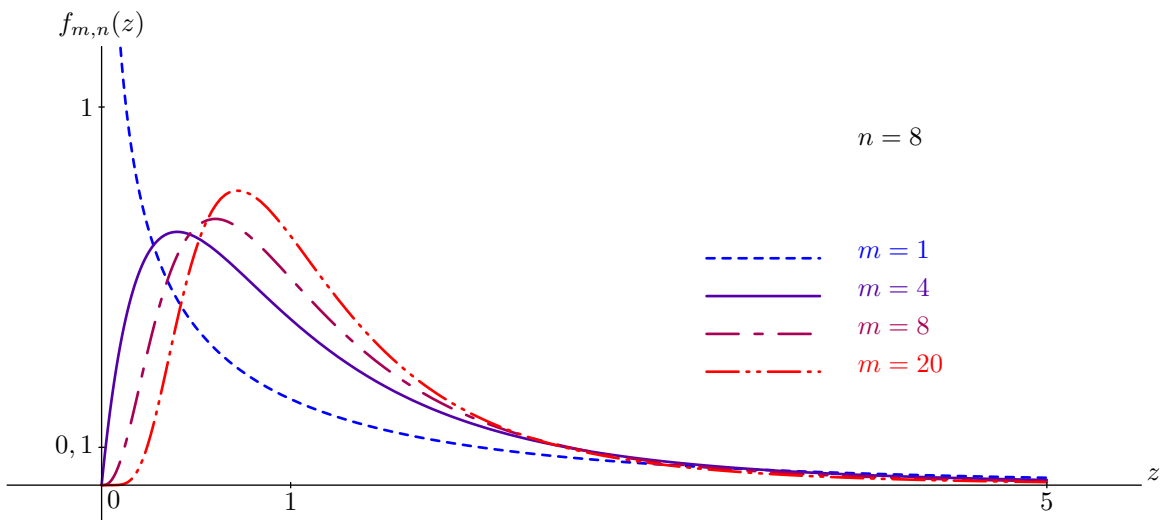
$$f_U(u) = \begin{cases} \frac{1}{B(m/2, n/2)} \frac{u^{\frac{m}{2}-1}}{(u + 1)^{\frac{m+n}{2}}} & \text{si } u > 0, \\ 0 & \text{sinon.} \end{cases}$$

Or : $Z = \frac{X/m}{Y/n} = \frac{n}{m}U$. La fonction de répartition de Z est donc :

$$F_{m,n}(z) = \mathbb{P}(Z < z) = \mathbb{P}(U < \frac{m}{n}z) = F_U(\frac{m}{n}z).$$

D'où la densité de probabilité cherchée :

$$f_{m,n}(z) = \frac{m}{n} f_U(\frac{m}{n}z) = \begin{cases} \frac{(m/n)^{\frac{m}{2}}}{B(m/2, n/2)} \frac{z^{\frac{m}{2}-1}}{(\frac{m}{n}z + 1)^{\frac{m+n}{2}}} & \text{si } z > 0, \\ 0 & \text{sinon.} \end{cases}$$



L'espérance mathématique et de la variance se déduisent immédiatement de celles de la loi bêta :

$$\mathbf{E}(Z) = \frac{n}{m} \mathbf{E}(U) = \frac{n}{m} \frac{m/2}{n/2 - 1} = \frac{n}{n - 2} \text{ pour } n > 2,$$

$$\mathbf{Var}(Z) = \left(\frac{n}{m}\right)^2 \mathbf{Var}(U) = \frac{n^2}{m^2} \frac{m/2 \left((m+n)/2 - 1\right)}{(n/2 - 1)^2 (n/2 - 2)} = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \text{ pour } n > 4.$$

Théorème 89. Si Z suit la loi $\mathcal{F}(m, n)$ alors $\frac{1}{Z}$ suit la loi $\mathcal{F}(n, m)$; donc :

$$F_{n,m}(t) = 1 - F_{m,n}\left(\frac{1}{t}\right).$$

En échangeant les rôles de X et Y on échange m et n . Alors :

$$F_{n,m}(t) = \mathbb{P}\left(\frac{1}{Z} < t\right) = \mathbb{P}\left(Z > \frac{1}{t}\right) = 1 - F_{m,n}\left(\frac{1}{t}\right).$$

Théorème 90. Le carré d'une variable de STUDENT à n degrés de liberté suit la loi de FISHER-SNEDECOR $\mathcal{F}(1, n)$.

Si T suit la loi de STUDENT à n degrés de liberté on peut poser $T = \frac{X}{\sqrt{Y/n}}$ où X suit la loi normale réduite et Y la loi du khi-deux à n degrés de liberté. On a alors $Z = T^2 = \frac{X^2}{Y/n}$ où X^2 suit la loi du khi-deux à un degré de liberté ; la conclusion découle de la définition de la loi de FISHER-SNEDECOR.

Chapitre 4

Convergences stochastiques

4.1 Différents types de convergence

4.1.1 Convergence en loi

Définition 36. Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace probabilisé ; soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires définies sur l'espace probabilisable (Ω, \mathcal{E}) et soit X une variable aléatoire définie sur le même espace. La suite $(X_n)_{n \geq 1}$ converge en loi vers X si la suite des valeurs $F_{X_n}(x)$ des fonctions de répartition converge vers $F_X(x)$ en tout point x où F_X est continue.

On note : $X_n \xrightarrow{L} X$.

Le théorème suivant, que nous admettons, est souvent utilisé pour démontrer la convergence en loi.

Théorème 91 (Paul LÉVY).

1. Soit (X_n) une suite de variables aléatoires qui converge en loi vers la variable aléatoire X . La suite des fonctions caractéristiques $(\widehat{f_{X_n}})$ converge vers la fonction caractéristique de X , la convergence étant uniforme sur tout intervalle borné.
2. Si la suite des fonctions caractéristiques $(\widehat{f_{X_n}})$ converge simplement vers une fonction φ continue à l'origine, alors :
 - (a) φ est une fonction caractéristique ;
 - (b) la suite (X_n) converge en loi vers une variable aléatoire X ayant φ pour fonction caractéristique.

4.1.2 Convergence en probabilité

Définition 37. Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace probabilisé ; soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires définies sur l'espace probabilisable (Ω, \mathcal{E}) et soit X une variable aléatoire définie sur le même espace. La suite $(X_n)_{n \geq 1}$ converge en probabilité vers X si

$$\forall \varepsilon > 0, \mathbb{P}(|X_n - X| > \varepsilon) \xrightarrow{n \rightarrow \infty} 0.$$

On note : $X_n \xrightarrow{P} X$.

Théorème 92. La convergence en probabilité entraîne la convergence en loi.

Démontrons d'abord le lemme suivant.

Lemme 1. Soit (X, Y) un couple aléatoire. On a, pour tout $\eta > 0$:

$$|F_X(x) - F_Y(x)| \leq F_X(x + \eta) - F_X(x - \eta) + \mathbb{P}(|X - Y| > \eta).$$

De

$$(Y < x) = ((Y < x) \cap (X < x + \eta)) \cup ((Y < x) \cap (X \geq x + \eta)) \subset (X < x + \eta) \cup (|X - Y| > \eta)$$

on déduit

$$F_Y(x) \leq F_X(x + \eta) + \mathbb{P}(|X - Y| > \eta).$$

De même, de

$$(X < x - \eta) = ((X < x - \eta) \cap (Y < x)) \cup ((X < x - \eta) \cap (Y \geq x)) \subset (Y < x) \cup (|X - Y| > \eta)$$

on déduit

$$F_X(x - \eta) \leq F_Y(x) + \mathbb{P}(|X - Y| > \eta).$$

D'où :

$$F_X(x - \eta) - \mathbb{P}(|X - Y| > \eta) \leq F_Y(x) \leq F_X(x + \eta) + \mathbb{P}(|X - Y| > \eta).$$

Par ailleurs, puisque F_X est croissante :

$$F_X(x - \eta) \leq F_X(x) \leq F_X(x + \eta).$$

Le lemme s'en déduit par soustraction membre à membre.

Montrons le théorème.

Remplaçant Y par X_n dans le lemme nous obtenons :

$$\forall n \geq 1, \forall \eta > 0, |F_X(x) - F_{X_n}(x)| \leq F_X(x + \eta) - F_X(x - \eta) + \mathbb{P}(|X - X_n| > \eta).$$

Soit $\varepsilon > 0$.

– Supposons que F_X est continue en x :

$$\exists \alpha > 0, 0 < \eta < \alpha \Rightarrow F_X(x + \eta) - F_X(x - \eta) \leq \frac{\varepsilon}{2}.$$

On choisit une valeur η_0 de η vérifiant cette inégalité :

$$F_X(x + \eta_0) - F_X(x - \eta_0) \leq \frac{\varepsilon}{2}.$$

– Supposons que $(X_n)_{n \geq 1}$ converge en probabilité vers X : $\mathbb{P}(|X_n - X| > \eta_0) \xrightarrow{n \rightarrow \infty} 0$.

Donc :

$$\exists N \in \mathbb{N}, n \geq N \Rightarrow \mathbb{P}(|X - X_n| > \eta_0) \leq \frac{\varepsilon}{2}.$$

On en déduit

$$n \geq N \Rightarrow |F_X(x) - F_{X_n}(x)| \leq \underbrace{F_X(x + \eta_0) - F_X(x - \eta_0)}_{\leq \varepsilon/2} + \underbrace{\mathbb{P}(|X - X_n| > \eta_0)}_{\leq \varepsilon/2} \leq \varepsilon,$$

c'est-à-dire : $F_{X_n}(x) \xrightarrow{n \rightarrow \infty} F_X(x)$; la suite $(X_n)_{n \geq 1}$ converge en loi vers X .

Il existe d'autres types de convergence stochastique : en moyenne quadratique, presque sûre... On pourra consulter à ce sujet [8] ou [16] par exemple.

4.2 Loi faible des grands nombres

Théorème 93. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires deux à deux non corrélées et suivant la même loi d'espérance mathématique m et d'écart-type σ . La moyenne $\overline{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ converge en probabilité vers m , c'est-à-dire que l'on a :

$$\forall \varepsilon > 0, \mathbb{P}(|\overline{X}_n - m| \geq \varepsilon) \xrightarrow{n \rightarrow \infty} 0.$$

On a immédiatement : $\mathbf{E}(\overline{X}_n) = m$ et $\mathbf{Var}(\overline{X}_n) = \frac{\sigma^2}{n}$. On applique l'inégalité de BIENAYMÉ-TCHEBYCHEV (théorème 24) :

$$\mathbb{P}(|\overline{X}_n - m| \geq \varepsilon) = \mathbb{P}(|\overline{X}_n - \mathbf{E}(\overline{X}_n)| \geq \varepsilon) \leq \frac{\mathbf{Var}(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0.$$

Théorème 94 (JACQUES BERNOULLI).

La fréquence d'apparition d'un événement dans un grand nombre d'épreuves indépendantes converge en probabilité vers la probabilité de cet événement lorsque le nombre d'épreuves tend vers l'infini.

On considère les variables aléatoires X_k définies par :

$$X_k = \begin{cases} 1 & \text{si l'événement est réalisé à la } k\text{-ième épreuve,} \\ 0 & \text{sinon.} \end{cases}$$

La fréquence d'apparition de l'événement en n épreuves est $\overline{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$. On a

$$\mathbf{E}(\overline{X}_n) = p \quad \text{et} \quad \mathbf{Var}(\overline{X}_n) = p(1 - p)$$

où p représente la probabilité de l'événement.

On applique le théorème précédent : la fréquence \overline{X}_n converge en probabilité vers la probabilité p de l'événement.

4.3 Théorème central limit

Théorème 95 (LINDBERG-LÉVY).

Soit une suite $(X_n)_{n \geq 1}$ de variables aléatoires indépendantes et de même loi, possédant une espérance mathématique m et une variance σ^2 . On pose : $S_n = \sum_{k=1}^n X_k$ et $\overline{X}_n = \frac{S_n}{n}$; la suite des variables réduites

$$S_n^* = \frac{S_n - \mathbf{E}(S_n)}{\sigma(S_n)} = \frac{S_n - nm}{\sigma\sqrt{n}} = \frac{\overline{X}_n - m}{\sigma/\sqrt{n}}$$

converge en loi vers une variable normale réduite.

D'après le théorème 49 la fonction caractéristique φ de $X_k - m$ s'écrit :

$$\varphi(t) = 1 - \frac{\sigma^2}{2}t^2 + o(t^2).$$

En appliquant les théorèmes 47 et 54 on obtient la fonction caractéristique de $S_n^* = \frac{\sum_{k=1}^n (X_k - m)}{\sigma\sqrt{n}}$:

$$\widehat{f_{S_n^*}}(t) = \left(\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \xrightarrow{n \rightarrow \infty} \exp\left(-\frac{t^2}{2}\right)$$

Le théorème 91 permet de conclure.

4.4 Application : approximations de la loi binomiale

Lorsque n est grand le calcul de la probabilité $b(k, n, p) = \binom{n}{k} p^k q^{n-k}$ est délicat ; on peut en donner deux approximations correspondant à deux domaines de validité.

4.4.1 Approximation normale

Théorème 96 (MOIVRE-LAPLACE).

Soit une suite de variables aléatoires S_n suivant la loi binomiale $\mathcal{B}(n, p)$; la suite des variables réduites $S_n^* = \frac{S_n - np}{\sqrt{npq}}$ converge en loi vers une variable normale réduite :

$$\mathbb{P}(a \leq S_n^* < b) \xrightarrow{n \rightarrow \infty} \Phi(b) - \Phi(a).$$

Il suffit d'appliquer le théorème 95 en prenant pour loi de X_n la loi de BERNOULLI $\mathcal{B}(1, p)$.

Une valeur approchée de $b(k, n, p)$ lorsque n est « grand » est : $\frac{1}{\sqrt{npq}} \varphi\left(\frac{k - np}{\sqrt{npq}}\right)$.

En pratique les approximations sont convenables

1. lorsque p est voisin de 0.5 et $n \geq 10$,
2. lorsque $n \geq 50$.

4.4.2 Approximation de POISSON

Théorème 97. Supposons que, pour tout n , la variable aléatoire X_n suit la loi binomiale $\mathcal{B}(n, p_n)$ avec $np_n \xrightarrow{n \rightarrow \infty} \mu$; la suite $(X_n)_{n \in \mathbb{N}^*}$ converge en loi vers une variable de POISSON de loi $\mathcal{P}(\mu)$:

$$\binom{n}{k} p_n^k (1 - p_n)^{n-k} \xrightarrow{n \rightarrow \infty} e^{-\mu} \frac{\mu^k}{k!}.$$

On pose : $u_n = np_n$; par hypothèse : $u_n \xrightarrow{n \rightarrow \infty} \mu$.

$$\binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{1}{k!} \frac{n!}{(n-k)!} \left(\frac{u_n}{n}\right)^k \left(1 - \frac{u_n}{n}\right)^{n-k}.$$

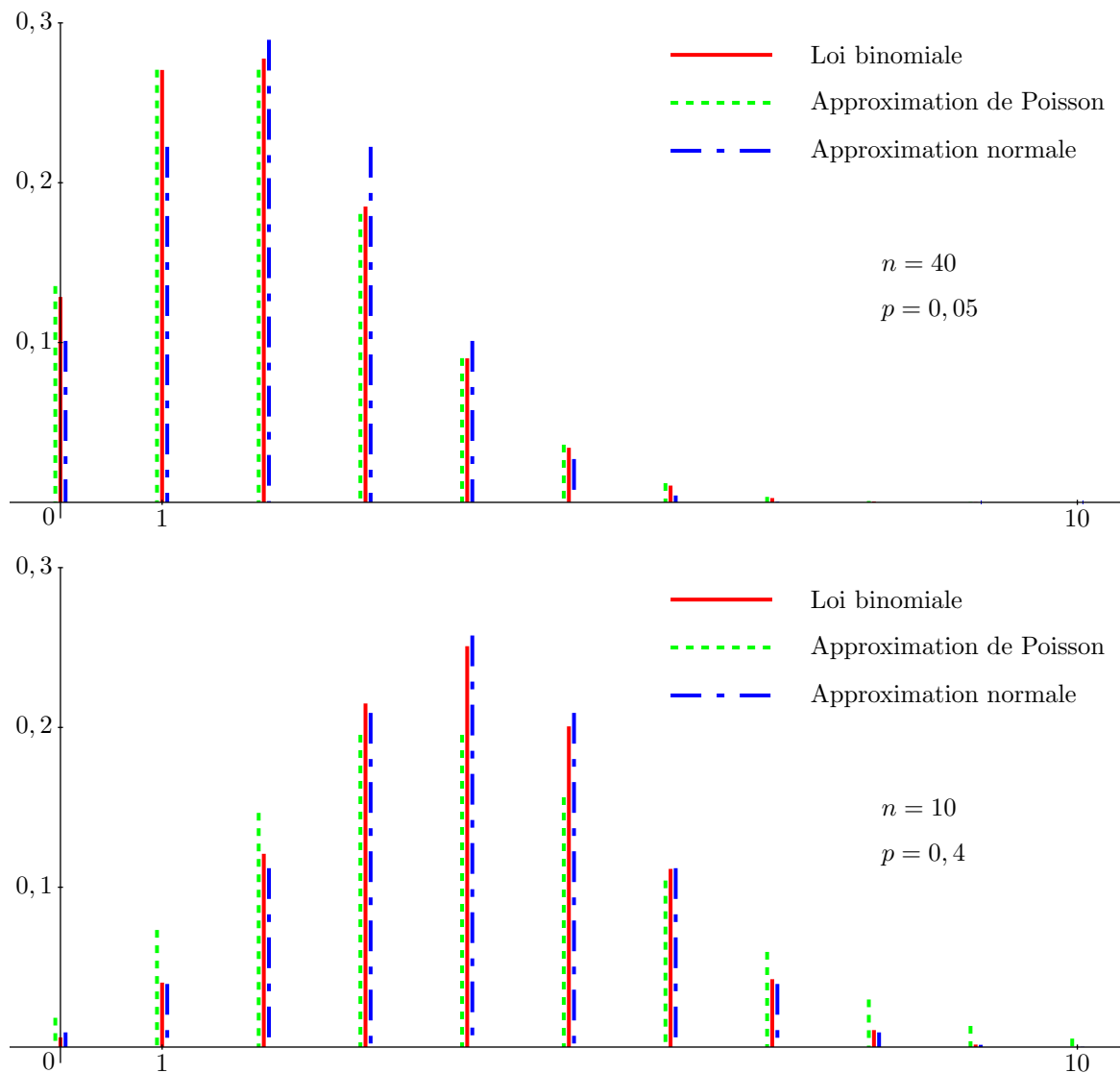
Or :

- $\frac{n!}{(n-k)!} = n^k \prod_{i=1}^{k-1} \left(1 - \frac{i}{n}\right) \sim n^k$;
- $\left(\frac{u_n}{n}\right)^k \sim \left(\frac{\mu}{n}\right)^k$;
- $\ln\left(\left(1 - \frac{u_n}{n}\right)^{n-k}\right) = (n-k) \ln\left(1 - \frac{u_n}{n}\right) \xrightarrow{n \rightarrow \infty} -\mu$.

Donc :

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} \xrightarrow{n \rightarrow \infty} e^{-\mu} \frac{\mu^k}{k!}.$$

Remarque : puisque np_n tend vers une limite finie, p_n tend vers zéro lorsque n tend vers l'infini ; l'approximation de POISSON de $b(k, n, p)$ est valable pour les « petites » valeurs de p . En pratique elle donne des résultats convenables lorsque $p < 0.2$ et $10 < np < 50$. Sinon il vaut mieux utiliser l'approximation normale.



Chapitre 5

Chaînes de MARKOV

5.1 Introduction

Définition 38. Un processus aléatoire est une application d'un ensemble \mathcal{T} inclus dans \mathbb{R} dans un ensemble de variables aléatoires prenant leurs valeurs dans un ensemble E :

$$t \in \mathcal{T} \longrightarrow (X_t : \Omega \rightarrow E)$$

On peut distinguer différents cas :

- \mathcal{T} est discret : le processus est dit discret. On peut avoir :
 - \mathcal{T} est fini : $\mathcal{T} = \{t_0, t_1, \dots, t_n\}$ ou, plus simplement, $\mathcal{T} = \{0, 1, \dots, n\}$,
 - \mathcal{T} est dénombrable : $\mathcal{T} = \{t_0, t_1, \dots, t_n, \dots\}$ ou $\mathcal{T} = \mathbb{N}$.
- \mathcal{T} est continu : le processus est dit permanent. En général \mathcal{T} est un intervalle de \mathbb{R} , par exemple $\mathcal{T} = [0, \infty[$.
- L'ensemble E des états (espace d'états) peut être fini, dénombrable ou continu.

Exemple 5.1.1. Lancers successifs d'un dé : on choisit $\mathcal{T} = \mathbb{N}^*$ et $E = \{1, 2, 3, 4, 5, 6\}$. C'est un processus discret à espace d'états fini.

$$\forall k \in \mathcal{T}, \forall i \in E, P(X_k = i) = \frac{1}{6}.$$

Exemple 5.1.2. Cheminement aléatoire dans le plan : aux instants $0, 1, \dots$ un mobile, initialement en O , se déplace d'une unité parallèlement à l'un des axes. On choisit $\mathcal{T} = \mathbb{N}$ et $E = \mathbb{Z}^2$. C'est un processus discret à espace d'états dénombrable.

Exemple 5.1.3. Processus de POISSON : $\mathcal{T} = [0, \infty[$ et $E = \mathbb{N}$ avec

$$P(X_t = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}.$$

C'est un processus permanent à espace d'états dénombrable.

5.2 Définition

Définition 39. Une chaîne de MARKOV est un processus aléatoire discret où $\mathcal{T} = \mathbb{N}$,

- $E = \mathbb{N}$ si l'espace d'états est dénombrable,
 - $E = \{1, 2, \dots, r\}$ si l'espace d'états est fini,
- tel que

$$\forall n \in \mathbb{N}, \forall m < n, \forall 0 \leq j_1 < j_2 < \dots < j_m < n, \forall i, i_1, i_2, \dots, i_m \in E,$$

$$P(X_n = i \mid \bigcap_{k=1}^m (X_{j_k} = i_k)) = P(X_n = i \mid X_{j_m} = i_m).$$

En d'autres termes l'état présent résume toute l'information utile pour connaître le comportement futur. On a :

$$P(X_{n+1} = j) = P\left(\bigcup_{i \in E} [(X_n = i) \cap (X_{n+1} = j)]\right) = \sum_{i \in E} P(X_n = i) P(X_{n+1} = j \mid X_n = i).$$

Si l'espace E des états est fini on pose :

$$a_{i,j}^{\{n\}} = P(X_{n+1} = j \mid X_n = i) \quad p_{i,n} = P(X_n = i)$$

et on introduit les matrices :

$$A_n = (a_{i,j}^{\{n\}})_{\substack{1 \leq i \leq r \\ 1 \leq j \leq r}} \quad \Pi_n = \begin{bmatrix} p_{1,n} & p_{2,n} & \dots & p_{r,n} \end{bmatrix}.$$

On a alors :

$$\forall j \in \{1, 2, \dots, r\}, p_{j,n+1} = \sum_{i=1}^r p_{i,n} a_{i,j}^{\{n\}}$$

ce qui s'écrit matriciellement :

$$\Pi_{n+1} = \Pi_n A_n.$$

D'où, par récurrence :

$$\Pi_n = \Pi_0 A_0 A_1 \dots A_{n-1} \quad (5.1)$$

Le processus est dit *homogène* (dans le temps) si la probabilité $a_{i,j}^{\{n\}}$ ne dépend pas de n ; la probabilité de transition en une étape est alors, *quelle que soit l'étape* :

$$a_{i,j} = P(X_{n+1} = j \mid X_n = i).$$

Si l'espace d'états E est fini on définit la *matrice (des probabilités) de transition* :

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,r} \\ a_{2,1} & a_{2,2} & \dots & a_{2,r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{r,1} & a_{r,2} & \dots & a_{r,r} \end{bmatrix}.$$

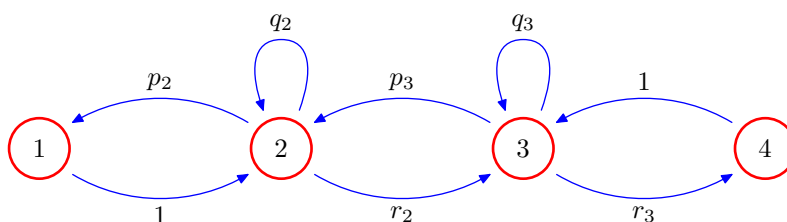
Nous nous limiterons par la suite aux chaînes homogènes à espace d'états fini.

Exemple 5.2.1. Promenade avec barrières réfléchissantes.

Un point peut occuper r positions sur une droite. Il se déplace d'une unité à chaque transition ; il est réfléchi aux extrémités. On aura par exemple (avec $r = 4$) : $\mathcal{T} = \mathbb{N}$ et $E = \{1, 2, 3, 4\}$ avec pour matrice de transition

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ p_2 & q_2 & r_2 & 0 \\ 0 & p_3 & q_3 & r_3 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad \text{avec} \quad \begin{cases} p_2 + q_2 + r_2 = 1 \\ p_3 + q_3 + r_3 = 1 \end{cases}$$

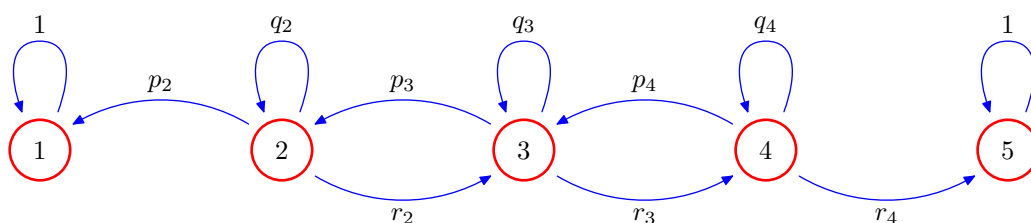
qui peut se représenter par le graphe :

**Exemple 5.2.2.** Promenade avec barrières absorbantes.

Un point peut occuper r positions sur une droite. Il se déplace d'une unité à chaque transition ; il est absorbé aux extrémités. On aura par exemple (avec $r = 5$) : $\mathcal{T} = \mathbb{N}$ et $E = \{1, 2, 3, 4, 5\}$ avec pour matrice de transition

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ p_2 & q_2 & r_2 & 0 & 0 \\ 0 & p_3 & q_3 & r_3 & 0 \\ 0 & 0 & p_4 & q_4 & r_4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{avec} \quad \begin{cases} p_2 + q_2 + r_2 = 1 \\ p_3 + q_3 + r_3 = 1 \\ p_4 + q_4 + r_4 = 1 \end{cases}$$

qui peut se représenter par le graphe :

**Exemple 5.2.3.** Cheminement aléatoire dénombrable.

$\mathcal{T} = \mathbb{N}$ et $E = \mathbb{Z}$ avec les probabilités de transition :

$$a_{i,j} = \begin{cases} p & \text{si } j = i + 1, \\ q & \text{si } j = i - 1, \\ 0 & \text{sinon,} \end{cases}$$

où l'on a posé $q = 1 - p$.

5.3 Méthodes algébriques

5.3.1 Relation fondamentale

On rappelle :

$$p_{i,n} = P(X_n = i) \quad \Pi_n = \begin{bmatrix} p_{1,n} & p_{2,n} & \dots & p_{r,n} \end{bmatrix}.$$

Théorème 98.

$$\Pi_n = \Pi_0 A^n$$

C'est un cas particulier de la formule (5.1) du paragraphe 5.2.

La matrice (des probabilités) de transition en n étapes est donc $A^n = (a_{i,j}^{(n)})_{\substack{1 \leq i \leq r \\ 1 \leq j \leq r}}$.

Le processus étant homogène on a :

$$\forall m \in \mathbb{N}, a_{i,j}^{(n)} = P(X_{m+n} = j \mid X_m = i).$$

Une chaîne de MARKOV est donc déterminée par la donnée de A et de Π_0 .

5.3.2 Matrices stochastiques

Définition 40. Une matrice stochastique est une matrice $(a_{i,j})_{\substack{1 \leq i \leq r \\ 1 \leq j \leq r}}$ telle que :

$$\forall i \in \{1, 2, \dots, r\}, \forall j \in \{1, 2, \dots, r\}, a_{i,j} \geq 0$$

et

$$\forall i \in \{1, 2, \dots, r\}, \sum_{j=1}^r a_{i,j} = 1.$$

Théorème 99. Toute matrice de transition est une matrice stochastique.

$$\forall i \in \{1, 2, \dots, r\}, \sum_{j=1}^r a_{i,j} = \sum_{j=1}^r P(X_{n+1} = j \mid X_n = i) = P_{(X_n=i)} \left(\underbrace{\bigcup_{j=1}^r (X_{n+1} = j)}_{\Omega} \right) = 1.$$

Théorème 100. Le produit de deux matrices stochastiques est une matrice stochastique.

Supposons que les matrices A et B sont stochastiques et posons $C = AB$. On a

$$\forall (i, j) \in \{1, 2, \dots, r\}^2, c_{i,j} = \sum_{k=1}^r a_{i,k} b_{k,j} \geq 0$$

et

$$\forall i \in \{1, 2, \dots, r\}, \sum_{j=1}^r c_{i,j} = \sum_{j=1}^r \sum_{k=1}^r a_{i,k} b_{k,j} = \sum_{k=1}^r a_{i,k} \underbrace{\sum_{j=1}^r b_{k,j}}_{=1, \forall k} = \sum_{k=1}^r a_{i,k} = 1.$$

La matrice A^n de transition en n étapes est donc stochastique.

Théorème 101. Soit A une matrice stochastique.

A admet 1 pour valeur propre (simple ou multiple).

Les autres valeurs propres de A ont un module inférieur ou égal à 1.

Il résulte directement de la définition des matrices stochastiques que 1 est valeur propre associée aux vecteurs propres ayant toutes leurs coordonnées égales.

Soit λ une valeur propre différente de 1 et $x = [x_1 \ x_2 \ \dots \ x_r]^T$ un vecteur propre associé ; il résulte de la définition

$$\forall i \in \{1, 2, \dots, r\}, \sum_{\substack{j=1 \\ j \neq i}}^r a_{i,j} x_j = (\lambda - a_{i,i}) x_i.$$

Soit k un indice tel que $\forall j \in \{1, 2, \dots, r\}, |x_j| \leq |x_k|$. On a alors :

$$|\lambda - a_{k,k}| |x_k| \leq \sum_{\substack{j=1 \\ j \neq k}}^r a_{k,j} |x_j| \leq |x_k| \sum_{\substack{j=1 \\ j \neq k}}^r a_{k,j} = |x_k| (1 - a_{k,k}).$$

D'où, puisque $x_k \neq 0 : |\lambda - a_{k,k}| \leq 1 - a_{k,k}$, et donc $|\lambda| \leq 1$.

Les éléments d'une matrice stochastique A sont tous positifs, un au moins par ligne étant non nul. Si aucun élément n'est nul nous noterons $A \gg 0$.

Théorème 102. Si A est une matrice stochastique les valeurs propres de module égal à 1 sont racines de l'unité.

Soit λ une valeur propre de A telle que $|\lambda| = 1$ et $\lambda \neq 1$: il existe un vecteur propre x non nul tel que $Ax = \lambda x$. D'où :

$$\forall i \in \{1, 2, \dots, r\}, \lambda x_i = \sum_{j=1}^r a_{i,j} x_j.$$

Soit k un indice tel que, pour tout i , on ait $|x_i| \leq |x_k| : \lambda x_k = \sum_{j=1}^r a_{k,j} x_j$.

