



**UNIVERSITE KASDI MERBAH
OUARGLA**

**Faculté des Mathématiques et des Sciences de la
Matière**

N° d'ordre :
N° de série :

DEPARTEMENT DE MATHEMATIQUES

MASTER

Spécialité : Mathématiques

Option : Probabilités et Statistique

Par : Bouroga Nacer Eddine

Thème

La loi normale multivariée

Soutenu publiquement le : 03/07/2019

Devant le jury composé de :

Pr. BAHADI Issa

Université KASDI Merbah-Ouargla

Président

Pr. Meraghni Djamel

Université Mohamed Khidar Biskra

Rapporteur

MCB. Mensoul Brahime

Université KASDI Merbah-Ouargla

Examineur

MCA. BOUSAAD Abdkader

Université KASDI Merbah-Ouargla

Rapporteur

Dédicace

Je dédie ce modeste travail à :
la mémoire de mon Père bien-aimé,
ma mère,
mes frères,
ma grande famille et mes amis,
mes collègues de la promotion "2018 - 2019",
et tous ceux qui sont proches à mon cœur.

Nacer Eddine

REMERCIEMENTS

J'aimerais en premier lieu remercier mon Dieu Allah qui m'a donné la volonté
et le courage pour la réalisation de ce travail que je dédie,

A ma source de bonheur :

Ma très chère mère et mes chers frères pour leur amour, leur bonté,
leur sacrifice, leurs encouragements perpétuels et leur soutien.

Puisse Allah prolonger leur vie dans le bonheur.

A ceux qui ont veillé pour mon bien être.

A ceux qui m'ont soutenu dans les moments les plus difficiles de ma vie.

A ceux que j'aime et je respecte infiniment.

Le jour est venu pour leur dire Merci. . .

Je voudrais remercier mon superviseur **Pr. Meraghni Djamel**,
qui a proposé le sujet de ce mémoire et à qui je me suis adressé tout le temps.

Je remercie également tous mes enseignants et particulièrement
ceux qui m'ont fait l'honneur d'être membres du jury de soutenance.

Le travail, les critiques et les conseils n'ont pas de prix.

Table des matières

Dédicace	i
Remerciements	ii
Table des matières	iii
Liste des figures	v
Liste des tableaux	vi
Introduction	1
1 Vecteurs aléatoires	3
1.1 Définitions - Notations	4
1.2 Vecteurs aléatoires	5
1.3 Loi de probabilités	5
1.3.1 Loi conjointe - lois marginales	5
1.3.2 Fonction de répartition	7
1.3.3 Fonction de densité	8
1.3.4 Fonction caractéristique	8
1.3.5 Coefficient de corrélation	9
1.4 Convergences	10
1.4.1 Convergence en loi	10

1.4.2	Convergence en probabilité	11
1.5	Espérance et matrice de covariance	13
1.5.1	Espérance mathématique	13
1.5.2	Variances-covariances	13
2	Vecteurs aléatoires gaussiens	15
2.1	Vecteur gaussien	15
2.1.1	Caractérisation	15
2.1.2	Vecteurs gaussiens à matrice de covariance diagonale	17
2.1.3	Conditionnement	20
2.1.4	Lois dérivées	20
2.1.5	Théorème Central Limite Vectoriel (TCLV)	22
2.2	Cas particulier : Couple aléatoire Gaussien	23
3	Estimation des paramètres	26
3.1	Preliminaires	26
3.2	Moyenne et matrice de covariance empiriques	29
3.3	Méthode du maximum de vraisemblance	32
3.4	Propriétés des $\bar{\mathbf{X}}$ et \mathbf{S}	33
3.5	Région de confiance pour μ	34
	Conclusion	36
	Bibliographie	37
	Annexe	39

Table des figures

2.1	Courbe de la densité de probabilité d'un vecteur gaussien bidimensionnel .	24
-----	--	----

Liste des tableaux

1.1	Probabilités conjointes	4
1.2	Lois marginales discrètes	7

Introduction

Plusieurs illustres scientifiques ont contribué à la naissance de ce qu'on appelle de nos jours la loi normale [18]. Les premières apparitions de cette loi peuvent être retracées aux années 1730, mais on dit couramment que la loi normale a été introduite au début du 19^e siècle indépendamment par deux mathématiciens, un mathématicien français, Pierre-Simon de Laplace et un mathématicien allemand, Carl Friedrich Gauss [19]. Le premier a utilisé une approche plutôt probabiliste, pendant que Gauss a eu recours à une approche statistique

La popularité du modèle normal multidimensionnel tient à ce qu'il se prête bien au calcul et les problèmes relatifs à un tel modèle reçoivent souvent une solution sous une forme agréable [1]. Cet avantage est dû principalement aux propriétés suivantes : d'une part les transformations affines des vecteurs gaussiens [16] produisent des vecteurs gaussiens, d'autre part lorsque deux vecteurs sont gaussiens, le conditionnement de l'un par l'autre conduit à un vecteur gaussien. Il y a deux raisons essentielles, l'une pratique et l'autre théorique, qui justifient l'usage de ce modèle.

De là je pose quelques questions

Quels sont les paramètres estimés dans cette étude ?

Quelle est la méthode pour cela ?

Chapitre 1 : Vecteurs aléatoires

Dans ce chapitre, on commence par présenter les vecteurs aléatoires en général. On rappelle leurs définitions et caractéristiques de base.

Chapitre 2 : Vecteurs aléatoires Gaussiens

Ce chapitre on s'intéresse plus particulièrement au vecteur aléatoire gaussien en définissant sa densité ainsi que les loi marginales et conditionnelles. On parle aussi des lois dérivées et on rappelle le théorème central limite qui donne à la loi multinormale toute son importance comme une distribution d'approximation.

Chapitre 3 : Estimation des paramètres

Au troisième chapitre, on construit par la méthode du maximum de vraisemblance, des estimateurs pour les deux paramètres de la distribution multinormale tout en étudiant leurs propriétés.

Chapitre 1

Vecteurs aléatoires

Dans des situations où interviennent plusieurs variables aléatoires, le calcul de la probabilité d'un événement dont la réalisation dépend des valeurs de ces variables doit faire intervenir ces variables considérées dans leur ensemble et non chacune isolément. On est amené à étudier ainsi une nouvelle notion, celle de vecteur aléatoire. On commence par préciser cette idée sur un exemple élémentaire.

Exemple 1.0.1 Une urne contient 7 boules : 2 bleues, 3 blanches et 2 rouges. On en prélève 3 d'un coup. On note respectivement X et Y les nombres de boules bleues et blanches dans l'échantillon tiré. Calculer les probabilités suivantes : $\mathbb{P}(X > Y)$, $\mathbb{P}(X = Y)$, \mathbb{P} (avoir 2 boules rouges). Il est clair que ces deux variables suffisent à décrire complètement l'expérience puisque la composition de l'échantillon est déterminée par les valeurs de X et Y : le nombre de rouges étant $(3 - X - Y)$.

L'espace probabilisé associé à cette expérience est l'ensemble de tous les échantillons possibles (il y en a $C_7^3 = 35$) muni de l'équiprobabilité. Les valeurs possibles du couple aléatoire (X, Y) sont dans l'ensemble $\{0, 1, 2\} \times \{0, 1, 2, 3\}$. Les probabilités d'observation de

$X \setminus Y$	0	1	2	3	$\mathbb{P}(X = i)$
0	0	3/35	6/35	1/35	10/35
1	2/35	12/35	6/35	0	20/35
2	2/35	3/35	0	0	5/35
$\mathbb{P}(Y = j)$	4/35	20/35	12/35	1/35	1

TAB. 1.1 – Probabilités conjointes

ces valeurs $\mathbb{P}(X = i, Y = j)$ se calculent en faisant par

$$\begin{aligned}
 \mathbb{P}(X = i, Y = j) &= \mathbb{P}(i \text{ bleues}, j \text{ blanches}, (3 - i - j) \text{ rouges}) \\
 &= \frac{C_2^i C_3^j C_2^{3-i-j}}{C_7^3}
 \end{aligned}$$

On résume les résultats obtenus dans le tableau 1.1.

