

# Table des matières

<b>I</b>	<b>Modèle Linéaire Univarié</b>	<b>3</b>
<b>1</b>	<b>Résultats préliminaires et notations</b>	<b>5</b>
1.1	Introduction et bibliographie. . . . .	5
1.2	Variables aléatoires, moments . . . . .	5
1.2.1	Probabilités et variables aléatoires. . . . .	5
1.2.2	Fonctions de répartition, percentiles. . . . .	7
1.2.3	Fonctions de quantile, médiane, quartiles. . . . .	8
1.2.4	Support et densité d'une loi de probabilité. . . . .	10
1.2.5	Lois conditionnelles. . . . .	10
1.2.6	Moments. . . . .	11
1.2.7	Fonctions caractéristiques et génératrices. . . . .	13
1.2.8	Coefficients de Pearson. . . . .	14
1.2.9	Système de Pearson . . . . .	15
1.3	Mesure empirique et échantillonnage. . . . .	18
1.3.1	Echantillons aléatoires. . . . .	18
1.3.2	Quelques théorèmes limites du calcul des probabilités. . . . .	21
1.3.3	Un exemple commenté. . . . .	22
<b>2</b>	<b>Matrices, notations et factorisations.</b>	<b>27</b>
2.1	Définitions générales et notations. . . . .	27
2.2	Matrices orthogonales. . . . .	29
2.3	Matrices symétriques positives. . . . .	30
2.4	Factorisations à l'aide de matrices triangulaires et orthogonales. . . . .	30
2.5	Produit de Kronecker. . . . .	32
2.6	Opérations Matricielles par Blocs . . . . .	34
2.7	Moments de vecteurs aléatoires . . . . .	36
<b>3</b>	<b>Lois normales.</b>	<b>41</b>
3.1	Lois normales réelles. . . . .	41
3.1.1	Propriétés Générales. . . . .	41
3.1.2	Quantiles de la Loi Normale Standard. . . . .	43
3.2	Lois normales vectorielles . . . . .	44
3.3	Dépendances mutuelles de vecteurs normaux. . . . .	46
3.4	Lois normales matricielles . . . . .	48
<b>4</b>	<b>Lois du <math>\chi^2</math> et de Fisher</b>	<b>53</b>
4.1	Lois du $\chi^2$ Centrées . . . . .	53
4.1.1	Propriétés Générales. . . . .	53
4.1.2	Quantiles supérieurs de la Loi du $\chi^2$ Centrée. . . . .	56
4.2	Lois du $\chi^2$ non Centrées . . . . .	57

4.3	Lois de Fisher . . . . .	65
4.3.1	A. Définition et propriétés élémentaires. . . . .	65
4.3.2	B. Application aux calculs de puissance. . . . .	67
4.4	Loi de Student . . . . .	67
4.4.1	A. Loi de Student centrée. . . . .	67
4.4.2	B. Loi de Student non centrée. . . . .	69
<b>5</b>	<b>Applications Statistiques.</b>	<b>71</b>
5.1	Estimation et tests. . . . .	71
5.2	Utilisation de la loi de Student. . . . .	74
5.2.1	A. Généralités. . . . .	74
5.2.2	B. Calculs de puissance et de tailles d'échantillons. . . . .	76
5.3	Comparaison des moyennes de deux échantillons. . . . .	80
5.3.1	Position du problème. . . . .	80
5.3.2	Cas où les variances sont égales. . . . .	80
5.3.3	Variances quelconques - le problème de Behrens-Fisher. . . . .	81
5.3.4	Utilisation de la loi de Fisher. . . . .	82
5.4	Intervalle de confiance simultanés. . . . .	83
5.5	Utilisation du $T^2$ de Hotelling. . . . .	84
5.6	Utilisation du $\Lambda$ de Wilks. . . . .	85
5.7	Analyse statistique du modèle linéaire simple. . . . .	86
5.7.1	Le modèle linéaire simple. . . . .	86
5.7.2	Plan complet à 0 ou 1 facteur - Echantillon d'une loi $N(\mu, \sigma^2)$ . . . . .	90
5.7.3	Evaluation de la précision des intervalles de confiance pour la moyenne. . . . .	92
5.7.4	Plan complet à 2 facteurs. . . . .	93
5.7.5	Régression linéaire multiple classique. . . . .	96
<b>6</b>	<b>Inverses généralisés</b>	<b>97</b>
6.1	Factorisations de plein rang. . . . .	97
6.2	Pseudoinverses . . . . .	98
<b>II</b>	<b>Modèle Linéaire Multivarié</b>	<b>105</b>
<b>1</b>	<b>Matrices aléatoires</b>	<b>107</b>
1.1	Règles de changement de variables-I. . . . .	107
1.2	Loi gamma multivariée et loi de Wishart. . . . .	110
1.3	Règles de changement de variables - II. . . . .	114
1.4	Variétés de Stiefel, groupe orthogonal et mesure invariante. . . . .	118
1.5	Règles de changement de variables-III. . . . .	120
<b>2</b>	<b>Lois de Wishart et Applications</b>	<b>121</b>
2.1	Loi de Wishart centrée. . . . .	121
2.1.1	Définition de la loi de Wishart centrée. . . . .	121
2.1.2	Loi de Wishart et échantillonnage de la loi normale. . . . .	122
2.1.3	Estimation du Maximum de Vraisemblance de $\mu$ et $\Sigma$ . . . . .	123
2.1.4	Test du rapport de vraisemblance de l'hypothèse de sphéricité. . . . .	126
2.1.5	Densité de la loi de Wishart centrée. . . . .	127
2.1.6	Décomposition de Bartlett associée à la loi de Wishart centrée. . . . .	128
2.1.7	Variance Généralisée. . . . .	129
2.2	Propriétés de la loi de Wishart centrée. . . . .	130
2.2.1	Fonction caractéristique de la loi de Wishart centrée. . . . .	130
2.2.2	Loi de Wishart et régression linéaire. . . . .	132

2.2.3	Loi de Wishart inversée. . . . .	135
2.2.4	La statistique du $T^2$ de Hotelling. . . . .	137
2.3	Le critère de Wilks et ses applications. . . . .	139
2.4	Analyse en composantes principales. . . . .	143
2.4.1	Valeurs propres de matrices aléatoires définies positives. . . . .	143
2.4.2	Densité jointe des valeurs propres. . . . .	145
2.4.3	Valeurs propres d'une matrice de Wishart. . . . .	146
2.4.4	Application au test de sphéricité. . . . .	147
2.4.5	Lois limites des valeurs propres de matrices de Wishart . . . . .	148
<b>3</b>	<b>Polynômes Zonaux et Applications.</b>	<b>151</b>
3.1	Polynômes Zonaux. . . . .	151
3.1.1	Introduction. . . . .	151
3.1.2	Partitions d'un entier. . . . .	151
3.1.3	Polynômes symétriques. . . . .	153
3.2	Polynômes zonaux . . . . .	156
3.2.1	Polynômes matriciels. . . . .	156
3.3	Fonctions hypergéométriques matricielles . . . . .	169
3.4	Loi de Wishart non centrée . . . . .	172
3.5	Propriétés de la loi de Wishart non centrée . . . . .	178
<b>4</b>	<b>Applications statistiques du modèle linéaire multivarié.</b>	<b>183</b>
4.1	Analyse du modèle linéaire multivarié classique. . . . .	183
4.1.1	Le modèle linéaire univarié classique. . . . .	183
4.1.2	Le modèle linéaire multivarié classique. . . . .	186
4.1.3	Analyse de variance multivariée. . . . .	190
4.2	Test d'homogénéité des moyennes (en cours de rédaction). . . . .	191
4.3	Test d'homogénéité des matrices de variances-covariances (en cours de rédaction). . . . .	191
<b>5</b>	<b>Distributions Elliptiques et Sphériques (en cours de rédaction).</b>	<b>193</b>
5.0.1	Estimation des Coefficients de Corrélacion. . . . .	194

# Cours d'Analyse Statistique Multivariée

Paul Deheuvels

2012-2013



Première partie

**Modèle Linéaire Univarié**



# Chapitre 1

## Résultats préliminaires et notations

### 1.1 Introduction et bibliographie.

Le présent ensemble de notes de cours, en deux chapitres I & II, ne représente qu'une partie de l'enseignement dispensé sur le sujet à l'Université Pierre et Marie Curie (Paris VI). Il est conseillé de le compléter par la lecture de tout ou partie des excellents ouvrages suivants.

- Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.
- Arnold, S. F. (1981). *The Theory of Linear Models and Multivariate Analysis*. Wiley, New York.
- Johnson, N. L., Kotz, S. et Balakrishnan, N. (1995). *Continuous Univariate Distributions*. Vol. 1 et 2 (2nd. Ed.), Wiley, New York.
- Johnson, N. L., Kotz, S. et Kemp, A. W. (1992). *Univariate Discrete Distributions*. (2nd. Ed.) Wiley, New York.
- Johnson, R. A., et Wichern, D. W. (1998). *Applied Multivariate Statistical Analysis*. (4th. Ed.). Prentice Hall, Upper Saddle River.
- Kendall, M. G. et Stuart, A. (1973). *The Advanced Theory of Statistics*. Vol. 1 : *Distribution Theory*, Vol. 2 : *Inference and Relationship*, Vol. 3 : *Design and Analysis, and Time-Series*. Griffin, Londres.
- Kshirsagar, A. M. (1972). *Multivariate Analysis*. Dekker, New York.
- Mace, A. E. (1964). *Sample Size Determination*. Reinhold Publishing Corporation, New York.
- Miller, R. G. Jr. (1981). *Simultaneous Statistical Inference*. (2nd. Ed). Springer, New York.
- Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.
- Patel, J. K. et Read, C. B. (1996). *Handbook of the Normal Distribution*. Dekker, New York.
- Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. (2nd. Ed.) Wiley, New York.
- Rao, C. R. et Mithra, S. K. (1971). *Generalized Inverse of Matrices and Its Applications*. Wiley, New York.
- Scheffé, H. (1959). *The Analysis of Variance*. Wiley, New York.
- Searle, S. R. (1971). *Linear Models*. Wiley, New York.
- Seber, G. A. F. (1977). *Linear Regression Analysis*. Wiley, New York.
- Srivastava, M. S. et Khatri, C. G. (1979). *An Introduction to Multivariate Analysis*. North Holland, Amsterdam.
- Timm, N. H. (2002). *Applied Multivariate Analysis*. Springer, New York.

### 1.2 Variables aléatoires, moments

#### 1.2.1 Probabilités et variables aléatoires.

Dans ce qui suit, nous supposons implicitement que les objets aléatoires considérés sont définis sur un même *espace de probabilités*  $(\Omega, \mathcal{A}, \mathbb{P})$ , où  $\Omega$  désigne un ensemble, dont les éléments sont interprétés comme des *événements élémentaires*, et  $\mathbb{P}$  une probabilité définie sur une *tribu*  $\mathcal{A}$  d'*événements probabilisables* (identifiés



à des parties de  $\Omega$ ). Par définition,  $\mathcal{A}$  est une *sous-tribu* de la tribu  $\mathcal{P}(\Omega)$  des parties de  $\Omega$ . Nous rappelons la signification de ce vocabulaire. Une *tribu* de parties de  $\Omega$  est une partie de l'ensemble  $\mathcal{P}(\Omega)$  des parties de  $\Omega$ , caractérisé par les propriétés (i – ii – iii) suivantes. On note  $\emptyset$  l'ensemble vide. Si  $A, B \in \mathcal{A}$  et  $A_1, A_2, \dots \in \mathcal{A}$ , alors

- (i)  $\Omega \in \mathcal{A}$  et  $\emptyset \in \mathcal{A}$ ;
- (ii)  $A \cup B \in \mathcal{A}$ ,  $A \cap B \in \mathcal{A}$ ,  $A - B \in \mathcal{A}$ ;
- (iii)  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$  et  $\bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$ .

Ici, et dans la suite du présent exposé, nous notons  $\infty$  pour  $+\infty$ . Une sous-tribu de  $\mathcal{A}$  est une tribu contenue dans  $\mathcal{A}$ .

Par convention, un *événement* quelconque est une partie  $A \subseteq \Omega$ , de  $\Omega$ , constitué d'une réunion d'événements élémentaires. Le formalisme probabiliste suppose que la probabilité est initialement définie, seulement, sur une partie restreinte  $\mathcal{A}$  de  $\mathcal{P}(\Omega)$ . Dans ce modèle, on ne connaît donc pas nécessairement la probabilité d'une partie (ou événement) quelconque  $A \subseteq \Omega$ .

On dira que  $A \in \mathcal{A}$  est un événement *mesurable* ou *probabilisable*, et on notera  $\mathbb{P}(A)$  sa probabilité. Par définition, une probabilité  $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$  est une application vérifiant les propriétés suivantes. Si  $A, B \in \mathcal{A}$  et  $A_1, A_2, \dots \in \mathcal{A}$ , alors

- (P.1)  $\mathbb{P}(\Omega) = 1$  et  $\mathbb{P}(\emptyset) = 0$ ;
- (P.2)  $A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$ ;
- (P.3)  $A \cap B = \emptyset \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ ;
- (P.4)  $A_1 \subseteq A_2 \subseteq \dots \subseteq A_n \subseteq \dots \Rightarrow \mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n)$ ;
- (P.5)  $A_i \cap A_j = \emptyset \quad \forall i \neq j \geq 1 \Rightarrow \mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n \geq 1} \mathbb{P}(A_n)$ .

Le cas particulier le plus important, pour ce qui concerne la suite de notre cours, est celui des lois de probabilité de *variables aléatoires* à valeurs dans  $\mathbb{R}^p$  (ou dans un espace de dimension finie sur  $\mathbb{R}$ ), auquel cas  $\Omega = \mathbb{R}^p$  et  $\mathcal{A}$  est (au minimum) la *tribu borélienne* de  $\mathbb{R}^p$ . D'une manière générale, la plus petite tribu de parties d'un espace topologique  $\mathcal{T}$ , comprenant les ouverts et les fermés de cet espace, est appelée *tribu borélienne* (de parties) de  $\mathcal{T}$ . Cette définition a un sens, dans la mesure où une intersection quelconque de tribus est toujours une tribu. La tribu borélienne est donc la tribu intersection de toutes les tribus contenant les ouverts et les fermés. Ce dernier ensemble n'est pas vide car il contient  $\mathcal{P}(\Omega)$ . Dans le cas où l'espace de référence est  $\mathcal{T} = \mathbb{R}^p$ , muni de sa topologie usuelle, la tribu borélienne est notée  $\mathcal{B}_p$  (ou  $\mathcal{B}$  lorsque la dimension de l'espace n'a pas besoin d'être spécifiée), et un élément  $B$  de  $\mathcal{B}_p$  est appelé *partie borélienne* de  $\mathbb{R}^p$ . Dans ce qui suit, nous supposons implicitement, et sauf mention du contraire, que l'espace de probabilité  $(\Omega, \mathcal{A}, \mathbb{P})$  est tel que  $\mathcal{B}_p \subseteq \mathcal{A}$ .

Nous ne nous étendons pas ici sur les aspects généraux de la théorie de la mesure, qui permettent de prolonger la définition d'une probabilité, initialement définie sur une tribu  $\mathcal{A}$  d'événements, à une tribu plus étendue  $\mathcal{A}_1 \supseteq \mathcal{A}$ , contenant la tribu originale  $\mathcal{A}$ . Une telle extension, pour  $\Omega$  et  $\mathcal{A}$  donnés, dépend naturellement de  $\mathbb{P}$ . De ce fait, si on considère plusieurs probabilités  $\mathbb{P}_1, \mathbb{P}_2, \dots$ , définies sur le même couple  $(\Omega, \mathcal{A})$ , il est plus commode de raisonner sur une même tribu de base, plutôt que sur celles qui correspondent aux extensions ultimes de la mesure, pour chacune des probabilités considérées. Tout particulièrement dans l'étude de variables aléatoires à valeurs dans  $\Omega = \mathbb{R}^p$ , il sera, dans la plupart des applications utiles, suffisant de se limiter à des propriétés caractérisées à partir de la *tribu borélienne*  $\mathcal{B}_p$  de  $\mathbb{R}^p$ . Une fonction  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  sera alors dite *mesurable* (pour *Borel-mesurable*) si, pour toute partie borélienne  $A \in \mathcal{B}$  de  $\mathbb{R}$ , l'image réciproque de  $A$  par  $f$ , soit  $f^{-1}(A) = \{x \in \mathbb{R}^p : f(x) \in A\} \in \mathcal{B}_p$ , est une partie borélienne de  $\mathbb{R}^p$ .

Un *vecteur aléatoire* [v.a.]  $Z \in \mathbb{R}^p$  (ou *variable aléatoire à valeurs dans  $\mathbb{R}^p$* ), défini sur un espace de probabilités général  $(\Omega, \mathcal{A}, \mathbb{P})$ , est constitué par une application  $Z : \Omega \rightarrow Z(\omega)$  de  $\Omega$  à valeurs dans  $\mathbb{R}^p$  tel que, pour tout

$B \in \mathcal{B}_p$ , l'événement défini par

$$\{Z \in B\} = Z^{-1}(B) = \{\omega \in \Omega : Z(\omega) \in B\}$$

appartienne à  $\mathcal{A}$ , de sorte qu'on puisse en définir sa probabilité  $\mathbb{P}(Z \in B)$ . La mesure borélienne  $\mathcal{M} = \mathcal{M}_Z$  induite sur  $(\mathbb{R}^p, \mathcal{B}_p)$  par les relations

$$B \in \mathcal{B}_p \rightarrow \mathbb{P}(Z \in B) = \mathcal{M}_Z(B) = \mathbb{P}_Z(B),$$

est appelée *loi* ou *loi de probabilité* de  $Z \in \mathbb{R}^p$ .

### 1.2.2 Fonctions de répartition, percentiles.

La loi de probabilité d'une variable aléatoire réelle  $Z$ , (souvent désignée par la suite par  $\mathcal{L}(Z)$  ou  $\mathcal{L}_Z$ , avec la notation  $Z \equiv \mathcal{L}(Z)$ , ou  $Z \stackrel{d}{=} \mathcal{L}(Z)$ , pour exprimer que  $Z$  suit la loi de probabilité  $\mathcal{L}(Z)$ ), est complètement caractérisée par la *fonction de répartition* de  $Z$ , définie, lorsque

$$Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_p \end{bmatrix} = [Z_1 \quad \cdots \quad Z_p]' \in \mathbb{R}^p$$

a pour coordonnées  $Z_1, \dots, Z_p$  dans la base canonique de  $\mathbb{R}^p$ , par

$$\mathbb{F}_Z(z_1, \dots, z_p) = \mathbb{P}(Z_1 \leq z_1, \dots, Z_p \leq z_p) = \mathcal{L}_Z\left(\prod_{i=1}^p (-\infty, z_i]\right). \quad (1.2.1)$$

Dans (1.2.1) et dans la suite de notre exposé,  $M'$  désigne la *matrice transposée* de la matrice  $M$ .

Dans le cas particulier où  $p = 1$ ,  $Z$  est une *variable aléatoire réelle* [en abrégé, variable aléatoire, v.a. ou v.a.r.], et sa fonction de répartition, sous sa version standard, continue à droite, soit

$$\mathbb{F}_Z(z) = \mathbb{P}(Z \leq z) = \mathcal{L}_Z((-\infty, z]),$$

est alors une fonction de variable réelle, définie sur  $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\} \cup \{\infty\}$ , et possédant les propriétés caractéristiques suivantes.

$$(F.1) \quad \forall x, y \in \mathbb{R} : x \leq y \implies 0 \leq \mathbb{F}_Z(x) \leq \mathbb{F}_Z(y) \leq 1;$$

$$(F.2) \quad \lim_{x \rightarrow -\infty} \mathbb{F}_Z(x) = 0 \text{ et } \lim_{x \rightarrow \infty} \mathbb{F}_Z(x) = 1;$$

$$(F.3) \quad \mathbb{F}_Z \text{ est continue à droite : } \forall z \in \mathbb{R},$$

$$\mathbb{F}_{Z+}(z) = \mathbb{F}_Z(z+0) = \lim_{\varepsilon \downarrow 0} \mathbb{F}_Z(z + \varepsilon) = \mathbb{F}_Z(z).$$

La propriété (F.2) (conséquence des propriétés de *continuité dénombrable* (P.4 – 5) de la probabilité) montre que la définition de  $\mathbb{F}_Z$  sur  $\mathbb{R}$  détermine les valeurs de  $\mathbb{F}_Z$  sur  $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\} \cup \{\infty\}$ . On a, en effet,

$$\mathbb{F}_Z(-\infty) = 0 \quad \text{et} \quad \mathbb{F}_Z(\infty) = 1.$$

Les conditions (F.1-2-3) sont *nécessaires et suffisantes* pour caractériser la fonction de répartition d'une v.a.r., au sens que, réciproquement, toute fonction  $\mathbb{F}_Z$  satisfaisant (F.1)–(F.2)–(F.3) est la fonction de répartition d'une loi de probabilité sur  $\mathbb{R}$ . Cette définition est rendue possible par l'intégrale de Lebesgue-Stieltjes

$$\mathbb{P}_Z(A) = \int_A d\mathbb{F}_Z(z), \quad \forall A \in \mathcal{B}, \quad (1.2.2)$$

qui permet de donner un sens à  $\mathbb{P}_Z(A)$  pour tout  $A \in \mathcal{B}$ . La *version régularisée à gauche* (notée  $\mathbb{F}_{Z-}(z)$  ou  $\mathbb{F}_Z(z-0)$ ) de la fonction de répartition  $\mathbb{F}_Z(z)$  est définie pour tout  $z \in \mathbb{R}$  par

$$\mathbb{F}_{Z-}(z) = \mathbb{F}_Z(z-0) = \lim_{\varepsilon \downarrow 0} \mathbb{F}_Z(z-\varepsilon) = \mathbb{P}(Z < z) \leq \mathbb{F}_Z(z). \quad (1.2.3)$$

On observera que  $\mathbb{F}_{Z-}(z)$ , définie par (1.2.3), est une fonction *continue à gauche* de  $z \in \mathbb{R}$ , c'est à dire, vérifiant l'identité  $\mathbb{F}_{Z-}(z) = \lim_{\varepsilon \downarrow 0} \mathbb{F}_{Z-}(z-\varepsilon)$  pour tout  $z \in \mathbb{R}$ . Le choix de l'une ou l'autre des versions  $\mathbb{F}_{Z-}(z)$  ou  $\mathbb{F}_Z(z)$  comme *forme standard* de la fonction de répartition de  $Z$ , est une affaire d'habitude, relevant des conventions et de l'usage le plus fréquent. Nous faisons ici usage de la notation, majoritairement utilisée dans la littérature scientifique, consistant à utiliser, comme forme standard de la fonction de répartition de  $Z$ , sa *version régularisée à droite*, donnée par la fonction  $\mathbb{F}_Z(z) = \mathbb{F}_Z(z+0) = \mathbb{P}(Z \leq z)$ .

**Définition 1.2.1.** *Pour tout  $x \in \mathbb{R}$ , on appelle percentile d'ordre  $x$  de la fonction de répartition  $\mathbb{F}_Z$ , tout nombre  $p_x$  vérifiant les inégalités*

$$\mathbb{F}_{Z-}(x) \leq p_x \leq \mathbb{F}_Z(x). \quad (1.2.4)$$

**Exercice 1.2.1.** 1°) *Vérifier que les propriétés (P.1)–(P.3)–(P.4) impliquent (P.2)–(P.5).*

2°) *Montrer que (F.1)–(F.2)–(F.3) sont conséquences de (P.1)–(P.3)–(P.4).*

**Exercice 1.2.2.** 1°) *Montrer que, pour toute fonction de répartition  $\mathbb{F}_Z$  d'une loi de probabilité sur  $\mathbb{R}$ , l'ensemble  $\mathcal{D}_{\mathbb{F}_Z} = \{x : \mathbb{F}_{Z-}(x) < \mathbb{F}_Z(x)\}$  des points de discontinuité de  $F$  est, ou bien vide, fini ou dénombrable.*

2°) *En déduire que  $\mathbb{F}_Z$  est parfaitement définie par la donnée des valeurs de  $\mathbb{F}_Z(x)$  lorsque  $x$  varie dans l'ensemble  $\mathcal{C}_{\mathbb{F}_Z} = \{x : \mathbb{F}_{Z-}(x) = \mathbb{F}_Z(x)\}$  des points de continuité de  $\mathbb{F}_Z$ .*

### 1.2.3 Fonctions de quantile, médiane, quartiles.

Comme nous l'avons mentionné plus haut, la loi de probabilité d'une v.a.r  $Z$ , de fonction de répartition  $\mathbb{F}_Z$ , est caractérisée par  $\mathbb{F}_Z$ . Cette loi se caractérise, de manière équivalente, par sa *fonction de quantile*. Cette dernière fonction est définie sur  $(0, 1)$  par

$$\mathbb{Q}_Z(s) = \mathbb{F}_Z^{\leftarrow}(s) = \inf\{x : \mathbb{F}_Z(x) \geq s\} = \sup\{x : \mathbb{F}_Z(x) < s\},$$

(souvent aussi notée  $\mathbb{F}^{-1}(s)$  lorsque cette notation n'est pas ambiguë). La fonction de quantile  $\mathbb{Q}_Z(s)$  est toujours définie en chaque valeur de  $s \in (0, 1)$ . Nous notons  $(a, b)$  (plutôt que  $]a, b[$ ) l'intervalle ouvert d'extrémités  $a, b$  et adoptons des notations analogues pour les intervalles semi-ouverts. Pour donner un sens à  $\mathbb{Q}_Z(1)$  et  $\mathbb{Q}_Z(0)$ , il est commode de prolonger la fonction de quantile  $\mathbb{Q}_Z$  aux extrémités de l'intervalle  $(0, 1)$ , en posant

$$\mathbb{Q}_Z(0) = \lim_{s \downarrow 0} \mathbb{Q}_Z(s) = \inf\{x : \mathbb{F}_Z(x) > 0\},$$

$$\mathbb{Q}_Z(1) = \lim_{s \uparrow 1} \mathbb{Q}_Z(s) = \sup\{x : \mathbb{F}_Z(x) < 1\}.$$

Il importe, au passage, de noter que les valeurs correspondantes de  $\mathbb{Q}_Z(0)$  et  $\mathbb{Q}_Z(1)$  peuvent être infinies.

Pour une valeur quelconque de  $\alpha \in (0, 1)$ , le nombre  $\mathbb{Q}_Z(\alpha)$  est appelé *quantile d'ordre  $\alpha$*  de  $Z$  (ou de la loi  $\mathcal{L}(Z)$  de  $Z$ ). De même, le nombre  $\mathbb{Q}_Z(1-\alpha)$  est appelé *quantile supérieur d'ordre  $\alpha$*  de  $Z$  (ou de la loi  $\mathcal{L}(Z)$  de  $Z$ ). D'une manière générale, le quantile d'ordre  $\alpha$  de  $Z$  n'est pas nécessairement unique. On adopte la définition suivante (qui reprend la définition 1.2.1). Dans ce qui suit, on écrit  $a := b$  (ou  $b =: a$ ) pour signifier que la valeur de  $a$  est définie par celle de  $b$ .

**Définition 1.2.2.** *Pour tout  $\alpha \in (0, 1)$ , on appelle quantile d'ordre  $\alpha$  de  $Z$  tout nombre  $q_\alpha \in \mathbb{R}$  vérifiant les inégalités*

$$\mathbb{Q}_Z(\alpha) \leq q_\alpha \leq \mathbb{Q}_{Z+}(\alpha) := \lim_{\varepsilon \downarrow 0} \mathbb{Q}_Z(\alpha + \varepsilon). \quad (1.2.5)$$

*Pour tout  $x \in \mathbb{R}$ , on appelle percentile d'ordre  $x$  de  $Z$  tout nombre  $p_x \in \mathbb{R}$  vérifiant les inégalités*

$$\mathbb{F}_{Z-}(x) \leq p_x \leq \mathbb{F}_Z(x). \quad (1.2.6)$$

Pour tout  $\alpha \in (0, 1)$ , les quantiles  $\mathbb{Q}_Z(\alpha)$  et  $\mathbb{Q}_Z(1 - \alpha)$  vérifient les relations

$$\mathbb{Q}_Z(\alpha) = \inf\{x : \mathbb{F}_Z(x) \geq \alpha\} = \inf\{x : 1 - \mathbb{F}_Z(x) \leq 1 - \alpha\}, \quad (1.2.7)$$

$$\mathbb{Q}_Z(1 - \alpha) = \inf\{x : \mathbb{F}_Z(x) \geq 1 - \alpha\} = \inf\{x : 1 - \mathbb{F}_Z(x) \leq \alpha\}. \quad (1.2.8)$$

Nous insistons sur le fait qu'avec les définitions adoptées ici, alors que la fonction  $\mathbb{F}_Z$  est *continue à droite*, c'est à dire telle que  $\mathbb{F}_Z(z) = \mathbb{F}_Z(z + 0) = \lim_{\varepsilon \downarrow 0} \mathbb{F}_Z(z + \varepsilon)$ , la fonction de quantile  $\mathbb{Q}_Z$  est, quant à elle, *continue à gauche* c'est à dire telle que

$$\mathbb{Q}_Z(s) = \mathbb{Q}_{Z-}(s) = \mathbb{Q}_Z(s - 0) = \lim_{\varepsilon \downarrow 0} \mathbb{Q}_Z(s - \varepsilon). \quad (1.2.9)$$

Il est également important de constater que la *version régularisée à droite*  $\mathbb{Q}_{Z+}(s)$  de la fonction de quantiles  $\mathbb{Q}_Z(s)$ , vérifie, pour tout  $s \in (0, 1)$ ,

$$\mathbb{Q}_{Z+}(s) = \mathbb{Q}_Z(s + 0) := \lim_{\varepsilon \downarrow 0} \mathbb{Q}_Z(s + \varepsilon) = \inf\{x : \mathbb{F}_Z(x) > s\} = \sup\{x : \mathbb{F}_Z(x) \leq s\}. \quad (1.2.10)$$

Les inégalités suivantes se déduisent sans grande difficulté des définitions précédentes. On a, pour tout  $s \in [0, 1]$  et  $x \in \mathbb{R}$ , en posant  $\mathbb{Q}_{Z+}(0) = \mathbb{Q}_Z(0)$  et  $\mathbb{Q}_{Z+}(1) = \mathbb{Q}_Z(1)$ ,

$$\mathbb{F}_{Z-}(\mathbb{Q}_Z(s)) \leq \mathbb{F}_{Z-}(\mathbb{Q}_{Z+}(s)) \leq s \leq \mathbb{F}_Z(\mathbb{Q}_Z(s)) \leq \mathbb{F}_Z(\mathbb{Q}_{Z+}(s)), \quad (1.2.11)$$

$$\mathbb{Q}_Z(\mathbb{F}_{Z-}(x)) \leq \mathbb{Q}_Z(\mathbb{F}_Z(x)) \leq x \leq \mathbb{Q}_{Z+}(\mathbb{F}_{Z-}(x)) \leq \mathbb{Q}_{Z+}(\mathbb{F}_Z(x)). \quad (1.2.12)$$

On en déduit la proposition suivante, qui complète la Définition 1.2.2.

**Proposition 1.2.1.** *Pour tout  $\alpha \in (0, 1)$ , un quantile d'ordre  $\alpha$  de  $Z$  est un nombre quelconque  $q_\alpha$ , vérifiant les inégalités*

$$\mathbb{F}_{Z-}(q_\alpha) = \mathbb{P}(Z < q_\alpha) \leq \alpha \leq \mathbb{P}(Z \leq q_\alpha) = \mathbb{F}_{Z+}(q_\alpha). \quad (1.2.13)$$

*Pour tout  $x \in \mathbb{R}$ , un percentile d'ordre  $x$  de  $Z$  est un nombre quelconque  $p_x \in \mathbb{R}$ , vérifiant les inégalités*

$$\mathbb{Q}_Z(p_x) \leq x \leq \mathbb{Q}_{Z+}(p_x). \quad (1.2.14)$$

**Preuve.** Par application de (1.2.11)–(1.2.12).□

Comme cas particulier de (1.2.13), on constate que, pour une valeur quelconque de  $\alpha \in (0, 1)$ , le *quantile supérieur*  $\mathbb{Q}_Z(1 - \alpha)$  d'ordre  $\alpha$  (ou, de manière équivalente, le *quantile d'ordre*  $1 - \alpha$ ) de  $Z$  (ou de la loi  $\mathcal{L}(Z)$  de  $Z$ ), vérifie les relations

$$\mathbb{Q}_Z(1 - \alpha) = \inf\{x : 1 - \mathbb{F}_Z(x) \leq \alpha\} = \inf\{x : \mathbb{F}_Z(x) \geq 1 - \alpha\}.$$

Les quantiles d'ordre  $\alpha$  pour les valeurs particulières de  $\alpha = 1/2$ , et  $\alpha = 1/4$  ou  $3/4$  jouent un rôle particulier en statistique, et sont appelés, respectivement, *médiane* et *quartiles*.

C'est ainsi qu'en spécialisant (1.2.13) au cas où  $\alpha = 1/2$ , on appelle habituellement *médiane* de  $Z$  tout nombre  $m$  tel que

$$\mathbb{P}(Z \leq m) \geq 1/2 \quad \text{et} \quad \mathbb{P}(Z \geq m) \geq 1/2. \quad (1.2.15)$$

On constate que cette définition a un sens, et que tout nombre  $m$  vérifiant

$$\mathbb{Q}_Z(1/2) \leq m \leq \mathbb{Q}_{Z+}(1/2), \quad (1.2.16)$$

est une valeur *admissible* (au sens de la définition (1.2.15)) de la médiane. Le choix *standard* de la médiane est habituellement donné par

$$\text{med}(Z) = \mathbb{Q}_Z(1/2), \quad (1.2.17)$$

mais ceci est une affaire de conventions, car on pourrait tout aussi bien choisir  $\mathbb{Q}_{Z+}(1/2)$ . Le choix de  $\mathbb{Q}_Z(1/2)$  représente l'unique possibilité lorsque  $\mathbb{Q}_Z$  est continue en  $1/2$ . Lorsque  $\mathbb{Q}_Z$  n'est pas continue en  $1/2$ , il existe plusieurs définitions alternatives possibles de la *médiane* de  $Z$ , la plus classique d'entre elles étant donnée par

$$\text{med}(Z) = \frac{1}{2} \left\{ \mathbb{Q}_Z(1/2) + \mathbb{Q}_{Z+}(1/2) \right\}. \quad (1.2.18)$$

**Exercice 1.2.3.** On considère une v.a. binomiale  $X \stackrel{d}{=} B(n, \frac{1}{2})$ , de loi de probabilité définie par

$$\mathbb{P}(X = k) = 2^{-n} \binom{n}{k} \quad \text{pour } k = 0, \dots, n.$$

Préciser, en fonction de la parité de  $n = 2m$  ou  $n = 2m + 1$ , les valeurs admissibles de la médiane de  $X$ .

### 1.2.4 Support et densité d'une loi de probabilité.

Le *support* de (la loi de probabilité) du vecteur aléatoire  $Z \in \mathbb{R}^p$  est, par définition, l'ensemble de tous les points  $x \in \mathbb{R}^p$ , tels que tout voisinage  $V$  de  $x$  ait une probabilité,  $\mathcal{L}_Z(V)$ , strictement positive, soit, telle que

$$\mathcal{L}_Z(V) = \mathbb{P}(Z \in V) > 0.$$

On désigne le support de  $Z$  par  $\text{supp}(Z)$  ou  $\text{supp } \mathcal{L}(Z)$ . Avec ces notations, la probabilité que  $Z$  appartienne à  $V$  est donc strictement positive pour tout choix du voisinage  $V$  de  $x \in \text{supp}(Z)$ . Dans tous les cas,  $\text{supp}(Z)$  constitue une partie fermée de  $\mathbb{R}^p$ .

Lorsque la loi de probabilité  $\mathcal{M}_Z$  de  $Z \in \mathbb{R}^p$  est *absolument continue* relativement à la mesure de Lebesgue  $\lambda$  sur  $\mathbb{R}^p$  (ce qui est noté  $\mathcal{M}_Z \ll \lambda$ , et qui exprime le fait que  $\lambda(A) = 0 \Rightarrow \mathcal{L}_Z(A) = 0$  pour tout  $A \in \mathcal{B}$ ), on désigne par

$$f = f_Z = \frac{d\mathcal{M}_Z}{d\lambda},$$

la *densité* de  $Z$  relativement à  $\lambda$ , dont l'existence est garantie par le théorème de Radon-Nikodym. Elle vérifie, par définition

$$\mathcal{L}_Z(A) = \mathbb{P}(Z \in A) = \int_A f_Z(z) dz \quad \text{pour tout } A \in \mathcal{B}.$$

La fonction  $f_Z$  est définie de manière unique sur  $\mathbb{R}^p$ , à une égalité près sur un ensemble de mesure de Lebesgue nulle (soit *presque partout*, en abrégé p.p.). L'une quelconque parmi l'ensemble des densités possibles de  $Z$  est appelée une *version* de la densité de  $Z$ . En particulier, lorsque  $f_Z$  existe, et lorsque la fonction de répartition  $\mathbb{F}_Z(z_1, \dots, z_p)$  admet des dérivées partielles continues d'ordre suffisant, on a l'identité p.p. sur  $\mathbb{R}^p$

$$f_Z(z_1, \dots, z_p) = \frac{\partial^p}{\partial z_1 \dots \partial z_p} \mathbb{F}_Z(z_1, \dots, z_p).$$

On dit que la densité  $f_Z$  de  $Z$  est *continue*, s'il existe une version continue de cette densité. Cette dernière est alors, nécessairement unique. On parlera alors de *la* densité de  $Z$  (par abus de langage). Dans ce dernier cas, le support de  $Z$  s'identifie avec l'adhérence (topologique) de l'ensemble  $\{z : f_Z(z) > 0\}$ .

**Exercice 1.2.4.** On considère une v.a.  $X$  de loi discrète, définie par

$$\mathbb{P}(X = x_j) = p_j \quad \text{pour } j = 1, \dots, N,$$

où  $x_1, \dots, x_N \in \mathbb{R}$  sont des réels distincts, et  $p_1, \dots, p_N$  des nombres tels que  $p_j \geq 0$  pour  $j = 1, \dots, N$ , et  $p_1 + \dots + p_N = 1$ . Déterminer le support  $\text{supp}(X)$  de (la loi de)  $X$ .

### 1.2.5 Lois conditionnelles.

Nous nous limiterons à la donnée de notations minimales, pour ce qui concerne la description des lois conditionnelles. Si  $\mathcal{A}_0 \subseteq \mathcal{A}$  désigne une sous-tribu particulière de la tribu  $\mathcal{A}$  des événements probabilisables de  $\Omega$ , pour tout événement probabilisable  $E \in \mathcal{A}$ , on note  $\mathbb{P}(E|\mathcal{A}_0)$  la *probabilité conditionnelle* de  $E$  sachant (la sous-tribu)  $\mathcal{A}_0$ . On note de même  $\mathbb{P}(E|Z)$  la probabilité de  $E$  sachant  $Z$  lorsque  $Z$  est une variable aléatoire (dans ce dernier cas,  $\mathcal{A}_0$  est la tribu engendrée par  $Z$ , c'est à dire la plus petite sous-tribu de  $\mathcal{A}$  contenant l'ensemble des  $Z^{-1}(B)$  pour  $B \in \mathcal{B}$ ).

### 1.2.6 Moments.

Soit  $X \in \mathbb{R}$  une variable aléatoire réelle de fonction de répartition  $\mathbb{F}$  et de fonction de quantile  $\mathbb{Q}$ . Le fait que  $d\mathbb{F}$  définisse une mesure de Lebesgue-Stieltjes sur  $\mathbb{R}$  permet de donner un sens à des intégrales de la forme

$$\int_{\mathbb{R}} g(x) d\mathbb{F}(x), \quad (1.2.19)$$

lorsque  $g(\cdot)$  est une fonction mesurable positive ou nulle. Plus généralement, si  $g(\cdot)$  est mesurable mais de signe quelconque, l'intégrale (1.2.19) a un sens (pour l'intégrale de Lebesgue-Stieltjes), *si et seulement si*

$$\int_{\mathbb{R}} |g(x)| d\mathbb{F}(x) < \infty. \quad (1.2.20)$$

Dans ce dernier cas, la valeur de (1.2.19) est appelée *espérance* de  $g(X)$ , et notée

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}} g(x) d\mathbb{F}(x). \quad (1.2.21)$$

Les *moments* de  $X$  sont obtenus lorsque  $g(x) = x^r$  est une puissance de  $x$ . Ils jouent un rôle théorique et pratique important. En spécialisant (1.2.19)–(1.2.21) au cas où  $g(x) = x$ , on définit, sous réserve que  $\mathbb{E}(|X|) < \infty$ , l'*espérance* de  $X$  par

$$\mu = \mu_X = \mathbb{E}(X) = \int_{-\infty}^{\infty} x d\mathbb{F}(x) = \int_0^1 \mathbb{Q}(s) ds.$$

On observera, comme cas particulier de (1.2.19)–(1.2.20), que la première des deux intégrales ci-dessus a un sens, en tant qu'intégrale de Lebesgue-Stieltjes, *si et seulement si*  $\mathbb{E}(|X|) < \infty$ . On a donc l'équivalence

$$\mathbb{E}(|X|) = \int_{-\infty}^{\infty} |x| d\mathbb{F}(x) < \infty \quad \Leftrightarrow \quad \mathbb{E}(X) = \int_{-\infty}^{\infty} x d\mathbb{F}(x) \quad \text{existe (dans } \mathbb{R}\text{)}.$$

Cependant, dans le cas où  $\mathbb{P}(X \geq 0) = 1$ , la définition de  $\mathbb{E}(X)$  a toujours un sens, mais cette fois-ci dans  $\mathbb{R} \cup \{\infty\}$ , en tant que nombre possiblement infini, c'est à dire, vérifiant  $\mathbb{E}(X) \in \mathbb{R} \cup \{\infty\}$ . D'une manière générale, nous dirons toutefois que  $\mathbb{E}(X)$  *existe* dans le *seul cas* où  $\mathbb{E}(X) \in \mathbb{R}$ , ce qui *équivalent* donc à  $\mathbb{E}(|X|) < \infty$ .

Lorsque  $\mathbb{E}(X^2) < \infty$  (ce qui implique l'existence de  $\mu = \mathbb{E}(X)$ , voir la Proposition 1.2.2), on note

$$\sigma^2 = \sigma_X^2 = \text{Var}(X) = \mathbb{E}((X - \mu)^2), \quad (1.2.22)$$

le moment d'ordre 2 centré, appelé communément *variance*, de  $X$ . L'*écart-type* de  $X$  est, par définition, la racine carrée  $\sigma = \sigma_X = \sqrt{\text{Var}(X)} \geq 0$  de la variance  $\sigma^2$  de  $X$ . Plus généralement, pour  $k \in \mathbb{N}$ , lorsque  $\mathbb{E}(|X|^k) < \infty$ , on définit le *moment d'ordre  $k$*  de  $X$  par

$$\mu_k = \mu_{X,k} = \mathbb{E}(X^k) = \int_{-\infty}^{\infty} x^k d\mathbb{F}(x), \quad k = 1, 2, \dots, \quad \text{et} \quad \mu_0 = \mu_{X,0} = 0, \quad (1.2.23)$$

en faisant usage de la convention  $0^0 = 1$ . Dans le cas particulier où  $\mathbb{P}(X \geq 0) = 1$ , on peut étendre cette définition au moment d'ordre  $s \in \mathbb{R}^+$  de  $X$ , pour  $s$  non nécessairement entier, en posant

$$\mu_s = \mu_{X,s} = \int_0^{\infty} x^s d\mathbb{F}(x),$$

expression qui a toujours un sens, soit comme un nombre réel lorsque  $\mathbb{E}(X^s) = \mathbb{E}(|X|^s) < \infty$ , soit comme  $\infty$ , lorsque  $\mathbb{E}(X^s) = \mathbb{E}(|X|^s) = \infty$ . Dans tous les autres cas, correspondant à  $\mathbb{P}(X \geq 0) < 1$ , le moment d'ordre  $s$  de  $X$  ne peut avoir de sens que pour  $s \in \mathbb{N}$ , et ceci, sous réserve que  $\mathbb{E}(|X|^s) < \infty$ .

On adoptera la notation particulière

$$\nu_s = \nu_{X,s}(X) = \mathbb{E}(|X|^s),$$

pour désigner le *moment absolu d'ordre*  $s \in \mathbb{R}^+$  de  $|X|$ . Ce moment a toujours un sens comme nombre, fini ou non, dans  $\mathbb{R}^+ \cup \{\infty\}$ . Toutefois, comme mentionné plus haut, il est d'usage de dire que  $X$  a des moments d'ordre  $s$ , ou encore que le *moment absolu d'ordre*  $s$  de  $X$  existe, dans le seul cas où la valeur de  $\mathbb{E}(|X|^s)$  est finie. Lorsque  $k \in \mathbb{N}$  est entier, on parlera de même de l'existence de moments d'ordre  $k$  pour  $X$  pour exprimer l'existence de  $\mathbb{E}(|X|^k) \in \mathbb{R}$ .

Les inégalités suivantes sont toujours satisfaites (que les valeurs de  $\mathbb{E}(|X|^r)$  et  $\mathbb{E}(|X|^s)$  soient finies ou non).

$$\forall r, s \in \mathbb{R} : \quad 1 \leq r \leq s \quad \implies \quad \mathbb{E}(|X|^r)^{1/r} \leq \mathbb{E}(|X|^s)^{1/s}. \quad (1.2.24)$$

Les inégalités (1.2.24) se déduisent, de façon générale de l'inégalité de convexité

$$\Psi(\mathbb{E}(X)) \leq \mathbb{E}(\Psi(X)), \quad (1.2.25)$$

vérifiée pour toute fonction convexe  $\Psi$ , sous réserve que les membres de l'inégalité aient un sens. On observera au passage que la fonction  $x \rightarrow x^r$  n'est convexe sur  $\mathbb{R}^+$  que si  $r \geq 1$ .