Puisque $|\lambda| = 1$ on a $|\lambda x_k| = |x_k|$: l'image de λx_k se trouve sur le cercle de centre O et de rayon $|x_k|$; mais λx_k est le barycentre, avec des coefficients positifs, des x_j dont les images sont sur ou dans ce même cercle. Donc λx_k est l'un des x_j , par exemple x_{j_1} . On a : $x_{j_1} = \lambda x_k \neq x_k$ car $\lambda \neq 1$. On recommence le même raisonnement. On a :

$$|\lambda x_{j_1}| = |x_{j_1}| = |x_k|, (\forall i, |x_i| \leq |x_{j_1}|) \text{ et } \lambda x_{j_1} = \sum_{j=1}^r a_{j_1,j} x_j.$$

Donc λx_{j_1} est l'un des x_j , par exemple $x_{j_2} : x_{j_2} = \lambda x_{j_1} = \lambda^2 x_k$. Distinguons deux cas :

1. soit $x_{j_2} = x_k$ et $\lambda^2 = 1$: λ est une racine de l'unité ;
2. soit $x_{j_2} \neq x_k$ et on itère le procédé. Le nombre de coordonnées de x étant fini on finira par obtenir $x_{j_p} = \lambda^p x_k = x_k$ et donc $\lambda^p = 1$: λ est racine de l'unité.

Théorème 103. Si $A \gg 0$ la seule valeur propre de module égal à 1 est 1 (elle est simple ou multiple).

Théorème admis.

5.3.3 Comportement asymptotique

Théorème 104. Si $A \gg 0$ alors A^n possède une limite A_∞ lorsque n tend vers l'infini ;

- A_∞ est stochastique,
- $A_\infty \gg 0$,
- A_∞ a toutes ses lignes identiques.

Si $A \gg 0$ il existe $\varepsilon > 0$ tel que $\forall (i, j) \in \{1, 2, \dots, r\}^2$, $a_{i,j} \geq \varepsilon$.

De $A^{n+1} = A A^n$ on déduit

$$\forall (i, j) \in \{1, 2, \dots, r\}^2, a_{i,j}^{(n+1)} = \sum_{k=1}^r a_{i,k} a_{k,j}^{(n)} \quad \text{avec} \quad \forall i \in \{1, 2, \dots, r\}, \sum_{k=1}^r a_{i,k} = 1,$$

c'est-à-dire que $a_{i,j}^{(n+1)}$ est le barycentre des $(a_{k,j}^{(n)})_{1 \leq k \leq r}$ affectés des poids $a_{i,k} \geq 0$.

Supposons j fixé.

Pour tout i appartenant à $\{1, 2, \dots, r\}$, $a_{i,j}^{(n+1)}$ appartient à l'intervalle qui contient les $(a_{k,j}^{(n)})_{1 \leq k \leq r}$.

Soient y_1, y_2, \dots, y_r les éléments de la j -ième colonne de A^n en supposant $0 \leq y_1 \leq y_2 \leq \dots \leq y_r \leq 1$: $\{y_1, y_2, \dots, y_r\} = \{a_{1,j}^{(n)}, a_{2,j}^{(n)}, \dots, a_{r,j}^{(n)}\}$.

$\ell_n = y_r - y_1$ est l'amplitude de l'intervalle qui contient tous les $a_{i,j}^{(n)}$.

De même notons z_1, z_2, \dots, z_r les éléments de la j -ième colonne de A^{n+1} en supposant $0 \leq z_1 \leq z_2 \leq \dots \leq z_r \leq 1$: $\{z_1, z_2, \dots, z_r\} = \{a_{1,j}^{(n+1)}, a_{2,j}^{(n+1)}, \dots, a_{r,j}^{(n+1)}\} \subseteq [y_1, y_r]$.

On a $\ell_{n+1} = z_r - z_1$ et

$$\forall h \in \{1, 2, \dots, r\}, z_h = \sum_{k=1}^r c_{h,k} y_k \quad \text{avec} \quad \left| \begin{array}{l} \forall (h, k), c_{h,k} \geq \varepsilon > 0, \\ \sum_{k=1}^r c_{h,k} = 1, \end{array} \right.$$

car les $c_{h,k}$ sont les éléments d'une ligne de A . D'où, pour tout h :

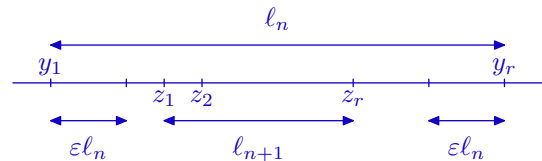
$$\begin{aligned} z_h - y_1 &= \sum_{k=1}^r c_{h,k} y_k - \sum_{k=1}^r c_{h,k} y_1 = \sum_{k=1}^r c_{h,k} (y_k - y_1) \\ &= \sum_{k=1}^{r-1} \underbrace{c_{h,k} (y_k - y_1)}_{\geq 0} + \underbrace{c_{h,r} (y_r - y_1)}_{\substack{\geq \varepsilon \\ = \ell_n}} \geq \varepsilon \ell_n. \end{aligned}$$

De même, pour tout h :

$$y_r - z_h = c_{h,1} (y_r - y_1) + \sum_{k=2}^r c_{h,k} (y_r - y_k) \geq \varepsilon \ell_n.$$

On en déduit :

$$\ell_{n+1} \leq \ell_n (1 - 2\varepsilon).$$



Or $\ell_1 \leq 1 - 2\varepsilon$ car le plus petit terme de toute colonne de A est supérieur ou égal à ε et le plus grand terme est inférieur ou égal à $(1 - \varepsilon)$ puisque la somme des termes de sa ligne, tous supérieurs ou égaux à ε , est égale à 1 (il est même inférieur à $(1 - (r - 1)\varepsilon)$...).

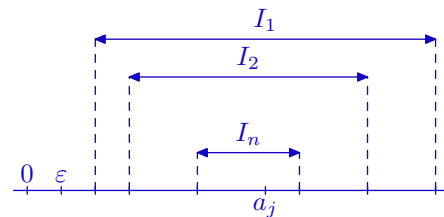
Donc : $\ell_n \leq (1 - 2\varepsilon)^n \xrightarrow{n \rightarrow \infty} 0$; tous les éléments de la j -ième colonne de A^n appartiennent à un intervalle I_n dont la longueur tend vers zéro : ils ont une limite commune a_j .

On a de plus

$$\forall n \geq 1, \forall (i, j) \in \{1, 2, \dots, r\}^2, |a_{i,j}^{(n)} - a_j| \leq (1 - 2\varepsilon)^n$$

car les intervalles I_n sont emboîtés ; donc :

$$\forall j \in \{1, 2, \dots, r\}, a_j > 0.$$



Enfin :

$$\sum_{j=1}^r a_j = \sum_{j=1}^r \left(\lim_{n \rightarrow \infty} a_{i,j}^{(n)} \right) = \lim_{n \rightarrow \infty} \left(\sum_{j=1}^r a_{i,j}^{(n)} \right) = 1.$$

Dans ce cas, en passant à la limite dans $\Pi_n = \Pi_0 A^n$, on voit que Π_n a une limite $\Pi_\infty = \Pi_0 A_\infty$ (vecteur d'état permanent) lorsque n tend vers l'infini ; cette limite est indépendante de Π_0 puisque les lignes de A_∞ sont égales (ergodicité).

En passant à la limite dans $\Pi_{n+1} = \Pi_n A$ on obtient $\Pi_\infty = \Pi_\infty A$, c'est-à-dire que le vecteur d'état permanent Π_∞ est un vecteur propre à gauche de A associé à la valeur propre 1 ; le vecteur d'état permanent correspond à une distribution stationnaire.

Attention : une chaîne de MARKOV peut avoir une distribution stationnaire Π_* qui ne soit pas un vecteur d'état permanent. Par exemple :

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \Pi_0 = \begin{bmatrix} 1/4 & 3/4 \end{bmatrix} \quad \Pi_* = \begin{bmatrix} 1/2 & 1/2 \end{bmatrix}.$$

Théorème 105. Soit A une matrice stochastique.

$\exists m \in \mathbb{N}, A^m \gg 0 \iff (A^n \text{ tend vers une matrice stochastique } A_\infty \text{ dont les lignes sont égales et } A_\infty \gg 0).$

Une condition *suffisante* pour que ce théorème s'applique est $|i - j| \leq 1 \implies a_{i,j} > 0$.

1. Supposons : $\exists m \in \mathbb{N}, A^m \gg 0$. D'après le théorème précédent $(A^m)^q$ a une limite B lorsque q tend vers l'infini où B est stochastique, $B \gg 0$ et B a toutes ses lignes égales.

Or, pour tout n , on a : $n = mq + s$ avec $0 \leq s < m$; d'où : $A^n = A^s (A^m)^q$.

Par ailleurs $A^s B = B$ puisque A^s est stochastique et que B a toutes ses lignes égales.

Donc $A^n \xrightarrow{n \rightarrow \infty} B$.

2. Supposons :

$$A^n \xrightarrow{n \rightarrow \infty} A_\infty = \begin{pmatrix} a_1 & a_2 & \dots & a_r \\ a_1 & a_2 & \dots & a_r \\ \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & \dots & a_r \end{pmatrix} \quad \text{où } \forall (i, j) \in \{1, 2, \dots, r\}^2, a_j = \lim_{n \rightarrow \infty} a_{i,j}^{(n)} > 0.$$

Soit $\varepsilon > 0$ tel que $\forall j \in \{1, 2, \dots, r\}, a_j \geq \varepsilon$. On a

$$\forall (i, j) \in \{1, 2, \dots, r\}^2, \exists N(i, j) \text{ tel que } n \geq N(i, j) \Rightarrow |a_{i,j}^{(n)} - a_j| \leq \frac{\varepsilon}{2}$$

et donc

$$\forall (i, j) \in \{1, 2, \dots, r\}^2, a_{i,j}^{(n)} \geq \frac{\varepsilon}{2}.$$

Soit $m = \max_{(i,j)} N(i, j)$; on a $\forall (i, j) \in \{1, 2, \dots, r\}^2, a_{i,j}^{(m)} \geq \frac{\varepsilon}{2}$ et donc $A^m \gg 0$.

5.4 Classification des états

5.4.1 Définitions

L'état j est accessible à partir de l'état i s'il existe un entier n tel que $a_{i,j}^{(n)} > 0$; on note $i \rightarrow j$.

Les états i et j communiquent si chacun est accessible à partir de l'autre : $i \rightarrow j$ et $j \rightarrow i$; on note $i \rightleftharpoons j$.

Théorème 106. *La relation \rightleftharpoons est une relation d'équivalence.*

La réflexivité est immédiate en posant (puisque $A^0 = I$) :

$$a_{i,j}^{(0)} = \delta_{i,j} = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{sinon.} \end{cases}$$

La symétrie résulte de la définition.

Pour la transitivité de

$$(\exists n, a_{i,j}^{(n)} > 0) \quad \text{et} \quad (\exists m, a_{j,k}^{(m)} > 0)$$

on déduit :

$$a_{i,k}^{(n+m)} = \sum_{l=1}^r a_{i,l}^{(n)} a_{l,k}^{(m)} \geq a_{i,j}^{(n)} a_{j,k}^{(m)} > 0.$$

Définition 41. *Une chaîne est irréductible si elle possède une seule classe d'équivalence.*

Définition 42. *Une classe dont on peut sortir est une classe de passage (on ne peut y entrer à nouveau...); sinon c'est une classe finale.*

Exemple 5.4.1. *Promenade avec barrières réfléchissantes (voir exemple 5.2.1) : la chaîne est irréductible.*

Exemple 5.4.2. *Promenade avec barrières absorbantes (voir exemple 5.2.2). Il y a trois classes :*

$$C_1 = \{1\} \quad C_2 = \{2, 3, 4\} \quad C_3 = \{5\}.$$

C_2 est une classe de passage, C_1 et C_3 des classes finales.

5.4.2 Périodicité

Définition 43. La période $d(i)$ de l'état i est le PGCD des entiers $n \geq 1$ tels que $a_{i,i}^{(n)} > 0$.

Si $a_{i,i}^{(n)}$ est nul pour tout $n \geq 1$ on note $d(i) = 0$ par convention.

Si tous les états ont pour période 1 la chaîne est apériodique.

Théorème 107. $\forall i \in \{1, 2, \dots, r\}, \exists N \in \mathbb{N}, n \geq N \implies a_{i,i}^{(nd(i))} > 0$.

Attention : on n'a pas en général $a_{i,i}^{(d(i))} > 0$. Dans l'exemple 5.2.1 l'état 4 a une période égale à 1 si $q_3 \neq 0$, mais on ne peut y revenir en une transition.

Théorème 108. Tous les états d'une même classe ont la même période : c'est la période de la classe.

Soient i et j deux états d'une même classe : $\exists m \geq 1, a_{i,j}^{(m)} > 0$ et $\exists n \geq 1, a_{j,i}^{(n)} > 0$.

Soit $s \geq 1$ tel que $a_{i,i}^{(s)} > 0$ ($s = m + n$ convient, par exemple). On a :

$$a_{j,j}^{(n+m+s)} = \sum_{k=1}^r a_{j,k}^{(n+s)} a_{k,j}^{(m)} = \sum_{k=1}^r \sum_{l=1}^r a_{j,l}^{(n)} a_{l,k}^{(s)} a_{k,j}^{(m)} \geq a_{j,i}^{(n)} a_{i,i}^{(s)} a_{i,j}^{(m)} > 0.$$

Donc $d(j)$ divise $n + m + s$. Par ailleurs :

$$a_{i,i}^{(2s)} = \sum_{k=1}^r a_{i,k}^{(s)} a_{k,i}^{(s)} \geq \left(a_{i,i}^{(s)}\right)^2 > 0$$

ce qui entraîne que $d(j)$ divise également $n + m + 2s$; donc $d(j)$ divise s .

Puisque $d(j)$ divise tous les s tels que $a_{i,i}^{(s)} > 0$ il divise leur PGCD $d(i)$; de la même façon $d(i)$ divise $d(j)$ et, finalement, $d(i) = d(j)$.

Exemple 5.4.3. Promenade avec barrières réfléchissantes (exemple 5.2.1) : la chaîne est apériodique, sauf si $q_2 = q_3 = 0$ auquel cas la période est égale à 2.

Exemple 5.4.4. Promenade avec barrières absorbantes (exemple 5.2.2) : les classes finales ont une période égale à 1, la classe de passage C_2 a une période égale à 1 si l'un des q_i est non nul et à 2 si $q_2 = q_3 = q_4 = 0$.

5.4.3 Récurrence

Définition 44. La probabilité de premier retour en i en n étapes est :

$$f_{i,i}^{(n)} = P\left((X_n = i) \cap \left[\bigcap_{j=1}^{n-1} (X_j \neq i)\right] \mid (X_0 = i)\right).$$

Notant $T_{i,j}$ le temps de premier passage en j partant de i on a $f_{i,i}^{(n)} = P(T_{i,i} = n)$.

On pose par convention : $f_{i,i}^{(0)} = 0$.

Théorème 109.

$$\forall n \geq 1, a_{i,i}^{(n)} = \sum_{k=0}^n f_{i,i}^{(k)} a_{i,i}^{(n-k)}.$$

$$\begin{aligned}
a_{i,i}^{(n)} &= P(X_n = i \mid X_0 = i) \\
&= P_{(X_0=i)} \left((X_n = i) \cap \left(\bigcup_{k=1}^n \left((X_k = i) \cap \left[\bigcap_{l=1}^{k-1} (X_l \neq i) \right] \right) \right) \right) \\
&= \sum_{k=1}^n P_{(X_0=i)} \left((X_n = i) \cap (X_k = i) \cap \left[\bigcap_{l=1}^{k-1} (X_l \neq i) \right] \right) \\
&= \sum_{k=1}^n P_{(X_0=i)} \left((X_k = i) \cap \left[\bigcap_{l=1}^{k-1} (X_l \neq i) \right] \right) P(X_n = i \mid X_k = i) \\
&= \sum_{k=1}^n f_{i,i}^{(k)} a_{i,i}^{(n-k)} = \sum_{k=0}^n f_{i,i}^{(k)} a_{i,i}^{(n-k)}
\end{aligned}$$

Cette relation permet de calculer les $f_{i,i}^{(k)}$ de proche en proche. On peut également utiliser les fonctions génératrices ; on pose :

$$F_{i,i}(z) = \sum_{k=0}^{\infty} f_{i,i}^{(k)} z^k = \sum_{k=1}^{\infty} f_{i,i}^{(k)} z^k \quad \text{et} \quad A_{i,i}(z) = \sum_{k=0}^{\infty} a_{i,i}^{(k)} z^k = 1 + \sum_{k=1}^{\infty} a_{i,i}^{(k)} z^k.$$

Théorème 110.

$$|z| < 1 \implies F_{i,i}(z) A_{i,i}(z) = A_{i,i}(z) - 1.$$

$$F_{i,i}(z) A_{i,i}(z) = \left(\sum_{k=0}^{\infty} f_{i,i}^{(k)} z^k \right) \left(\sum_{k=0}^{\infty} a_{i,i}^{(k)} z^k \right) = \sum_{n=0}^{\infty} c_n z^n$$

avec $c_0 = 0$ car $f_{i,i}^{(0)} = 0$ et

$$c_n = \sum_{k=0}^n f_{i,i}^{(k)} a_{i,i}^{(n-k)} = a_{i,i}^{(n)} \quad \text{si } n \geq 1$$

d'après le théorème 109. D'où le résultat en reportant.

Définition 45. L'état i est dit récurrent si

$$f_{i,i}^* = \sum_{n=1}^{\infty} f_{i,i}^{(n)} = 1.$$

Le temps de récurrence moyen est alors :

$$\mu_{i,i} = E(T_{i,i}) = \sum_{n=1}^{\infty} n f_{i,i}^{(n)} = F'_{i,i}(1).$$

Un état qui n'est pas récurrent est dit transitoire : $\sum_{n=1}^{\infty} f_{i,i}^{(n)} < 1$.

Théorème 111. *L'état i est récurrent $\iff \sum_{n=1}^{\infty} a_{i,i}^{(n)} = \infty$ (la série diverge).*

On rappelle le théorème :

Soit une série entière dont le rayon de convergence est égal à R . Si la série converge pour $x = R$ sa somme est continue sur l'intervalle $[0, R]$.

1. Supposons que l'état i est récurrent : $\sum_{n=1}^{\infty} f_{i,i}^{(n)} = 1$. La série entière $F_{i,i}(z) = \sum_{n=1}^{\infty} f_{i,i}^{(n)} z^n$ converge pour $z = 1$; sa somme $F_{i,i}$ est donc continue sur l'intervalle $[0, 1]$: $F_{i,i}(z) \xrightarrow{z \rightarrow 1^-} 1$.

Donc $A_{i,i}(z) \xrightarrow{z \rightarrow 1^-} +\infty$ et $\sum_{n=0}^{\infty} a_{i,i}^{(n)} = +\infty$ car sinon la série $\sum_{n=0}^{\infty} a_{i,i}^{(n)} z^n$ converge pour $z = 1$ et la fonction $A_{i,i}$, continue sur $[0, 1]$, a une limite finie lorsque z tend vers 1 par valeur inférieure.

2. Si la série $\sum_{n=0}^{\infty} a_{i,i}^{(n)}$ diverge on a $A_{i,i}(z) \xrightarrow{z \rightarrow 1^-} +\infty$, donc $F_{i,i}(z) \xrightarrow{z \rightarrow 1^-} 1$ et finalement $\sum_{n=1}^{\infty} f_{i,i}^{(n)} = 1$.

Théorème 112. *Les états d'une même classe sont*

- ou bien tous récurrents : la classe est dite *récurrente* (les classes récurrentes sont les classes finales : si on y arrive on n'en sort plus) ;
- ou bien tous transitoires : la classe est dite *transitoire*.

Les états i et j sont dans la même classe ($i \rightleftharpoons j$) si et seulement si

$$\exists (m, n) \in (\mathbb{N}^*)^2, a_{i,j}^{(m)} > 0 \text{ et } a_{j,i}^{(n)} > 0.$$

Supposons que l'état i est récurrent : $\sum_{s=1}^{\infty} a_{i,i}^{(s)} = +\infty$. Alors :

$$\begin{aligned} \sum_{s=1}^{\infty} a_{j,j}^{(m+n+s)} &= \sum_{s=1}^{\infty} \sum_{k=1}^r a_{j,k}^{(m)} a_{k,j}^{(n+s)} = \sum_{s=1}^{\infty} \sum_{k=1}^r \sum_{l=1}^r a_{j,k}^{(m)} a_{k,l}^{(s)} a_{l,j}^{(n)} \\ &\geq \sum_{s=1}^{\infty} a_{j,i}^{(m)} a_{i,i}^{(s)} a_{i,j}^{(n)} = a_{j,i}^{(m)} a_{i,j}^{(n)} \sum_{s=1}^{\infty} a_{i,i}^{(s)} = +\infty \end{aligned}$$

Donc $\sum_{k=1}^{\infty} a_{j,j}^{(k)} = +\infty$: l'état j est récurrent.

Théorème 113. *Le nombre de classes récurrentes est égal à la multiplicité de 1 comme valeur propre de A .*

Voir [10], page 84.

Probabilités d'absorption par une classe récurrente

Soit C une classe récurrente et T l'ensemble des états transitoires. Soit $i \in T$; on note $x_i^{(n)}(C)$ la probabilité d'être absorbé par la classe C en n transitions à partir de l'état transitoire i :

$$x_i^{(1)}(C) = \sum_{k \in C} a_{i,k} \quad \text{et} \quad x_i^{(n+1)}(C) = \sum_{j \in T} a_{i,j} x_j^{(n)}(C).$$

La probabilité d'absorption par la classe C à partir de l'état i est :

$$\begin{aligned} x_i(C) &= \sum_{n=1}^{\infty} x_i^{(n)}(C) = x_i^{(1)}(C) + \sum_{n=2}^{\infty} x_i^{(n)}(C) = x_i^{(1)}(C) + \sum_{n=1}^{\infty} x_i^{(n+1)}(C) \\ &= \sum_{k \in C} a_{i,k} + \sum_{n=1}^{\infty} \sum_{j \in T} a_{i,j} x_j^{(n)}(C) = \sum_{k \in C} a_{i,k} + \sum_{j \in T} a_{i,j} \underbrace{\sum_{n=1}^{\infty} x_j^{(n)}(C)}_{=x_j(C)} \\ &= \sum_{k \in C} a_{i,k} + \sum_{j \in T} a_{i,j} x_j(C). \end{aligned}$$

Notant $r_i(C) = \sum_{k \in C} a_{i,k}$ la probabilité d'absorption par la classe C en une transition à partir de l'état transitoire i , on a :

$$x_i(C) - \sum_{j \in T} a_{i,j} x_j(C) = r_i(C).$$

Quitte à renuméroter les états si besoin on note $T = \{1, 2, \dots, \nu\}$ et on introduit les matrices :

$$Q = [a_{i,j}]_{1 \leq i,j \leq \nu}, \quad X(C) = [x_1(C) \quad \dots \quad x_\nu(C)]^T \quad \text{et} \quad R(C) = [r_1(C) \quad \dots \quad r_\nu(C)]^T.$$

On a :

$$X(C) = (I - Q)^{-1} R(C).$$

Exemple 5.4.5. Barrières absorbantes

On reprend l'exemple 5.2.2 en fixant des valeurs numériques pour les probabilités de transition :

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 2/3 & 0 & 1/3 & 0 & 0 \\ 0 & 2/3 & 0 & 1/3 & 0 \\ 0 & 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Le spectre de A est $\{0, -2/3, 2/3, 1\}$, la valeur propre 1 étant double : il y a donc deux classes récurrentes (théorème 113), les classes $C_1 = \{1\}$ et $C_2 = \{5\}$; l'ensemble des états transitoires est $T = \{2, 3, 4\}$. On peut vérifier ces résultats en appliquant le théorème 111; le calcul de A^n , aisé, quoique fastidieux (mais on peut utiliser Mathematica), donne

$$a_{1,1}^{(n)} = a_{5,5}^{(n)} = 1, \quad a_{2,2}^{(n)} = a_{4,4}^{(n)} = \frac{1}{2} a_{3,3}^{(n)} = \frac{1}{4} \left(\frac{2}{3}\right)^n (1 + (-1)^n).$$

D'où :

$$1. \quad \sum_{n=1}^{\infty} a_{1,1}^{(n)} = \sum_{n=1}^{\infty} a_{5,5}^{(n)} = +\infty : \text{les états 1 et 5 sont récurrents};$$

$$2. \sum_{n=1}^{\infty} a_{2,2}^{(n)} = \sum_{n=1}^{\infty} a_{4,4}^{(n)} = \frac{1}{2} \sum_{n=1}^{\infty} a_{3,3}^{(n)} = \frac{2}{5} : \text{les états 2, 3 et 4 sont transitoires.}$$

Calculons les probabilités d'absorption par les classes récurrentes. On a

$$Q = \begin{pmatrix} 0 & 1/3 & 0 \\ 2/3 & 0 & 1/3 \\ 0 & 2/3 & 0 \end{pmatrix}, \quad \text{d'où} \quad (I - Q)^{-1} = \frac{1}{5} \begin{pmatrix} 7 & 3 & 1 \\ 6 & 9 & 3 \\ 4 & 6 & 7 \end{pmatrix}.$$

Par ailleurs :

$$R(C_1) = [2/3 \quad 0 \quad 0]^T \quad \text{et} \quad R(C_2) = [0 \quad 0 \quad 1/3]^T.$$

On en déduit immédiatement en appliquant la formule ci-dessus :

$$X(C_1) = \frac{1}{15} [14 \quad 12 \quad 8]^T \quad \text{et} \quad X(C_2) = \frac{1}{15} [1 \quad 3 \quad 7]^T.$$

On peut en déduire (mais ce résultat peut également être obtenu en passant à la limite dans A^n) :

$$A_{\infty} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 14/15 & 0 & 0 & 0 & 1/15 \\ 12/15 & 0 & 0 & 0 & 3/15 \\ 8/15 & 0 & 0 & 0 & 7/15 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Calculons les probabilités de premier retour.

On applique le théorème 110.

1. États 1 et 5 :

$$A_{1,1}(s) = 1 + \sum_{n=1}^{\infty} a_{1,1}^{(n)} s^n = 1 + \sum_{n=1}^{\infty} s^n = \frac{1}{1-s} \quad \text{pour} \quad |s| < 1.$$

D'où :

$$F_{1,1}(s) = \frac{A_{1,1}(s) - 1}{A_{1,1}(s)} = s = \sum_{n=1}^{\infty} f_{1,1}^{(n)} s^n$$

Donc

$$f_{1,1}^{(1)} = 1 \text{ et } f_{1,1}^{(n)} = 0 \text{ pour } n \geq 2,$$

ce qui est assez évident...

$f_{1,1}^* = \sum_{n=1}^{\infty} f_{1,1}^{(n)} = 1$, ce qui confirme, d'après la définition 45, que l'état 1 est récurrent. Le temps de récurrence moyen est (évidemment...) $\mu_{1,1} = 1$.

On a un résultat identique pour l'état 5.

2. États 2 et 4 :

$$A_{2,2}(s) = 1 + \sum_{k=1}^{\infty} \frac{1}{2} \left(\frac{2}{3}\right)^{2k} s^{2k} = \frac{1 - \frac{2s^2}{9}}{1 - \frac{4s^2}{9}} \quad \text{pour} \quad |s| < \frac{3}{2}.$$

D'où :

$$F_{2,2}(s) = \frac{2s^2}{9-2s^2} = \sum_{k=1}^{\infty} \left(\frac{2}{9}\right)^k s^{2k}.$$

Donc

$$f_{2,2}^{(2k)} = \left(\frac{2}{9}\right)^k \text{ et } f_{2,2}^{(2k-1)} = 0 \text{ pour } k \in \mathbb{N}^*.$$

$f_{2,2}^* = \sum_{n=1}^{\infty} f_{2,2}^{(n)} = \frac{2}{7}$, ce qui confirme que l'état 2 est transitoire.

On a un résultat identique pour l'état 4.

3. État 3 :

$$A_{3,3}(s) = 1 + \sum_{k=1}^{\infty} \left(\frac{2}{3}\right)^{2k} s^{2k} = \frac{1}{1 - \frac{4s^2}{9}} \quad \text{pour } |s| < \frac{3}{2}.$$

D'où :

$$F_{3,3}(s) = \frac{4s^2}{9}.$$

Donc

$$f_{3,3}^{(2)} = \frac{4}{9} \text{ et } f_{3,3}^{(n)} = 0 \text{ pour } n \neq 2,$$

ce qui est assez évident...

$f_{3,3}^* = \sum_{n=1}^{\infty} f_{3,3}^{(n)} = \frac{4}{9}$, ce qui confirme que l'état 3 est transitoire.

Temps total de séjour dans chaque état transitoire

Soit $N_{i,j}$ le temps total de séjour dans l'état transitoire j en partant de l'état transitoire i . Posons :

$$r_{i,j}^{(k)} = \mathbb{P}(N_{i,j} = k), \quad m_{i,j} = \mathbf{E}(N_{i,j}) \quad \text{et} \quad M = (m_{i,j})_{1 \leq i,j \leq \nu}.$$

On a, en notant R l'ensemble des états récurrents et T celui des états transitoires :

- Si $i \neq j$:

$$r_{i,j}^{(k)} = \sum_{s \in T} a_{i,s} r_{s,j}^{(k)}.$$

Donc :

$$m_{i,j} = \sum_{k=0}^{\infty} k r_{i,j}^{(k)} = \sum_{k=0}^{\infty} k \sum_{s \in T} a_{i,s} r_{s,j}^{(k)} = \sum_{s \in T} a_{i,s} \sum_{k=0}^{\infty} k r_{s,j}^{(k)} = \sum_{s \in T} a_{i,s} m_{s,j}.$$

- Si $i = j$:

$$\begin{aligned} r_{i,i}^{(0)} &= 0, \\ r_{i,i}^{(1)} &= \sum_{s \in R} a_{i,s} + \sum_{s \in T} a_{i,s} r_{s,i}^{(0)}, \\ r_{i,i}^{(k)} &= \sum_{s \in T} a_{i,s} r_{s,i}^{(k-1)} \text{ pour } k \geq 2. \end{aligned}$$

Donc :

$$\begin{aligned}
 m_{i,i} &= \sum_{k=0}^{\infty} k r_{i,i}^{(k)} = \sum_{s \in R} a_{i,s} + \sum_{s \in T} a_{i,s} r_{s,i}^{(0)} + \sum_{k=2}^{\infty} k \sum_{s \in T} a_{i,s} r_{s,i}^{(k-1)} \\
 &= \sum_{s \in R} a_{i,s} + \sum_{s \in T} a_{i,s} r_{s,i}^{(0)} + \sum_{s \in T} a_{i,s} \sum_{k=2}^{\infty} k r_{s,i}^{(k-1)} \\
 &= \sum_{s \in R} a_{i,s} + \sum_{s \in T} a_{i,s} r_{s,i}^{(0)} + \sum_{s \in T} a_{i,s} \sum_{l=1}^{\infty} (l+1) r_{s,i}^{(l)} \\
 &= \sum_{s \in R} a_{i,s} + \underbrace{\sum_{s \in T} a_{i,s} r_{s,i}^{(0)} + \sum_{s \in T} a_{i,s} \sum_{k=1}^{\infty} r_{s,i}^{(k)}}_{= \sum_{s \in T} a_{i,s} \sum_{k=0}^{\infty} r_{s,i}^{(k)} = 1} + \sum_{s \in T} a_{i,s} \underbrace{\sum_{k=1}^{\infty} k r_{s,i}^{(k)}}_{= m_{s,i}} \\
 &= \underbrace{\sum_{s \in R} a_{i,s} + \sum_{s \in T} a_{i,s}}_{=1} + \sum_{s \in T} a_{i,s} m_{s,i} \\
 &= 1 + \sum_{s \in T} a_{i,s} m_{s,i}.
 \end{aligned}$$

Ces résultats peuvent s'écrire matriciellement $M = I + Q M$, ce qui donne :

$$M = (I - Q)^{-1}.$$

Exemple 5.4.6. Barrières absorbantes

On a vu dans l'exemple 5.4.5 :

$$M = (I - Q)^{-1} = \frac{1}{5} \begin{pmatrix} 7 & 3 & 1 \\ 6 & 9 & 3 \\ 4 & 6 & 7 \end{pmatrix}.$$

D'où, par exemple : $\mathbf{E}(N_{1,2}) = \frac{3}{5}$.