Le tableau des $\mathbb{P}(X = i, Y = j)$ permet le calcul des lois marginales \mathbb{P}_X et \mathbb{P}_Y . Plus généralement, on peut à l'aide de ce tableau calculer la probabilité de tout événement dont la réalisation ne dépend que de la valeur du couple (X, Y) . Ainsi on a :

$$\mathbb{P}(X > Y) = \frac{2}{35} + \frac{2}{35} + \frac{3}{35} = \frac{1}{5}$$

$$\mathbb{P}(X = Y) = 0 + \frac{12}{35} + 0 = \frac{12}{35}$$

$$\mathbb{P}(2 \text{ rouges}) = \mathbb{P}(X + Y = 1) = \frac{2}{35} + \frac{3}{35} = \frac{1}{7}$$

1.1 Définitions - Notations

Une expérience aléatoire est modélisée par un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, où :

- Ω désigne l'ensemble de tous les résultats possibles, appelé ensemble fondamental : $\Omega = \{\omega / \omega \text{ résultat}\}$.
- \mathcal{F} une tribu ou σ -algèbre sur Ω appelée tribu des événements.
- \mathbb{P} mesure de probabilité sur (Ω, \mathcal{F}) .

Exemple 1.1.1 Les expériences classiques du lancé de dé, du tirage de boules dans une

urne... Expériences aléatoires plus complexes : sondages, études de fiabilité, durées de vie, ...Nécessité d'introduire la notion de variable aléatoire.

1.2 Vecteurs aléatoires

On considère une expérience aléatoire modélisée par un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

Définition 1.2.1 (Variable aléatoire) Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé [7]. On appelle variable aléatoire réelle X toute application mesurable de (Ω, \mathcal{F}) vers $(\mathbb{R}, \mathcal{B})$, où \mathcal{B} est la tribu de Borel de \mathbb{R} . La loi de probabilité de X est définie par

$$\mathbb{P}_X(B) = \mathbb{P}(\omega \mid X(\omega) \in B) = \mathbb{P}(X^{-1}(B)), \quad B \in \mathcal{B}.$$

Définition 1.2.2 Si X_1, \dots, X_p sont des variables aléatoires sur le même espace $(\Omega, \mathcal{F}, \mathbb{P})$, on définit le vecteur aléatoire $X = (X_1, \dots, X_p)^t$ comme l'application

$$\begin{aligned} \Omega &\rightarrow \mathbb{R}^p \\ \omega &\mapsto (X_1(\omega), \dots, X_p(\omega))^t \end{aligned}$$

- La variable aléatoire X_i est la i -ème marginale du vecteur X .
- Chaque X_i est une variable aléatoire
- Comme dans le cas univarié, on associe à chaque vecteur aléatoire une fonction de répartition et une densité.

1.3 Loi de probabilités

1.3.1 Loi conjointe - lois marginales

La notion de vecteur aléatoire [7] permet de probabiliser l'espace mesurable $(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p))$, où $\mathcal{B}(\mathbb{R}^p)$ est la tribu des boréliens de \mathbb{R}^p .

Proposition 1.3.1 Soit $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p))$ un vecteur aléatoire à p dimensions. Alors l'application $\mathbb{P}_X : \mathcal{B}(\mathbb{R}^p) \rightarrow [0, 1]$ définie pour tout borélien B de \mathbb{R}^p par

$$\mathbb{P}_X(B) := \mathbb{P}(X^{-1}(B)), \quad (1.1)$$

est une mesure de probabilité sur $(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p))$.

Définition 1.3.1 La mesure de probabilité \mathbb{P}_X définie par (1.1) est appelée loi de probabilité ou loi conjointe du vecteur aléatoire X . On dit aussi que X suit la loi de probabilité \mathbb{P}_X ou que X a pour loi conjointe \mathbb{P}_X , et on peut noter $X \sim \mathbb{P}_X$.

Remarque 1.3.1 L'espace $(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), \mathbb{P}_X)$ est un nouvel espace de probabilité. Comme pour les variables aléatoires réelles, on peut donc aussi parler de vecteur aléatoire X ayant une loi conjointe \mathbb{P}_X sans spécifier l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ sur lequel X est défini à la base.

Définition 1.3.2 (Loi discrète) La loi conjointe \mathbb{P}_X de X est dite discrète [7] s'il existe un sous-ensemble S_X de \mathbb{R}^p fini ou dénombrable tel que pour tout $x = (x_1, \dots, x_p)^t \in S_X$, $\mathbb{P}_X(x) > 0$ et $\mathbb{P}_X(S_X) = \sum_{x \in S_X} \mathbb{P}_X(x) = 1$. L'ensemble S_X est alors appelé support de la loi conjointe de X .

Définition 1.3.3 (Loi marginale discrète) Si X est un vecteur aléatoire [7] de loi \mathbb{P}_X discrète, si S_{X_i} désigne le support de X_i , alors $S_X \subset \prod_{i=1}^p S_{X_i}$ et la loi marginale de X_i est donnée par :

$$\mathbb{P}_{X_i}(x_i) = \sum_{x_j \in S_{X_j}, j \neq i} \mathbb{P}_X(x_1, x_2), \quad x_i \in S_{X_i}.$$

Remarque 1.3.2 Un vecteur aléatoire a une loi conjointe discrète si et seulement si ses lois marginales sont discrètes.

Exemple 1.3.1 On lance successivement trois pièces de monnaie (équilibrées) et on désigne respectivement par X et Y les variables aléatoires représentant le nombre de faces

$Y \setminus X$	0	1	2	marge
0	0	1/8	1/8	1/4
1	1/8	2/8	1/8	1/2
2	1/8	1/8	0	1/4
marge	1/4	1/2	1/4	1

TAB. 1.2 – Lois marginales discrètes

apparues sur les deux premières pièces et le nombre de piles sur les deux dernières. La loi du couple (X, Y) et les lois marginales de X et Y sont données dans le tableau 1.2.

Définition 1.3.4 (Loi absolument continue) La loi conjointe \mathbb{P}_X d'un vecteur aléatoire X est dite absolument continue [7] si elle est absolument continue par rapport à la mesure de Lebesgue sur $(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p))$. On dit aussi que le vecteur X est absolument continu.

1.3.2 Fonction de répartition

La fonction de répartition [6] conjointe de X est une application \mathbb{F}_X de \mathbb{R}^p dans $[0, 1]$ définie par :

$$\mathbb{F}_X(x_1, \dots, x_p) := \mathbb{P}(X_1 \leq x_1, \dots, X_p \leq x_p).$$

Elle permet de caractériser la loi de probabilité conjointe.

Proposition 1.3.2 La fonction de répartition [13] conjointe \mathbb{F}_X d'un vecteur aléatoire X satisfait les propriétés suivantes :

1. Pour tout $x \in \mathbb{R}^p$, $0 \leq \mathbb{F}_X(x) \leq 1$.
2. \mathbb{F}_X est une fonction croissante, continue à droite en chacun de ses arguments x_1, \dots, x_p .
3. $\mathbb{F}_X(x)$ tend vers 0 lorsque l'un des arguments x_i tend vers $-\infty$, et tend vers 1 lorsque tous les x_i tendent vers $+\infty$.

Définition 1.3.5 La fonction de répartition [14] conjointe d'un vecteur aléatoire $X = (X_1, \dots, X_p)^t$ détermine les fonctions de répartition $\mathbb{F}_{X_1}, \dots, \mathbb{F}_{X_p}$ des variables aléatoires

marginales X_1, \dots, X_p (appelées fonctions de répartition marginales). En effet, pour tout $i \in \{1, \dots, p\}$.

$$\mathbb{F}_{X_i}(x_i) = \mathbb{P}(X_i \leq x_i) := \lim_{x_j \rightarrow +\infty, j \neq i} \mathbb{F}_X(x_1, \dots, x_p) = \mathbb{F}_X(+\infty, \dots, +\infty, x_i, +\infty, \dots, +\infty).$$

1.3.3 Fonction de densité

Lorsque le vecteur aléatoire X est absolument continu, on appelle densité [15] conjointe de X toute fonction f_X définie sur \mathbb{R}^p par

$$f_X(x_1, \dots, x_p) := \frac{\partial^p \mathbb{F}_X(x_1, \dots, x_p)}{\partial x_1 \dots \partial x_p}.$$

Proposition 1.3.3 Une densité [15] conjointe f_X vérifie les propriétés suivantes :

- $f_X(x_1, \dots, x_p) \geq 0, (x_1, \dots, x_p) \in \mathbb{R}^p$.
- $\underbrace{\int_{\mathbb{R}} \dots \int_{\mathbb{R}}}_{p} f_X(x_1, \dots, x_p) dx_1 \dots dx_p = 1$.
- Si \mathbb{F}_X désigne la fonction de répartition conjointe de X , alors

$$\mathbb{F}_X(x_1, \dots, x_p) = \int \dots \int_{]-\infty, x_1] \times \dots \times]-\infty, x_p]} f_X(u_1, \dots, u_p) du_1 \dots du_p.$$