En appliquant la formule (1.2.24) pour  $r = i \in \mathbb{N}$  et  $s = k \in \mathbb{N}$ , on obtient la proposition utile suivante.

**Proposition 1.2.2.** *lorsque  $k \in \mathbb{N}$  est entier, l'existence de moments d'ordre  $k$  pour  $X$  implique l'existence de tous les moments de  $X$  d'ordre  $j$  entier, positif ou nul et inférieur ou égal à  $k$ , c'est à dire, tel que  $0 \leq j \leq k$ .*

Soit  $k \in \mathbb{N}^* = \{1, 2, \dots\}$  un entier positif. L'existence de  $\mathbb{E}(|X|^k)$  implique (voir la Proposition 1.2.3) l'existence de  $\mathbb{E}(X^\ell)$  pour  $\ell = 1, \dots, k$ . Par conséquent, lorsque  $\mathbb{E}(|X|^k) < \infty$ , en faisant usage des notations (1.2.23), et en posant  $\mu_k = \mu_{X,k} = \mathbb{E}(X^k)$  pour  $k \geq 1$ ,  $\mu = \mu_1 = \mu_{X,1}$ , et  $\mu_0 = \mu_{X,0} = 1$ , on peut définir le *moment centré d'ordre*  $k$  de  $X$  par

$$M_k = M_{X,k} = \mu_{X-\mu,k} = \mathbb{E}((X - \mu)^k) = \sum_{i=0}^k \binom{k}{i} (-1)^i \mu_{k-i} \mu^i, \quad M_1 = 0, \quad M_0 = 1. \quad (1.2.26)$$

Dans (1.2.26), nous nous sommes servis de la formule du binôme pour exprimer le moment centré d'ordre  $k$ ,  $M_k$  en fonction des moments non centrés d'ordre inférieur ou égal à  $k$ ,  $\{\mu_j : 0 \leq j \leq k\}$ . En développant de même la relation  $\mu_k = \mathbb{E}((X - \mu) + \mu)^k$  par la formule du binôme, on obtient l'expression réciproque de  $\mu_k$  en fonction de  $\mu$  et  $\{M_j : 0 \leq j \leq k\}$  (rappelons que, selon (1.2.26),  $M_1 = 0$  et  $M_0 = 1$ ), donnée par

$$\mu_k = \sum_{i=0}^k \binom{k}{i} M_{k-i} \mu^i, \quad k = 0, 1, \dots. \quad (1.2.27)$$

Lorsque  $0 < \sigma < \infty$  (rappelons la notation (1.2.22) exprimant que  $\sigma = \sqrt{\text{var}(X)}$ ), on définit le *moment centré réduit* d'ordre  $j$  de  $X$  par

$$\lambda_j = \lambda_{j,X} = \frac{M_j}{\sigma^j} \quad \text{pour } j = 1, 2, \dots, \quad \text{et } \lambda_0 = 1.$$

On notera au passage que le fait que  $\sigma = 0$  est équivalent au fait que la v.a.  $X$  est *dégénérée*, c'est à dire, presque sûrement constante. On a, en effet,

$$\sigma = 0 \quad \Leftrightarrow \quad \mathbb{P}(X = \mu) = 1 \quad \text{avec } \mu = \mathbb{E}(X).$$

On dira d'une v.a. qu'elle est *non dégénérée*, si elle n'est pas presque sûrement constante. Ceci est équivalent à la propriété

$$\forall x \in \mathbb{R}, \quad \mathbb{P}(X = x) < 1.$$

**Proposition 1.2.3.** *Pour tout  $k \in \mathbb{N}$  tel que les moments  $\mu_0, \mu_1, \dots, \mu_{2k}$  de  $X$  existent, on a*

$$\det \begin{bmatrix} \mu_0 & \mu_1 & \cdots & \mu_k \\ \mu_1 & \mu_2 & \cdots & \mu_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_k & \mu_{k+1} & \cdots & \mu_{2k} \end{bmatrix} \geq 0, \quad (1.2.28)$$

*l'inégalité étant stricte sauf, lorsque, pour  $k \geq 1$ , il existe un ensemble fini  $\mathcal{F}_k = \{x_1, \dots, x_k\} \subset \mathbb{R}$ , formé de  $k$  constantes, distinctes ou confondues, tel que  $\mathbb{P}(Z \in \mathcal{F}_k) = 1$ .*

**Preuve.** Soit

$$M_k = \begin{bmatrix} \mu_0 & \mu_1 & \cdots & \mu_k \\ \mu_1 & \mu_2 & \cdots & \mu_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_k & \mu_{k+1} & \cdots & \mu_{2k} \end{bmatrix},$$

la *matrice de Hankel* intervenant dans (1.2.28). Pour tout  $u = [u_0 \ \cdots \ u_k]' \in \mathbb{R}^{k+1}$ , on a

$$\begin{aligned} u' M_k u &= [u_0 \ \cdots \ u_k] \begin{bmatrix} \mu_0 & \cdots & \mu_k \\ \vdots & \ddots & \vdots \\ \mu_k & \cdots & \mu_{2k} \end{bmatrix} \begin{bmatrix} u_0 \\ \vdots \\ u_k \end{bmatrix} \\ &= \sum_{i=0}^k \sum_{j=0}^k u_i u_j \mu_{i+j} = \mathbb{E} \left( \left\{ \sum_{i=0}^k u_i X^i \right\}^2 \right) \geq 0, \end{aligned} \quad (1.2.29)$$

ce qui montre que  $M_k$  est une *matrice positive* (i.e., la matrice symétrique d'une forme quadratique positive), et vérifie donc  $\det(M_k) \geq 0$  (car ses valeurs propres sont positives ou nulles). Plus précisément, on constate, d'une part, que  $\det M_k > 0$  si  $M_k$  est définie positive, et, d'autre part, que  $\det M_k = 0$  si  $M_k$  est positive sans être définie. Or, par (1.2.29), ce dernier cas ne peut se produire que s'il existe des valeurs non toutes nulles de  $u_0, \dots, u_k$  telles que

$$\mathbb{E} \left( \left\{ \sum_{i=0}^k u_i X^i \right\}^2 \right) = 0 \iff \mathbb{P} \left( \sum_{i=0}^k u_i X^i = 0 \right) = 1.$$

Cette dernière égalité implique qu'avec probabilité 1,  $X$  appartient à l'ensemble, composé au plus de  $k$  éléments, des racines du polynôme  $u_0 + u_1 x + \dots + u_k x^k$ .  $\square$

### 1.2.7 Fonctions caractéristiques et génératrices.

La *fonction caractéristique* du vecteur aléatoire  $Z = [Z_1 \ \dots \ Z_d]' \in \mathbb{R}^d$  (ou de la loi  $\mathcal{L}(Z)$  de ce vecteur aléatoire) est la fonction de  $\mathbf{u} = [u_1 \ \dots \ u_d]' \in \mathbb{R}^d$  définie par

$$\phi_Z(\mathbf{u}) = \phi_Z(u_1, \dots, u_d) = \mathbb{E} \left( \exp \{ \mathbf{i} \mathbf{u}' Z \} \right) = \mathbb{E} \left( \exp \left\{ \mathbf{i} \sum_{j=1}^d u_j Z_j \right\} \right). \quad (1.2.30)$$

Ici, et dans la suite, on notera  $\mathbf{i} = \sqrt{-1} \in \mathbb{C}$  le nombre complexe racine de  $-1$ , pour le distinguer de  $i$ , utilisé comme indice. L'utilisation des fonctions caractéristiques pour étudier les lois de probabilités est particulièrement commode, grâce aux propriétés (1–2–3–4) suivantes. Pour leur démonstration, on consultera l'ouvrage suivant.

Lukacs, E. (1970). *Characteristic Functions*. 2nd. Ed. Griffin, Londres).

Dans la suite, on prendra garde à ne pas confondre les propriétés (1.2.31) et (1.2.32) ci-dessous.

(C.1) La fonction  $\phi_Z(\mathbf{u})$ ,  $\mathbf{u} \in \mathbb{R}^d$ , détermine complètement la loi  $\mathcal{L}(Z)$  de  $Z \in \mathbb{R}^d$ ;

(C.2) Si  $Z_1 \in \mathbb{R}^d$  et  $Z_2 \in \mathbb{R}^d$  sont indépendants, alors,  $\forall \mathbf{u} \in \mathbb{R}^d$ ,



$$\phi_{Z_1+Z_2}(\mathbf{u}) = \mathbb{E}(\exp\{i\mathbf{u}'(Z_1 + Z_2)\}) = \phi_{Z_1}(\mathbf{u})\phi_{Z_2}(\mathbf{u}); \quad (1.2.31)$$

(C.3) Pour que  $Z_1 \in \mathbb{R}^p$  et  $Z_2 \in \mathbb{R}^q$  soient indépendants, il faut et il suffit que la fonction caractéristique jointe de  $Z_1$  et  $Z_2$  vérifie,  $\forall \mathbf{u}_1 \in \mathbb{R}^p$  et  $\forall \mathbf{u}_2 \in \mathbb{R}^q$ ,

$$\phi_{Z_1, Z_2}(\mathbf{u}_1, \mathbf{u}_2) = \mathbb{E}(\exp\{i\mathbf{u}_1'Z_1 + i\mathbf{u}_2'Z_2\}) = \phi_{Z_1}(\mathbf{u}_1)\phi_{Z_2}(\mathbf{u}_2); \quad (1.2.32)$$

(C.4) Pour tout  $N \in \mathbb{N}$ , les conditions suivantes sont équivalentes, pour toute variable aléatoire  $Z \in \mathbb{R}$  (voir Lukacs, E. (1970), op. Cit. Theorem 2.3.3, p.23).

(i) La v.a.  $Z$  possède des moments  $\mu_k = \mathbb{E}(Z^k)$  d'ordre  $k$ , finis pour tout  $k \in \mathbb{N}$ ,  $0 \leq k \leq 2N$ ;

(ii) La fonction caractéristique  $\phi_Z(u)$  de  $Z$  possède un développement limité

d'ordre  $2N$  en 0, soit  $\phi_Z(u) = \sum_{k=1}^{2N} \frac{c_k}{k!} u^k + o(u^{2N})$ , lorsque  $u \rightarrow 0$ ;

(iii) De plus, dans ce dernier cas, on a  $c_k = i^k \mathbb{E}(Z^k)$  pour  $k = 0, \dots, 2N$ .

Ces dernières propriétés sont particulièrement utiles pour le calcul des moments de  $Z$ . Sous réserve de conditions générales de régularité (voir ci-dessus), on a donc

$$\mathbb{E}(Z^k) = i^{-k} \frac{d^k}{du^k} \phi_Z(0). \quad (1.2.33)$$

La *fonction génératrice des moments* de  $Z \in \mathbb{R}^p$  est une version de la fonction caractéristique de  $Z$ , dont la définition est étendue à une partie de  $\mathbb{C}^p$ . Elle est définie par

$$\psi_Z(\mathbf{z}) = \mathbb{E}(\exp\{\mathbf{z}'Z\}), \quad (1.2.34)$$

pour tout  $\mathbf{z} \in \mathbb{C}^p$  tel que le membre de droite de (1.2.34) ait un sens. Le domaine minimal de définition de la fonction génératrice des moments  $\psi_Z(z)$  est  $i\mathbb{R}^p$ . Il est d'ailleurs possible qu'il n'existe pas d'autres valeurs de  $\mathbf{z}$  où  $\psi_Z(\mathbf{z})$  soit défini que celles de la forme  $\mathbf{z} = i\mathbf{u} \in i\mathbb{R}^p$  pour  $\mathbf{u} \in \mathbb{R}^p$ . Pour celles-ci, on a l'identité évidente  $\phi_Z(\mathbf{u}) = \psi_Z(i\mathbf{u})$ , qui lie la fonction génératrice des moments  $\psi_Z$  à la fonction caractéristique  $\phi_Z$  de  $Z$ .

## 1.2.8 Coefficients de Pearson.

Les *coefficients de Pearson*  $\beta_1$  et  $\beta_2$  d'une variable aléatoire réelle  $X$  non dégénérée sont définis, à partir des moments centrés  $M_k = \mathbb{E}((X - \mathbb{E}(X))^k)$ ,  $k = 2, 3, \dots$ , (avec  $\sigma = \sqrt{M_2} > 0$ ), de cette v.a., comme suit (sous réserve d'existence de ces moments). On consultera, à cet effet, la p.85 de l'ouvrage

Kendall, M.G., et Stuart, A. (1969). *The Advanced Theory of Statistics*, Vol.1, Griffin, Londres).

On pose, sous réserve que  $\sigma > 0$  et que les moments correspondants ( $M_3$  ou  $M_4$ ) soient définis,

$$\sqrt{\beta_1} = \frac{M_3}{\sigma^3} = \lambda_3, \quad \beta_1 = \frac{M_3^2}{\sigma^6} = \lambda_3^2, \quad \beta_2 = \frac{M_4}{\sigma^4} = \lambda_4, \quad (1.2.35)$$

en notant que cette définition autorise  $\sqrt{\beta_1}$  à avoir un signe quelconque (celui de  $M_3$ ), par dérogation à l'usage établi pour les radicaux, qui leur affecte habituellement une valeur positive ou nulle. Plus généralement, pour  $k = 0, 1, 2, \dots$ , on pose

$$\beta_{2k+1} = \frac{M_3 M_{2k+3}}{\sigma^{2k+6}} = \lambda_3 \lambda_{2k+3}, \quad \beta_{2k} = \frac{M_{2k+2}}{\sigma^{2k+2}} = \lambda_{2k+2}, \quad (1.2.36)$$

et en particulier

$$\beta_3 = \frac{M_3 M_5}{\sigma^8} = \lambda_3 \lambda_5, \quad \beta_4 = \frac{M_6}{\sigma^6} = \lambda_6. \quad (1.2.37)$$

Le coefficient  $\beta_1$  est appelé *coefficient de dissymétrie* (en anglais, skewness). Il est en effet nul pour les v.a. symétriques autour de leurs moyennes, c'est à dire, telles que  $X - \mathbb{E}(X) \stackrel{d}{=} -(X - \mathbb{E}(X))$ . Le coefficient  $\beta_2$

est appelé *coefficient d'aplatissement* (en anglais, kurtosis). On observera ici que ces appellations varient, dans la mesure où certains auteurs appellent coefficient de dissymétrie,  $\sqrt{\beta_1}$  (avec les conventions ci-dessus), et coefficient d'aplatissement,  $\gamma = \beta_2 - 3$ . Il convient donc de préciser, dans chaque application, la définition utilisée. Rappelons qu'une v.a.r.  $X$  est dite non dégénérée si elle n'est pas constante. Si  $X$  possède des moments d'ordre 2, cette propriété est équivalente au fait que  $\text{Var}(X) > 0$ .

**Proposition 1.2.4.** *Pour toute v.a.r. non dégénérée  $X$  ayant des moments d'ordre 4, les coefficients de Pearson  $\beta_1$  et  $\beta_2$  sont bien définis, et vérifient l'inégalité*

$$\beta_2 - \beta_1 - 1 \geq 0, \quad (1.2.38)$$

cette inégalité étant stricte, sauf s'il existe deux constantes  $x_1, x_2$  telles que

$$\mathbb{P}(X \in \{x_1, x_2\}) = 1.$$

**Preuve.** L'existence de  $\beta_1$  et  $\beta_2$  n'a de sens que si  $\sigma^2 = M_2 > 0$ , ce qui exclut les distributions dégénérées (i.e. concentrées en un seul point). Sous l'hypothèse que  $\sigma^2 = M_2 > 0$ , on applique la Proposition 1.2.3 aux moments centrés  $M_0 = 1, M_1 = 0, M_2, M_3, M_4$ , pour obtenir que

$$\det \begin{bmatrix} 1 & 0 & M_2 \\ 0 & M_2 & M_3 \\ M_2 & M_3 & M_4 \end{bmatrix} = M_2^3 (\beta_2 - \beta_1 - 1) \geq 0,$$

l'inégalité étant stricte sauf lorsque le support de la loi de  $X$  est réduit à deux points, au plus. On en déduit directement (1.2.38).  $\square$

### 1.2.9 Système de Pearson

Le statisticien britannique Karl Pearson (1857-1936) a introduit aux alentours de 1900 un système de lois de probabilités, connu sous le nom de *système de Pearson*, et particulièrement commode dans certaines applications. On consultera à ce sujet :

Elderton, W.P. et Johnson, N.L. (1969). *Systems of Frequency Curves*. Cambridge University Press, Cambridge.

Le système de Pearson comprend des distributions de variables aléatoires réelles  $X$  ayant une densité  $f$  relativement à la mesure de Lebesgue sur  $\mathbb{R}$ , un support défini par un intervalle  $[A, B] \subseteq \mathbb{R}$  avec  $A < B$ , et vérifiant, pour des constantes réelles  $a, b_0, b_1, b_2$  convenables, l'équation différentielle d'ordre 1

$$\frac{df}{dx} = \frac{(x-a)f}{b_0 + b_1x + b_2x^2}. \quad (1.2.39)$$

Sous réserve que le support de la loi de  $X$  soit un intervalle  $[A, B]$  (aux extrémités éventuellement infinies), et que la densité  $f(x)$  de  $X$  soit telle que, pour un entier convenable  $N \geq 0$ , on ait

$$\lim_{x \rightarrow A} x^m f(x) = \lim_{x \rightarrow B} x^m f(x) = 0 \quad \text{pour } m = 0, 1, 2, \dots, N+2,$$

on peut alors intégrer sur  $[A, B]$ , pour  $n = 0, \dots, N$ , les deux membres de l'identité (conséquence de (1.2.39))

$$x^n (b_0 + b_1x + b_2x^2) df(x) = (x-a)f(x) dx. \quad (1.2.40)$$

Ceci permet d'obtenir, après intégration par parties du premier membre de (1.2.40), la relation suivante liant les moments  $\mu_k = \mathbb{E}(X^k)$ , d'ordres  $k = n-1, n, n+1$  de  $X$  (pour  $n = 0$ , la relation reste valable en posant  $0 \times b_0 \mu_{-1} = 0$ ). On obtient ainsi, pour  $n = 0, \dots, N$ ,

$$nb_0 \mu_{n-1} + \{(n+1)b_1 - a\} \mu_n + \{(n+2)b_2 + 1\} \mu_{n+1} = 0. \quad (1.2.41)$$

Il est intéressant d'écrire les équations (1.2.41) dans le cas de distributions *sous forme centrée réduite*, c'est à dire telles que

$$\mu_1 = \mathbb{E}(X) = 0 \quad \text{et} \quad \mu_2 = M_2 = \sigma^2 = \text{Var}(X) = 1.$$

dans ce cas, en posant  $\mu_0 = \mu_2 = 1$  et  $\mu_1 = 0$  dans (1.2.41) pour  $n = 0, 1, 2, 3$ , on aboutit au système

$$\begin{aligned} & \{b_1 - a\} & & = 0 \\ b_0 & & + \{3b_2 + 1\} & = 0 \\ & \{3b_1 - a\} & + \{4b_2 + 1\}M_3 & = 0 \\ 3b_0 & + \{4b_1 - a\}M_3 & + \{5b_2 + 1\}M_4 & = 0 \end{aligned} \tag{1.2.42}$$

On constate, de plus, que

$$\mu_1 = 0 \quad \text{et} \quad \mu_2 = 1 \quad \Leftrightarrow \quad a = b_1 \quad \text{et} \quad b_0 + 3b_2 + 1 = 0. \tag{1.2.43}$$

Comme, lorsque ces conditions sont satisfaites,  $M_3 = \sqrt{\beta_1}$  et  $M_4 = \beta_2$ , il s'ensuit que, d'une manière générale, les coefficients  $a, b_0, b_1, b_2$  de (1.2.39)–(1.2.40) sont parfaitement déterminés par  $\mu_1, \sigma^2, \sqrt{\beta_1}$  et  $\beta_2$ , et réciproquement. Par exemple, pour une loi centrée, avec  $\mu_1 = 0$ , on obtient les formules

$$a = b_1 = -\frac{\sigma\sqrt{\beta_1}(\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)}, \quad b_0 = -\frac{\sigma^2(4\beta_2 - 3\beta_1)}{2(5\beta_2 - 6\beta_1 - 9)}, \quad b_2 = -\frac{(2\beta_2 - 3\beta_1 - 6)}{2(5\beta_2 - 6\beta_1 - 9)}, \tag{1.2.44}$$

qui permettent de déterminer  $f$ , et donc la loi de  $X$ , à partir de  $\mu, \sigma^2$  et des coefficients de Pearson  $\sqrt{\beta_1}$  et  $\beta_2$ . L'application statistique de cette méthode consiste à évaluer une forme approximative de cette loi à partir des estimations fournies par la *méthode des moments*. Ayant observé un échantillon  $X_1, \dots, X_n$  de taille  $n \geq 4$ , et composé d'observations non toutes identiques (pour éviter tout problème de dégénérescence), on évalue successivement les *moments empiriques* et les *coefficients de Pearson empiriques*, par

$$\begin{aligned} \hat{\mu}_1 &= \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, & \hat{\sigma} &= \left\{ \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right\}^{1/2}, & \hat{M}_k &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k, \\ \sqrt{\hat{\beta}_1} &= \frac{\hat{M}_3}{\hat{\sigma}^3}, & \hat{\beta}_1 &= \frac{\hat{M}_3^2}{\hat{\sigma}^6}, & \hat{\beta}_2 &= \frac{\hat{M}_4}{\hat{\sigma}^4}. \end{aligned}$$

Ceci permet d'obtenir, après remplacement de  $\sqrt{\beta_1}$  et  $\beta_2$  par  $\sqrt{\hat{\beta}_1}$  et  $\hat{\beta}_2$  dans (1.2.44), des valeurs estimées de  $a, b_0, b_1, b_2$  permettant, après intégration de la version de l'équation (1.2.39) correspondante, une estimation de la loi de  $(X - \mu_1)/\sigma$ . Cette méthode est d'une application délicate à cause, notamment, de l'instabilité des estimations de  $\sqrt{\beta_1}$  et  $\beta_2$  en présence de *valeurs aberrantes* [on désigne ainsi des observations parmi  $X_1, \dots, X_n$  qui n'appartiennent pas à la loi de l'échantillon]. De telles valeurs aberrantes sont fréquemment rencontrées dans la pratique statistique, et notamment lorsque se produisent des erreurs dans le recueil des données. Dans le cas présent, une seule valeur aberrante située au delà des valeurs extrêmes engendrées par les autres observations, peut suffire parfois à fausser profondément les estimations  $\hat{M}_3$  et  $\hat{M}_4$  des moments supérieurs  $M_3$  et  $M_4$ . Les estimations correspondantes perdent alors toute utilité statistique. Cependant, et malgré ses défauts, la *méthode de Pearson* est à recommander systématiquement dans le cadre d'une analyse exploratoire des données, ayant pour but, entre autres, de se faire une idée des modèles probabilistes devant être utilisée dans la description des observations.

Il est parfois intéressant de mesurer la *dissymétrie* d'une loi de probabilité par d'autres indicateurs que  $\beta_1$  ou  $\sqrt{\beta_1}$ . Pour une loi de densité  $f$  sur  $\mathbb{R}$ , on définit le *mode* de  $f$  (lorsque cette définition a un sens) par la relation

$$\text{mode} = \text{argmax}(f).$$

D'une manière générale, on note  $\text{argmax}(f)$  (resp.  $\text{argmin}(f)$ ) toute valeur de  $x$  (si elle existe) telle que  $f(x) = \sup_{t \in \mathbb{R}} f(t)$  (resp.  $f(x) = \inf_{t \in \mathbb{R}} f(t)$ ).

Une autre définition du *coefficient de dissymétrie* (en anglais *skewness*) de Pearson (à ne pas confondre avec  $\sqrt{\beta_1}$  ou  $\beta_1$ , qui relèvent aussi de cette appellation) est donnée par le rapport

$$[Sk]_1 = \frac{\text{moyenne} - \text{mode}}{\sigma}.$$

Un autre indicateur de dissymétrie (voir p.85 dans : Kendall, M. G. et Stuart, A. (1969). *The Advanced Theory of Statistics*. Vol.1) est donné par l'expression

$$[Sk]_2 = \frac{\text{moyenne} - \text{médiane}}{\sigma}.$$

Il est intéressant de constater que, pour les lois du système de Pearson, on a l'identité

$$[Sk]_1 = \frac{\sqrt{\beta_1} (\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)}.$$

On utilise tantôt  $[Sk]_1$  ou  $[Sk]_2$ , tantôt  $\sqrt{\beta_1}$  ou  $\beta_1$  sous l'appellation ambiguë de *coefficient de dissymétrie*, et il convient de ne pas confondre ces différents coefficients, en précisant le vocabulaire utilisé dans chaque cas.

**Exercice 1.2.5.** Une variable aléatoire  $X \in \mathbb{N}$  à valeurs entières, positives ou nulles, suit une loi de Poisson  $\text{Po}(\lambda)$  de paramètre  $\lambda \geq 0$  (ce qui est noté  $X \stackrel{d}{=} \text{Po}(\lambda)$ ), si (avec la convention  $0^0 = 1$ ),

$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{pour } k = 0, 1, \dots \quad (1.2.45)$$

1°) Déterminer l'expression, en fonction de  $u \in \mathbb{R}$  et  $\lambda$ , de la fonction caractéristique  $\phi_X(u) = \mathbb{E}(\exp(iuX))$  de  $X$ . En déduire que les premiers moments  $\mu_k = \mathbb{E}(X^k)$ , et les moments centrés  $M_k = \mathbb{E}((X - \mathbb{E}(X))^k)$  d'ordre  $k$  de  $X$  sont donnés par

$$\mu_0 = 1, \quad \mu_1 = \lambda, \quad M_2 = \lambda, \quad M_3 = \lambda, \quad M_4 = 3\lambda^2 + \lambda. \quad (1.2.46)$$

2°) En déduire l'expression des coefficients de Pearson  $\beta_1$  et  $\beta_2$  en fonction de  $\lambda$ . Montrer que, lorsque  $\lambda > 0$  varie dans  $\mathbb{R}^{+*}$ , le couple  $(\beta_1, \beta_2)$  parcourt la demi-droite ouverte d'équation

$$\beta_2 - \beta_1 - 3 = 0, \quad \beta_1 > 0. \quad (1.2.47)$$

On trouvera la solution de cet exercice, ainsi que davantage de détails sur la loi de Poisson dans la référence (voir p.157)

Johnson, N. L., Kotz, S. et Kemp, A. W. (1992). *Univariate Discrete Distributions*. John Wiley & Sons, New York.

**Exercice 1.2.6.** Une variable aléatoire réelle  $X$  est dite suivre une loi gamma standard  $\Gamma(r)$  (ou  $\Gamma(r, 1)$ ), ce qui est noté  $X \stackrel{d}{=} \Gamma(r)$  (ou  $X \stackrel{d}{=} \Gamma(r, 1)$ ), si, pour  $r > 0$ , la densité  $f_X(x)$  est donnée par

$$\begin{aligned} f_X(x) &= \frac{x^{r-1} e^{-x}}{\Gamma(r)} \quad \text{pour } x > 0, \\ &= 0 \quad \text{pour } x \leq 0. \end{aligned} \quad (1.2.48)$$

1°) Montrer que, pour tout  $\rho > 0$ , le moment d'ordre  $\rho$  de  $X$  est défini par

$$\mu_\rho = \mathbb{E}(X^\rho) = \frac{\Gamma(r + \rho)}{\Gamma(r)}. \quad (1.2.49)$$

En déduire l'expression de  $\mu_1, \mu_2, \mu_3, \mu_4$  en fonction de  $r > 0$ .

2°) Montrer que les coefficients de Pearson  $\sqrt{\beta_1}$  et  $\beta_2$  de  $X \stackrel{d}{=} \Gamma(r)$  sont donnés par

$$\sqrt{\beta_1} = \frac{2}{\sqrt{r}} \quad \text{et} \quad \beta_2 = 3 + \frac{6}{r}. \quad (1.2.50)$$

En déduire que, lorsque  $r > 0$  varie dans  $\mathbb{R}^{+*}$ , le couple  $(\beta_1, \beta_2)$  parcourt la demi-droite ouverte d'équation

$$\beta_2 - \frac{3}{2}\beta_1 - 3 = 0, \quad \beta_1 > 0. \quad (1.2.51)$$

3°) Une variable aléatoire réelle  $Y$  est dite suivre une loi exponentielle  $E(\lambda)$ , ce qui est noté  $X \stackrel{d}{=} E(\lambda)$ , si, pour  $\lambda > 0$ , la densité  $f_Y(y)$  est donnée par

$$\begin{aligned} f_Y(y) &= e^{-y} \quad \text{pour } y > 0, \\ &= 0 \quad \text{pour } y \leq 0. \end{aligned} \quad (1.2.52)$$

Montrer que les coefficients de Pearson d'une loi exponentielle sont donnés par

$$\sqrt{\beta_1} = 2, \quad \beta_1 = 4, \quad \beta_2 = 9. \quad (1.2.53)$$

On trouvera la solution de cet exercice, ainsi que davantage de détails sur les lois gamma et exponentielles dans la référence (voir p.338)

Johnson, N. L., Kotz, S. et Balakrishnan, N. (1994). *Continuous Univariate Distributions*. 2nd. Ed., John Wiley & Sons, New York.

## 1.3 Mesure empirique et échantillonnage.

### 1.3.1 Echantillons aléatoires.

Une suite  $Z_1, \dots, Z_n$  de répliques indépendantes de même loi qu'un vecteur aléatoire  $Z \in \mathbb{R}^p$  est appelé *échantillon (aléatoire)* de taille  $n$  de  $Z$  (ou de la loi  $\mathcal{L}(Z)$ ). On lui associe la *mesure empirique* (ou *loi de probabilité empirique*) définie par

$$\mathbb{P}_n(A) = n^{-1} \#\{Z_i \in A : 1 \leq i \leq n\} \quad \text{pour } A \in \mathcal{B}_p. \quad (1.3.1)$$

Ici,  $\#E$  désigne la *cardinalité* de  $E$  (le nombre d'éléments de  $E$ , lorsque  $E$  est fini). La mesure empirique affecte une masse  $1/n$  à chacune des  $n$  observations  $Z_1, \dots, Z_n$ .

Conditionnellement à l'échantillon  $Z_1, \dots, Z_n$ , la mesure empirique définit une loi de probabilité, notée  $\mathcal{L}_n(Z)$ , sur  $\mathbb{R}^d$ . On peut donc lui associer tous les éléments caractéristiques correspondants (moments, fonction de répartition, fonction de quantiles, ...), qui sont alors qualifiés d'*empiriques*.

Le même principe s'applique aux *paramètres* d'une distribution. Un paramètre  $\theta \in \mathbb{R}^p$  est une fonctionnelle  $\theta = \Theta(\mathcal{L}(Z))$  de  $\mathcal{L}(Z)$ , qui détermine, partiellement ou complètement,  $\mathcal{L}(Z)$ . L'*estimateur empirique*  $\hat{\theta}_n$  de  $\theta$  (lorsqu'il existe) est défini par le remplacement formel de  $\mathcal{L}(Z)$  par  $\mathcal{L}_n(Z)$ , de sorte que

$$\hat{\theta}_n = \Theta(\mathcal{L}_n(Z)).$$

La statistique repose sur le principe général que, sous certaines réserves et conditions, l'estimateur  $\hat{\theta}_n$  de  $\theta$  converge vers  $\theta$  lorsque  $n \rightarrow \infty$ . Quand on parle, ici, de *convergence*, il faut entendre, selon l'un des modes de convergence que définit le calcul des probabilités. Le plus utile d'entre eux, en statistique, est celui de la *convergence en probabilité*. Celle-ci se produit si et seulement si, pour tout  $\varepsilon > 0$ , on a

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\hat{\theta}_n - \theta| > \varepsilon) = 0,$$

où  $|\cdot|$  désigne la norme (euclidienne) usuelle sur  $\mathbb{R}^p$ . L'estimateur empirique  $\widehat{\theta}_n$  de  $\theta$  (lorsqu'il existe) n'est pas toujours le plus intéressant du point de vue statistique. Il a, par contre, l'intérêt d'être généralement facile à construire, et possédant une vitesse de convergence bien souvent optimale, à une constante multiplicative près.

Par exemple, lorsque  $X_1, \dots, X_n$  est un échantillon aléatoire de taille  $n$  d'une variable aléatoire réelle  $X$ , on définit sa *fonction de répartition empirique* par

$$\mathbb{F}_n(x) := \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{X_i \leq x\}} = n^{-1} \#\{X_i \leq x : 1 \leq i \leq n\} \quad \text{pour } x \in \mathbb{R}, \quad (1.3.2)$$

où  $\mathbb{I}_A$  désigne la fonction indicatrice de  $A$  (égale à 1 si  $A$  a lieu, et à 0, autrement). Les *moments empiriques* de cet échantillon sont alors définis, pour  $k = 0, 1, 2, \dots$ , par

$$\widehat{\mu}_k = \frac{1}{n} \sum_{i=1}^n X_i^k = \int_{-\infty}^{\infty} x^k d\mathbb{F}_n(x). \quad (1.3.3)$$

Le plus important parmi ces moments empiriques est la *moyenne empirique*, définie par

$$\bar{X} = \widehat{\mu} = \widehat{\mu}_1 = \frac{1}{n} \sum_{i=1}^n X_i. \quad (1.3.4)$$

Les *moments empiriques centrés* sont définis, de même, par

$$\widehat{M}_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k. \quad (1.3.5)$$

Le plus important d'entre eux est la *variance empirique*, notée

$$\widehat{\sigma}^2 = \widehat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \widehat{\mu}_2 - \widehat{\mu}_1^2. \quad (1.3.6)$$

Comme les moments empiriques sont les moments d'une loi de probabilité, ils satisfont les propriétés générales correspondantes. Par exemple, si on note les *moments de Pearson empiriques* par

$$\widehat{\beta}_1 = \frac{\widehat{M}_3^2}{\widehat{\sigma}^6} \quad \text{et} \quad \widehat{\beta}_2 = \frac{\widehat{M}_4}{\widehat{\sigma}^4}, \quad (1.3.7)$$

on a (de même qu'en (1.2.38))

$$\widehat{\beta}_2 - \widehat{\beta}_1 - 1 \geq 0. \quad (1.3.8)$$

On notera, toutefois, que (1.3.7) n'a de sens que pour  $n \geq 2$ , et sous réserve de l'existence de deux observations  $X_i \neq X_j$  distinctes parmi  $X_1, \dots, X_n$  (autrement, on aurait  $\widehat{\sigma} = 0$ ).

On constate facilement que, sous réserve d'existence des moments apparaissant dans les membres de droite ci-dessous,

$$\mathbb{E}(\widehat{\mu}_k) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = \mu_k, \quad (1.3.9)$$

$$\begin{aligned} \mathbb{E}(\widehat{\mu}^2) &= \frac{1}{n^2} \left\{ \sum_{i=1}^n \mathbb{E}(X_i^2) + \sum_{1 \leq i \neq j \leq n} \mathbb{E}(X_i) \mathbb{E}(X_j) \right\} \\ &= \frac{1}{n^2} \left\{ n\mu_2 + (n^2 - n)\mu^2 \right\} = \mu^2 + \frac{\sigma^2}{n}, \end{aligned} \quad (1.3.10)$$

$$\mathbb{E}(\widehat{\sigma}^2) = \mathbb{E}(\widehat{\mu}_2) - \mathbb{E}(\widehat{\mu}^2) = \frac{n-1}{n} \sigma^2. \quad (1.3.11)$$

Cette dernière relation mène à introduire, pour  $n \geq 2$ , un estimateur sans biais de la variance  $\sigma^2$ . Celui-ci est, habituellement, noté  $s^2$  ou  $\tilde{\sigma}^2$  (par opposition à  $\hat{\sigma}^2$ ), et défini par

$$s^2 = \tilde{\sigma}^2 := \frac{n}{n-1} \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2. \quad (1.3.12)$$

Toutefois, alors que  $\hat{\sigma}^2$  est défini pour tout  $n \geq 1$ ,  $s^2$  n'a de sens que pour  $n \geq 2$ . Dans ce dernier cas,  $s^2$  définit un *estimateur sans biais* de  $\sigma^2$ , puisqu'il vérifie  $\mathbb{E}(s^2) = \sigma^2$ .

D'une manière générale, on a les relations suivantes, exprimant les moments croisés des moments empiriques.

$$\begin{aligned} \mathbb{E}(\hat{\mu}_k \hat{\mu}_\ell) &= \frac{1}{n^2} \left\{ \sum_{i=1}^n \mathbb{E}(X_i^{k+\ell}) + \sum_{1 \leq i \neq j \leq n} \mathbb{E}(X_i^k) \mathbb{E}(X_j^\ell) \right\} \\ &= \mu_k \mu_\ell + \frac{1}{n} \left\{ \mu_{k+\ell} - \mu_k \mu_\ell \right\}, \end{aligned} \quad (1.3.13)$$

de sorte que

$$\text{Cov}(\hat{\mu}_k, \hat{\mu}_\ell) = \frac{1}{n} \left\{ \mu_{k+\ell} - \mu_k \mu_\ell \right\}. \quad (1.3.14)$$

Le *théorème central limite* [CLT] permet d'établir que, sous réserve d'existence de moments finis d'ordre  $2d$  de  $X$ , on a la convergence en loi, lorsque  $n \rightarrow \infty$ ,

$$\sqrt{n} \begin{bmatrix} \hat{\mu}_1 - \mu_1 \\ \vdots \\ \hat{\mu}_d - \mu_d \end{bmatrix} \xrightarrow{d} N_d \left( \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} \mu_2 - \mu_1^2 & \cdots & \mu_{d+1} - \mu_1 \mu_d \\ \vdots & \ddots & \vdots \\ \mu_{d+1} - \mu_1 \mu_d & \cdots & \mu_{2d} - \mu_d^2 \end{bmatrix} \right). \quad (1.3.15)$$

On se référera à la définition 3.2.2 (dans la suite de ce texte), pour la description des lois normales multivariées utilisées dans le membre de droite de (1.3.15). Le calcul des moments des moments empiriques centrés d'ordre  $k \geq 2$  est, en général, plus délicat (voir p. 230 dans Kendall, M. G. et Stuart, A. (1969). *The Advanced Theory of Statistics*. Vol. 1, 3<sup>ème</sup> ed.). On obtient que, lorsque  $n \rightarrow \infty$ ,

$$\mathbb{E}(\widehat{M}_k) = M_k + o(n^{-1/2}), \quad (1.3.16)$$

$$\begin{aligned} \text{Cov}(\widehat{M}_k, \widehat{M}_\ell) &= \frac{1}{n} \left\{ M_{k+\ell} - M_k M_\ell + k\ell \sigma^2 M_{k-1} M_{\ell-1} \right. \\ &\quad \left. - k M_{k-1} M_{\ell+1} - \ell M_{\ell-1} M_{k+1} \right\} + o(n^{-1}). \end{aligned} \quad (1.3.17)$$

Il est intéressant d'évaluer la covariance  $\text{Cov}(\hat{\mu}, \widehat{M}_k)$ . On obtient que (voir p.231 et p.243 dans Kendall, M. G. et Stuart, A. (1969)), lorsque  $n \rightarrow \infty$ ,

$$\text{Cov}(\hat{\mu}_k, \widehat{M}_\ell) = \frac{1}{n} \left\{ M_{k+\ell} - M_k M_\ell - \ell M_{k+1} M_{\ell-1} \right\} + o(n^{-1}). \quad (1.3.18)$$

**Exercice 1.3.1.** Soit un échantillon  $X_1, \dots, X_n$  d'observations, d'effectif  $n \geq 2$ , et composé au moins de 2

observations distinctes. On pose (avec  $\hat{\sigma} \geq 0$  et  $s \geq 0$ )

$$\begin{aligned}\bar{X} &= \frac{1}{n} \sum_{i=1}^n nX_i, & \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, & s^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \\ \widehat{M}_k &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k, & \text{pour } k &= 1, 2, \dots, & \text{et } \widehat{M}_0 &= 1, \\ \widehat{\beta}_1 &= \frac{\widehat{M}_3^2}{\widehat{\sigma}^6}, & \widehat{\beta}_2 &= \frac{\widehat{M}_4}{\widehat{\sigma}^4}, & \widetilde{\beta}_1 &= \frac{\widehat{M}_3^2}{s^6}, & \widetilde{\beta}_2 &= \frac{\widehat{M}_4}{s^4}.\end{aligned}$$

1°) Justifier les inégalités  $\widehat{\beta}_2 - \widehat{\beta}_1 - 1 \geq 0$  et  $\widehat{\beta}_1 - 1 \geq 0$ . Est-il possible d'avoir l'égalité  $\widehat{\beta}_2 - \widehat{\beta}_1 - 1 = 0$ ? L'inégalité  $\widetilde{\beta}_2 - \widetilde{\beta}_1 - 1 \geq 0$  est elle nécessairement satisfaite?

2°) On suppose que  $n = 2$ . Déterminer  $\widehat{\beta}_1$  et  $\widehat{\beta}_2$ . En général, pour  $n \geq 2$  fixé, est-ce que  $\widehat{\beta}_1$  et  $\widehat{\beta}_2$  peuvent prendre des valeurs arbitraires dans le plan de Pearson, caractérisé par  $\widehat{\beta}_2 - \widehat{\beta}_1 - 1 \geq 0$  et  $\widehat{\beta}_1 - 1 \geq 0$ ?

### 1.3.2 Quelques théorèmes limites du calcul des probabilités.

La statistique des échantillons aléatoires repose sur un ensemble de théorèmes limites du calcul des probabilités. Nous en rappelons quelques uns parmi les plus utiles.

La loi des grands nombres [LLN] (pour "law of large numbers") est due, sous sa forme presque sûre, à Kolmogorov (1928). On consultera, par exemple, à ce sujet, le livre :

Révész, P. (1968). *The Laws of Large Numbers*. Academic Press, New York.

Sous sa forme de base, la LLN exprime que si  $X = X_1, X_2, \dots$  est une suite de v.a.r. indépendantes de même loi, et si  $S_n = X_1 + \dots + X_n$ , alors, l'existence d'une constante  $c$  telle que, lorsque  $n \rightarrow \infty$ ,

$$n^{-1}S_n \rightarrow c \quad \text{p.s.},$$

est équivalente au fait que  $\mathbb{E}(|X|) < \infty$ , avec  $c = \mu = \mathbb{E}(X)$ .

Rappelons que p.s. est l'abrégié de *presque sûrement* (ou de *presque sûr*), qui exprime ici que la convergence de  $n^{-1}S_n$  vers  $\mu = \mathbb{E}(X)$  a lieu avec probabilité 1. La *convergence presque sûre* implique la *convergence en probabilité*. Cette dernière apparaît, quant à elle, dans la *loi faible des grands nombres* (initialement découverte par Bernoulli en 1713). Celle-ci exprime que, pour tout  $\varepsilon > 0$  fixé, lorsque  $n \rightarrow \infty$ ,

$$\mathbb{P}(|n^{-1}S_n - \mu| > \varepsilon) \rightarrow 0.$$

Comme application statistique de ces résultats, on déduit de la LLN la convergence, p.s. lorsque  $n \rightarrow \infty$ , des moments empiriques vers les moments exacts *sous réserve d'existence de ces derniers*. En particulier, lorsque la taille  $n$  d'un échantillon croît indéfiniment, on a les convergences p.s. des estimateurs définis en (1.3.4)–(1.3.6)–(1.3.12), soit

$$\widehat{\mu} \rightarrow \mu, \quad \widehat{\sigma}^2 \rightarrow \sigma^2 \quad \text{et} \quad s^2 \rightarrow \sigma^2.$$

La vitesse de convergence dans la LLN est fournie par le *théorème central limite* [CLT] (pour "central limit theorem"). Ce dernier exprime, dans sa version de base, que, sous réserve d'existence (et de non-dégénérescence) de  $0 < \sigma^2 = \text{Var}(X) < \infty$ , on a, pour tout  $x \in \mathbb{R}$ , lorsque  $n \rightarrow \infty$ , la convergence

$$\mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

Cette propriété est équivalente à la convergence en loi

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0, 1),$$



où  $N(0, 1)$  désigne la *loi normale standard* (voir le §3.1 pour les définitions et notations appropriées).

La convergence asymptotique des sommes partielles  $S_n$ , dans une échelle convenable, vers la loi normale, est caractérisée par le théorème de Lindeberg-Feller. On consultera, par exemple, pour un exposé détaillé de cette théorie, le livre :

Chow, Y. S. et Teicher, H. (1988). *Probability Theory*. Springer, New York.

Des théorèmes comme le CLT peuvent être utilisés sans modification en remplaçant des paramètres inconnus par des estimateurs convergents (en probabilité) de ces derniers. Pour cela, le théorème (ou lemme) de Slutsky (voir Chow et Teicher, op. cit., p.254) s'avère être particulièrement commode. Sous une forme générale, ce résultat exprime que si  $\mathbf{X}_n$  et  $\mathbf{Y}_n$  sont deux suites de v.a. (définies sur le même espace de probabilités), telles que, pour une loi de probabilité  $\mathcal{L}$  donnée,

$$\mathbf{X}_n \xrightarrow{d} \mathcal{L} \quad \text{et} \quad \mathbf{Y}_n \xrightarrow{\mathbb{P}} 1,$$

où " $\xrightarrow{\mathbb{P}}$ " désigne la convergence en probabilité, alors, lorsque  $n \rightarrow \infty$ ,

$$\mathbf{X}_n \mathbf{Y}_n \xrightarrow{d} \mathcal{L}.$$

Comme application directe de tout ceci, on a, lorsque  $n \rightarrow \infty$ , les convergences en loi

$$\frac{S_n - n\mu}{\hat{\sigma}\sqrt{n}} \xrightarrow{d} N(0, 1) \quad \text{et} \quad \frac{S_n - n\mu}{s\sqrt{n}} \xrightarrow{d} N(0, 1).$$

De telles formules permettent d'obtenir des *intervalles de confiance asymptotiques* pour des paramètres de distribution inconnus (ici  $\mu$ ). Un exemple d'une telle application est fourni au §1.3.3.