Chapitre 6

Estimation

6.1 Introduction

Soit X une variable aléatoire dont la densité de probabilité $f(x, \theta)$ dépend d'un paramètre θ appartenant à \mathbb{R}^k .

Au vu d'un échantillon de X il s'agit de déterminer au mieux la « vraie » valeur θ_0 de θ . On peut utiliser deux méthodes :

- Estimation ponctuelle : on calcule une valeur vraisemblable $\hat{\theta}$ de θ_0 .
- Estimation par intervalle : on cherche un intervalle dans lequel θ_0 se trouve avec une probabilité élevée.

6.2 Estimation ponctuelle

6.2.1 Définitions

Soit X une variable aléatoire de densité $f(x, \theta)$, θ appartenant à un intervalle I de \mathbb{R} .

Définition 46. Un n -échantillon de X est un n -uplet (X_1, X_2, \dots, X_n) tel que les X_k ont la même loi que X et sont indépendantes.

Définition 47. Une réalisation de l'échantillon est un n -uplet (x_1, x_2, \dots, x_n) où les x_k sont des nombres réels, résultats d'expériences.

Définition 48. Une statistique de l'échantillon est une variable aléatoire $\varphi(X_1, X_2, \dots, X_n)$ où φ est une application de \mathbb{R}^n dans \mathbb{R} .

Définition 49. Un estimateur T de θ est une statistique à valeurs dans I .

Exemple : $T_n = \frac{1}{n} \sum_{k=1}^n X_k$ est un estimateur de l'espérance mathématique de X .

Définition 50. Le biais de l'estimateur est $(E(T) - \theta_0)$. S'il est nul T est un estimateur sans biais.

Définition 51. L'estimateur est asymptotiquement sans biais si

$$\lim_{n \rightarrow \infty} E(T_n) = \theta_0.$$

Définition 52. L'estimateur est convergent si la suite (T_n) converge en probabilité vers θ_0 :

$$\forall \varepsilon > 0, \mathbb{P}(|T_n - \theta_0| > \varepsilon) \xrightarrow{n \rightarrow \infty} 0$$

Théorème 114. Un estimateur convergent est asymptotiquement sans biais. Pour qu'un estimateur asymptotiquement sans biais soit convergent il suffit que $\text{Var}(T_n) \xrightarrow{n \rightarrow \infty} 0$.

Définition 53. Soient T_1 et T_2 deux estimateurs de θ . On dira que T_1 est plus efficace que T_2 si $\mathbb{E}((T_1 - \theta_0)^2) \leq \mathbb{E}((T_2 - \theta_0)^2)$.

Si T_1 et T_2 sont des estimateurs sans biais cette condition s'écrit : $\text{Var}(T_1) \leq \text{Var}(T_2)$.

Définition 54. Une estimation est la valeur de l'estimateur correspondant à une réalisation de l'échantillon.

6.2.2 Exemples fondamentaux

Soit X une variable aléatoire telle que $\mathbb{E}(X) = m$ et $\text{Var}(X) = \sigma^2$.

Estimation de m

Théorème 115. La moyenne empirique $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ est un estimateur sans biais et convergent de m .

On a

$$\mathbb{E}(\bar{X}_n) = \frac{1}{n} \sum_{k=1}^n \mathbb{E}(X_k) = m \quad \text{et} \quad \text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{k=1}^n \text{Var}(X_k) = \frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0.$$

Estimation de σ^2 en supposant m connu

Théorème 116. Lorsque m est connu

$$\overline{S_n''^2} = \frac{1}{n} \sum_{k=1}^n (X_k - m)^2$$

est un estimateur sans biais et convergent de σ^2 .

1. C'est un estimateur sans biais de σ^2 :

$$\begin{aligned} \mathbb{E}(\overline{S_n''^2}) &= \frac{1}{n} \mathbb{E}\left(\sum_{k=1}^n X_k^2 - 2m \sum_{k=1}^n X_k + nm^2\right) = \frac{1}{n} (n(\sigma^2 + m^2) - 2m(nm) + nm^2) \\ &= \sigma^2. \end{aligned}$$

2. Par ailleurs, les variables $(X_k - m)^2$ étant indépendantes :

$$\begin{aligned}\text{Var}(\overline{S_n''^2}) &= \frac{1}{n^2} \sum_{k=1}^n \text{Var}((X_k - m)^2) = \frac{1}{n} \left(\mathbf{E}((X - m)^4) - \mathbf{E}((X - m)^2)^2 \right) \\ &= \frac{1}{n} (\mu_4 - \mu_2^2) \quad \text{où} \quad \mu_k = \mathbf{E}((X - \mathbf{E}(X))^k).\end{aligned}$$

Donc $\overline{S_n''^2}$ est un estimateur convergent.

Estimation de σ^2 (m supposé inconnu)

En pratique m est rarement connu exactement ; on le remplace par un estimateur et on introduit la *variance empirique* :

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

Théorème 117. *La variance empirique $\overline{S_n^2}$ est un estimateur biaisé de σ^2 , asymptotiquement sans biais ; il est convergent.*

1. $\overline{S_n^2}$ est un estimateur biaisé de σ^2 :

$$\begin{aligned}\overline{S_n^2} &= \frac{1}{n} \left(\sum_{k=1}^n X_k^2 - 2\overline{X}_n \sum_{k=1}^n X_k + n(\overline{X}_n)^2 \right) = \frac{1}{n} \sum_{k=1}^n X_k^2 - (\overline{X}_n)^2 \\ \mathbf{E}(\overline{S_n^2}) &= \frac{1}{n} \sum_{k=1}^n \mathbf{E}(X_k^2) - \mathbf{E}((\overline{X}_n)^2) = \frac{1}{n} (n(m^2 + \sigma^2)) - (m^2 + \frac{\sigma^2}{n}) \\ &= \frac{n-1}{n} \sigma^2 \neq \sigma^2.\end{aligned}$$

2. On montre (c'est fastidieux...) :

$$\text{Var}(\overline{S_n^2}) = \frac{1}{n} (\mu_4 - \mu_2^2) - \frac{2}{n^2} (\mu_4 - 2\mu_2^2) + \frac{1}{n^3} (\mu_4 - 3\mu_2^2) \sim \frac{1}{n} (\mu_4 - \mu_2^2) \xrightarrow{n \rightarrow \infty} 0.$$

Donc $\overline{S_n^2}$ est un estimateur convergent.

Théorème 118. *Un estimateur sans biais et convergent de σ^2 est la variance empirique corrigée :*

$$\overline{S_n'^2} = \frac{1}{n-1} \sum_{k=1}^n (X_k - \overline{X}_n)^2.$$

Puisque $\overline{S_n'^2} = \frac{n}{n-1} \overline{S_n^2}$ on a :

$$\mathbf{E}(\overline{S_n'^2}) = \frac{n}{n-1} \mathbf{E}(\overline{S_n^2}) = \sigma^2 \quad \text{et} \quad \text{Var}(\overline{S_n'^2}) = \left(\frac{n}{n-1} \right)^2 \text{Var}(\overline{S_n^2}) \xrightarrow{n \rightarrow \infty} 0.$$

Remarque : on ne saurait déduire de ce théorème que $\overline{S_n'^2}$ est un estimateur sans biais de σ .

Cas particulier de la loi normale

On suppose dans ce paragraphe que X suit la loi normale $\mathcal{N}(m, \sigma)$.

On sait que $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ suit la loi normale $\mathcal{N}(m, \frac{\sigma}{\sqrt{n}})$; c'est donc un estimateur sans biais et convergent de m .

Les résultats obtenus au paragraphe précédent pour l'estimation de σ^2 sont encore valables; on peut préciser en outre :

$$\text{Var}(\bar{S}_n^2) = \frac{2(n-1)}{n^2} \mu_2^2.$$

1. Calcul de μ_k .

$$\begin{aligned} \mu_k &= \mathbf{E}((X-m)^k) = \int_{-\infty}^{\infty} (x-m)^k \frac{\exp(-\frac{(x-m)^2}{2\sigma^2})}{\sigma\sqrt{2\pi}} dx \\ &= \int_{-\infty}^{\infty} \frac{(\sigma\sqrt{2}u)^k}{\sigma\sqrt{2\pi}} e^{-u^2} (\sigma\sqrt{2} du) \text{ en posant } x = m + u\sigma\sqrt{2} \\ &= 0 \text{ si } k \text{ est impair.} \end{aligned}$$

Si k est pair :

$$\begin{aligned} \mu_{2p} &= \frac{2^p \sigma^{2p}}{\sqrt{\pi}} \int_{-\infty}^{\infty} u^{2p} e^{-u^2} du = \frac{2^{p+1} \sigma^{2p}}{\sqrt{\pi}} \int_0^{\infty} u^{2p} e^{-u^2} du \\ &= \frac{2^p \sigma^{2p}}{\sqrt{\pi}} \int_0^{\infty} v^{p-\frac{1}{2}} e^{-v} dv \text{ en posant } u = \sqrt{v}, \\ &= \frac{2^p \sigma^{2p}}{\sqrt{\pi}} \Gamma(p + \frac{1}{2}) = \frac{(2p)!}{2^p (p!)} \sigma^{2p}. \end{aligned}$$

2. Lorsque X suit la loi normale $\mathcal{N}(m, \sigma)$ on a $\mu_2 = \sigma^2$ et $\mu_4 = 3\sigma^4 = 3\mu_2^2$. En reportant :

$$\text{Var}(\bar{S}_n^2) = \frac{1}{n}(\mu_4 - \mu_2^2) - \frac{2}{n^2}(\mu_4 - 2\mu_2^2) + \frac{1}{n^3}(\mu_4 - 3\mu_2^2) = \frac{2(n-1)}{n^2} \mu_2^2.$$

Théorème 119. Soit (X_1, X_2, \dots, X_n) un échantillon d'une variable normale réduite; $n(\bar{X}_n)^2$ suit une loi du χ^2 à un degré de liberté.

On a : $n(\bar{X}_n)^2 = \left(\frac{X_1 + X_2 + \dots + X_n}{\sqrt{n}} \right)^2$. Or les X_k suivent la loi normale réduite et sont indépendantes; donc leur somme suit la loi $\mathcal{N}(0, \sqrt{n})$ et $\frac{1}{\sqrt{n}} \sum_{k=1}^n X_k$ suit la loi normale réduite.

Théorème 120 (FISHER). Soit (X_1, X_2, \dots, X_n) un échantillon d'une variable normale réduite; les variables aléatoires

$$\sqrt{n} \bar{X}_n \quad \text{et} \quad \sum_{k=1}^n (X_k - \bar{X}_n)^2 = n\bar{S}_n^2 = (n-1)\bar{S}_n'^2$$

sont indépendantes et suivent respectivement la loi normale réduite et la loi du χ^2 à $(n - 1)$ degrés de liberté.

1. Montrons que \overline{X}_n et $\sum_{k=1}^n (X_k - \overline{X}_n)^2$ sont indépendantes.

On a

$$\begin{bmatrix} \overline{X}_n \\ X_1 - \overline{X}_n \\ \vdots \\ X_n - \overline{X}_n \end{bmatrix} = a \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \quad \text{où : } a = \begin{pmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ \frac{n-1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & \frac{n-1}{n} & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\frac{1}{n} \\ -\frac{1}{n} & \cdots & -\frac{1}{n} & \frac{n-1}{n} \end{pmatrix}.$$

Le vecteur aléatoire $[X_1 \ X_2 \ \dots \ X_n]^T$ est gaussien de loi $\mathcal{N}(0, I)$; d'après le corollaire 3 le vecteur $[\overline{X}_n \ X_1 - \overline{X}_n \ \dots \ X_n - \overline{X}_n]^T$ est également gaussien, de loi $\mathcal{N}(0, a a^T)$. Or :

$$a a^T = \begin{pmatrix} \frac{1}{n} & 0 & 0 & \cdots & 0 \\ 0 & \frac{n-1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ 0 & -\frac{1}{n} & \frac{n-1}{n} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & -\frac{1}{n} \\ 0 & -\frac{1}{n} & \cdots & -\frac{1}{n} & \frac{n-1}{n} \end{pmatrix}$$

La variable aléatoire \overline{X}_n est indépendante du vecteur aléatoire $[X_1 - \overline{X}_n \ X_2 - \overline{X}_n \ \dots \ X_n - \overline{X}_n]^T$: elle est donc indépendante de $\sum_{k=1}^n (X_k - \overline{X}_n)^2$.

2. On sait que \overline{X}_n suit la loi normale $\mathcal{N}(0, 1/\sqrt{n})$. Donc $\sqrt{n}\overline{X}_n$ suit la loi normale réduite.
3. Montrons que $n\overline{S}_n^2 = \sum_{k=1}^n (X_k - \overline{X}_n)^2$ suit la loi du χ^2 à $(n - 1)$ degrés de liberté.

$$\underbrace{[X_1 - \overline{X}_n \ X_2 - \overline{X}_n \ \dots \ X_n - \overline{X}_n]^T}_{=Y} = b \underbrace{[X_1 \ X_2 \ \dots \ X_n]^T}_{=X}$$

où b est la matrice obtenue en supprimant la première ligne de a .

b est symétrique : il existe une matrice orthogonale u telle que $b = u d u^T$ où d est la matrice diagonale des valeurs propres de b . Or celles-ci sont :

- la valeur propre simple égale à 0 dont le sous-espace propre associé a pour équation $x_1 = x_2 = \cdots = x_n$;
- la valeur propre d'ordre $(n - 1)$ égale à 1 dont le sous-espace propre a pour équation $x_1 + x_2 + \cdots + x_n = 0$.

En ordonnant convenablement la base de vecteurs propres on peut choisir :

$$d = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 \\ 0 & \cdots & 0 & 0 \end{pmatrix}.$$

On a : $Y = bX = u d u^T X$ et

$$\sum_{k=1}^n (X_k - \bar{X}_n)^2 = Y^T Y = X^T \underbrace{u d \underbrace{u^T u}_{=I} d}_{=d} u^T X = (u^T X)^T d (u^T X).$$

Or le vecteur aléatoire $Z = u^T X$ est gaussien de loi $\mathcal{N}(0, \underbrace{u^T I u}_{=I})$. Donc

$$\sum_{k=1}^n (X_k - \bar{X}_n)^2 = Z^T d Z = \sum_{i=1}^{n-1} Z_i^2$$

suit la loi du khi-deux à $(n - 1)$ degrés de liberté.

On en déduit immédiatement que si (X_1, X_2, \dots, X_n) est un échantillon d'une variable normale de loi $\mathcal{N}(m, \sigma)$ la variable aléatoire

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

suit la loi du χ^2 à $(n - 1)$ degrés de liberté.

On pose : $Y_k = \frac{X_k - m}{\sigma}$. Les variables Y_k sont indépendantes et suivent la loi normale réduite. De $X_k = \sigma Y_k + m$ on déduit immédiatement $\bar{X}_n = \sigma \bar{Y}_n + m$; donc

$$\frac{1}{\sigma^2} \sum_{k=1}^n (X_k - \bar{X}_n)^2 = \sum_{k=1}^n (Y_k - \bar{Y}_n)^2$$

suit la loi du khi-deux à $(n - 1)$ degrés de liberté.

6.2.3 Maximum de vraisemblance

La méthode du *maximum de vraisemblance* consiste à choisir comme estimateurs des paramètres inconnus les valeurs qui rendent maximum la probabilité d'avoir obtenu l'échantillon. Soit X une variable aléatoire dont la densité $f(x, \theta)$ dépend d'un paramètre $\theta = (\theta_1, \theta_2, \dots, \theta_k) \in \Theta \subset \mathbb{R}^k$ et soit (x_1, x_2, \dots, x_n) une réalisation de l'échantillon (X_1, X_2, \dots, X_n) ; la *vraisemblance* est la fonction

$$\theta \rightarrow \mathcal{L}(\theta) = f(x_1, \theta) f(x_2, \theta) \cdots f(x_n, \theta).$$

Dans le cas d'une loi discrète c'est la fonction

$$\theta \rightarrow \mathcal{L}(\theta) = p(x_1, \theta) p(x_2, \theta) \cdots p(x_n, \theta)$$

où $p(x, \theta) = \mathbb{P}(X = x)$.

Un estimateur du maximum de vraisemblance $\hat{\theta}$ vérifie donc

$$\forall \theta \in \Theta, \mathcal{L}(\theta) \leq \mathcal{L}(\hat{\theta}).$$

Remarques

1. Un estimateur du maximum de vraisemblance, s'il existe, n'est pas nécessairement unique.
2. Un estimateur du maximum de vraisemblance peut être biaisé.

Lorsque $f > 0$ il est équivalent, et généralement plus facile, de chercher le maximum de $\ln(\mathcal{L}(\theta))$. Si \mathcal{L} n'est pas différentiable la recherche du maximum peut être délicate ; cependant il est fréquent que l'on puisse appliquer le théorème suivant.

Théorème 121. Si $f > 0$ et si $\frac{\partial f}{\partial \theta}$ et $\frac{\partial^2 f}{\partial \theta^2}$ sont continues l'estimateur $\hat{\theta}$ du maximum de vraisemblance vérifie

1. les équations de vraisemblance

$$\forall i \in \{1, 2, \dots, k\}, \quad \frac{\partial \ln \mathcal{L}}{\partial \theta_i}(\hat{\theta}) = 0,$$

2. la condition suffisante de maximum : la matrice hessienne de $\ln \mathcal{L}$, $\left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta_i \partial \theta_j}(\hat{\theta}) \right)_{1 \leq i, j \leq k}$, est définie négative.

Dans le cas où le paramètre θ est scalaire les conditions s'écrivent simplement

$$\frac{d(\ln \mathcal{L})}{d\theta}(\hat{\theta}) = 0, \quad \frac{d^2(\ln \mathcal{L})}{d\theta^2}(\hat{\theta}) < 0.$$

6.3 Estimation par intervalle

6.3.1 Définitions

Définition 55. Soit X une variable aléatoire dont la loi dépend d'un paramètre θ .

T_1 et T_2 étant deux statistiques définies sur un n -échantillon de X on dit que l'intervalle $[T_1, T_2]$ est un intervalle de confiance au niveau α pour θ si :

$$\mathbb{P}(T_1 \leq \theta \leq T_2) = \alpha.$$

La probabilité d'erreur, c'est-à-dire que θ ne se trouve pas dans l'intervalle $[T_1, T_2]$, est égale à $(1 - \alpha)$.

Définition 56. Si $\mathbb{P}(\theta < T_1) = \mathbb{P}(\theta > T_2) = \frac{1 - \alpha}{2}$ on dit que l'intervalle $[T_1, T_2]$ est symétrique en probabilité.

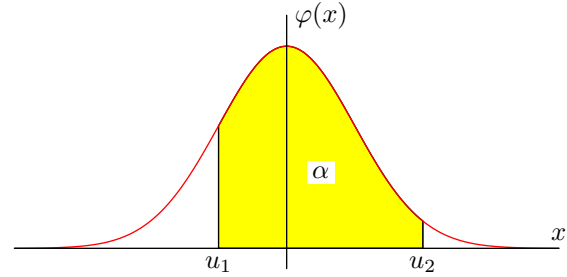
6.3.2 Exemples fondamentaux

Estimation de m pour la loi normale $\mathcal{N}(m, \sigma)$ avec σ connu

Théorème 122. Lorsque σ est connu un intervalle de confiance au niveau α de m est

$$\left[\bar{X}_n - u_2 \frac{\sigma}{\sqrt{n}}, \bar{X}_n - u_1 \frac{\sigma}{\sqrt{n}} \right]$$

où u_1 et u_2 vérifient $\Phi(u_2) - \Phi(u_1) = \alpha$ (voir table à la page 150).



Un estimateur ponctuel de m est \bar{X}_n qui suit la loi normale $\mathcal{N}(m, \frac{\sigma}{\sqrt{n}})$. On cherche l'intervalle de confiance sous la forme $[\bar{X}_n - a, \bar{X}_n + b]$. On veut obtenir :

$$\begin{aligned} \alpha &= \mathbb{P}(\bar{X}_n - a \leq m \leq \bar{X}_n + b) = \mathbb{P}(-b \leq \bar{X}_n - m \leq a) \\ &= \mathbb{P}\left(\frac{-b\sqrt{n}}{\sigma} \leq \frac{\bar{X}_n - m}{\sigma/\sqrt{n}} \leq \frac{a\sqrt{n}}{\sigma}\right) = \Phi\left(\frac{a\sqrt{n}}{\sigma}\right) - \Phi\left(\frac{-b\sqrt{n}}{\sigma}\right) \end{aligned}$$

où Φ représente la fonction de répartition de la loi normale réduite. On choisit u_1 et u_2 tels que

$$\Phi(u_2) - \Phi(u_1) = \alpha.$$

De

$$\frac{a\sqrt{n}}{\sigma} = u_2 \quad \text{et} \quad \frac{-b\sqrt{n}}{\sigma} = u_1$$

on déduit un intervalle de confiance au niveau α :

$$\left[\bar{X}_n - u_2 \frac{\sigma}{\sqrt{n}}, \bar{X}_n - u_1 \frac{\sigma}{\sqrt{n}} \right].$$

Notons z_α le fractile d'ordre α de la loi normale réduite, c'est-à-dire le nombre qui vérifie :

$$\Phi(z_\alpha) = \alpha.$$

L'intervalle de confiance symétrique en probabilité au niveau α de m est :

$$\left[\bar{X}_n - z_{\frac{1+\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{\frac{1+\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right].$$

Exemple : puisque $z_{0,975} = 1.96$ l'intervalle de confiance de m au niveau 95%, symétrique en probabilité, est

$$\left[\bar{X}_n - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{X}_n + 1.96 \frac{\sigma}{\sqrt{n}} \right],$$

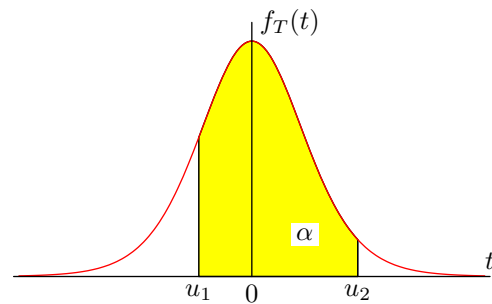
ce qui signifie que sur un grand nombre d'expériences cet intervalle contiendra effectivement m dans 95% des cas en moyenne.

Estimation de m pour la loi normale $\mathcal{N}(m, \sigma)$ avec σ inconnu

Théorème 123. Lorsque σ est inconnu un intervalle de confiance au niveau α de m est

$$\left[\bar{X}_n - u_2 \frac{\bar{S}'_n}{\sqrt{n}}, \bar{X}_n - u_1 \frac{\bar{S}'_n}{\sqrt{n}} \right]$$

où u_1 et u_2 vérifient $\mathbb{P}(u_1 \leq T \leq u_2) = \alpha$, T suivant la loi de STUDENT à $(n - 1)$ degrés de liberté (voir table à la page 154).



On estime σ par \bar{S}'_n tel que $\bar{S}'_n{}^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$. On veut obtenir :

$$\alpha = \mathbb{P}(\bar{X}_n - a \leq m \leq \bar{X}_n + b) = \mathbb{P}(-b \leq \bar{X}_n - m \leq a) = \mathbb{P}\left(\frac{-b\sqrt{n}}{\bar{S}'_n} \leq \frac{\bar{X}_n - m}{\bar{S}'_n/\sqrt{n}} \leq \frac{a\sqrt{n}}{\bar{S}'_n}\right)$$

Posons

$$T = \frac{\bar{X}_n - m}{\bar{S}'_n/\sqrt{n}} = \frac{\bar{X}_n - m}{\sigma/\sqrt{n}} \bigg/ \sqrt{\frac{\bar{S}'_n{}^2}{\sigma^2}} \quad \text{et} \quad \frac{\bar{S}'_n{}^2}{\sigma^2} = \frac{U}{n-1}.$$

Alors, d'après le théorème 120 (page 90), $U = \frac{1}{\sigma^2} \sum_{k=1}^n (X_k - \bar{X}_n)^2$ suit la loi du χ^2 à $(n - 1)$ degrés de liberté. Donc T suit la loi de STUDENT à $(n - 1)$ degrés de liberté (voir § 3.6.7, page 60).

On détermine dans la table (page 154) les valeurs de u_1 et u_2 telles que : $\mathbb{P}(u_1 \leq T \leq u_2) = \alpha$. L'intervalle de confiance est alors :

$$\left[\bar{X}_n - u_2 \frac{\bar{S}'_n}{\sqrt{n}}, \bar{X}_n - u_1 \frac{\bar{S}'_n}{\sqrt{n}} \right].$$

Notons $t_{n-1, \alpha}$ le fractile d'ordre α de la loi de STUDENT à $(n - 1)$ degrés de liberté, c'est-à-dire le nombre qui vérifie

$$F_T(t_{n-1, \alpha}) = \alpha,$$

où T suit la loi de STUDENT à $(n - 1)$ degrés de liberté.

L'intervalle de confiance symétrique en probabilité au niveau α de m est :

$$\left[\bar{X}_n - t_{n-1, \frac{1+\alpha}{2}} \frac{\bar{S}'_n}{\sqrt{n}}, \bar{X}_n + t_{n-1, \frac{1+\alpha}{2}} \frac{\bar{S}'_n}{\sqrt{n}} \right].$$

Exemple : pour $n = 10$, avec un niveau de confiance de 95% et un intervalle symétrique on obtient l'intervalle

$$\left[\bar{X}_n - 2.26 \frac{\bar{S}'_n}{\sqrt{n}}, \bar{X}_n + 2.26 \frac{\bar{S}'_n}{\sqrt{n}} \right].$$

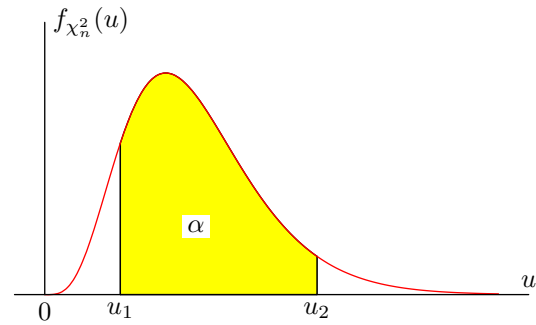
Remarque : l'intervalle de confiance est plus grand que celui obtenu lorsqu'on connaît σ .

Estimation de σ^2 pour la loi normale $\mathcal{N}(m, \sigma)$ avec m connu

Théorème 124. Lorsque m est connu un intervalle de confiance au niveau α de σ^2 est

$$\left[\frac{1}{u_2} \sum_{k=1}^n (X_k - m)^2, \frac{1}{u_1} \sum_{k=1}^n (X_k - m)^2 \right]$$

où u_1 et u_2 vérifient $\mathbb{P}(u_1 \leq \chi_n^2 \leq u_2) = \alpha$, χ_n^2 suivant la loi du khi-deux à n degrés de liberté (voir table à la page 152).



Si X_k suit la loi normale $\mathcal{N}(m, \sigma)$ alors $\frac{X_k - m}{\sigma}$ suit la loi normale réduite et donc $\chi_n^2 = \sum_{k=1}^n \left(\frac{X_k - m}{\sigma} \right)^2$

suit la loi du χ^2 à n degrés de liberté. On cherche dans les tables (page 152) les valeurs de u_1 et u_2 telles que

$$\mathbb{P}(u_1 \leq \chi_n^2 \leq u_2) = \alpha.$$

De

$$\alpha = \mathbb{P}(u_1 \leq \frac{1}{\sigma^2} \sum_{k=1}^n (X_k - m)^2 \leq u_2) = \mathbb{P}(\frac{1}{u_2} \sum_{k=1}^n (X_k - m)^2 \leq \sigma^2 \leq \frac{1}{u_1} \sum_{k=1}^n (X_k - m)^2)$$

on déduit l'intervalle de confiance au niveau α pour σ^2 :

$$\left[\frac{1}{u_2} \sum_{k=1}^n (X_k - m)^2, \frac{1}{u_1} \sum_{k=1}^n (X_k - m)^2 \right].$$

En pratique on choisira généralement un intervalle symétrique en probabilité c'est-à-dire tel que

$$\mathbb{P}(\chi_n^2 < u_1) = \frac{1 - \alpha}{2} \quad \text{et} \quad \mathbb{P}(\chi_n^2 > u_2) = \frac{1 - \alpha}{2}.$$

En notant $u_{n,\alpha}$ le fractile d'ordre α de la loi du khi-deux à n degrés de liberté, c'est-à-dire le nombre qui vérifie

$$F_{\chi_n^2}(u_{n,\alpha}) = \alpha,$$

l'intervalle de confiance symétrique en probabilité au niveau α de σ^2 est :

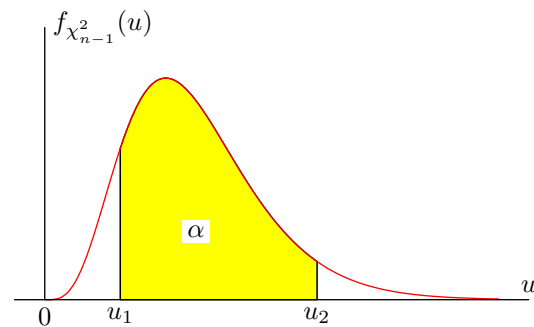
$$\left[\frac{1}{u_{n, \frac{1+\alpha}{2}}} \sum_{k=1}^n (X_k - m)^2, \frac{1}{u_{n, \frac{1-\alpha}{2}}} \sum_{k=1}^n (X_k - m)^2 \right].$$

Estimation de σ^2 pour la loi normale $\mathcal{N}(m, \sigma)$ avec m inconnu

Théorème 125. Lorsque m est inconnu un intervalle de confiance au niveau α de σ^2 est

$$\left[\frac{1}{u_2} \sum_{k=1}^n (X_k - \bar{X}_n)^2, \frac{1}{u_1} \sum_{k=1}^n (X_k - \bar{X}_n)^2 \right]$$

où u_1 et u_2 vérifient $\mathbb{P}(u_1 \leq \chi_{n-1}^2 \leq u_2) = \alpha$, χ_{n-1}^2 suivant la loi du khi-deux à $(n-1)$ degrés de liberté (voir table à la page 152).



On estime m par \bar{X}_n . D'après le théorème 120 (page 90) la variable aléatoire $\sum_{k=1}^n \left(\frac{X_k - \bar{X}_n}{\sigma} \right)^2$ suit la loi du khi-deux à $(n-1)$ degrés de liberté.

Par le même raisonnement qu'au paragraphe précédent on obtient l'intervalle de confiance au niveau α pour σ^2 .

L'intervalle de confiance au niveau α de σ^2 symétrique en probabilité est :

$$\left[\frac{1}{u_{n-1, \frac{1+\alpha}{2}}} \sum_{k=1}^n (X_k - \bar{X}_n)^2, \frac{1}{u_{n-1, \frac{1-\alpha}{2}}} \sum_{k=1}^n (X_k - \bar{X}_n)^2 \right] = \left[\frac{(n-1)\bar{S}_n'^2}{u_{n-1, \frac{1+\alpha}{2}}}, \frac{(n-1)\bar{S}_n'^2}{u_{n-1, \frac{1-\alpha}{2}}} \right].$$

Remarque : un intervalle de confiance pour l'écart-type σ est obtenu à partir de celui de la variance en prenant les racines carrées des bornes.