Définition 1.3.6 (Densités marginales) Lorsque X admet [16] une densité f_X , la densité marginale de X_i est définie par

$$f_{X_i}(x_i) = \int \dots \int_{\underbrace{\mathbb{R}}_{p-1}} f_X(x_1, \dots, x_p) dx_1, \dots, dx_{i-1}, dx_{i+1}, \dots, dx_p.$$

1.3.4 Fonction caractéristique

Soit $t = (t_1, \dots, t_p)^t \in \mathbb{R}^p$. On appelle fonction caractéristique [15] du vecteur aléatoire X la fonction φ_X définie par :

$$\varphi_X(t) = \mathbb{E} [\exp(it^t X)] = \mathbb{E} \left[\exp \left(i \sum_{j=1}^p t_j X_j \right) \right],$$

où $i^2 = -1$. [14]

Proposition 1.3.4 (Indépendance) *Les composantes X_1, \dots, X_p de X sont indépendantes [14] si et seulement si la fonction caractéristique de X est égale au produit des fonctions caractéristiques de ses composantes, c'est à dire :*

$$\varphi_X(t) = \prod_{j=1}^p \varphi_{X_j}(t_j).$$

Preuve. Si les X_j sont indépendantes l'espérance [14] d'un produit de fonctions des X_j est égale au produit des espérances, donc

$$\mathbb{E} [\exp(it^t X)] = \mathbb{E} [\exp(it_1 X_1)] \mathbb{E} [\exp(it_2 X_2)] \dots \mathbb{E} [\exp(it_p X_p)].$$

Ceci démontre une partie de la proposition. La réciproque, plus délicate, utilise l'inversion de la fonction caractéristique et est omise. ■

1.3.5 Coefficient de corrélation

Le coefficient de corrélation du couple (X, Y) est le réel [2] noté ρ_{XY} et égal à

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}} = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

La matrice de covariance s'écrit alors

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Proposition 1.3.5 *Soient deux variables aléatoires réelles X et Y possédant des variances*

non nulles.

1. $|\rho_{XY}| \leq 1$
2. $|\rho_{XY}| = 1$ si et seulement si Y est une fonction quasi-affine de X .
3. Posons $U = aX + b$ et $V = cY + d$ avec a et c dans \mathbb{R}^* et b et d dans \mathbb{R} alors

$$|\rho_{UV}| = |\rho_{XY}|.$$

1.4 Convergences

1.4.1 Convergence en loi

Définition

Définition 1.4.1 Soit (X_n) et X des vecteurs aléatoires à valeurs dans l'espace probabilisable $(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p))$. On dit que la suite (X_n) converge en loi vers X si, pour toute fonction h de \mathbb{R}^p vers \mathbb{R} , continue et bornée, on a [8]

$$\lim_{n \rightarrow +\infty} \mathbb{E}[h(X_n)] = \mathbb{E}[h(X)]$$

Caractérisation de la convergence en loi

Voyons dans un premier temps une condition nécessaire et suffisante pour la convergence en loi dans le cas de v.a.r.

Proposition 1.4.1 Soit (X_n) et X des v.a.r. de fonction de répartition (\mathbb{F}_{X_n}) et \mathbb{F}_X respectivement. La suite (X_n) converge en loi [8] vers X si, et seulement si,

$$\lim_{n \rightarrow +\infty} \mathbb{F}_{X_n}(x) = \mathbb{F}_X(x)$$

en tout point x où \mathbb{F} est continue.

Théorème 1.4.1 Soit (X_n) et X des vecteurs aléatoires de \mathbb{R}^p , absolument [8] continus de densité (f_{X_n}) et f_X par rapport à la mesure de Lebesgue dans \mathbb{R}^p . Si on a

$$\lim_{n \rightarrow +\infty} f_{X_n}(x) = f_X(x),$$

alors

$$X_n \xrightarrow{\text{loi}} X.$$

Théorème 1.4.2 (Cramer-Wold, 1936) X_n et X deux vecteurs aléatoires [14] à valeurs dans \mathbb{R}^p . Alors pour tout $u \in \mathbb{R}^p$, on a

$$X_n \xrightarrow{\text{loi}} X \iff u^t X_n \xrightarrow{\text{loi}} u^t X.$$

Preuve. Voir [16, page 217]. ■

1.4.2 Convergence en probabilité

Définition 1.4.2 Soit (X_n) et X des vecteurs aléatoires à valeurs dans \mathbb{R}^p . On dit que (X_n) converge en probabilité [8] vers le vecteur aléatoire X si ses p composantes $X_{n,1}, \dots, X_{n,p}$ convergent en probabilité vers les composantes X_1, \dots, X_p de X

Le théorème suivant permet de donner une définition équivalente à cette convergence en probabilité .

Théorème 1.4.3 Soit $\|\cdot\|$ une norme quelconque dans \mathbb{R}^p et (X_n) et X des vecteurs aléatoires dans \mathbb{R}^p . La suite (X_n) converge en probabilité [8] vers la v.a. X si, et seulement si,

$$\|X_n - X\| \rightarrow 0, \text{ quand } n \rightarrow +\infty$$

Preuve. Démontrons ce résultat pour la norme supérieure dans \mathbb{R}^p . Supposons en premier lieu que X_n converge en probabilité vers X et notons

$$Y = \|X_n - X\| = \max_i |X_{n,i} - X_i|.$$

De l'inclusion

$$\{Y > \epsilon\} \subset \bigcup_{i=1}^p \{|X_{n,i} - X_i| > \epsilon\},$$

on tire

$$\mathbb{P}(Y > \epsilon) \leq \sum_{i=1}^p \mathbb{P}\{|X_{n,i} - X_i| > \epsilon\}.$$

Par hypothèse, le terme de droite de cette dernière inégalité converge en probabilité vers 0, d'où on a la convergence

$$Y \xrightarrow{p} 0, \text{ quand } n \rightarrow +\infty.$$

Réciproquement, supposons que la v.a.r. Y converge en probabilité vers 0. Ayant, pour tout $i = 1, \dots, p$,

$$\{|X_{n,i} - X_i| > \epsilon\} \subset \{Y > \epsilon\},$$

et donc

$$\mathbb{P}\{|X_{n,i} - X_i| > \epsilon\} \leq \mathbb{P}(Y > \epsilon).$$

On a bien la convergence de X_n vers X en probabilité. ■

Convergence en probabilité et convergence en loi

La proposition suivante donne la relation entre les deux types de convergence [8].

Proposition 1.4.2

$$X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{\text{loi}} X.$$

1.5 Espérance et matrice de covariance

1.5.1 Espérance mathématique

Définition 1.5.1 (Espérance mathématique) *On dit que $X = (X_1, \dots, X_p)$ possède une espérance mathématique μ si ses variables marginales X_1, \dots, X_p possèdent chacune une espérance mathématique, et on note [14]*

$$\mu = \mathbb{E}(X) := \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix},$$

où $\mu_i = \mathbb{E}(X_i)$, $i = 1, \dots, p$.

Les propriétés de l'espérance mathématique pour les vecteurs aléatoires découlent donc directement de celles de l'espérance mathématique des variables aléatoires réelles. En particulier, on a le résultat de linéarité suivant :

Proposition 1.5.1 *Si A est une matrice réelle de taille $n \times p$, alors*

$$\mathbb{E}(AX) = A\mathbb{E}(X).$$

1.5.2 Variances-covariances

Définition 1.5.2 (Variations-covariances) *La matrice [2] de variance-covariance Σ de X est définie par*

$$\Sigma := \mathbb{E}((X - \mathbb{E}(X))(X - \mathbb{E}(X))^t).$$

Remarque 1.5.1

1. Après développement, on trouve que $\Sigma = \mathbb{E}(XX^t) - \mu\mu^t$.

2. L'écriture matricielle de la matrice de covariance est

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_p^2 \end{pmatrix},$$

où $\sigma_i^2 = \sigma_{ii} := \mathbb{E}(X_i - \mu_i)^2$ représente la variance de la v.a.r X_j et $\sigma_{ij} = \text{cov}(X_i, X_j) := \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)]$ représente la covariance entre X_i et X_j .

3. Σ est une matrice carrée symétrique d'ordre p .

Proposition 1.5.2 La matrice de covariance Σ est **définie positive**.