### 1.3.3 Un exemple commenté.

Les calculs statistiques nécessitent l'emploi de logiciels spécialisés. Il convient, cependant, de faire attention aux définitions adoptées par ces logiciels, qui ne sont pas toujours identiques. De plus, les calculs qui y sont réalisés sont, le plus souvent, arrondis, et il faut en tenir compte. L'exemple suivant illustre ces problèmes. On considère l'échantillon ci-dessous, noté  $X_1, \dots, X_n$ , et composé de  $n = 19$  observations, numérotées de 1 à 19.

1	0.152	2	0.422
3	0.368	4	0.109
5	0.922	6	0.331
7	0.948	8	0.526
9	5.191	10	2.830
11	1.973	12	0.687
13	1.308	14	0.383
15	0.084	16	0.377
17	0.425	18	1.086
19	0.703		

(1.3.19)

On obtient les statistiques

$$\sum_{i=1}^n X_i = 20.827, \quad \sum_{i=1}^n X_i^2 = 52.414\,269, \quad \sum_{i=1}^n X_i^3 = 195.626\,395\,117, \quad \sum_{i=1}^n X_i^4 = 863.843\,487\,407\,673$$

Pour cet échantillon, la moyenne empirique  $\hat{\mu}$  et la variance empirique  $\hat{\sigma}^2$ , sont données, respectivement, par

$$\begin{aligned} \hat{\mu} &= 1,096157895 \quad \text{à } 10^{-9} \text{ près;} \\ \hat{\sigma}^2 &= 1.557\,083\,607 \quad \text{à } 10^{-9} \text{ près;} \\ \widehat{M}_3 &= 3.858\,595\,779 \quad \text{à } 10^{-9} \text{ près;} \\ \widehat{M}_4 &= 15.877\,578\,077 \quad \text{à } 10^{-9} \text{ près.} \end{aligned}$$

Il est intéressant de comparer ces valeurs précises à celles que fournissent les logiciels usuels de statistique. Nous traitons, tout d'abord, ces observations par *Statistica 6.1*. Comme tous les logiciels, les résultats y sont donnés avec une approximation, ici, à la 6<sup>ème</sup> décimale. Nous obtenons, par *Statistica 6.1*, les valeurs numériques (arrondies à la sixième décimale)

$$\begin{aligned} \text{Moyenne estimée} &= 1.096\,158 \quad \text{à } 10^{-6} \text{ près;} \\ \text{Variance estimée} &= 1.643\,588 \quad \text{à } 10^{-6} \text{ près;} \\ \text{Asymétrie estimée} &= 2.160\,357 \quad \text{à } 10^{-6} \text{ près;} \\ \text{Aplatissement estimé} &= 5.093\,973 \quad \text{à } 10^{-6} \text{ près;.} \end{aligned}$$

L'explication de la différence entre l'estimation de  $\sigma^2$ , égale à 1.557 084, obtenue plus haut, l'estimation  $\hat{\sigma}^2 \simeq 1.643\,588$  ci-dessus, vient du fait que le logiciel calcule l'estimation centrée de la variance, donnée par

$$s^2 = \tilde{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n X_i^2 = \frac{n}{n-1} \hat{\sigma}^2 = 1.643\,588\,251 \quad \text{à } 10^{-9} \text{ près.}$$

Il est très important, dans la pratique, de lire attentivement les notices de logiciels pour savoir l'estimateur de la variance qu'ils utilisent (en général  $s^2 = \tilde{\sigma}^2$ ). Ce point est encore davantage mis en évidence par les estimations des coefficients de Pearson. Un calcul des moments empiriques fournit, dans cet exemple,

$$\begin{aligned} \sqrt{\hat{\beta}_1} &= \frac{\widehat{M}_3}{\hat{\sigma}^3} = 1.985\,918\,339 \quad \text{à } 10^{-9} \text{ près,} \\ \hat{\beta}_2 &= \frac{\widehat{M}_4}{\hat{\sigma}^4} = 6.548\,779\,869 \quad \text{à } 10^{-9} \text{ près.} \end{aligned}$$

On notera que les estimations trouvées ci-dessus sont, encore une fois, sensiblement différentes de celles utilisées par *Statistica 6.1*. En effet, ce dernier logiciel calcule les statistiques

$$\begin{aligned} \text{Asymétrie} &= \frac{n^2}{(n-1)(n-2)} \frac{\widehat{M}_3}{\tilde{\sigma}^3} = \frac{n^2(n-1)\sqrt{n-1}}{(n-1)(n-2)n\sqrt{n}} \frac{\widehat{M}_3}{\hat{\sigma}^3} = \frac{\sqrt{n(n-1)}}{n-2} \sqrt{\hat{\beta}_1} \\ &\simeq \frac{\sqrt{18 \times 19}}{17} \times 1.985\,918\,339 = 2.160\,356\,968 \quad \text{à } 10^{-9} \text{ près,} \end{aligned}$$

et

$$\begin{aligned} \text{Aplatissement} &= \frac{(n-1)(n+1)\hat{\beta}_2 - 3(n-1)^2}{(n-2)(n-3)} \\ &= \frac{18 \times 20 \times \hat{\beta}_2 - 3 \times 18^2}{17 \times 16} = 5.093\,973\,357 \quad \text{à } 10^{-9} \text{ près.} \end{aligned}$$

En fait, le coefficient d'aplatissement estimé par *Statistica 6.1* est le coefficient  $\gamma = \beta_2 - 3$ . Le coefficient d'aplatissement empirique est donné par

$$\hat{\beta}_2 = 6.548\,779\,869 \quad \text{à } 10^{-9} \text{ près,}$$

ce qui mène à l'estimateur empirique de  $\gamma$ , égal à

$$\hat{\gamma} = \hat{\beta}_2 - 3 = 3.548\,779\,869,$$

qui est très sensiblement différent de l'estimateur 5.093 973 de *Statistica*. Un traitement semblable, effectué par un autre logiciel, ici *StatXact 6*, fournit les valeurs suivantes (arrondies, cette fois-ci, à 3 décimales) :

$$\begin{aligned} \text{Moyenne estimée} &= 1.096, \\ \text{Variance estimée} &= 1.644, \\ \text{Dissymétrie estimée} &= 1.986, \\ \text{Aplatissement estimé} &= 3.549. \end{aligned}$$

On observe ici que, tout comme *Statistica 6.1*, le logiciel *StatXact 6* estime  $\sigma^2$  par  $\tilde{\sigma}^2$ . Par contre, son coefficient de dissymétrie est le coefficient empirique  $\sqrt{\hat{\beta}_1}$ , alors que son coefficient d'aplatissement coïncide avec le coefficient empirique  $\hat{\beta}_2 - 3$ . Ces exemples illustrent bien la nécessité de bien apprécier ce que calcule effectivement le logiciel qui est employé.

Les résultats précédents permettent d'obtenir un *intervalle de confiance asymptotique* pour la valeur inconnue de la moyenne  $\mu$  de la loi commune des observations. On écrit, tout d'abord, en faisant usage du *théorème central limite* (voir le §1.3.2), que, pour tout  $x \in \mathbb{R}$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{\sqrt{n}(\hat{\mu} - \mu)}{\sigma} \leq x\right) = \Phi(x),$$

où  $\Phi(\cdot)$  est la fonction de répartition de la loi normale  $N(0, 1)$  (voir (3.1.2)). On se sert ensuite de la *loi des grands nombres*, pour écrire que (sous réserve de l'existence de  $\sigma^2 = \text{Var}(Z)$ ), que

$$\lim_{n \rightarrow \infty} \frac{\hat{\sigma}^2}{\sigma^2} = 1 \quad \text{en probabilité.}$$

Le *lemme de Slutsky* (voir le §1.3.2) implique alors que

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{\sqrt{n}(\hat{\mu} - \mu)}{\hat{\sigma}} \leq x\right) = \Phi(x).$$

Cette propriété permet d'écrire (par exemple, compte tenu des valeurs numériques des quantiles de la loi normale  $N(0, 1)$ , voir la Table 1, au §3.1.2) la formule *approximative*

$$\mathbb{P}\left(\mu \in \left[\hat{\mu} - 1.960 \times \frac{\hat{\sigma}}{\sqrt{n}}, \hat{\mu} + 1.960 \times \frac{\hat{\sigma}}{\sqrt{n}}\right]\right) \simeq 95\%. \quad (1.3.20)$$

En prenant les valeurs numériques obtenues précédemment, on obtient

$$\mathbb{P}\left(\mu \in \left[1.096 - 1.960 \times \frac{1.248}{\sqrt{19}}, 1.096 + 1.960 \times \frac{1.248}{\sqrt{19}}\right]\right) \simeq 95\%,$$

soit encore

$$\mathbb{P}\left(\mu \in [0.535, 1.657]\right) \simeq 95\%.$$

La large "fourchette" obtenue pour  $\mu$  montre le peu d'intérêt de conserver un grand nombre de décimales dans ces calculs, ce qui justifie, a posteriori, les usages des logiciels les plus courants. On remarquera, au passage, que l'on aurait pu, tout aussi bien, remplacer dans la formule précédente,  $\hat{\sigma}$  par  $s = \tilde{\sigma}$ . L'intervalle ainsi obtenu reste voisin du précédent. On obtient, avec le choix de  $s \simeq 1.282$

$$\mathbb{P}\left(\mu \in \left[\hat{\mu} - 1.960 \times \frac{s}{\sqrt{n}}, \hat{\mu} + 1.960 \times \frac{s}{\sqrt{n}}\right]\right) \simeq 95\%. \quad (1.3.21)$$

En prenant les valeurs numériques obtenues précédemment, on obtient

$$\mathbb{P}\left(\mu \in \left[1.096 - 1.960 \times \frac{1.282}{\sqrt{19}}, 1.096 + 1.960 \times \frac{1.282}{\sqrt{19}}\right]\right) \simeq 95\%,$$

soit encore

$$\mathbb{P}\left(\mu \in [0.520, 1.672]\right) \simeq 95\%.$$

Le logiciel *Statistica 6.1* fournit un intervalle de confiance de *Student* pour  $\mu$ . Cet intervalle est, également, asymptotique, dans la mesure où la loi des observations n'est pas gaussienne. Cet intervalle est donné par

$$\mathbb{P}\left(\mu \in [0, 478242, 1, 714074]\right) \simeq 95\%.$$

Un autre logiciel, *SPSS 11.5.1* fournit un intervalle de Student analogue, arrondi, cette fois-ci, à la quatrième décimale, et donné par

$$\mathbb{P}\left(\mu \in \left[0,4782, 1,7140\right]\right) \simeq 95\%.$$

On constate aisément que les trois intervalles de confiance, donnés par l'une ou l'autre des méthodes, sont voisins.

En fait, l'application de ces techniques à l'échantillon étudié peut être discutée, du fait de la présence éventuelle de *valeurs aberrantes*. Pour illustrer ce problème, nous fournissons ci-dessous le tableau des valeurs de  $(X_i - \bar{X})^2$ , comparé à leur somme. Ce dernier est utilisé pour le calcul de  $\widehat{M}_4$ . On obtient, en arrondissant chaque entrée à la sixième décimale la plus voisine,

1	0.794655	11	0.591133
2	0.206560	12	6.404840
3	0.281127	13	0.002014
4	0.949613	14	0.258668
5	0.000920	15	1.049526
6	0.342771	16	0.267484
7	0.000482	17	0.200500
8	0.105677	18	0,000000
9	281.156832	19	0.023893
10	9.037289		
		Total	300.879329

Ce qui est le plus remarquable dans ce tableau est que la contribution de la seule donnée n°9 fournit une proportion de plus de 93% de la somme totale ! On constate donc que le fait de priver l'échantillon de l'observation n°9 modifie complètement certaines estimations. On est donc amené à suspecter que cette valeur puisse être *aberrante*.

La théorie des valeurs aberrante a donné lieu à une ample littérature. On consultera, en particulier :

Barnett, V. et Lewis, T. (1995). *Outliers in Statistical Data*. 3<sup>ème</sup> Ed., Wiley, New York.

Le fait qu'une observation soit *aberrante*, se réfère à un *modèle admis* pour les observations. Le statisticien a le choix entre l'une ou l'autre des alternatives suivantes :

- Soit il admet que les observations  $X_1, \dots, X_n$  contiennent une petite proportion de variables suivant une autre distribution que la loi de la majorité des variables de l'échantillon, supposées être conformes à un *modèle admis*. Il s'agit alors de détecter ces *observations aberrantes*, et de les *éliminer* de l'échantillon ;

- Soit il rejette le *modèle admis*, comme n'étant pas adapté à la description des observations, si des valeurs aberrantes sont détectées, alors qu'il n'y a pas de raison *a priori* de les éliminer de l'échantillon.

Il n'y a pas de règle spécifique justifiant qu'on privilégie un choix par rapport à l'autre. Par exemple, une méthode classique de détection des valeurs aberrantes consiste à supposer que la loi (de la majorité) des observations soit normale, et fait appel aux "résidus studentisés" (Studentized Residuals, voir Barnett et Lewis, op. cit. p.329). L'application de cette technique par le logiciel *SysStat*, identifie catégoriquement l'observation n°9 comme aberrante, avec une valeur de résidu studentisé égale à 5.031. Un échantillon de ce type peut donc faire l'objet d'une étude très complexe, menant à analyser sa structure dans le détail. Quelques uns de ces prolongements sont proposés en exercice.

**Exercice 1.3.2.** On considère l'échantillon présenté dans ce paragraphe, après élimination de l'observation n°9.

1°) Déterminer, sur ce nouvel échantillon, les estimations empiriques  $\widehat{\mu}$ ,  $\widehat{\sigma}^2$ ,  $\widehat{\beta}_1$ ,  $\widehat{\beta}_2$ .

2°) Déterminer, à partir des valeurs numériques de  $\hat{\mu}$  et  $\hat{\sigma}$ , un intervalle de confiance asymptotique à 95% pour la moyenne inconnue  $\mu$  des observations.

3° Comparer ces résultats à ceux obtenus à partir de l'échantillon complet.

## Chapitre 2

# Matrices, notations et factorisations.

### 2.1 Définitions générales et notations.

Les matrices considérées dans ce qui suit seront toujours supposées à coefficients réels, sauf mention du contraire. Dans la suite, on identifiera l'espace  $\mathbb{R}^p$  à l'espace des matrices colonnes d'ordre  $p$ . Cet espace est muni de la topologie usuelle, pouvant être commodément définie par la norme euclidienne

$$\|z\| = (z'z)^{1/2}, \quad (2.1.1)$$

où  $z'$  désigne la transposée de  $z \in \mathbb{R}^p$ .

Une matrice  $A$  à  $p$  lignes et  $q$  colonnes est dite  $(p \times q)$ . Il est commode de noter  $A = [a_{i,j}]$  lorsque  $A$  est une matrice dont l'élément générique de la  $i^{\text{ème}}$  ligne et de la  $j^{\text{ème}}$  colonne est  $a_{i,j}$ , et

$$A = [a_j] = [a_1 \quad \dots \quad a_q] \quad \text{où les } a_j = \begin{bmatrix} a_{1,j} \\ \vdots \\ a_{p,j} \end{bmatrix} \in \mathbb{R}^p, \quad \text{pour } j = 1, \dots, q,$$

désignent les *vecteurs colonnes* de  $A$ . Dans le cas où la matrice  $A = [a_{k,\ell}]$  est à coefficients complexes, avec  $a_{k,\ell} = u_{k,\ell} + iv_{k,\ell}$ , où  $u_{k,\ell} = \text{Re}(a_{k,\ell}) \in \mathbb{R}$  et  $v_{k,\ell} = \text{Im}(a_{k,\ell}) \in \mathbb{R}$  désignent respectivement la *partie réelle* et la *partie imaginaire* de  $a_{k,\ell}$ , on posera

$$\text{Re}(A) = [\text{Re}(a_{k,\ell})] = [u_{k,\ell}] \quad \text{et} \quad \text{Im}(A) = [\text{Im}(a_{k,\ell})] = [v_{k,\ell}]. \quad (2.1.2)$$

pour désigner respectivement la partie réelle et la partie imaginaire de la matrice  $A$ .

Avec ces notations, si  $A = [a_1 \quad \dots \quad a_q]$  est une matrice  $p \times q$ , on désigne par

$$\text{Vec}(A) = \begin{bmatrix} a_1 \\ \vdots \\ a_q \end{bmatrix} \in \mathbb{R}^{pq} \quad (2.1.3)$$

la matrice colonne, de paramètres  $(pq \times 1)$ , obtenue en empilant, dans l'ordre naturel, les colonnes de  $A$  les unes au dessus des autres.

On note  $\mathcal{M}_{p,q} = \mathcal{M}_{p,q}(\mathbb{R})$  l'espace vectoriel réel des matrices  $(p \times q)$  à coefficients réels. L'application

$$\text{Vec} : A \in \mathcal{M}_{p,q} \rightleftarrows \text{Vec}(A) \in \mathbb{R}^{pq} = \mathcal{M}_{pq,1},$$

définie par (2.1.3) est un isomorphisme d'espaces vectoriels entre  $\mathcal{M}_{p,q}$  et  $\mathbb{R}^{pq}$  (ce dernier espace étant, par convention, identifié à l'espace vectoriel  $\mathcal{M}_{pq,1}$  des matrices colonnes d'ordre  $pq$ ).

La *transposée* d'une matrice  $(p \times q)$ ,  $A = [a_{i,j}]$ , est la matrice  $(q \times p)$ , notée  $A' = [a_{i,j}^*]$ , telle que, pour tout  $1 \leq i \leq q$  et  $1 \leq j \leq p$ ,  $a_{i,j}^* = a_{j,i}$ .

Une matrice réelle (ou complexe)  $A$  est dite *carrée* lorsqu'elle est  $(p \times p)$  pour une valeur convenable de l'entier  $p \geq 1$ .

Une matrice carrée  $B$  est dite *symétrique* si elle est égale à sa transposée  $B'$ .

Une matrice carrée  $C$  est dite *antisymétrique*, si  $C = -C'$ . Une matrice carrée  $A$  se décompose de manière unique sous la forme  $A = A_1 + A_2 = \frac{1}{2}(A + A') + \frac{1}{2}(A - A')$ , où  $A_1 = \frac{1}{2}(A + A')$  est symétrique, et  $A_2 = \frac{1}{2}(A - A')$ , antisymétrique.

La *trace*, notée  $\text{tr}(A) = \sum_{i=1}^p a_{i,i}$  d'une matrice réelle  $(p \times p)$  carrée,  $A = [a_{i,j}]$  est définie de manière équivalente comme la somme des éléments diagonaux de  $A$  ou comme la somme des valeurs propres (qui existent toujours dans  $\mathbb{C}$ ) de  $A$ . Lorsque  $A$  est  $(p \times q)$  et  $B$  est  $(q \times p)$ , on peut définir, aussi bien  $AB$  (qui est  $(p \times p)$ ), que  $BA$  (qui est  $(q \times q)$ ). On a alors l'égalité

$$\text{tr}(AB) = \text{tr}(BA). \quad (2.1.4)$$

Dans la suite, nous ferons un usage fréquent de la notation, ayant un sens pour toute matrice carrée  $(p \times p)$   $A \in \mathcal{M}_{p,p}$ ,

$$\text{etr}(A) = \exp(\text{tr}(A)). \quad (2.1.5)$$

On note  $\text{diag}(\lambda_1, \dots, \lambda_d)$ , ou  $\text{diag}_d(\lambda_1, \dots, \lambda_d)$ , la matrice carrée  $(d \times d)$  ayant  $\lambda_1, \dots, \lambda_d$  sur sa diagonale principale et 0 ailleurs, soit

$$\text{diag}(\lambda_1, \dots, \lambda_d) = \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_d \end{bmatrix}$$

La matrice identité d'ordre  $d$  est notée  $\mathbb{I}_d = \text{diag}_d(1, \dots, 1)$ , ou tout simplement  $\mathbb{I}$ , lorsqu'il n'y a pas de doute sur sa dimension.

La matrice  $(p \times q)$ , dont les éléments sont tous égaux à 1, est notée  $E_{p,q}$  ou  $\mathbb{1}_{p,q}$ . On note  $\mathbb{1}_p = \mathbb{1}_{p,1}$ , ou  $\mathbb{1}$  lorsque la dimension  $p$  n'est pas ambiguë, la matrice colonne d'ordre  $p$  composée de 1. On observera que  $\mathbb{1}_{p,q} = \mathbb{1}_p \mathbb{1}_q'$ .

La matrice  $(p \times q)$ , dont les éléments sont tous nuls et égaux à 0, est notée  $\mathbb{O}_{p,q}$ , ou  $\mathbb{O}$  lorsque les dimensions  $p$  et  $q$  ne sont pas ambiguës. On note de même  $\mathbb{O}_p = \mathbb{O}_{p,1}$  la matrice colonne d'ordre  $p$  composée de 0. On observera que  $\mathbb{O}_{p,q} = \mathbb{O}_p \mathbb{O}_q'$ .

On identifie  $\mathbb{R}^p$  à l'espace  $\mathcal{M}_{p,1}$  des matrices colonnes  $(p \times 1)$ . Pour  $p \geq 1$  et  $q \geq 1$  quelconques, on munit l'espace  $\mathcal{M}_{p,q}$  du produit scalaire euclidien canonique

$$\langle A, B \rangle = \langle [a_{i,j}], [b_{i,j}] \rangle = \sum_{i=1}^p \sum_{j=1}^q a_{i,j} b_{i,j} = \text{tr}(A'B) = \text{tr}(B'A) = \text{tr}(AB') = \text{tr}(BA'). \quad (2.1.6)$$

En particulier, pour  $u \in \mathbb{R}^p$ , on retrouve dans (2.1.6) la forme habituelle du produit scalaire euclidien  $\langle u, v \rangle = u'v = v'u$ . Compte tenu des identités fournies par (2.1.6), il est aussi possible d'exprimer ce dernier sous la forme  $\langle u, v \rangle = \text{tr}(uv') = \text{tr}(vu')$ . On note  $\|A\| = \langle A, A \rangle^{1/2}$  pour  $A \in \mathcal{M}_{p,q}$  la norme euclidienne définie sur  $\mathcal{M}_{p,q}$  et correspondant au produit scalaire (2.1.6). On écrit  $A \perp B$  lorsque les vecteurs  $A, B \in \mathcal{M}_{p,q}$  sont orthogonaux, c'est à dire tels que  $\langle A, B \rangle = 0$ . On écrit  $E \perp F$ , pour exprimer le fait que  $E$  et  $F$  sont des sous-espaces vectoriels de  $\mathcal{M}_{p,q}$  vérifiant  $A \perp B$  pour tout  $A \in E$  et  $B \in F$ . De même, la relation  $A \perp F$  signifie que  $A \perp B$  pour tout  $B \in F$ .

Une suite finie  $F_1, \dots, F_k$  de sous-espaces vectoriels de  $\mathcal{M}_{p,q}$  est dite en *somme directe*, et égale à un sous-espace vectoriel  $E$  de  $\mathcal{M}_{p,q}$ , ce qui est noté

$$E = \bigoplus_{i=1}^k F_i,$$

si tout  $A \in E$  s'écrit de manière unique sous la forme  $A = \sum_{i=1}^k A_i$ , avec  $A_i \in F_i$  pour  $i = 1, \dots, k$ .

Si  $E$  est un sous-espace vectoriel de  $\mathcal{M}_{p,q}$ , on note  $E^\perp$  l'orthogonal de  $E$  dans  $\mathcal{M}_{p,q}$ , c'est à dire, le sous-espace vectoriel de  $\mathcal{M}_{p,q}$  composé des matrices  $A \in \mathcal{M}_{p,q}$  telles que  $A \perp E$ .

Si  $A$  est une matrice  $(p \times q)$ , on désigne par  $\text{Im}(A)$  (ne pas confondre avec  $\text{Im}(A)$  qui désigne la partie imaginaire de  $A$ ) le sous-espace de  $\mathbb{R}^p$  engendré par les colonnes de  $A$ , et par  $\text{Or}(A) = \text{Im}(A)^\perp$  l'orthogonal de  $\text{Im}(A)$  dans  $\mathbb{R}^p$ . Il revient au même de dire que  $\text{Im}(A)$  est l'espace image de  $\mathbb{R}^q$  dans  $\mathbb{R}^p$  par l'application linéaire

$$x \in \mathbb{R}^q \longrightarrow Ax \in \mathbb{R}^p.$$

Le *rang* de la matrice  $A$ , noté  $\text{rg}(A)$ , est, par définition, la dimension de l'espace image  $\text{Im}(A)$ . On a

$$\text{rg}(A) = \text{rg}(A') = \text{rg}(AA') = \text{rg}(A'A).$$

## 2.2 Matrices orthogonales.

Une matrice carrée  $H$ , réelle  $(d \times d)$ , est dite *orthogonale*, si  $HH' = H'H = \mathbb{I}_d$ , ou, ce qui revient au même, si ses colonnes  $h_1, \dots, h_d$ , correspondant à la décomposition  $H = [h_1 \ \dots \ h_d]$ , sont *orthonormales*, vérifiant

$$h'_i h_j = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i \neq j. \end{cases} \quad (2.2.7)$$

Une matrice orthogonale est associée à une transformation orthogonale de l'espace. Une telle transformation préserve les relations d'orthogonalité pour le produit scalaire usuel. On vérifie, en effet, que, si  $H$  est orthogonale,

$$u'v = 0 \Leftrightarrow (Hu)'(Hv) = u'H'Hv = 0. \quad (2.2.8)$$

Plus généralement, une tranformation orthogonale conserve le produit scalaire euclidien usuel. En effet, si  $H$  est orthogonale, on a, pour tout  $u, v \in \mathbb{R}^d$ ,

$$(Hu)'(Hv) = u'H'Hv = u'v. \quad (2.2.9)$$

Bien évidemment, si  $H$  est orthogonale, il en est de même de sa transposée  $H'$ , et réciproquement. Une matrice orthogonale admet des valeurs propres réelles ou complexes de module 1. De ce fait, son déterminant est un nombre de module 1, nécessairement réel puisque la matrice est réelle. Il ne peut donc être égal qu'à 1 ou  $-1$ . Dans tous les cas, si  $H$  est orthogonale, on a, ou bien  $\det(H) = 1$ , ou bien  $\det(H) = -1$ , et, dans tous les cas,  $|\det(H)| = 1$ .

Une matrice carrée  $A$ , réelle, symétrique,  $(d \times d)$ , possède toujours des valeurs propres  $\lambda_1, \dots, \lambda_d$  réelles, et une base de vecteurs propres orthonormés dans  $\mathbb{R}^d$ . Elle peut donc se factoriser en

$$A = H \text{diag}(\lambda_1, \dots, \lambda_d) H^{-1} = H \text{diag}(\lambda_1, \dots, \lambda_d) H', \quad (2.2.10)$$

où  $H = [h_1 \ \dots \ h_d]$  est une matrice orthogonale, composée de vecteurs propres de la transformation  $x \rightarrow Ax$ , et tels que

$$Ah_i = \lambda_i h_i \quad \text{pour } i = 1, \dots, d,$$

avec

$$h'_i h_j = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i \neq j. \end{cases}$$

La factorisation (2.2.10) n'est pas unique, même lorsque les valeurs propres de  $A$  sont distinctes. En effet, si  $H = [h_1 \ \dots \ h_d]$  vérifie (2.2.10), il en est de même de  $H = [\pm h_1 \ \dots \ \pm h_d]$ . Réciproquement, toute matrice



$A$  possédant la factorisation (2.2.10) est réelle symétrique. Il est souvent commode d'écrire (2.2.10) sous la forme équivalente, développée à partir de (2.2.10),

$$A = \sum_{j=1}^d \lambda_j h_j h_j' = [h_1 \ \dots \ h_d] \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_d \end{bmatrix} \begin{bmatrix} h_1' \\ \vdots \\ h_d' \end{bmatrix} \quad \text{avec} \quad h_j' h_\ell = \begin{cases} 1 & \text{si } j = \ell, \\ 0 & \text{si } j \neq \ell. \end{cases} \quad (2.2.11)$$

### 2.3 Matrices symétriques positives.

Une matrice carrée réelle  $A$ , symétrique ( $d \times d$ ), est dite *positive*, ce qui est noté  $A \geq 0$ , si  $u' A u \geq 0 \ \forall u \in \mathbb{R}^d$ . Elle est dite *définie positive*, ce qui est noté  $A > 0$ , si elle est positive, et vérifie, de plus, la propriété  $u' A u = 0 \Leftrightarrow u = \mathbb{O}$ . Une condition nécessaire et suffisante pour que la matrice ( $d \times d$ ) symétrique  $A$  soit positive (resp. définie positive) est que ses valeurs propres  $\lambda_1, \dots, \lambda_d$  soient positives ou nulles (resp. strictement positives).

**Proposition 2.3.1.** *Si  $A \geq 0$  est une matrice positive, alors, pour tout entier  $N \in \mathbb{N}^*$ , il existe une matrice positive  $X$  unique, notée  $X = A^{1/N}$ , telle que  $X^N = A$ .*

**Preuve.** L'existence de  $X$  découle de (2.2.10), en posant  $A^{1/N} = H \text{diag}(\lambda_1^{1/N}, \dots, \lambda_d^{1/N}) H^{-1}$  qui se trouve nécessairement être symétrique positive du fait que  $H^{-1} = H'$ . L'unicité se déduit du fait que, si  $X^N = A$ , alors  $X$  et  $A$  commutent (on a  $AX = XA$ ). Cette propriété implique que  $X$  et  $A$  sont triangulaires dans une même base, et, par conséquent, que  $X_1 = X_2$  chaque fois que  $X_j \geq 0$  et  $X_j^N = A$  pour  $j = 1, 2, \dots, d$ .

La propriété précédente permet de définir, d'une manière générale,  $\forall r \in \mathbb{R}^+$  lorsque  $A \geq 0$  et  $\forall r \in \mathbb{R}$  lorsque  $A > 0$ , la puissance  $r^{\text{ème}}$  de la matrice symétrique positive  $A = H \text{diag}(\lambda_1, \dots, \lambda_d) H'$  par

$$A^r = H \text{diag}(\lambda_1^r, \dots, \lambda_d^r) H^{-1} = H \text{diag}(\lambda_1^r, \dots, \lambda_d^r) H'. \quad (2.3.1)$$

La convention  $0^0 = 1$  implique, en particulier, que  $A^0 = \mathbb{I}_p$  (et ceci, même lorsque  $A = \mathbb{O}_{p,p}$ ). Avec ces conventions, la matrice  $A^r$ , lorsqu'elle est définie, est toujours symétrique positive (respectivement symétrique définie positive) lorsque  $A$  est positive (respectivement définie positive).

### 2.4 Factorisations à l'aide de matrices triangulaires et orthogonales.

**Proposition 2.4.2.** 1°) *Une matrice  $A > 0$ , définie positive ( $d \times d$ ), se factorise de manière unique en*

$$A = T' T, \quad \text{où la matrice} \quad T = \begin{bmatrix} t_{1,1} & t_{1,2} & \dots & t_{1,d} \\ 0 & t_{2,2} & \dots & t_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & t_{d,d} \end{bmatrix} \quad (2.4.2)$$

*est triangulaire supérieure, à éléments diagonaux strictement positifs, c'est à dire, telle que  $t_{i,i} > 0$  pour  $i = 1, \dots, d$  et  $t_{i,j} = 0$  pour  $i > j$ .*

2°) *Une matrice  $A \geq 0$ , positive ( $d \times d$ ), peut être factorisée en*

$$A = T' T, \quad \text{où la matrice} \quad T = \begin{bmatrix} t_{1,1} & t_{1,2} & \dots & t_{1,d} \\ 0 & t_{2,2} & \dots & t_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & t_{d,d} \end{bmatrix} \quad (2.4.3)$$

*est triangulaire supérieure à éléments diagonaux positifs ou nuls, c'est à dire, telle que  $t_{i,i} \geq 0$  pour  $i = 1, \dots, d$ . Dans ce dernier cas, la matrice  $T$  n'est pas nécessairement définie de manière unique.*

**Preuve.** Les propriétés 1°) et (2°) sont évidentes lorsque  $d = 1$  puisqu'alors  $A = [a] = T'T = [t_{1,1}^2]$  et le seul choix possible pour  $T = [t_{1,1}]$  est obtenu pour  $t_{1,1} = \sqrt{a}$ , si on impose la positivité des termes diagonaux. Supposons maintenant que les propriétés 1°) et 2°) aient été établies pour toutes les matrices  $((d-1) \times (d-1))$ . Ecrivons

$$A = \begin{bmatrix} a_{11} & \alpha'_1 \\ \alpha_1 & A_{2,2} \end{bmatrix} = T'T = \begin{bmatrix} t_{1,1} & \mathbb{O} \\ \tau_1 & T'_{2,2} \end{bmatrix} \begin{bmatrix} t_{1,1} & \tau'_1 \\ \mathbb{O} & T_{2,2} \end{bmatrix} = \begin{bmatrix} t_{1,1}^2 & t_{1,1}\tau'_1 \\ \tau_1 t_{1,1} & \tau_1 \tau'_1 + T'_{2,2} T_{2,2} \end{bmatrix}. \quad (2.4.4)$$

Considérons deux cas possibles, en fonction de la valeur de  $a_{1,1}$ . Notons que, nécessairement, la condition  $A \geq 0$  (resp.  $A > 0$ ) impose que  $a_{1,1} \geq 0$  et  $A_{2,2} \geq 0$  (resp.  $a_{1,1} > 0$  et  $A_{2,2} > 0$ ).

**Cas 1.** Supposons  $a_{1,1} = 0$  (ceci ne peut se produire que si  $A \geq 0$  n'est pas définie). Dans ce cas, nécessairement  $\alpha_1 = \mathbb{O}$ . Autrement, on obtiendrait, pour tout choix de  $y$  tel que  $y < -(\alpha'_1 A_{2,2} \alpha_1)/(2\alpha'_1 \alpha_1)$ ,

$$\begin{bmatrix} y & \alpha'_1 \end{bmatrix} A \begin{bmatrix} y \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} y & \alpha'_1 \end{bmatrix} \begin{bmatrix} 0 & \alpha'_1 \\ \alpha_1 & A_{2,2} \end{bmatrix} \begin{bmatrix} y \\ \alpha_1 \end{bmatrix} = 2y\alpha'_1 \alpha_1 + \alpha'_1 A_{2,2} \alpha_1 < 0,$$

ce qui contredirait la positivité de  $A$ . Ce cas n'est donc possible que si  $A \geq 0$  n'est pas définie positive. On constate alors que la relation (2.4.4) impose  $t_{1,1} = 0$ , mais permet des choix arbitraires de  $\tau_1$  et  $T_{2,2}$ , pourvu que ceux-ci vérifient

$$\tau_1 \tau'_1 + T'_{2,2} T_{2,2} = A_{2,2}.$$

On peut toujours prendre  $\tau_1 = \mathbb{O}$ , et choisir pour  $T_{2,2}$  une matrice triangulaire supérieure à diagonale positive ou nulle telle que  $T'_{2,2} T_{2,2} = A_{2,2}$ , qui existe par l'hypothèse de récurrence. Toutefois, ce n'est, en général pas la seule possibilité. Par exemple, la factorisation

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \cos \theta & \sin \theta \end{bmatrix} \begin{bmatrix} 0 & \cos \theta \\ 0 & \sin \theta \end{bmatrix},$$

répond à la question 2°) indépendamment de  $\theta \in (-\pi, \pi)$  pour la matrice

$$A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \geq 0.$$

Bien que l'unicité de  $T$  ne soit pas garantie, on établit néanmoins ainsi (2.4.3), et donc la propriété 2°) pour les matrices symétriques positives  $(d \times d)$  telles que  $a_{1,1} = 0$ .

**Cas 2.** Supposons maintenant que  $a_{1,1} > 0$ . Dans ce dernier cas, la relation (2.4.4) définit de manière unique  $t_{1,1} > 0$  et  $\tau_1$ , grâce aux formules (se déduisant de (2.4.4))

$$t_{1,1} = \sqrt{a_{1,1}} \quad \text{et} \quad \tau_1 = \frac{1}{t_{1,1}} \alpha_1 = \frac{1}{\sqrt{a_{1,1}}} \alpha_1.$$

Il reste à établir l'existence de  $T_{2,2}$  (et son unicité lorsque  $A > 0$ ). Or celle-ci se déduit de l'hypothèse de récurrence jointe à la vérification que la matrice

$$M = A_{2,2} - \tau_1 \tau'_1 = A_{2,2} - \frac{1}{a_{1,1}} \alpha_1 \alpha'_1$$

est positive lorsque  $A \geq 0$  (respectivement définie positive lorsque  $A > 0$ ). Nous nous limitons à la vérification du fait que  $A > 0$  implique que  $M > 0$ . Pour établir cette propriété, on constate que, pour tout  $v \in \mathbb{R}^{d-1}$ ,  $v \neq \mathbb{O}$ , le choix de  $x = -(v' \alpha_1)/t_{1,1}^2 = -(v' \alpha_1)/a_{1,1}$ , implique que

$$\begin{aligned} \begin{bmatrix} x & v' \end{bmatrix} A \begin{bmatrix} x \\ v \end{bmatrix} &= \begin{bmatrix} x & v' \end{bmatrix} \begin{bmatrix} a_{1,1} & \alpha'_1 \\ \alpha_1 & A_{2,2} \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix} = a_{1,1} x^2 + 2x \alpha'_1 v + v' A_{2,2} v \\ &= v' \left\{ A_{2,2} - \frac{1}{a_{1,1}} \alpha_1 \alpha'_1 \right\} v = v' M v > 0, \end{aligned}$$

ce qui suffit pour établir que  $M$  est définie positive. En factorisant  $M$  sous la forme  $M = T'_{2,2} T_{2,2}$ , ce qui est possible par l'hypothèse de récurrence, on montre ainsi que les propriétés 1°) et 2°) sont satisfaites pour les matrices  $(d \times d)$  telles que  $a_{1,1} > 0$ . La conclusion est évidente.  $\square$

**Proposition 2.4.3.** *Toute matrice  $(n \times d)$   $Z$  de rang  $\text{rg}(Z) = d \leq n$  se factorise de manière unique sous la forme*

$$Z = H_1 T, \quad (2.4.5)$$

où  $T$  est une matrice triangulaire supérieure  $(d \times d)$  à éléments diagonaux strictement positifs et  $H_1$  est une matrice  $(n \times d)$  telle que  $H_1' H_1 = \mathbb{I}_d$ .

**Preuve.** On constate que  $Z'Z \geq 0$  est une matrice symétrique positive  $(d \times d)$  de rang  $\text{rg}(Z'Z) = \text{rg}(Z) = d$ . Comme ceci implique que  $Z'Z > 0$ , on en déduit par (2.4.2) l'existence et l'unicité d'une matrice triangulaire supérieure à diagonale positive  $T$  telle que  $Z'Z = T'T$ . En posant  $H_1 = ZT^{-1}$ , on constate que  $H_1' H_1 = (T')^{-1} Z T^{-1} = (T')^{-1} T' T T^{-1} = \mathbb{I}_d$ . On a ainsi établi l'existence de  $T$  et de  $H_1$ . L'unicité de ces deux matrices est évidente par la construction précédente, qui implique que  $Z'Z = T'T$  et  $H_1 = ZT^{-1}$ .  $\square$

## 2.5 Produit de Kronecker.

Le *produit de Kronecker* entre une matrice  $(p \times q)$   $A = [a_{ij}]$  et une matrice  $(r \times s)$   $B$ , est la matrice  $(pr \times qs)$ , notée  $A \otimes B$ , et définie par

$$A \otimes B = \begin{bmatrix} a_{11}B & \dots & a_{1q}B \\ \vdots & \ddots & \vdots \\ a_{p1}B & \dots & a_{pq}B \end{bmatrix}. \quad (2.5.1)$$

Le produit de Kronecker est, bien évidemment, une opération non commutative. Il y a cependant des cas où  $A \otimes B = B \otimes A$ . Par exemple, toute matrice  $A$  peut être factorisée en un produit de Kronecker de la forme

$$A = A \otimes [1] = [1] \otimes A.$$

Lorsque  $D = d'$ , pour  $d \in \mathbb{R}^p$ , est une matrice ligne  $(1 \times m)$ , on peut écrire  $[1] D = D$ , et la propriété ci-dessus implique alors, par le (iv) de la Propriété 1.1 ci-dessous, que

$$A(C \otimes D) = (A \otimes [1])(C \otimes D) = AC \otimes ([1] D) = AC \otimes D \quad \forall D (1 \times m). \quad (2.5.2)$$

**Propriété 2.5.1.** *Le produit de Kronecker vérifie les propriétés suivantes. Sous réserve que les quantités correspondantes soient définies, on a :*

- (i) Pour tout couple de scalaires  $a, b \in \mathbb{R}$ ,  $(aA) \otimes (bB) = ab(A \otimes B)$ .
- (ii)  $(A + B) \otimes C = A \otimes C + B \otimes C$ , et  $A \otimes (C + D) = A \otimes C + A \otimes D$ .
- (iii)  $(A \otimes B) \otimes C = A \otimes (B \otimes C)$ .
- (iv)  $(A \otimes B)(C \otimes D) = AC \otimes BD$ .
- (v) Si la matrice carrée  $(m \times m)$   $A$  a pour valeurs propres  $\{a_1, \dots, a_m\}$  et la matrice carrée  $(n \times n)$   $B$  a pour valeurs propres  $\{b_1, \dots, b_n\}$ , alors la matrice carrée  $(mn \times mn)$   $A \otimes B$  a pour valeurs propres  $\{a_i b_j : 1 \leq i \leq m, 1 \leq j \leq n\}$ .
- (vi) Si  $A$  et  $B$  sont des matrices carrées inversibles, alors  $(A \otimes B)^{-1} = (A^{-1} \otimes B^{-1})$ .
- (vii) Si  $A$  est  $(m \times m)$  et  $B$   $(n \times n)$ , alors  $\det(A \otimes B) = (\det A)^n (\det B)^m$ .
- (viii) Si  $A > 0$  et  $B > 0$ , alors  $A \otimes B > 0$ .
- (ix) Si  $A$  et  $B$  sont carrées, alors  $\text{tr}(A \otimes B) = (\text{tr } A)(\text{tr } B)$ .
- (x)  $(A \otimes B)' = (A' \otimes B')$ .
- (xi)  $\mathbb{I}_p \otimes \mathbb{I}_q = \mathbb{I}_{pq}$ .
- (xii)  $\mathbb{1}_p \otimes \mathbb{1}_q = \mathbb{1}_{pq}$ .

Le produit de Kronecker  $\otimes$ , défini en (2.5.1), est lié à l'opération  $\text{Vec}(\cdot)$ , définie en (2.1.3), par les propriétés suivantes. Celles-ci sont parfois peu évidentes, et les énoncés correspondants pourront être omis en première lecture.

**Lemme 2.5.1.** *Si  $P$  est une matrice  $(r \times m)$ ,  $\mathbb{X}'$  une matrice  $(m \times n)$  et  $Q$  une matrice  $(n \times s)$ , alors*

$$\text{Vec}(P\mathbb{X}'Q) = (Q' \otimes P)\text{Vec}(\mathbb{X}'). \quad (2.5.3)$$

**Preuve.** Posons  $Q = [q_{ij}] \in \mathcal{M}_{n,s}$  et  $\mathbb{X}' = [X_1 \ \dots \ X_n] \in \mathcal{M}_{m,n}$ . Comme

$$P\mathbb{X}' = [PX_1 \ \dots \ PX_n], \quad Q' \otimes P = \begin{bmatrix} q_{11}P & \dots & q_{n1}P \\ \vdots & \ddots & \vdots \\ q_{1s}P & \dots & q_{ns}P \end{bmatrix}, \quad \text{Vec}(\mathbb{X}') = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix},$$

on a

$$\begin{aligned} (Q' \otimes P)\text{Vec}(\mathbb{X}') &= \begin{bmatrix} q_{11}P & \dots & q_{n1}P \\ \vdots & \ddots & \vdots \\ q_{1s}P & \dots & q_{ns}P \end{bmatrix} \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n PX_i q_{i1} \\ \vdots \\ \sum_{i=1}^n PX_i q_{is} \end{bmatrix} \\ &= \text{Vec}([\sum_{i=1}^n PX_i q_{i1} \ \dots \ \sum_{i=1}^n PX_i q_{is}]) \\ &= \text{Vec}([PX_1 \ \dots \ PX_n] Q) = \text{Vec}(P\mathbb{X}'Q), \end{aligned}$$

ce qu'il fallait démontrer.  $\square$

**Lemme 2.5.2.** *Sous réserve que les produits correspondants soient définis, on a :*

- (i)  $\text{Vec}(BC) = (\mathbb{I} \otimes B)\text{Vec}(C) = (C' \otimes \mathbb{I})\text{Vec}(B) = (C' \otimes B)\text{Vec}(\mathbb{I})$  ;
- (ii)  $\text{tr}(BCD) = (\text{Vec}(B'))'(\mathbb{I} \otimes C)\text{Vec}(D)$  ;
- (iii)  $\text{tr}(BX'CXD) = (\text{Vec}(X))'(B'D' \otimes C)\text{Vec}(X) = (\text{Vec}(X))'(DB \otimes C')\text{Vec}(X)$ .