6.4 Estimation d'une proportion

Dans la population étudiée la proportion d'individus ayant une certaine propriété est égale à p . Soit X le nombre d'individus d'un échantillon de taille n ayant la propriété.

6.4.1 Estimation ponctuelle

Théorème 126. Un estimateur sans biais et convergent de p est :

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

Le nombre X d'individus de l'échantillon ayant la propriété suit la loi binomiale $\mathcal{B}(n, p)$. On a :

$$\mathbf{E}(T) = \frac{1}{n} \mathbf{E}(X) = p \quad \text{et} \quad \mathbf{Var}(T) = \frac{1}{n^2} \mathbf{Var}(X) = \frac{p(1-p)}{n} \xrightarrow[n \rightarrow \infty]{} 0.$$

6.4.2 Estimation par intervalle

On ne sait pas déterminer exactement un intervalle de confiance. On utilise des solutions approchées.

Échantillon de grande taille

Lorsque n est grand et p voisin de 0,5 on peut approcher la loi binomiale par une loi normale (voir le théorème 96).

Théorème 127. *Un intervalle de confiance approché, symétrique en probabilité, de p au niveau α est*

$$\left[T - z_{\frac{1+\alpha}{2}} \sqrt{\frac{T(1-T)}{n}}, T + z_{\frac{1+\alpha}{2}} \sqrt{\frac{T(1-T)}{n}} \right].$$

(On rappelle que z_α représente le fractile d'ordre α de la loi normale réduite.)

$$\frac{X - np}{\sqrt{np(1-p)}} = \frac{T - p}{\sqrt{\frac{p(1-p)}{n}}} \text{ suit approximativement la loi normale réduite } \mathcal{N}(0, 1).$$

$[T - a, T + b]$ est un intervalle de confiance de p au niveau α si

$$\begin{aligned} \alpha &= \mathbb{P}(T - a < p < T + b) = \mathbb{P}\left(-\frac{b}{\sqrt{\frac{p(1-p)}{n}}} < \frac{T - p}{\sqrt{\frac{p(1-p)}{n}}} < \frac{a}{\sqrt{\frac{p(1-p)}{n}}}\right) \\ &\simeq \Phi\left(\frac{a}{\sqrt{\frac{p(1-p)}{n}}}\right) - \Phi\left(-\frac{b}{\sqrt{\frac{p(1-p)}{n}}}\right). \end{aligned}$$

En choisissant un intervalle symétrique en probabilité on détermine u tel que $2\Phi(u) - 1 = \alpha$, c'est-à-dire $u = z_{\frac{1+\alpha}{2}}$; on doit avoir

$$a = b = u \sqrt{\frac{p(1-p)}{n}}.$$

Puisque p est inconnu on n'a pas a et b ...

On obtient un intervalle de confiance *approché* en remplaçant p par l'estimateur T :

$$\left[T - z_{\frac{1+\alpha}{2}} \sqrt{\frac{T(1-T)}{n}}, T + z_{\frac{1+\alpha}{2}} \sqrt{\frac{T(1-T)}{n}} \right].$$

Théorème 128. *Un intervalle de confiance approché, symétrique en probabilité, de p au niveau α est*

$$\left[\frac{1}{1 + \frac{u^2}{n}} \left(T + \frac{u^2}{2n} - \frac{u}{\sqrt{n}} \sqrt{T(1-T) + \frac{u^2}{4n}} \right), \frac{1}{1 + \frac{u^2}{n}} \left(T + \frac{u^2}{2n} + \frac{u}{\sqrt{n}} \sqrt{T(1-T) + \frac{u^2}{4n}} \right) \right]$$

où $u = z_{\frac{1+\alpha}{2}}$.

Cette approximation de l'intervalle de confiance est meilleure que celle fournie par le théorème précédent.

On a :

$$\begin{aligned} -u < \frac{T - p}{\sqrt{\frac{p(1-p)}{n}}} < u &\Leftrightarrow (T - p)^2 < u^2 \frac{p(1-p)}{n} \\ &\Leftrightarrow p^2 \left(1 + \frac{u^2}{n} \right) - 2p \left(T + \frac{u^2}{2n} \right) + T^2 < 0. \end{aligned}$$

p doit donc être compris entre les racines du trinôme du second degré. On vérifie aisément que ce trinôme a deux racines réelles appartenant à l'intervalle $[0, 1]$. D'où l'intervalle de confiance indiqué.

Échantillon de petite taille

Lorsque n est faible la solution analytique est délicate : on utilise des abaques... ou un logiciel spécialisé.

6.5 Comparaison de moyennes et de variances

Soient $(X_1, X_2, \dots, X_{n_1})$ un échantillon d'une population suivant la loi normale $\mathcal{N}(m_1, \sigma_1)$ et $(Y_1, Y_2, \dots, Y_{n_2})$ un échantillon d'une population suivant la loi normale $\mathcal{N}(m_2, \sigma_2)$; ces deux échantillons sont supposés indépendants. Nous souhaitons comparer les moyennes, m_1 et m_2 , et les variances, σ_1^2 et σ_2^2 , à l'aide de ces échantillons. Pour cela nous allons construire des intervalles de confiance pour $m_1 - m_2$ et pour σ_1^2/σ_2^2 .

6.5.1 Intervalle de confiance de la différence de deux moyennes

Posons :

$$\bar{X} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i, \quad \bar{Y} = \frac{1}{n_2} \sum_{j=1}^{n_2} Y_j \quad \text{et} \quad D = \bar{X} - \bar{Y}.$$

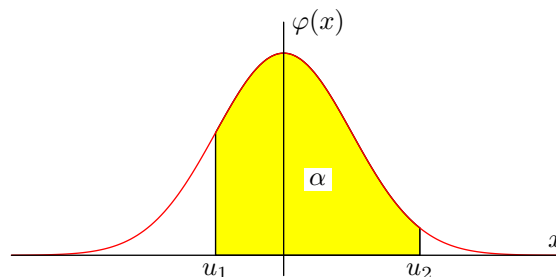
Théorème 129. D est un estimateur sans biais de $(m_1 - m_2)$.

\bar{X} suit la loi normale $\mathcal{N}(m_1, \sigma_1/\sqrt{n_1})$ et \bar{Y} suit la loi normale $\mathcal{N}(m_2, \sigma_2/\sqrt{n_2})$; elles sont indépendantes. D'après les théorèmes 64 et 65 D suit donc la loi normale $\mathcal{N}(m_1 - m_2, \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2})$.

Théorème 130. Si σ_1 et σ_2 sont connus un intervalle de confiance de $(m_1 - m_2)$ au niveau α est

$$\left[D - u_2 \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, D - u_1 \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$$

où u_1 et u_2 vérifient $\Phi(u_2) - \Phi(u_1) = \alpha$.



Un intervalle de confiance de $(m_1 - m_2)$ au niveau α est $[D - a, D + b]$ si :

$$\begin{aligned} \alpha &= \mathbb{P}(D - a \leq m_1 - m_2 \leq D + b) = \mathbb{P}(-b \leq D - (m_1 - m_2) \leq a) \\ &= \mathbb{P}\left(\frac{-b}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq D^* \leq \frac{a}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right) = \Phi(u_2) - \Phi(u_1) \end{aligned}$$

où l'on a posé

$$u_2 = \frac{a}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad \text{et} \quad u_1 = \frac{-b}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

On en déduit l'intervalle de confiance annoncé.

On choisit généralement un intervalle symétrique en probabilité :

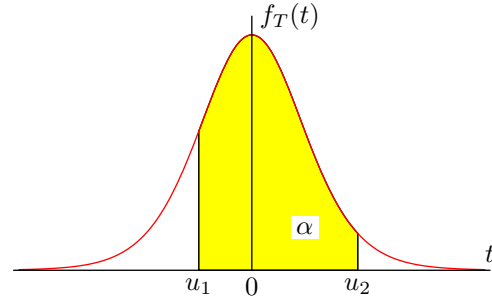
$$\left[D - z_{\frac{1+\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, D + z_{\frac{1+\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$$

Théorème 131. Si σ_1 et σ_2 sont inconnus mais égaux un intervalle de confiance de $(m_1 - m_2)$ au niveau α est

$$\left[D - u_2 S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, D - u_1 S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right].$$

où u_1 et u_2 vérifient $\mathbb{P}(u_1 \leq T \leq u_2) = \alpha$, T suivant la loi de STUDENT à $(n_1 + n_2 - 2)$ degrés de liberté (voir table à la page 154) et

$$S^2 = \frac{1}{n_1 + n_2 - 2} \left(\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{j=1}^{n_2} (Y_j - \bar{Y})^2 \right).$$



L'hypothèse d'égalité des variances peut être vérifiée en appliquant le théorème 134.

On pourrait remplacer σ_1 et σ_2 par les estimateurs S_1 et S_2 définis par :

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2 \quad \text{et} \quad S_2^2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (Y_j - \bar{Y})^2.$$

Mais il est préférable d'utiliser un estimateur basé sur la réunion des deux échantillons. D'après le théorème de FISHER (th. 120 page 90) $U_1 = \sum_{i=1}^{n_1} \left(\frac{X_i - \bar{X}}{\sigma_1} \right)^2$ suit la loi du khi-deux à $(n_1 - 1)$ degrés

de liberté et $U_2 = \sum_{j=1}^{n_2} \left(\frac{Y_j - \bar{Y}}{\sigma_2} \right)^2$ suit la loi du khi-deux à $(n_2 - 1)$ degrés de liberté ; de plus ces deux variables sont indépendantes : d'après le théorème 81, $U = U_1 + U_2$ suit la loi du khi-deux à $(n_1 + n_2 - 2)$ degrés de liberté. En posant $\sigma_1 = \sigma_2 = \sigma$ on obtient :

$$\mathbf{E}(U) = n_1 + n_2 - 2 = \frac{1}{\sigma^2} \mathbf{E} \left(\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{j=1}^{n_2} (Y_j - \bar{Y})^2 \right).$$

Un estimateur sans biais de σ^2 est donc :

$$S^2 = \frac{1}{n_1 + n_2 - 2} \left(\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{j=1}^{n_2} (Y_j - \bar{Y})^2 \right) = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

On remplace l'écart-type de D , $\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$, par $S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$:

$$\alpha = \mathbb{P}(D - a \leq m_1 - m_2 \leq D + b) = \mathbb{P} \left(\frac{-b}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leq \frac{D - (m_1 - m_2)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leq \frac{a}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right).$$

On pose :

$$T = \frac{D - (m_1 - m_2)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{D - (m_1 - m_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \bigg/ \sqrt{\frac{S^2}{\sigma^2}}.$$

Le numérateur $\frac{D - (m_1 - m_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ suit la loi normale réduite. Au dénominateur on a $\frac{S^2}{\sigma^2} = \frac{U}{n_1 + n_2 - 2}$

où U suit la loi du khi-deux à $(n_1 + n_2 - 2)$ degrés de liberté ; donc, par définition, T suit la loi de STUDENT à $(n_1 + n_2 - 2)$ degrés de liberté.

En choisissant u_1 et u_2 tels que $\mathbb{P}(u_1 \leq T \leq u_2) = \alpha$ on obtient l'intervalle de confiance annoncé.

On choisit généralement un intervalle symétrique en probabilité :

$$\left[D - t_{n_1+n_2-2, \frac{1+\alpha}{2}} S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, D + t_{n_1+n_2-2, \frac{1+\alpha}{2}} S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right].$$

Lorsque σ_1 et σ_2 sont inconnus, non nécessairement égaux, on utilise l'une des deux *méthodes approchées* suivantes.

Théorème 132. Si les échantillons ont des tailles importantes et égales ($n_1 = n_2 = n > 30$) un intervalle de confiance approché de $(m_1 - m_2)$ au niveau α est

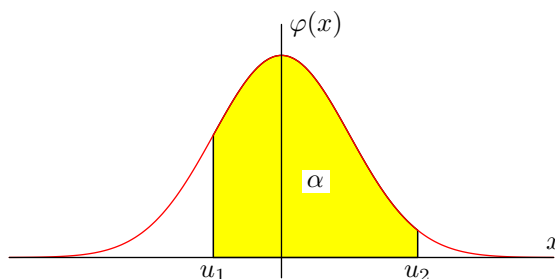
$$\left[D - \frac{u_2}{\sqrt{n}} \sqrt{S_1^2 + S_2^2}, D - \frac{u_1}{\sqrt{n}} \sqrt{S_1^2 + S_2^2} \right]$$

où u_1 et u_2 vérifient $\Phi(u_2) - \Phi(u_1) = \alpha$ et où

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2$$

et

$$S_2^2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (Y_j - \bar{Y})^2.$$



Il suffit de remarquer que la variable aléatoire $\frac{D - (m_1 - m_2)}{\sqrt{S_1^2 + S_2^2}/\sqrt{n}}$ suit sensiblement la loi normale réduite.

L'intervalle de confiance *approché*, symétrique en probabilité, de $(m_1 - m_2)$ au niveau α est

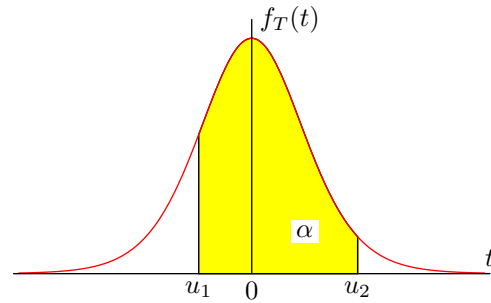
$$\left[D - z_{\frac{1+\alpha}{2}} \frac{1}{\sqrt{n}} \sqrt{S_1^2 + S_2^2}, D + z_{\frac{1+\alpha}{2}} \frac{1}{\sqrt{n}} \sqrt{S_1^2 + S_2^2} \right].$$

Théorème 133 (Méthode de WELCH). *Un intervalle de confiance approché de $(m_1 - m_2)$ au niveau α est*

$$\left[D - u_2 \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}, D - u_1 \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right]$$

où u_1 et u_2 vérifient $\mathbb{P}(u_1 \leq T \leq u_2) = \alpha$, la variable aléatoire T suivant la loi de STUDENT avec un nombre ν de degrés de liberté estimé à partir des échantillons :

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\frac{S_1^4}{n_1^2(n_1 - 1)} + \frac{S_2^4}{n_2^2(n_2 - 1)}}.$$



Ce résultat est admis.

L'intervalle de confiance *approché*, symétrique en probabilité, de $(m_1 - m_2)$ au niveau α fourni par la méthode de WELCH est :

$$\left[D - t_{\nu, \frac{1+\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}, D + t_{\nu, \frac{1+\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right].$$

6.5.2 Intervalle de confiance du rapport de deux variances

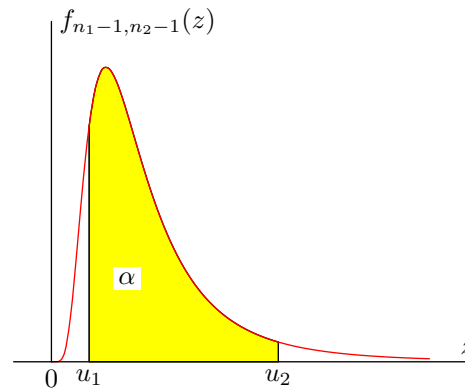
Théorème 134. *Un intervalle de confiance au niveau α de σ_1^2/σ_2^2 est*

$$\left[\frac{1}{u_2} \frac{S_1^2}{S_2^2}, \frac{1}{u_1} \frac{S_1^2}{S_2^2} \right]$$

où u_1 et u_2 sont donnés par la loi de FISHER-SNEDECOR $\mathcal{F}(n_1 - 1, n_2 - 1)$:

$$F_{n_1-1, n_2-1}(u_2) - F_{n_1-1, n_2-1}(u_1) = \alpha.$$

(Voir les tables page 155 et suivantes.)



On a vu au paragraphe précédent que $U_1 = \sum_{i=1}^{n_1} \left(\frac{X_i - \bar{X}}{\sigma_1} \right)^2$ et $U_2 = \sum_{j=1}^{n_2} \left(\frac{Y_j - \bar{Y}}{\sigma_2} \right)^2$ suivent les lois du khi-deux à $(n_1 - 1)$ et $(n_2 - 1)$ degrés de liberté respectivement et sont indépendantes.

On pose :

$$Z = \frac{\frac{U_1}{n_1 - 1}}{\frac{U_2}{n_2 - 1}} = \frac{\frac{1}{(n_1 - 1)\sigma_1^2} \sum_{i=1}^{n_1} (X_i - \bar{X})^2}{\frac{1}{(n_2 - 1)\sigma_2^2} \sum_{j=1}^{n_2} (Y_j - \bar{Y})^2} = \frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}.$$

Par définition (voir § 3.6.8) Z suit la loi de FISHER-SNEDECOR $\mathcal{F}(n_1 - 1, n_2 - 1)$. On cherche a et b tels que :

$$\alpha = \mathbb{P}(a \leq \frac{\sigma_1^2}{\sigma_2^2} \leq b) = \mathbb{P}(u_1 \leq Z \leq u_2) = F_{n_1-1, n_2-1}(u_2) - F_{n_1-1, n_2-1}(u_1)$$

où l'on a posé

$$u_1 = \frac{S_1^2/S_2^2}{b} \quad \text{et} \quad u_2 = \frac{S_1^2/S_2^2}{a}.$$

L'intervalle de confiance indiqué s'en déduit immédiatement.

On note $f_{n,m,\alpha}$ le fractile d'ordre α de la loi de FISHER-SNEDECOR $\mathcal{F}(n, m)$. On obtient un intervalle symétrique en probabilité en choisissant $u_1 = f_{n_1-1, n_2-1, \frac{1-\alpha}{2}}$ et $u_2 = f_{n_1-1, n_2-1, \frac{1+\alpha}{2}}$:

$$\left[\frac{1}{f_{n_1-1, n_2-1, \frac{1+\alpha}{2}}} \frac{S_1^2}{S_2^2}, \frac{1}{f_{n_1-1, n_2-1, \frac{1-\alpha}{2}}} \frac{S_1^2}{S_2^2} \right].$$

6.6 Méthode du « Bootstrap »

À partir d'un échantillon $X = (X_1, X_2, \dots, X_n)$ on détermine un estimateur ponctuel $s(X)$ d'un paramètre θ . Sauf dans quelques cas particuliers ($s(X) = \bar{X}_n$ par exemple) le calcul de la variance n'est pas aisé, ce qui rend problématique la détermination d'intervalles de confiance pour θ .

En 1979 EFRON [13] a développé une méthode s'appuyant sur des concepts simples permettant, à partir d'une réalisation (x_1, x_2, \dots, x_n) de l'échantillon, d'obtenir une estimation de la variance de $s(X)$ et un intervalle de confiance pour θ .

On considère que la réalisation de l'échantillon (x_1, x_2, \dots, x_n) est représentative de la population et on tire parmi les x_k , au hasard et *avec remise*, un échantillon « bootstrapé » $X^* = (X_1^*, X_2^*, \dots, X_n^*)$; en pratique on tire n nombres au hasard entre 1 et n et on associe au nombre tiré k la valeur x_k . Sur cet échantillon « bootstrapé » on peut calculer un estimateur $s(X^*)$ par le même algorithme que celui qui donne $s(X)$.

On répète le tirage un « grand » nombre de fois, B , ce qui donne une population de valeurs de $s(X^*)$

$$\mathcal{S} = \{s_1, s_2, \dots, s_B\}$$

que l'on peut représenter par un histogramme. Sur cette population on peut calculer une estimation de la moyenne et de l'écart-type :

$$\bar{s} = \frac{1}{B} \sum_{k=1}^B s_k, \quad \sqrt{\frac{1}{B-1} \sum_{k=1}^B (s_k - \bar{s})^2}.$$

La population \mathcal{S} peut être triée par valeurs croissantes ce qui permet de déterminer un intervalle de confiance en gardant une certaine proportion des valeurs « centrales ». Par exemple si $B = 1000$ et si les valeurs triées de \mathcal{S} sont $\alpha_1 \leq \alpha_2 \leq \alpha_3 \leq \dots \leq \alpha_{1000}$, l'intervalle de confiance à 95% est $[\alpha_{25}, \alpha_{975}]$.

Chapitre 7

Tests

7.1 Introduction

Un test d'hypothèse a pour but d'effectuer un choix entre deux hypothèses statistiques portant sur les caractéristiques d'une population. Ces caractéristiques peuvent dépendre, ou non, de paramètres :

- tests paramétriques : comparaison de moyennes. . .
- tests non paramétriques : tests d'adéquation (χ^2 par exemple). . .

Les deux hypothèses ne jouent pas des rôles symétriques. L'une des deux est supposée vraie *a priori* et est soumise au test : c'est l'hypothèse « nulle », notée H_0 ; l'hypothèse alternative est notée H_1 . Le test pourra conduire à rejeter H_0 ou à ne pas la rejeter ; il ne peut pas permettre de conclure que H_0 est vraie, mais seulement d'affirmer qu'il n'y a pas de raison de la rejeter.

Exemple : test d'un lot de pièces reçues d'un fournisseur. L'hypothèse nulle H_0 est de considérer que le lot est bon ; le rejet à tort de H_0 est plus grave que celui de H_1 .

Il existe deux types d'erreurs de décision :

- rejet à tort de l'hypothèse nulle ; $\mathbb{P}(\text{rejeter } H_0 \mid H_0)$ est le risque de première espèce,
- acceptation à tort de l'hypothèse nulle ; $\mathbb{P}(\text{rejeter } H_1 \mid H_1)$ est le risque de deuxième espèce.

L'idéal serait de réduire ces deux risques au minimum (et même de les annuler. . .) ; ce n'est pas possible car, pour diminuer le risque de première espèce, il faut éviter de rejeter H_0 , ce qui augmente le risque de deuxième espèce. Le risque de première espèce est considéré comme le plus important : on va le limiter à un seuil et on va alors tenter de minimiser le risque de deuxième espèce.

La *puissance du test* est $1 - \mathbb{P}(\text{rejeter } H_1 \mid H_1) = \mathbb{P}(\text{rejeter } H_0 \mid \overline{H_0})$.

7.2 Tests paramétriques

La variable aléatoire X a pour densité de probabilité $f(x, \theta)$ où le paramètre θ appartient à une partie Θ de \mathbb{R} . Le test est basé sur une statistique d'un n -échantillon de X ; on accepte H_0 si $\varphi(X_1, X_2, \dots, X_n)$ appartient à une partie A de \mathbb{R} et on la rejette sinon : A est la *région d'acceptation* du test et $\mathbb{R} \setminus A$ est la *région critique* ou *région de rejet*. La définition du test consiste à choisir la région de rejet.

Nous allons développer quelques exemples.

7.2.1 Test de la moyenne pour une loi normale (σ connu)

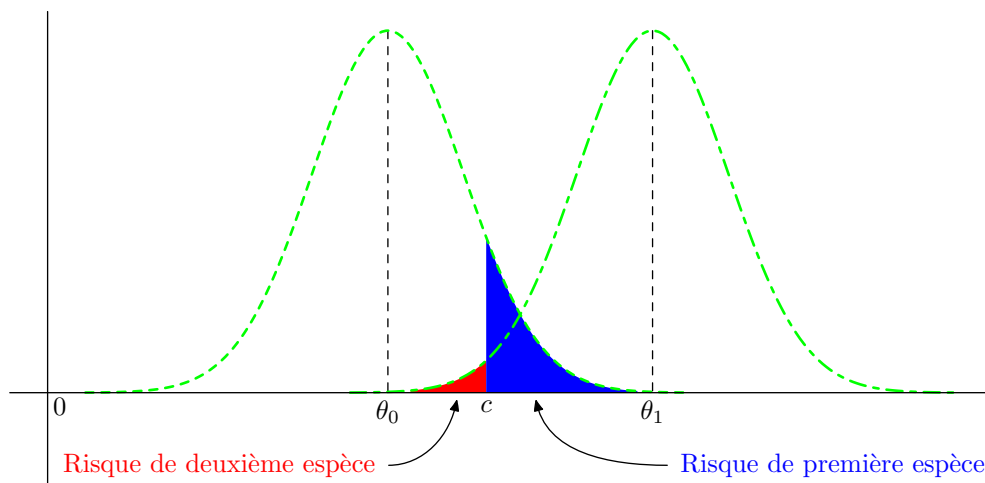
Supposons que X suit la loi normale $\mathcal{N}(\theta, \sigma)$ où σ est connu.

Hypothèses simples

Nous choisissons $\Theta = \{\theta_0, \theta_1\}$ avec $\theta_0 < \theta_1$ et :

$$H_0 : \theta = \theta_0, \quad H_1 : \theta = \theta_1.$$

Or θ est l'espérance mathématique de X : on peut l'estimer par $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$. Puisque $\theta_0 < \theta_1$ on choisit comme région de rejet $[c, +\infty[$, c'est-à-dire que la région d'acceptation est $A =]-\infty, c]$. Représentons la densité de probabilité de \bar{X}_n , qui suit la loi normale $\mathcal{N}(\theta, \sigma/\sqrt{n})$ sous les deux hypothèses H_0 et H_1 .



Déterminons c ; on limite le risque de première espèce au seuil α :

$$\mathbb{P}(\bar{X}_n \geq c \mid \theta = \theta_0) \leq \alpha$$

et on cherche à minimiser le risque de deuxième espèce

$$\mathbb{P}(\bar{X}_n < c \mid \theta = \theta_1).$$

On cherche donc c tel que

$$\mathbb{P}(\bar{X}_n \geq c \mid \theta = \theta_0) = 1 - \underbrace{\Phi\left(\frac{c - \theta_0}{\sigma/\sqrt{n}}\right)}_{\psi_0(c)} \leq \alpha$$

et

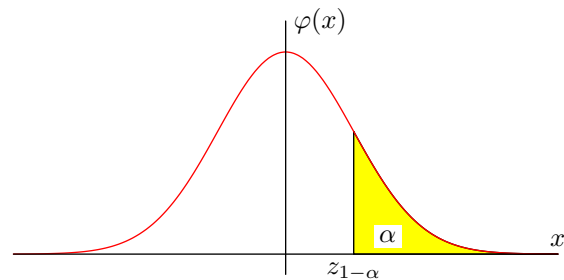
$$\mathbb{P}(\bar{X}_n < c \mid \theta = \theta_1) = \underbrace{\Phi\left(\frac{c - \theta_1}{\sigma/\sqrt{n}}\right)}_{\psi_1(c)} \text{ est minimum.}$$

La fonction ψ_1 étant croissante, on choisit c le plus petit possible vérifiant l'inégalité.

Or la fonction ψ_0 est décroissante ; donc on choisit c tel que $1 - \Phi\left(\frac{c - \theta_0}{\sigma/\sqrt{n}}\right) = \alpha$, soit :

$$c = \theta_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}}.$$

Le risque de deuxième espèce est alors :



$$\mathbb{P}(\overline{X}_n < c \mid \theta = \theta_1) = \Phi\left(\frac{c - \theta_1}{\sigma/\sqrt{n}}\right) = \Phi\left(z_{1-\alpha} - \frac{\theta_1 - \theta_0}{\sigma/\sqrt{n}}\right).$$

La puissance du test est d'autant plus grande que ce risque est plus petit et donc

- que $\theta_1 - \theta_0$ est grand (valeurs testées nettement distinctes),
- que n est grand (échantillon de taille importante).

Application numérique : $\theta_0 = 1$, $\theta_1 = 2$, $\sigma = 1$, $n = 20$, $\alpha = 0,05$.

On a $z_{1-\alpha} = 1,645$ et $c = 1,368$; la puissance du test est égale à 0,998.

Si on choisit un seuil $\alpha = 0,01$ on obtient $c = 1,52$ et la puissance du test est égale à 0,984.

Dans cet exemple les hypothèses sont *simples*, c'est-à-dire correspondent à une seule valeur du paramètre ; sinon l'hypothèse est *composite*.

Une hypothèse composite

On remarque que la valeur de c trouvée ci-dessus ne dépend pas de θ_1 ; on peut donc choisir $\Theta = [\theta_0, +\infty[$, prendre pour hypothèses

$$H_0 : \theta = \theta_0, \quad H_1 : \theta > \theta_0 \quad (\text{test unilatéral})$$

et décider de rejeter H_0 si $\overline{X}_n \geq c$ où $c = \theta_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$.

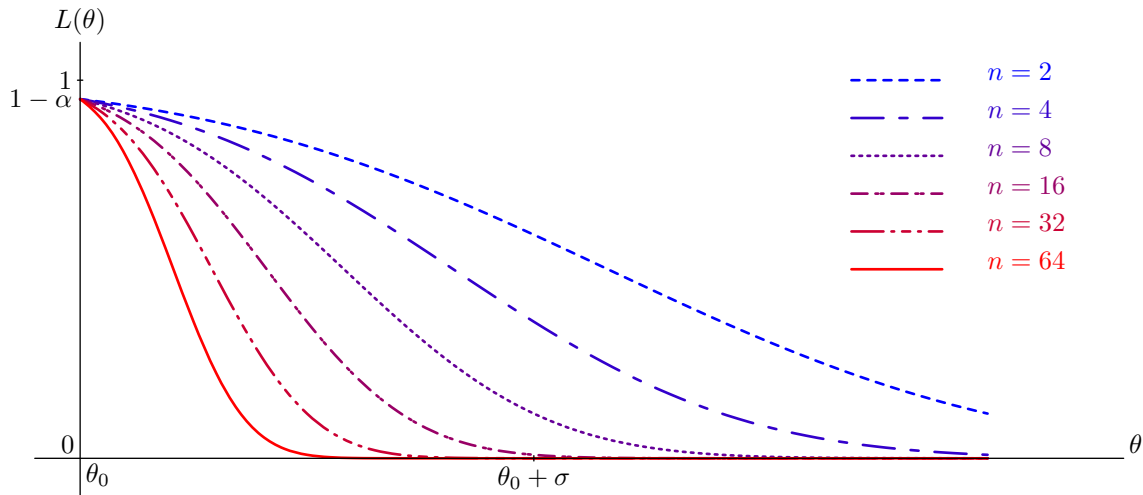
Le risque de première espèce $\mathbb{P}(\overline{X}_n \geq c \mid H_0)$ est encore égal à α .

Le risque de deuxième espèce est une fonction de θ :

$$\mathbb{P}(\overline{X}_n < c \mid H_1) = \Phi\left(\frac{c - \theta}{\sigma/\sqrt{n}}\right) = \Phi\left(z_{1-\alpha} - \frac{\theta - \theta_0}{\sigma/\sqrt{n}}\right).$$

La courbe d'efficacité du test est le graphe de la fonction

$$L(\theta) = \mathbb{P}(\text{ne pas rejeter } H_0 \mid \theta \text{ fixé}) = \mathbb{P}(\overline{X}_n < c \mid \theta \text{ fixé}) = \Phi\left(\frac{c - \theta}{\sigma/\sqrt{n}}\right).$$



Deux hypothèses composites

On choisit $\Theta = \mathbb{R}$, on prend pour hypothèses

$$H_0 : \theta \leq \theta_0, \quad H_1 : \theta > \theta_0 \quad (\text{test unilatéral})$$

et on décide de rejeter H_0 si $\overline{X}_n \geq c$. Déterminons c .