Preuve. Soit $x = (x_1, \dots, x_p)^t \in \mathbb{R}^p - \{0\}$, on montre que $x^t \Sigma x > 0$. On a

$$\begin{aligned} x^t \Sigma x &= \sum_{i=1}^p \sum_{j=1}^p x_i x_j \sigma_{ij} \\ &= \sum_{i=1}^p \sum_{j=1}^p x_i x_j \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] \\ &= \mathbb{E} \left[\left(\sum_{i=1}^p x_i (X_i - \mu_i) \right)^2 \right]. \end{aligned}$$

D'où le résultat. ■

Chapitre 2

Vecteurs aléatoires gaussiens

Dans ce chapitre, on présente les définitions et caractéristiques principales d'un vecteur aléatoire particulier, à savoir le vecteur aléatoire multinormal.

2.1 Vecteur gaussien

2.1.1 Caractérisation

Définition 2.1.1 Un vecteur aléatoire $X = (X_1, \dots, X_p)^t$ de \mathbb{R}^p est dit gaussien [14] si et seulement si, pour tout $\lambda = (\lambda_1, \dots, \lambda_p)^t$ de $\mathbb{R}^p - \{0\}$, la v.a

$$\lambda^t X = \sum_{i=1}^p \lambda_i X_i,$$

est une v.a de loi normale. Autrement dit, si et seulement si toute combinaison linéaire des composantes de (X_1, \dots, X_p) est gaussienne.

Remarque 2.1.1 Si X est un vecteur aléatoire gaussien alors toutes ses composantes sont gaussiennes. La réciproque est fausse [14].

Proposition 2.1.1 Un vecteur aléatoire X est gaussien si et seulement si il existe $\mu \in \mathbb{R}^p$ et une matrice carrée d'ordre p , symétrique, définie positive Σ tels que pour tout $t =$

$(t_1, \dots, t_p)^t \in \mathbb{R}^p$, on a

$$\varphi_X(t) = \exp \left\{ it^t \mu - \frac{t^t \Sigma t}{2} \right\}.$$

Dans ce cas, on note $X \sim \mathcal{N}_p(\mu, \Sigma)$, avec μ et Σ désignant le vecteur moyen et la matrice de covariance de X respectivement [14].

Remarque 2.1.2 Si $p = 1$, alors X est une variable aléatoire (unidimensionnelle) normale d'espérance m et de variance σ^2 : $X \sim \mathcal{N}(m, \sigma^2)$. Dans ce cas, la fonction caractéristique est donnée par [14]

$$\varphi_X(t) = \exp \left\{ imt - \frac{\sigma^2 t^2}{2} \right\}, \quad t \in \mathbb{R}. \quad (2.1)$$

Preuve. (\implies) Par définition de φ_X , on a

$$\varphi_X(t) = \mathbb{E}(\exp(it^t X)) = \mathbb{E}(\exp(iY)) = \varphi_Y(1),$$

où $Y = t^t X = \sum_{j=1}^p t_j X_j$ est une variable aléatoire gaussienne (combinaison linéaire des coordonnées du vecteur gaussien X). D'après (2.1), on a

$$\varphi_Y(1) = \exp \left\{ i\mathbb{E}(Y) - \frac{\text{Var}(Y)}{2} \right\},$$

avec

$$\mathbb{E}(Y) = \sum_{j=1}^p t_j \mathbb{E}(X_j) = t^t \mu,$$

et

$$\begin{aligned} \text{Var}(Y) &= \text{Var} \left(\sum_{i=1}^p t_i X_i \right) = \sum_{i=1}^p \sum_{j=1}^p \text{cov}(t_i X_i, t_j X_j) \\ &= \sum_{i=1}^p \sum_{j=1}^p t_i t_j \text{cov}(X_i, X_j) = t^t \Sigma t. \end{aligned}$$

Ceci entraîne la forme voulue de la fonction caractéristique de X .

(\Leftarrow) Soit $\lambda = (\lambda_1, \dots, \lambda_p)^t \in \mathbb{R}^p$. On pose $Y = \sum_{i=1}^p \lambda_i X_i = \lambda^t X$. On doit montrer que Y est une variable aléatoire gaussienne. Pour tout $s \in \mathbb{R}$, la fonction caractéristique de Y est donnée par

$$\begin{aligned} \varphi_Y(s) &= \mathbb{E}(\exp(is\lambda^t X)) = \varphi_X(s\lambda) \\ &= \exp\left(i(s\lambda)^t \mu - \frac{1}{2}(s\lambda)^t \Sigma (s\lambda)\right) \\ &= \exp\left(is\lambda^t \mu - \frac{1}{2}s^2 \lambda^t \Sigma \lambda\right). \end{aligned}$$

On reconnaît la fonction caractéristique de la loi gaussienne $\mathcal{N}_p(\lambda^t \mu, \lambda^t \Sigma \lambda)$. Donc Y est bien une variable aléatoire gaussienne. ■

Proposition 2.1.2 (Densité de probabilité) *La loi gaussienne $\mathcal{N}_p(\mu, \Sigma)$ sur \mathbb{R}^p admet une densité de probabilité par rapport à la mesure de Lebesgue de \mathbb{R}^p si et seulement si Σ est inversible. Dans ce cas la densité est donnée, pour tout $x \in \mathbb{R}^p$, par [14]*

$$f_X(x_1, \dots, x_p) = \frac{1}{(\sqrt{2\pi})^p \sqrt{\det(\Sigma)}} \exp\left\{-\frac{1}{2}(x - \mu)^t \Sigma^{-1} (x - \mu)\right\}.$$

2.1.2 Vecteurs gaussiens à matrice de covariance diagonale

Le résultat suivant caractérise les vecteurs gaussiens à matrice de covariance diagonale, c-à-d à composantes non corrélées.

Théorème 2.1.1 *Pour tout vecteur gaussien X de \mathbb{R}^p , les trois propriétés suivantes sont équivalentes :*

- a) *Les composantes X_1, \dots, X_p sont mutuellement indépendantes.*
- b) *Les composantes X_1, \dots, X_p sont deux à deux indépendantes.*
- c) *La matrice de covariance Σ de X est diagonale.*

Preuve. Les implications $\mathbf{a} \Rightarrow \mathbf{b} \Rightarrow \mathbf{c}$. sont évidentes. Montrons que $\mathbf{c} \Rightarrow \mathbf{a}$. Si $\Sigma = \text{Diag}(\sigma_1^2, \dots, \sigma_p^2)$, alors pour tout $u \in \mathbb{R}^p$, on a

$$\begin{aligned} \varphi_X(u) &= \exp\left(iu^t\mu - \frac{1}{2}u^t\Sigma u\right) \\ &= \exp\left(i\sum_{k=1}^p u_k\mu_k - \frac{1}{2}\sum_{k=1}^p \sigma_k^2 u_k^2\right) \\ &= \prod_{k=1}^p \exp\left(iu_k\mu_k - \frac{1}{2}\sigma_k^2 u_k^2\right) \\ &= \prod_{k=1}^p \varphi_{X_k}(u_k). \end{aligned}$$

Le corollaire suivant permet de caractériser l'indépendance des copposantes d'un vecteur aléatoire Gaussien. ■

Corollaire 2.1.1 *Soit $X = (X_1, \dots, X_p)^t$ un vecteur gaussien. Pour tout $(i, j) \in \{1, \dots, p\}^2$ tel que $i \neq j$, X_i et X_j sont indépendantes si et seulement si $\text{cov}(X_i, X_j) = 0$.*

Exemple 2.1.1 *Considérons p variables aléatoires X_1, \dots, X_p indépendantes et de loi respectivement $\mathcal{N}(\mu_1, \sigma_1^2), \dots, \mathcal{N}(\mu_p, \sigma_p^2)$. Pour $i = 1, \dots, p$, la variable aléatoire X_i est donc de densité*

$$f_{X_i} = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{1}{2}\left(\frac{x - \mu_i}{\sigma_i}\right)^2\right\},$$

par rapport à la mesure de Lebesgue λ sur \mathbb{R} . En raison de l'indépendance des variables aléatoires X_i , la densité conjointe du vecteur (X_1, \dots, X_p) est

$$f_{X_1, \dots, X_p}(x_1, \dots, x_p) = \frac{1}{(\sqrt{2\pi})^p} \frac{1}{\prod_{i=1}^p \sigma_i} \exp\left\{-\frac{1}{2}\sum_{i=1}^p \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right\}.$$

L'espérance du vecteur $X = (X_1, \dots, X_p)^t$ et sa matrice de covariance sont

$$\mathbb{E}[X] = \mathbb{E}[X_1, \dots, X_p] = \mu = [\mu_1, \dots, \mu_p]^t,$$

et

$$\Sigma = \begin{pmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_p^2 \end{pmatrix}.$$