**Preuve.** Ce lemme est dû à Neudecker (Neudecker, H. (1969). Some theorems on matrix differentiation with special reference to Kronecker matrix products. *Journal of the American Statistical Association.* **64** 953–963).

– La propriété (i) est une conséquence directe de (2.5.3). En écrivant cette dernière relation sous la forme

$$\text{Vec}(\mathcal{P}\mathcal{X}'\mathcal{Q}) = (\mathcal{Q}' \otimes \mathcal{P})\text{Vec}(\mathcal{X}'), \quad (2.5.4)$$

– On obtient la relation  $\text{Vec}(BC) = (\mathbb{I} \otimes B)\text{Vec}(C)$  en posant  $\mathcal{P} = B$ ,  $\mathcal{X}' = C$  et  $\mathcal{Q} = \mathbb{I} = \mathbb{I}'$  dans (2.5.4).

– On obtient la relation  $\text{Vec}(BC) = (C' \otimes \mathbb{I})\text{Vec}(B)$  en posant  $\mathcal{P} = \mathbb{I}$ ,  $\mathcal{X}' = B$  et  $\mathcal{Q} = C$  dans (2.5.4).

– On obtient la relation  $\text{Vec}(BC) = (C' \otimes B)\text{Vec}(\mathbb{I})$  en posant  $\mathcal{P} = B$ ,  $\mathcal{X}' = \mathbb{I}$  et  $\mathcal{Q} = C$  dans (2.5.4).

– La propriété (ii) se vérifie directement, d'abord en effectuant les produits considérés dans chaque membre, puis en constatant que les résultats sont identiques. En effet, en posant

$$B' = [b_1 \ \dots \ b_n] \quad \text{et} \quad D = [d_1 \ \dots \ d_n],$$

on constate que  $(\text{Vec}(B'))' = [b'_1 \ \dots \ b'_n]$ , et donc, que

$$(\text{Vec}(B'))'(\mathbb{I} \otimes C)\text{Vec}(D) = [b'_1 \ \dots \ b'_n] \begin{bmatrix} C & \dots & \mathbb{O} \\ \vdots & \ddots & \vdots \\ \mathbb{O} & \dots & C \end{bmatrix} \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix} = \sum_{i=1}^n b'_i C d_i = \text{tr}(BCD).$$

– La première partie de la propriété (iii) est conséquence de la relation (ii), écrite sous la forme

$$\text{tr}(\mathcal{B}\mathcal{C}\mathcal{D}) = (\text{Vec}(\mathcal{B}'))'(\mathbb{I} \otimes \mathcal{C})\text{vec}(\mathcal{D}), \quad (2.5.5)$$

ainsi que de (i), écrite sous la forme

$$\text{vec}(\mathcal{U}\mathcal{V}) = (\mathcal{V}' \otimes \mathbb{I})\text{vec}(\mathcal{U}). \quad (2.5.6)$$

On écrit alors les identités successives

$$\text{tr}(BX'CXD) = \text{tr}\{(BX')C(XD)\}$$

$$\begin{aligned}
&= (\text{Vec}(XB'))'(\mathbb{I} \otimes C)(\text{Vec}(XD)) \\
(\text{en posant } \mathbf{B} = BX', \mathbf{C} = C \text{ et } \mathbf{D} = XD \text{ dans (2.5.5)}), \\
&= ((B \otimes \mathbb{I})\text{Vec}(X))'(\mathbb{I} \otimes C)((D' \otimes \mathbb{I})\text{Vec}(X)) \\
(\text{en posant } \mathbf{U} = X, \mathbf{V} = B', \text{ puis } \mathbf{U} = X, \mathbf{V} = D \text{ dans (2.5.6)}) \\
&= ((\text{Vec}(X))'(B' \otimes \mathbb{I}))(\mathbb{I} \otimes C)(D' \otimes \mathbb{I})\text{Vec}(X) \\
(\text{par le (x) de la propriété 1.1, qui montre que } (B \otimes \mathbb{I})' = B' \otimes \mathbb{I}) \\
&= (\text{Vec}(X))'(B'D' \otimes C)\text{Vec}(X) \\
(\text{par le (iv) de la propriété 1.1 qui montre que } (\mathcal{A} \otimes \mathcal{B})(\mathcal{C} \otimes \mathcal{D}) = \mathcal{A}\mathcal{C} \otimes \mathcal{B}\mathcal{D}).
\end{aligned}$$

On établit ainsi que

$$\text{tr}(BX'CXD) = (\text{Vec}(X))'(B'D' \otimes C)\text{Vec}(X). \quad (2.5.7)$$

Comme  $\text{tr}(\mathcal{M}) = \text{tr}(\mathcal{M}')$ , on en déduit que

$$\text{tr}(BX'CXD) = \text{tr}(D'X'C'XB') = (\text{Vec}(X))'(DB \otimes C')\text{Vec}(X),$$

la dernière égalité étant obtenue en remplaçant formellement  $B, C, D$  par  $D', B', C'$  dans la relation (2.5.7). Ceci achève la démonstration du lemme.  $\square$

## 2.6 Opérations Matricielles par Blocs

Soit  $A \in \mathcal{M}_{n,n}$  une matrice carrée ( $n \times n$ ), avec  $n = p + q$ ,  $p \geq 1, q \geq 1$ , partitionnée en

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad \text{où } A_{11} \in \mathcal{M}_{p,p} \text{ et } A_{22} \in \mathcal{M}_{q,q}. \quad (2.6.1)$$

L'inverse de  $A$ , lorsqu'il existe, est noté

$$A^{-1} = \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{bmatrix}, \quad \text{où } A^{11} \in \mathcal{M}_{p,p} \text{ et } A^{22} \in \mathcal{M}_{q,q}. \quad (2.6.2)$$

**Proposition 2.6.1.** *Sous réserve que les inverses correspondants existent, on a les identités*

$$A^{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}, \quad A^{12} = -A^{11}A_{12}A_{22}^{-1} = -A_{11}^{-1}A_{12}A^{22}, \quad (2.6.3)$$

$$A^{22} = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}, \quad A^{21} = -A^{22}A_{21}A_{11}^{-1} = -A_{22}^{-1}A_{21}A^{11}. \quad (2.6.4)$$

**Preuve.** Pour établir (2.6.3) et (2.6.4), il suffit de montrer que les matrices définies par

$$\mathcal{A}^{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}, \quad \mathcal{A}^{12} = -A_{11}^{-1}A_{12}\mathcal{A}^{22}, \quad (2.6.5)$$

$$\mathcal{A}^{22} = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}, \quad \mathcal{A}^{21} = -A_{22}^{-1}A_{21}\mathcal{A}^{11}. \quad (2.6.6)$$

vérifient l'égalité

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \mathcal{A}^{11} & \mathcal{A}^{12} \\ \mathcal{A}^{21} & \mathcal{A}^{22} \end{bmatrix} = \begin{bmatrix} A_{11}\mathcal{A}^{11} + A_{12}\mathcal{A}^{21} & A_{11}\mathcal{A}^{12} + A_{12}\mathcal{A}^{22} \\ A_{21}\mathcal{A}^{11} + A_{22}\mathcal{A}^{21} & A_{21}\mathcal{A}^{12} + A_{22}\mathcal{A}^{22} \end{bmatrix} = \begin{bmatrix} \mathbb{I} & \mathbb{O} \\ \mathbb{O} & \mathbb{I} \end{bmatrix}. \quad (2.6.7)$$

En effet, si celle-ci est satisfaite, on a nécessairement  $\mathcal{A}^{ij} = A^{ij}$  pour  $1 \leq i, j \leq 2$ , et, par voie de conséquence, (2.6.5) et (2.6.6). Une fois ces relations obtenues, on en déduit directement (2.6.3) et (2.6.4). Pour établir (2.6.7), on commence par écrire

$$\begin{aligned}
&A_{11}\mathcal{A}^{12} + A_{12}\mathcal{A}^{22} = \mathbb{O} \quad \text{et} \quad A_{21}\mathcal{A}^{12} + A_{22}\mathcal{A}^{22} = \mathbb{I} \\
\Leftrightarrow &\mathcal{A}^{12} = -A_{11}^{-1}A_{12}\mathcal{A}^{22} \quad \text{et} \quad A_{21}\mathcal{A}^{12} + A_{22}\mathcal{A}^{22} = \mathbb{I} \\
\Leftrightarrow &\mathcal{A}^{12} = -A_{11}^{-1}A_{12}\mathcal{A}^{22} \quad \text{et} \quad (A_{22} - A_{21}A_{11}^{-1}A_{12})\mathcal{A}^{22} = \mathbb{I},
\end{aligned}$$

ce qui permet de vérifier que  $\mathcal{A}^{12} = A^{12}$  et  $\mathcal{A}^{22} = A^{22}$ . De même, on constate que

$$\begin{aligned} A_{22}\mathcal{A}^{21} + A_{21}\mathcal{A}^{11} &= \mathbb{O} \quad \text{et} \quad A_{12}\mathcal{A}^{21} + A_{11}\mathcal{A}^{11} = \mathbb{I} \\ \Leftrightarrow \mathcal{A}^{21} &= -A_{22}^{-1}A_{21}\mathcal{A}^{11} \quad \text{et} \quad A_{12}\mathcal{A}^{21} + A_{11}\mathcal{A}^{11} = \mathbb{I} \\ \Leftrightarrow \mathcal{A}^{21} &= -A_{22}^{-1}A_{21}\mathcal{A}^{11} \quad \text{et} \quad (A_{11} - A_{12}A_{22}^{-1}A_{21})\mathcal{A}^{11} = \mathbb{I}, \end{aligned}$$

ce qui permet de conclure que  $\mathcal{A}^{21} = A^{21}$  et  $\mathcal{A}^{11} = A^{11}$ , achevant la démonstration.  $\square$

**Proposition 2.6.2.** *On a les identités*

$$\det \begin{bmatrix} A_{11} & \mathbb{O} \\ A_{21} & A_{22} \end{bmatrix} = \det(A_{11}) \times \det(A_{22}) = \det \begin{bmatrix} A_{11} & A_{21} \\ \mathbb{O} & A_{22} \end{bmatrix}, \quad (2.6.8)$$

et, sous réserve que les inverses  $A_{11}^{-1}$  et  $A_{22}^{-1}$  existent,

$$\det \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \det(A_{11}) \times \det\{A_{22} - A_{21}A_{11}^{-1}A_{12}\} \quad (2.6.9)$$

$$= \det(A_{22}) \times \det\{A_{11} - A_{12}A_{22}^{-1}A_{21}\}. \quad (2.6.10)$$

**Preuve.** Supposons que  $A$  soit  $(n \times n)$ , que  $A_{11}$  soit  $(p \times p)$ , et que  $A_{22}$  soit  $((n-p) \times (n-p))$ . La première égalité dans la relation (2.6.8) se démontre en posant

$$A = \begin{bmatrix} A_{11} & \mathbb{O} \\ A_{21} & A_{22} \end{bmatrix} = [a_1 \quad \dots \quad a_n] \quad \text{et} \quad \begin{bmatrix} \mathbb{O} \\ A_{22} \end{bmatrix} = [a_{p+1} \quad \dots \quad a_n].$$

De deux chose l'une. Ou bien  $\det(A_{22}) = 0$ , et alors nécessairement  $\det(A) = 0$ , ou alors  $\det(A_{22}) \neq 0$ , et on peut retrancher aux  $p$  premières colonnes de  $A$  des combinaisons linéaires de  $a_{p+1}, \dots, a_n$  de sorte à construire une nouvelle matrice de la forme

$$A^* = \begin{bmatrix} A_{11} & \mathbb{O} \\ \mathbb{O} & A_{22} \end{bmatrix}.$$

En effet, pour chaque  $i = 1, \dots, p$ , il est possible de choisir  $\lambda_{p+1,i}, \dots, \lambda_{n,i}$  tels que, si

$$a_i = \begin{bmatrix} a_{1i} \\ a_{2i} \end{bmatrix} \quad \text{pour} \quad i = 1, \dots, p \quad \text{et} \quad A_{21} = [a_{21} \quad \dots \quad a_{2p}],$$

alors, pour  $i = 1, \dots, p$ ,

$$a_{2i} = \sum_{j=p+1}^n \lambda_j a_j = [a_{p+1} \quad \dots \quad a_n] \begin{bmatrix} \lambda_{p+1,i} \\ \vdots \\ \lambda_{n,i} \end{bmatrix} = A_{22} \begin{bmatrix} \lambda_{p+1,i} \\ \vdots \\ \lambda_{n,i} \end{bmatrix},$$

ce qui revient à poser, pour  $i = 1, \dots, p$ ,

$$\begin{bmatrix} \lambda_{p+1,i} \\ \vdots \\ \lambda_{n,i} \end{bmatrix} = A_{22}^{-1}a_{2i}.$$

Comme, dans ce cas,  $\det(A) = \det(A^*) = \det(A_{11})\det(A_{22})$ , on obtient la première partie de (2.6.8). La seconde partie se démontre par un procédé analogue.

Pour établir (2.6.9) et (2.6.10), on pose

$$C = \begin{bmatrix} \mathbb{I} & -A_{12}A_{22}^{-1} \\ \mathbb{O} & \mathbb{I} \end{bmatrix} \quad \text{et} \quad D = \begin{bmatrix} \mathbb{I} & -A_{11}^{-1}A_{12} \\ \mathbb{O} & \mathbb{I} \end{bmatrix},$$

on écrit

$$\begin{aligned}
CAD &= \begin{bmatrix} \mathbb{I} & -A_{12}A_{22}^{-1} \\ \mathbb{O} & \mathbb{I} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \mathbb{I} & -A_{11}^{-1}A_{12} \\ \mathbb{O} & \mathbb{I} \end{bmatrix} \\
&= \begin{bmatrix} A_{11} - A_{12}A_{22}^{-1}A_{21} & \mathbb{O} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \mathbb{I} & -A_{11}^{-1}A_{12} \\ \mathbb{O} & \mathbb{I} \end{bmatrix} \\
&= \begin{bmatrix} \mathbb{I} & -A_{12}A_{22}^{-1} \\ \mathbb{O} & \mathbb{I} \end{bmatrix} \begin{bmatrix} A_{11} & \mathbb{O} \\ A_{21} & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix}.
\end{aligned}$$

Comme, par (2.6.8), on a

$$\det \begin{bmatrix} \mathbb{I} & -A_{11}^{-1}A_{12} \\ \mathbb{O} & \mathbb{I} \end{bmatrix} = \det \begin{bmatrix} \mathbb{I} & -A_{12}A_{22}^{-1} \\ \mathbb{O} & \mathbb{I} \end{bmatrix} = 1,$$

on constate que  $\det(CAD) = \det(A)$ , puis que

$$\det(A) = \det \begin{bmatrix} A_{11} - A_{12}A_{22}^{-1}A_{21} & \mathbb{O} \\ A_{21} & A_{22} \end{bmatrix} = \det \begin{bmatrix} A_{11} & \mathbb{O} \\ A_{21} & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix}.$$

Une nouvelle utilisation de (2.6.8), mène alors à (2.6.9), puis à (2.6.10).  $\square$

**Corollaire 2.6.1.** *Pour tout couple de matrices  $C \in \mathcal{M}_{p,q}$  et  $D \in \mathcal{M}_{q,p}$ , on a*

$$\det\{\mathbb{I}_p + CD\} = \det\{\mathbb{I}_q + DC\}. \quad (2.6.11)$$

De plus, si  $A \in \mathcal{M}_{p,p}$  est inversible, on a

$$\begin{aligned}
\det\{A + CD\} &= \det(A) \times \det\{\mathbb{I} + A^{-1}CD\} \\
&= \det(A) \times \det\{\mathbb{I} + DA^{-1}C\} \\
&= \det(A) \times \det\{\mathbb{I} + CDA^{-1}\}
\end{aligned} \quad (2.6.12)$$

**Preuve.** Pour établir (2.6.11), il suffit de poser  $A_{11} = \mathbb{I}_p$ ,  $A_{22} = \mathbb{I}_q$ ,  $A_{12} = -C$  et  $A_{21} = D$  dans (2.6.9) et (2.6.10). Les relations (2.6.12) sont obtenues comme cas particuliers de (2.6.11), après avoir effectué la factorisation

$$A + CD = A(\mathbb{I} + A^{-1}CD),$$

et avoir calculé le déterminant des deux membres. On utilise, en particulier, le fait que, par (2.6.11),

$$\det(\mathbb{I} + A^{-1}\{CD\}) = \det(\mathbb{I} + \{A^{-1}C\}D) = \det(\mathbb{I} + \{CD\}A^{-1}) = \det(\mathbb{I} + D\{A^{-1}C\}),$$

pour conclure.  $\square$

**Corollaire 2.6.2.** *Si  $C$  et  $D$  sont des matrices carrées ( $m \times m$ ), les valeurs propres  $y_1, \dots, y_m$  de  $CD$  sont identiques à celles de  $DC$ .*

**Preuve.** Par le corollaire 1.1, le polynôme caractéristique de  $CD$  est donné par la formule

$$\det(\mathbb{I}_m - \lambda CD) = \det(\mathbb{I}_m - C\lambda D) = \det(\mathbb{I}_m - \lambda DC).$$

Il coïncide donc avec le polynôme caractéristique de  $DC$ , ce qui permet de conclure.  $\square$

## 2.7 Moments de vecteurs aléatoires

Soit  $d \geq 1$  un entier positif, et soit  $X = [X_1 \ \dots \ X_d]^\top \in \mathbb{R}^d$  un vecteur aléatoire réel. Pour tout  $k \in \mathbb{N}$ , l'existence de moments d'ordre  $k$  de  $X$  est équivalente à l'une des propriétés suivantes.

- (1)  $\mathbb{E}(\|X\|^k) = \mathbb{E}(\{X'X\}^{k/2}) < \infty$ ;
- (2)  $\mathbb{E}(|X_1|^{m_1} \dots |X_p|^{m_p}) < \infty \quad \forall m_1 \geq 0, \dots, m_p \geq 0$  entiers tels que  $m_1 + \dots + m_p = k$ ;
- (3)  $\mathbb{E}(|X_i|^k) < \infty \quad \forall i = 1, \dots, d$ ;
- (4)  $\mathbb{E}(\{u'X\}^k) \in \mathbb{R} \quad \forall u \in \mathbb{R}^d$ ;
- (5)  $\mathbb{E}(\{u'_1 X\} \dots \{u'_k X\}) \in \mathbb{R} \quad \forall u_1, \dots, u_k \in \mathbb{R}^d$

Formellement, il est commode de se baser sur (5) pour définir le *moment d'ordre  $k$*  de  $X$  comme la forme  $k$ -linéaire symétrique

$$(u_1, \dots, u_k) \in (\mathbb{R}^d)^k \rightarrow \mathbb{E}(\{u'_1 X\} \dots \{u'_k X\}). \quad (2.7.1)$$

Plus explicitement, si  $u_j = [u_{1,j} \ \dots \ u_{d,j}]' \in \mathbb{R}^d$  pour  $j = 1, \dots, k$ , la forme  $k$ -linéaire symétrique ci-dessus est égale à

$$\sum_{i_1=1}^d \dots \sum_{i_k=1}^d u_{i_1,1} \dots u_{i_k,k} \mathbb{E}(X_{i_1} \dots X_{i_k}). \quad (2.7.2)$$

Dans ce qui suit, nous accorderons une importance toute particulière au cas des moments d'ordre  $k = 1$  et  $k = 2$  de  $X$ , qui peuvent être identifiés respectivement par la relation ci-dessus à des formes linéaires et à des formes bilinéaires symétriques sur  $\mathbb{R}^p$ .

– L'existence de moments d'ordre 1 pour  $X$  est, par définition, équivalente à l'existence de moments d'ordre 1 pour chacune des coordonnées  $X_1, \dots, X_d$  de  $X$ . Par identification de  $\mathbb{R}^d$  et de son dual, ceci permet de définir le moment d'ordre 1 de  $X \in \mathbb{R}^d$ , appelé aussi *espérance de  $X$* , par

$$\mu_X = \mu = \mathbb{E}(X) = \begin{bmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_d) \end{bmatrix} \in \mathbb{R}^d. \quad (2.7.3)$$

La même formule est valable si  $X$  s'exprime dans une base quelconque, et en particulier, dans le cas d'une matrice aléatoire  $M = [m_{i,j}] \in \mathcal{M}_{p,q}$ , dont on définit l'espérance par  $\mathbb{E}(M) = [\mathbb{E}(m_{i,j})]$ . D'une manière générale, si  $\mathfrak{X}$  est un espace vectoriel réel de dimension finie, et si  $X \in \mathfrak{X}$  admet des moments d'ordre 1, l'application qui à  $X \in \mathfrak{X}$  associe  $\mathbb{E}(X)$  est un opérateur linéaire. En particulier, lorsque  $M \in \mathcal{M}_{p,q}$  est une matrice aléatoire possédant des moments d'ordre 1, on a l'identité générale, valable pour tout couple de matrices constantes (non aléatoires)  $A \in \mathcal{M}_{r,p}$  et  $B \in \mathcal{M}_{q,s}$ ,

$$\mathbb{E}(AMB) = A\mathbb{E}(M)B. \quad (2.7.4)$$

– L'existence de moments d'ordre 2 pour  $X \in \mathbb{R}^d$  est équivalente au fait que  $\mathbb{E}(X_i X_j) \in \mathbb{R} \ \forall 1 \leq i, j \leq d$ . Il est commode d'exprimer ceci à l'aide de la matrice

$$\mathbb{E}(XX') = \mathbb{E} \left( \begin{bmatrix} X_1^2 & \dots & X_1 X_d \\ \vdots & \ddots & \vdots \\ X_d X_1 & \dots & X_d^2 \end{bmatrix} \right) = \begin{bmatrix} \mathbb{E}(X_1^2) & \dots & \mathbb{E}(X_1 X_d) \\ \vdots & \ddots & \vdots \\ \mathbb{E}(X_d X_1) & \dots & \mathbb{E}(X_d^2) \end{bmatrix}. \quad (2.7.5)$$

**Proposition 2.7.1.** *Pour tout vecteur aléatoire  $X \in \mathbb{R}^d$  possédant des moments d'ordre 2, la matrice  $(d \times d)$  définie par  $\mathbb{E}(XX')$  est symétrique positive.*

**Preuve.** Il suffit de constater, par (2.7.4), que,  $\forall u \in \mathbb{R}^d$ ,

$$u' \mathbb{E}(XX') u = \mathbb{E}(u' X X' u) = \mathbb{E}(\{X' u\} \{X' u\}) = \mathbb{E}(\|X' u\|^2) = \mathbb{E}(\|u' X\|^2) \geq 0.$$

On remarquera, de plus, que  $\mathbb{E}(XX') > 0$  sauf s'il existe un  $u \in \mathbb{R}^d$  tel que  $\mathbb{P}(u' X = 0) = 1$ .  $\square$



– D'une manière générale, étant donné un vecteur aléatoire réel  $X = [X_1 \dots X_d]'$   $\in \mathbb{R}^d$ , les moments  $\mu_{k,\ell;i,j} = \mathbb{E}(X_i^k X_j^\ell)$ , lorsqu'ils existent, sont appelés moments croisés d'ordre  $k, \ell$  de  $X_i$  et  $X_j$ . On note, de même, les moments centrés correspondants par  $M_{k,\ell;i,j} = \mathbb{E}((X_i - \mathbb{E}(X_i))^k (X_j - \mathbb{E}(X_j))^\ell)$ .

En particulier, pour  $k = \ell = 1$ , on obtient la *covariance* entre  $X_i$  et  $X_j$ , définie  $\forall 1 \leq i, j \leq d$ , par

$$\text{Cov}(X_i, X_j) = M_{1,1;i,j} = \mathbb{E}((X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))). \quad (2.7.6)$$

En prenant  $i = j$  dans cette expression, on obtient la *variance* de  $X_i$ , définie,  $\forall 1 \leq i \leq d$ , par

$$\text{Cov}(X_i, X_i) = \text{Var}(X_i) = M_{1,1;i,i} = \mathbb{E}((X_i - \mathbb{E}(X_i))^2).$$

**Définition 2.7.1.** *Pour tout vecteur aléatoire  $X \in \mathbb{R}^d$  possédant des moments d'ordre 2, la matrice ( $d \times d$ ) symétrique positive définie par*

$$\text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))(X - \mathbb{E}(X))') = \begin{bmatrix} \text{Var}(X_1) & \dots & \text{Cov}(X_1, X_d) \\ \vdots & \ddots & \vdots \\ \text{Cov}(X_d, X_1) & \dots & \text{Var}(X_d) \end{bmatrix}, \quad (2.7.7)$$

*est appelée variance de  $X$ , ou matrice de variances-covariances de  $X$ .*

**Définition 2.7.2.** *Pour tout couple de vecteurs aléatoires  $X \in \mathbb{R}^p$  et  $Y \in \mathbb{R}^q$ , possédant chacun des moments d'ordre 2, la matrice ( $p \times q$ ) définie par*

$$\text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))') = \begin{bmatrix} \text{Cov}(X_1, Y_1) & \dots & \text{Cov}(X_1, Y_q) \\ \vdots & \ddots & \vdots \\ \text{Cov}(X_p, Y_1) & \dots & \text{Cov}(X_p, Y_q) \end{bmatrix}, \quad (2.7.8)$$

*est appelée matrice de covariance entre  $X$  et  $Y$ .*

Cette définition appelle plusieurs remarques. Tout d'abord, pour tout couple de vecteurs aléatoires  $X \in \mathbb{R}^p$  et  $Y \in \mathbb{R}^q$ , on a toujours, lorsque ces quantités existent,

$$\text{Cov}(X, Y) = \text{Cov}(Y, X)' \quad \text{avec} \quad \text{Cov}(X, Y) \in \mathcal{M}_{p,q} \quad \text{et} \quad \text{Cov}(Y, X) \in \mathcal{M}_{q,p}.$$

D'autre part, l'existence de moments d'ordre 2 pour  $X$  et  $Y$  équivaut à ce que,  $\forall u \in \mathbb{R}^p$  et  $\forall v \in \mathbb{R}^q$ ,

$$\mathbb{E}(\{u'(X - \mathbb{E}(X))\}^2) = u' \text{Var}(X) u < \infty \quad \text{et} \quad \mathbb{E}(\{v'(Y - \mathbb{E}(Y))\}^2) = v' \text{Var}(Y) v < \infty.$$

Par l'inégalité de Schwarz, on en déduit que,  $\forall u \in \mathbb{R}^p$  et  $\forall v \in \mathbb{R}^q$ ,

$$\begin{aligned} |u' \text{Cov}(X, Y) v| &= |\mathbb{E}(\{u'(X - \mathbb{E}(X))\} \{(Y - \mathbb{E}(Y))' v\})| \\ &\leq \{u' \text{Var}(X) u\}^{1/2} \{v' \text{Var}(Y) v\}^{1/2} < \infty, \end{aligned} \quad (2.7.9)$$

ce qui implique l'existence de  $\text{Cov}(X, Y)$ . En appliquant (2.7.9) dans le cas particulier des coordonnées de  $X$  et  $Y$ , on obtient les inégalités, valables  $\forall 1 \leq i \leq p$  et  $\forall 1 \leq j \leq q$ ,

$$\text{Cov}(X_i, Y_j) \leq \sqrt{\text{Var}(X_i) \text{Var}(Y_j)}. \quad (2.7.10)$$

Sous réserve que  $\text{Var}(X_i) \text{Var}(Y_j) > 0$ , on peut alors définir le *coefficient de corrélation* entre  $X_i$  et  $Y_j$  par le rapport

$$\rho_{X_i, Y_j} = \text{Corr}(X_i, Y_j) = \frac{\text{Cov}(X_i, Y_j)}{\sqrt{\text{Var}(X_i) \text{Var}(Y_j)}} \in [-1, 1], \quad (2.7.11)$$

et la *matrice de corrélation* entre  $X$  et  $Y$ , donnée par

$$\text{Corr}(X, Y) = \begin{bmatrix} \text{Corr}(X_1, Y_1) & \dots & \text{Corr}(X_1, Y_q) \\ \vdots & \ddots & \vdots \\ \text{Corr}(X_p, Y_1) & \dots & \text{Corr}(X_p, Y_q) \end{bmatrix}. \quad (2.7.12)$$

Dans le cas particulier où  $X = Y$ , en notant  $\sigma_i = \text{Var}(X_i)^{1/2} > 0$ ,  $\sigma_{i,j} = \text{Cov}(X_i, X_j)$  et  $\rho_{i,j} = \text{Corr}(X_i, X_j) = \sigma_{i,j}/(\sigma_i\sigma_j)$  pour  $1 \leq i, j \leq p$ , on obtient la *matrice des corrélations* de  $X$ , qui est nécessairement une matrice  $(p \times p)$ , symétrique et positive,

$$\begin{aligned} \text{Corr}(X) &= \text{Corr}(X, X) = \begin{bmatrix} 1 & \dots & \rho_{1,p} \\ \vdots & \ddots & \vdots \\ \rho_{p,1} & \dots & 1 \end{bmatrix} \\ &= \text{diag}(1/\sigma_1, \dots, 1/\sigma_p) \text{Var}(X) \text{diag}(1/\sigma_1, \dots, 1/\sigma_p) \geq 0. \end{aligned}$$

**Définition 2.7.3.** Deux vecteurs aléatoires  $X \in \mathbb{R}^p$  et  $Y \in \mathbb{R}^q$  possédant chacun des moments d'ordre 2 sont dits non corrélés (ou de covariance nulle) si et seulement si

$$\text{Cov}(X, Y) = \text{Cov}(Y, X)' = \mathbb{O} = \mathbb{O}_{p,q}. \quad (2.7.13)$$

**Proposition 2.7.2.** Si les vecteurs aléatoires  $X \in \mathbb{R}^p$  et  $Y \in \mathbb{R}^q$  possédant des moments d'ordre 2 sont indépendants, alors, ils sont non corrélés.

**Preuve.** Il suffit d'écrire, par (2.7.8), que

$$\text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))') = (\mathbb{E}(X) - \mathbb{E}(X))(\mathbb{E}(Y) - \mathbb{E}(Y))' = \mathbb{O}_{p,q},$$

cette dernière égalité se déduisant de l'indépendance de  $X$  et  $Y$ .  $\square$

**Proposition 2.7.3.** Si  $X \in \mathbb{R}^d$  est un vecteur aléatoire possédant des moments d'ordre 2, alors, pour tout couple de matrices constantes (non aléatoires)  $\nu \in \mathbb{R}^q$  et  $P \in \mathcal{M}_{q,d}$ , on a

$$\mathbb{E}(\nu + PX) = \nu + P\mathbb{E}(X) \quad \text{et} \quad \text{Var}(\nu + PX) = P \text{Var}(X) P'. \quad (2.7.14)$$

**Proposition 2.7.4.** Si  $Z_1, \dots, Z_n$  est une suite de vecteurs aléatoires non corrélés de  $\mathbb{R}^d$ , alors

$$\text{Var}(Z_1 + \dots + Z_n) = \text{Var}(Z_1) + \dots + \text{Var}(Z_n). \quad (2.7.15)$$

**Preuve.** On constate, par (2.7.8), que, lorsque  $Z_1, \dots, Z_n$  possèdent des moments d'ordre 2,

$$\text{Var}\left(\sum_{i=1}^n Z_i\right) = \sum_{i=1}^n \text{Var}(Z_i) + \sum_{1 \leq i \neq j \leq n} \text{Cov}(Z_i, Z_j), \quad (2.7.16)$$

qui se ramène à (2.7.15) lorsque  $Z_i$  et  $Z_j$  sont non corrélés  $\forall 1 \leq i \neq j \leq n$ . On remarquera que cette dernière condition est satisfaite lorsque  $Z_i$  et  $Z_j$  sont indépendants  $\forall 1 \leq i \neq j \leq n$ . L'hypothèse de non-corrélation des couples de vecteurs  $Z_i$  et  $Z_j$  pour  $i \neq j$  est, bien entendu, plus faible que celle qui consiste à supposer que  $Z_1, \dots, Z_n$  sont indépendants deux à deux, et, *a priori*, dans leur ensemble.  $\square$



# Chapitre 3

## Lois normales.

### 3.1 Lois normales réelles.

#### 3.1.1 Propriétés Générales.

**Définition 3.1.1.** Une variable aléatoire réelle  $Y$  suit une loi normale standard, ou loi normale centrée réduite, ce qui sera noté symboliquement par  $Y \stackrel{d}{=} N(0, 1)$ , ou  $Y \stackrel{d}{=} N_1(0, 1)$ , si la densité de  $Y$  sur  $\mathbb{R}$  est donnée par

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}, \quad t \in \mathbb{R}. \quad (3.1.1)$$

On note habituellement la densité d'une loi  $N(0, 1)$  par  $\varphi(\cdot)$ , et sa fonction de répartition par  $\Phi(\cdot)$ , où

$$\Phi(y) = \mathbb{P}(Y \leq y) = \int_{-\infty}^y \varphi(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{1}{2}t^2} dt, \quad \forall y \in \mathbb{R}, \quad (3.1.2)$$

où  $\varphi$  est comme en (3.1.1). Ces notations ne doivent pas être confondues avec celles qui désignent les *fonctions caractéristiques*. On note habituellement  $\phi_X(u) = \mathbb{E}(\exp(iuX))$ , pour  $u \in \mathbb{R}$ , la fonction caractéristique d'une variable aléatoire réelle quelconque  $X$ . Pour les valeurs de  $z \in \mathbb{C}$  où elle est définie, on désigne habituellement par  $\psi_X(z) = \mathbb{E}(\exp(zX))$  la *fonction génératrice des moments* de  $X$ .

**Proposition 3.1.1.** La fonction caractéristique d'une variable aléatoire  $Y \stackrel{d}{=} N(0, 1)$ , suivant la loi normale  $N(0, 1)$  standard, est donnée par

$$\phi_Y(u) = \mathbb{E}(\exp(iuY)) = \exp\left(-\frac{1}{2}u^2\right), \quad u \in \mathbb{R}. \quad (3.1.3)$$

**Preuve.** On écrit que

$$\phi_Y(u) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(iuz - \frac{1}{2}z^2\right) dz = \exp\left(-\frac{1}{2}u^2\right) \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}-iu} \exp\left(-\frac{1}{2}z^2\right) dz\right). \quad (3.1.4)$$

Du fait que  $z \rightarrow \exp(-\frac{1}{2}z^2)$  est une fonction analytique de  $z \in \mathbb{C}$ , le théorème de Cauchy montre que

$$\int_{-r-iu}^{r-iu} + \int_{r-iu}^r + \int_r^{-r} + \int_{-r}^{-r-iu} \exp\left(-\frac{1}{2}z^2\right) dz = 0.$$

Si  $z = x + iy \in \mathbb{C}$ , avec  $x, y \in \mathbb{R}$ , on note  $Re(z) = x$  et  $Im(z) = y$ , respectivement, la partie réelle et la partie

imaginaire de  $z$ . La propriété ci-dessus, jointe au fait que  $|\exp(-\frac{1}{2}z^2)| = \exp(\operatorname{Re}(-\frac{1}{2}z^2))$ , permet de vérifier que

$$\begin{aligned} & \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}-iu} \exp\left(-\frac{1}{2}z^2\right) dz - 1 = \frac{1}{\sqrt{2\pi}} \left( \int_{\mathbb{R}-iu} - \int_{\mathbb{R}} \exp\left(-\frac{1}{2}z^2\right) dz \right) \\ & = \lim_{r \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \left( \int_{-r-iu}^{r-iu} + \int_r^{-r} \exp\left(-\frac{1}{2}z^2\right) dz \right) = \lim_{r \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \left( \int_r^{r-iu} + \int_{-r-iu}^{-r} \exp\left(-\frac{1}{2}z^2\right) dz \right) \\ & \leq \lim_{r \rightarrow \infty} \left( \frac{4|u|}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}r^2\right) \right) = 0. \end{aligned} \quad (3.1.5)$$

On obtient alors directement (3.1.3) à partir de (3.1.4) et (3.1.5).□

Rappelons les notations (1.2.35)–(1.2.36)–(1.2.37).

**Proposition 3.1.2.** *Les moments d'une variable  $Y \stackrel{d}{=} N(0, 1)$  sont donnés, pour tout entier  $k \in \mathbb{N}$ , par*

$$\mu_{2k+1} = M_{2k+1} = \mathbb{E}(Y^{2k+1}) = 0, \quad \mu_{2k} = M_{2k} = \mathbb{E}(Y^{2k}) = \frac{(2k)!}{2^k k!}, \quad (3.1.6)$$

$$\lambda_{2k+1} = 0, \quad \lambda_{2k} = \frac{(2k)!}{2^k k!}, \quad \beta_{2k+1} = 0, \quad \beta_{2k} = \frac{(2k+2)!}{2^{k+1} (k+1)!}. \quad (3.1.7)$$

En particulier,

$$\mu_1 = \mathbb{E}Y = 0, \quad M_2 = \sigma_Y^2 = \operatorname{Var}(Y) = 1, \quad M_3 = 0, \quad M_4 = 3, \quad \beta_1 = 0, \quad \beta_2 = 3. \quad (3.1.8)$$

**Preuve.** Par (3.1.1), et en effectuant le changement de variable  $t = \sqrt{2u}$ , on constate que le moment  $\mu_{2k} = \mathbb{E}(Y^{2k})$  d'ordre  $2k$  de  $Y$  vérifie les identités

$$\mu_{2k} = \frac{2}{\sqrt{2\pi}} \int_0^\infty t^{2k} e^{-\frac{1}{2}t^2} dt = \frac{2^k}{\sqrt{\pi}} \int_0^\infty u^{k-\frac{1}{2}} e^{-u} du = 2^k \frac{\Gamma(k+\frac{1}{2})}{\Gamma(\frac{1}{2})} = \frac{(2k)!}{2^k k!},$$

ce qui, compte tenu du fait que, par symétrie,  $\mu_{2k+1} = 0$ , donne bien (3.1.6), puis (3.1.8) pour  $k = 1, 2$ . Ici, on a fait usage des formules  $\Gamma(\frac{1}{2}) = \sqrt{\pi}$  et  $\Gamma(r) = (r-1)\Gamma(r-1)$  pour  $r > 1$ .□

**Remarque 3.1.1.** On peut utiliser la même démonstration pour montrer que les moments absolus  $\nu_r = \mathbb{E}|Y|^r$  d'ordre  $r > 0$  de  $Y$  sont donnés par (nous renvoyons à (4.2.1) pour la définition du symbole de Pochhammer  $(a)_p = \Gamma(a+p)/\Gamma(a)$ )

$$\nu_r = \mathbb{E}|Y|^r = 2^{r/2} \frac{\Gamma(\frac{r+1}{2})}{\Gamma(\frac{1}{2})} = 2^{r/2} \left(\frac{1}{2}\right)_{\frac{r}{2}}, \quad \text{pour } r \in \mathbb{R}^+. \quad (3.1.9)$$

On obtient, en particulier, que pour  $k \in \mathbb{N}$ ,

$$\nu_{2k+1} = \frac{2^{k+1} k!}{\sqrt{2\pi}},$$

alors que la relation (3.1.6) fournit

$$\nu_{2k} = \mu_{2k} = \frac{(2k)!}{2^k k!}.$$

On vérifie aisément, à l'aide de ces formules, que

$$\lim_{n \rightarrow \infty} \frac{\nu_{n+1}/(n+1)!}{\nu_n/n!} = 0.$$

Ceci implique que la série entière  $\sum_{n=0}^{\infty} z^n (\nu_n/n!)$  a un rayon de convergence infini. Cette propriété permet d'établir l'identité, pour tout  $z \in \mathbf{C}$

$$\mathbb{E}\left(\exp(zY)\right) = \mathbb{E}\left(\sum_{n=0}^{\infty} \frac{(zY)^n}{n!}\right) = \sum_{n=0}^{\infty} \frac{z^n}{n!} \mu_n = \sum_{k=0}^{\infty} \frac{z^{2k}}{(2k)!} \frac{(2k)!}{2^k k!} = \exp\left(\frac{1}{2}z^2\right). \quad (3.1.10)$$

On déduit de (3.1.10) l'existence,  $\forall z \in \mathbf{C}$ , de la fonction génératrice des moments de  $Y$ , définie par

$$\psi_Y(z) = \mathbb{E}(\exp(zY)) = \exp\left(\frac{1}{2}z^2\right). \quad (3.1.11)$$

La notation  $X \stackrel{d}{=} Y$  est utilisée pour exprimer que les deux variables aléatoires  $X$  et  $Y$  possèdent la même loi de probabilité.

**Définition 3.1.2.** Soient  $\mu \in \mathbb{R}$  et  $\sigma \in \mathbb{R}^+$ . On dira qu'un vecteur aléatoire réel  $X$  suit une loi normale  $N(\mu, \sigma^2)$ , ce qui sera noté symboliquement par  $Y \stackrel{d}{=} N(\mu, \sigma^2)$ , si  $X \stackrel{d}{=} \sigma Y + \mu$ , où  $Y \stackrel{d}{=} N(0, 1)$ .

**Proposition 3.1.3.** Soit une variable aléatoire  $X \stackrel{d}{=} N(\mu, \sigma^2)$ . Alors,  $\mu = \mathbb{E}X$  et  $\sigma^2 = \text{Var}(X)$ . De plus, la fonction caractéristique de  $X$  est donnée par

$$\phi_X(u) = \mathbb{E}(e^{iuX}) = \exp\left(iu\mu - \frac{1}{2}\sigma^2 u^2\right), \quad \text{pour } u \in \mathbf{C}, \quad (3.1.12)$$

et, lorsque  $\sigma > 0$ , la densité de  $X$  (relativement à la mesure de Lebesgue) existe, et est donnée par

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad \text{pour } x \in \mathbb{R}. \quad (3.1.13)$$

**Preuve.** Par la Définition 3.1.2, on a  $\phi_X(u) = \mathbb{E}(\exp(iuX)) = \mathbb{E}(\exp(iu(\sigma Y + \mu))) = \phi_Y(u\sigma) \exp(iu\mu)$ , ce qui, grâce à (3.1.3) donne directement (3.1.12).

Pour (3.1.13), il suffit de faire le changement de variable  $y = (x - \mu)/\sigma$  dans (3.1.1) pour obtenir

$$f_X(x) = f_Y(y) \left| \frac{dy}{dx} \right| = \sigma^{-1} f_Y\left(\frac{x-\mu}{\sigma}\right) = \sigma^{-1} \varphi\left(\frac{x-\mu}{\sigma}\right),$$

qui donne le résultat.  $\square$

**Proposition 3.1.4.** Soient  $X_1, \dots, X_n$  des variables aléatoires réelles indépendantes, de lois respectives  $N(\mu_1, \sigma_1^2), \dots, N(\mu_n, \sigma_n^2)$ , et soient des constantes réelles  $a_1, \dots, a_n$ . Alors  $S = \sum_{j=1}^n a_j X_j$  suit une loi normale de paramètres

$$S \stackrel{d}{=} N\left(\sum_{j=1}^n a_j \mu_j, \sum_{j=1}^n a_j^2 \sigma_j^2\right). \quad (3.1.14)$$

**Preuve.** Comme les variables aléatoires  $X_1, \dots, X_n$  sont indépendantes, la fonction caractéristique de leur somme  $S$  est égale à  $\phi_S(u) = \mathbb{E}(\exp(iuS)) = \prod_{j=1}^n \mathbb{E}(\exp(iua_j X_j))$ . Par (3.1.12), cette dernière expression vaut  $\prod_{j=1}^n \exp(iu\mu_j - \frac{1}{2}\sigma_j^2 a_j^2 u^2) = \exp(iu \sum_{j=1}^n \mu_j - \frac{1}{2} \sum_{j=1}^n a_j^2 \sigma_j^2 u^2)$ , ce qui permet de conclure.  $\square$

### 3.1.2 Quantiles de la Loi Normale Standard.

La table suivante fournit les valeurs numériques, à 0.001 près, des quantiles supérieures d'ordre  $\alpha$  de la loi normale standard  $N(0, 1)$ . Si  $Y \stackrel{d}{=} N(0, 1)$  désigne une variable aléatoire suivant une loi normale standard (ou centrée réduite), il est d'usage de noter  $\Phi(y) = \mathbb{P}(Y \leq y)$  la fonction de répartition correspondante, et, pour tout  $0 < \alpha < 1$ , de noter par  $\nu_\alpha$ , le quantile supérieur d'ordre  $\alpha$  de  $Y$  (ou  $\Phi$ ) défini (ici, de manière unique) par

$$\mathbb{P}(Y > \nu_\alpha) = 1 - \Phi(\nu_\alpha) = \alpha. \quad (3.1.15)$$

Lorsque  $0 < \alpha < 1$ , on a, de plus,

$$\mathbb{P}(|Y| \leq \nu_{\alpha/2}) = 1 - \alpha. \quad (3.1.16)$$

$\alpha$	$\nu_{\alpha}$
10%	1.282
5%	1.645
2.5%	1.960
1%	2.326
0.5%	2.576
0.25%	2.807
0.1%	3.090
0.05%	3.291

**Table 1. Quantiles supérieurs de la Loi Normale Standard.**

La valeur de  $\nu_{2.5\%} \simeq 1.960$ , voisine de 2, est particulièrement utile dans les applications. On a, en effet

$$\mathbb{P}(Y \in [-1.960, 1.960]) \simeq 95\% = 0.95.$$

## 3.2 Lois normales vectorielles

On considère, dans ce paragraphe, le cas de variables aléatoires à valeurs vectorielles dans l'espace  $\mathbb{R}^d = \mathcal{M}_{d,1}$  des matrices colonnes à  $d$  lignes. Tout vecteur de cet espace sera, par convention, identifié à la matrice colonne  $(d \times 1)$  de ses composantes dans la base canonique de  $\mathbb{R}^d$ .