Le risque de première espèce doit être inférieur au seuil α quel que soit θ inférieur ou égal à θ_0 :

$$\forall \theta \leq \theta_0, \mathbb{P}(\overline{X}_n \geq c \mid H_0) = \mathbb{P}(\overline{X}_n \geq c \mid \theta \leq \theta_0) = \underbrace{1 - \Phi\left(\frac{c - \theta}{\sigma/\sqrt{n}}\right)}_{\varrho(\theta)} \leq \alpha.$$

La fonction $\theta \mapsto \varrho(\theta)$ étant croissante il faut et il suffit que l'inégalité soit vérifiée lorsque $\theta = \theta_0$; donc c doit vérifier

$$\underbrace{1 - \Phi\left(\frac{c - \theta_0}{\sigma/\sqrt{n}}\right)}_{\psi(c)} \leq \alpha.$$

Le risque de deuxième espèce

$$\mathbb{P}(\overline{X}_n < c \mid \theta > \theta_0) = \Phi\left(\frac{c - \theta}{\sigma/\sqrt{n}}\right)$$

est une fonction croissante de c : il est minimum lorsque c est minimum.

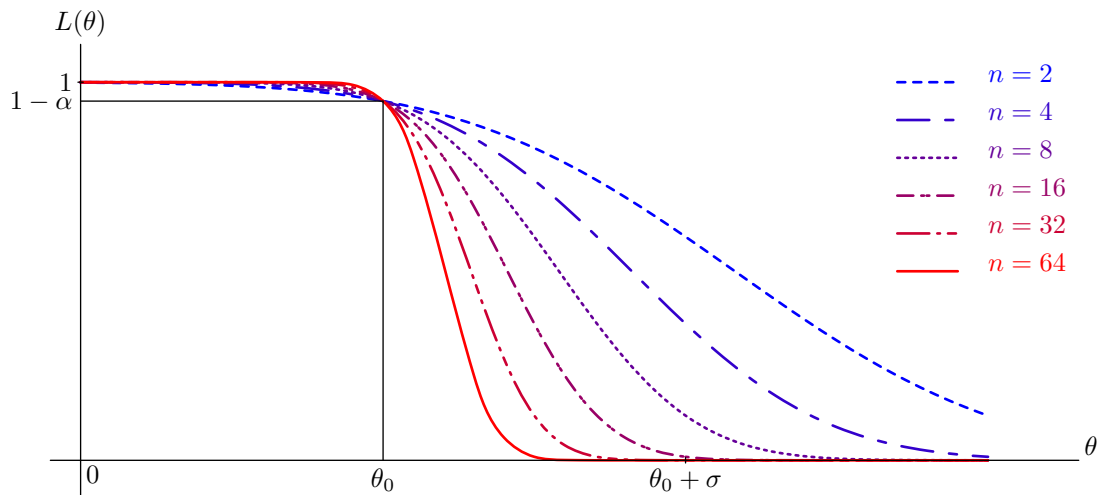
Puisque $c \mapsto \psi(c)$ est une fonction décroissante, on choisit c tel que $\psi(c) = \alpha$ ce qui donne encore :

$$c = \theta_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}}.$$

On a alors :

$$L(\theta) = \mathbb{P}(\overline{X}_n < c \mid \theta \text{ fixé}) = \Phi\left(\frac{c - \theta}{\sigma/\sqrt{n}}\right) = \Phi\left(z_{1-\alpha} + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right).$$

D'où la courbe d'efficacité :



On constate que le test est d'autant plus efficace que n est grand... ce qui n'est pas surprenant.

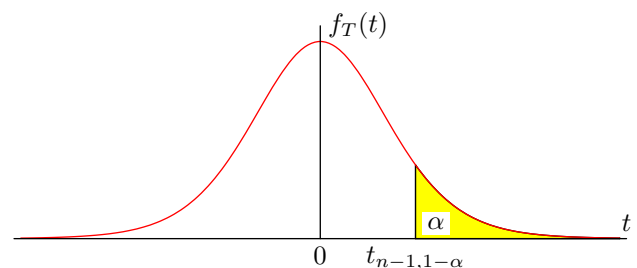
7.2.2 Test de la moyenne pour une loi normale (σ inconnu)

Lorsque σ est inconnu on le remplace par l'estimateur $\overline{S'_n}$. La statistique de test $\frac{\overline{X}_n - \theta}{\sigma/\sqrt{n}}$, qui suivait la loi normale réduite, est alors remplacée par $\frac{\overline{X}_n - \theta}{\overline{S'_n}/\sqrt{n}}$ qui suit la loi de STUDENT à $(n - 1)$ degrés de liberté (voir la démonstration du théorème 123).

Par les mêmes raisonnements que ceux du paragraphe 7.2.1 on décide de rejeter l'hypothèse nulle H_0 si

$$\overline{X}_n \geq \theta_0 + t_{n-1,1-\alpha} \frac{\overline{S'_n}}{\sqrt{n}}$$

où $t_{n-1,1-\alpha}$ est le fractile d'ordre $(1 - \alpha)$ de la loi de STUDENT à $(n - 1)$ degrés de liberté.



7.2.3 Test de l'écart-type pour une loi normale (m inconnue)

Supposons que X suit la loi normale $\mathcal{N}(m, \theta)$ où m est inconnue.

Test bilatéral (une hypothèse composite)

Nous choisissons $\Theta = \mathbb{R}$ et

$$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0.$$

La statistique de test est $\overline{S'_n}^2$, estimateur sans biais de la variance.

Sous H_0 la variable aléatoire $\frac{(n-1)\overline{S'_n}^2}{\theta_0^2}$ suit la loi du khi-deux à $(n - 1)$ degrés de liberté (voir théorème 120).

On rejette H_0 si $\frac{\overline{S'_n}^2}{\theta_0^2} < l_1 < 1$ ou si $\frac{\overline{S'_n}^2}{\theta_0^2} > l_2 > 1$. En partageant équitablement le risque de première espèce α entre les deux cas on aura

$$\mathbb{P}\left(\frac{\overline{S'_n}^2}{\theta_0^2} < l_1\right) = \mathbb{P}\left(\frac{(n-1)\overline{S'_n}^2}{\theta_0^2} < (n-1)l_1\right) = \frac{\alpha}{2}, \text{ d'où } l_1 = \frac{u_{n-1, \frac{\alpha}{2}}}{n-1},$$

et

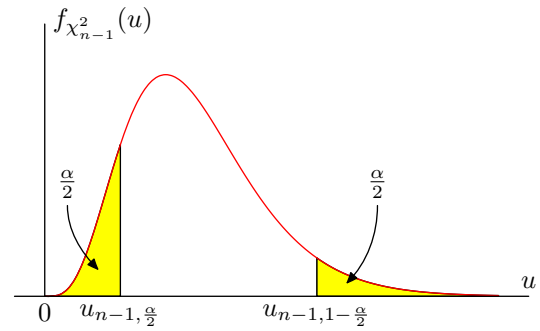
$$\mathbb{P}\left(\frac{\overline{S'_n}^2}{\theta_0^2} > l_2\right) = \mathbb{P}\left(\frac{(n-1)\overline{S'_n}^2}{\theta_0^2} > (n-1)l_2\right) = \frac{\alpha}{2}, \text{ d'où } l_2 = \frac{u_{n-1, 1-\frac{\alpha}{2}}}{n-1}.$$

On rejette donc H_0 si

$$\frac{\overline{S'_n}^2}{\theta_0^2} < \frac{u_{n-1, \frac{\alpha}{2}}}{n-1}$$

ou

$$\frac{\overline{S'_n}^2}{\theta_0^2} > \frac{u_{n-1, 1-\frac{\alpha}{2}}}{n-1}.$$



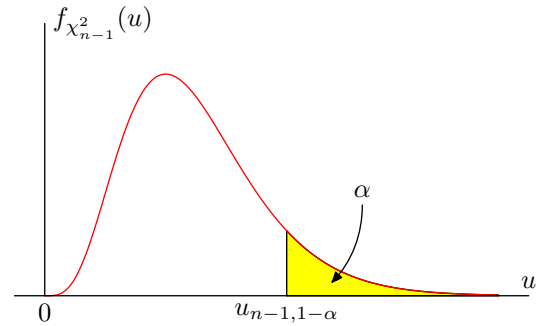
Test unilatéral (deux hypothèses composites)

Nous choisissons $\Theta = \mathbb{R}$ et

$$H_0 : \theta \leq \theta_0, \quad H_1 : \theta > \theta_0.$$

Par le même raisonnement que ci-dessus on rejette H_0 si

$$\frac{\overline{S'_n}^2}{\theta_0^2} > \frac{u_{n-1, 1-\alpha}}{n-1}.$$



7.2.4 Comparaison de proportions

Les populations A et B contiennent des proportions p_1 et p_2 d'éléments d'un certain type (pièces défectueuses par exemple). On tire de A et B des échantillons de tailles n_1 et n_2 respectivement : on obtient X (resp. Y) éléments de type cherché dans A (resp. B). On veut savoir si p_1 et p_2 peuvent être considérées comme égales ; on choisit donc :

$$H_0 : p_1 = p_2 \quad \text{et} \quad H_1 : p_1 \neq p_2.$$

Les variables X et Y suivent respectivement les lois binomiales $\mathcal{B}(n_1, p_1)$ et $\mathcal{B}(n_2, p_2)$. On suppose que n_1 et n_2 sont assez grands pour que l'on puisse approcher les lois binomiales par des lois

normales ; alors X/n_1 et Y/n_2 suivent respectivement les lois normales $\mathcal{N}(p_1, \sqrt{p_1(1-p_1)/n_1})$ et $\mathcal{N}(p_2, \sqrt{p_2(1-p_2)/n_2})$.

La statistique de test

$$T = \frac{X}{n_1} - \frac{Y}{n_2}$$

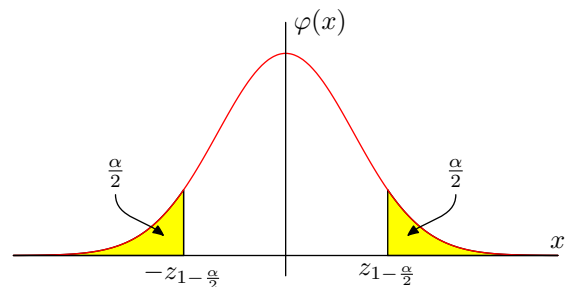
suit alors la loi normale $\mathcal{N}(p_1 - p_2, \sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2})$.

Sous H_0 , en notant $p = p_1 = p_2$, T suit la loi normale $\mathcal{N}(0, \sqrt{p(1-p)(1/n_1 + 1/n_2)})$.

On estime p par $P = (X+Y)/(n_1+n_2)$ et on rejette H_0 si

$$|T| > z_{1-\frac{\alpha}{2}} \sqrt{P(1-P) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

où $z_{1-\frac{\alpha}{2}}$ est le fractile d'ordre $(1 - \frac{\alpha}{2})$ de la loi normale réduite.



7.2.5 Comparaison de deux moyennes (observations appariées)

Soient X_i et Y_i les résultats de deux mesures sur une même unité $i \in \{1, 2, \dots, n\}$ (ou sur deux unités appariées) ; par exemple X_i et Y_i sont les notes obtenues par un même étudiant dans deux matières. On suppose que les variables X_i (resp. Y_i) suivent la même loi normale $\mathcal{N}(m_1, \sigma_1)$ (resp. $\mathcal{N}(m_2, \sigma_2)$) et que ces variables sont indépendantes.

On choisit :

$$H_0 : m_1 = m_2 \quad \text{et} \quad H_1 : m_1 \neq m_2.$$

Sous H_0 les variables $D_i = X_i - Y_i$ suivent la loi normale $\mathcal{N}(0, \sqrt{\sigma_1^2 + \sigma_2^2})$ et sont indépendantes.

On pose :

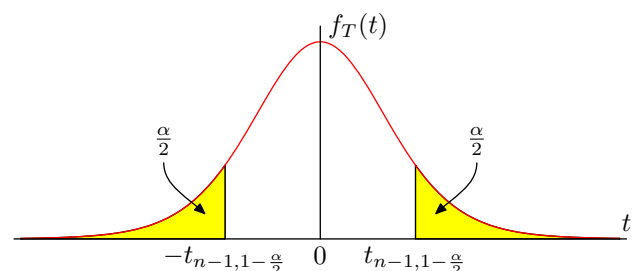
$$\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i \quad \text{et} \quad \overline{S_n'}^2 = \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2.$$

La statistique de test $\frac{\bar{D}}{\overline{S_n'}/\sqrt{n}}$ suit la loi de STUDENT à $(n-1)$ degrés de liberté (voir la démonstration du théorème 123).

Avec un risque égal à α on rejette donc H_0 si

$$|\bar{D}| > t_{n-1, 1-\frac{\alpha}{2}} \frac{\overline{S_n'}}{\sqrt{n}}$$

où $t_{n-1, 1-\frac{\alpha}{2}}$ représente le fractile d'ordre $1 - \frac{\alpha}{2}$ de la loi de STUDENT à $(n-1)$ degrés de liberté.



7.2.6 Comparaison de plusieurs moyennes (observations non appariées) – analyse de variance pour un facteur

On prélève indépendamment des n -échantillons dans k populations ; on note X_{ij} la j -ième valeur obtenue ($j \in \{1, 2, \dots, n\}$) dans la population n° i ($i \in \{1, 2, \dots, k\}$). Ces données peuvent être regroupées dans un tableau :

	1	2	...	j	...	n	Moyenne
1	X_{11}	X_{12}	...	X_{1j}	...	X_{1n}	\bar{X}_1
2	X_{21}	X_{22}	...	X_{2j}	...	X_{2n}	\bar{X}_2
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
i	X_{i1}	X_{i2}	...	X_{ij}	...	X_{in}	\bar{X}_i
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
k	X_{k1}	X_{k2}	...	X_{kj}	...	X_{kn}	\bar{X}_k

On suppose que, pour tout i , les variables aléatoires X_{ij} suivent, pour tout j , la même loi normale $\mathcal{N}(m_i, \sigma)$. On désire tester l'hypothèse nulle :

$$H_0 : m_1 = m_2 = \dots = m_k.$$

On rejettera cette hypothèse si la variance des moyennes des échantillons est grande par rapport à la variance dans les échantillons.

On pose :

$$\bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{ij} \quad \text{et} \quad \bar{\bar{X}} = \frac{1}{k} \sum_{i=1}^k \bar{X}_i.$$

Sous H_0 et en notant m la valeur commune des m_i , la variable aléatoire \bar{X}_i suit, pour tout i , la loi normale $\mathcal{N}(m, \sigma/\sqrt{n})$; donc, d'après le théorème de FISHER (voir th. 120, page 90), la variable aléatoire

$$\frac{1}{(\sigma/\sqrt{n})^2} \sum_{i=1}^k (\bar{X}_i - \bar{\bar{X}})^2 = \frac{n}{\sigma^2} \sum_{i=1}^k (\bar{X}_i - \bar{\bar{X}})^2$$

suit la loi du khi-deux à $(k - 1)$ degrés de liberté. En posant

$$SC_E = n \sum_{i=1}^k (\bar{X}_i - \bar{\bar{X}})^2 \quad \text{"sommes des carrés entre échantillons"}$$

la variable aléatoire $(1/\sigma^2)SC_E$ suit la loi du khi-deux à $(k - 1)$ degrés de liberté.

D'après le même théorème de FISHER les variables aléatoires $(1/\sigma^2) \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2$ suivent, pour tout i , la loi du khi-deux à $(n - 1)$ degrés de liberté ; elles sont indépendantes. Donc la variable aléatoire

$$\sum_{i=1}^k \left(\frac{1}{\sigma^2} \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2 \right) = \frac{1}{\sigma^2} SC_I,$$

en posant

$$SC_I = \sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 \quad (\text{"somme des carrés à l'intérieur des échantillons"}),$$

suit la loi du khi-deux à $(k(n-1))$ degrés de liberté.

Le rapport des variances

$$F = \frac{\frac{SC_E}{k-1}}{\frac{SC_I}{k(n-1)}} = \frac{s_E^2}{s_I^2}$$

suit donc la loi de FISHER-SNEDECOR à $(k-1, k(n-1))$ degrés de liberté (voir page 62).

On rejette H_0 au risque α si la réalisation de F pour les échantillons dépasse le fractile $f_{k-1, k(n-1), 1-\alpha}$ d'ordre $(1-\alpha)$ de cette loi.

Généralisation

Le test s'applique encore lorsque les effectifs des échantillons ne sont pas égaux ; soit n_i la taille du i -ième échantillon. Les formules sont alors :

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}, \quad \bar{\bar{x}} = \frac{\sum_{i=1}^k n_i \bar{x}_i}{\sum_{i=1}^k n_i},$$

$$SC_E = \sum_{i=1}^k n_i (\bar{x}_i - \bar{\bar{x}})^2, \quad SC_I = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2, \quad F = \frac{SC_E/k - 1}{SC_I / \sum_{i=1}^k (n_i - 1)}.$$

La statistique de test F suit la loi de FISHER-SNEDECOR à $(k-1, \sum_{i=1}^k (n_i - 1))$ degrés de liberté. Les résultats sont résumés dans le tableau d'analyse de variance qui se présente ainsi :

Source de variation	Degrés de liberté	Somme des carrés	Moyenne des carrés	F	Probabilité critique
Entre les échantillons	$k - 1$	SC_E	s_E^2	$\frac{s_E^2}{s_I^2}$	p_c
À l'intérieur des échantillons	$\sum_{i=1}^k (n_i - 1)$	SC_I	s_I^2		
Total	$\sum_{i=1}^k n_i - 1$	$SC_T = SC_E + SC_I$			

7.2.7 Probabilité critique

Soit t la réalisation de la statistique de test T obtenue à partir d'un échantillon. La *probabilité critique* p_c est la probabilité d'obtenir t ou une valeur plus éloignée de θ_0 lorsque H_0 est vraie.

– Test bilatéral

On rejette H_0 si $|T| > \ell$; la probabilité critique est

$$p_c(t) = \mathbb{P}(|T| \geq t \mid \theta = \theta_0).$$

– Test unilatéral

On rejette H_0 si $T > \ell$; la probabilité critique est

$$p_c(t) = \mathbb{P}(T \geq t \mid \theta = \theta_0).$$

Lorsque la probabilité critique est faible la contradiction est forte entre H_0 et les valeurs obtenues dans l'échantillon : si le risque de première espèce est égal à α on décide de rejeter H_0 si $p_c < \alpha$. Les logiciels de statistique fournissent généralement la probabilité critique (« p -value »), laissant l'utilisateur conclure en fonction du risque de première espèce choisi.

7.3 Test du khi-deux

Une épreuve peut fournir m résultats différents A_1, A_2, \dots, A_m . On désire tester l'hypothèse suivante (« hypothèse nulle ») :

$$H_0 : \forall j \in \{1, 2, \dots, m\}, \mathbb{P}(A_j) = p_j$$

où les probabilités p_j sont fixées avec $\sum_{j=1}^m p_j = 1$.

On réalise n épreuves indépendantes ; soit N_j le nombre de réalisations de A_j . Si l'hypothèse H_0 est vérifiée on a (voir la loi multinomiale au § 2.7.2, page 25) :

$$\mathbb{P}\left(\bigcap_{j=1}^m (N_j = n_j)\right) = \frac{n!}{n_1! n_2! \dots n_m!} p_1^{n_1} p_2^{n_2} \dots p_m^{n_m}.$$

Théorème 135. La variable $\sum_{j=1}^m \frac{(N_j - np_j)^2}{np_j}$ converge en loi vers la loi du khi-deux à $(m-1)$ degrés de liberté lorsque n tend vers l'infini.

Remarque : la loi limite est indépendante des probabilités p_j .

On reprend les notations du § 2.7.2.

Pour tout $j \in \{1, 2, \dots, m\}$ et tout $k \in \{1, 2, \dots, n\}$ soit $X_j^{(k)}$ le nombre (égal à 0 ou 1) d'élément de type j tiré au k -ième tirage. Pour j fixé les variables $(X_j^{(k)})_{1 \leq k \leq n}$ sont indépendantes et, pour tout k , $\mathbb{P}(X_j^{(k)} = 1) = p_j$.

Nous introduisons alors, pour $k \in \{1, 2, \dots, n\}$, les vecteurs aléatoires

$$Y_k = \left(\frac{X_j^{(k)} - p_j}{\sqrt{np_j}} \right)_{1 \leq j \leq m}$$

et le vecteur aléatoire

$$Z_n = \sum_{k=1}^n Y_k = \left(\sum_{k=1}^n \frac{X_j^{(k)} - p_j}{\sqrt{np_j}} \right)_{1 \leq j \leq m} = \left(\frac{N_j - np_j}{\sqrt{np_j}} \right)_{1 \leq j \leq m}.$$

Les vecteurs Y_k étant indépendants et de même loi la fonction caractéristique de Z_n est définie par (voir théorème 54) :

$$\forall t \in \mathbb{R}^m, \widehat{f_{Z_n}}(t) = (\widehat{f_{Y_1}}(t))^n.$$

Par définition

$$\widehat{f_{Y_1}}(t) = \mathbf{E} \left(\exp \left(i \sum_{j=1}^m t_j \frac{X_j^{(1)} - p_j}{\sqrt{np_j}} \right) \right).$$

Or le vecteur $(X_1^{(1)}, X_2^{(1)}, \dots, X_m^{(1)})$ prend, pour tout j , la valeur $(0, \dots, 0, 1, 0, \dots, 0)$ où le 1 est à la j -ième place avec la probabilité p_j . Donc :

$$\begin{aligned} \widehat{f_{Y_1}}(t) &= \sum_{j=1}^m p_j \exp \left(i \left(\sum_{k \neq j} t_k \frac{-p_k}{\sqrt{np_k}} + t_j \frac{1-p_j}{\sqrt{np_j}} \right) \right) = \sum_{j=1}^m p_j \exp \left(i \left(- \sum_{k=1}^m \frac{t_k \sqrt{p_k}}{\sqrt{n}} + \frac{t_j}{\sqrt{np_j}} \right) \right) \\ &= \exp \left(-i \sum_{k=1}^m \frac{t_k \sqrt{p_k}}{\sqrt{n}} \right) \sum_{j=1}^m p_j \exp \left(i \frac{t_j}{\sqrt{np_j}} \right). \end{aligned}$$

D'où :

$$\begin{aligned} \log \widehat{f_{Z_n}}(t) &= n \log \widehat{f_{Y_1}}(t) = -in \sum_{k=1}^m \frac{t_k \sqrt{p_k}}{\sqrt{n}} + n \log \sum_{j=1}^m p_j \exp \left(i \frac{t_j}{\sqrt{np_j}} \right) \\ &= -i\sqrt{n} \sum_{k=1}^m t_k \sqrt{p_k} + n \log \left\{ 1 + \sum_{j=1}^m p_j \left(\exp \left(i \frac{t_j}{\sqrt{np_j}} \right) - 1 \right) \right\} \\ &= -i\sqrt{n} \sum_{k=1}^m t_k \sqrt{p_k} + n \log \left\{ 1 + \sum_{j=1}^m p_j \left(i \frac{t_j}{\sqrt{np_j}} - \frac{t_j^2}{2np_j} + o\left(\frac{1}{n}\right) \right) \right\} \\ &= -i\sqrt{n} \sum_{k=1}^m t_k \sqrt{p_k} + n \log \left(1 + \frac{i}{\sqrt{n}} \sum_{j=1}^m t_j \sqrt{p_j} - \frac{1}{2n} \sum_{j=1}^m t_j^2 + o\left(\frac{1}{n}\right) \right) \\ &= n \left\{ -\frac{1}{2n} \sum_{j=1}^m t_j^2 - \frac{1}{2} \left(-\frac{1}{n} \left(\sum_{j=1}^m t_j \sqrt{p_j} \right)^2 \right) + o\left(\frac{1}{n}\right) \right\} \\ &= -\frac{1}{2} \sum_{j=1}^m t_j^2 + \frac{1}{2} \left(\sum_{j=1}^m t_j \sqrt{p_j} \right)^2 + o(1) \end{aligned}$$

Donc

$$\widehat{f_{Z_n}}(t) \xrightarrow{n \rightarrow \infty} \exp \left(-\frac{1}{2} \sum_{j=1}^m t_j^2 + \frac{1}{2} \left(\sum_{j=1}^m t_j \sqrt{p_j} \right)^2 \right) = \exp \left(-\frac{1}{2} t^T c t \right)$$

où

$$c = I - v v^T \quad \text{avec } v^T = \begin{bmatrix} \sqrt{p_1} & \sqrt{p_2} & \cdots & \sqrt{p_m} \end{bmatrix}$$

D'après le théorème 91 le vecteur aléatoire Z_n converge en loi vers un vecteur gaussien Z de loi $\mathcal{N}(0, c)$.

La matrice c est symétrique : il existe une matrice orthogonale u telle que $d = u^T c u$ soit la matrice diagonale des valeurs propres de c . Or celles-ci sont :

- la valeur propre simple égale à 0 dont le sous-espace propre associé est engendré par v ;
- la valeur propre d'ordre $(m-1)$ égale à 1 dont le sous-espace propre a pour équation $\sum_{j=1}^m \sqrt{p_j} x_j = 0$.

En ordonnant convenablement la base de vecteurs propres on peut choisir :

$$d = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 \\ 0 & \dots & 0 & 0 \end{pmatrix}.$$

D'après le corollaire 3 le vecteur aléatoire $W = u^T Z$ est gaussien de loi $\mathcal{N}(0, d)$. La variable aléatoire

$$\sum_{j=1}^m \frac{(N_j - np_j)^2}{np_j} = Z_n^T Z_n$$

converge en loi vers

$$Z^T Z = (uW)^T (uW) = W^T u^T u W = W^T W = \sum_{j=1}^m W_j^2 = \sum_{j=1}^{m-1} W_j^2,$$

puisque W_m est presque sûrement nulle car elle a une espérance mathématique et une variance nulles. Pour $j \in \{1, 2, \dots, m-1\}$ la variable aléatoire W_j suit la loi normale réduite et ces variables sont indépendantes ; par définition $Z^T Z$ suit la loi du khi-deux à $(m-1)$ degrés de liberté.

Lorsque n est grand la variable aléatoire $\sum_{j=1}^m \frac{(N_j - np_j)^2}{np_j}$ suit donc sensiblement la loi du khi-deux à $(m-1)$ degrés de liberté.

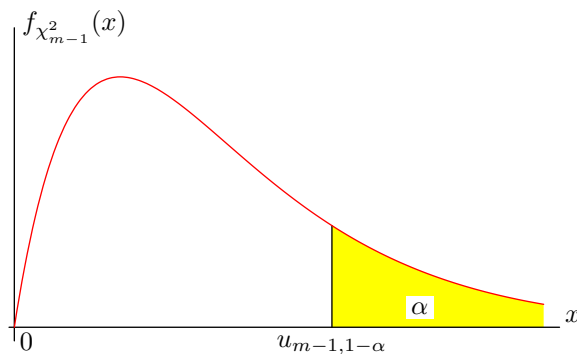
Soit (n_1, n_2, \dots, n_m) une réalisation de l'échantillon (N_1, N_2, \dots, N_m) . On évalue la somme

$$t = \sum_{j=1}^m \frac{(n_j - np_j)^2}{np_j}.$$

L'hypothèse H_0 est rejetée si t est trop grande : on fixe un niveau α « petit » et on ne rejette pas H_0 si t est inférieure au fractile $u_{m-1, 1-\alpha}$ tel que

$$\mathbb{P}(\chi_{m-1}^2 \geq u_{m-1, 1-\alpha}) = \alpha.$$

(La valeur $u_{m-1, 1-\alpha}$ est lue dans la table des fractiles de la loi du χ^2 , page 152).



L'éventualité de rejeter H_0 lorsqu'elle est vraie constitue le risque de première espèce ; on choisit une probabilité α « petite » afin de limiter ce risque. On ne sait rien (a priori) sur le risque d'accepter H_0 lorsqu'elle est fausse (risque de seconde espèce) ; il est supposé de moindre importance que le risque de première espèce.

Remarques

1. En pratique les valeurs des produits np_j ne doivent pas être trop petites : $np_j \geq 5$ est raisonnable. Si ce n'est pas le cas il faut regrouper les (données en) classes.
2. Si certains paramètres p_j sont inconnus et doivent être estimés le nombre de degrés de liberté de la loi du χ^2 est diminué d'autant.

Chapitre 8

Régression linéaire

8.1 Introduction : lignes de régression

Soient X et Y deux variables aléatoires ; on cherche à caractériser la relation entre ces variables : on peut, par exemple, associer à une valeur de X la moyenne des valeurs correspondantes de Y . On obtient ainsi les lignes de régression de Y en X . On peut également chercher la régression de X en Y .

8.1.1 Cas discret

Pour $(i, j) \in I \times J$ on pose : $p_{ij} = \mathbb{P}((X = x_i) \cap (Y = y_j))$. Les lois de X et de Y sont définies par

$$p_{i.} = \sum_{j \in J} p_{ij}, \quad p_{.j} = \sum_{i \in I} p_{ij}.$$

La loi conditionnelle de Y sachant que $X = x_i$ est définie par

$$\forall j \in J, \mathbb{P}(Y = y_j \mid X = x_i) = \frac{p_{ij}}{p_{i.}}.$$

On pose :

$$\forall i \in I, m_i = \mathbf{E}(Y \mid X = x_i) = \sum_{j \in J} y_j \frac{p_{ij}}{p_{i.}} = \frac{1}{p_{i.}} \sum_{j \in J} y_j p_{ij}.$$

Les points $(x_i, m_i)_{i \in I}$ définissent la régression de Y en X .

De même, en posant $m'_j = \mathbf{E}(X \mid Y = y_j) = \frac{1}{p_{.j}} \sum_{i \in I} x_i p_{ij}$, la régression de X en Y est définie par les points $(m'_j, y_j)_{j \in J}$.

8.1.2 Cas continu

Le couple (X, Y) ayant une densité de probabilité $(x, y) \mapsto f(x, y)$ la régression de Y en X est la courbe d'équation

$$y = m(x) = \frac{1}{f_X(x)} \int_{-\infty}^{\infty} y f(x, y) dy.$$

On définit de manière analogue la régression de X en Y .

Cas particulier : loi normale

La densité conditionnelle de Y sachant que $X = x$ est

$$\begin{aligned} f_{Y|X=x}(y) &= \frac{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left[-\frac{1}{2(1-\rho^2)} \left\{ \left(\frac{x-m_1}{\sigma_1} \right)^2 - 2\rho \frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \left(\frac{y-m_2}{\sigma_2} \right)^2 \right\} \right]}{\frac{1}{\sigma_1\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x-m_1}{\sigma_1} \right)^2 \right]} \\ &= \frac{1}{(\sigma_2\sqrt{1-\rho^2})\sqrt{2\pi}} \exp \left[-\frac{1}{2(\sigma_2\sqrt{1-\rho^2})^2} \left(y - \left(m_2 + \rho \frac{\sigma_2}{\sigma_1} (x - m_1) \right) \right)^2 \right] \end{aligned}$$

La loi conditionnelle de Y sachant que $X = x$ est donc la loi normale

$$\mathcal{N}\left(m_2 + \rho \frac{\sigma_2}{\sigma_1} (x - m_1), \sigma_2 \sqrt{1 - \rho^2}\right).$$

La régression de Y en X est donc définie par

$$y = m(x) = \mathbf{E}(Y | X = x) = m_2 + \rho \frac{\sigma_2}{\sigma_1} (x - m_1).$$

La courbe de régression de Y en X est la droite d'équation

$$\frac{y - m_2}{\sigma_2} = \rho \frac{x - m_1}{\sigma_1}.$$

De même la courbe de régression de X en Y est la droite d'équation

$$\frac{x - m_1}{\sigma_1} = \rho \frac{y - m_2}{\sigma_2}.$$

Ces deux droites sont distinctes si $|\rho| \neq 1$.

Le fait que les courbes de régression soient des droites dans le cas de la loi normale conduit naturellement à s'intéresser à la régression linéaire.