Notons que la matrice Σ est diagonale en raison de l'indépendance des v.a.r. X . Comme toutes les variances σ_i^2 sont strictement positives, on obtient aisément la matrice inverse

$$\Sigma^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2} & & & 0 \\ & \frac{1}{\sigma_2^2} & & \\ & & \ddots & \\ 0 & & & \frac{1}{\sigma_p^2} \end{pmatrix}.$$

On peut alors réécrire la densité conjointe du vecteur $X = (X_1, \dots, X_p)^t$ sous la forme

$$f_X(x_1, \dots, x_p) = \frac{1}{(\sqrt{2\pi})^p} \frac{1}{\sqrt{\det[\Sigma]}} \exp \left\{ -\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu) \right\},$$

car

$$(x - \mu)^t \Sigma^{-1} (x - \mu) =$$

$$(x_1 - \mu_1, \dots, x_p - \mu_p) \begin{pmatrix} \frac{1}{\sigma_1^2} & & & 0 \\ & \frac{1}{\sigma_2^2} & & \\ & & \ddots & \\ 0 & & & \frac{1}{\sigma_p^2} \end{pmatrix} \begin{pmatrix} x_1 - \mu_1 \\ \vdots \\ \vdots \\ \vdots \\ x_p - \mu_p \end{pmatrix}$$

$$= \sum_{i=1}^p \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2.$$

2.1.3 Conditionnement

Partitionnons X en deux sous-vecteurs X_1 et X_2 à k et $p - k$ composantes [14] respectivement d'espérance μ_1 et μ_2 :

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

La matrice de variance-covariance se partitionne en 4 blocs :

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

La loi du vecteur X_1 conditionnée par X_2 est donnée par la proposition suivante :

Proposition 2.1.3 *La loi de X_1/X_2 est une loi multinormale à k dimensions d'espérance $\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$ et de matrice variance de covariance $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}$.*

2.1.4 Lois dérivées

Loi de Wishart

Définition 2.1.2 *On définit la matrice aléatoire $W := \sum_{i=1}^n X_i X_i^t$, où X_1, \dots, X_n sont n vecteurs aléatoires indépendants de même loi $\mathcal{N}_p(0, \Sigma)$. Alors la loi de probabilité de W est appelée loi de **Wishart** [14] de paramètres p, n, Σ , où n est degré de liberté et on écrit $W \sim \mathcal{W}_p(n, \Sigma)$.*

Remarque 2.1.3 *La loi de Wishart est une généralisation multivariée de celle du khi-deux.*

Proposition 2.1.4

1. *L'espérance de distribution de Wishart est $\mathbb{E}(W) = n\Sigma$.*

2. Si $W \sim \mathcal{W}_p(n, \Sigma)$, alors pour tout vecteur a fixé, le rapport des deux formes quadratiques $a^t \Sigma^{-1} a$ et $a^t W^{-1} a$ suit la loi du khi-deux à $n - p + 1$ degré de liberté :

$$\frac{a^t \Sigma^{-1} a}{a^t W^{-1} a} \sim \chi_{n-p+1}^2.$$

Preuve. Voir [14, page 104]. ■

Loi T^2 de Hotelling

Cette distribution généralise celle de Student.

Définition 2.1.3 Soit X un vecteur aléatoire normale $\mathcal{N}_p(0, I)$ et W une matrice de Wishart $\mathcal{W}_p(n, I)$ indépendante de X alors la variable aléatoire $nX^t W^{-1} X$ suit la loi du T^2 de **Hotelling** [14] de paramètres p et n , notée $T^2(p, n)$.

Proposition 2.1.5 Soient $X \sim \mathcal{N}_p(\mu, \Sigma)$ et $W \sim \mathcal{W}_p(n, \Sigma)$ indépendants. Alors

$$n(X - \mu)^t W^{-1} (X - \mu) \sim T^2(p, n).$$

Preuve. Voir [14, page 104]. ■

Il existe une relation entre la loi de Hotelling et celle de Fisher.

Proposition 2.1.6 La loi de **Hotelling** $T^2 = n(X - \mu)^t W^{-1} (X - \mu) \sim T^2(p, n)$ à une loi de **Fisher**, soit

$$\frac{n - p + 1}{np} T^2 \sim \mathcal{F}(p, n - p + 1).$$

Preuve. Voir [20, page 155]. ■

Loi du lamda Λ de Wilks

Définition 2.1.4 Soit A et B deux matrice de Wishart $\mathcal{W}_p(m, \Sigma)$ et $\mathcal{W}_p(n, \Sigma)$ indépendantes où $m \geq p$ alors le quotient [14] :

$$\Lambda := \frac{\det(A)}{\det(A+B)},$$

a une distribution de Wilks de paramètres p, m et n , notée $\Lambda(p, m, n)$.

Remarque 2.1.4 Λ , qui s'écrit aussi $1/\det(A^{-1}B + I)$, ne dépend pas de Σ . Il est comprise entre 0 et 1.

2.1.5 Théorème Central Limite Vectoriel (TCLV)

Le résultat suivant généralise celui du cas unidimensionnel. C'est un résultat de base en statistique.

Théorème 2.1.2 Soient X_1, \dots, X_n des vecteurs aléatoires de \mathbb{R}^p i.i.d. admettant un moment d'ordre 2. On note par μ leur espérance et par Σ leur matrice de covariance. Alors [8]

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\text{loi}} \mathcal{N}_p(0, \Sigma), \text{ quand } n \rightarrow \infty.$$

Preuve. Soit $u \in \mathbb{R}^p$, on pose $X_n = u^t X_n$. Alors $(X_n)_{n \geq 1}$ est une suite de v.a.r i.i.d avec

$$\mathbb{E}(X_n) = u^t \mu \text{ et } \text{Var}(X_n) = u^t \Sigma u.$$

En utilisant le théorème central limite unidimensionnel classique, on a

$$\sqrt{n}(\bar{X}_n - u^t \mu) \xrightarrow{\text{loi}} \mathcal{N}_p(0, u^t \Sigma u).$$

Comme $\overline{X_n} = u^t \overline{X_n}$, alors on peut écrire que

$$u^t \sqrt{n} (\overline{X_n} - \mu) \xrightarrow{\text{loi}} u^t \mathcal{N}_p(0, \Sigma).$$

Enfin, on applique le théorème de Cramer-Wold pour obtenir le résultat. ■

2.2 Cas particulier : Couple aléatoire Gaussien

Soit le vecteur aléatoire gaussien $X = (X_1, X_2)^t$ dibimensionnel de paramètres

$$\mu = (\mu_1, \mu_2)^t \text{ et } \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

où $\rho := \sigma_{12}/(\sigma_1\sigma_2)$ est le coefficient de corrélation linéaire des deux v.a X_1 et X_2 . On a

$$\det \Sigma = \sigma_1^2\sigma_2^2 - \sigma_{12}^2 = \sigma_1^2\sigma_2^2 (1 - \rho^2),$$

et

$$\Sigma^{-1} = \begin{pmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{pmatrix} = \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}.$$

Par conséquent, la densité de probabilité du couple X , dont le graphe est illustré par la figure 2.1 [14], s'écrit sous les deux formes suivantes :

$$f_X(x_1, x_2) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2 - \sigma_{12}^2}} \exp \left\{ -\frac{1}{2(\sigma_1^2\sigma_2^2 - \sigma_{12}^2)} \times \right. \\ \left. -\frac{1}{2(\sigma_1^2\sigma_2^2 - \sigma_{12}^2)} [(x_1 - \mu_1)^2\sigma_2^2 - 2(x_1 - \mu_1)(x_2 - \mu_2)\sigma_{12} + (x_2 - \mu_2)^2\sigma_1^2] \right\},$$