Les propriétés suivantes des matrices symétriques seront utiles par la suite (voir le §2.3).

**Propriété 3.2.1.** Une matrice  $\Sigma$ , symétrique réelle,  $(d \times d)$ , est dite positive, ce qui est noté  $\Sigma \geq 0$ , si elle vérifie l'une des conditions équivalentes suivantes :

- (i)  $u' \Sigma u \geq 0$  pour tout  $u \in \mathbb{R}^d$  ;
- (ii) Il existe une matrice réelle  $(d \times d)$   $P$  telle que  $\Sigma = PP'$  ;
- (iii) Il existe une matrice réelle  $(d \times r)$   $Q$ , de rang  $r = \text{rg}(Q) = \text{rg}(\Sigma) \leq d$ , telle que  $\Sigma = QQ'$  ;
- (iv) Il existe une matrice orthogonale  $(d \times d)$   $H$  (c'est à dire, telle que  $H^{-1} = H'$ ), telle que

$$\Sigma = H \text{diag}(\lambda_1, \dots, \lambda_m) H' = H \text{diag}(\lambda_1, \dots, \lambda_m) H^{-1}, \text{ avec } \lambda_i \geq 0 \text{ pour } i = 1, \dots, d.$$

**Définition 3.2.1.** Un vecteur aléatoire réel  $Y \in \mathbb{R}^d$  suit une loi normale standard dans  $\mathbb{R}^d$ , ce qui est noté symboliquement par  $Y \stackrel{d}{=} N_d(\mathbb{O}_d, \mathbb{I}_d)$  ou  $Y \stackrel{d}{=} N_m(\mathbb{O}, \mathbb{I})$ , si les coordonnées  $Y_1, \dots, Y_d$  de  $Y = [Y_1 \ \dots \ Y_d]'$  sont des variables aléatoires indépendantes, de même loi  $N(0, 1)$ , normale centrée réduite.

**Proposition 3.2.1.** Soit une variable aléatoire  $Y \stackrel{d}{=} N_d(\mathbb{O}_d, \mathbb{I}_d)$ . Alors,  $\mathbb{E}(Y) = \mathbb{O}_d$  et  $\text{Var}(Y) = \mathbb{I}_d$ . De plus, la fonction caractéristique de  $Y$  est donnée par

$$\phi_Y(u) = \mathbb{E}(e^{iu'Y}) = \exp\left(-\frac{1}{2}u'u\right) \quad \text{pour } u \in \mathbb{R}^d, \quad (3.2.1)$$

et la densité de  $Y$  est égale à

$$f_Y(y) = (2\pi)^{-d/2} \exp\left(-\frac{1}{2}y'y\right), \quad \text{pour } y \in \mathbb{R}^d. \quad (3.2.2)$$

**Preuve.** Tout d'abord, on a, par (3.1.6) et l'indépendance de  $Y_1, \dots, Y_d$ ,

$$\mathbb{E}(Y) = \begin{bmatrix} \mathbb{E}(Y_1) \\ \vdots \\ \mathbb{E}(Y_d) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix},$$

et

$$\text{Var}(Y) = \mathbb{E}(YY') = \begin{bmatrix} \mathbb{E}(Y_1^2) & \dots & \mathbb{E}(Y_1 Y_d) \\ \vdots & \ddots & \vdots \\ \mathbb{E}(Y_d Y_1) & \dots & \mathbb{E}(Y_d^2) \end{bmatrix} = \begin{bmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{bmatrix} = \mathbb{I}_d.$$

L'indépendance de  $Y_1, \dots, Y_d$ , combinée avec (3.1.3), montre de même que, si  $u = [u_1 \ \dots \ u_d]'$ , alors

$$\begin{aligned} \phi_Y(u) &= \mathbb{E}\left(\exp(iu'Y)\right) = \mathbb{E}\left(\exp\left(i\sum_{j=1}^d u_j Y_j\right)\right) = \prod_{j=1}^d \mathbb{E}\left(\exp(iu_j Y_j)\right) \\ &= \exp\left(-\frac{1}{2}\sum_{j=1}^d u_j^2\right) = \exp\left(-\frac{1}{2}u'u\right), \end{aligned}$$

ce qui établit (3.2.1). Par indépendance de  $Y_1, \dots, Y_d$ , on obtient, en posant  $y = [y_1 \ \dots \ y_d]'$ , que la densité de  $Y$  est égale à

$$f_Y(y) = \prod_{j=1}^d f_{Y_j}(y_j) = \prod_{j=1}^d \varphi(y_j) = \prod_{j=1}^d \left((2\pi)^{-1/2} \exp\left(-\frac{1}{2}y_j^2\right)\right),$$

ce qui mène à (3.2.2).  $\square$

**Définition 3.2.2.** Soient  $\mu \in \mathbb{R}^d$ , soit  $\Sigma \geq 0$  une matrice  $(d \times d)$  positive, et soit  $Q$  une matrice  $(d \times d)$  telle que  $\Sigma = QQ'$ . On dit qu'un vecteur aléatoire réel  $X$  suit une loi normale  $N_d(\mu, \Sigma)$ , ce qui est noté symboliquement par  $X \stackrel{d}{=} N_d(\mu, \Sigma)$ , si  $X \stackrel{d}{=} QY + \mu$ , où  $Y \stackrel{d}{=} N_d(\mathbb{O}_d, \mathbb{I}_d)$ .

Pour vérifier que la définition 3.2.2 a un sens, il suffit de constater que la loi de  $QY + \mu$ , ainsi définie, est indépendante du choix de la matrice  $Q$  vérifiant  $\Sigma = QQ'$ . La proposition suivante établit cette propriété ainsi que d'autres résultats utiles.

**Proposition 3.2.2.** Soit une variable aléatoire  $X \stackrel{d}{=} N_d(\mu, \Sigma)$ . Alors,  $\mathbb{E}(X) = \mu$  et  $\text{Var}(X) = \Sigma$ . De plus, la fonction caractéristique de  $X$  est donnée par

$$\phi_X(u) = \mathbb{E}\left(e^{iu'X}\right) = \exp\left(iu'\mu - \frac{1}{2}u'\Sigma u\right) \quad \text{pour } u \in \mathbb{R}^d \quad (3.2.3)$$

Dans le cas où la matrice  $\Sigma > 0$  est définie positive, la densité de  $X$  existe, et est donnée par

$$f_X(x) = (2\pi)^{-d/2} (\det \Sigma)^{-1/2} \exp\left(-\frac{1}{2}(x - \mu)'\Sigma^{-1}(x - \mu)\right) \quad \text{pour } x \in \mathbb{R}^d. \quad (3.2.4)$$

**Preuve.** Comme  $\mathbb{E}(Y) = \mathbb{O}_d$  et  $\text{Var}(Y) = \mathbb{I}_d$ , par la définition 3.2.2, et la formule (2.7.14) de la proposition 2.7.3, on voit que  $\mathbb{E}(X) = Q\mathbb{E}(Y) + \mu = \mu$ , et que

$$\text{Var}(X) = \text{Var}(QY) = Q\text{Var}(Y)Q' = QQ' = \Sigma.$$

De même, par (3.2.1), on a les relations

$$\begin{aligned} \phi_X(u) &= \mathbb{E}\left(\exp(iu'X)\right) = \mathbb{E}\left(\exp(iu'(QY + \mu))\right) \\ &= \phi_Y(Q'u) \exp(iu\mu) = \exp\left(-\frac{1}{2}u'QQ'u\right) \exp(iu\mu), \end{aligned}$$

ce qui donne directement (3.2.3).

Pour établir (3.2.4), on constate tout d'abord que le fait que  $\Sigma = QQ'$  soit régulière implique que  $Q$  est  $(d \times d)$ , inversible, telle que  $\det(Q) = (\det \Sigma)^{1/2}$ , et vérifiant  $\Sigma^{-1} = \{Q'\}^{-1}Q^{-1}$ . Par le changement de variable



$y = Q^{-1}(x - \mu)$ , effectué dans (3.2.2), on en déduit que

$$\begin{aligned} f_X(x) &= f_Y(y) \left| \frac{dy}{dx} \right| = (\det Q)^{-1} f_Y(Q^{-1}(x - \mu)) \\ &= (\det \Sigma)^{-1/2} (2\pi)^{-d/2} \exp \left( -\frac{1}{2} (Q^{-1}(x - \mu))' (Q^{-1}(x - \mu)) \right), \end{aligned}$$

d'où la conclusion (3.2.4).  $\square$

**Proposition 3.2.3.** Soit  $X \stackrel{d}{=} N_d(\mu, \Sigma)$  un vecteur normal de  $\mathbb{R}^d$ . Alors, pour toute matrice  $(p \times d)$  constante (non aléatoire)  $P$  et pour tout vecteur constant  $\nu \in \mathbb{R}^p$ , on a

$$PX + \nu \stackrel{d}{=} N_p(P\mu + \nu, P\Sigma P'). \quad (3.2.5)$$

**Preuve.** Compte tenu de (3.2.3), il suffit d'écrire que

$$\mathbb{E} \left( \exp(iv'(PX + \nu)) \right) = \exp \left( iv'(P\mu + \nu) - \frac{1}{2} v' P \Sigma P' v \right),$$

pour conclure à la validité de (3.2.5).  $\square$

**Corollaire 3.2.1.** Si  $Y \stackrel{d}{=} N_d(\mathbb{0}_d, \mathbb{I}_d)$ , alors,  $HY \stackrel{d}{=} N_d(\mathbb{0}_d, \mathbb{I}_d)$  pour toute matrice  $H$ , orthogonale  $(d \times d)$ .

**Preuve.** Par (3.2.5), on a  $HY \stackrel{d}{=} N_d(0, HH')$ , d'où la conclusion, du fait que  $HH' = \mathbb{I}$  pour toute matrice orthogonale  $H$ .  $\square$

**Corollaire 3.2.2.** Soit  $\mathbb{R}^d = E_1 \oplus \dots \oplus E_k$  une décomposition de  $\mathbb{R}^d$  en somme directe de sous-espaces orthogonaux, tels que  $E_i \perp E_j$  pour tout  $1 \leq i \neq j \leq k$ , relativement au produit scalaire euclidien  $\langle u, v \rangle = u'v$  de  $\mathbb{R}^d$ . Soit  $Y \stackrel{d}{=} N_d(\mathbb{0}, \mathbb{I})$  un vecteur aléatoire suivant une loi normale standard dans  $\mathbb{R}^d$ , et soit  $Y = Y_1 + \dots + Y_k$  son unique décomposition en somme de vecteurs tels que, pour chaque  $i = 1, \dots, k$ ,  $Y_i = \mathcal{P}_{E_i}(Y) \in E_i$ , où  $\mathcal{P}_{E_i}$  désigne la projection orthonormale de  $\mathbb{R}^d$  sur  $E_i$ . Alors

1°) Les vecteurs aléatoires  $Y_1, \dots, Y_k$  sont indépendants ;

2°) Si, pour  $i = 1, \dots, k$ ,  $d_i = \dim(E_i)$  et  $e_{i,1}, \dots, e_{i,d_i}$  désigne une base orthonormée de  $E_i$ , alors, les coordonnées  $Y_{i,1}, \dots, Y_{i,d_i}$  de  $Y_i$  dans la base  $e_{i,1}, \dots, e_{i,d_i}$  sont des variables aléatoires indépendantes de même loi  $N(0, 1)$ .

**Preuve.** Il est possible de construire une base orthonormée de  $\mathbb{R}^d$ , composée des vecteurs  $e_{i,j} \in \mathbb{R}^d$ , pour  $1 \leq j \leq d_i$ , et  $1 \leq i \leq k$ , telle que, pour chaque choix de  $i = 1, \dots, k$ ,  $e_{i,1}, \dots, e_{i,d_i}$  constitue une base orthonormée de l'espace vectoriel  $E_i$ . Le changement de base, passant de la base canonique de  $\mathbb{R}^d$  à cette nouvelle base, ayant alors une matrice de passage orthogonale  $H$ , les coordonnées de  $Y$  dans la nouvelle base seront données par  $HY \stackrel{d}{=} N_d(\mathbb{0}_d, \mathbb{I}_d)$ , ceci, par application du corollaire 3.2.1. Ceci implique que ces coordonnées sont des v.a.r., mutuellement indépendantes et de même loi  $N(0, 1)$ . On en déduit à la fois le 1°) et le 2°).  $\square$

### 3.3 Dépendances mutuelles de vecteurs normaux.

Dans ce qui suit, nous considérons une suite  $X_1 \in \mathbb{R}^{d_1}, \dots, X_N \in \mathbb{R}^{d_N}$  de vecteurs aléatoires suivant une loi jointe normale. Il est commode d'exprimer ceci en écrivant que

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix} \stackrel{d}{=} N_d \left( \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_N \end{bmatrix}, \begin{bmatrix} \Sigma_{1,1} & \dots & \Sigma_{1,N} \\ \vdots & \ddots & \vdots \\ \Sigma_{N,1} & \dots & \Sigma_{N,N} \end{bmatrix} \right), \quad (3.3.1)$$

les entiers  $d_1 \geq 1, \dots, d_N \geq 1$  étant tels que  $d_1 + \dots + d_N = d$ , et

$$\begin{aligned} \mu_j &= \mathbb{E}(X_j) \in \mathbb{R}^{d_j}, \quad X_j \in \mathbb{R}^{d_j}, \quad X_j \stackrel{d}{=} N_{d_j}(\mu_j, \Sigma_{j,j}), \\ \Sigma_{j,j} &= \text{Var}(X_j) \in \mathcal{M}_{d_j, d_j}, \quad \Sigma_{j,\ell} = \text{Cov}(X_j, X_\ell) \in \mathcal{M}_{d_j, d_\ell}, \quad \forall 1 \leq j, \ell \leq N. \end{aligned}$$

**Proposition 3.3.1.** *Sous les hypothèses (3.3.1), une condition nécessaire et suffisante pour que les vecteurs aléatoires  $X_1, \dots, X_N$  soient indépendants est qu'ils soient non corrélés, i.e. que*

$$\Sigma_{j,\ell} = \text{Cov}(X_j, X_\ell) = \mathbb{O}_{d_j, d_\ell} \quad \forall 1 \leq j \neq \ell \leq N. \quad (3.3.2)$$

**Preuve.** L'indépendance de  $X_1, \dots, X_N$  a lieu si et seulement si on a l'identité,  $\forall u_1 \in \mathbb{R}^{d_1}, \dots, \forall u_N \in \mathbb{R}^{d_N}$ ,

$$\mathbb{E}\left(\exp\left\{i \sum_{j=1}^N u'_j X_j\right\}\right) = \prod_{j=1}^N \mathbb{E}\left(\exp\left\{i u'_j X_j\right\}\right).$$

Par (3.3.1), cette relation s'écrit

$$\exp\left(i \sum_{j=1}^N u'_j \mu_j - \frac{1}{2} \sum_{j=1}^N \sum_{\ell=1}^N u'_j \Sigma_{j,\ell} u_\ell\right) = \exp\left(i \sum_{j=1}^N u'_j \mu_j - \frac{1}{2} \sum_{j=1}^N u'_j \Sigma_{j,j} u_j\right),$$

qui équivaut donc bien à (3.3.1). On observera que *l'hypothèse que la loi jointe de  $X_1, \dots, X_N$  est normale est essentielle pour la validité de la démonstration.* En général, des vecteurs aléatoires non corrélés peuvent très bien être dépendants.□

Dans ce qui suit, nous considérons le cas particulier où  $N = 2$  dans (3.3.1), ce qui revient à poser

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \stackrel{d}{=} N_m \left( \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix} \right). \quad (3.3.3)$$

Dans ce qui précède, et par la suite, nous supposons que  $X_1 \in \mathbb{R}^p$  et  $X_2 \in \mathbb{R}^q$ , pour  $p \geq 1$  et  $q \geq 1$ . On pose  $m = p + q$ .

**Lemme 3.3.1.** *Sous l'hypothèse que  $\Sigma_{2,2}$  est définie positive, les vecteurs aléatoires  $X_1 - \Sigma_{1,2} \Sigma_{2,2}^{-1} X_2$  et  $X_2$  sont indépendants, de lois respectives  $N_p(\mu_1 - \Sigma_{1,2} \Sigma_{2,2}^{-1} \mu_2, \Sigma_{1,1} - \Sigma_{1,2} \Sigma_{2,2}^{-1} \Sigma_{2,1})$  et  $N_q(\mu_2, \Sigma_{2,2})$ . De plus, la loi conditionnelle  $\mathcal{L}(X_1 | X_2 = x_2)$  de  $X_1$  sachant  $X_2 = x_2$  est donnée par*

$$\mathcal{L}(X_1 | X_2 = x_2) \stackrel{d}{=} N_p\left(\mu_1 + \Sigma_{1,2} \Sigma_{2,2}^{-1} (x_2 - \mu_2), \Sigma_{1,1} - \Sigma_{1,2} \Sigma_{2,2}^{-1} \Sigma_{2,1}\right). \quad (3.3.4)$$

**Preuve.** On constate, tout d'abord, que

$$\text{Cov}\left(X_1 - \Sigma_{1,2} \Sigma_{2,2}^{-1} X_2, X_2\right) = \Sigma_{1,2} - \Sigma_{1,2} \Sigma_{2,2}^{-1} \Sigma_{2,2} = \mathbb{O}_{p,q},$$

ce qui établit l'indépendance de  $X_1 - \Sigma_{1,2} \Sigma_{2,2}^{-1} X_2$  et  $X_2$ . Par hypothèse,  $X_2 \equiv N_q(\mu_2, \Sigma_{2,2})$ . Enfin,

$$\mathbb{E}(X_1 - \Sigma_{1,2} \Sigma_{2,2}^{-1} X_2) = \mu_1 - \Sigma_{1,2} \Sigma_{2,2}^{-1} \mu_2,$$

et

$$\begin{aligned} \text{Var}(X_1 - \Sigma_{1,2} \Sigma_{2,2}^{-1} X_2) &= \Sigma_{1,1} + \Sigma_{1,2} \Sigma_{2,2}^{-1} \Sigma_{2,1} - \Sigma_{1,2} \Sigma_{2,2}^{-1} \Sigma_{2,1} - \Sigma_{1,2} \Sigma_{2,2}^{-1} \Sigma_{2,1} \\ &= \Sigma_{1,1} - \Sigma_{1,2} \Sigma_{2,2}^{-1} \Sigma_{2,1}, \end{aligned}$$

ce qui permet d'établir la première partie du lemme. Pour montrer (3.3.4), on observe que

$$\begin{aligned} \mathcal{L}\left(X_1 - \Sigma_{1,2} \Sigma_{2,2}^{-1} x_2 \mid X_2 = x_2\right) &= \mathcal{L}\left(X_1 - \Sigma_{1,2} \Sigma_{2,2}^{-1} X_2 \mid X_2 = x_2\right) = \mathcal{L}\left(X_1 - \Sigma_{1,2} \Sigma_{2,2}^{-1} X_2\right) \\ &= N_p\left(\mu_1 - \Sigma_{1,2} \Sigma_{2,2}^{-1} \mu_2, \Sigma_{1,1} - \Sigma_{1,2} \Sigma_{2,2}^{-1} \Sigma_{2,1}\right), \end{aligned}$$

ce qui permet de conclure à la validité de la relation (3.3.4).□

### 3.4 Lois normales matricielles

L'introduction de lois normales matricielles est particulièrement utile pour traiter des échantillons de vecteurs aléatoires normaux dans  $\mathbb{R}^d$ . A cet effet, il conviendra de garder à l'esprit le cas particulier important suivant. Dans le cas où l'on observe un échantillon de vecteurs  $X_1, \dots, X_n \in \mathbb{R}^d$ , il est commode de ranger l'ensemble de ces données sous la forme de la matrice ( $n \times d$ )

$$\mathbb{X} = \begin{bmatrix} X'_1 \\ \vdots \\ X'_n \end{bmatrix} \in \mathcal{M}_{n,d} \quad \Leftrightarrow \quad \mathbb{X}' = [X_1 \quad \dots \quad X_n] \in \mathcal{M}_{d,n}. \quad (3.4.1)$$

La formulation (3.4.1) présente, entre autres, l'avantage de permettre des écritures matricielles commodes, comme

$$\mathbb{X}'\mathbb{X} = \sum_{i=1}^n X_i X'_i \in \mathcal{M}_{d,d}. \quad (3.4.2)$$

Il conviendra donc de s'accoutumer au fait que, dans un tel formalisme, la *taille*  $n$  de l'échantillon correspond au *nombre de lignes* de  $\mathbb{X}$ , tandis que la *dimension*  $d$  de l'espace où les observations prennent leurs valeurs correspond au *nombre de colonnes* de  $\mathbb{X}$ . L'opération  $\text{Vec}(\cdot)$  permet de passer de la forme matricielle à la forme vectorielle de ces objets aléatoires. Compte tenu de l'application considérée, il convient de l'appliquer non pas à  $\mathbb{X}$ , mais à sa transposée  $\mathbb{X}'$ , du fait que, avec les notations (3.4.1),

$$\text{Vec}(\mathbb{X}') = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \in \mathbb{R}^{dn}. \quad (3.4.3)$$

Dans le cas d'un échantillon  $X_1, \dots, X_n$  issu de la loi normale  $N_d(\mu, \Sigma)$ , les moments de  $\text{Vec}(\mathbb{X}')$  s'expriment simplement, puisque

$$\mathbb{E}(\text{Vec}(\mathbb{X}')) = \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix} = \mathbb{1}_n \otimes \mu \quad \text{et} \quad \text{Var}(\text{Vec}(\mathbb{X}')) = \begin{bmatrix} \Sigma & \dots & \mathbb{O}_d \\ \vdots & \ddots & \vdots \\ \mathbb{O}_d & \dots & \Sigma \end{bmatrix} = \mathbb{I}_n \otimes \Sigma. \quad (3.4.4)$$

On peut donc écrire, compte tenu de (3.4.4), que

$$\text{Vec}(\mathbb{X}') \stackrel{d}{=} N_{dn}(\mathbb{1}_n \otimes \mu, \mathbb{I}_n \otimes \Sigma). \quad (3.4.5)$$

Nous considérons donc ici, dans un cadre plus général que celui de l'échantillon, des *matrices aléatoires*  $\mathbb{X}$  à valeurs dans l'espace  $\mathcal{M}_{n,d}$  des matrices ( $n \times d$ ). Compte tenu de (3.4.3), nous dirons que  $\mathbb{X}$  suit une *loi normale matricielle* dans  $\mathcal{M}_{n,d}$  si et seulement si  $\text{Vec}(\mathbb{X}')$  suit une loi normale vectorielle dans  $\mathbb{R}^{dn}$ . Dans la suite, il sera commode d'adopter le formalisme suivant.

**Définition 3.4.1.** Soit  $M \in \mathcal{M}_{n,d}$  une matrice réelle ( $n \times d$ ) (à  $n$  lignes et  $m$  colonnes), et  $S$  une matrice symétrique positive ( $dn \times dn$ ). On dit que la matrice aléatoire ( $n \times d$ ),  $\mathbb{X}$ , suit la loi normale matricielle  $N_{n,d}(M, S)$ , ce qu'on note  $\mathbb{X} \stackrel{d}{=} N_{n,d}(M, S)$  si  $\text{Vec}(\mathbb{X}') \stackrel{d}{=} N_{dn}(\text{Vec}(M'), S)$ . En résumé,

$$\mathbb{X} \stackrel{d}{=} N_{n,d}(M, S), \quad \mathbb{X} \in \mathcal{M}_{n,d} \quad \Leftrightarrow \quad \text{Vec}(\mathbb{X}') \stackrel{d}{=} N_{dn}(\text{Vec}(M'), S). \quad (3.4.6)$$

Sous ces hypothèses, on dit que  $\mathbb{X}$  est une matrice aléatoire gaussienne à valeurs dans  $\mathcal{M}_{n,d}$

On remarquera que, dans cette définition,  $M = \mathbb{E}(\mathbb{X})$  et  $S = \text{Var}(\text{Vec}(\mathbb{X}'))$ . Dans la suite, compte tenu de (3.4.5), nous nous limiterons essentiellement au cas où cette dernière matrice est de la forme  $S = C \otimes D$ ,  $C \geq 0$  et  $D \geq 0$  étant des matrices symétriques positives. Les exemples qui suivent illustrent les cas particuliers les plus importants où cette propriété est vérifiée.

**Exemple 3.4.1.** Nous reprenons le cas, décrit dans (3.4.4) et (3.4.5), où  $\mathbb{X}' = [X_1 \ \dots \ X_n]$ , dans le cas où les  $X_1, \dots, X_n$  forment une suite de vecteurs aléatoires indépendants de même loi  $N_d(\mu, \Sigma)$ . Alors,

$$M = \mathbb{E}(\mathbb{X}) = \mathbb{1}_n \otimes \mu' = \mathbb{1}_n \mu', \quad \mathbb{E}(\text{Vec}(\mathbb{X}')) = \text{Vec}(M') = \mathbb{1}_n \otimes \mu, \quad \text{et} \quad \text{Var}(\text{Vec}(\mathbb{X}')) = \mathbb{I}_n \otimes \Sigma.$$

Compte tenu de (3.4.5), on en déduit la version correspondante de (3.4.6), soit

$$\mathbb{X} = \begin{bmatrix} X'_1 \\ \vdots \\ X'_n \end{bmatrix} \stackrel{d}{=} N_{n,d}(\mathbb{1}_n \mu', \mathbb{I}_n \otimes \Sigma) \quad \Leftrightarrow \quad \text{Vec}(\mathbb{X}') = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \stackrel{d}{=} N_{dn}(\mathbb{1}_n \otimes \mu, \mathbb{I}_n \otimes \Sigma). \quad (3.4.7)$$

Plus généralement, si  $X_1 \stackrel{d}{=} N_d(\mu_1, \Sigma), \dots, X_n \stackrel{d}{=} N_d(\mu_n, \Sigma)$ , et si

$$\mathbb{X} = \begin{bmatrix} X'_1 \\ \vdots \\ X'_n \end{bmatrix} \quad \text{et} \quad M = \mathbb{E}(\mathbb{X}) = \begin{bmatrix} \mu'_1 \\ \vdots \\ \mu'_n \end{bmatrix}, \quad (3.4.8)$$

on a

$$\mathbb{X} = \begin{bmatrix} X'_1 \\ \vdots \\ X'_n \end{bmatrix} \stackrel{d}{=} N_{n,d}(M, \mathbb{I}_n \otimes \Sigma) \quad \Leftrightarrow \quad \text{Vec}(\mathbb{X}') = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \stackrel{d}{=} N_{dn}(\text{Vec}(M'), \mathbb{I}_n \otimes \Sigma). \quad (3.4.9)$$

**Exemple 3.4.2.** Nous considérons maintenant un cas sensiblement généralisé par rapport à celui de l'exemple précédent. Partant, comme dans ce dernier, de  $\mathbb{X}' = [X_1 \ \dots \ X_n] \in \mathcal{M}_{d,n}$ , où  $X_1, \dots, X_n$  composent une suite de vecteurs aléatoires indépendants de même loi  $N_m(\mu, \Sigma)$ , nous posons

$$\mathbb{Y}' = P\mathbb{X}'Q = [PX_1 \ \dots \ PX_n]Q \in \mathcal{M}_{r,s} \quad \text{et} \quad \mathbb{Y} = Q'\mathbb{X}P' \in \mathcal{M}_{s,r}, \quad (3.4.10)$$

où  $P \in \mathcal{M}_{r,d}$  est une matrice ( $r \times d$ ) et  $Q \in \mathcal{M}_{n,s}$  une matrice ( $n \times s$ ). Par (2.5.3), en supposant que les matrices  $P, Q$  soient constantes (non aléatoires), on obtient que

$$\text{Vec}(\mathbb{Y}') = \text{Vec}(P\mathbb{X}'Q) = (Q' \otimes P)\text{Vec}(\mathbb{X}'),$$

et donc, en posant  $M = \mathbb{E}(\mathbb{X})$ , et en écrivant (comme dans (3.4.4)) que  $\mathbb{E}(\text{Vec}(\mathbb{X}')) = \text{Vec}(M') = \mathbb{1}_n \otimes \mu$ , on obtient, compte tenu de (3.4.4), et par une application du (iii) de la propriété 1.1, que

$$\begin{aligned} \mathbb{E}(\text{Vec}(\mathbb{Y}')) &= \mathbb{E}(\text{Vec}(P\mathbb{X}'Q)) = \mathbb{E}((Q' \otimes P)\text{Vec}(\mathbb{X}')) = (Q' \otimes P)\text{Vec}(M') \\ &= (Q' \otimes P)(\mathbb{1}_n \otimes \mu) = Q' \mathbb{1}_n \otimes P\mu. \end{aligned}$$

De même, compte tenu de (3.4.4), en faisant usage des (iv) et (x) de la propriété 1.1, on obtient que

$$\begin{aligned} \text{Var}(\text{Vec}(\mathbb{Y}')) &= \text{Var}((Q' \otimes P)\text{Vec}(\mathbb{X}')) = (Q' \otimes P)\text{Var}(\text{Vec}(\mathbb{X}'))(Q' \otimes P)' \\ &= (Q' \otimes P)(\mathbb{I}_n \otimes \Sigma)(Q \otimes P') = Q'Q \otimes P\Sigma P'. \end{aligned}$$

Comme, par (3.4.4),  $\mathbb{E}(\mathbb{X}) = \mathbb{1}_n \otimes \mu' = \mathbb{1}_n \mu'$ , de toute évidence,

$$\mathbb{E}(\mathbb{Y}) = \mathbb{E}(Q'\mathbb{X}P') = Q'(\mathbb{1}_n \mu')P' = Q' \mathbb{1}_n (P\mu)',$$

on en déduit que

$$\mathbb{Y} = Q'\mathbb{X}P' \stackrel{d}{=} N_{s,r}(Q' \mathbb{1}_n (P\mu)', Q'Q \otimes P\Sigma P'), \quad (3.4.11)$$

et

$$\text{Vec}(\mathbb{Y}') \stackrel{d}{=} N_{sr}(Q' \mathbb{1}_n \otimes P\mu, Q'Q \otimes P\Sigma P'). \quad (3.4.12)$$

On remarquera ici que les matrices  $Q'Q$  et  $P\Sigma P'$  sont toujours positives. Ceci implique, par la propriété 2.5.1 (viii), que  $Q'Q \otimes P\Sigma P'$  est positive.

**Exemple 3.4.3.** Nous considérons maintenant le cas particulier de l'exemple 3.4.2, obtenu en choisissant  $P = \mathbb{I}_d$  et  $Q \in \mathcal{M}_{n,s}$  dans (3.4.10). Dans ce cas,  $\mathbb{Y} = Q'\mathbb{X}$  avec  $\mathbb{X}' = [X_1 \ \dots \ X_n] \in \mathcal{M}_{n,d}$ , où  $X_1, \dots, X_n$  désignent des vecteurs aléatoires indépendants de lois respectives  $N_d(\mu_1, \Sigma), \dots, N_d(\mu_n, \Sigma)$ . Posons  $M' = \mathbb{E}(\mathbb{X}') = [\mu_1 \ \dots \ \mu_n]$ . Soit  $Q \in \mathcal{M}_{n,s}$  une matrice  $(n \times s)$  constante (non aléatoire), et soit la matrice  $(s \times s)$  constante  $C = Q'Q$ . Alors, par (3.4.11),

$$\mathbb{Y} = Q'\mathbb{X} \stackrel{d}{=} N_{s,d}(Q'M, C \otimes \Sigma). \quad (3.4.13)$$

Nous revenons maintenant au cas général où  $X \in \mathcal{M}_{n,d}$  est comme en (3.4.6). Il se trouve que la loi de la matrice aléatoire normale  $\mathbb{X} \stackrel{d}{=} N_{n,d}(M, S)$  s'exprime alors simplement lorsque  $S$  est de la forme  $S = C \otimes D$ , comme le montre le théorème suivant. On retiendra de l'exemple 3.4.3 que si  $\mathbb{Z} \stackrel{d}{=} N_{r,d}(R, \mathbb{I}_n \otimes \Sigma)$  et  $Q \in \mathcal{M}_{r,n}$  est une matrice constante, alors  $\mathbb{X} = Q'\mathbb{Z} \stackrel{d}{=} N_{n,d}(Q'R, (Q'Q) \otimes \Sigma)$  est bien de la forme  $N_{n,d}(M, C \otimes D)$ , avec  $M = Q'R$ ,  $C = Q'Q$  et  $D = \Sigma$ .

**Théorème 3.4.1.** Soit  $\mathbb{X} \stackrel{d}{=} N_{n,d}(M, C \otimes D)$  une matrice normale aléatoire  $(n \times d)$ , où  $C > 0$  et  $D > 0$  sont des matrices définies positives, de dimensions respectives  $(n \times n)$  et  $(d \times md)$ . Alors,  $\mathbb{X}$  a une densité relativement à la mesure de Lebesgue sur  $\mathcal{M}_{n,d} \leftrightarrow \mathbb{R}^{dn}$  donnée par

$$\begin{aligned} f(\mathbb{X}) &= (2\pi)^{-dn/2} (\det C)^{-d/2} (\det D)^{-n/2} \operatorname{etr} \left( -\frac{1}{2} C^{-1} \{\mathbb{X} - M\} D^{-1} \{\mathbb{X} - M\}' \right) \\ &= (2\pi)^{-dn/2} (\det C)^{-m/2} (\det D)^{-n/2} \operatorname{etr} \left( -\frac{1}{2} D^{-1} \{\mathbb{X} - M\}' C^{-1} \{\mathbb{X} - M\} \right), \end{aligned} \quad (3.4.14)$$

où  $\mathbb{X}$  varie dans l'espace  $\mathcal{M}_{n,d}$  des matrices réelles  $(n \times d)$ .

**Preuve.** Posons  $Z = \operatorname{Vec}(\mathbb{X}')$  et  $\mu = \operatorname{Vec} M'$ . Par (3.4.6), l'hypothèse que  $\mathbb{X} \stackrel{d}{=} N_{n,d}(M, C \otimes D)$  équivaut à ce que  $Z = \operatorname{Vec}(\mathbb{X}') \stackrel{d}{=} N_{dn}(\operatorname{vec}(M'), C \otimes D)$ . Comme (voir la propriété 2.5.1 (viii))  $C > 0$  et  $D > 0 \Rightarrow C \otimes D > 0$ , ceci implique que la densité de  $Z$  dans  $\mathbb{R}^{dn}$  est donnée par

$$\begin{aligned} g(Z) &= (2\pi)^{-dn/2} (\det C)^{-d/2} (\det D)^{-n/2} \exp \left( -\frac{1}{2} (Z - \mu)' (C \otimes D)^{-1} (Z - \mu) \right) \\ &= (2\pi)^{-dn/2} (\det C)^{-d/2} (\det D)^{-n/2} \\ &\quad \times \exp \left( -\frac{1}{2} \left( \operatorname{Vec}(\{\mathbb{X} - M\}') \right)' (C \otimes D)^{-1} \left( \operatorname{Vec}(\{\mathbb{X} - M\}') \right) \right), \end{aligned}$$

en posant  $\mu = \operatorname{Vec} M'$ , et faisant usage de la formule  $\det(C \otimes D) = (\det C)^d (\det D)^n$  (propriété 2.5.1 (vii)). Nous écrivons maintenant l'identité du lemme 2.5.2 (iii) sous la forme particulière

$$\operatorname{tr}(\mathbf{B}\mathbf{X}'\mathbf{C}\mathbf{X}) = (\operatorname{Vec}(\mathbf{X}))' (\mathbf{B}' \otimes \mathbf{C}) \operatorname{Vec}(\mathbf{X}). \quad (3.4.15)$$

Comme, par la propriété 2.5.1 (v),  $(C \otimes D)^{-1} = C^{-1} \otimes D^{-1}$ , l'observation que les matrices  $C^{-1}$  et  $D^{-1}$  sont définies positives, permet d'appliquer (3.4.15) avec  $\mathbf{X} = \{\mathbb{X} - M\}'$ ,  $\mathbf{B} = \mathbf{B}' = C^{-1}$  et  $\mathbf{C} = D^{-1}$ , pour obtenir l'identité

$$\begin{aligned} (Z - \mu)' (C \otimes D)^{-1} (Z - \mu) &= \left( \operatorname{Vec}(\{\mathbb{X} - M\}') \right)' (C^{-1} \otimes D^{-1}) \left( \operatorname{Vec}(\{\mathbb{X} - M\}') \right) \\ &= \operatorname{tr} \left( C^{-1} \{\mathbb{X} - M\} D^{-1} \{\mathbb{X} - M\}' \right), \end{aligned}$$

ce qui permet d'établir la première partie de (3.4.14). La seconde partie de (3.4.14) se déduit de la première en appliquant l'identité  $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ , avec  $A = C^{-1} \{\mathbb{X} - M\}$  et  $B = D^{-1} \{\mathbb{X} - M\}'$ .  $\square$

**Remarque 3.4.1.** Considérons à nouveau le cas particulier où

$$\mathbb{X} = [X_1 \ \dots \ X_n]' \stackrel{d}{=} N_{n,d}(\mathbb{1}_n \otimes \mu', \mathbb{I}_n \otimes \Sigma),$$

ce qui correspond (voir l'exemple 3.4.1) à des vecteurs aléatoires  $X_1, \dots, X_n$  indépendants de même loi  $N_d(\mu, \Sigma)$ . Une démonstration plus directe de la version correspondante de (3.4.14) est alors possible en constatant que la densité jointe de  $X_1, \dots, X_n$  est donnée par

$$\begin{aligned} f(\mathbb{X}) &= (2\pi)^{-dn/2} (\det \Sigma)^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^n (X_i - \mu)' \Sigma^{-1} (X_i - \mu)\right) \\ &= (2\pi)^{-dn/2} (\det \Sigma)^{-n/2} \text{etr}\left(-\frac{1}{2} \sum_{i=1}^n (X_i - \mu)' \Sigma^{-1} (X_i - \mu)\right) \\ &= (2\pi)^{-dn/2} (\det \Sigma)^{-n/2} \text{etr}\left(-\frac{1}{2} \Sigma^{-1} \sum_{i=1}^n (X_i - \mu)(X_i - \mu)'\right) \\ &= (2\pi)^{-dn/2} (\det \Sigma)^{-n/2} \text{etr}\left(-\frac{1}{2} \Sigma^{-1} (\mathbb{X} - \mathbb{1}'_n \mu)' (\mathbb{X} - \mathbb{1}'_n \mu)\right), \end{aligned}$$

ce qui s'identifie au cas particulier de (3.4.14) obtenu pour  $C = \mathbb{I}$ ,  $D = \Sigma$  et  $M = \mathbb{1}'_n \mu$ .



# Chapitre 4

## Lois du $\chi^2$ et de Fisher

### 4.1 Lois du $\chi^2$ Centrées

#### 4.1.1 Propriétés Générales.

Nous commençons par effectuer quelques rappels concernant les fonctions gamma et béta d'Euler (dues au mathématicien suisse Leonhard Euler (1707-1783)) univariées, qui sont respectivement définies par

$$\Gamma(r) = \Gamma_1(r) = \int_0^\infty t^{r-1} e^{-t} dt \quad \text{et} \quad \beta(r, s) = \beta_1(r, s) = \int_0^1 t^{r-1} (1-t)^{s-1} dt,$$

pour  $Re(r) > 0$  et  $Re(s) > 0$ .

Il y a, naturellement, d'autres moyens de définir ces fonctions, mais ces derniers ne seront pas utiles dans le contexte présent. On se référera, par exemple, au Ch. 6 de la référence ci-dessous (voir pp. 255-293), pour davantage de détails à ce sujet.

Abramowitz, M. et Stegun; I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York.

**Propriété 4.1.1.** *Les fonctions gamma et béta satisfont les identités suivantes.*

$$(i) \quad \Gamma(n) = (n-1)! \quad \text{pour} \quad n \in \mathbb{N}^*; \tag{4.1.1}$$

$$(ii) \quad \Gamma(r) = (r-1)\Gamma(r-1) \quad \text{pour} \quad Re(r) > 1; \tag{4.1.2}$$

$$(iii) \quad \Gamma(r)\Gamma(s) = \Gamma(r+s)\beta(r, s) \quad \text{pour} \quad Re(r) > 0 \quad \text{et} \quad Re(s) > 0; \tag{4.1.3}$$

$$(iv) \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}. \tag{4.1.4}$$

La relation (4.1.2) permet de prolonger la définition de la fonction  $\Gamma(r)$  pour des valeurs de  $r \in \mathbb{C}$  telles que  $Re(r) \notin \{0, -1, -2, \dots\}$ . Nous serons ici principalement intéressés par la définition de la fonction gamma sur  $\mathbb{R}^+$ . Celle-ci permet de définir la *loi gamma* comme suit.

**Définition 4.1.1.** *Soient  $r > 0$  et  $\lambda > 0$ . On dit qu'une variable aléatoire  $X$  suit une loi  $\Gamma(r, \lambda)$ , ce qui est noté symboliquement par  $X \stackrel{d}{=} \Gamma(r, \lambda)$ , si l'une des propriétés équivalentes suivantes est vérifiée.*

(i) *La densité de  $X$  est donnée par*

$$f_X(x) = \begin{cases} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, & \text{pour } x > 0, \\ 0 & \text{pour } x \leq 0. \end{cases} \tag{4.1.5}$$



(ii) La fonction caractéristique de  $X$  est donnée par

$$\phi_X(u) = \left(1 - \frac{iu}{\lambda}\right)^{-r}, \quad \text{pour } u \in \mathbb{R}. \quad (4.1.6)$$

**Proposition 4.1.1.** Les conditions (i) et (ii) de la définition 4.1.1 de la loi  $\Gamma(r, \lambda)$  sont équivalentes et définissent une loi de probabilité. De plus, on a les propriétés suivantes.

(i)  $X \stackrel{d}{=} \Gamma(r, \lambda) \iff \lambda X \stackrel{d}{=} \Gamma(r, 1)$ ;

(ii) Pour tout  $s \in \mathbb{R}^+$ , le moment d'ordre  $s$  de  $X \stackrel{d}{=} \Gamma(r, \lambda)$  est égal à

$$\mu_s = \mathbb{E}(X^s) = \lambda^{-s} \left\{ \frac{\Gamma(r+s)}{\Gamma(r)} \right\} = \lambda^{-s} (r)_s. \quad (4.1.7)$$

En particulier,

$$\mu_n = \lambda^{-n} (r)_n = r(r+1) \dots (r+n-1) \quad \text{pour } n \in \mathbb{N}, \quad \mathbb{E}(X) = \frac{r}{\lambda} \quad \text{et} \quad \text{Var}(X) = \frac{r}{\lambda^2}. \quad (4.1.8)$$

(iii) Si  $X_1 \stackrel{d}{=} \Gamma(r_1, \lambda)$  et  $X_2 \stackrel{d}{=} \Gamma(r_2, \lambda)$  sont indépendantes, alors  $X_1 + X_2 \stackrel{d}{=} \Gamma(r_1 + r_2, \lambda)$ .

**Preuve.** Nous renvoyons à (4.2.1), au paragraphe 4.2, pour une définition générale du *symbole de Pochhammer*  $(r)_n$ , utilisé dans (4.1.7) et (4.1.8). Nous nous limiterons ici à la formule, pour  $r > 0$ ,

$$(r)_n = r(r+1) \dots (r+n-1) = \frac{\Gamma(r+n)}{\Gamma(r)}.$$

Si  $X$  suit une loi  $\Gamma(r, \lambda)$ , la densité de  $Z = \lambda X$ , obtenue par le changement de variable  $z = \lambda x$ , est donnée par

$$\begin{aligned} f_Z(z) &= f_X(x) \left| \frac{dx}{dz} \right| = f_X(\lambda^{-1}z) \lambda^{-1} \\ &= \frac{\lambda^r}{\Gamma(r)} (\lambda^{-1}z)^{r-1} e^{-z} \lambda^{-1} = \frac{1}{\Gamma(r)} z^{r-1} e^{-z} \quad \text{pour } z > 0, \end{aligned}$$

d'où le (i). En faisant usage de cette propriété, nous pouvons vérifier le calcul de la fonction caractéristique, énoncé dans la définition 4.1.1, en nous limitant, sans perte de généralité, au cas où  $\lambda = 1$ . Nous nous servons alors de l'analyticité de la fonction  $z \rightarrow z^{r-1} e^{-z}$  dans le domaine  $\mathbb{C} - \{0\}$ , pour constater, à l'aide du théorème de Cauchy, que, pour tout choix de  $R > 0$  et de  $\theta \in (-\pi/2, \pi/2)$ , on a l'identité

$$\int_{z \in R \times [0,1]} z^{r-1} e^{-z} dz - \int_{z \in Re^{i\theta} \times [0,1]} z^{r-1} e^{-z} dz + \int_{z \in \{Re^{i\alpha} : \alpha \in [0, \theta]\}} z^{r-1} e^{-z} dz = 0.$$

Comme, lorsque  $R \rightarrow \infty$ , quelque soit  $\theta \in (-\pi/2, \pi/2)$  fixé,

$$\begin{aligned} \left| \int_{z \in \{Re^{i\alpha} : \alpha \in [0, \theta]\}} z^{r-1} e^{-z} dz \right| &\leq R^{r-1} \exp(-R \cos \theta) \left| \int_{z \in \{Re^{i\alpha} : \alpha \in [0, \theta]\}} dz \right| \\ &\leq (\pi R) R^{r-1} \exp(-R \cos \theta) \rightarrow 0, \end{aligned}$$

on en déduit que

$$\begin{aligned} \lim_{R \rightarrow \infty} \int_{z \in R \times [0,1]} z^{r-1} e^{-z} dz &= \int_{z \in \mathbb{R}^+} z^{r-1} e^{-z} dz \\ &= \lim_{R \rightarrow \infty} \int_{z \in \{Re^{i\theta} \times [0,1]\}} z^{r-1} e^{-z} dz = \int_{z \in e^{i\theta} \times \mathbb{R}^+} z^{r-1} e^{-z} dz. \end{aligned}$$

On constate ensuite que, pour tout  $u \in \mathbb{R}$ ,  $\theta = \arg(1 - iu) \in (-\pi/2, \pi/2) [2\pi]$ . De la sorte, le changement de variable  $z \rightarrow x$ , défini par  $z = (1 - iu)x$ , permet d'obtenir, à l'aide de l'identité ci-dessus, avec  $\theta = \arg(1 - iu)$ ,

$$\begin{aligned}\phi_X(u) &= \frac{1}{\Gamma(r)} \int_0^\infty x^{r-1} e^{-(1-iu)x} dx = (1 - iu)^{-r} \frac{1}{\Gamma(r)} \int_{z \in e^{i\theta} \times \mathbb{R}^+} z^{r-1} e^{-z} dz \\ &= (1 - iu)^{-r} \frac{1}{\Gamma(r)} \int_{z \in \mathbb{R}^+} z^{r-1} e^{-z} dz = (1 - iu)^{-r},\end{aligned}$$

ce qui permet de conclure à la validité de la formule (4.1.6).