8.2 Régression linéaire simple

Les valeurs Y_1, Y_2, \dots, Y_n d'une *variable dépendante* (ou *expliquée*) correspondent aux valeurs x_1, x_2, \dots, x_n de la *variable indépendante* (ou *explicative*) ; on postule le modèle linéaire

$$\forall i \in \{1, 2, \dots, n\}, Y_i = a + bx_i + U_i$$

où les variables aléatoires U_i , qui représentent les *erreurs résiduelles*, vérifient

$$\forall i \in \{1, 2, \dots, n\}, \mathbf{E}(U_i) = 0, \mathbf{Var}(U_i) = \sigma^2 \text{ et, pour } i \neq j, \mathbf{Cov}(U_i, U_j) = 0.$$

Remarque : les Y_i sont des variables aléatoires, mais pas les x_i qui sont fixés par l'expérimentateur.

On va estimer les paramètres a, b et σ^2 en envisageant deux cas :

1. aucune hypothèse n'est faite sur la loi de probabilité des U_i : on estime a et b par la méthode des moindres carrés ;
2. les erreurs résiduelles U_i suivent une loi normale : on estime a, b et σ^2 par la méthode du maximum de vraisemblance.

8.2.1 Estimation des paramètres par les moindres carrés

Détermination des estimateurs

On minimise la somme des carrés des erreurs :

$$\varphi(a, b) = \sum_{i=1}^n U_i^2 = \sum_{i=1}^n (Y_i - a - bx_i)^2.$$

Théorème 136. *Les estimateurs des moindres carrés de a et b sont*

$$\hat{b} = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{et} \quad \hat{a} = \bar{Y} - \hat{b} \bar{x}$$

où l'on a posé

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{et} \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i.$$

Une condition nécessaire de minimum

$$\frac{\partial \varphi(a, b)}{\partial a} = \frac{\partial \varphi(a, b)}{\partial b} = 0$$

conduit au système des *équations normales* :

$$\begin{cases} na + b \sum_{i=1}^n x_i = \sum_{i=1}^n Y_i, \\ a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i Y_i. \end{cases}$$

Leur résolution donne immédiatement le résultat annoncé.

On montre à l'aide des équations normales :

$$\varphi(a, b) - \varphi(\hat{a}, \hat{b}) = n((a - \hat{a}) + (b - \hat{b})\bar{x})^2 + (b - \hat{b})^2 \sum_{i=1}^n (x_i - \bar{x})^2.$$

Donc (\hat{a}, \hat{b}) est bien le minimum de φ .

Remarque : malgré l'abandon des conventions typographiques utilisées dans ce cours et l'adoption du « chapeau » habituel dans la littérature, il ne faut pas oublier que les estimateurs des paramètres sont des variables aléatoires.

Propriétés des estimateurs

Théorème 137. *Les estimateurs des moindres carrés sont sans biais :*

$$\mathbf{E}(\hat{a}) = a \quad \text{et} \quad \mathbf{E}(\hat{b}) = b.$$

Pour tout $i \in \{1, 2, \dots, n\}$ on a $Y_i = a + bx_i + U_i$; on en déduit

$$\bar{Y} = a + b\bar{x} + \bar{U} \quad \text{où} \quad \bar{U} = \frac{1}{n} \sum_{i=1}^n U_i$$

puis

$$\forall i \in \{1, 2, \dots, n\}, Y_i - \bar{Y} = b(x_i - \bar{x}) + (U_i - \bar{U}).$$

On reporte dans l'expression de \hat{b} :

$$\begin{aligned} \hat{b} &= \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{b \sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (x_i - \bar{x})(U_i - \bar{U})}{\sum_{i=1}^n (x_i - \bar{x})^2} = b + \frac{\sum_{i=1}^n (x_i - \bar{x})(U_i - \bar{U})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= b + \frac{\sum_{i=1}^n (x_i - \bar{x})U_i - \bar{U} \sum_{i=1}^n (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = b + \frac{\sum_{i=1}^n (x_i - \bar{x})U_i}{\sum_{i=1}^n (x_i - \bar{x})^2}. \end{aligned} \quad (8.1)$$

D'où :

$$\mathbf{E}(\hat{b}) = b + \frac{\sum_{i=1}^n (x_i - \bar{x})\mathbf{E}(U_i)}{\sum_{i=1}^n (x_i - \bar{x})^2} = b, \quad \text{car, pour tout } i, \mathbf{E}(U_i) = 0.$$

Par ailleurs

$$\hat{a} = \bar{Y} - \hat{b}\bar{x} = a + (b - \hat{b})\bar{x} + \bar{U}$$

et donc

$$\mathbf{E}(\hat{a}) = a + \bar{x}(b - \mathbf{E}(\hat{b})) + \mathbf{E}(\bar{U}) = a.$$

Définition 57. Pour tout i on note $\hat{Y}_i = \hat{a} + \hat{b}x_i$ l'estimateur de Y_i .

Corollaire 4. On a :

$$\mathbf{E}(\hat{Y}_i) = \mathbf{E}(Y_i) = a + bx_i.$$

D'après le théorème précédent

$$\mathbf{E}(\hat{Y}_i) = \mathbf{E}(\hat{a} + \hat{b}x_i) = a + bx_i.$$

D'autre part

$$\mathbf{E}(Y_i) = \mathbf{E}(a + bx_i + U_i) = a + bx_i \quad \text{puisque} \quad \mathbf{E}(U_i) = 0.$$

Théorème 138. Les variances et covariance sont données par

$$\mathbf{Var}(\hat{a}) = \frac{\sigma^2}{n} \frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \mathbf{Var}(\hat{b}) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{et} \quad \mathbf{Cov}(\hat{a}, \hat{b}) = -\frac{\sigma^2 \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

1. D'après la relation 8.1 on a : $\hat{b} - b = \frac{\sum_{i=1}^n (x_i - \bar{x})U_i}{\sum_{i=1}^n (x_i - \bar{x})^2}.$

D'où :

$$\begin{aligned} \mathbf{Var}(\hat{b}) &= \mathbf{E}((\hat{b} - b)^2) = \frac{\mathbf{E}\left(\left(\sum_{i=1}^n (x_i - \bar{x})U_i\right)^2\right)}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} \\ &= \frac{1}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} \mathbf{E}\left(\sum_{i=1}^n ((x_i - \bar{x})U_i)^2 + 2 \sum_{1 \leq i < j \leq n} (x_i - \bar{x})(x_j - \bar{x})U_i U_j\right) \\ &= \frac{1}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} \left(\sum_{i=1}^n (x_i - \bar{x})^2 \underbrace{\mathbf{E}(U_i^2)}_{=\sigma^2} + 2 \sum_{1 \leq i < j \leq n} (x_i - \bar{x})(x_j - \bar{x}) \underbrace{\mathbf{E}(U_i U_j)}_{=0}\right) \\ &= \frac{\sigma^2 \sum_{i=1}^n (x_i - \bar{x})^2}{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}. \end{aligned}$$

2. De la relation $\hat{a} = a - (\hat{b} - b)\bar{x} + \bar{U}$ on déduit :

$$\mathbf{Var}(\hat{a}) = \mathbf{Var}(\hat{a} - a) = \bar{x}^2 \mathbf{Var}(\hat{b}) + \mathbf{Var}(\bar{U}) - 2\bar{x} \mathbf{Cov}(\hat{b} - b, \bar{U}).$$

On connaît

$$\mathbf{Var}(\hat{b}) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{et} \quad \mathbf{Var}(\bar{U}) = \frac{\sigma^2}{n}.$$

Calculons la covariance. On a, pour tout i ,

$$\mathbf{Cov}(U_i, \bar{U}) = \frac{1}{n} \sum_{j=1}^n \mathbf{Cov}(U_i, U_j) = \frac{\sigma^2}{n}.$$

D'où, d'après 8.1 :

$$\text{Cov}(\hat{b} - b, \bar{U}) = \frac{\sum_{i=1}^n (x_i - \bar{x}) \text{Cov}(U_i, \bar{U})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sigma^2 \sum_{i=1}^n (x_i - \bar{x})}{n \sum_{i=1}^n (x_i - \bar{x})^2} = 0.$$

En reportant : $\text{Var}(\hat{a}) = \frac{\sigma^2}{n} \left(1 + \frac{n\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)$. On obtient l'expression annoncée de $\text{Var}(\hat{a})$

en remarquant :

$$\sum_{i=1}^n (x_i - \bar{x})^2 + n\bar{x}^2 = \left(\sum_{i=1}^n x_i^2 - 2\bar{x} \underbrace{\sum_{i=1}^n x_i}_{=n\bar{x}} + n\bar{x}^2 \right) + n\bar{x}^2 = \sum_{i=1}^n x_i^2.$$

3. De la relation $\hat{a} - a = -(\hat{b} - b)\bar{x} + \bar{U}$ on déduit

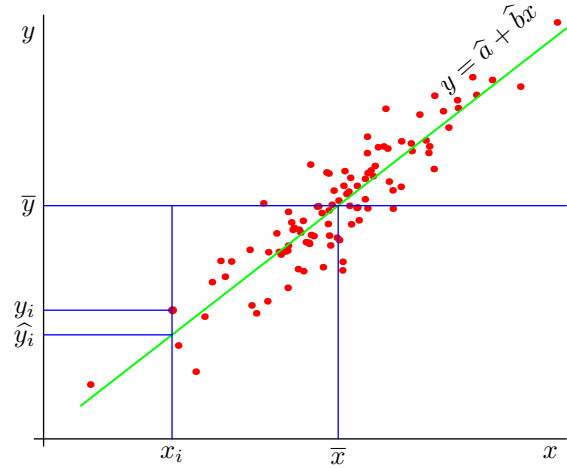
$$\text{Cov}(\hat{a}, \hat{b}) = \mathbf{E}((\hat{a} - a)(\hat{b} - b)) = -\bar{x} \mathbf{E}((\hat{b} - b)^2) + \mathbf{E}((\hat{b} - b)\bar{U}) = -\bar{x} \text{Var}(\hat{b})$$

car $\mathbf{E}((\hat{b} - b)\bar{U}) = \text{Cov}(\hat{b} - b, \bar{U}) = 0$ comme on l'a vu ci-dessus. D'où le résultat.

Qualité de l'estimation — Analyse de variance

Théorème 139. Soit $\hat{U}_i = Y_i - \hat{Y}_i$ le résidu en x_i . On a la relation d'analyse de la variance :

$$\underbrace{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2}_{\text{variance « totale »}} = \underbrace{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}_{\text{variance « expliquée »}} + \underbrace{\frac{1}{n} \sum_{i=1}^n (\hat{U}_i - \bar{\hat{U}})^2}_{\text{variance « résiduelle »}}$$



Remarque : les notations y_i , \hat{y}_i et \bar{y} sont utilisées dans le graphique pour souligner que ce sont des réalisations des variables aléatoires Y_i , \hat{Y}_i et \bar{Y} .

1. Le résidu en x_i peut s'écrire

$$\hat{U}_i = Y_i - (\hat{a} + \hat{b} x_i) = Y_i - \bar{Y} - \hat{b}(x_i - \bar{x}), \quad (\text{car } \hat{a} = \bar{Y} - \hat{b} \bar{x})$$

et on a par ailleurs

$$\bar{\hat{Y}} = \frac{1}{n} \sum_{i=1}^n \hat{Y}_i = \hat{a} + \hat{b} \bar{x} = \bar{Y}$$

et

$$\bar{\hat{U}} = \frac{1}{n} \sum_{i=1}^n \hat{U}_i = \frac{1}{n} \sum_{i=1}^n Y_i - \frac{1}{n} \sum_{i=1}^n \hat{Y}_i = \bar{Y} - \bar{\hat{Y}} = 0.$$

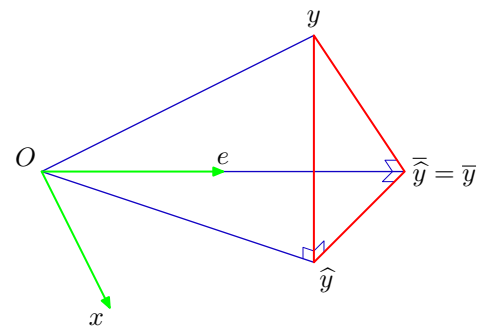
D'où :

$$\begin{aligned} & \sum_{i=1}^n (\hat{U}_i - \bar{\hat{U}})^2 + \sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})^2 \\ &= \sum_{i=1}^n \hat{U}_i^2 + \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \\ &= \sum_{i=1}^n (Y_i - \bar{Y} - \hat{b}(x_i - \bar{x}))^2 + \sum_{i=1}^n ((\hat{a} + \hat{b} x_i) - (\hat{a} + \hat{b} \bar{x}))^2 \\ &= \left(\sum_{i=1}^n (Y_i - \bar{Y})^2 + \hat{b}^2 \sum_{i=1}^n (x_i - \bar{x})^2 - 2\hat{b} \underbrace{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}_{=\hat{b} \sum_{i=1}^n (x_i - \bar{x})^2} \right) + \hat{b}^2 \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \sum_{i=1}^n (Y_i - \bar{Y})^2. \end{aligned}$$

2. On peut donner une autre démonstration, de nature géométrique.

La méthode des moindres carrés consiste à minimiser la distance de Y au plan engendré par les vecteurs $e = [1 \ 1 \ \dots \ 1]^T$ et x ; \hat{Y} est donc la projection orthogonale de Y sur ce plan.

La projection orthogonale de Y sur la droite engendrée par e est \bar{Y} ; de même la projection orthogonale de \hat{Y} sur cette même droite est $\bar{\hat{Y}}$. D'après le théorème des trois perpendiculaires¹ on a $\bar{\hat{Y}} = \bar{Y}$.



Le triangle (Y, \hat{Y}, \bar{Y}) est rectangle (en \hat{Y}); d'après le théorème de PYTHAGORE

$$(Y - \bar{Y})^2 = (Y - \hat{Y})^2 + (\hat{Y} - \bar{\hat{Y}})^2,$$

ce qui donne l'équation d'analyse de la variance.

¹Si dans un tétraèdre $ABCD$ les angles \widehat{ABC} , \widehat{ABD} et \widehat{BCD} sont droits, alors l'angle \widehat{ACD} est également droit.

$$\overrightarrow{AC} \cdot \overrightarrow{CD} = (\overrightarrow{AB} + \overrightarrow{BC}) \cdot \overrightarrow{CD} = \overrightarrow{AB} \cdot (\overrightarrow{BD} - \overrightarrow{BC}) = 0.$$

Définition 58. Le coefficient de détermination est le rapport de la variance expliquée à la variance totale :

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}.$$

D'après le théorème précédent

$$0 \leq R^2 \leq 1.$$

Le coefficient de détermination donne une bonne indication sur la qualité de l'ajustement : plus il est grand, plus la variance résiduelle est faible.

8.2.2 Estimation des paramètres par le maximum de vraisemblance

On suppose que le vecteur aléatoire $U = (U_1, U_2, \dots, U_n)$ est gaussien de loi $\mathcal{N}(0, \sigma^2 I_n)$, c'est-à-dire que les variables aléatoires U_i sont normales, centrées et indépendantes, d'écart-type σ .

Détermination des estimateurs

Théorème 140. Les estimateurs du maximum de vraisemblance de a , b et σ^2 sont

$$\hat{b} = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \hat{a} = \bar{Y} - \hat{b} \bar{x}$$

et

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{a} - \hat{b}x_i)^2 = \frac{1}{n} \sum_{i=1}^n \hat{U}_i^2.$$

La vraisemblance de l'échantillon (Y_1, Y_2, \dots, Y_n) est

$$\mathcal{L}(a, b, \sigma^2) = (\sigma\sqrt{2\pi})^{-n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - a - bx_i)^2\right).$$

On pose pour simplifier $v = \sigma^2$ et

$$\psi(a, b, v) = \ln(\mathcal{L}(a, b, v)) = -\frac{n}{2} \ln(2\pi v) - \frac{1}{2v} \sum_{i=1}^n (Y_i - a - bx_i)^2.$$

Les équations de vraisemblance s'écrivent :

$$\begin{aligned}\frac{\partial \psi(a, b, v)}{\partial a} &= \frac{1}{v} \sum_{i=1}^n (Y_i - a - bx_i) = 0, \\ \frac{\partial \psi(a, b, v)}{\partial b} &= \frac{1}{v} \sum_{i=1}^n (Y_i - a - bx_i)x_i = 0, \\ \frac{\partial \psi(a, b, v)}{\partial v} &= -\frac{n}{2v} + \frac{1}{2v^2} \sum_{i=1}^n (Y_i - a - bx_i)^2 = 0.\end{aligned}$$

Les deux premières donnent les équations normales de la méthode des moindres carrés ; la dernière fournit l'estimateur de $v = \sigma^2$.

On peut vérifier que la matrice hessienne est définie négative.

Propriétés des estimateurs

Les estimateurs du maximum de vraisemblance de a et b sont les estimateurs des moindres carrés (voir th. 136), ce qui n'était pas évident *a priori* : ils sont donc sans biais (voir th. 137).

Théorème 141. *L'estimateur du maximum de vraisemblance de σ^2 est biaisé :*

$$\mathbf{E}(\widehat{\sigma^2}) = \frac{n-2}{n} \sigma^2.$$

D'après le théorème 139 :

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n \hat{U}_i^2.$$

Or

$$\left. \begin{aligned} \hat{Y}_i &= \hat{a} + \hat{b} x_i \\ \bar{Y} &= \hat{a} + \hat{b} \bar{x} \end{aligned} \right\} \Rightarrow \hat{Y}_i - \bar{Y} = \hat{b}(x_i - \bar{x}) \text{ et } \left. \begin{aligned} Y_i &= a + bx_i + U_i \\ \bar{Y} &= a + b\bar{x} + \bar{U} \end{aligned} \right\} \Rightarrow Y_i - \bar{Y} = b(x_i - \bar{x}) + (U_i - \bar{U}).$$

D'où, en reportant :

$$\sum_{i=1}^n \hat{U}_i^2 = b^2 \sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (U_i - \bar{U})^2 + 2b \sum_{i=1}^n (x_i - \bar{x})(U_i - \bar{U}) - \hat{b}^2 \sum_{i=1}^n (x_i - \bar{x})^2.$$

Or

$$\mathbf{E}\left(\sum_{i=1}^n (x_i - \bar{x})(U_i - \bar{U})\right) = \sum_{i=1}^n (x_i - \bar{x}) \mathbf{E}(U_i - \bar{U}) = 0$$

et

$$\mathbf{E}\left(\sum_{i=1}^n (U_i - \bar{U})^2\right) = \mathbf{E}\left(\sum_{i=1}^n U_i^2 - n\bar{U}^2\right) = \sum_{i=1}^n \mathbf{E}(U_i^2) - n\mathbf{E}(\bar{U}^2) = n\sigma^2 - n\frac{\sigma^2}{n} = (n-1)\sigma^2.$$

D'où enfin :

$$\begin{aligned} \mathbf{E}\left(\sum_{i=1}^n \hat{U}_i^2\right) &= b^2 \sum_{i=1}^n (x_i - \bar{x})^2 + (n-1)\sigma^2 - \mathbf{E}(\hat{b}^2) \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= b^2 \sum_{i=1}^n (x_i - \bar{x})^2 + (n-1)\sigma^2 - \left(\frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} + b^2\right) \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= (n-2)\sigma^2. \end{aligned}$$

Le théorème s'en déduit immédiatement.

Théorème 142. Un estimateur sans biais de σ^2 est :

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{U}_i^2.$$

Remarque : pour calculer une estimation de σ^2 on utilise la relation d'analyse de la variance

$$\sum_{i=1}^n \hat{U}_i^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2 - \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2 - \hat{b}^2 \sum_{i=1}^n (x_i - \bar{x})^2.$$

Lois des estimateurs

Théorème 143. Le couple (\hat{a}, \hat{b}) des estimateurs de a et b est gaussien de loi $\mathcal{N}(m, c)$ où

$$m = \begin{pmatrix} a \\ b \end{pmatrix} \quad \text{et} \quad c = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}.$$

Le couple (\hat{a}, \hat{b}) s'exprime linéairement en fonction des U_i , qui sont gaussiennes ; d'après le corollaire 3 il est donc gaussien. Le calcul des espérances mathématiques, variances et covariance a été effectué aux théorèmes 137 et 138 sans supposer que les U_i sont gaussiennes (mais le résultat est encore valable...).

Théorème 144. La loi de l'estimateur $\hat{\sigma}^2$ est telle que

$$(n-2) \frac{\hat{\sigma}^2}{\sigma^2} \text{ suit la loi du khi-deux à } (n-2) \text{ degrés de liberté.}$$

Ce théorème est un cas particulier du théorème 161.

Corollaire 5. La variance de l'estimateur $\hat{\sigma}^2$ est

$$\mathbf{Var}(\hat{\sigma}^2) = \frac{2\sigma^4}{n-2}.$$

$$\mathbf{Var}\left((n-2) \frac{\hat{\sigma}^2}{\sigma^2}\right) = \mathbf{Var}(\chi_{n-2}^2) = 2(n-2) = \frac{(n-2)^2}{\sigma^4} \mathbf{Var}(\hat{\sigma}^2).$$

D'où le résultat.

Théorème 145. Les estimateurs \hat{a} et \hat{b} sont indépendants de chacun des résidus \hat{U}_i .

De

$$\hat{U}_i = Y_i - \hat{Y}_i = (a + bx_i + U_i) - (\hat{a} + \hat{b} x_i) = -(\hat{a} - a) - (\hat{b} - b)x_i + U_i$$

on déduit

$$\mathbf{E}(\hat{U}_i) = -(\mathbf{E}(\hat{a}) - a) - (\mathbf{E}(\hat{b}) - b)x_i + \mathbf{E}(U_i) = 0$$

et

$$\mathbf{Cov}(\hat{b}, \hat{U}_i) = \mathbf{E}(\hat{b} \hat{U}_i) = -\mathbf{E}(\hat{b}(\hat{a} - a)) - x_i \mathbf{E}(\hat{b}(\hat{b} - b)) + \mathbf{E}(\hat{b} U_i).$$

Or :

$$\begin{aligned} \mathbf{E}(\hat{b}(\hat{a} - a)) &= \mathbf{Cov}(\hat{a}, \hat{b}) + b\mathbf{E}(\hat{a} - a) = \mathbf{Cov}(\hat{a}, \hat{b}) = -\frac{\sigma^2 \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2}, \\ \mathbf{E}(\hat{b}(\hat{b} - b)) &= \mathbf{Var}(\hat{b}) + b\mathbf{E}(\hat{b} - b) = \mathbf{Var}(\hat{b}) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \text{ et} \\ \mathbf{E}(\hat{b} U_i) &= \mathbf{E}\left(\left(b + \frac{\sum_{j=1}^n (x_j - \bar{x}) U_j}{\sum_{i=1}^n (x_i - \bar{x})^2}\right) U_i\right) \quad (\text{d'après 8.1}) \\ &= b\mathbf{E}(U_i) + \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \sum_{j=1}^n (x_j - \bar{x}) \underbrace{\mathbf{E}(U_i U_j)}_{\begin{cases} \sigma^2 & \text{si } j = i, \\ 0 & \text{sinon.} \end{cases}} \\ &= \frac{\sigma^2 (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}. \end{aligned}$$

En reportant on obtient

$$\mathbf{Cov}(\hat{b}, \hat{U}_i) = 0.$$

On procède de même pour \hat{a} ; on obtient

$$\mathbf{Cov}(\hat{a}, \hat{U}_i) = \mathbf{E}(\hat{a} \hat{U}_i) = -\mathbf{E}(\hat{a}(\hat{a} - a)) - x_i \mathbf{E}(\hat{a}(\hat{b} - b)) + \mathbf{E}(\hat{a} U_i) = -\mathbf{Var}(\hat{a}) - x_i \mathbf{Cov}(\hat{a}, \hat{b}) + \mathbf{E}(\hat{a} U_i).$$

Or

$$\mathbf{E}(\hat{a} U_i) = \mathbf{E}((\bar{Y} - \hat{b} \bar{x}) U_i) = \mathbf{E}(\bar{Y} U_i) - \bar{x} \mathbf{E}(\hat{b} U_i)$$

et

$$\mathbf{E}(\bar{Y} U_i) = \frac{1}{n} \mathbf{E}\left(\sum_{j=1}^n (a + bx_j + U_j) U_i\right) = \frac{1}{n} \sum_{j=1}^n (a + bx_j) \mathbf{E}(U_i) + \frac{1}{n} \sum_{j=1}^n \mathbf{E}(U_i U_j) = \frac{\sigma^2}{n}.$$

En reportant on obtient

$$\mathbf{Cov}(\hat{a}, \hat{U}_i) = 0.$$

Domaines de confiance sur les paramètres

Théorème 146. Un intervalle de confiance symétrique en probabilité au niveau α de b est

$$\left[\hat{b} - t_{n-2, \frac{1+\alpha}{2}} \frac{S}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}, \hat{b} + t_{n-2, \frac{1+\alpha}{2}} \frac{S}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \right]$$

où $t_{n-2, \frac{1+\alpha}{2}}$ représente le fractile d'ordre $\frac{1+\alpha}{2}$ de la loi de STUDENT à $(n-2)$ degrés de liberté et où $S = \sqrt{\widehat{\sigma^2}}$.

On sait que $\frac{\hat{b} - b}{\sigma / \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$ suit la loi normale réduite et que $(n-2) \frac{\widehat{\sigma^2}}{\sigma^2}$ suit la loi du khi-deux à $(n-2)$ degrés de liberté ; donc

$$\frac{\hat{b} - b}{\sigma / \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \bigg/ \sqrt{\frac{\widehat{\sigma^2}}{\sigma^2}} = \frac{\hat{b} - b}{S / \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

suit la loi de STUDENT à $(n-2)$ degrés de liberté. L'intervalle de confiance s'en déduit immédiatement.

Théorème 147. Un intervalle de confiance symétrique en probabilité au niveau α de a est

$$\left[\hat{a} - t_{n-2, \frac{1+\alpha}{2}} \frac{S \sqrt{\sum_{i=1}^n x_i^2}}{\sqrt{n \sum_{i=1}^n (x_i - \bar{x})^2}}, \hat{a} + t_{n-2, \frac{1+\alpha}{2}} \frac{S \sqrt{\sum_{i=1}^n x_i^2}}{\sqrt{n \sum_{i=1}^n (x_i - \bar{x})^2}} \right]$$

où $t_{n-2, \frac{1+\alpha}{2}}$ représente le fractile d'ordre $\frac{1+\alpha}{2}$ de la loi de STUDENT à $(n-2)$ degrés de liberté et où $S = \sqrt{\widehat{\sigma^2}}$.

Identique à celle du théorème précédent.

On peut donner un résultat pour le couple (a, b) .

Théorème 148. Une région de confiance au niveau α de (a, b) est l'ellipse dont l'équation dans le plan (a, b) s'écrit

$$(\hat{a} - a)^2 + 2\bar{x}(\hat{a} - a)(\hat{b} - b) + \frac{1}{n} \sum_{i=1}^n x_i^2 (\hat{b} - b)^2 < \frac{2\widehat{\sigma^2}}{n} f_{2, n-2, \alpha}$$

où $f_{2, n-2, \alpha}$ représente le fractile d'ordre α de la loi de FISHER-SNEDECOR $\mathcal{F}(2, n-2)$.

On a vu (théorème 143) que le couple $\mathbf{V} = (\hat{a}, \hat{b})$ est gaussien de loi $\mathcal{N}(m, c)$; d'après le théorème 82 la variable aléatoire

$$\mathbf{W} = (\mathbf{V} - m)^T c^{-1} (\mathbf{V} - m)$$

¹Il y a une subtilité à ce sujet : $\widehat{\sigma^2}$ est un estimateur sans biais de la variance σ^2 , ce qui n'entraîne pas que S soit un estimateur sans biais de l'écart-type ; nous évitons la notation $\hat{\sigma}$ qui est souvent employée bien qu'elle soit trompeuse.

suit la loi du khi-deux à deux degrés de liberté. On montre immédiatement que

$$c^{-1} = \frac{n}{\sigma^2} \begin{pmatrix} 1 & \bar{x} \\ \bar{x} & \frac{\sum_{i=1}^n x_i^2}{n} \end{pmatrix}.$$

Donc :

$$W = \frac{n}{\sigma^2} ((\hat{a} - a)^2 + 2\bar{x}(\hat{a} - a)(\hat{b} - b) + \frac{\sum_{i=1}^n x_i^2}{n} (\hat{b} - b)^2).$$

Par ailleurs (théorème 144) $Z = (n - 2) \frac{\widehat{\sigma^2}}{\sigma^2}$ suit la loi du khi-deux à $(n - 2)$ degrés de liberté. Donc (définition 35) le rapport

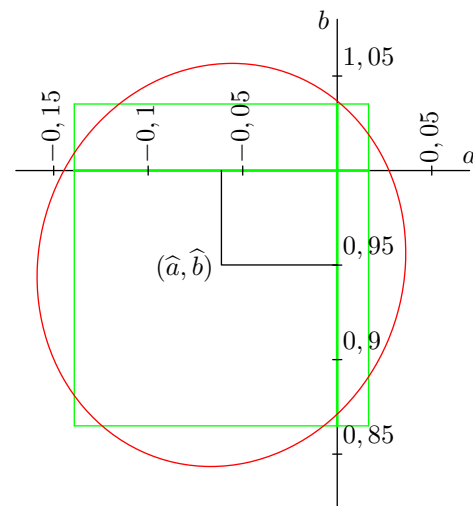
$$Q = \frac{W/2}{Z/(n - 2)} = \frac{n}{2\widehat{\sigma^2}} ((\hat{a} - a)^2 + 2\bar{x}(\hat{a} - a)(\hat{b} - b) + \frac{\sum_{i=1}^n x_i^2}{n} (\hat{b} - b)^2)$$

suit la loi de FISHER-SNEDECOR à 2 et $(n - 2)$ degrés de liberté $\mathcal{F}(2, n - 2)$. On en déduit immédiatement l'ellipse de confiance.

À partir des données représentées à la page 124 on a tracé sur le graphique ci-contre :

- l'ellipse de confiance du couple (a, b) ;
- les deux intervalles de confiance de a et b .

Le niveau de confiance est égal à 95 %.



Théorème 149. Un intervalle de confiance symétrique en probabilité au niveau α de σ^2 est

$$\left[\frac{(n - 2)\widehat{\sigma^2}}{k_{n-2, \frac{1+\alpha}{2}}}, \frac{(n - 2)\widehat{\sigma^2}}{k_{n-2, \frac{1-\alpha}{2}}} \right]$$

où $k_{n-2, \frac{1-\alpha}{2}}$ et $k_{n-2, \frac{1+\alpha}{2}}$ représentent les fractiles d'ordres $\frac{1-\alpha}{2}$ et $\frac{1+\alpha}{2}$ de la loi du khi-deux à $(n - 2)$ degrés de liberté.

Ce résultat est une conséquence immédiate du théorème 144.

Tests sur les paramètres

La connaissance des lois de probabilité de \hat{a} , \hat{b} , (\hat{a}, \hat{b}) et $\widehat{\sigma^2}$ permet, en appliquant les méthodes étudiées au paragraphe 7.2, de définir des tests sur ces paramètres.