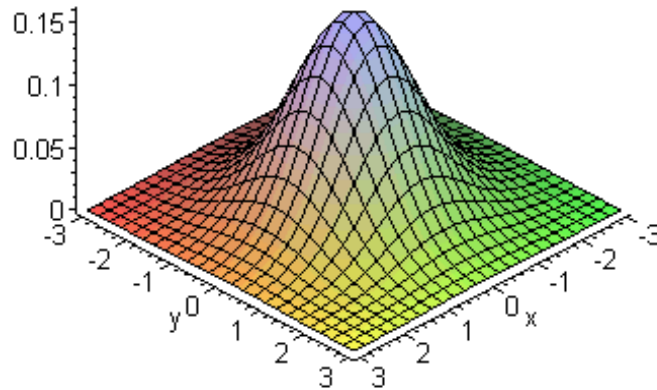


FIG. 2.1 – Courbe de la densité de probabilité d'un vecteur gaussien bidimensionnel

et

$$f_X(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right] \right\}.$$

Exemple 2.2.1 (Vecteur Gaussien à deux dimensions) Soit $X = (X_1, X_2)^t$ une vecteur gaussien bidimensionnel de moyenne nulle et de matrice de covariance

$$\Sigma = \begin{pmatrix} 1 & 0.75 \\ 0.75 & 1 \end{pmatrix}.$$

On a $\det(\Sigma) > 0$. Alors la matrice de covariance est inversible et son inverse est

$$\Sigma^{-1} = (\det(\Sigma))^{-1} \times \begin{pmatrix} 1 & -0.75 \\ -0.75 & 1 \end{pmatrix},$$

d'où

$$f_X(x_1, x_2) = \frac{1}{2\pi\sqrt{0.44}} \exp\left(-\frac{1}{0.88} (x_1^2 - 1.5x_1x_2 + x_2^2)\right).$$

Chapitre 3

Estimation des paramètres

De chapitre est consacré à l'estimation de l'espérance et la matrice de covariance d'un vecteur aléatoire gaussien.

3.1 Préliminaires

Définition 3.1.1 Soit X un vecteur aléatoire réel de dimension p . Un échantillon, de taille $n \geq 1$, de X est une suite des vecteurs aléatoires X_1, \dots, X_n indépendants de même loi de probabilité que X . On le note par

$$\mathbf{X} = \begin{bmatrix} X_1^t \\ \vdots \\ X_n^t \end{bmatrix} = \begin{bmatrix} X_{11} & \cdots & X_{1p} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{np} \end{bmatrix}.$$

La matrice

$$\mathbf{x} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix},$$

est une **réalisation** de l'échantillon \mathbf{X} .

La fonction de répartition \mathbb{F} des X_i appartient à une famille paramétrique des fonctions de répartition

$$\mathcal{F} := \{\mathbb{F}_\theta, \theta \in \Theta\},$$

où Θ est un ensemble connu, appelé **ensemble des paramètres**, et $\mathbb{F}_\theta(x)$ est connu comme fonction de x et θ . La seule inconnue dans cette construction est la vraie valeur du paramètre $\theta = (\theta_1, \dots, \theta_k)^t$. La famille \mathcal{F} est appelée **modèle statistique** (où **modèle statistique paramétrique**). Le problème de l'estimation statistique [12] consiste à construire une statistique (fonction mesurable des X_i), dite estimateur, $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_k)^t$ qui soit proche de θ en un sens probabiliste. Deux des propriétés les plus importantes d'un estimateur sont données dans la définition suivante.

Définition 3.1.2

Soit $\hat{\theta}$ un estimateur θ [12]

1. $\hat{\theta}$ est dit **sans biais** pour θ si $\mathbb{E}(\hat{\theta}) = \theta$. Dans le cas contraire, on dit que $\hat{\theta}$ est **biaisé**.
2. $\hat{\theta}$ est dit **convergent** (où **consistant**) s'il converge vers θ en probabilité : $\hat{\theta} \xrightarrow{P} \theta$, i.e.

$$\lim_{n \rightarrow +\infty} \mathbb{P}_\theta \left(\left| \hat{\theta} - \theta \right| > \epsilon \right) = 0, \text{ pour tout } \epsilon > 0.$$

Définition 3.1.3 (Fonction de vraisemblance) Soit X_1, \dots, X_n un échantillon de densité $f(x, \theta)$. La fonction de vraisemblance [12] de cet échantillon est définie comme la densité jointe des vecteurs X_i :

$$L(\mathbf{X}, \theta) = L(X_1, \dots, X_n, \theta) := \prod_{i=1}^n f(x_i, \theta).$$

Remarque 3.1.1 Une statistique $\hat{\theta}$ est exhaustive pour un paramètre θ si sa fonction du

vraisemblance [12] écrire comme

$$L(X_1, \dots, X_n, \theta) = g(\hat{\theta}, \theta) h(X_1, \dots, X_n).$$

où $h(X_1, \dots, X_n)$ ne dépend pas de θ .

L'importance d'une statistique exhaustive est que toutes les informations apportées par l'échantillon sur le paramètre sont contenues dans cette statistique.

Définition 3.1.4 *L'estimateur du maximum de vraisemblance (emv) de θ est défini comme*

$$\hat{\theta}^{MV} := \arg \max_{\theta} L(\mathbf{X}, \theta).$$

Remarque 3.1.2 *Souvent, il est plus facile de maximiser la fonction log-vraisemblance,*

$$l(\mathbf{X}, \theta) := \log L(\mathbf{X}, \theta).$$

Puisque le logarithme est une fonction croissante, alors on a

$$\hat{\theta}^{MV} := \arg \max_{\theta} l(\mathbf{X}, \theta).$$

L'estimation ponctuelle d'un paramètre θ qui donne une valeur numérique unique à ce paramètre. Pour évaluer la confiance que l'on peut avoir en une estimation, il est nécessaire de lui associer un intervalle qui contient, avec une certaine probabilité, la vraie valeur du paramètre, c'est **l'estimation par intervalle de confiance**.

Définition 3.1.5 *Une région d'estimation, de niveau [12] de confiance $1 - \alpha$, pour le paramètre θ est un sous-ensemble aléatoire de \mathbb{R}^k contenant θ avec une probabilité égale à $1 - \alpha$. Si on la note par RC , alors on a*

$$\mathbb{P}(\theta \in RC) = 1 - \alpha$$

C'est tout simplement une généralisation au cas multidimensionnel de l'intervalle de confiance univarié.

3.2 Moyenne et matrice de covariance empiriques

Définition 3.2.1 Soit X_1, \dots, X_n un échantillon de taille n d'une population X de dimension p . Les statistiques [12]

$$\bar{X} := (\bar{X}_1, \dots, \bar{X}_p)^t = \frac{1}{n} X^t I_n \text{ et } S := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^t.$$

sont respectivement appelées moyenne et matrice de covariance empiriques.

Remarque 3.2.1 S est une matrice symétrique semi définitive positive. La quantités

$$\bar{x} = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{bmatrix} \text{ et } s = \begin{bmatrix} s_{11} & \cdots & s_{1p} \\ \vdots & \ddots & \vdots \\ s_{p1} & \cdots & s_{pp} \end{bmatrix},$$

sont des réalisations de \bar{X} et S respectivement, où [12]

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \text{ et } s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)^t.$$

Pour ($j = k$) on retrouve les variances empiriques.

Proposition 3.2.1 Soit X_1, \dots, X_n un échantillon d'une distribution multivariée [8] de moyenne μ et de matrice de covariance Σ . Alors on a

1. $\mathbb{E}(\bar{X}) = \mu$ et $\text{cov}(\bar{X}) = \Sigma/n$.
2. $\mathbb{E}(S) = \Sigma$.

Preuve. [8, page 121]

Maintenant, $\bar{X} = (X_1 + \dots + X_n) / n$.

$$\begin{aligned} \mathbb{E}(\bar{X}) &= \mathbb{E}\left(\frac{1}{n}X_1 + \dots + \frac{1}{n}X_n\right) \\ &= \mathbb{E}\left(\frac{1}{n}X_1\right) + \dots + \mathbb{E}\left(\frac{1}{n}X_n\right) \\ &= \frac{1}{n}\mathbb{E}(X_1) + \dots + \frac{1}{n}\mathbb{E}(X_n) = \frac{1}{n}\mu + \dots + \frac{1}{n}\mu \\ &= \mu \end{aligned}$$

Prochain,

$$\begin{aligned} (\bar{X} - \mu)(\bar{X} - \mu)^t &= \left(\frac{1}{n}\sum_{i=1}^n (X_i - \mu)\right) \left(\frac{1}{n}\sum_{j=1}^n (X_j - \mu)\right)^t \\ &= \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n (X_i - \mu)(X_j - \mu)^t \end{aligned}$$

alors

$$\text{cov}(\bar{X}) = \mathbb{E}(\bar{X} - \mu)(\bar{X} - \mu)^t = \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n \mathbb{E}(X_i - \mu)(X_j - \mu)^t$$

Pour $i \neq j$, $\mathbb{E}(\bar{X} - \mu)(\bar{X} - \mu)^t$ est zéro parce que l'entrée est la covariance entre un composant de X_i et un composant de X_j , et ceux-ci sont indépendants.