Pour établir le (ii), on observe tout d'abord que l'égalité  $\mathbb{E}(X^s) = \lambda^{-s} \Gamma(r+s)/\Gamma(s)$  est triviale, par la définition même de la fonction gamma. On obtient ensuite  $\mathbb{E}(X^n) = \lambda^{-n} (r)_n$  par (4.2.2) lorsque  $n \in \mathbb{N}$ . En appliquant cette formule pour  $n = 1$  et  $n = 2$ , on achève la démonstration de (4.1.8).

Enfin le (iii) est une conséquence directe de la relation (4.1.6), qui implique que  $\phi_{X_1} \phi_{X_2} = \phi_{X_1+X_2}$ .  $\square$

Nous sommes maintenant en mesure de donner la définition des lois du  $\chi^2$  centrées, en nous appuyant sur la définition 4.1.1 et la proposition 4.1.1 ci-dessus.

**Définition 4.1.2.** Une variable aléatoire  $Z$  suit une loi du  $\chi_n^2$ , dite loi du Khi-deux centrée à  $n$  degrés de liberté, ce qui est noté par  $Z \stackrel{d}{=} \chi_n^2$ , ou  $Z \stackrel{d}{=} \chi_n^2(0)$ , si l'une des propriétés équivalentes suivantes est vérifiée.

(i)  $Z \stackrel{d}{=} \Gamma(\frac{n}{2}, \frac{1}{2})$ ;

(ii) La densité de  $Z$  est donnée par

$$g_n(z) = \begin{cases} \frac{2^{-n/2}}{\Gamma(n/2)} z^{n/2-1} \exp(-z/2), & \text{pour } z > 0, \\ 0 & \text{pour } z \leq 0. \end{cases} \quad (4.1.9)$$

(iii) La fonction caractéristique de  $Z$  est donnée par

$$\phi_Z(u) = (1 - 2iu)^{-n/2}, \quad \text{pour } u \in \mathbb{R}. \quad (4.1.10)$$

**Proposition 4.1.2.** Pour tout  $n \geq 1$ , le moment  $\mu_m = \mathbb{E}(Z^m)$ , d'ordre  $m$ , de  $Z \stackrel{d}{=} \chi_n^2$  est donné, pour tout  $m \in \mathbb{N}$ , par

$$\mu_m = 2^m (n/2)_m = n(n+2) \dots (n+2m-2), \quad \mathbb{E}(Z) = n, \quad \text{Var}(Z) = 2n. \quad (4.1.11)$$

**Preuve.** On constate que (4.1.11) est un cas particulier de (4.1.8) correspondant à  $r = n/2$  et  $\lambda = 1/2$ .  $\square$

**Théorème 4.1.1.** Soit  $X \stackrel{d}{=} N_d(\mathbb{O}, \Sigma)$ , où  $\Sigma > 0$  est une matrice définie positive. Alors  $X' \Sigma^{-1} X \stackrel{d}{=} \chi_d^2$ .

**Preuve.** Posons  $X = PY$ , où  $P$  est une matrice telle que  $PP' = \Sigma$ , et donc  $P'^{-1}P^{-1} = \Sigma^{-1}$ . On a alors  $Y \stackrel{d}{=} N_d(\mathbb{O}, \mathbb{I})$ , et  $X' \Sigma^{-1} X = Y' P' P'^{-1} P^{-1} P Y = Y' Y$ . Maintenant, si  $Y = [Y_1 \dots Y_d]'$ , le fait que  $Y \stackrel{d}{=} N_d(\mathbb{O}, \mathbb{I})$  équivaut au fait que les v.a.r.  $Y_1, \dots, Y_d$  sont  $N(0, 1)$  et indépendantes. Comme  $Y' Y = \sum_{i=1}^d Y_i^2$ , compte tenu du (iii) de la proposition 4.1.1, il nous suffit d'établir que, si  $T \stackrel{d}{=} N(0, 1)$ , alors  $Z = T^2 \stackrel{d}{=} \chi_1^2$ .

Pour établir cette dernière propriété, on constate, par (3.1.1), et en effectuant le changement de variable  $t = \sqrt{z}$  dans la densité  $\varphi(t)$  de la loi normale  $N(0, 1)$ , que la densité de  $Z$  est donnée par

$$f_Z(z) = 2\varphi(t) \left| \frac{dt}{dz} \right| = \varphi(\sqrt{z}) z^{-1/2} = \frac{2^{-1/2}}{\Gamma(1/2)} z^{1/2-1} \exp(-z/2) \quad \text{pour } z > 0,$$

où nous nous sommes servis du fait que  $\Gamma(1/2) = \sqrt{\pi}$  (propriété 4.1.1(iv)). Compte tenu de la définition 4.1.2, on obtient bien ainsi le résultat voulu.  $\square$

### 4.1.2 Quantiles supérieurs de la Loi du $\chi^2$ Centrée.

Soit  $Z \stackrel{d}{=} \chi_n^2$  une variable aléatoire suivant une loi du  $\chi^2$  (centrée) à  $n$  degrés de liberté. Pour chaque valeur de  $0 < \alpha < 1$ , on désigne habituellement par  $\chi_{n,\alpha}^2$  le quantile supérieur d'ordre  $\alpha$  de la loi du  $\chi_n^2$ , c'est à dire, la valeur telle que

$$\mathbb{P}(Z > \chi_{n,\alpha}^2) = \alpha. \quad (4.1.12)$$

Plutôt que  $\chi_{n,\alpha}^2$ , il est commode de tabuler le rapport

$$\frac{\chi_{n,\alpha}^2}{n} =: F_{n,\infty}^\alpha. \quad (4.1.13)$$

En effet, si  $Z_n \stackrel{d}{=} \chi^2(n)$ , la loi des grands nombres implique que  $Z_n/n \rightarrow 1$ , lorsque  $n \rightarrow \infty$ . Le choix donné par (4.1.13) implique ainsi que, indépendamment de  $\alpha \in (0, 1)$ , lorsque  $n \rightarrow \infty$ ,

$$\frac{\chi_{n,\alpha}^2}{n} = F_{n,\infty}^\alpha \rightarrow 1 =: F_{\infty,\infty}^\alpha. \quad (4.1.14)$$

Lorsque  $n \geq 30$ , la formule approchée<sup>1</sup>

$$\chi_{n,\alpha}^2 \simeq \frac{1}{2} \left\{ \sqrt{2n-1} + \nu_\alpha \right\}^2, \quad (4.1.15)$$

fournit une excellente approximation de la valeur numérique de  $\chi_{n,\alpha}^2$ . Celle-ci s'avère bien meilleure que l'approximation fournie par le théorème central limite  $(Z_n - n)/\sqrt{2n} \stackrel{d}{\rightarrow} N(0, 1)$ , soit

$$\chi_{n,\alpha}^2 \simeq n + \nu_\alpha \sqrt{2n}. \quad (4.1.16)$$

n	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 1\%$
1	2.71	3.84	6.63
2	2.30	3.00	4.61
3	2.08	2.60	3.78
4	1.94	2.37	3.32
5	1.85	2.21	3.02
6	1.77	2.10	2.80
7	1.72	2.01	2.64
8	1.67	1.94	2.51
9	1.63	1.88	2.41
10	1.60	1.83	2.32
12	1.55	1.75	2.18
15	1.49	1.67	2.04
20	1.42	1.57	1.88
24	1.38	1.52	1.79
30	1.34	1.46	1.70
40	1.30	1.39	1.59
60	1.24	1.32	1.47
120	1.17	1.22	1.32
$\infty$	1.00	1.00	1.00

Table des quantiles supérieurs pondérés  $\frac{\chi_{n,\alpha}^2}{n}$  de la loi du  $\chi_n^2$ .

**Exercice 4.1.1.** Vérifier que les approximations fournies par (4.1.15) et (4.1.16) sont équivalentes lorsque  $n \rightarrow \infty$ . Comparer les résultats qu'elles fournissent, pour  $n = 30$  et  $\alpha = 5\%$ , à la valeur fournie par la table ci-dessus.

1. Voir : Thompson, C. (1941/42). Tables of percentage points of the  $\chi^2$  distribution. *Biometrika*. **32** 188-191.

## 4.2 Lois du $\chi^2$ non Centrées

Les propriétés suivantes des *fonctions hypergéométriques* univariées seront utiles par la suite. Tout d'abord, le *symbole de Pochhammer* univarié  $(a)_k$  est défini, pour  $a \in \mathbb{C}$  et  $k \in \mathbb{N}$ , par

$$(a)_k := \begin{cases} 1 & \text{si } k = 0, \\ a(a+1) \dots (a+k-1) & \text{si } k \geq 1. \end{cases} \quad (4.2.1)$$

Dans le cas où  $Re(a) > 0$  on peut remplacer cette définition par

$$(a)_k := \frac{\Gamma(a+k)}{\Gamma(a)}. \quad (4.2.2)$$

On notera que la définition (4.2.2) donne un sens à  $(a)_k$  pour  $a, k \in \mathbb{R}$ , avec  $\min(a, a+k) > 0$ .

La fonction hypergéométrique univariée généralisée  ${}_pF_q$  est définie, pour tout choix de  $p, q \in \mathbb{N}$ ,  $a_1, \dots, a_p \in \mathbb{C}$ , et  $b_1, \dots, b_q \in \mathbb{C} - \{-\mathbb{N}\}$ , comme la somme (lorsqu'elle converge) de la série entière

$${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; z) := \sum_{n=0}^{\infty} \frac{(a_1)_n \dots (a_p)_n}{(b_1)_n \dots (b_q)_n} \frac{z^n}{n!}, \quad \text{pour } z \in \mathbb{C}. \quad (4.2.3)$$

**Remarque 4.2.1.** La propriété suivante, dont la démonstration, élémentaire, est omise, montre que  ${}_pF_q$  ne présente d'intérêt pratique que si  $q \geq p-1$ .

**Propriété 4.2.1.** (i) *Le rayon de convergence  $r$ , de la série entière  ${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; z)$ , vérifie*

$$r = \begin{cases} \infty & \text{si } q \geq p, \\ 1 & \text{si } q = p-1, \\ 0 & \text{si } q \leq p-2. \end{cases}$$

(ii) *On a les cas particuliers suivants, pour  ${}_pF_q$ ,  $p, q = 0, 1$ , avec  $a \in \mathbb{C}$  et  $z \in \mathbb{C}$ .*

$$\begin{aligned} {}_0F_0(z) &= e^z; \\ {}_1F_0(a; z) &= (1-z)^{-a}; \\ {}_0F_1\left(\frac{1}{2}; \frac{1}{4}z\right) &= \cosh \sqrt{z} = (e^{\sqrt{z}} + e^{-\sqrt{z}})/2. \end{aligned}$$

**Définition 4.2.1.** *Soit, pour  $n \in \mathbb{N}^* = \{1, 2, \dots\}$ ,*

$$g_n(z) = \frac{2^{-\frac{1}{2}n}}{\Gamma\left(\frac{1}{2}n\right)} z^{\frac{1}{2}n-1} \exp\left(-\frac{1}{2}z\right) \quad \text{pour } z > 0, \quad (4.2.4)$$

la densité d'une loi du  $\chi_n^2$  centrée, et soit  $K \stackrel{d}{=} \text{Po}(\delta/2)$  une variable aléatoire suivant une loi de Poisson d'espérance  $\delta/2$ , avec  $\delta \in \mathbb{R}^+$ , c'est à dire, telle que

$$\mathbb{P}(K = k) = \frac{\left(\frac{1}{2}\delta\right)^k}{k!} \exp\left(-\frac{1}{2}\delta\right), \quad \text{pour } k \in \mathbb{N}. \quad (4.2.5)$$

On dit qu'une variable aléatoire  $Z$  suit une loi du *Khi-deux non centrée*, à  $n$  degrés de liberté, de paramètre de décentrement  $\delta$ , ce qu'on note symboliquement par  $Z \stackrel{d}{=} \chi_n^2(\delta)$ , si la densité de  $Z$  est définie par

$$f_Z(z) = \sum_{k=0}^{\infty} \mathbb{P}(K = k) g_{n+2k}(z), \quad \text{pour } z > 0, \quad (4.2.6)$$

**Remarque 4.2.2.** On retiendra de cette définition une construction commode de  $Z \stackrel{d}{=} \chi_n^2(\delta)$ . On définit  $Z$  et  $K$  sur le même espace de probabilités, de telle sorte que  $K$  suive une loi de Poisson d'espérance  $\delta/2$ , et que la loi conditionnelle de  $Z$  sachant  $\{K = k\}$  soit une loi du  $\chi_{n+2k}^2(0)$ , centrée à  $n + 2k$  degrés de liberté. Ceci justifie la formule symbolique

$$\chi_n^2(\delta) \stackrel{d}{=} \chi_{n+2K}^2(0). \quad (4.2.7)$$

**Remarque 4.2.3.** Une expression analytique de la densité  $f_Z$  d'une loi du  $\chi_n^2(\delta)$  est aisément obtenue en explicitant (4.2.5) et (4.2.6). En remarquant que, par (4.2.1)–(4.2.2),  $(\frac{1}{2}n)_k = \Gamma(\frac{1}{2}n + k)/\Gamma(\frac{1}{2}n)$ , et en faisant usage de (4.2.4), on obtient ainsi les identités

$$\begin{aligned} f_Z(z) &= \exp\left(-\frac{1}{2}\delta\right) \sum_{k=0}^{\infty} \frac{\left(\frac{1}{2}\delta\right)^k}{k!} \frac{2^{-\frac{1}{2}n-k}}{\Gamma\left(\frac{1}{2}n+k\right)} z^{\frac{1}{2}n+k-1} \exp\left(-\frac{1}{2}z\right) \\ &= \exp\left(-\frac{1}{2}\delta\right) \left( \sum_{k=0}^{\infty} \frac{\left(\frac{1}{4}\delta z\right)^k}{k! \left(\frac{1}{2}n\right)_k} \right) \frac{2^{-\frac{1}{2}n}}{\Gamma\left(\frac{1}{2}n\right)} z^{\frac{1}{2}n-1} \exp\left(-\frac{1}{2}z\right) \\ &= \exp\left(-\frac{1}{2}\delta\right) {}_0F_1\left(\frac{1}{2}n; \frac{1}{4}\delta z\right) g_n(z). \end{aligned} \quad (4.2.8)$$

**Proposition 4.2.1.** La fonction caractéristique de  $Z \stackrel{d}{=} \chi_n^2(\delta)$  est donnée par

$$\phi_Z(u) = (1 - 2iu)^{-n/2} \exp\left(\frac{iu\delta}{1 - 2iu}\right) \quad \text{pour } u \in \mathbb{R} \quad (4.2.9)$$

**Preuve.** Par les relations (4.1.10) et (4.2.6), et en remplaçant  $\mathbb{P}(K = k)$  par sa valeur, donnée en (4.2.5), on obtient que

$$\begin{aligned} \phi_Z(u) &= \sum_{k=0}^{\infty} \mathbb{P}(K = k) (1 - 2iu)^{-n/2-k} = (1 - 2iu)^{-n/2} \sum_{k=0}^{\infty} \frac{\left(\frac{1}{2}\delta\right)^k}{k!} (1 - 2iu)^{-k} \exp\left(-\frac{1}{2}\delta\right) \\ &= (1 - 2iu)^{-n/2} \exp\left(\frac{\frac{1}{2}\delta}{1 - 2iu} - \frac{1}{2}\delta\right), \end{aligned}$$

ce qui donne directement (4.2.9).  $\square$

**Remarque 4.2.4.** Il est important de constater que, si  $Z \stackrel{d}{=} \chi_n^2(\delta)$ , alors, pour toute constante  $c \neq 0$ ,  $Z + c$  ne peut pas suivre une loi du  $\chi_n^2(\delta')$ , quelle que soit la valeur de  $\delta'$ . En effet, par (4.2.9),

$$\begin{aligned} \phi_{Z+c}(u) &= (1 - 2iu)^{-n/2} \exp\left(iu\left\{\frac{\delta}{1 - 2iu} + c\right\}\right) \\ &= (1 - 2iu)^{-n/2} \exp\left(iu\left\{\frac{\delta + c - 2iuc}{1 - 2iu}\right\}\right), \end{aligned}$$

expression qui ne peut se mettre sous la forme

$$(1 - 2iu)^{-n/2} \exp\left(iu\left\{\frac{\delta'}{1 - 2iu}\right\}\right),$$

pour un choix convenable de  $\delta'$ , à moins que l'on ait  $c = 0$ .

**Proposition 4.2.2.** Si  $Z_1 \stackrel{d}{=} \chi_{n_1}^2(\delta_1)$  et  $Z_2 \stackrel{d}{=} \chi_{n_2}^2(\delta_2)$  sont indépendantes, alors  $Z_1 + Z_2 \stackrel{d}{=} \chi_{n_1+n_2}^2(\delta_1 + \delta_2)$

**Preuve.** Il suffit de constater que  $\phi_{Z_1+Z_2} = \phi_{Z_1}\phi_{Z_2}$ , ce qui s'obtient comme conséquence directe de (4.2.9).  $\square$

**Proposition 4.2.3.** *Les moments d'ordre 1 et 2 de  $Z \stackrel{d}{=} \chi_n^2(\delta)$  sont donnés par*

$$\mathbb{E}(Z) = n + \delta \quad \text{et} \quad \text{Var}(Z) = 2n + 4\delta. \quad (4.2.10)$$

**Preuve.** Soit  $V_n$  une v.a. suivant une loi du  $\chi_n^2$  et  $K$  une v.a. de Poisson d'espérance  $\delta/2$ . On déduit de la Remarque 4.2.3 que

$$\mathbb{E}(Z^m) = \mathbb{E}\left(\mathbb{E}(Z^m \mid K)\right) = \sum_{k=0}^{\infty} \mathbb{E}(V_{n+2k}^m) P(K = k) = \sum_{k=0}^{\infty} 2^m \left(\frac{1}{2}(n+2k)\right)_m \frac{\left(\frac{1}{2}\delta\right)^k}{k!} \exp\left(-\frac{1}{2}\delta\right).$$

Un calcul simple de cette expression est possible pour  $m = 1$ , ce qui donne, avec  $\mathbb{E}(V_{n+2k}) = n + 2k$ , et  $\mathbb{E}(K) = \delta/2$ ,

$$\mathbb{E}(Z) = \mathbb{E}(n + 2K) = n + \delta.$$

Pour  $m = 2$ , on procède de même, posant  $\mathbb{E}(V_{n+2k}^2) = (n+2k)(n+2k+2)$  et  $\mathbb{E}(K^2) = (\delta/2)(1 + \delta/2)$ , puis écrivant que

$$\begin{aligned} \mathbb{E}(Z^2) &= n(n+2) + 4(n+1)\mathbb{E}(K) + 4\mathbb{E}(K^2) = n(n+2) + 2(n+1)\delta + \delta(2+\delta) \\ &= (n+\delta)^2 + 2n + 4\delta = \mathbb{E}(Z)^2 + 2n + 4\delta, \end{aligned}$$

d'où l'on déduit (4.2.10).  $\square$

**Théorème 4.2.1.** *Soit  $X \stackrel{d}{=} N_d(\mu, \Sigma)$ , où  $\Sigma > 0$  est une matrice symétrique, ( $d \times d$ ), définie positive. Alors :*

$$(i) \quad X' \Sigma^{-1} X \stackrel{d}{=} \chi_d^2(\mu' \Sigma^{-1} \mu); \quad (4.2.11)$$

$$(ii) \quad (X - \mu)' \Sigma^{-1} (X - \mu) \stackrel{d}{=} \chi_d^2. \quad (4.2.12)$$

**Preuve.** Pour établir le (i), on commence par poser  $X = PY$ , où  $P$  est une matrice telle que  $PP' = \Sigma$ , et  $Y$ , telle que  $Y \stackrel{d}{=} N_d(\nu, \mathbb{I})$ . On obtient alors que  $X' \Sigma^{-1} X = Y' Y$ , et  $\mu' \Sigma^{-1} \mu = \nu' \nu$ . Le problème se ramène donc à démontrer que  $Y' Y \stackrel{d}{=} \chi_d^2(\nu' \nu)$ .

Il est toujours possible de choisir une matrice  $C$ , orthogonale ( $d \times d$ ), telle que

$$C\nu = C \begin{bmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_d \end{bmatrix} = \begin{bmatrix} \sqrt{\nu' \nu} \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

En posant  $Z = CY$ , on constate, d'une part, que  $Z' Z = Y' Y$ , et d'autre part, que

$$Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_d \end{bmatrix} \stackrel{d}{=} N_d(C\nu, C\mathbb{I}C') \stackrel{d}{=} N_d(C\nu, \mathbb{I}),$$

ce qui implique que les v.a.r.  $Z_1, Z_2, \dots, Z_d$  sont indépendantes, de lois respectives  $Z_1 \stackrel{d}{=} N(\sqrt{\nu' \nu}, 1)$ , et  $Z_j \stackrel{d}{=} N(0, 1)$  pour  $j = 2, \dots, d$ . Comme alors, par le théorème 4.1.1,  $\sum_{j=2}^d Z_j^2 \stackrel{d}{=} \chi_{d-1}^2 \stackrel{d}{=} \chi_{d-1}^2(0)$ , et compte tenu de la proposition 4.2.2, le problème se ramène à montrer que si  $V \stackrel{d}{=} N(\delta^{1/2}, 1)$ , alors  $T = V^2 \stackrel{d}{=} \chi_1^2(\delta)$ . Pour

établir cette dernière propriété, on obtient d'abord, à partir de la densité de  $V \stackrel{d}{=} N(v, 1)$ , donné par (3.1.13), et en posant  $z = x^2$  (ce qui équivaut à  $x = \sqrt{z}$ , ou  $x = -\sqrt{z}$ ), que

$$\begin{aligned} f_T(z) &= (f_V(x) + f_V(-x)) \left| \frac{dx}{dz} \right| = \frac{z^{-1/2}}{2\sqrt{2\pi}} \left[ \exp\left(-\frac{(x - \delta^{1/2})^2}{2}\right) + \exp\left(-\frac{(-x - \delta^{1/2})^2}{2}\right) \right] \\ &= \exp\left(-\frac{1}{2}\delta\right) \frac{1}{2} \left[ \exp((\delta z)^{1/2}) + \exp(-(\delta z)^{1/2}) \right] \frac{2^{-\frac{1}{2}}}{\Gamma(\frac{1}{2})} z^{\frac{1}{2}-1} \exp\left(-\frac{1}{2}z\right) \\ &= \exp\left(-\frac{1}{2}\delta\right) \cosh\left(\sqrt{\delta z}\right) \left( \frac{2^{-\frac{1}{2}}}{\Gamma(\frac{1}{2})} z^{\frac{1}{2}-1} \exp\left(-\frac{1}{2}z\right) \right). \end{aligned}$$

Pour conclure, il suffit de vérifier par identification des développements, que

$$\cosh(u) = \sum_{n=0}^{\infty} \frac{u^{2n}}{(2n)!} = {}_0F_1\left(\frac{1}{2}, \frac{1}{4}u^2\right). \quad (4.2.13)$$

Finalement, pour établir le (ii), on applique le (i) à  $X - \mu \stackrel{d}{=} N_d(0, \Sigma)$ .  $\square$

**Corollaire 4.2.1.** *Soit  $X \stackrel{d}{=} N_d(\mu, \sigma^2\mathbb{I})$ , où  $\sigma > 0$  est une constante non nulle. Alors, en notant  $\|u\| = \{u'u\}^{1/2}$  la norme euclidienne sur  $\mathbb{R}^d$ , on a les identités en loi*

$$(i) \quad \frac{1}{\sigma^2} \|X\|^2 \stackrel{d}{=} \chi_d^2\left(\frac{1}{\sigma^2} \|\mu\|^2\right); \quad (4.2.14)$$

$$(ii) \quad \frac{1}{\sigma^2} \|X - \mu\|^2 \stackrel{d}{=} \chi_d^2. \quad (4.2.15)$$

**Preuve.** Il suffit d'appliquer le théorème 4.2.1 à la variable  $\sigma^{-1}X \stackrel{d}{=} N_d(\sigma^{-1}\mu, \mathbb{I})$ .  $\square$

**Corollaire 4.2.2.** *Soit  $\mathbb{R}^d = E_1 \oplus \dots \oplus E_k$  une décomposition de  $\mathbb{R}^d$  en somme directe de sous-espaces orthogonaux, tels que  $E_i \perp E_j$  pour tout  $1 \leq i \neq j \leq k$ , relativement au produit scalaire euclidien  $\langle u, v \rangle = u'v$  de  $\mathbb{R}^d$ . Soit  $Y \stackrel{d}{=} N_d(M, \sigma^2\mathbb{I})$  un vecteur aléatoire suivant une loi normale de matrice de variances-covariances proportionnelle à l'identité dans  $\mathbb{R}^d$ , avec  $\sigma > 0$ . Soit  $Y = Y_1 + \dots + Y_k$  l'unique décomposition de  $Y$  en somme de vecteurs tels que, pour  $j = 1, \dots, k$ ,  $Y_j = \mathcal{P}_{E_j}(Y) \in E_j$ , où  $\mathcal{P}_{E_j}$  désigne la projection orthogonale de  $\mathbb{R}^d$  sur  $E_j$ . Désignons de manière analogue par  $M = M_1 + \dots + M_k$  l'unique décomposition de  $M$  en somme de vecteurs tels que, pour  $j = 1, \dots, k$ ,  $M_j = \mathcal{P}_{E_j}(M) \in E_j$  soit la projection orthogonale de  $M$  sur  $E_j$ . On a alors les résultats suivants, en notant  $\|u\| = \{u'u\}^{1/2} = \langle u, u \rangle^{1/2}$  la norme euclidienne sur  $\mathbb{R}^d$ .*

1°)  $Y_1, \dots, Y_k$  (et, par conséquent,  $Y_1'Y_1 = \|Y_1\|^2, \dots, Y_k'Y_k = \|Y_k\|^2$ ) sont indépendants;

2°) Si, pour  $j = 1, \dots, k$ ,  $d_j = \dim(E_j)$ , avec  $d_1 + \dots + d_k = d$ , alors, pour chaque  $j = 1, \dots, k$ ,

$$\frac{1}{\sigma^2} Y_j'Y_j = \frac{1}{\sigma^2} \langle Y_j, Y_j \rangle = \frac{1}{\sigma^2} \|Y_j\|^2 \stackrel{d}{=} \chi_{d_j}^2 \left( \frac{1}{\sigma^2} M_j' M_j \right) \stackrel{d}{=} \chi_{d_j}^2 \left( \frac{1}{\sigma^2} \langle M_j, M_j \rangle \right) \stackrel{d}{=} \chi_{d_j}^2 \left( \frac{1}{\sigma^2} \|M_j\|^2 \right). \quad (4.2.16)$$

3°) On a l'identité

$$Y'Y = \langle Y, Y \rangle = \sum_{j=1}^k Y_j'Y_j = \sum_{j=1}^k \langle Y_j, Y_j \rangle \quad \text{et} \quad \frac{1}{\sigma^2} Y'Y \stackrel{d}{=} \chi_d^2 \left( \frac{1}{\sigma^2} \|M\|^2 \right). \quad (4.2.17)$$

**Preuve.** On se ramène au cas  $\sigma^2 = 1$ . Il est possible de construire une base orthonormée de  $\mathbb{R}^d$ , soit  $\{e_{j,i} : 1 \leq i \leq d_j, 1 \leq j \leq k\}$ , telle que, pour chaque choix de  $j = 1, \dots, k$ ,  $\{e_{j,1}, \dots, e_{j,d_j}\}$  constitue une base orthonormée de  $E_j$ . Le changement de base, passant de la base canonique de  $\mathbb{R}^d$  à cette nouvelle base, ayant alors une matrice de passage orthogonale  $H$ , les coordonnées de  $Y$  dans la nouvelle base seront données par

$HY \stackrel{d}{=} N_d(HM, \mathbb{I})$ , cette dernière identité se déduisant du corollaire 3.2.1. Ceci implique que les coordonnées de  $Y$  dans cette nouvelle base sont mutuellement indépendantes, suivant chacune une loi normale de variance 1. On en déduit le 1°).

Pour établir le 2°), par linéarité de la projection, on constate que, pour  $j = 1, \dots, k$ ,  $M_j = \mathcal{P}_{E_j}(\mathbb{E}(Y)) = \mathbb{E}(\mathcal{P}_{E_j}(Y)) = \mathbb{E}(Y_j)$ . Soit  $\mathcal{Y}_j$  la matrice colonne des coordonnées de  $Y_j$  dans la base  $e_{j,1}, \dots, e_{j,d_j}$ . On a, comme ci-dessus,  $\mathcal{Y}_j = HY_j$ , de sorte que  $\mathcal{M}_j = \mathbb{E}(\mathcal{Y}_j) = HM_j$ . Comme la matrice de variances-covariances de  $\mathcal{Y}_j$  est la matrice identité  $\mathbb{I}$  de dimension  $d_j$ , on a, par (4.2.14),  $\mathcal{Y}'_j \mathcal{Y}_j \stackrel{d}{=} \chi_{d_j}^2(\mathcal{M}'_j \mathcal{M}_j)$ . On conclut à la validité de (4.2.18) en constatant (du fait que  $HH' = \mathbb{I}$ ) que, pour  $j = 1, \dots, d$ ,  $\mathcal{Y}'_j \mathcal{Y}_j = Y'_j Y_j$ , et  $\mathcal{M}'_j \mathcal{M}_j = M'_j M_j$ .

Finalement, le 3°) ne fait qu'exprimer la relation de Pythagore entre les composantes orthogonales  $Y_1, \dots, Y_k$  de  $Y$ .  $\square$

**Corollaire 4.2.3.** *Soit  $\Sigma > 0$  une matrice  $(d \times d)$  définie positive, et soit  $\mathbb{R}^d = F_1 \oplus \dots \oplus F_k$  une décomposition de  $\mathbb{R}^d$  en somme directe de sous-espaces orthogonaux, tels que  $F_i \perp F_j$  pour tout  $1 \leq i \neq j \leq k$ , relativement au produit scalaire  $\langle u, v \rangle_{\Sigma^{-1}} = u' \Sigma^{-1} v$  de  $\mathbb{R}^d$ . Soit  $X \stackrel{d}{=} N_d(M, \Sigma)$  un vecteur aléatoire normal de matrice de variances-covariances  $\Sigma$  dans  $\mathbb{R}^d$ , et soit  $X = X_1 + \dots + X_k$  son unique décomposition en somme de vecteurs tels que, pour  $i = 1, \dots, k$ ,  $X_i = \mathcal{P}_{F_i}(X) \in E_i$  désigne la projection orthogonale (relativement au produit scalaire  $\langle u, v \rangle_{\Sigma^{-1}} = u' \Sigma^{-1} v$  de  $\mathbb{R}^d$ ) de  $X$  sur  $F_i$ . Désignons de manière analogue par  $M = M_1 + \dots + M_k$  l'unique décomposition de  $M$  en somme de vecteurs tels que, pour  $i = 1, \dots, k$ ,  $M_i = \mathcal{P}_{E_i}(M) \in E_i$  soit la projection orthogonale (relativement au produit scalaire  $\langle u, v \rangle_{\Sigma^{-1}} = u' \Sigma^{-1} v$  de  $\mathbb{R}^d$ ) de  $M$  sur  $E_i$ . On a alors les résultats suivants.*

1°)  $X_1, \dots, X_k$  (et, par conséquent,  $X'_1 \Sigma^{-1} X_1, \dots, X'_k \Sigma^{-1} X_k$ ) sont indépendants ;

2°) Si, pour  $j = 1, \dots, k$ , on désigne par  $d_j = \dim(F_j)$  la dimension de  $F_j$ , alors

$$X'_j \Sigma^{-1} X_j = \langle X_j, X_j \rangle_{\Sigma^{-1}} \stackrel{d}{=} \chi_{d_j}^2(M'_j \Sigma^{-1} M_j) = \chi_{d_j}^2(\langle M_j, M_j \rangle_{\Sigma^{-1}}). \quad (4.2.18)$$

3°) On a l'identité

$$X' \Sigma^{-1} X = \langle X, X \rangle_{\Sigma^{-1}} = \sum_{i=1}^k X'_i \Sigma^{-1} X_i = \sum_{i=1}^k \langle X_i, X_i \rangle_{\Sigma^{-1}} \stackrel{d}{=} \chi_d^2(M' \Sigma^{-1} M). \quad (4.2.19)$$

**Proof.** On se ramène au Corollaire 4.2.1, en posant  $X = PY$  pour une matrice  $P$  telle que  $PP' = \Sigma$ .  $\square$

Le théorème suivant donne des conditions nécessaires et suffisantes simples pour que  $X'BX$  suive une loi du  $\chi^2$ , lorsque  $X \stackrel{d}{=} N_d(\mu, \mathbb{I}_d)$ .

**Théorème 4.2.2.** *Si  $X \stackrel{d}{=} N_d(\mu, \mathbb{I}_d)$ , et si  $B$  est une matrice symétrique  $(d \times d)$ , alors,  $X'BX$  suit une loi du  $\chi_k^2(\delta)$  si et seulement si la matrice  $B$  est idempotente, i.e. telle que  $B^2 = B$ . Dans ce cas, le nombre de degrés de liberté  $k$  et le paramètre de décentrement  $\delta$  sont donnés respectivement par*

$$X'BX \stackrel{d}{=} \chi_k^2(\delta) \quad \text{avec} \quad k = \text{rg}(B) = \text{tr}(B) \quad \text{et} \quad \delta = \mu' B \mu. \quad (4.2.20)$$

**Preuve.** Supposons tout d'abord que  $B$  soit idempotente de rang  $k$ . Comme  $B$  est, de plus, symétrique, il existe, par (2.2.10), une base  $h_1, \dots, h_d$  orthonormée de  $\mathbb{R}^d$ , composée de vecteurs propres de  $B$ , telle que, la matrice  $H = [h_1 \ \dots \ h_d]$  soit une matrice orthogonale vérifiant les identités

$$B = H \begin{bmatrix} \mathbb{I}_k & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} H', \quad \text{tr}(B) = \text{rg}(B) = k, \quad (4.2.21)$$

et, par conséquent

$$X'BX = (H'X)' \begin{bmatrix} \mathbb{I}_k & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} (H'X). \quad (4.2.22)$$



Comme  $H$  est orthogonale,  $Y = H'X = [Y_1 \ \dots \ Y_d]'$  est de loi  $Y \stackrel{d}{=} N_d(H'\mu, H'\mathbb{I}_dH) = N_d(H'\mu, \mathbb{I}_d)$ . Ceci implique que les v.a.r.  $Y_1, \dots, Y_d$  sont indépendantes et de lois respectives  $N(\nu_1, 1), \dots, N(\nu_d, 1)$ . Posons  $\nu = H'\mu = [\nu_1 \ \dots \ \nu_d]'$ . La relation (4.2.21) montre alors que

$$\mu' B \mu = (H'\mu)' \begin{bmatrix} \mathbb{I}_k & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} (H'\mu) = \nu' \begin{bmatrix} \mathbb{I}_k & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \nu = \sum_{j=1}^k \nu_j^2.$$

Cette relation, combinée à (4.2.22) implique que

$$X' B X = Y_1^2 + \dots + Y_k^2 \stackrel{d}{=} \chi_k^2 \left( \sum_{j=1}^k \nu_j^2 \right) \stackrel{d}{=} \chi_k^2 (\mu' B \mu),$$

qui établit (4.2.20).

Réciproquement, supposons que  $X' B X \stackrel{d}{=} \chi_k^2(\delta)$ . En posant  $\text{rg}(B) = p$ , on peut toujours trouver une matrice orthononale  $H$ , et des constantes  $\lambda_1, \dots, \lambda_p \neq 0$ , telles que

$$B = H \begin{bmatrix} \text{diag}(\lambda_1, \dots, \lambda_p) & \mathbb{O} \\ & \mathbb{O} \end{bmatrix} H'. \quad (4.2.23)$$

En posant, comme précédemment,  $Y = H'X = [Y_1 \ \dots \ Y_d]'$   $\stackrel{d}{=} N_d(H'\mu, \mathbb{I}_d)$ , on constate que

$$X' B X = \sum_{j=1}^p \lambda_j Y_j^2. \quad (4.2.24)$$

Comme les  $Y_j^2$  sont des v.a. indépendantes, de lois données par  $Y_j^2 \stackrel{d}{=} \chi_1^2(\nu_j^2)$ , on déduit de (4.2.9) que la fonction caractéristique de  $X' B X$  est donnée par

$$\begin{aligned} \mathbb{E} \left( \exp(iu X' B X) \right) &= \prod_{j=1}^p \mathbb{E} \left( \exp(iu \lambda_j Y_j^2) \right) \\ &= \prod_{j=1}^p \left\{ (1 - 2iu \lambda_j)^{-1/2} \exp \left( \frac{iu \lambda_j \nu_j^2}{1 - 2iu \lambda_j} \right) \right\} \\ &= \left\{ \prod_{j=1}^p (1 - 2iu \lambda_j)^{-1/2} \right\} \exp \left( iu \sum_{j=1}^p \frac{\lambda_j \nu_j^2}{1 - 2iu \lambda_j} \right). \end{aligned}$$

Or, on a supposé que  $X' B X \stackrel{d}{=} \chi_k^2(\delta)$ , ce qui implique, compte tenu de ce qui précède, l'identité

$$\begin{aligned} \mathbb{E} \left( \exp(iu X' B X) \right) &= (1 - 2iu)^{-k/2} \exp \left( \frac{iu \delta}{1 - 2iu} \right) \\ &= \left\{ \prod_{j=1}^p (1 - 2iu \lambda_j)^{-1/2} \right\} \exp \left( iu \sum_{j=1}^p \frac{\lambda_j \nu_j^2}{1 - 2iu \lambda_j} \right). \end{aligned}$$

A l'évidence, cette relation ne peut avoir lieu  $\forall u \in \mathbb{R}$  et pour une valeur convenable de  $\delta$ , que si  $p = k$  et  $\lambda_1 = \dots = \lambda_k = 1$ , ce qui ramène au cas considéré antérieurement.  $\square$

Le résultat ci-dessous fait appel à la notion d'*inverse généralisé*, développée au §6. Il pourra donc être évité en première lecture.

**Théorème 4.2.3.** Soit  $\mathbf{X} \stackrel{d}{=} N_d(\mathbb{O}, \Sigma)$ , où  $\Sigma \geq 0$  est une matrice positive de rang  $r = \text{rg}(\Sigma) \leq d$ , et soit  $\Sigma^-$  un pseudoinverse (ou inverse généralisé) de  $\Sigma$  (c'est à dire, une matrice  $\Sigma^-$  vérifiant  $\Sigma \Sigma^- \Sigma = \Sigma$ ). Alors

$$\mathbf{X}' \Sigma^- \mathbf{X} \stackrel{d}{=} \chi_r^2. \quad (4.2.25)$$

Plus généralement, si  $\mathbf{X} \stackrel{d}{=} N_d(\mathbf{M}, \Sigma)$ , et si  $\Sigma^\dagger$  désigne le pseudo-inverse de Moore-Penrose de  $\Sigma$ , alors,

$$\mathbf{X}'\Sigma^\dagger\mathbf{X} \stackrel{d}{=} \chi_r^2(\mathbf{M}'\Sigma^\dagger\mathbf{M}). \quad (4.2.26)$$

**Preuve.** Supposons que  $\mathbf{X} \stackrel{d}{=} N_d(\mathbf{M}, \Sigma)$ , et posons  $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = C\mathbf{X} \Leftrightarrow \mathbf{X} = C^{-1}\mathbf{y}$ , où  $C = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}$  est une matrice  $(d \times d)$  régulière, telle que

$$C\Sigma C' = \begin{bmatrix} \mathbb{I}_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}. \quad (4.2.27)$$

Ceci implique que

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = C\mathbf{X} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} \mathbf{X} \stackrel{d}{=} N_d(C\mathbf{M}, C\Sigma C') = N_d\left(\begin{bmatrix} C_1\mathbf{M} \\ C_2\mathbf{M} \end{bmatrix}, \begin{bmatrix} \mathbb{I}_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}\right),$$

On a donc, nécessairement,  $\mathbf{y}_2 = C_2\mathbf{M}$  et  $\mathbf{y}_1 \stackrel{d}{=} N_r(C_1\mathbf{M}, \mathbb{I}_r)$ . Par hypothèse,  $\Sigma^-$  vérifie  $\Sigma\Sigma^-\Sigma = \Sigma$ . Par conséquent,

$$\begin{aligned} \begin{bmatrix} \mathbb{I}_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} &= C\Sigma C' = C(\Sigma\Sigma^-\Sigma)C' = (C\Sigma C')(C'^{-1}\Sigma^-C^{-1})(C\Sigma C') \\ &= \begin{bmatrix} \mathbb{I}_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} C'^{-1}\Sigma^-C^{-1} \begin{bmatrix} \mathbb{I}_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}. \end{aligned}$$

Supposons maintenant que  $C_2\mathbf{M} = \mathbb{O}$ , de sorte que  $\mathbf{y}_2 = \mathbb{O}$ . On en déduit que

$$\begin{aligned} \mathbf{X}'\Sigma^-\mathbf{X} &= \mathbf{y}'C'^{-1}\Sigma^-C^{-1}\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbb{O} \end{bmatrix}' C'^{-1}\Sigma^-C^{-1} \begin{bmatrix} \mathbf{y}_1 \\ \mathbb{O} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{y}_1 \\ \mathbb{O} \end{bmatrix}' \begin{bmatrix} \mathbb{I}_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} C'^{-1}\Sigma^-C^{-1} \begin{bmatrix} \mathbb{I}_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbb{O} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbb{O} \end{bmatrix}' \begin{bmatrix} \mathbb{I}_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbb{O} \end{bmatrix} \\ &= \mathbf{y}_1'\mathbf{y}_1 \equiv \chi_r^2(\mathbf{M}'C_1'C_1\mathbf{M}). \end{aligned}$$

De même,

$$\begin{aligned} \mathbf{M}'\Sigma^-\mathbf{M} &= \{C\mathbf{M}\}'C'^{-1}\Sigma^-C^{-1}\{C\mathbf{M}\} = \begin{bmatrix} C_1\mathbf{M} \\ \mathbb{O} \end{bmatrix}' C'^{-1}\Sigma^-C^{-1} \begin{bmatrix} C_1\mathbf{M} \\ \mathbb{O} \end{bmatrix} \\ &= \begin{bmatrix} C_1\mathbf{M} \\ \mathbb{O} \end{bmatrix}' \begin{bmatrix} \mathbb{I}_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} C'^{-1}\Sigma^-C^{-1} \begin{bmatrix} \mathbb{I}_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \begin{bmatrix} C_1\mathbf{M} \\ \mathbb{O} \end{bmatrix} = \begin{bmatrix} C_1\mathbf{M} \\ \mathbb{O} \end{bmatrix}' \begin{bmatrix} \mathbb{I}_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \begin{bmatrix} C_1\mathbf{M} \\ \mathbb{O} \end{bmatrix} \\ &= \{C_1\mathbf{M}\}'\{C_1\mathbf{M}\} = \mathbf{M}'C_1'C_1\mathbf{M}. \end{aligned}$$

On en déduit que

$$\mathbf{X}'\Sigma^-\mathbf{X} = \mathbf{y}_1'\mathbf{y}_1 \stackrel{d}{=} \chi_r^2(\mathbf{M}'C_1'C_1\mathbf{M}) = \chi_r^2(\mathbf{M}'\Sigma^-\mathbf{M}), \quad (4.2.28)$$

qui fournit le résultat (4.2.25) dans le cas où  $\mathbf{M} = \mathbb{O}$ .