Théorème 150 (Test bilatéral sur b). On teste l'hypothèse nulle

$$H_0 : b = b_0 \quad \text{contre l'hypothèse alternative} \quad H_1 : b \neq b_0.$$

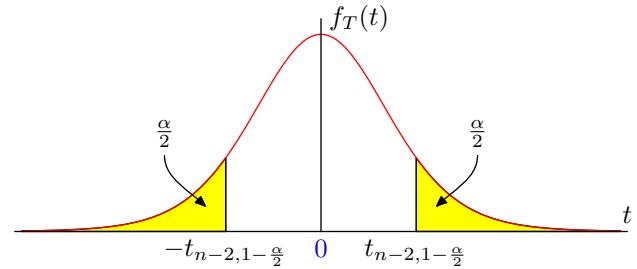
Au risque α on rejette H_0 si

$$\frac{|\hat{b} - b_0|}{S / \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} > t_{n-2, 1-\frac{\alpha}{2}}$$

où $S = \sqrt{\widehat{\sigma^2}}$ et où $t_{n-2, 1-\frac{\alpha}{2}}$ représente le fractile d'ordre $1 - \frac{\alpha}{2}$ de la loi de STUDENT à $(n - 2)$ degrés de liberté.

Remarque : ce test est surtout utilisé avec $b_0 = 0$, c'est-à-dire pour savoir si la variable expliquée est indépendante de la variable explicative.

On sait que \hat{b} suit la loi normale $\mathcal{N}(b, \sigma_b)$ où $\sigma_b^2 = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$; donc, sous l'hypothèse H_0 , $\frac{\hat{b} - b_0}{\sigma_b}$ suit la loi normale réduite. Par ailleurs (voir le théorème 144) $\widehat{\sigma^2}$ suit la loi du khi-deux à $(n - 2)$ degrés de liberté. Donc (voir la démonstration du théorème 146) le rapport $\frac{\hat{b} - b_0}{S / \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$ suit la loi de STUDENT à $(n - 2)$ degrés de liberté. La région de rejet s'en déduit immédiatement.



Théorème 151 (Test bilatéral sur b (variante)). On teste l'hypothèse nulle

$$H_0 : b = b_0 \quad \text{contre l'hypothèse alternative} \quad H_1 : b \neq b_0.$$

Au risque α on rejette H_0 si

$$\frac{(\hat{b} - b_0)^2}{\widehat{\sigma^2} / \sum_{i=1}^n (x_i - \bar{x})^2} > f_{1, n-2, 1-\alpha}$$

où $f_{1, n-2, 1-\alpha}$ représente le fractile d'ordre $1 - \alpha$ de la loi de FISHER-SNEDECOR à 1 et $(n - 2)$ degrés de liberté $\mathcal{F}(1, n - 2)$.

Il suffit d'appliquer le théorème 90.

On peut définir de même (c'est un bon exercice...) des test unilatéraux sur b correspondants aux hypothèses nulles

$$H_0 : b \geq b_0 \quad \text{ou} \quad H_0 : b \leq b_0.$$

Théorème 152 (Test bilatéral sur a). On teste l'hypothèse nulle

$$H_0 : a = a_0 \quad \text{contre l'hypothèse alternative} \quad H_1 : a \neq a_0.$$

Au risque α on rejette H_0 si

$$\frac{|\hat{a} - a_0|}{\frac{S}{\sqrt{n}} \sqrt{\frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} > t_{n-2, 1-\frac{\alpha}{2}}$$

où $S = \sqrt{\widehat{\sigma^2}}$ et où $t_{n-2, 1-\frac{\alpha}{2}}$ représente le fractile d'ordre $1 - \frac{\alpha}{2}$ de la loi de STUDENT à $(n - 2)$ degrés de liberté.

La démonstration est analogue à celle du théorème 150.

Théorème 153 (Test sur (a, b)). On teste l'hypothèse nulle

$$H_0 : a = a_0 \text{ et } b = b_0 \quad \text{contre l'hypothèse alternative} \quad H_1 : a \neq a_0 \text{ ou } b \neq b_0.$$

Au risque α on rejette H_0 si

$$Q = \frac{n}{2\widehat{\sigma^2}} ((\hat{a} - a_0)^2 + 2\bar{x}(\hat{a} - a_0)(\hat{b} - b_0) + \frac{\sum_{i=1}^n x_i^2}{n} (\hat{b} - b_0)^2) > f_{2, n-2, 1-\alpha}$$

où $f_{2, n-2, 1-\alpha}$ représente le fractile d'ordre $1 - \alpha$ de la loi de FISHER-SNEDECOR à 2 et $(n - 2)$ degrés de liberté $\mathcal{F}(2, n - 2)$.

Ce théorème est fréquemment utilisé pour tester l'hypothèse $H_0 : a = b = 0$.

On a vu dans la démonstration du théorème 148 que Q suit la loi de FISHER-SNEDECOR à 2 et $(n - 2)$ degrés de liberté $\mathcal{F}(2, n - 2)$; le résultat s'en déduit immédiatement.

Prédiction de la variable expliquée

On peut prédire la valeur de la variable expliquée pour une valeur x_0 de la variable explicative par

$$\hat{Y}_0 = \hat{a} + \hat{b} x_0.$$

D'après le modèle on a

$$Y_0 = a + bx_0 + U_0 \quad \text{où} \quad \left| \begin{array}{l} U_0 \text{ suit la loi normale } \mathcal{N}(0, \sigma), \\ \forall i \in \{1, 2, \dots, n\}, \mathbf{E}(U_0 U_i) = 0. \end{array} \right.$$

La différence $\hat{Y}_0 - Y_0 = (\hat{a} - a) + (\hat{b} - b)x_0 - U_0$ suit la loi normale $\mathcal{N}(0, s)$ où

$$\begin{aligned} s^2 &= \mathbf{Var}(\hat{Y}_0 - Y_0) = \mathbf{Var}(\hat{a}) + x_0^2 \mathbf{Var}(\hat{b}) + 2x_0 \mathbf{Cov}(\hat{a}, \hat{b}) + \mathbf{Var}(U_0) \\ &= \sigma^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right). \end{aligned}$$

La variance σ^2 , inconnue, est estimée par $\widehat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n \widehat{U}_i^2$; on pose

$$S^2 = \widehat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)$$

Donc (voir la démonstration du théorème 146) le rapport $\frac{\widehat{Y}_0 - Y_0}{S}$ suit la loi de STUDENT à $(n - 2)$ degrés de liberté.

Un intervalle de confiance au niveau α de Y_0 (intervalle de prévision) est donc :

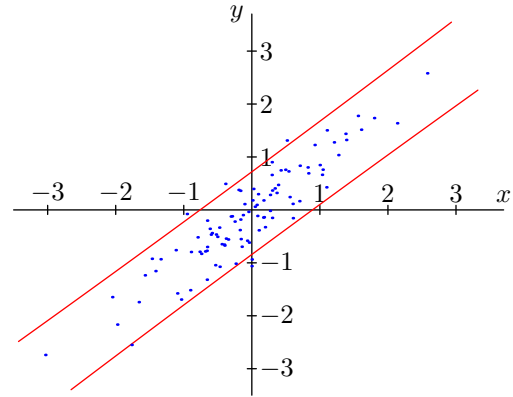
$$\left[\widehat{Y}_0 - t_{n-2, \frac{1+\alpha}{2}} S, \widehat{Y}_0 + t_{n-2, \frac{1+\alpha}{2}} S \right].$$

Lorsque la valeur x_0 varie, les deux extrémités de l'intervalle de prévision décrivent deux courbes d'équations

$$y = \widehat{a} + \widehat{b} x_0 \pm t_{n-2, \frac{1+\alpha}{2}} \sqrt{\frac{\sum_{i=1}^n \widehat{U}_i^2}{n-2}} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}.$$

On montre sans (trop de) peine que ce sont les deux branches d'une hyperbole.

Sur le graphique ci-contre nous avons représenté les données de la page 124 et l'« hyperbole de confiance » délimitant les intervalles de confiance au niveau 95 % pour la prédiction de la variable expliquée.



8.3 Régression linéaire multiple

La régression linéaire multiple généralise la régression linéaire simple : on suppose que la *variable dépendante* Y est fonction de k *variables indépendantes* (x_1, x_2, \dots, x_k) ; on postule le modèle linéaire

$$\forall i \in \{1, 2, \dots, n\}, Y_i = \sum_{j=1}^k b_j x_{i,j} + U_i$$

où les variables aléatoires U_i , qui représentent les *erreurs résiduelles*, vérifient

$$\forall i \in \{1, 2, \dots, n\}, \mathbf{E}(U_i) = 0, \mathbf{Var}(U_i) = \sigma^2 \text{ et, pour } i \neq j, \mathbf{Cov}(U_i, U_j) = 0.$$

On a noté $x_{i,j}$ la valeur de la variable x_j pour l'individu i ; l'ensemble de ces valeurs est regroupé dans une matrice

$$x = \begin{pmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,k} \\ x_{2,1} & x_{2,2} & \dots & x_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \dots & x_{n,k} \end{pmatrix}.$$

La colonne j représente les valeurs de la variable x_j pour tous les individus. On suppose que $n > k$ et que la matrice x est de rang k .

En introduisant les vecteurs

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, U = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix} \text{ et } b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{pmatrix}$$

on peut écrire le modèle sous forme matricielle

$$Y = x b + U \quad \text{avec} \quad \mathbf{E}(U) = 0 \text{ et } \mathbf{Var}(U) = \sigma^2 I_n.$$

Remarque : on obtient un modèle affine (ou « avec constante ») en choisissant $x_{i,1} = 1$ pour tout i ; on a alors

$$\forall i \in \{1, 2, \dots, n\}, Y_i = b_1 + \sum_{j=2}^k b_j x_{i,j} + U_i$$

8.3.1 Estimation des paramètres par les moindres carrés

On ne fait pas d'hypothèse sur la loi de probabilité des erreurs U_i .

Détermination des estimateurs

On minimise la somme des carrés des erreurs :

$$\varphi(b) = \sum_{i=1}^n U_i^2 = \sum_{i=1}^n \left(Y_i - \sum_{j=1}^k b_j x_{i,j} \right)^2$$

Théorème 154. *L'estimateur des moindres carrés de b est*

$$\hat{b} = (x^T x)^{-1} (x^T Y).$$

On a, pour tout ℓ dans $\{1, 2, \dots, k\}$:

$$\frac{\partial \varphi(b)}{\partial b_\ell} = 2 \sum_{i=1}^n \left(Y_i - \sum_{j=1}^k b_j x_{i,j} \right) (-x_{i,\ell}).$$

La condition nécessaire de minimum s'écrit donc

$$\forall \ell \in \{1, 2, \dots, k\}, \sum_{i=1}^n \sum_{j=1}^k x_{i,j} b_j x_{i,\ell} = \sum_{i=1}^n Y_i x_{i,\ell}$$

ou encore

$$\forall \ell \in \{1, 2, \dots, k\}, \sum_{j=1}^k \left(\sum_{i=1}^n x_{i,j} x_{i,\ell} \right) b_j = \sum_{i=1}^n x_{i,\ell} Y_i$$

ce qui s'écrit matriciellement

$$(x^T x)b = x^T Y.$$

On en déduit le résultat annoncé si $(x^T x)$ est inversible, ce que nous supposons.

On a :

$$\forall (\ell, m) \in \{1, 2, \dots, k\}^2, \frac{\partial^2 \varphi(b)}{\partial b_\ell \partial b_m} = 2 \sum_{i=1}^n x_{i,\ell} x_{i,m}.$$

La matrice hessienne est donc égale à $2x^T x$; \hat{b} est un minimum de φ si elle est définie positive, ce que nous supposons.

La valeur ajustée de Y est

$$\hat{Y} = x\hat{b} = x(x^T x)^{-1}x^T Y = pY \quad \text{en posant} \quad p = x(x^T x)^{-1}x^T.$$

Le résidu est le vecteur

$$\hat{U} = Y - \hat{Y} = (I_n - p)Y = qY \quad \text{en posant} \quad q = I_n - p.$$

Lemme 2. p et q sont des projecteurs symétriques.

Le sous-espace propre de p (resp. q) associé à la valeur propre 1 a pour dimension k (resp. $(n - k)$)

1. p est un projecteur :

$$p^2 = (x(x^T x)^{-1}x^T)(x(x^T x)^{-1}x^T) = x(x^T x)^{-1}x^T = p.$$

2. q est un projecteur :

$$q^2 = (I_n - p)^2 = I_n - 2p + p^2 = I_n - p = q.$$

3. p et q sont symétriques :

$$p^T = (x(x^T x)^{-1}x^T)^T = x(x^T x)^{-1}x^T = p$$

et

$$q^T = (I_n - p)^T = I_n - p = q.$$

4. Sous-espace propre E_1 de p associé à la valeur propre 1 ; on a :

$$px = (x(x^T x)^{-1}x^T)x = x,$$

donc toutes les colonnes de la matrice x sont des vecteurs propres de p associés à la valeur propre 1. La dimension du sous-espace propre E_1 est au moins égale à k . Le sous-espace orthogonal à E_1 , ensemble des vecteurs y qui vérifient $x^T y = 0$, est de dimension $(n - k)$; c'est le noyau de p . La dimension de E_1 est égale à k . Donc

$$\text{rg}(p) = \text{tr}(p) = k.$$

Même démonstration pour q .

Propriétés des estimateurs

Théorème 155. \hat{b} est un estimateur sans biais de b et

$$\mathbf{Var}(\hat{b}) = \sigma^2 (x^T x)^{-1}.$$

On a :

$$\hat{b} = (x^T x)^{-1} x^T Y = (x^T x)^{-1} x^T (xb + U) = b + (x^T x)^{-1} x^T U.$$

Donc :

$$\mathbf{E}(\hat{b}) = b + (x^T x)^{-1} x^T \mathbf{E}(U) = b$$

Par ailleurs :

$$\begin{aligned} \mathbf{Var}(\hat{b}) &= \mathbf{E}((\hat{b} - b)(\hat{b} - b)^T) = \mathbf{E}((x^T x)^{-1} x^T U U^T x (x^T x)^{-1}) \\ &= (x^T x)^{-1} x^T \underbrace{\mathbf{E}(U U^T)}_{=\sigma^2 I_n} x (x^T x)^{-1} \\ &= \sigma^2 (x^T x)^{-1}. \end{aligned}$$

Remarque : le théorème 138 est un cas particulier de ce résultat.

Théorème 156. On a :

$$\mathbf{Cov}(\hat{b}, \hat{U}) = 0.$$

Par définition :

$$\mathbf{Cov}(\hat{b}, \hat{U}) = \mathbf{E}((\hat{b} - b)(\hat{U} - \mathbf{E}(\hat{U}))^T).$$

Or :

$$\hat{U} = qY = q(xb + U) = qU, \text{ car } qx = (I_n - p)x = x - px = 0.$$

D'où $\mathbf{E}(\hat{U}) = 0$ et

$$\begin{aligned} \mathbf{Cov}(\hat{b}, \hat{U}) &= \mathbf{E}((\hat{b} - b)\hat{U}^T) = (x^T x)^{-1} x^T \underbrace{\mathbf{E}(U U^T)}_{=\sigma^2 I_n} q^T \\ &= \sigma^2 (x^T x)^{-1} x^T (I_n - p^T) = \sigma^2 ((x^T x)^{-1} x^T - (x^T x)^{-1} x^T x (x^T x)^{-1} x^T) \\ &= 0. \end{aligned}$$

Théorème 157. On a :

$$\mathbf{E}(\hat{U}) = 0 \quad \text{et} \quad \mathbf{Var}(\hat{U}) = \sigma^2 (I_n - x(x^T x)^{-1} x^T)$$

L'espérance a été calculée au cours de la démonstration précédente.

Par ailleurs :

$$\mathbf{Var}(\hat{U}) = \mathbf{E}(\hat{U}\hat{U}^T) = \mathbf{E}(qU U^T q^T) = q \underbrace{\mathbf{E}(U U^T)}_{=\sigma^2 I_n} q^T = \sigma^2 q^2 = \sigma^2 q.$$

Qualité de l'ajustement

1. Modèle affine (avec « terme constant »)

On a l'équation d'analyse de la variance :

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})^2 + \sum_{i=1}^n \hat{U}_i^2.$$

Comme au théorème 139 \hat{Y} est la projection orthogonale de Y sur le sous-espace engendré par les vecteurs e, x_2, x_3, \dots, x_k ; l'équation d'analyse de la variance résulte du théorème de PYTHAGORE.

Le coefficient de détermination est le rapport de la « variance expliquée » à la « variance totale » :

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}.$$

Il est compris entre 0 et 1.

On a :

$$R^2 = 1 - \frac{\sum_{i=1}^n \hat{U}_i^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \leq 1.$$

Ce coefficient de détermination croît avec k ; on « corrige » cette croissance (sans signification statistique) en introduisant le *coefficient de détermination ajusté* :

$$\bar{R}^2 = 1 - \frac{\sum_{i=1}^n \hat{U}_i^2 / (n - k)}{\sum_{i=1}^n (Y_i - \bar{Y})^2 / (n - 1)}$$

On a :

$$\bar{R}^2 \leq R^2.$$

$$\bar{R}^2 - R^2 = (1 - R^2) \left(1 - \frac{n-1}{n-k}\right) = -\frac{k-1}{n-k} (1 - R^2) \leq 0.$$

2. Modèle linéaire (sans « terme constant »)

On n'a plus l'équation d'analyse de la variance, ni l'égalité $\bar{\hat{Y}} = \bar{Y}$.

\hat{Y} est la projection orthogonale de Y sur le sous-espace engendré par les vecteurs x_1, x_2, \dots, x_k dont e ne fait pas partie ; le théorème des trois perpendiculaires ne s'applique plus.

Le coefficient de détermination, que l'on peut calculer par la formule habituelle, n'est pas toujours compris entre 0 et 1, ce qui lui enlève beaucoup d'intérêt.

On a :

$$\sum_{i=1}^n Y_i^2 = \sum_{i=1}^n \hat{Y}_i^2 + \sum_{i=1}^n \hat{U}_i^2.$$

Le triangle (O, Y, \hat{Y}) est rectangle ; on applique le théorème de PYTHAGORE.

Comme indicateur de la qualité de l'ajustement on peut utiliser le rapport :

$$\cos^2 \alpha = \frac{\sum_{i=1}^n \hat{Y}_i^2}{\sum_{i=1}^n Y_i^2} = 1 - \frac{\sum_{i=1}^n \hat{U}_i^2}{\sum_{i=1}^n Y_i^2}.$$

8.3.2 Estimation des paramètres par le maximum de vraisemblance

On suppose que le vecteur aléatoire $U = (U_1, U_2, \dots, U_n)$ est gaussien de loi $\mathcal{N}(0, \sigma^2 I_n)$, c'est-à-dire que les variables aléatoires U_i sont normales, centrées et indépendantes, d'écart-type σ .

Détermination des estimateurs

Théorème 158. *Les estimateurs du maximum de vraisemblance de b et σ^2 sont*

$$\hat{b} = (x^T x)^{-1} x^T Y \quad \text{et} \quad \hat{v} = \frac{1}{n} \sum_{i=1}^n \hat{U}_i^2.$$

La vraisemblance de l'échantillon (Y_1, Y_2, \dots, Y_n) est, en posant $v = \sigma^2$,

$$\mathcal{L}(b, v) = (2\pi v)^{-\frac{n}{2}} \exp\left(-\frac{1}{2v} \sum_{i=1}^n \left(Y_i - \sum_{j=1}^k x_{i,j} b_j\right)^2\right).$$

On pose

$$\psi(b, v) = \ln(\mathcal{L}(b, v)) = -\frac{n}{2} \ln(2\pi v) - \frac{1}{2v} \sum_{i=1}^n \left(Y_i - \sum_{j=1}^k x_{i,j} b_j\right)^2.$$

Les équations de vraisemblance s'écrivent :

$$\begin{aligned} \forall \ell \in \{1, 2, \dots, k\}, \quad \frac{\partial \psi(b, v)}{\partial b_\ell} &= \frac{1}{v} \sum_{i=1}^n \left(Y_i - \sum_{j=1}^k x_{i,j} b_j\right) x_{i,\ell} = 0, \\ \frac{\partial \psi(b, v)}{\partial v} &= -\frac{n}{2v} + \frac{1}{2v^2} \sum_{i=1}^n \left(Y_i - \sum_{j=1}^k x_{i,j} b_j\right)^2 = 0. \end{aligned}$$

Les premières équations donnent les équations normales de la méthode des moindres carrés ; on en déduit donc l'estimateur des moindres carrés de b :

$$\hat{b} = (x^T x)^{-1} x^T Y.$$

La dernière équation fournit l'estimateur du maximum de vraisemblance de v :

$$\hat{v} = \frac{1}{n} \sum_{i=1}^n \left(Y_i - \sum_{j=1}^k x_{i,j} b_j\right)^2 = \frac{1}{n} \sum_{i=1}^n \hat{U}_i^2.$$

Théorème 159. *Un estimateur sans biais de σ^2 est*

$$\widehat{\sigma^2} = \frac{1}{n-k} \sum_{i=1}^n \hat{U}_i^2.$$

On a :

$$\sum_{i=1}^n \hat{U}_i^2 = \hat{U}^T \hat{U} = (qU)^T (qU) = U^T q^T q U = U^T q U = \sum_{i=1}^n \sum_{j=1}^n q_{i,j} U_i U_j.$$

D'où :

$$\begin{aligned} \mathbf{E}\left(\sum_{i=1}^n \hat{U}_i^2\right) &= \sum_{i=1}^n \sum_{j=1}^n q_{i,j} \underbrace{\mathbf{E}(U_i U_j)}_{\begin{cases} \sigma^2 & \text{si } j = i, \\ 0 & \text{sinon.} \end{cases}} = \sigma^2 \sum_{i=1}^n q_{i,i} = \sigma^2 \text{tr}(q) = (n - k) \sigma^2. \end{aligned}$$

Donc :

$$\mathbf{E}(\hat{v}) = \frac{n - k}{n} \sigma^2.$$

L'estimateur sans biais de σ^2 déduit de l'estimateur du maximum de vraisemblance est

$$\hat{\sigma}^2 = \frac{n}{n - k} \hat{v} = \frac{1}{n - k} \sum_{i=1}^n \hat{U}_i^2.$$

Lois des estimateurs

Théorème 160. L'estimateur \hat{b} de b est gaussien de loi $\mathcal{N}(b, \sigma^2(x^T x)^{-1})$.

\hat{b} est linéaire en Y , donc en U qui est gaussien : d'après le corollaire 3 il est gaussien. Nous avons déjà calculé son espérance mathématique et sa matrice de covariance (voir théorème 155).

Théorème 161. L'estimateur $\hat{\sigma}^2$ de σ^2 est tel que $(n - k) \frac{\hat{\sigma}^2}{\sigma^2}$ suit la loi du khi-deux à $(n - k)$ degrés de liberté.

Nous avons montré (démonstration du théorème 159) que $\sum_{i=1}^n \hat{U}_i^2 = U^T q U$. Or q , symétrique, est diagonalisable et ses valeurs propres sont 1, d'ordre $(n - k)$, et 0, d'ordre k . Il existe donc une matrice orthogonale ω telle que

$$\omega^{-1} q \omega = \omega^T q \omega = d = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix}.$$

où la diagonale contient $(n - k)$ fois 1 et k fois 0. Alors :

$$\sum_{i=1}^n \hat{U}_i^2 = U^T \omega d \omega^T U = Z^T d Z = \sum_{i=1}^{n-k} Z_i^2 \quad \text{où} \quad Z = \omega^T U.$$

D'après le corollaire 3 Z est gaussien, de loi $\mathcal{N}(0, \sigma^2 I_n)$. Donc, pour tout i , le rapport Z_i/σ suit la loi normale réduite et

$$(n - k) \frac{\hat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n \hat{U}_i^2 = \sum_{i=1}^{n-k} \left(\frac{Z_i}{\sigma}\right)^2$$

suit la loi du khi-deux à $(n - k)$ degrés de liberté.

Tests sur les paramètres

Théorème 162 (Test bilatéral sur un paramètre). Pour une valeur de j comprise entre 1 et k on teste l'hypothèse nulle

$$H_0 : b_j = b_j^0 \quad \text{contre l'hypothèse alternative} \quad H_1 : b_j \neq b_j^0.$$

On rejette H_0 avec un risque α si

$$\frac{|\hat{b}_j - b_j^0|}{S\sqrt{v_j}} > t_{n-k, 1-\frac{\alpha}{2}}$$

où

1. $S = \sqrt{\widehat{\sigma^2}}$;
2. v_j est le j -ième terme de la diagonale de la matrice $(x^T x)^{-1}$;
3. $t_{n-k, 1-\frac{\alpha}{2}}$ est le fractile d'ordre $(1 - \frac{\alpha}{2})$ de la loi de STUDENT à $(n - k)$ degrés de liberté.

On sait (théorème 160) que, sous H_0 , le rapport $\frac{\hat{b}_j - b_j^0}{\sigma\sqrt{v_j}}$ suit la loi normale réduite. Par ailleurs (théorème 161) $(n - k) \frac{\widehat{\sigma^2}}{\sigma^2}$ suit la loi du khi-deux à $(n - k)$ degrés de liberté. Donc $\frac{\hat{b}_j - b_j^0}{S\sqrt{v_j}}$ suit la loi de STUDENT à $(n - k)$ degrés de liberté. La règle de décision s'en déduit immédiatement.

Pour $b_j^0 = 0$ on teste si la présence de b_j est significative dans le modèle.

Théorème 163 (Test sur l'ensemble des paramètres). On teste l'hypothèse nulle

$$H_0 : b = b^0 \in \mathbb{R}^k \quad \text{contre l'hypothèse alternative} \quad H_1 : b \neq b^0.$$

On rejette H_0 avec un risque α si

$$\frac{(\hat{b} - b^0)^T (x^T x) (\hat{b} - b^0)}{k\widehat{\sigma^2}} > f_{k, n-k, 1-\alpha}$$

où $f_{k, n-k, 1-\alpha}$ est le fractile d'ordre $(1 - \alpha)$ de la loi de FISHER-SNEDECOR à k et $(n - k)$ degrés de liberté $\mathcal{F}(k, n - k)$.

1. Sous H_0 le vecteur \hat{b} est gaussien, de loi $\mathcal{N}(b^0, \sigma^2(x^T x)^{-1})$; donc, d'après le théorème 82, $(\hat{b} - b^0)^T \frac{x^T x}{\sigma^2} (\hat{b} - b^0)$ suit la loi du khi-deux à k degrés de liberté.
2. On sait (théorème 161) que $(n - k) \frac{\widehat{\sigma^2}}{\sigma^2}$ suit la loi du khi-deux à $(n - k)$ degrés de liberté.
3. Le vecteur $(\hat{b}, \hat{U}) = (b + (x^T x)^{-1} x^T U, qU)$ s'exprime de façon affine en fonction de U qui est gaussien; donc (corollaire 3) il est gaussien. D'après le théorème 68, puisque $\text{Cov}(\hat{b}, \hat{U}) = 0$, \hat{b} et \hat{U} sont indépendantes et donc également \hat{b} et $\widehat{\sigma^2} = \frac{1}{n-k} \hat{U}^T \hat{U}$.

D'après la définition 35 le rapport

$$\frac{(\hat{b} - b^0)^T (x^T x) (\hat{b} - b^0) / k \sigma^2}{\widehat{\sigma^2} / \sigma^2} = \frac{(\hat{b} - b^0)^T (x^T x) (\hat{b} - b^0)}{k \widehat{\sigma^2}}$$

suit la loi de FISHER-SNEDECOR à k et $(n - k)$ degrés de liberté $\mathcal{F}(k, n - k)$. La règle de décision s'en déduit.

Lorsque $b^0 = 0$ on teste si la présence de b est significative dans le modèle.

Annexe A

Dénombrements

Quelques techniques de dénombrement élémentaires sont indispensables dans les calculs de probabilités discrètes.

A.1 Suites. Arrangements sans répétition

Théorème 164. Soient A_1, A_2, \dots, A_k des ensembles ayant respectivement n_1, n_2, \dots, n_k éléments ; le nombre de suites (a_1, a_2, \dots, a_k) telles que, pour tout i , l'élément a_i appartient à A_i est égal à

$$\prod_{i=1}^k n_i = n_1 n_2 \dots n_k.$$

Récurrance sur k .

Ces suites sont les éléments, appelés aussi k -uplets, de l'ensemble produit $A_1 \times A_2 \times \dots \times A_k$.

Corollaire 6. Soit A un ensemble ayant n éléments ; le nombre de suites de k éléments de A est égal à n^k .

C'est le cardinal de l'ensemble produit A^k .

C'est aussi le nombre d'applications d'un ensemble de k éléments dans un ensemble de n éléments.

Corollaire 7. Soit A un ensemble ayant n éléments ; le nombre de suites de k éléments distincts de A est égal à

$$n(n-1)(n-2) \dots (n-k+1) = \frac{n!}{(n-k)!}$$

que l'on note $(n)_k$ ou bien A_n^k .

Une telle suite est un *arrangement sans répétition* de k éléments pris parmi n .

$(n)_k$ est le nombre d'applications injectives d'un ensemble de k éléments dans un ensemble de n éléments.

Lorsque $n = k$ on obtient le nombre de bijections entre deux ensembles ayant n éléments (ou de permutations d'un ensemble ayant n éléments) : $(n)_n = n!$.

A.2 Combinaisons sans répétition

Théorème 165. *Le nombre de sous-ensembles ayant k éléments d'un ensemble de n éléments est égal à*

$$\frac{n!}{k!(n-k)!}$$

que l'on note $\binom{n}{k}$ ou bien C_n^k .

Pour chacun de ces sous-ensembles on obtient en permutant les k éléments $k!$ arrangements sans répétition : le nombre de sous-ensembles cherché est donc égal à $\frac{(n)_k}{k!}$.

Par convention $\binom{n}{k} = 0$ si $k \notin \{0, 1, \dots, n\}$.

Théorème 166. *Quelques propriétés élémentaires des coefficients du binôme :*

- $\binom{n}{k} = \binom{n}{n-k}$;
- $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$;
- $0 < k < n \Rightarrow \binom{n}{k} = \frac{n}{k} \binom{n-1}{k-1} = \frac{n}{n-k} \binom{n-1}{k} = \frac{n-k+1}{k} \binom{n}{k-1}$;
- $(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$.

A.3 Arrangements avec répétition

Théorème 167. *Soit A un ensemble ayant n éléments et soit (r_1, r_2, \dots, r_k) une suite d'entiers positifs ou nuls dont la somme est égale à n ; le nombre de partitions de A en k sous-ensembles A_1, A_2, \dots, A_k tels que, pour tout i , A_i a r_i éléments est*

$$\frac{n!}{r_1! r_2! \dots r_k!}.$$

Réurrence sur k .

Définition 59. *Soit $A = \{a_1, a_2, \dots, a_k\}$ et soit (r_1, r_2, \dots, r_k) une suite d'entiers positifs ou nuls dont la somme est égale à n ; un arrangement avec répétition des a_i associé aux facteurs de répétition r_i est une suite d'éléments de A où l'on trouve r_i fois a_i .*

L'ordre des éléments est important : les suites (b, a, c, a, c, c) et (a, b, c, c, a, c) sont différentes bien qu'elles soient formées des mêmes éléments.

Théorème 168. *Le nombre d'arrangements avec répétition associés à la suite (r_1, r_2, \dots, r_k) est égal à*

$$\frac{(r_1 + r_2 + \dots + r_k)!}{r_1! r_2! \dots r_k!}.$$

Soit (x_1, x_2, \dots, x_n) un tel arrangement avec répétition. En posant $j \in A_i \Leftrightarrow x_j = a_i$ on définit une bijection de l'ensemble des arrangements avec répétition associés à la suite (r_1, r_2, \dots, r_k) dans l'ensemble des partitions de $\{1, 2, \dots, n\}$ en k sous-ensembles A_1, A_2, \dots, A_k ayant r_1, r_2, \dots, r_k éléments. On est ramené au théorème précédent.

Il y a, par exemple $\frac{6!}{2!1!3!} = 60$ suites contenant deux fois a , une fois b et trois fois c .