Donc,

$$\begin{aligned} \text{cov}(\bar{X}) &= \mathbb{E}(\bar{X} - \mu)(\bar{X} - \mu)^t = \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}(X_i - \mu)(X_i - \mu)^t \\ &= \frac{1}{n^2}\sum_{i=1}^n \Sigma = \frac{1}{n^2}n\Sigma = \frac{1}{n}\Sigma \end{aligned}$$

Pour obtenir la valeur attendue de S , nous notons d'abord que $(X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)^t$ est le (j,k) ème élément de $(X_i - \bar{X})(X_i - \bar{X})^t$. La matrice représentant les sommes des

carrés et des produits croisés peut alors être écrite ainsi

$$\begin{aligned}\sum_{i=1}^n (X_i - \bar{X}) (X_i - \bar{X})^t &= \sum_{i=1}^n (X_i - \bar{X}) X_i^t + \sum_{i=1}^n (X_i - \bar{X}) (-\bar{X})^t \\ &= \sum_{i=1}^n X_i X_i^t - n \bar{X} \bar{X}^t\end{aligned}$$

puisque $\sum_{i=1}^n (X_i - \bar{X}) = 0$ et $n \bar{X}^t = \sum_{i=1}^n X_i^t$. Par conséquent, sa valeur attendue est

$$\mathbb{E} \left(\sum_{i=1}^n X_i X_i^t - n \bar{X} \bar{X}^t \right) = \sum_{i=1}^n \mathbb{E} (X_i X_i^t) - n \mathbb{E} \left(\bar{X} \bar{X}^t \right)$$

Pour tout vecteur aléatoire \mathbf{V} avec $\mathbb{E}(\mathbf{V}) = \mu_{\mathbf{V}}$ et $\text{cov}(\mathbf{V}) = \Sigma_{\mathbf{V}}$, on a $\mathbb{E}(\mathbf{V}\mathbf{V}^t) = \Sigma_{\mathbf{V}} + \mu_{\mathbf{V}}\mu_{\mathbf{V}}^t$. Consequently,

$$\mathbb{E}(X_i X_i^t) = \Sigma + \mu\mu^t \text{ et } \mathbb{E}(\bar{X} \bar{X}^t) = \frac{1}{n}\Sigma + \mu\mu^t$$

Utiliser ces résultats, on obtient

$$\sum_{i=1}^n \mathbb{E}(X_i X_i^t) - n \mathbb{E}(\bar{X} \bar{X}^t) = n\Sigma + n\mu\mu^t + n \left(\frac{1}{n}\Sigma + \mu\mu^t \right) = (n-1)\Sigma$$

et donc, depuis $S = (1/n-1) \left(\sum_{i=1}^n X_i X_i^t - n \bar{X} \bar{X}^t \right)$, il s'ensuit immédiatement que

$$\mathbb{E}(S) = \Sigma$$

■

Proposition 3.2.2 *Soit X_1, \dots, X_n un échantillon d'une population multivariée de moyenne μ et de matrice de covariance Σ finie (non singulière). Alors, pour $n-p$ grand, les distributions des statistiques $\sqrt{n}(\bar{X} - \mu)^t$ et $n(\bar{X} - \mu)^t S^{-1}(\bar{X} - \mu)$ sont approximativement $\mathcal{N}_p(0, \Sigma)$ et χ_p^2 respectivement.*

Preuve. Voir [8, page 176] ■

Remarque 3.2.2 Si $n \leq p$ i.e. (taille de l'échantillon) \leq (nombre des variables), alors $\det(S) = 0$ pour tous les échantillons (voir [8, page 71]).

3.3 Méthode du maximum de vraisemblance

Dans ce cas, le vecteur des paramètres θ regroupe à la fois les composantes de l'espérance multidimensionnelle μ et de la matrice de covariance Σ . On cherche à estimer ces deux paramètres. La densité de probabilité de la loi normale multidimensionnelle de paramètres θ est

$$f(x, \mu, \Sigma) = \frac{1}{(\sqrt{2\pi})^p} \frac{1}{\sqrt{\det(\Sigma)}} \exp \left\{ -\frac{1}{2} (x - \mu)^t \Sigma^{-1} (x - \mu) \right\}.$$

Proposition 3.3.1 Les estimateurs du maximum de vraisemblance de μ et Σ sont respectivement

$$\hat{\mu} = \bar{X} \text{ et } \hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^t = \frac{n-1}{n} S.$$

Les valeurs observées \bar{x} et $(n-1)s/n$ sont appelées estimations du maximum de vraisemblance de μ et Σ respectivement.

Preuve. On cherche à maximiser la fonction de vraisemblance La fonction de vraisemblance s'écrit

$$L(\mu, \Sigma) = \frac{1}{(\sqrt{2\pi})^{np}} \frac{1}{(\sqrt{\det(\Sigma)})^n} \exp \left\{ -\frac{1}{2} \left[(n-1) \operatorname{tr}(\Sigma^{-1} S) + n (\bar{X} - \mu)^t \Sigma^{-1} (\bar{X} - \mu) \right] \right\},$$

par rapport à μ puis Σ . Les détails de la démonstration se trouvent dans [8, page 171]. ■

Proposition 3.3.2

1. L'emv $\hat{\mu}$ est sans biais pour μ : $\mathbb{E}(\hat{\mu}) = \mu$.
2. L'emv $\hat{\Sigma}$ est biaisé : $\mathbb{E}(\hat{\Sigma}) = (n-1)\Sigma/n$.

Preuve. Les deux résultats découlent directement du théorème 3.2.1 ■

Remarque 3.3.1

1. Comme S est un estimateur sans biais pour Σ , alors souvent on l'utilise à la place de l'emv $\widehat{\Sigma}$ qui n'est que asymptotiquement sans biais.
2. Lorsque la moyenne μ est connue, la statistique

$$U := \frac{1}{n} \sum_{i=1}^n (\bar{X}_i - \mu) (\bar{X}_i - \mu)^t,$$

fournit une estimation pour Σ meilleure que celle obtenue à partir de S .

3.4 Propriétés des \bar{X} et S

On résume les propriétés des \bar{X} et S dans la proposition suivante.

Proposition 3.4.1 Soit X_1, \dots, X_n un échantillon de loi $\mathcal{N}_p(\mu, \Sigma)$. Alors

1. \bar{X} est distribuée comme $\mathcal{N}_p(\mu, \Sigma/n)$.
2. $(n - 1)S$ suit la loi de Wishart à $(n - 1)$ degrés de liberté.
3. \bar{X} et S sont indépendantes.

Preuve. Voir [8, page 174]. ■

Proposition 3.4.2 Si \bar{X} et S sont basées sur un échantillon multinormal, alors \bar{X} et S sont des statistiques conjointement exhaustives pour μ et Σ .

Preuve. La fonction de vraisemblance s'écrit

$$\begin{aligned} L(\mu, \Sigma) &= \frac{1}{(\sqrt{2\pi})^{np}} \frac{1}{(\sqrt{\det(\Sigma)})^n} \exp \left\{ -\frac{1}{2} \left[(n-1) \operatorname{tr}(\Sigma^{-1}S) + n(\bar{X} - \mu)^t \Sigma^{-1}(\bar{X} - \mu) \right] \right\} \\ &= g(\bar{X}, S, \mu, \Sigma) h(X_1, \dots, X_n), \end{aligned}$$

avec $h(X_1, \dots, X_n) = 1$. ■

Remarque 3.4.1 \bar{X} et S sont conjointement exhaustives pour μ et Σ mais pas individuellement exhaustives.

Proposition 3.4.3 Soit X_1, \dots, X_n un échantillon de $\mathcal{N}_p(\mu, \Sigma)$, alors

$$n(\bar{X} - \mu)^t \Sigma^{-1} (\bar{X} - \mu) \sim \chi_p^2 \text{ et } n(\bar{X} - \mu)^t S^{-1} (\bar{X} - \mu) \sim \mathcal{T}^2(p, n-1).$$

Preuve. Voir [17, page 50]. ■

3.5 Région de confiance pour μ

On distingue deux formes à la région de confiance (appelée hyperellipsoïde) de la moyenne μ , selon que Σ soit connue ou non.

Proposition 3.5.1 Soit $0 < \alpha < 1$, la région de confiance pour μ de niveau $(1 - \alpha)$ est définie par :

1. Σ connue :

$$n(\bar{x} - \mu)^t \Sigma^{-1} (\bar{x} - \mu) \leq \chi_p^2(\alpha),$$

où $\chi_p^2(\alpha)$ est le quantile d'ordre $(1 - \alpha)$ de la loi du Khi-deux à p degrés de liberté (χ_p^2).