Comme, en général, il existe une matrice  $H$ , orthogonale  $d \times d$ , telle que

$$H\Sigma H' = \text{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0),$$

il est facile de constater qu'un choix possible de  $C$  vérifiant (4.2.27) est donné par

$$C = \text{diag}(1/\sqrt{\lambda_1}, \dots, 1/\sqrt{\lambda_r}, 0, \dots, 0)H.$$

Cette matrice n'est pas inversible, mais son pseudo-inverse de Moore-Penrose existe, et est donné par

$$C^\dagger = H' \text{diag}(\sqrt{\lambda_1}, \dots, 1/\sqrt{\lambda_r}, 0, \dots, 0).$$

Le pseudo-inverse de Moore-Penrose  $\Sigma^\dagger$ , de  $\Sigma$ , vérifie

$$\Sigma^\dagger = H' \text{diag}(1/\lambda_1, \dots, 1/\lambda_r, 0, \dots, 0)H = C'C = C' \begin{bmatrix} \mathbb{I}_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} C,$$

ce qui permet d'écrire que

$$\begin{aligned} \mathbf{M}'\Sigma^\dagger\mathbf{M} &= \{C\mathbf{M}\}' \begin{bmatrix} \mathbb{I}_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \{C\mathbf{M}\} = \begin{bmatrix} C_1\mathbf{M}' \\ C_2\mathbf{M}' \end{bmatrix}' \begin{bmatrix} \mathbb{I}_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \begin{bmatrix} C_1\mathbf{M} \\ C_2\mathbf{M} \end{bmatrix} \\ &= \{C_1\mathbf{M}'\}' \{C_1\mathbf{M}\} = \mathbf{M}'C_1'C_1\mathbf{M}. \end{aligned}$$

On en déduit que

$$\mathbf{X}'\Sigma^\dagger\mathbf{X} \stackrel{d}{=} \chi_r^2(\mathbf{M}'C_1'C_1\mathbf{M}) = \chi_r^2(\mathbf{M}'\Sigma^\dagger\mathbf{M}), \quad (4.2.29)$$

qui fournit le résultat (4.2.26).□

**Remarque 4.2.5.** Considérons le cas particulier où

$$\mathbf{x} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ \vdots \\ X_r \\ X_{r+1} \\ \vdots \\ X_d \end{bmatrix} \stackrel{d}{=} N_d(\mathbf{M}, \Sigma) = N_d\left(\begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \end{bmatrix}, \begin{bmatrix} \mathbb{I} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}\right) \quad \text{et} \quad \mathbf{M} = \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_r \\ \mu_{r+1} \\ \vdots \\ \mu_d \end{bmatrix}.$$

De toute évidence,  $\mathbb{E}(\mathbf{X}_1) = \mathbf{M}_1$  et  $\mathbf{X}_2 = \mathbf{M}_2$ , car les variables  $X_{r+1} = \mu_{r+1}, \dots, X_d = \mu_d$  sont constantes. Considérons l'inverse généralisé  $\Sigma^-$  de  $\Sigma$  défini par

$$\Sigma^- = \begin{bmatrix} \mathbb{I} & \mathbb{O} \\ \mathbb{O} & \lambda\mathbb{I} \end{bmatrix}.$$

On a alors

$$\mathbf{X}'\Sigma^-\mathbf{X} = \mathbf{X}'_1\mathbf{X}_1 + \lambda\mathbf{M}'_2\mathbf{M}_2. \quad (4.2.30)$$

Comme, de toute évidence,  $\mathbf{X}'_1\mathbf{X}_1 \stackrel{d}{=} \chi_r^2(\mathbf{M}'_1\mathbf{M}_1)$ , il ressort de la remarque 3.4 que  $\mathbf{X}'\Sigma^-\mathbf{X} = \mathbf{X}'_1\mathbf{X}_1 + c$ , où  $c = \lambda\mathbf{M}'_2\mathbf{M}_2$ , ne peut suivre une loi du Khi-deux que lorsque  $c = 0$ . On constate donc que, sans condition restrictive portant, soit sur  $\mathbf{M} = \mathbb{E}(\mathbf{X})$ , soit sur  $\Sigma^-$ , la loi de  $\mathbf{X}'\Sigma^-\mathbf{X}$  n'est, en général, pas une loi du Khi-deux. La formule (4.2.28) n'est donc pas valable en général, indépendamment des choix de  $\Sigma^-$  et  $\mathbf{M}$ . Notons toutefois que le cas particulier où  $\lambda = 0$  donne

$$\Sigma^\dagger = \begin{bmatrix} \mathbb{I} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix},$$

ce qui permet de vérifier (4.2.26), en écrivant que

$$\mathbf{X}'\Sigma^\dagger\mathbf{X} = \mathbf{X}'_1\mathbf{X}_1 \stackrel{d}{=} \chi_r^2(\mathbf{M}'_1\mathbf{M}_1) = \chi_r^2(\mathbf{M}'\Sigma^\dagger\mathbf{M}). \quad (4.2.31)$$

### 4.3 Lois de Fisher

#### 4.3.1 A. Définition et propriétés élémentaires.

**Définition 4.3.1.** Soient  $Z_1 \stackrel{d}{=} \chi_{n_1}^2(\delta)$  et  $Z_2 \stackrel{d}{=} \chi_{n_2}^2$  deux variables aléatoires suivant des lois du Khi-deux indépendantes, avec  $n_1 \geq 1$  et  $n_2 \geq 1$ . Alors le rapport

$$F = \frac{Z_1/n_1}{Z_2/n_2}$$

suit une loi de Fisher à  $n_1$  et  $n_2$  degrés de liberté, et de paramètre de décentrement  $\delta$  (resp. une loi de Fisher centrée si  $\delta = 0$ ), ce qui est noté symboliquement par  $F \stackrel{d}{=} F_{n_1, n_2}(\delta)$  (resp.  $F \stackrel{d}{=} F_{n_1, n_2} \stackrel{d}{=} F_{n_1, n_2}(0)$ ).

**Théorème 4.3.1.** Pour tout couple d'entiers  $m \geq 1$  et  $n \geq 1$ , désignons par  $Z_{m,n}$  une variable aléatoire de loi  $F_{m,n}$ . Alors, si  $Z \stackrel{d}{=} F_{n_1, n_2}(\delta)$ , pour tout  $x \geq 0$ ,

$$\mathbb{P}(Z \leq x) = \exp(-\frac{1}{2}\delta) \sum_{k=0}^{\infty} \frac{(\frac{1}{2}\delta)^k}{k!} \mathbb{P}\left(Z_{n_1+2k, n_2} \leq \frac{n_1 x}{n_1 + 2k}\right). \quad (4.3.1)$$

La densité de  $Z \stackrel{d}{=} F_{n_1, n_2}(\delta)$  est donnée par

$$\begin{aligned} f_Z(z) &= \exp(-\frac{1}{2}\delta) {}_1F_1\left(\frac{1}{2}(n_1 + n_2); \frac{1}{2}n_1; \frac{\frac{1}{2}\frac{n_1}{n_2}\delta z}{1 + \frac{n_1}{n_2}z}\right) \\ &\times \frac{\Gamma(\frac{1}{2}(n_1 + n_2))}{\Gamma(\frac{1}{2}n_1)\Gamma(\frac{1}{2}n_2)} \frac{z^{n_1/2-1}(\frac{n_1}{n_2})^{n_1/2}}{(1 + \frac{n_1}{n_2}z)^{(n_1+n_2)/2}} \quad \text{pour } z > 0. \end{aligned} \quad (4.3.2)$$

Si  $n_2 \geq 3$ , l'espérance de  $Z$  existe et donnée par

$$\mathbb{E}(Z) = \frac{n_2(n_1 + \delta)}{n_1(n_2 - 2)}.$$

Si  $n_2 \geq 5$ , la variance de  $Z$  existe et donnée par

$$\text{Var}(Z) = 2\left(\frac{n_2}{n_1}\right)^2 \left(\frac{(n_1 + \delta)^2 + (n_1 + 2\delta)(n_2 - 2)}{(n_2 - 2)^2(n_2 - 4)}\right).$$

**Preuve.** Admise.  $\square$

Les deux propositions suivantes sont extrêmement utiles pour calculer les quantiles de la loi de Fisher centrée.

**Proposition 4.3.1.** Si les variables aléatoires  $X \stackrel{d}{=} \Gamma(r, \lambda)$  et  $Y \stackrel{d}{=} \Gamma(s, \lambda)$  sont indépendantes, alors les variables aléatoires  $X/(X + Y)$  et  $X + Y$  sont indépendantes, et  $X/(X + Y) \stackrel{d}{=} \beta(r, s)$ .

**Preuve.** La densité jointe de  $(X, Y)$  est

$$f_{X,Y}(x, y) = \lambda^{r+s} \frac{x^{r-1} y^{s-1}}{\Gamma(r)\Gamma(s)} \exp(-\lambda(x + y)).$$

En effectuant le changement de variables  $(x, y) \rightarrow (x, z = \frac{x}{x+y})$ , de jacobien

$$\left| \frac{D(x, z)}{D(x, y)} \right| = \left| \det \begin{bmatrix} \frac{\partial x}{\partial x} & \frac{\partial x}{\partial z} \\ \frac{\partial z}{\partial x} & \frac{\partial z}{\partial z} \end{bmatrix} \right| = \frac{z^2}{x},$$

on obtient la densité de  $(X, Z = \frac{X}{X+Y})$  :

$$f_{X,Z}(x, z) = f_{X,Y}(x, y) \left| \frac{D(x, y)}{D(x, z)} \right| = \lambda^{r+s} \frac{x^{r+s-1} (\frac{1}{z} - 1)^{s-1} \Gamma(r) \Gamma(s)}{\exp} \left( -\lambda \frac{x}{z} \right) z^{-2},$$

ce qui, après intégration par rapport à  $x$  sur  $(0, \infty)$ , donne, compte tenu du (iii) de la Propriété 4.1.1,

$$\frac{z^{r-1} (1-z)^{s-1} \Gamma(r+s)}{\Gamma(r) \Gamma(s)} = \frac{z^{r-1} (1-z)^{s-1}}{\beta(r, s)},$$

ce qui achève la démonstration.  $\square$

**Proposition 4.3.2.** Si  $Z \stackrel{d}{=} F_{n_1, n_2}$ , alors

$$\frac{1}{1 + \frac{n_1}{n_2} Z} \stackrel{d}{=} \beta\left(\frac{1}{2}n_2, \frac{1}{2}n_1\right).$$

**Preuve.** On applique la proposition 4.3.1 avec  $r = n_2/2$ ,  $s = n_1/2$  et  $\lambda = 1/2$ .  $\square$

**Preuve du théorème 4.3.1.** Nous donnons quelques éléments de sa démonstration. Tout d'abord, nous calculons les moments de  $Z \stackrel{d}{=} F_{n_1, n_2}$ . Par la proposition 4.3.2,

$$V = \frac{1}{1 + \frac{n_1}{n_2} Z} \stackrel{d}{=} \beta\left(\frac{1}{2}n_2, \frac{1}{2}n_1\right) \quad \text{et} \quad Z = \frac{n_2}{n_1} \left[ \frac{1}{V} - 1 \right].$$

On en déduit que

$$\begin{aligned} \mathbb{E}(Z^k) &= \mathbb{E}\left(\left\{ \frac{n_2}{n_1} \left[ \frac{1}{V} - 1 \right] \right\}^k\right) \\ &= \frac{1}{\beta\left(\frac{1}{2}n_2, \frac{1}{2}n_1\right)} \frac{n_2^k}{n_1^k} \int_0^1 v^{\frac{1}{2}n_2 - k - 1} (1-v)^{\frac{1}{2}n_1 + k - 1} dv \\ &= \frac{n_2^k}{n_1^k} \left\{ \frac{\beta\left(\frac{1}{2}n_2 - k, \frac{1}{2}n_1 + k\right)}{\beta\left(\frac{1}{2}n_2, \frac{1}{2}n_1\right)} \right\} = \frac{n_2^k}{n_1^k} \left\{ \frac{\Gamma\left(\frac{1}{2}n_2 - k\right) \Gamma\left(\frac{1}{2}n_1 + k\right)}{\Gamma\left(\frac{1}{2}n_2\right) \Gamma\left(\frac{1}{2}n_1\right)} \right\}. \end{aligned}$$

En particulier, pour  $k = 1$ , on obtient

$$\mathbb{E}(Z) = \frac{n_2}{n_1} \left\{ \frac{\Gamma\left(\frac{1}{2}n_2 - 1\right) \Gamma\left(\frac{1}{2}n_1 + 1\right)}{\Gamma\left(\frac{1}{2}n_2\right) \Gamma\left(\frac{1}{2}n_1\right)} \right\} = \frac{n_2}{n_1} \left\{ \frac{\frac{1}{2}n_1}{\frac{1}{2}n_2 - 1} \right\} = \frac{n_2}{n_2 - 2}.$$

On notera que, si  $Z \stackrel{d}{=} F_{n_1, n_2}$ , alors  $E(Z) = \infty$  si  $n_2 \leq 2$ . Pour  $n_2 \geq 3$ , l'espérance  $E(Z) = n_2/(n_2 - 2) < \infty$  est finie et *indépendante de  $n_1$* . Pour  $k = 2$ , on obtient, de même,

$$\begin{aligned} \mathbb{E}(Z^2) &= \frac{n_2^2}{n_1^2} \left\{ \frac{\Gamma\left(\frac{1}{2}n_2 - 2\right) \Gamma\left(\frac{1}{2}n_1 + 2\right)}{\Gamma\left(\frac{1}{2}n_2\right) \Gamma\left(\frac{1}{2}n_1\right)} \right\} = \frac{n_2^2}{n_1^2} \left\{ \frac{\frac{1}{2}n_1 \left(\frac{1}{2}n_1 + 1\right)}{\left(\frac{1}{2}n_2 - 1\right) \left(\frac{1}{2}n_2 - 2\right)} \right\} \\ &= \frac{n_2^2 (n_1 + 2)}{n_1 (n_2 - 2) (n_2 - 4)}, \end{aligned}$$

expression qui n'est finie que pour  $n_2 \geq 5$ . On en déduit que

$$\text{Var}(Z) = \frac{n_2^2 (n_1 + 2)}{n_1 (n_2 - 2) (n_2 - 4)} - \frac{n_2^2}{(n_2 - 2)^2} = \frac{2n_2^2 (n_1 + n_2 - 2)}{n_1 (n_2 - 2)^2 (n_2 - 4)}.$$

**Exercice 4.3.1.** Déterminer, par une expression simple s'exprimant en fonction de  $\alpha \in (0, 1)$ , le quantile supérieur  $f_{2,2;\alpha}$  d'ordre  $\alpha$  de la loi de Fisher  $F_{2,2}$ .

### 4.3.2 B. Application aux calculs de puissance.

Sous sa forme générale, un *test de Fisher* considère deux variables aléatoires indépendantes

$$X_1 \stackrel{d}{=} \chi_{n_1}^2(\delta) \quad \text{et} \quad X_2 \stackrel{d}{=} \chi_{n_2}^2. \quad (4.3.3)$$

Il s'agit de tester l'hypothèse  $(H.0) : \{\delta = 0\}$  contre l'alternative que  $\delta \geq 0$  est quelconque. On rejette  $(H.0)$  au seuil  $\alpha \in (0, 1)$ , lorsque

$$F_{n_1, n_2}(\delta) := \frac{X_1/n_1}{X_2/n_2} \geq f_{n_1, n_2; \alpha}, \quad (4.3.4)$$

où  $f_{n_1, n_2; \alpha}$  désigne le quantile supérieur d'ordre  $\alpha$  de la loi de Fisher  $F_{n_1, n_2}$ . La *puissance* du test est, par définition, la probabilité de rejeter  $(H.0)$  sachant la valeur de  $\delta$ , soit la fonction de  $\delta \geq 0$  définie par

$$\mathcal{P}_\alpha(\delta) = \mathbb{P}(F_{n_1, n_2}(\delta) \geq f_{n_1, n_2; \alpha}). \quad (4.3.5)$$

Pour  $\alpha \in (0, 1)$  fixé,  $\mathcal{P}_\alpha(\delta)$  est une fonction croissante de  $\delta \geq 0$ , telle que  $\mathcal{P}_\alpha(0) = \alpha$ , et  $\mathcal{P}_\alpha(\delta) \rightarrow 1$  lorsque  $\delta \rightarrow \infty$  (voir Johnson, Kotz et Balakrishnan (1995), p. 487). La fonction de puissance  $\mathcal{P}_\alpha(\delta)$  définie en (4.3.5) se calcule, soit à l'aide de logiciels (voir le § 5.2.2), soit à l'aide de tables, dont les plus classiques sont dues à Tiku (1967). On consultera à cet effet :

Tiku, M. L. (1967). Tables of the power of the  $F$ -test. *Journal of the American Statist. Assoc.* **62** 525–539.

Johnson, N. L., Kotz, S. et Balakrishnan, N. (1995). *Continuous Univariate Distributions*. 2<sup>nd</sup> Ed., Wiley, New York.

## 4.4 Loi de Student

### 4.4.1 A. Loi de Student centrée.

La loi de Student, introduite en 1908 par William Gosset, sous le pseudonyme de "Student" (l'étudiant), constitue l'un des outils les plus utiles de la statistique.

"Student" (1908). On the probable error of the mean. *Biometrika*. **6** 1-25.

**Définition 4.4.1.** Soient  $U \stackrel{d}{=} N(0, 1)$  et  $V \stackrel{d}{=} \chi_n^2$  deux variables aléatoires indépendantes, avec  $n \geq 1$ . Par définition, une variable aléatoire  $T$  suit une loi de Student à  $n$  degrés de liberté, si

$$T \stackrel{d}{=} \frac{U}{\sqrt{V/n}}, \quad (4.4.1)$$

ce qui est noté  $T \stackrel{d}{=} t_n$ .

Si  $T$  suit une loi de Student à  $n$  degrés de liberté, alors  $T^2 \stackrel{d}{=} F_{1, n}$ . De ce fait, pour tout  $0 < \alpha \leq 1/2$ , le quantile supérieur d'ordre  $\alpha$ , noté  $t_{n; \alpha}$  de la loi de Student à  $n$  degré de liberté est lié au quantile supérieur d'ordre  $\alpha/2$ , noté  $f_{1, n; \alpha/2}$  de la loi de Fisher  $F_{1, n}$  par l'identité

$$t_{n; \alpha}^2 = f_{1, n; \alpha/2}. \quad (4.4.2)$$

On vérifie aisément, par la loi des grands nombres, que, si  $\nu_\alpha$  désigne le quantile supérieur d'ordre  $\alpha$  de la loi normale  $N(0, 1)$  (tel que  $\Phi(\nu_\alpha) = 1 - \alpha$ ), on a

$$t_{\infty; \alpha} = \lim_{n \rightarrow \infty} t_{n; \alpha} = \nu_\alpha. \quad (4.4.3)$$

En fait, il est possible de montrer que les inégalités suivantes sont toujours satisfaites, pour tout  $n \geq 1$  et  $0 < \alpha \leq \frac{1}{2}$  (voir Johnson, Kotz et Balakrishnan (1995), p.364). On a

$$t_{n; \alpha} > \nu_\alpha > 0 \quad \text{et} \quad t_{n; 1-\alpha} = -t_{n; \alpha} < -\nu_\alpha < 0. \quad (4.4.4)$$

L'importance des inégalités (4.4.4) est considérable, en pratique. Elles impliquent que les intervalles de confiance de Student pour la moyenne (voir le §5.2), de la forme

$$\mathbb{P}\left(\mu \in \left[\bar{X} - \frac{s}{\sqrt{n}} t_{n-1;\alpha/2}, \bar{X} + \frac{s}{\sqrt{n}} t_{n-1;\alpha/2}\right]\right) = 1 - \alpha,$$

sont systématiquement plus étendus que les intervalles de confiance asymptotiques, de la forme

$$\mathbb{P}\left(\mu \in \left[\bar{X} - \frac{s}{\sqrt{n}} \nu_{\alpha/2}, \bar{X} + \frac{s}{\sqrt{n}} \nu_{\alpha/2}\right]\right) \approx 1 - \alpha.$$

On a, en effet, pour tout  $n \geq 1$  et  $0 < \alpha \leq 1$ ,

$$\left[\bar{X} - \frac{s}{\sqrt{n}} \nu_{\alpha/2}, \bar{X} + \frac{s}{\sqrt{n}} \nu_{\alpha/2}\right] \subset \left[\bar{X} - \frac{s}{\sqrt{n}} t_{n-1;\alpha/2}, \bar{X} + \frac{s}{\sqrt{n}} t_{n-1;\alpha/2}\right]. \quad (4.4.5)$$

Johnson, N., Kotz, S., Balakrishnan, N. (1995). *Continuous Univariate Distributions*. Vol.2, Second Edition. Wiley, New York.

n	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 2.5\%$	$\alpha = 1\%$
1	3.0777	6.3138	12.7062	31.8207
2	1.8856	2.9200	4.3027	6.9646
3	1.6377	2.3534	3.1824	4.5407
4	1.5332	2.1318	2.7764	3.7469
5	1.4759	2.0150	2.5706	3.3649
6	1.4398	1.9432	2.4469	3.1427
7	1.4149	1.8946	2.3646	2.9980
8	1.3968	1.8595	2.3060	2.8965
9	1.3830	1.8331	2.2622	2.8214
10	1.3722	1.8125	2.2281	2.7638
11	1.3634	1.7559	2.2010	2.7181
12	1.3562	1.7823	2.1788	2.6810
13	1.3502	1.7709	2.1604	2.6503
14	1.3450	1.7613	2.1448	2.6245
15	1.3406	1.7531	2.1315	2.6025
20	1.3253	1.7247	2.0860	2.5280
25	1.3163	1.7081	2.0595	2.4851
30	1.3104	1.6973	2.0423	2.4573
40	1.3031	1.6839	2.0211	2.4233
60	1.2958	1.6706	2.0003	2.3901
120	1.2886	1.6577	1.9784	2.3578
$\infty$	1.2816	1.6449	1.9600	2.3263

**Table des quantiles supérieurs  $t_{n;\alpha}$  de la loi de Student à  $n$  degrés de liberté.**

Faisant usage de la proposition 4.3.1 et de (4.4.2), on obtient la proposition suivante, qui permet un calcul numérique aisé des quantiles de la loi de Student.

**Proposition 4.4.1.** *Pour tout  $n \geq 1$ , et  $t \geq 0$ , si  $T \stackrel{d}{=} t_n$ , on a l'identité*

$$\mathbb{P}(T \leq t) = 1 - \frac{1}{\beta\left(\frac{1}{2}n, \frac{1}{2}\right)} \int_0^{n/(n+t^2)} x^{\frac{1}{2}n-1} (1-x)^{\frac{1}{2}-1} dx. \quad (4.4.6)$$

**Preuve.** Omise (voir p. 364 dans Johnson, Kotz et Balakrishnan (1995) pour plus de détails).□

**Exercice 4.4.1.** On note  $\tan^{-1}(u) = \text{arctg}(u)$  la fonction Arc Tangente (inverse de la fonction tangente  $\tan$ ).

1°) Montrer que, si  $T_n \stackrel{d}{=} t_n$ , alors, pour  $t \in \mathbb{R}$ ,

$$\mathbb{P}(T_1 \leq t) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}(t). \quad (4.4.7)$$

2°) Vérifier les formules suivantes, pour  $n = 2$  et  $n = 3$ . On pose  $\theta = \tan^{-1}(t/\sqrt{n})$ .

$$\mathbb{P}(T_{2k} \geq t) = \frac{1}{2} + \frac{1}{2} \left\{ 1 + \sum_{j=0}^{k-1} \frac{(2j)!}{2^{2j} \{j!\}^2} \cos^{2j} \theta \right\} \sin \theta \quad \text{si } n = 2k \geq 2, \quad (4.4.8)$$

$$\mathbb{P}(T_{2k+1} \geq t) = \frac{1}{2} + \frac{1}{\pi} \left\{ \theta + \sum_{j=0}^{k-1} \frac{2^{2j} \{j!\}^2}{(2j+1)!} \cos^{2j+1} \theta \right\} \sin \theta \quad \text{si } n = 2k+1 \geq 3. \quad (4.4.9)$$

Voir p. 365 dans Johnson, Kotz et Balakrishnan (1995), *Op. cit.*, pour plus de détails

#### 4.4.2 B. Loi de Student non centrée.

Soit  $\theta \in \mathbb{R}$  une constante arbitraire, et  $n \geq 1$ . La v.a.  $Z$  suit une loi de Student non centrée à  $N$  degrés de liberté, de coefficient de décentrement  $\theta$ , ce qui est noté  $Z \stackrel{d}{=} t_N(\theta)$ , si  $Z \stackrel{d}{=} U/V$ , où  $U, V$  sont des v.a. indépendantes, de lois respectives  $U \stackrel{d}{=} N(\theta, 1)$ , et  $NV^2 \stackrel{d}{=} \chi_N^2$ .

Il est clair que si  $Z \stackrel{d}{=} t_N(\theta)$ , alors  $Z^2 \stackrel{d}{=} F_{1,N}(\theta^2)$ . Ceci permet de ramener l'étude numérique des lois de Student décentrées à celles des lois de Fisher décentrées. Il est à noter, toutefois, que la définition du paramètre de décentrement diffère, dans chaque cas.





## Chapitre 5

# Applications Statistiques.

### 5.1 Estimation et tests.

Nous faisons ici quelques brefs rappels de statistique classique, en vue des utilisations qui vont suivre. Nous nous plaçons dans le cas où la suite de vecteurs  $X_1, \dots, X_n$  est un échantillon aléatoire observé de taille  $n$ , composé d'une suite de  $n$  répliques indépendantes de même loi d'un vecteur aléatoire  $X \in \mathbb{R}^d$ . Pour simplifier, nous supposons que  $X$  possède une densité  $f(x; \theta)$  relativement à la mesure de Lebesgue (notée  $dx$ ) dans  $\mathbb{R}^d$  (en fait, on peut remplacer la mesure de Lebesgue par une mesure différente, telle que, par exemple, la mesure discrète qui affecte la masse 1 à chaque point de  $\mathbb{R}^d$  de coordonnées entières. Nous ne rentrerons pas ici dans ce cadre élargi par souci de simplification de l'exposé). Nous avons donc, pour tout  $A \in \mathcal{B}_d$ ,

$$\mathbb{P}(X \in A) = \int_A f(x; \theta) dx. \quad (5.1.1)$$

Ici,  $\theta \in \Theta \in \mathbb{R}^p$  est un *paramètre* inconnu. On appelle *estimateur*  $\hat{\theta}$  de  $\theta$  une *statistique*  $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n) \in \mathbb{R}^p$ , fonction mesurable de  $n$  et des observations  $X_1, \dots, X_n$ , dont le but est de fournir une évaluation numérique de la quantité inconnue  $\theta$  à partir des données de l'échantillon. Sans autre précision, un estimateur de  $\theta$  peut être à peu près n'importe quoi. Son intérêt dépend de propriétés additionnelles qu'on lui impose dans un but d'efficacité, au sens large du terme. Parmi celles-ci, on retiendra les suivantes. L'estimateur  $\hat{\theta}_n$  de  $\theta$  est dit (en notant  $\|\cdot\|$  la norme euclidienne dans  $\mathbb{R}^p$ ) :

- *consistant*, si  $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$  lorsque  $n \rightarrow \infty$ , c'est à dire, si  $\mathbb{P}(\|\hat{\theta}_n - \theta\| > \varepsilon) \rightarrow 0$  pour tout  $\varepsilon > 0$ ;
- *sans biais*, si  $\mathbb{E}(\hat{\theta}_n) = \theta$ , indépendamment de la valeur de  $\theta \in \Theta$ .

Il est utile de comparer les performances de deux estimateurs en faisant usage de leurs matrices de variances-covariances respectives (supposées existantes). Nous dirons que l'estimateur  $\hat{\theta}$  est *plus concentré* que l'estimateur  $\tilde{\theta}$ , si

$$\text{Var}(\hat{\theta}) \leq \text{Var}(\tilde{\theta}), \quad (5.1.2)$$

ou, ce qui revient au même, si, pour tout  $u \in \mathbb{R}^p$ ,

$$\text{Var}(u'\hat{\theta}) = u'\text{Var}(\hat{\theta})u \leq u'\text{Var}(\tilde{\theta})u = \text{Var}(u'\tilde{\theta}). \quad (5.1.3)$$

Un estimateur sans biais  $\hat{\theta}$  de  $\theta$  satisfait, sous des hypothèses générales de régularité, l'*inégalité de Cramér-Rao*, supposant l'existence et la régularité de la *matrice d'information de Fisher*  $I(\theta)$ . Pour définir celle-ci, il est nécessaire d'introduire quelques concepts utiles. Tout d'abord, la *vraisemblance* de l'échantillon est la fonction

aléatoire du paramètre inconnu

$$\theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_p \end{pmatrix},$$

définie par

$$L(\theta) = f(\mathbf{X}; \theta) = \prod_{i=1}^n f(X_i; \theta), \quad (5.1.4)$$

en faisant usage des notations  $\mathbf{X} = [X_1 \ \dots \ X_n]^\top \in \mathbb{R}^n$ , et  $f(\mathbf{X}; \theta) = \prod_{i=1}^n f(X_i; \theta)$ . La *matrice d'information de Fisher* est alors la matrice  $(p \times p)$  définie (sous réserve d'existence des moments appropriés) par

$$I(\theta) = \begin{pmatrix} \vdots & & \\ \dots & -\mathbb{E} \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log L(\theta) \right) & \dots \\ \vdots & & \end{pmatrix}. \quad (5.1.5)$$

On notera que  $I(\theta)$  s'exprime de manière simple à partir de la densité  $f(x; \theta)$ . On a, en effet,

$$I(\theta) = nI_0(\theta) = n \begin{pmatrix} \vdots & & \\ \dots & -\mathbb{E} \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(X, \theta) \right) & \dots \\ \vdots & & \end{pmatrix}, \quad (5.1.6)$$

où  $X \stackrel{d}{=} X_1$  est une variable générique, de même loi que  $X_1, \dots, X_n$ .

On a alors le résultat suivant. Les hypothèses de validité du théorème ci-dessous ne seront pas explicitées. On comprendra mieux leur nature en détaillant la démonstration simplifiée qui en est présentée plus loin.

**Théorème 5.1.1.** *Sous réserve de régularité et d'existence des expressions correspondantes, tout estimateur sans biais,  $\hat{\theta}$  de  $\theta$  vérifie*

$$\text{Var}(\hat{\theta}) \geq I(\theta)^{-1} = n^{-1}I_0(\theta)^{-1}. \quad (5.1.7)$$

**Preuve.** Nous donnons une version simplifiée de la démonstration pour  $p = 1$ . Celle-ci donnera également des indications utiles sur les conditions de régularité requises dans le théorème. On suppose que l'estimateur  $\hat{\theta} = \hat{\theta}(\mathbf{X})$  est fonction de l'observation  $\mathbf{X} = [X_1 \ \dots \ X_n]^\top \in \mathbb{R}^n$ , qui, elle-même, possède une densité  $f(x; \theta)$  dans  $\mathbb{R}^n$ . Dans ce cas, on a  $L(\theta) = f(\mathbf{X})$  ainsi que les identités

$$\begin{aligned} \mathbb{E}(1) &= \int_{\mathbb{R}^n} f(x; \theta) dx = 1, \\ \mathbb{E}(\hat{\theta}) &= \int_{\mathbb{R}^n} \hat{\theta}(x) f(x; \theta) dx = \theta, \end{aligned}$$

ce qui, en dérivant sous le signe somme par rapport à  $\theta$  (ce qui impose des conditions de régularité rendant cette opération possible), implique que

$$\mathbb{E} \left( \frac{d}{d\theta} \log f(\mathbf{X}; \theta) \right) = \int_{\mathbb{R}^n} \frac{1}{f(x; \theta)} \left\{ \frac{\partial}{\partial \theta} f(x; \theta) \right\} f(x; \theta) dx = \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta} f(x; \theta) dx = 0, \quad (5.1.8)$$

$$\mathbb{E} \left( \hat{\theta}(\mathbf{X}) \frac{d}{d\theta} \log f(\mathbf{X}; \theta) \right) = \int_{\mathbb{R}^n} \hat{\theta}(x) \frac{\partial}{\partial \theta} f(x; \theta) dx = 1, \quad (5.1.9)$$

$$\mathbb{E} \left( \frac{d^2}{d\theta^2} \log f(\mathbf{X}; \theta) \right) + \mathbb{E} \left( \left\{ \frac{d}{d\theta} \log f(\mathbf{X}; \theta) \right\}^2 \right) = \int_{\mathbb{R}^n} \frac{\partial^2}{\partial \theta^2} f(x; \theta) dx = 0. \quad (5.1.10)$$

On notera également que les opérations ci-dessus supposent implicitement que le *support* des lois considérées (c'est à dire, l'adhérence de l'ensemble  $\{x : f(x; \theta) > 0\}$ ), ne dépend pas du paramètre  $\theta$ . En combinant les relations (5.1.8)–(5.1.9), on obtient que

$$\mathbb{E} \left( \left\{ \hat{\theta}(\mathbf{X}) - \theta \right\} \frac{d}{d\theta} \log f(\mathbf{X}; \theta) \right) = 1.$$

On applique ensuite l'inégalité de Cauchy-Schwarz à cette relation, pour aboutir à

$$\mathbb{E} \left( \left\{ \hat{\theta}(\mathbf{X}) - \theta \right\}^2 \right) \mathbb{E} \left( \left\{ \frac{d}{d\theta} \log f(\mathbf{X}; \theta) \right\}^2 \right) \geq 1.$$

Finalement, l'observation, par (5.1.10), que

$$\mathbb{E} \left( \left\{ \frac{d}{d\theta} \log f(\mathbf{X}; \theta) \right\}^2 \right) = -\mathbb{E} \left( \frac{d^2}{d\theta^2} \log f(\mathbf{X}; \theta) \right),$$

montre que

$$\text{Var}(\hat{\theta}) \geq \left\{ -\mathbb{E} \left( \frac{d^2}{d\theta^2} \log f(\mathbf{X}; \theta) \right) \right\}^{-1},$$

ce qui est la version cherchée de l'inégalité (5.1.7) de Cramér-Rao. On remarquera au passage que les conditions de régularité justifiant la démonstration se limitent, pour l'essentiel, à la justification des dérivations successives sous le signe somme ayant été utilisées.  $\square$

Une des contributions majeures du statisticien anglais R. A. Fisher (1890-1962) est d'avoir introduit, à partir de 1923, la *méthode du maximum de vraisemblance*, qui fournit, sous des hypothèses assez générales, des estimateurs atteignant asymptotiquement la borne (5.1.7) de Cramér-Rao. C'est le statisticien français Daniel Dugué (1912-1987) qui, dans sa thèse, publiée en 1937 (Dugué, D. (1937). Applications de la limite au sens du calcul des probabilités à l'étude de diverses questions d'estimation. *Journal de l'Ecole Polytechnique*. pp.305-374), démontra, pour la première fois, la normalité asymptotique et l'efficacité des estimateurs du maximum de vraisemblance. Ces propriétés généralisées par la suite par d'autres auteurs, sont résumées dans les théorèmes 5.1.1-5.1.2-5.1.3 du présent paragraphe.

L'estimateur du maximum de vraisemblance  $\hat{\theta}$  de  $\theta$  est, par définition, une statistique telle que

$$L(\hat{\theta}) = \sup_{\theta \in \Theta} L(\theta), \tag{5.1.11}$$

où la *vraisemblance*  $L(\theta)$  est définie par (5.1.4). On notera au passage que l'existence de  $\hat{\theta}$  n'est pas garantie, et que, même dans ce cas, et lorsque l'équation (5.1.11) a un sens, l'estimateur du maximum de vraisemblance  $\hat{\theta}$  peut ne pas être défini de manière unique. Toutefois, sous des hypothèses assez générales (que nous ne détaillerons pas ici), il a une loi limite asymptotique normale centrée, ce qui est exprimé dans le théorème suivant.

**Théorème 5.1.2.** *Sous réserve de conditions de régularité et d'existence des expressions correspondantes, l'estimateur du maximum de vraisemblance  $\hat{\theta}$  de  $\theta$  vérifie asymptotiquement, lorsque la taille  $n$  de l'échantillon tend vers l'infini, la convergence en loi*

$$n^{1/2}(\hat{\theta} - \theta) \xrightarrow{d} N_p(\mathbb{O}, I_0(\theta)^{-1}). \tag{5.1.12}$$

La notion de vraisemblance intervient aussi de manière utile dans la théorie des *tests d'hypothèses*. Nous considérerons essentiellement ici le cas où  $\theta \in \mathbb{R}^p$  se décompose sous la forme

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix},$$

où  $\theta_1 \in \mathbb{R}^q$  et  $\theta_2 \in \mathbb{R}^{p-q}$  sont des paramètres composants de  $\theta$ . Le problème se pose alors, pour  $\theta_{10} \in \mathbb{R}^q$  spécifié, de tester l'hypothèse

(H.0)  $\theta_1 = \theta_{10}$ , contre (H.1)  $\theta_1 \neq \theta_{10}$ .

Pour cela, on se sert de la vraisemblance, écrite, avec des notations évidentes, sous la forme

$$L(\theta_1, \theta_2) = \prod_{i=1}^n f(X_i; \theta_1, \theta_2),$$

pour définir la statistique du *rapport de vraisemblance*

$$\mathcal{R} = \frac{\sup_{\theta_2} L(\theta_{10}, \theta_2)}{\sup_{\theta_1, \theta_2} L(\theta_1, \theta_2)}. \quad (5.1.13)$$

Sous des hypothèses de régularité générales, que nous ne préciserons pas, le théorème suivant est vrai.

**Théorème 5.1.3.** *Sous réserve de conditions générales de régularité portant sur  $f$ , sous (H.0), on a la convergence en loi, lorsque  $n \rightarrow \infty$ ,*

$$-2 \log \mathcal{R} \xrightarrow{d} \chi_p^2. \quad (5.1.14)$$

La vérification de la validité des résultats ci-dessus, dans chaque cas particulier d'intérêt, a fait l'objet de très nombreux travaux de recherche, au cours des cinquante dernières années. Les calculs correspondants étant souvent fastidieux, l'habitude des statisticiens appliqués est d'appliquer les formules asymptotiques correspondantes, un peu comme un principe physique, sans prendre toujours le soin d'en vérifier l'exactitude. Il faut demeurer conscient de l'existence d'exceptions notables à ces principes généraux. L'exemple le plus simple en est donné par un échantillon  $X_1, \dots, X_n$  de la loi uniforme sur  $[0, \theta]$ , pour lequel l'estimateur du maximum de vraisemblance est donné par  $\hat{\theta} = \max(X_1, \dots, X_n)$ . On vérifie aisément dans cet exemple que  $n(\theta - \hat{\theta})$  converge, lorsque  $n \rightarrow \infty$ , vers une loi exponentielle. On est loin de la loi normale, et, de plus, la convergence a lieu avec le facteur  $n$ , au lieu de  $\sqrt{n}$ . Lorsque tout se passe bien, on dit que l'estimateur  $\hat{\theta}$  est *régulier*, et on parlera de *non régularité* lorsque ce n'est pas le cas. Il importe donc de vérifier la régularité de  $\hat{\theta}$  dans chaque application particulière.

## 5.2 Utilisation de la loi de Student.

### 5.2.1 A. Généralités.

**Définition 5.2.1.** *Soit  $U, V$  deux variables aléatoires indépendantes telles que  $U \stackrel{d}{=} N(0, 1)$  et  $NV^2 \stackrel{d}{=} \chi_N^2$  (avec  $V > 0$ ). Par définition, le rapport*

$$T = \frac{U}{V} \stackrel{d}{=} \frac{N(0, 1)}{\sqrt{\chi_N^2/N}}, \quad (5.2.1)$$

*suit la loi de Student à  $N$  degrés de liberté, ce qui est noté symboliquement par  $T \stackrel{d}{=} t_N$ .*

Pour tout  $0 < \alpha < 1$ , on désigne par  $t_{N;\alpha}$  le quantile supérieur d'ordre  $\alpha$  de la loi de Student à  $N$  degrés de liberté, c'est à dire, tel que, si  $T \stackrel{d}{=} t_N$ ,

$$\mathbb{P}(T \geq t_{N;\alpha}) = \alpha. \quad (5.2.2)$$

**Remarque 5.2.1.** L'approximation suivante (due à Peiser, A. M. (1943). Asymptotic formulas for significance levels of certain distributions. *Ann. Math. Statist.* **14** 56–62, et Ibid. Correction (1949), **20** 128–129), est particulièrement commode pour évaluer les niveaux critiques de la loi de Student. Si  $\nu_\alpha$  désigne la quantile supérieure d'ordre  $\alpha$  de la loi normale  $N(0, 1)$ , alors

$$t_{N;\alpha} \simeq \nu_\alpha + \frac{1}{4N} \{ \nu_\alpha^3 + \nu_\alpha \}. \quad (5.2.3)$$

La loi de Student est particulièrement utile pour établir des *intervalles de confiance* pour la moyenne.

**Exemple 5.2.1.** Soit  $X_1, \dots, X_n$  un échantillon de  $n \geq 2$  variables aléatoires indépendantes de même loi normale  $N(\mu, \sigma^2)$  (avec  $\sigma > 0$ ). On estime  $\mu$  et  $\sigma^2$  par les moments empiriques  $\hat{\mu}$  et  $\hat{\sigma}^2$  (avec  $\hat{\sigma} > 0$  p.s.), définis par

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2. \quad (5.2.4)$$

L'estimateur  $\hat{\sigma}^2$  de  $\sigma^2$  n'est pas centré. On le remplace avantageusement pour certaines applications par sa variante centrée  $s^2$  (avec  $s \geq 0$ ), définie par

$$s^2 = \frac{n\hat{\sigma}^2}{n-1} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{qui est tel que} \quad \mathbb{E}(s^2) = \sigma^2 \quad \text{pour} \quad n \geq 2. \quad (5.2.5)$$

Sous ces hypothèses,  $\bar{X}$  et  $s^2$  sont indépendants, de lois respectives données par

$$\sqrt{n}(\bar{X} - \mu) \stackrel{d}{=} N(0, \sigma^2) \quad \text{et} \quad \frac{(n-1)s^2}{\sigma^2} \stackrel{d}{=} \chi_{n-1}^2. \quad (5.2.6)$$

Par conséquent,

$$T_{n-1}^2 = \frac{\sqrt{n}(\bar{X} - \mu)}{s} \stackrel{d}{=} t_{n-1}, \quad (5.2.7)$$

suit, sous l'hypothèse de normalité de l'échantillon, *exactement* une loi de Student à  $n-1$  degrés de liberté. Ceci peut alors être utilisé pour donner un intervalle de confiance exact à l'ordre  $1 - \alpha$  pour la moyenne (inconnue)  $\mu$ . Ce dernier est de la forme

$$\mathbb{P} \left( \mu \in \left[ \bar{X} - \frac{s}{\sqrt{n}} t_{n-1}^{\alpha/2}, \bar{X} + \frac{s}{\sqrt{n}} t_{n-1}^{\alpha/2} \right] \right) = 1 - \alpha. \quad (5.2.8)$$

**Exemple 5.2.2.** Le tableau ci-dessus fournit un échantillon simulé de taille  $n = 20$  d'une loi normale standard  $N(0, 1)$ . Il s'agit de vérifier si les résultats obtenus sont bien représentatifs, en procédant à diverses vérifications statistiques. Nous réalisons ici une partie de ces opérations en construisant un intervalle de confiance à 95% pour la moyenne  $\mu$  commune de ces observations.

-0.029881	-0.708300	-0.336880	-0.059962	-0.962360
1.383800	-1.043900	1.078500	1.584300	-0.062007
-0.503330	-0.810920	1.077800	-0.501430	0.561640
-1.031500	0.151550	0.859030	-0.156390	2.213200

Désignant ces observations par  $X_1, \dots, X_n$  avec  $n = 20$ , on constate que

$$\begin{aligned} \min_{1 \leq i \leq n} X_i &= -1.0439, & \max_{1 \leq i \leq n} X_i &= 2.2132, \\ \bar{X} &= 0.135137, & \hat{\sigma}^2 &= 0.907844, & \hat{\sigma} &= 0.952808, \\ \sqrt{\tilde{\beta}_1} &= 0.637846, & \tilde{\beta}_2 - 3 &= -0.731732, \\ \hat{\nu}_1 &= \frac{1}{n} \sum_{i=1}^n |X_i - \bar{X}| = 0.78286. \end{aligned}$$

Le logiciel Statistica 6.1 (Copyright © StatSoft, Inc. 1984-2003) fournit la valeur  $t_{19;2.5\%} \simeq 2.093024$ . On obtient donc, dans cet exemple précis,

$$\begin{aligned} \frac{s}{\sqrt{n}} t_{n-1;\alpha/2} &= \left\{ \frac{n}{n-1} \right\}^{1/2} \frac{\hat{\sigma}}{\sqrt{n}} t_{n-1;\alpha/2} = \frac{\hat{\sigma}}{\sqrt{n-1}} t_{n-1;\alpha/2} \\ &\simeq \left\{ \frac{0.86674}{19} \right\}^{1/2} \times 2.093024 \simeq 0.447035, \end{aligned}$$

ce qui donne un intervalle de confiance (en arrondissant à la troisième décimale) pour la moyenne  $\mu$  des variables,

$$\mathbb{P}(\mu \in [-0.196, 0.675]) \simeq 95\%.$$

Ce dernier montre que, pour ce qui est de la moyenne, la simulation est correcte, puisque l'hypothèse  $\mu = 0$  n'est pas rejetée au seuil  $\alpha = 5\%$ .

Il est intéressant, dans cet exemple, de se poser la question suivante. La valeur maximale (en valeur absolue) 2.213200 observée est-elle acceptable pour un tel échantillon? On peut faire le raisonnement suivant. Une valeur  $z$  ayant une probabilité  $\alpha = 5\%$  d'être dépassée par  $\max_{1 \leq i \leq n} |X_i|$  est telle que

$$\Phi(z)^n = 1 - \alpha \quad \Leftrightarrow \quad \Phi(z) = (1 - \alpha)^{1/n}.$$

Pour  $n = 20$  et  $\alpha = 5\%$ , on trouve  $z = 3.019709$ , ce qui montre que la valeur observée de  $2.213200 < z$  est parfaitement acceptable, au sens du critère précédent.