Annexe B

Fonctions eulériennes

Les fonctions eulériennes interviennent dans des domaines nombreux et variés ; elles sont très utiles, et même indispensables, dans l'étude de certaines lois de probabilité.

Définition 60.

– La fonction eulérienne de première espèce

$$B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt$$

est définie pour x et y réels strictement positifs.

– La fonction eulérienne de deuxième espèce

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$$

est définie pour x strictement positif.

Théorème 169.

$$\forall x > 0, \Gamma(x+1) = x \Gamma(x).$$

Théorème 170.

$$\forall n \in \mathbb{N}^*, \Gamma(n) = (n-1)!$$

Théorème 171.

$$B(x, y) = 2 \int_0^{\frac{\pi}{2}} \cos^{2x-1} \theta \sin^{2y-1} \theta d\theta.$$

On effectue le changement de variable défini par $\theta = \arccos \sqrt{t}$.

Théorème 172.

$$B(x, y) = \frac{\Gamma(x) \Gamma(y)}{\Gamma(x+y)}.$$

En effectuant le changement de variable défini par $u = \sqrt{t}$ dans la définition on obtient

$$\Gamma(x) = 2 \int_0^\infty e^{-u^2} u^{2x-1} du$$

et de la même façon

$$\Gamma(y) = 2 \int_0^\infty e^{-v^2} v^{2y-1} dv.$$

Posant $\Delta = (\mathbb{R}^{+*})^2$ et $D =]0, \pi/2[\times \mathbb{R}^{+*}$ on obtient

$$\begin{aligned} \Gamma(x)\Gamma(y) &= 4 \iint_{\Delta} e^{-(u^2+v^2)} u^{2x-1} v^{2y-1} du dv \\ &= 4 \iint_D e^{-r^2} (r \cos \theta)^{2x-1} (r \sin \theta)^{2y-1} r dr d\theta \end{aligned}$$

en passant en coordonnées polaires ; d'où :

$$\begin{aligned} \Gamma(x)\Gamma(y) &= 4 \int_0^{\pi/2} \cos^{2x-1} \theta \sin^{2y-1} \theta d\theta \underbrace{\int_0^\infty e^{-r^2} r^{2(x+y)-1} dr}_{\frac{1}{2}\Gamma(x+y)} \\ &= B(x, y)\Gamma(x+y) \end{aligned}$$

Théorème 173 (Intégrales de GAUSS).

$$\int_{-\infty}^\infty e^{-x^2} dx = \sqrt{\pi}, \quad \int_{-\infty}^\infty x^2 e^{-x^2} dx = \frac{\sqrt{\pi}}{2}.$$

On a :

$$B\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2})}{\Gamma(1)} = \left(\Gamma\left(\frac{1}{2}\right)\right)^2 = 2 \int_0^{\pi/2} d\theta = \pi.$$

Donc : $\Gamma(1/2) = \sqrt{\pi}$. On en déduit le théorème en effectuant le changement de variables défini par $x = \sqrt{t}$.

Terminons par quelques formules données sans démonstration.

– Formule des compléments

$$\forall x \in]0, 1[, \Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin \pi x}.$$

– Formule de STIRLING

Au voisinage de l'infini

$$\Gamma(x) \sim \sqrt{2\pi} x^{x-\frac{1}{2}} e^{-x}.$$

Plus précisément :

$$\ln(\Gamma(x)) = x \ln x - x - \frac{1}{2} \ln x + \frac{1}{2} \ln(2\pi) + \sum_{k=1}^p \frac{b_{2k}}{2k(2k-1)} \frac{1}{x^{2k-1}} + O\left(\frac{1}{x^{2p}}\right)$$

où les b_{2k} sont les nombres de BERNOULLI définis par

$$\frac{x}{e^x - 1} = \sum_{n=0}^\infty b_n \frac{x^n}{n!}.$$

– Formule de GAUSS

$$\forall x > 0, \Gamma(x) = \lim_{n \rightarrow \infty} \frac{n^x n!}{x(x+1) \cdots (x+n)}.$$

– Formule de WEIERSTRASS

$$\forall x > 0, \Gamma(x) = \frac{e^{-\gamma x}}{x} \prod_{n=1}^{\infty} \frac{e^{\frac{x}{n}}}{1 + \frac{x}{n}}$$

où γ , constante d'EULER, est définie par $\gamma = \lim_{n \rightarrow \infty} \left(\sum_{k=1}^n \frac{1}{k} - \ln n \right)$.

– Formule de LEGENDRE-GAUSS

$$\forall z \in \mathbb{C} - \mathbb{Z}^-, \forall p \in \mathbb{N}^*, \Gamma\left(\frac{z}{p}\right) \Gamma\left(\frac{z+1}{p}\right) \cdots \Gamma\left(\frac{z+p-1}{p}\right) = (2\pi)^{\frac{p-1}{2}} p^{\frac{1}{2}-z} \Gamma(z).$$

– Considérons la dérivée logarithmique de Γ :

$$\forall x \in \mathbb{R}^{+*}, \psi(x) = \frac{d(\ln(\Gamma(x)))}{dx} = \frac{\Gamma'(x)}{\Gamma(x)}.$$

On a les propriétés suivantes :

$$\psi(x) = -\gamma - \frac{1}{x} + \sum_{n=1}^{\infty} \left(\frac{1}{n} - \frac{1}{n+x} \right).$$

$$\psi^{(k-1)}(x) = (-1)^k (k-1)! \sum_{n=0}^{\infty} \frac{1}{(x+n)^k}, \quad (k \geq 2).$$

$$\psi(x) = -\gamma + \int_0^1 \frac{1 - (1-t)^{x-1}}{t} dt, \quad (\text{GAUSS}).$$

$$\int_x^{x+1} \ln(\Gamma(t)) dt = x(\ln x - 1) + \frac{1}{2} \ln(2\pi), \quad (\text{RAABE}).$$

Annexe C

Tables numériques

Ces tables ont été établies à l'aide du logiciel *Mathematica*.

- Loi normale
- Loi du khi-deux
- Loi de STUDENT
- Loi de FISHER-SNEDECOR

C.1 Loi normale

Fonction de répartition de la loi normale réduite : $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.	0.5	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861

Table de 1 – $\Phi(x)$ pour les grandes valeurs de x

x	0.	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
3.	{1.35 ; 3}	{9.68 ; 4}	{6.87 ; 4}	{4.83 ; 4}	{3.37 ; 4}	{2.33 ; 4}	{1.59 ; 4}	{1.08 ; 4}	{7.23 ; 5}	{4.81 ; 5}
4.	{3.17 ; 5}	{2.07 ; 5}	{1.33 ; 5}	{8.54 ; 6}	{5.41 ; 6}	{3.40 ; 6}	{2.11 ; 6}	{1.30 ; 6}	{7.93 ; 7}	{4.79 ; 7}
5.	{2.87 ; 7}	{1.70 ; 7}	{9.96 ; 8}	{5.79 ; 8}	{3.33 ; 8}	{1.90 ; 8}	{1.07 ; 8}	{5.99 ; 9}	{3.32 ; 9}	{1.82 ; 9}

N.B. La notation {m ; k} signifie $m \cdot 10^{-k}$.

Densité de la loi normale réduite : $\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$

x	0.	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.	0.39894	0.39892	0.39886	0.39876	0.39862	0.39844	0.39822	0.39797	0.39767	0.39733
0.1	0.39695	0.39654	0.39608	0.39559	0.39505	0.39448	0.39387	0.39322	0.39253	0.39181
0.2	0.39104	0.39024	0.38940	0.38853	0.38762	0.38667	0.38568	0.38466	0.38361	0.38251
0.3	0.38139	0.38023	0.37903	0.37780	0.37654	0.37524	0.37391	0.37255	0.37115	0.36973
0.4	0.36827	0.36678	0.36526	0.36371	0.36213	0.36053	0.35889	0.35723	0.35553	0.35381
0.5	0.35207	0.35029	0.34849	0.34667	0.34482	0.34294	0.34105	0.33912	0.33718	0.33521
0.6	0.33322	0.33121	0.32918	0.32713	0.32506	0.32297	0.32086	0.31874	0.31659	0.31443
0.7	0.31225	0.31006	0.30785	0.30563	0.30339	0.30114	0.29887	0.29659	0.29431	0.29200
0.8	0.28969	0.28737	0.28504	0.28269	0.28034	0.27798	0.27562	0.27324	0.27086	0.26848
0.9	0.26609	0.26369	0.26129	0.25888	0.25647	0.25406	0.25164	0.24923	0.24681	0.24439
1.	0.24197	0.23955	0.23713	0.23471	0.23230	0.22988	0.22747	0.22506	0.22265	0.22025
1.1	0.21785	0.21546	0.21307	0.21069	0.20831	0.20594	0.20357	0.20121	0.19886	0.19652
1.2	0.19419	0.19186	0.18954	0.18724	0.18494	0.18265	0.18037	0.17810	0.17585	0.17360
1.3	0.17137	0.16915	0.16694	0.16474	0.16256	0.16038	0.15822	0.15608	0.15395	0.15183
1.4	0.14973	0.14764	0.14556	0.14350	0.14146	0.13943	0.13742	0.13542	0.13344	0.13147
1.5	0.12952	0.12758	0.12566	0.12376	0.12188	0.12001	0.11816	0.11632	0.11450	0.11270
1.6	0.11092	0.10915	0.10741	0.10567	0.10396	0.10226	0.10059	0.09893	0.09728	0.09566
1.7	0.09405	0.09246	0.09089	0.08933	0.08780	0.08628	0.08478	0.08329	0.08183	0.08038
1.8	0.07895	0.07754	0.07614	0.07477	0.07341	0.07206	0.07074	0.06943	0.06814	0.06687
1.9	0.06562	0.06438	0.06316	0.06195	0.06077	0.05959	0.05844	0.05730	0.05618	0.05508
2.	0.05399	0.05292	0.05186	0.05082	0.04980	0.04879	0.04780	0.04682	0.04586	0.04491
2.1	0.04398	0.04307	0.04217	0.04128	0.04041	0.03955	0.03871	0.03788	0.03706	0.03626
2.2	0.03547	0.03470	0.03394	0.03319	0.03246	0.03174	0.03103	0.03034	0.02965	0.02898
2.3	0.02833	0.02768	0.02705	0.02643	0.02582	0.02522	0.02463	0.02406	0.02349	0.02294
2.4	0.02239	0.02186	0.02134	0.02083	0.02033	0.01984	0.01936	0.01888	0.01842	0.01797
2.5	0.01753	0.01709	0.01667	0.01625	0.01585	0.01545	0.01506	0.01468	0.01431	0.01394
2.6	0.01358	0.01323	0.01289	0.01256	0.01223	0.01191	0.01160	0.01130	0.01100	0.01071
2.7	0.01042	0.01014	0.00987	0.00961	0.00935	0.00909	0.00885	0.00861	0.00837	0.00814
2.8	0.00792	0.00770	0.00748	0.00727	0.00707	0.00687	0.00668	0.00649	0.00631	0.00613
2.9	0.00595	0.00578	0.00562	0.00545	0.00530	0.00514	0.00499	0.00485	0.00470	0.00457

Table de $\varphi(x)$ pour les grandes valeurs de x

x	0.	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
3.	{4.43 ; 3}	{3.27 ; 3}	{2.38 ; 3}	{1.72 ; 3}	{1.23 ; 3}	{8.73 ; 4}	{6.12 ; 4}	{4.25 ; 4}	{2.92 ; 4}	{1.99 ; 4}
4.	{1.34 ; 4}	{8.93 ; 5}	{5.89 ; 5}	{3.85 ; 5}	{2.49 ; 5}	{1.60 ; 5}	{1.01 ; 5}	{6.37 ; 6}	{3.96 ; 6}	{2.44 ; 6}
5.	{1.49 ; 6}	{8.97 ; 7}	{5.36 ; 7}	{3.17 ; 7}	{1.86 ; 7}	{1.08 ; 7}	{6.18 ; 8}	{3.51 ; 8}	{1.98 ; 8}	{1.10 ; 8}

N.B. La notation {m ; k} signifie $m \cdot 10^{-k}$.

C.2 Loi du khi-deux

Fractiles de la loi du khi-deux

$\nu \setminus p$	0.0005	0.001	0.005	0.01	0.025	0.05	0.1	0.5
1	0.000	0.000	0.000	0.000	0.001	0.004	0.016	0.455
2	0.001	0.002	0.010	0.020	0.051	0.103	0.211	1.386
3	0.015	0.024	0.072	0.115	0.216	0.352	0.584	2.366
4	0.064	0.091	0.207	0.297	0.484	0.711	1.064	3.357
5	0.158	0.210	0.412	0.554	0.831	1.145	1.610	4.351
6	0.299	0.381	0.676	0.872	1.237	1.635	2.204	5.348
7	0.485	0.598	0.989	1.239	1.690	2.167	2.833	6.346
8	0.710	0.857	1.344	1.646	2.180	2.733	3.490	7.344
9	0.972	1.152	1.735	2.088	2.700	3.325	4.168	8.343
10	1.265	1.479	2.156	2.558	3.247	3.940	4.865	9.342
11	1.587	1.834	2.603	3.053	3.816	4.575	5.578	10.341
12	1.934	2.214	3.074	3.571	4.404	5.226	6.304	11.340
13	2.305	2.617	3.565	4.107	5.009	5.892	7.042	12.340
14	2.697	3.041	4.075	4.660	5.629	6.571	7.790	13.339
15	3.108	3.483	4.601	5.229	6.262	7.261	8.547	14.339
16	3.536	3.942	5.142	5.812	6.908	7.962	9.312	15.338
17	3.980	4.416	5.697	6.408	7.564	8.672	10.085	16.338
18	4.439	4.905	6.265	7.015	8.231	9.390	10.865	17.338
19	4.912	5.407	6.844	7.633	8.907	10.117	11.651	18.338
20	5.398	5.921	7.434	8.260	9.591	10.851	12.443	19.337
21	5.896	6.447	8.034	8.897	10.283	11.591	13.240	20.337
22	6.404	6.983	8.643	9.542	10.982	12.338	14.041	21.337
23	6.924	7.529	9.260	10.196	11.689	13.091	14.848	22.337
24	7.453	8.085	9.886	10.856	12.401	13.848	15.659	23.337
25	7.991	8.649	10.520	11.524	13.120	14.611	16.473	24.337
26	8.538	9.222	11.160	12.198	13.844	15.379	17.292	25.336
27	9.093	9.803	11.808	12.879	14.573	16.151	18.114	26.336
28	9.656	10.391	12.461	13.565	15.308	16.928	18.939	27.336
29	10.227	10.986	13.121	14.256	16.047	17.708	19.768	28.336
30	10.804	11.588	13.787	14.953	16.791	18.493	20.599	29.336

Le fractile $u_{\nu,p}$ est tel que $\mathbb{P}(\chi_\nu^2 < u_{\nu,p}) = p$.

Lorsque le nombre de degrés de liberté, ν , est supérieur à 30 on peut admettre que les variables aléatoires

$$\sqrt{2\chi_\nu^2} - \sqrt{2\nu - 1} \quad \text{et} \quad \sqrt{\frac{9\nu}{2}} \left(\sqrt{\frac{\chi_\nu^2}{n}} - \left(1 - \frac{2}{9n}\right) \right)$$

suivent approximativement la loi normale réduite (voir théorèmes 84 et 85).

Fractiles de la loi du khi-deux (suite)

$\nu \backslash p$	0.5	0.9	0.95	0.975	0.99	0.995	0.999	0.9995
1	0.455	2.706	3.841	5.024	6.635	7.879	10.828	12.116
2	1.386	4.605	5.991	7.378	9.210	10.597	13.816	15.202
3	2.366	6.251	7.815	9.348	11.345	12.838	16.266	17.730
4	3.357	7.779	9.488	11.143	13.277	14.860	18.467	19.997
5	4.351	9.236	11.070	12.833	15.086	16.750	20.515	22.105
6	5.348	10.645	12.592	14.449	16.812	18.548	22.458	24.103
7	6.346	12.017	14.067	16.013	18.475	20.278	24.322	26.018
8	7.344	13.362	15.507	17.535	20.090	21.955	26.124	27.868
9	8.343	14.684	16.919	19.023	21.666	23.589	27.877	29.666
10	9.342	15.987	18.307	20.483	23.209	25.188	29.588	31.420
11	10.341	17.275	19.675	21.920	24.725	26.757	31.264	33.137
12	11.340	18.549	21.026	23.337	26.217	28.300	32.909	34.821
13	12.340	19.812	22.362	24.736	27.688	29.819	34.528	36.478
14	13.339	21.064	23.685	26.119	29.141	31.319	36.123	38.109
15	14.339	22.307	24.996	27.488	30.578	32.801	37.697	39.719
16	15.338	23.542	26.296	28.845	32.000	34.267	39.252	41.308
17	16.338	24.769	27.587	30.191	33.409	35.718	40.790	42.879
18	17.338	25.989	28.869	31.526	34.805	37.156	42.312	44.434
19	18.338	27.204	30.144	32.852	36.191	38.582	43.820	45.973
20	19.337	28.412	31.410	34.170	37.566	39.997	45.315	47.498
21	20.337	29.615	32.671	35.479	38.932	41.401	46.797	49.011
22	21.337	30.813	33.924	36.781	40.289	42.796	48.268	50.511
23	22.337	32.007	35.172	38.076	41.638	44.181	49.728	52.000
24	23.337	33.196	36.415	39.364	42.980	45.559	51.179	53.479
25	24.337	34.382	37.652	40.646	44.314	46.928	52.620	54.947
26	25.336	35.563	38.885	41.923	45.642	48.290	54.052	56.407
27	26.336	36.741	40.113	43.195	46.963	49.645	55.476	57.858
28	27.336	37.916	41.337	44.461	48.278	50.993	56.892	59.300
29	28.336	39.087	42.557	45.722	49.588	52.336	58.301	60.735
30	29.336	40.256	43.773	46.979	50.892	53.672	59.703	62.162

C.3 Loi de STUDENT

Fractiles de la loi de STUDENT

p	0.5	0.6	0.7	0.8	0.9	0.95	0.975	0.99	0.995	0.999	0.9995
1	0.	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657	318.309	636.619
2	0.	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.	0.254	0.526	0.846	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.	0.254	0.526	0.845	1.290	1.660	1.984	2.364	2.626	3.174	3.390
200	0.	0.254	0.525	0.843	1.286	1.653	1.972	2.345	2.601	3.131	3.340
∞	0.	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Le fractile $t_{\nu,p}$ est tel que $\mathbb{P}(T_\nu < t_{\nu,p}) = p$.

C.4 Loi de FISHER-SNEDECOR

Fractiles de la loi de FISHER-SNEDECOR

$$p = 0.95$$

$n_2 \backslash n_1$	1	2	3	4	5	6	7	8	10	12	20	40	60	100	∞
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.40	19.41	19.45	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.79	8.74	8.66	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.96	5.91	5.80	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.74	4.68	4.56	4.46	4.43	4.41	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.06	4.00	3.87	3.77	3.74	3.71	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.64	3.57	3.44	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.35	3.28	3.15	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.14	3.07	2.94	2.83	2.79	2.76	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	2.98	2.91	2.77	2.66	2.62	2.59	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.85	2.79	2.65	2.53	2.49	2.46	2.41
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.75	2.69	2.54	2.43	2.38	2.35	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.67	2.60	2.46	2.34	2.30	2.26	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.60	2.53	2.39	2.27	2.22	2.19	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.54	2.48	2.33	2.20	2.16	2.12	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.49	2.42	2.28	2.15	2.11	2.07	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.45	2.38	2.23	2.10	2.06	2.02	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.41	2.34	2.19	2.06	2.02	1.98	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.38	2.31	2.16	2.03	1.98	1.94	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.35	2.28	2.12	1.99	1.95	1.91	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.32	2.25	2.10	1.96	1.92	1.88	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.30	2.23	2.07	1.94	1.89	1.85	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.27	2.20	2.05	1.91	1.86	1.82	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.25	2.18	2.03	1.89	1.84	1.80	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.24	2.16	2.01	1.87	1.82	1.78	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.22	2.15	1.99	1.85	1.80	1.76	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.20	2.13	1.97	1.84	1.79	1.74	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.19	2.12	1.96	1.82	1.77	1.73	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.18	2.10	1.94	1.81	1.75	1.71	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.16	2.09	1.93	1.79	1.74	1.70	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.08	2.0	1.84	1.69	1.64	1.59	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	1.99	1.92	1.75	1.59	1.53	1.48	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.91	1.83	1.66	1.50	1.43	1.37	1.26
∞	3.84	3.00	2.61	2.37	2.21	2.10	2.01	1.94	1.83	1.75	1.57	1.40	1.32	1.25	1.03

Le fractile $f_{n_1, n_2, p}$ est tel que $\mathbb{P}(Z_{n_1, n_2} < f_{n_1, n_2, p}) = p$.

Fractiles de la loi de FISHER-SNEDECOR (suite)

$$p = 0.975$$

$n_2 \backslash n_1$	1	2	3	4	5	6	7	8	10	12	20	40	60	100	∞
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.40	39.41	39.45	39.47	39.48	39.49	39.50
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.42	14.34	14.17	14.04	13.99	13.96	13.90
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.84	8.75	8.56	8.41	8.36	8.32	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.62	6.52	6.33	6.18	6.12	6.08	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.46	5.37	5.17	5.01	4.96	4.92	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.76	4.67	4.47	4.31	4.25	4.21	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.30	4.20	4.00	3.84	3.78	3.74	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	3.96	3.87	3.67	3.51	3.45	3.40	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.72	3.62	3.42	3.26	3.20	3.15	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.53	3.43	3.23	3.06	3.00	2.96	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.37	3.28	3.07	2.91	2.85	2.80	2.73
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.25	3.15	2.95	2.78	2.72	2.67	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.15	3.05	2.84	2.67	2.61	2.56	2.49
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.06	2.96	2.76	2.59	2.52	2.47	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	2.99	2.89	2.68	2.51	2.45	2.40	2.32
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.92	2.82	2.62	2.44	2.38	2.33	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.87	2.77	2.56	2.38	2.32	2.27	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.82	2.72	2.51	2.33	2.27	2.22	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.77	2.68	2.46	2.29	2.22	2.17	2.09
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.73	2.64	2.42	2.25	2.18	2.13	2.04
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.70	2.60	2.39	2.21	2.14	2.09	2.00
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.67	2.57	2.36	2.18	2.11	2.06	1.97
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.64	2.54	2.33	2.15	2.08	2.02	1.94
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.61	2.51	2.30	2.12	2.05	2.00	1.91
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.59	2.49	2.28	2.09	2.03	1.97	1.88
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.57	2.47	2.25	2.07	2.00	1.94	1.85
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.55	2.45	2.23	2.05	1.98	1.92	1.83
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.53	2.43	2.21	2.03	1.96	1.90	1.81
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.51	2.41	2.20	2.01	1.94	1.88	1.79
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.39	2.29	2.07	1.88	1.80	1.74	1.64
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.27	2.17	1.94	1.74	1.67	1.60	1.48
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.16	2.05	1.82	1.61	1.53	1.45	1.31
∞	5.03	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.05	1.95	1.71	1.49	1.39	1.30	1.04

Fractiles de la loi de FISHER-SNEDECOR (suite et fin)

$$p = 0.99$$

$n_2 \backslash n_1$	1	2	3	4	5	6	7	8	10	12	20	40	60	100	∞
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.40	99.42	99.45	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.23	27.05	26.69	26.41	26.32	26.24	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.55	14.37	14.02	13.75	13.65	13.58	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.05	9.89	9.55	9.29	9.20	9.13	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.87	7.72	7.40	7.14	7.06	6.99	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.62	6.47	6.16	5.91	5.82	5.75	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.81	5.67	5.36	5.12	5.03	4.96	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.26	5.11	4.81	4.57	4.48	4.41	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.85	4.71	4.41	4.17	4.08	4.01	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.54	4.40	4.10	3.86	3.78	3.71	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.30	4.16	3.86	3.62	3.54	3.47	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.10	3.96	3.66	3.43	3.34	3.27	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	3.94	3.80	3.51	3.27	3.18	3.11	3.01
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.80	3.67	3.37	3.13	3.05	2.98	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.69	3.55	3.26	3.02	2.93	2.86	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.59	3.46	3.16	2.92	2.83	2.76	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.51	3.37	3.08	2.84	2.75	2.68	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.43	3.30	3.00	2.76	2.67	2.6	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.37	3.23	2.94	2.69	2.61	2.54	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.31	3.17	2.88	2.64	2.55	2.48	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.26	3.12	2.83	2.58	2.50	2.42	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.21	3.07	2.78	2.54	2.45	2.37	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.17	3.03	2.74	2.49	2.40	2.33	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.13	2.99	2.70	2.45	2.36	2.29	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.09	2.96	2.66	2.42	2.33	2.25	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.06	2.93	2.63	2.38	2.29	2.22	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.03	2.90	2.60	2.35	2.26	2.19	2.07
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.00	2.87	2.57	2.33	2.23	2.16	2.04
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	2.98	2.84	2.55	2.30	2.21	2.13	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.80	2.66	2.37	2.11	2.02	1.94	1.81
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.63	2.50	2.20	1.94	1.84	1.75	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.47	2.34	2.03	1.76	1.66	1.56	1.38
∞	6.64	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.32	2.19	1.88	1.59	1.48	1.36	1.05

Bibliographie

- [1] Gérard BAILLARGEON. *Méthodes statistiques de l'ingénieur*. Les éditions SMG, 1990.
- [2] Jean BASS. *Éléments de calcul des probabilités théorique et appliqué*. Masson, 1967.
- [3] Jean BASS. *Cours de mathématiques (tome 2)*. Masson, 1968.
- [4] Jean BASS. *Cours de mathématiques (tome 3)*. Masson, 1971.
- [5] Jean BASS. *Éléments de calcul des probabilités*. Masson, 1974.
- [6] S. BLUMENTHAL. *Statistique appliquée*. Les éditions d'organisation, 1989.
- [7] N. BOULEAU. *Probabilités de l'ingénieur*. Hermann, 1986.
- [8] Pierre BRÉMAUD. *Introduction aux probabilités*. Springer-Verlag, 1988.
- [9] D. CARTON. *Processus aléatoires utilisés en recherche opérationnelle*. Masson, 1975.
- [10] P. CHRÉTIENNE et R. FAURE. *Processus stochastiques, leurs graphes, leurs usages*. Gauthier-Villars, 1974.
- [11] G. CULLMANN. *Initiation aux chaînes de Markov — Méthodes et applications*. Masson, 1975.
- [12] Yadolah DODGE. *Premiers pas en statistique*. Springer, 2000.
- [13] Bradley EFRON et Robert TIBSHIRANI. *An introduction to the Bootstrap*. Chapman and Hall, 1993.
- [14] Arthur ENGEL. *Les certitudes du hasard*. Aléas, 1990.
- [15] William FELLER. *An introduction to probability Theory and its applications*. John Wiley & Sons, 1968.
- [16] Dominique FOATA et Aimé FUCHS. *Calcul des probabilités*. Masson, 1996.
- [17] Maurice GIRAULT. *Processus aléatoires*. Dunod, 1965.
- [18] Maurice GIRAULT. *Calcul des probabilités en vue des applications*. Dunod, 1972.
- [19] P.-G. HOEL. *Statistique mathématique*. Armand Colin, 1984.
- [20] S. KARLIN. *Initiation aux processus aléatoires*. Dunod, 1969.
- [21] S. LESSARD et MONGA. *Statistique. Concepts et méthodes*. Masson, 1993.
- [22] M. MÉTIVIER. *Notions fondamentales de la théorie des probabilités*. Dunod, 1968.
- [23] M. MÉTIVIER. *Probabilités : dix leçons d'introduction*. Ellipses, 1987.
- [24] J. NEVEU. *Bases mathématiques du calcul des probabilités*. Masson, 1971.
- [25] Alain REY. « *Analyse 3 – Notions sur la théorie des distributions* ». ENSAM CER d'Angers, 1988.

- [26] A. RUEGG. *Probabilités et statistique*. Presses polytechniques romandes, 1985.
- [27] Gilbert SAPORTA. *Probabilités, analyse des données et statistique*. Technip, 1990.
- [28] Laurent SCHWARTZ. *Méthodes mathématiques pour les sciences physiques*. Hermann, 1965.
- [29] Laurent SCHWARTZ. *Théorie des distributions*. Hermann, 1966.
- [30] L. TAKÀCS. *Processus stochastiques*. Dunod, 1964.
- [31] Bernard VAUQUOIS. *Probabilités*. Hermann, 1969.
- [32] Hélène VENTSEL. *Théorie des probabilités*. Mir, 1973.
- [33] van der WAERDEN. *Statistique mathématique*. Dunod, 1967.
- [34] Thomas WONNACOTT et Ronald WONNACOTT. *Statistique*. Economica, 1991.

Index

- analyse de variance, 112, 124
- « bootstrap », 103
- chaîne de MARKOV, 71
- coefficient
 - de corrélation, 20
 - de détermination, 126, 138
 - de détermination ajusté, 138
- convergence
 - en loi, 65
 - en probabilité, 65
- covariance, 19
- densité de probabilité, 35
- écart-type, 18
- échantillon, 87
 - réalisation de l'–, 87
- ellipse de confiance, 130
- épreuve, 7
- équations normales, 121
- espace
 - probabilisable, 8
 - probabilisé, 8
- espérance mathématique, 17
- estimateur, 87
 - asymptotiquement sans biais, 87
 - convergent, 88
 - des moindres carrés, 135
 - du maximum de vraisemblance, 92
 - sans biais, 87
- estimateurs
 - des moindres carrés, 121
 - du maximum de vraisemblance, 126, 139
- estimation, 88
 - par intervalle, 93
 - ponctuelle, 87
- événement, 7
 - certain, 7
 - impossible, 7
- événements
 - incompatibles, 7
 - indépendants, 10
 - mutuellement indépendants, 11
- fonction
 - caractéristique, 40
 - de répartition, 14, 33, 35
 - génératrice, 22
- intervalle de confiance, 93
- inégalité de BIENAYMÉ-TCHEBYCHEV, 21
- issues, 7
 - équiprobables, 9
- lignes de régression, 119
- loi
 - binomiale, 24
 - bêta (de deuxième espèce), 56
 - conjointe, 15
 - de FISHER-SNEDECOR, 62
 - de POISSON, 29
 - de STUDENT, 60
 - de probabilité, 14
 - des grands nombres, 67
 - du khi-deux, 57
 - exponentielle, 45
 - gamma, 52
 - géométrique, 28
 - hypergéométrique, 27
 - multinomiale, 25
 - normale, 47
 - uniforme, 44
- lois marginales, 16
- matrice
 - de covariance, 21

- de transition, 72
- stochastique, 74
- moment, 17, 38
- moyenne empirique, 88
- prédiction de la variable expliquée, 133
- probabilité, 8
 - conditionnelle, 10
 - critique, 113
- probabilités
 - totales, 8
- processus
 - aléatoire, 71
 - de POISSON, 31, 71
 - homogène, 72
- régression linéaire
 - multiple, 134
 - simple, 120
- risque
 - de deuxième espèce, 105
 - de première espèce, 105
- statistique, 87
- système complet, 11
- test
 - du khi-deux, 114
- tests
 - paramétriques, 105
- théorème
 - de FISHER, 90
 - de la limite centrée, 67
- tribu, 7
 - de BOREL, 33
- variable aléatoire
 - continue, 33
 - discrète, 13
 - réduite, 19
- variables aléatoires
 - indépendantes, 16, 36
 - mutuellement indépendantes, 16
 - non-corrélées, 20
- variance, 18
 - empirique, 89
 - empirique corrigée, 89
- vecteur
 - aléatoire, 16, 21
 - d'état permanent, 77
 - gaussien, 49
 - vraisemblance, 92