2. Σ inconnue :

$$n(\bar{x} - \mu)^t s^{-1} (\bar{x} - \mu) \leq \mathcal{T}_{p, n-1}^2(\alpha),$$

où $\mathcal{T}_{p, n-1}^2(\alpha)$ est le quantile d'ordre $(1 - \alpha)$ de la loi $\mathcal{T}^2(p, n-1)$.

Preuve.

1. En utilisant le résultat du théorème 3.4.4, on construit la région de confiance sous la condition $\mathbb{P}(n(\bar{x} - \mu)^t \Sigma^{-1} (\bar{x} - \mu) \leq \chi_p^2(\alpha)) = 1 - \alpha$

2. En utilisant le deuxième résultat du théorème 3.4.4, on construit la région de confiance sous la condition $\mathbb{P} \left(n (\bar{x} - \mu)^t s^{-1} (\bar{x} - \mu) \leq \mathcal{T}_{p,n-1}^2 (\alpha) \right) = 1 - \alpha$.

■

Intervalle de confiance pour une combinaison linéaire des μ_j

Soit $a = (a_1, \dots, a_p)^t$ un vecteur de \mathbb{R}^p et $a^t \mu$ une combinaison linéaire de μ_1, \dots, μ_p

Proposition 3.5.2 *Si X_1, \dots, X_n est un échantillon de $\mathcal{N}_p(\mu, \Sigma)$, alors l'intervalle de confiance, de niveau de confiance $(1 - \alpha)$, pour $a^t \mu$ est donné par*

$$a^t \bar{x} - t_{n-1}(\alpha/2) \sqrt{\frac{a^t s a}{n}} \leq a^t x \leq a^t \bar{x} + t_{n-1}(\alpha/2) \sqrt{\frac{a^t s a}{n}}.$$

Preuve. Voir [15, page 81]. ■

Remarque 3.5.1 *En prenant \mathbf{a} telque $\mathbf{a}_j = 1$ et $\mathbf{a}_i = 0$ pour $i \neq j$, on construit un intervalle de confiance pour chaque μ_j $j = 1, \dots, p$.*

Conclusion

Dans ce travail, on a passé en revue les résultats fondamentaux relatifs à la distribution multinormale. On s'est particulièrement intéressé à l'estimation, par la méthode du maximum de vraisemblance, de sa moyenne et sa matrice de covariance. L'autre aspect important de l'inférence statistique, à savoir les tests d'hypothèses sur ces deux paramètres, reste à développer. Il est aussi utile de faire des illustrations sur des données simulées pour valider les résultats théoriques obtenus. Des applications sur des données réelles permettront de mettre en valeur l'importance de cette loi dans la vie socio-économique.

Bibliographie

- [1] Anderson, T.W. (2003) . An Introduction to multivariate statistical analysis. Third edition. Wiley.
- [2] Arjun, K.G., Taras, B. (2014) . An exact test about the covariance matrix. Journal of multivariate analysis. 125(2014) 176 – 189.
- [3] Brémaud, P. (2009) . Initiation aux probabilités et aux chaines de markov. Springer.
- [4] Box, P. (1949) . A general distribution theory for class of likelihood criteria. Biometrika. 317-346
- [5] Casella, G.Fienberg, S. Olkin, I. (2002) . Applied multivariate analysis. Springer.
- [6] Caumel, Y. (2011) . Probabilités et processus stochastiques. Springer.
- [7] Helbling, J.M. (2016) . Statistique multivariée, MATH 444.
- [8] Johnson, R., Wichern, D. (2004). Applied multivariate statistical analysis. Sixth edition. Pearson Education.
- [9] Lecoutre, J.P. (2008) . Statistique et probabilités. Dunod.
- [10] Lejeune, M. (2010) . Statistique : la théorie et ses application. Springer.
- [11] Rencher, A.C. (1998) . Multivariate statistical inference and application. Wiley.
- [12] Rencher, A.C., Christensen, W.F. (2012) ; Method of multivariate analysis. Third edition. Wiley ;
- [13] Robbj, M.(1982) . Aspects of multivariate statistical theory. Wiley.
- [14] Saporta, G.(2006). Probabilités analyse des données et statistique. Technip.

- [15] Seber, G.A.F.(1984) . Multivariate observations. Wiley.
- [16] Tassi, P.Legait, S. (1990). Théorie des probabilités en vue des applications statistiques. Technip.
- [17] Tong, Y.L.(1990) . The multivariate normal distribution. Springer.
- [18] Bernard Bru, « La courbe de Gauss ou le théorème de Bernoulli raconté aux enfants », Mathematics and Social Sciences, vol. 175, no 3, 2006, p. 5-23 (lire en ligne [archive]).
- [19] Aimé Fuchs, « Plaidoyer pour la loi normale », Pour la Science, 1995, p. 17 (lire en ligne [archive] [PDF]).
- [20] Wolfgrang, H.Léopold, S.(2007) . Applied multivariate statistical analysis. Springer.

Annexe

A. Rappel d'algèbre linéaire : matrice symétrique définie positive

Une matrice réelle $A = (\alpha_{ij})$, de taille $d \times d$, est dite Symétrique Définie Positive si et seulement si :

1. A est symétrique : $A = A^t$.

2. A est définie positive : pour tout vecteur $x \in \mathbb{R}^d$, $x^t Ax = \sum_{i=1}^d \alpha_{ii} x_i^2 + \sum_{i \neq j}^d \alpha_{ij} x_i x_j > 0$.

Exemple 3.5.1 La matrice $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ est définie positive. En effet,

- A est bien symétrique.
- Pour tout $x = (x_1, x_2, x_3)^t$, on a :

$$\begin{aligned} x^t Ax &= (x_1, x_2, x_3) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\ &= x_1^2 + 3x_2^2 + 2x_3^2 > 0. \end{aligned}$$

B. Abréviations et notations

Les différentes abréviations et notations utilisées dans ce mémoire sont expliquées dans cet annexe.

Abreviation ou Notations	Explication.
c-à-d	C'est a dire.
<i>cov</i>	Covariance.
\xrightarrow{loi}	Converge en loi.
\xrightarrow{p}	Converge en probabilité.
$\det(A)$	Détermine de la matrice A .
\mathbb{R}	Ensemble des nombres réels.
$\mathbb{E}[X]$	Espérance mathématique ou moyenne de vecteur aléatoire.
exp	Exponentiel.
\mathbb{F}	Fonction de répartition.
<i>i.i.d</i>	Indépendantes identiquement distribuées.
A^{-1}	Matrice inverse.
A^t	Matrice transposée.
$\langle \cdot, \cdot \rangle$	Produit scalaire.
\mathbb{P}	Probabilité.
\sum	Somme.
<i>TCLV</i>	Théorème central limite vectoriel.
$tr(\cdot)$	Trace d'une matrice.
<i>Var</i>	Variance.
<i>v.a</i>	variable aléatoire.

المخلص

التوزيع الطبيعي متعدد المتغيرات هو تعميم لتوزيع احتمالي طبيعي احادي الابعاد. و يتميز باهمية كبيرة التي تتركز على مبرهنة النهاية المركزية. في هذه المذكرة سنعينا الى استقراء احصائي الذي يشمل استخدام احصائيات و اخذ العينات العشوائية من المجتمعات التي تتميز بكونها تابعة للتوزيع الطبيعي متعدد المتغيرات. و بهذه الطريقة سنحصل على استنتاجات على بعض الخواص المجهولة للمجتمع.

الكلمات المفتاحية

التوزيع الطبيعي متعدد المتغيرات - التقدير الاحصائي - الاستقلال - مبرهنة النهايات المركزية.

Abdtract

The multivariate normal distribution is a generalization of the very popular unidimensional Gaussian distribution. It has got a very big importance due to the famous central limit theorem. In this dissertation, we were interested in the estimation of the unknown mean and covariance matrix of this distribution.

Keywords

Central limit theorem , Independence , Multivariate normal distribution , Statistical . estimation

Résumé

La distribution normale multivariée est une généralisation de la distribution gaussienne unidimensionnelle très populaire. Il a une très grande importance en raison du célèbre théorème de la limite centrale. Dans cette thèse, nous nous sommes intéressés à l'estimation de la matrice de moyenne et de covariance inconnue de cette distribution.

Mots clés

Estimation statistique ; Indépendance ; Loi normale .multidimensionnelle ; Théorème central limite