### 5.2.2 B. Calculs de puissance et de tailles d'échantillons.

L'utilisation d'un logiciel de calcul s'avère extrêmement commode pour réaliser ces opérations qui, autrefois, nécessitaient un emploi de tables long et fastidieux. Toutefois, avant de se lancer dans des manipulations de ce type, il importe de consacrer tout le temps nécessaire à la lecture approfondie du mode d'emploi du logiciel utilisé (ce dernier n'étant parfois seulement disponible qu'au prix d'une connexion internet). Cet apprentissage initial doit permettre à l'utilisateur de saisir des notations et conventions, qui peuvent varier sensiblement d'un logiciel à l'autre.

Pour faciliter la compréhension du sujet, nous donnons ci-dessous des exemples de calculs de puissance et de tailles d'échantillons réalisés avec le logiciel Statistica 6.1 (Copyright © StatSoft, Inc. 1984-2003). Les valeurs numériques traitées sont artificielles, et ne servent qu'à illustrer le procédé. Nous considérons ainsi un échantillon  $X_1, \dots, X_n$  de taille  $n = 25$  pour lequel la moyenne empirique, dite *moyenne de la population*, est

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = 3.1,$$

et l'écart-type empirique, dit *écart-type de la population*,

$$\hat{\sigma} = \left\{ \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right\}^{1/2} = 8.5.$$

On suppose que l'échantillon est issu d'une loi normale  $N(\mu, \sigma^2)$ , et on désire tester, par un test de Student bilatéral, l'hypothèse ( $H_0$ ) que  $\mu = \mu_0 = 0.0$ , avec un seuil  $\alpha = 5\%$ . On désire évaluer les performances du test lorsque la *moyenne vraie*  $\mu$  est égale à  $\mu = 2.0$ , et l'*écart-type exact*  $\sigma$  est  $\sigma = 10.0$ . Par ailleurs, avec cette même configuration de paramètres, on souhaite connaître la taille  $N$  d'échantillon nécessaire pour obtenir une probabilité  $1 - \beta = 65\%$  de rejeter ( $H_0$ ) par un test de Student bilatéral de seuil  $\alpha = 5\%$ .

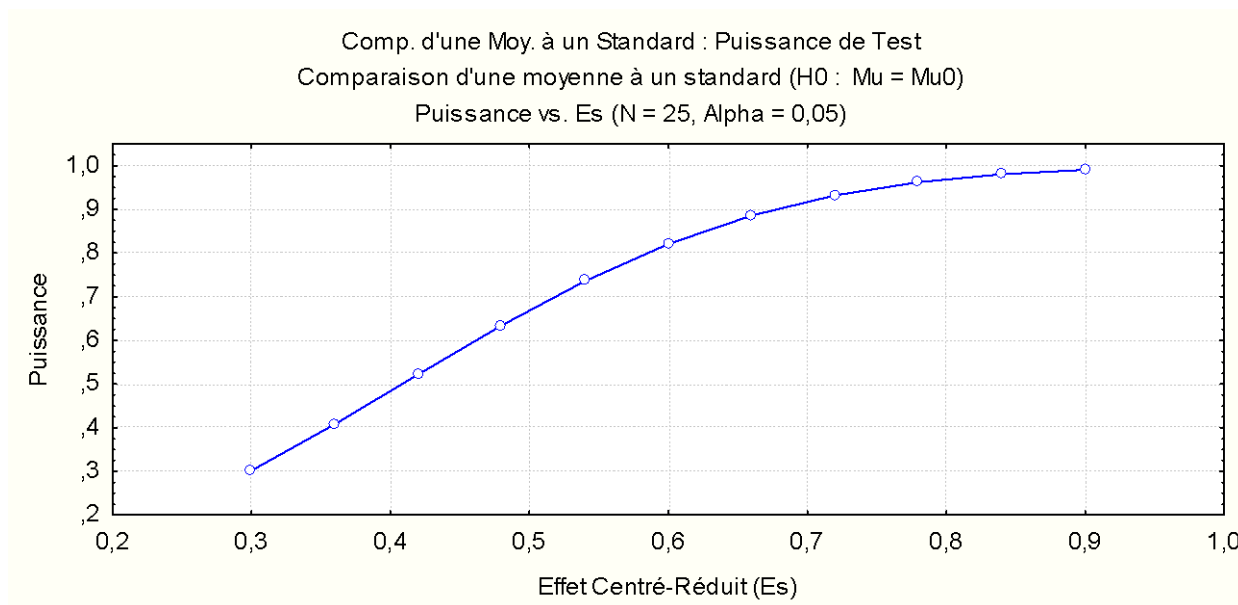
On obtient (par le logiciel) la valeur de  $t_{n-1; \alpha/2} = t_{24; 2.5\%} \simeq 2.063918$  qui permet de construire le test cherché. Le principe du test consiste à rejeter ( $H_0$ ) pour les valeurs observées de

$$\sqrt{n} |\bar{X} - \mu_0| \geq \left\{ \frac{n}{n-1} \right\}^{1/2} \hat{\sigma} t_{n-1; \alpha/2} = \left\{ \frac{25}{24} \right\}^{1/2} \times 10.0 \times 2.063918 \simeq 17.905059.$$

Le seuil  $\alpha = 5\%$  du test correspond donc à la probabilité

$$\mathbb{P}(\sqrt{n} |\bar{X} - \mu_0| \geq 17.905059 \mid (H_0)) \simeq \alpha = 5\%.$$

Dans le cas présent,  $n = 25$ ,  $\bar{X} = 3.1$  et  $\mu_0 = 0$ , de sorte que  $\sqrt{n} |\bar{X} - \mu_0| = 15.5$ , et le test *accepte* ( $H_0$ ), au seuil  $\alpha = 5\%$ .

FIGURE 5.1 – Puissance fonction de l'effet centré réduit  $\delta$ .

La valeur de  $\alpha$  représente le *risque de première espèce*, c'est à dire, la probabilité de rejeter l'hypothèse ( $H_0$ ) que  $\mu = \mu_0$  alors que celle-ci est vraie. La *puissance* du test représente inversement la probabilité  $1 - \beta$  de rejeter l'hypothèse ( $H_0$ ) lorsqu'elle est fausse. Ce n'est plus un nombre mais une fonction des paramètres du modèle. Plus précisément, ici,

$$1 - \beta = \mathbb{P} \left( \sqrt{n} |\bar{X} - \mu_0| \geq \left\{ \frac{n}{n-1} \right\}^{1/2} \hat{\sigma} t_{n-1; \alpha/2} \mid \mu, \mu_0, \sigma \right).$$

Pour évaluer cette quantité, on observe que

$$t = \frac{\sqrt{n} \{\bar{X} - \mu_0\}}{\hat{\sigma} \sqrt{n/(n-1)}} \stackrel{d}{=} t_{n-1} \left( \frac{\mu - \mu_0}{\sigma} \right),$$

suit une loi de Student non centrée à  $n - 1$  degrés de liberté, et de paramètre de décentrement

$$\theta = \frac{\mu - \mu_0}{\sigma}.$$

Les valeurs numériques fournies par les observations et le modèle postulé est sont de

$$t = \frac{\sqrt{25} \{3.1 - 0.0\}}{8.5 \times \sqrt{25/24}} \simeq 1.786687 \quad \text{et} \quad \theta = \frac{2.0 - 0.0}{10} = 0.20.$$

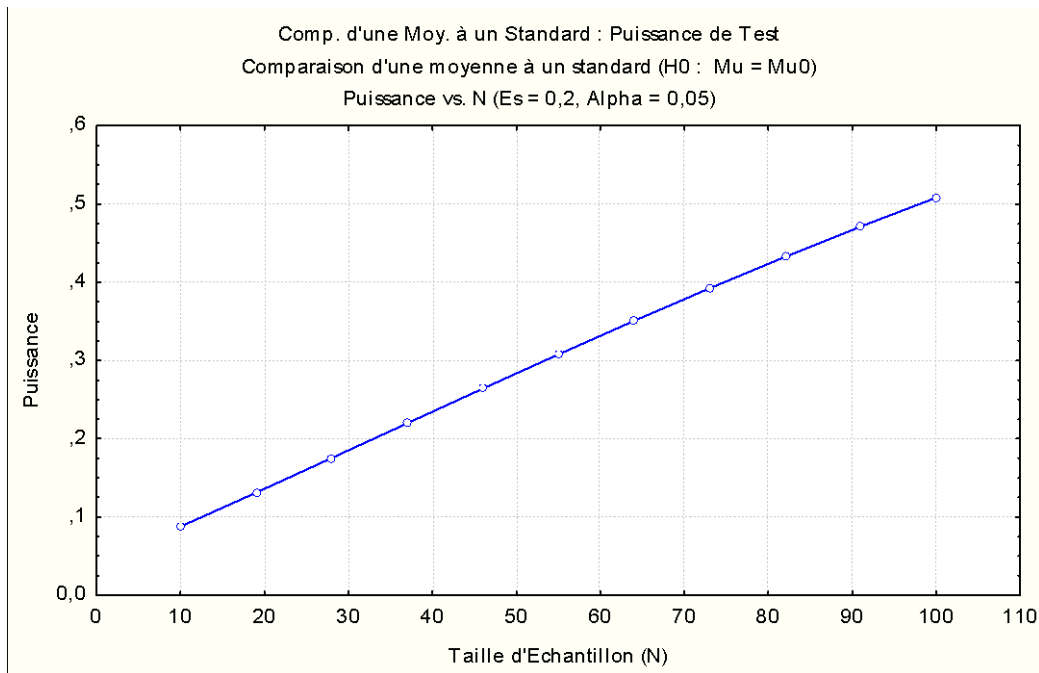
On voit apparaître ici une quantité  $\theta = \frac{\mu - \mu_0}{\sigma}$  communément appelée *effet centré réduit*. Adoptons la notation  $T_{n-1}(\theta)$  pour désigner une variable aléatoire de loi de Student non centrée, à  $n - 1$  degrés de liberté, de paramètre de décentrement (ou un *effet centré réduit*)  $\theta$ . La puissance du test est alors donnée par la formule

$$1 - \beta = \mathbb{P} (|T_{n-1}(\theta)| \geq t_{n-1; \alpha/2}). \quad (5.2.9)$$

On observera que, pour  $\mu = \mu_0$ ,  $\theta = 0$  et  $1 - \beta = \alpha$ . Dans l'alternative où  $\mu \neq \mu_0$ ,  $\beta$  représente le *risque de deuxième espèce*, consistant à accepter à tort l'hypothèse ( $H_0$ ). D'une manière générale, on peut considérer  $1 - \beta$  dans l'expression (5.2.9), comme une fonction de :



- L'effet centré réduit  $\theta = \frac{\mu - \mu_0}{\sigma}$ , pour  $n$  et  $\alpha$  fixés (voir la Figure 5.1) ;
- La taille  $n$  de l'échantillon, pour  $\theta = \frac{\mu - \mu_0}{\sigma}$  et  $\alpha$  fixés (voir la Figure 5.2) ;
- Le seuil  $\alpha$  du test, pour  $\theta = \frac{\mu - \mu_0}{\sigma}$  et  $n$  fixés (voir la Figure 5.3).

FIGURE 5.2 – Puissance fonction de la taille  $n$  d'échantillon.

Le logiciel fournit un tableau de la forme suivante pour afficher les paramètres du calcul de puissance.

```

Comp. d'une Moy. à un Standard : Puissance de Test
H0 : Mu = Mu0
Risque 1ère espèce (Alpha) : 0,05
Taille d'Ech. Groupe (N) : 25
Moyenne sous H0 (Mu0) : 0
Moyenne Population (Mu) : 2
Ecart-Type Population (Sigma) : 10
Effet Centré-Réduit (Es) : 0,2

```

Un deuxième tableau, voisin du premier, fournit le calcul de puissance pour les valeurs de  $\mu$  et  $\sigma$  choisies.

```

Comp. d'une Moy. à un Standard
Moyenne sous H0 (Mu0) : 0,0000
Moyenne Population (Mu) : 2,0000
Ecart-type Population (Sigma) : 10,0000
Effet centré réduit (Es) : 0,2000
Taille d'Ech. Groupe (N) : 25
Risque 1ère Espèce (Alpha) : 0,0500
Valeur Critique de t : 2,0639
Puissance : 0,1606

```

On constate ici que le test a une probabilité  $\alpha = 5\%$  de rejeter ( $H_0$ ) lorsque  $\mu = 0,0$ , mais seulement une probabilité de 16,06% de rejeter cette hypothèse lorsque  $\mu = 2,0$ . La conclusion est que, pour une taille d'échantillon

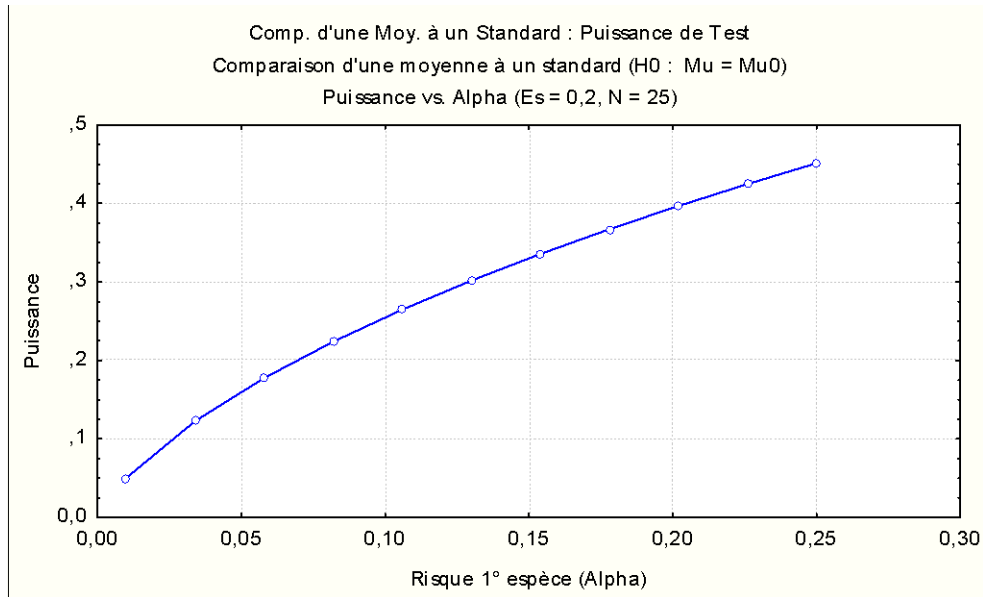


FIGURE 5.3 – Puissance fonction du seuil  $\alpha$  du test.

$n = 25$  on aura environ  $84\% \simeq 83.94\%$  de chances d'accepter ( $H_0$ ), que cette hypothèse soit vraie ou non (avec, soit  $\mu = 0.0$ , soit  $\mu = 2.0$ ). Une telle situation est assez fréquente, et montre l'intérêt des calculs précédents. En effet, la taille d'échantillon  $n = 25$  ne permet pas de discriminer correctement les deux hypothèses avec un test de seuil  $\alpha = 5\%$ .

Pour obtenir la taille  $N$  de l'échantillon nécessaire pour obtenir une puissance  $1 - \beta$  au moins égale à  $65\%$ , on peut, au moins dans un premier stade, réaliser d'une extrapolation graphique à partir de la Figure 5.2. Celle-ci montre qu'une telle performance devrait être atteinte pour  $N$  voisin de  $135 - 140$ . Un calcul plus précis de ce nombre est donné par le logiciel qui fournit  $N = 140$ , correspondant à  $1 - \beta \simeq 65.18\%$  (voir Figure 5.4). L'expression exacte de cette quantité étant

$$N = \inf \{ n : \mathbb{P} ( |T_{n-1}(\theta)| \geq t_{n-1;\alpha/2} ) \geq 1 - \beta \}, \tag{5.2.10}$$

qui vaut  $N = 140$  pour  $\theta = 0.2$  et  $\alpha = 5\%$ .

Calculs de taille d'échantillon (Feuille de données3 Comparaison d'une moyenne à un standard H0 : Mu = Mu0)	
	Valeur
Moyenne sous H0 (Mu0)	0,0000
Moyenne Population (Mu)	2,0000
Ecart-Type Population (Sigma)	10,0000
Effet Centré-Réduit (Es)	0,2000
Risque 1° espèce (Alpha)	0,0500
Puissance Cible	0,6500
Puissance correspondant au N Souhaité	0,6518
Taille d'Ech. Requise (N)	140,0000

FIGURE 5.4 – Taille  $n$  de l'échantillon nécessaire pour une puissance  $\geq 65\%$ .

### 5.3 Comparaison des moyennes de deux échantillons.

#### 5.3.1 Position du problème.

L'un des problèmes les plus fréquents de la statistique est celui de la comparaison des moyennes de deux échantillons. Nous considérerons ici deux échantillons indépendants  $X_1, \dots, X_m$ , et  $Y_1, \dots, Y_n$ , composés, respectivement, de  $m \geq 2$  répliques aléatoires indépendantes d'une loi normale  $N(\mu_X, \sigma_X^2)$ , et de  $n \geq 2$  répliques aléatoires indépendantes d'une loi normale  $N(\mu_Y, \sigma_Y^2)$ . Le problème est alors d'obtenir un intervalle de confiance de niveau de confiance  $1 - \alpha$ , spécifié à l'avance, pour la différence

$$d = \mu_X - \mu_Y.$$

On résume les échantillons par les statistiques

$$\hat{\mu}_X = \bar{X} = \frac{1}{m} \sum_{i=1}^m X_i, \quad \hat{\mu}_Y = \bar{Y} = \frac{1}{n} \sum_{j=1}^n Y_j, \quad (5.3.1)$$

$$s_X^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2, \quad s_Y^2 = \frac{1}{n-1} \sum_{j=1}^n (Y_j - \bar{Y})^2. \quad (5.3.2)$$

On estime ensuite  $d = \mu_X - \mu_Y$  par

$$\hat{d} = \hat{\mu}_X - \hat{\mu}_Y. \quad (5.3.3)$$

Sous les hypothèses admises, de normalité et d'indépendance des échantillons, les statistiques  $\hat{\mu}_X = \bar{X}$ ,  $\hat{\mu}_Y = \bar{Y}$ ,  $s_X^2$  et  $s_Y^2$  sont indépendantes. En posant  $\lambda_X = 1/m$ ,  $\lambda_Y = 1/n$ ,  $f_X = m - 1$  et  $f_Y = n - 1$ , les statistiques  $\hat{\mu}_X = \bar{X}$ ,  $\hat{\mu}_Y = \bar{Y}$ ,  $s_X^2$ ,  $s_Y^2$  et  $\hat{d}$ , ont des lois données par les formules

$$\begin{aligned} \hat{\mu}_X - \mu_X &\stackrel{d}{=} N(0, \lambda_X \sigma_X^2), & \hat{\mu}_Y - \mu_Y &\stackrel{d}{=} N(0, \lambda_Y \sigma_Y^2), & \hat{d} - d &\stackrel{d}{=} N(0, \lambda_X \sigma_X^2 + \lambda_Y \sigma_Y^2), \\ \frac{(m-1)s_X^2}{\sigma_X^2} &= \frac{f_X s_X^2}{\sigma_X^2} \stackrel{d}{=} \chi_{f_X}^2, & \frac{(n-1)s_Y^2}{\sigma_Y^2} &= \frac{f_Y s_Y^2}{\sigma_Y^2} \stackrel{d}{=} \chi_{f_Y}^2. \end{aligned}$$

#### 5.3.2 Cas où les variances sont égales.

Lorsque les variances sont égales,  $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ , et la variance commune  $\sigma^2$  possède un estimateur sans biais  $s^2$ , donné par

$$s^2 = \frac{f_X s_X^2 + f_Y s_Y^2}{f_X + f_Y},$$

tel que

$$\frac{(f_X + f_Y)s^2}{\sigma^2} = \frac{f_X s_X^2 + f_Y s_Y^2}{\sigma^2} \stackrel{d}{=} \chi_{f_X + f_Y}^2.$$

Comme, par ailleurs,

$$\hat{d} - d \stackrel{d}{=} N(0, \sigma^2(\lambda_X + \lambda_Y)),$$

on en déduit que

$$\frac{\hat{d} - d}{s\sqrt{\lambda_X + \lambda_Y}} \stackrel{d}{=} t_{f_X + f_Y}.$$

On obtient donc l'intervalle de confiance suivant pour  $d$ , quel que soit  $\alpha \in (0, 1)$ ,

$$\mathbb{P}\left(d \in \left[\hat{d} - st_{f_X + f_Y; \alpha/2} \sqrt{\lambda_X + \lambda_Y}, \hat{d} + st_{f_X + f_Y; \alpha/2} \sqrt{\lambda_X + \lambda_Y}\right]\right) = 1 - \alpha. \quad (5.3.4)$$

### 5.3.3 Variances quelconques - le problème de Behrens-Fisher.

Le cas où  $\sigma_X^2$  et  $\sigma_Y^2$  sont quelconques est beaucoup plus délicat. Il n'existe pas alors de procédure exacte permettant d'obtenir un intervalle de confiance de seuil  $1 - \alpha$  fixé pour  $d$ , indépendamment de  $\sigma_X^2$  et  $\sigma_Y^2$ . Notons  $\theta = \sigma_X^2/\sigma_Y^2$ . Comme, en général,

$$\frac{\hat{d} - d}{\sqrt{\lambda_X \sigma_X^2 + \lambda_Y \sigma_Y^2}} \stackrel{d}{=} N(0, 1),$$

il est naturel de construire des intervalles de confiance pour  $d$  de la forme

$$I(v) = \left[ \hat{d} - v \sqrt{\lambda_X s_X^2 + \lambda_Y s_Y^2}, \hat{d} + v \sqrt{\lambda_X s_X^2 + \lambda_Y s_Y^2} \right]$$

On utilise à cet effet l'une des méthodes suivantes, qui consistent à choisir  $v$  comme la quantile supérieure d'une loi de Student, dont le nombre de degrés de liberté est convenablement choisi (dont on trouvera les justifications dans Scheffé (1970), voir ci-dessous).

La méthode approximative de Welch (voir aussi les tables d'Aspin et Welch, dans les références ci-dessous) consiste à approximer la statistique

$$\mathcal{J} := \frac{\hat{d} - d}{\sqrt{\lambda_X s_X^2 + \lambda_Y s_Y^2}} = \frac{\hat{d} - d}{\sqrt{\frac{s_X^2}{m} + \frac{s_Y^2}{n}}},$$

par une loi de Student  $t_{r_{AW}}$ , dont le nombre de degrés de liberté, non entier,  $r$ , est donné par la formule

$$r_{AW} := \frac{1}{\frac{c^2}{m-1} + \frac{(1-c)^2}{n-1}} \quad \text{où} \quad c := \frac{\frac{s_X^2}{m}}{\frac{s_X^2}{m} + \frac{s_Y^2}{n}}.$$

Il n'y a pas d'inconvénient à utiliser une loi de Student à paramètre non entier, si on définit formellement la loi  $t_r$  par le rapport  $U/(V/r)$ , où les variables aléatoires  $U \stackrel{d}{=} N(0, 1)$  et  $V \stackrel{d}{=} \Gamma(\frac{r}{2}, \frac{1}{2})/r$  sont indépendantes. On peut vérifier que la valeur de  $r$  ainsi trouvée vérifie les inégalités

$$r_{min} := \min\{m - 1, n - 1\} \leq r_{AW} \leq r_{max} := (m - 1) + (n - 1).$$

On peut utiliser les relations

$$\mathbb{P}(|\mathcal{J}| \geq t_{r_{min}; \alpha/2}) \leq \alpha \simeq \mathbb{P}(|\mathcal{J}| \geq t_{r_{AW}; \alpha/2}) \leq \mathbb{P}(|\mathcal{J}| \geq t_{r_{max}; \alpha/2}) \leq \mathbb{P}(|\mathcal{J}| \geq t_{\infty; \alpha/2}).$$

On observera, en passant, que la valeur de la quantile supérieure d'ordre  $\alpha/2$  de la loi de Student  $t_r$ , notée  $t_{r; \alpha/2}$ , est une fonction décroissante de  $r$ , ayant pour limite  $\nu_{\alpha/2}$ , quantile supérieure d'ordre  $\alpha/2$  de la loi normale  $N(0, 1)$ , lorsque  $r \rightarrow \infty$ . Pour cette raison, on peut poser  $t_{\infty; \alpha/2} = \nu_{\alpha/2}$ . L'intervalle de confiance asymptotique basé sur l'approximation normale est donc toujours plus étroit que celui basé sur l'une ou l'autre des lois  $t_{r_{max}}$ ,  $t_{r_{AW}}$  ou  $t_{r_{min}}$ .

On consultera, sur ce problème, les articles de :

Aspin, A. A. (1948). An examination and further development of a formula arising in the problem of comparing two mean values. *Biometrika*. **35** 88-96.

Scheffé, H. (1970). Practical solutions of the Behrens-Fisher problem. *Journal of the American Statistical Association*. **65** 1501-1508.

Welch, B. L. (1947). The generalization of 'Students' problem when several different population variances are involved. *Biometrika*. **34** 28-35.

### 5.3.4 Utilisation de la loi de Fisher.

La loi de Fisher  $F_{p,q}$  est la loi du rapport  $Z = U/V$ , où  $U, V$  sont indépendantes, et de lois respectives

$$pU \stackrel{d}{=} \chi_p^2 \quad \text{et} \quad qV \stackrel{d}{=} \chi_q^2.$$

Il est commode d'adopter la notation

$$Z \stackrel{d}{=} F_{p,q} \quad \Leftrightarrow \quad Z \stackrel{d}{=} \frac{\chi_p^2/p}{\chi_q^2/q}, \quad (5.3.5)$$

pour désigner cette propriété. Pour tout  $0 < \alpha < 1$ , on note  $F_{p,q}^\alpha$  le quantile supérieur d'ordre  $\alpha$  de la loi de Fisher  $F_{p,q}$ , défini par, si  $Z \stackrel{d}{=} F_{p,q}$ ,

$$\mathbb{P}(Z \geq F_{p,q}^\alpha) = \alpha. \quad (5.3.6)$$

**Exemple 5.3.1.** Soit  $\mathbf{X} \equiv N_m(\mathbf{M}, \sigma^2 \mathbb{I})$  un vecteur normal tel que  $\mathbf{M} = A\theta$ , avec  $\theta \in \mathbb{R}^p$ . On suppose que  $1 \leq p < m$ , et que la matrice  $A \in \mathcal{M}_{m,p}$  est de rang  $p = \text{rg}(A)$ . On estime alors  $\theta$  et  $\sigma^2$  par les statistiques mutuellement indépendantes

$$\hat{\theta} = (A'A)^{-1} A' \mathbf{X} \stackrel{d}{=} N_p(\theta, \sigma^2 (A'A)^{-1}), \quad (5.3.7)$$

$$s^2 = \frac{1}{m-p} \|\mathbf{X} - A\hat{\theta}\|^2 = \frac{1}{m-p} \mathbf{X}' \left\{ \mathbb{I}_m - A(A'A)^{-1} A' \right\} \mathbf{X} \quad (5.3.8)$$

$$= \frac{1}{m-p} \left\{ \mathbf{X}' \mathbf{X} - (A\hat{\theta})' (A\hat{\theta}) \right\}. \quad (5.3.9)$$

On utilise alors le fait que

$$SS_1 = \left\{ A(\hat{\theta} - \theta) \right\}' \left\{ A(\hat{\theta} - \theta) \right\} = (\hat{\theta} - \theta)' (A'A) (\hat{\theta} - \theta) \stackrel{d}{=} \sigma^2 \chi_p^2, \quad (5.3.10)$$

$$SS_2 = (m-p)s^2 = \left\{ \mathbf{X}' \mathbf{X} - (A\hat{\theta})' (A\hat{\theta}) \right\} \stackrel{d}{=} \sigma^2 \chi_{m-p}^2, \quad (5.3.11)$$

sont indépendants pour conclure que

$$Z = \frac{SS_1/p}{SS_2/(m-p)} = \frac{SS_1}{ps^2} \stackrel{d}{=} F_{p,m-p}. \quad (5.3.12)$$

Ceci permet de construire un domaine de confiance d'ordre  $1 - \alpha$  pour  $\theta$ , en écrivant que

$$\mathbb{P} \left( (\hat{\theta} - \theta)' (A'A) (\hat{\theta} - \theta) \leq ps^2 F_{p,m-p}^\alpha \right) = 1 - \alpha. \quad (5.3.13)$$

A partir de cette relation, on peut obtenir des *intervalles de confiance de type Scheffé* pour les contrastes  $u'\theta$ , en écrivant, par l'inégalité de Schwarz que

$$\begin{aligned} \sup_{u \in \mathbb{R}^p} \left| u'(\hat{\theta} - \theta) \right|^2 &= \sup_{u \in \mathbb{R}^p} \left| \{ (A'A)^{-1} u \}' (A'A) (\hat{\theta} - \theta) \right|^2 \\ &= \sup_{u \in \mathbb{R}^p} \left\{ \{ (A'A)^{-1} u \}' (A'A) \{ (A'A)^{-1} u \} \right\} (\hat{\theta} - \theta)' (A'A) (\hat{\theta} - \theta) \\ &= \sup_{u \in \mathbb{R}^p} \left\{ u' (A'A)^{-1} u \right\} (\hat{\theta} - \theta)' (A'A) (\hat{\theta} - \theta). \end{aligned}$$

On en déduit les intervalles de confiance simultanés (avec la convention  $s \geq 0$ )

$$\begin{aligned} \mathbb{P} \left( u'\hat{\theta} \in \left[ u'\hat{\theta} - s \sqrt{u'(A'A)^{-1}u} \sqrt{pF_{p,m-p}^\alpha}, \right. \right. \\ \left. \left. u'\hat{\theta} + s \sqrt{u'(A'A)^{-1}u} \sqrt{pF_{p,m-p}^\alpha} \right] \quad \forall u \in \mathbb{R}^p \right) = 1 - \alpha. \end{aligned} \quad (5.3.14)$$

On comparera ces intervalles *simultanés* aux intervalles *individuels*, obtenus pour  $u \in \mathbb{R}^p$  fixé. Comme

$$u' \widehat{\theta} \stackrel{d}{=} N(u' \theta, \sigma^2 u' (A' A)^{-1} u),$$

on obtient que, pour tout  $u \in \mathbb{R}^p$ ,

$$\mathbb{P} \left( u' \theta \in \left[ u' \widehat{\theta} - \widehat{\sigma} \sqrt{u' (A' A)^{-1} u} t_{m-p}^{\alpha/2}, \right. \right. \\ \left. \left. u' \widehat{\theta} + \widehat{\sigma} \sqrt{u' (A' A)^{-1} u} t_{m-p}^{\alpha/2} \right] \right) = 1 - \alpha, \quad (5.3.15)$$

ou, de manière équivalente,

$$\mathbb{P} \left( u' \theta \in \left[ u' \widehat{\theta} - s \sqrt{u' (A' A)^{-1} u} \sqrt{F_{1, m-p}^{\alpha}}, \right. \right. \\ \left. \left. u' \widehat{\theta} + s \sqrt{u' (A' A)^{-1} u} \sqrt{F_{1, m-p}^{\alpha}} \right] \right) = 1 - \alpha. \quad (5.3.16)$$

## 5.4 Intervalles de confiance simultanés.

Le lemme suivant (cf. p.74 dans Scheffé, H. (1959) *The Analysis of Variance*, Wiley, New York) est extrêmement utile.

**Lemme 5.4.1.** *Les deux propriétés suivantes (A) et (B) sont équivalentes.*

(A) On a  $|\theta_i - \theta_j| \leq \delta$  pour tout couple d'indices  $i, j$  tel que  $1 \leq i, j \leq p$ ;

(B) On a  $|\sum_{i=1}^p u_i \theta_i| \leq \delta \left\{ \frac{1}{2} \sum_{i=1}^p |u_i| \right\}$  pour toute suite de coefficients telle que  $\sum_{i=1}^p u_i = 0$ .

**Preuve.** Il est clair que (B)  $\Rightarrow$  (A) (il suffit de prendre  $u_i = 1, u_j = -1$  et  $u_\ell = 0$  pour  $\ell \neq i, j$ ), et la difficulté de la démonstration se limite à vérifier que (A)  $\Rightarrow$  (B). Supposons donc (A) satisfaite, et donnons nous une suite de coefficients  $u_1, \dots, u_p$  tels que  $\sum_{i=1}^p u_i = 0$ . Deux cas sont possibles.

– Ou bien  $u_1 = \dots = u_p = 0$ , auquel cas l'inégalité  $|\sum_{i=1}^p u_i \theta_i| \leq \delta \left\{ \frac{1}{2} \sum_{i=1}^p |u_i| \right\}$  est triviale.

– Ou bien les  $u_1, \dots, u_p$  ne sont pas tous nuls. Dans ce cas, si on désigne par  $\mathcal{P}$  l'ensemble des indices  $i$  tels que  $u_i > 0$  et par  $\mathcal{N}$  l'ensemble des indices  $j$  tels que  $u_j \leq 0$ ,  $\mathcal{P}$  et  $\mathcal{N}$  ne sont pas vides, et vérifient

$$\sum_{i=1}^p u_i = \sum_{i \in \mathcal{P}} u_i - \sum_{j \in \mathcal{N}} (-u_j) = 0, \quad (5.4.1)$$

$$\sum_{i=1}^p |u_i| = \sum_{i \in \mathcal{P}} u_i + \sum_{j \in \mathcal{N}} (-u_j) = 2s > 0, \quad (5.4.2)$$

où  $s$  vérifie (en conséquence de (5.4.1)–(5.4.2)),

$$s = \frac{1}{2} \sum_{i=1}^p |u_i| = \sum_{i \in \mathcal{P}} u_i = \sum_{j \in \mathcal{N}} (-u_j) > 0. \quad (5.4.3)$$

Ecrivons maintenant

$$\begin{aligned} \sum_{i=1}^p u_i \theta_i &= \frac{1}{s} \left\{ \left( \sum_{i \in \mathcal{P}} u_i \theta_i \right) \left( \sum_{j \in \mathcal{N}} (-u_j) \right) - \left( \sum_{j \in \mathcal{N}} (-u_j) \theta_j \right) \left( \sum_{i \in \mathcal{P}} u_i \right) \right\} \\ &= \frac{1}{s} \sum_{i \in \mathcal{P}} \sum_{j \in \mathcal{N}} u_i (-u_j) (\theta_i - \theta_j). \end{aligned}$$

Comme, pour  $i \in \mathcal{P}$  et  $j \in \mathcal{N}$ ,  $u_i(-u_j) \geq 0$ , on déduit de l'hypothèse (A) et de l'expression précédente que

$$\begin{aligned} \left| \sum_{i=1}^p u_i \theta_i \right| &\leq \sum_{i \in \mathcal{P}} \sum_{j \in \mathcal{N}} |u_i(-u_j) (\theta_i - \theta_j)| \\ &\leq \delta \left\{ \frac{1}{s} \sum_{i \in \mathcal{P}} \sum_{j \in \mathcal{N}} u_i(-u_j) \right\} = \delta s, \end{aligned}$$

ce qu'il fallait démontrer, compte tenu de (5.4.3).  $\square$

**Définition 5.4.1.** Soient  $Y_1, \dots, Y_m$  des variables aléatoires indépendantes de même loi normale  $N(0, 1)$ , et  $s^2$  une variable aléatoire indépendante, telle que  $ns^2 \stackrel{d}{=} \chi_n^2$ . On définit alors les variables

$$M_{m,n} = \frac{1}{s} \max_{1 \leq i \leq m} |Y_i| \quad (\textit{Studentized Maximum}), \quad (5.4.4)$$

et

$$Q_{m,n} = \frac{1}{s} \max_{1 \leq i, j \leq m} |Y_i - Y_j| \quad (\textit{Studentized Range}), \quad (5.4.5)$$

dites respectivement du "maximum studentisé" et de l' "étendue studentisée". Pour  $0 < \alpha < 1$ , on note, respectivement,  $m_{m,n}^\alpha$  et  $q_{m,n}^\alpha$  les valeurs telles que

$$\mathbb{P}(M_{m,n} \geq m_{m,n}^\alpha) = \alpha \quad \text{et} \quad \mathbb{P}(Q_{m,n} \geq q_{m,n}^\alpha) = \alpha. \quad (5.4.6)$$

**Remarque 5.4.1.** 1°) Les valeurs de  $q_{m,n}^\alpha$  sont tabulées, en particulier, pp. 434–436 dans :

Scheffé, H. (1959) *The Analysis of Variance*, Wiley, New York.

2°) Les valeurs de  $m_{m,n}^\alpha$  sont tabulées, en particulier, dans :

Pillai, K. C. S. et Ramachandran, K. V. (1954). On the distribution of the ratio of the  $i$ -th observation in an ordered sample from a normal population to an independent estimate of the standard deviation. *Ann. Math. Statist.* **25** 565-572.

## 5.5 Utilisation du $T^2$ de Hotelling.

Soient  $X_1, \dots, X_n$  un échantillon de taille  $n \geq d$  de la loi normale  $N_d(\mu, \Sigma)$ . On suppose que  $\mu \in \mathbb{R}^d$  et que  $\Sigma > 0$  est une matrice  $(d \times d)$  définie positive. On estime  $\mu$  et  $\Sigma$ , respectivement, par les statistiques

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad \hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})'. \quad (5.5.1)$$

L'estimateur  $\hat{\Sigma}$  de  $\Sigma$  n'est pas centré. Sous réserve que  $n \geq 2$ , on le remplace avantageusement pour certaines applications par l'estimateur alternatif  $S$  de  $\Sigma$ , défini par

$$S = \frac{n}{n-1} \hat{\Sigma} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})', \quad \text{qui est tel que} \quad \mathbb{E}(S) = \Sigma \quad \text{pour} \quad n \geq 2. \quad (5.5.2)$$

Sous ces hypothèses, les matrices aléatoires  $\bar{X}$  et  $S$  sont indépendantes, de lois respectives données par

$$\sqrt{n}(\bar{X} - \mu) \stackrel{d}{=} N_d(0, \Sigma) \quad \text{et} \quad S \stackrel{d}{=} W_d(n-1, \Sigma). \quad (5.5.3)$$

D'une manière générale, on dit qu'une matrice aléatoire symétrique positive  $A$  suit une *loi de Wishart*  $W_d(N, \Sigma)$  si on a les identités en loi

$$A \stackrel{d}{=} \sum_{j=1}^N Y_j Y_j' \stackrel{d}{=} W_d(N, \Sigma), \quad (5.5.4)$$

où  $Y_1, \dots, Y_N$  sont des vecteurs aléatoires de  $\mathbb{R}^d$ , indépendants de même loi normale  $N_d(\mathbb{O}, \Sigma)$ . On admettra ici l'équivalence (sous la condition  $\Sigma > 0$ )

$$A \stackrel{d}{=} W_d(N, \Sigma) > 0 \text{ p.s.} \Leftrightarrow N \geq d. \quad (5.5.5)$$

La statistique du  $T^2$  de Hotelling généralise en dimension  $d \geq 2$  la statistique  $T$  de Student. Elle est basée sur le fait (admis provisoirement) que, sous les hypothèses ci-dessus, pour  $n \geq d + 1$ ,

$$T^2 = n(\bar{X} - \mu)' S^{-1} (\bar{X} - \mu) \stackrel{d}{=} \left\{ \frac{d(n-1)}{n-d} \right\} F_{d, n-d}. \quad (5.5.6)$$

On peut ainsi donner un domaine de confiance de degré  $1 - \alpha$  pour  $\mu \in \mathbb{R}^d$  en écrivant que

$$\mathbb{P} \left( n(\bar{X} - \mu)' S^{-1} (\bar{X} - \mu) \leq \left\{ \frac{d(n-1)}{n-d} \right\} f_{d, n-d; \alpha} \right) = 1 - \alpha. \quad (5.5.7)$$

On remarquera que, pour  $d = 1$ , (5.5.7) se ramène à un intervalle de confiance de Student symétrique.

On obtient des généralisations de la statistique du  $T^2$  de Hotelling, en considérant, de manière plus générale, deux matrices quelconques  $Z \stackrel{d}{=} N_d(0, \Sigma)$  et  $A \stackrel{d}{=} W_d(N, \Sigma)$ . Lorsque les matrices  $Z$  et  $A$  sont indépendantes, les propriétés suivantes sont satisfaites.

$$(a) \quad Y' \Sigma^{-1} Y \stackrel{d}{=} \chi_d^2; \quad (5.5.8)$$

$$(b) \quad Y' \Sigma^{-1} Y / Y' A^{-1} Y \stackrel{d}{=} \chi_{N-d+1}^2 \quad (5.5.9)$$

$$(c) \quad Y' \Sigma^{-1} Y \text{ et } Y' \Sigma^{-1} Y / Y' A^{-1} Y \text{ sont indépendants.} \quad (5.5.10)$$

Une conséquence directe de ces propriétés est que

$$\begin{aligned} Y' A^{-1} Y \times \left\{ \frac{N-d+1}{d} \right\} &= \frac{Y' \Sigma^{-1} Y}{Y' \Sigma^{-1} Y / Y' A^{-1} Y} \times \left\{ \frac{N-d+1}{d} \right\} \\ &\stackrel{d}{=} \frac{\chi_d^2 / d}{\chi_{N-d+1}^2 / (N-d+1)} \stackrel{d}{=} F_{d, N-d+1}, \end{aligned} \quad (5.5.11)$$

suit une loi de Fisher  $F_{d, N-d+1}$ .

## 5.6 Utilisation du $\Lambda$ de Wilks.

La statistique  $\Lambda$  de Wilks joue le même rôle que les statistiques de Fisher dans le cas où les carrés scalaires (fournissant des lois du  $\chi^2$ ) sont remplacés par des matrices de Wishart, générées par des données vectorielles. On suppose que  $\Sigma > 0$  est une matrice  $(d \times d)$  symétrique, définie positive, et on dispose de deux variables matricielles  $A$  et  $B$  indépendantes, suivant des lois de Wishart, de paramètres respectifs

$$A \stackrel{d}{=} W_d(r, \Sigma) \text{ et } B \stackrel{d}{=} W_d(n, \Sigma) \text{ avec } \min\{r, n\} \geq d. \quad (5.6.1)$$

On désire tester l'hypothèse  $(H_0)$  que  $A$  suit une loi de Wishart centrée, c'est à dire de la forme

$$A \stackrel{d}{=} \sum_{j=1}^r Y_j Y_j',$$

où  $Y_1, \dots, Y_r$  sont des vecteurs aléatoires indépendants de même loi  $N_d(\mathbb{O}, \Sigma)$ , contre l'hypothèse que, dans cette même représentation,  $Y_1, \dots, Y_r$  sont des vecteurs aléatoires indépendants, de lois respectives  $N_m(\mu_1, \Sigma), \dots, N(\mu_r, \Sigma)$ , où les  $\mu_1, \dots, \mu_r$  sont quelconques. Pour cela, on calcule la statistique

$$U_{r, n; d} = \frac{\det(A)}{\det(A+B)} = \Lambda^{2/(n+d)}, \quad (5.6.2)$$



et on rejette l'hypothèse lorsque  $U_{r,n;d} \leq c_\alpha$  pour une valeur convenable du niveau critique  $c_\alpha$ . On utilisera à cet effet l'approximation suivante, valable pour les grandes valeurs de

$$N = n - \frac{1}{2} \{d - r + 1\}.$$

On a (voir Theorem 8.6.2, p.208 dans : Anderson, T. W. (1958). *An Introduction to Multivariate Analysis*. Wiley, New York)

$$\begin{aligned} -N \log \mathbb{P}(U_{r,n;d} \leq z) &= \mathbb{P}(\chi_{dr}^2 \leq z) \\ &+ \frac{\gamma_2}{N^2} \left\{ \mathbb{P}(\chi_{dr+4}^2 \leq z) - \mathbb{P}(\chi_{dr}^2 \leq z) \right\} + O\left(\frac{1}{N^3}\right) \\ &+ \frac{1}{N^4} \left\{ \gamma_4 \left\{ \mathbb{P}(\chi_{dr+4}^2 \leq z) - \mathbb{P}(\chi_{dr}^2 \leq z) \right\} \right. \\ &\quad \left. - \gamma_2^2 \left\{ \mathbb{P}(\chi_{dr+4}^2 \leq z) - \mathbb{P}(\chi_{dr}^2 \leq z) \right\} \right\} + O\left(\frac{1}{N^6}\right), \end{aligned} \quad (5.6.3)$$

où

$$\gamma_2 = \frac{dr(d^2 + r^2 - 5)}{48}, \quad (5.6.4)$$

$$\gamma_4 = \frac{\gamma_2^2}{2} + \frac{dr}{1920} \left\{ 3r^2 + 3d^2 + 10r^2d^2 - 50(r^2 + d^2) + 159 \right\}. \quad (5.6.5)$$

## 5.7 Analyse statistique du modèle linéaire simple.

### 5.7.1 Le modèle linéaire simple.

Nous considérons un vecteur aléatoire  $\mathbf{X}$  à valeurs dans  $\mathbb{R}^n$ . Nous raisonnerons en supposant l'une ou l'autre des hypothèses suivantes.

$$(\mathcal{C}) \quad \mathbb{E}(\mathbf{X}) = \mathbf{M} \in \mathbb{R}^n \text{ et } \text{Var}(\mathbf{X}) = \sigma^2 \mathbb{I}_n,$$

$$(\mathcal{N}) \quad \mathbf{X} \stackrel{d}{=} N_n(\mathbf{M}, \sigma^2 \mathbb{I}_n).$$

Dans ces relations,  $\sigma > 0$  est un paramètre inconnu. De toute évidence,  $(\mathcal{N}) \Rightarrow (\mathcal{C})$ .

Nous parlerons de *modèle linéaire simple* pour exprimer la condition

$$(\mathcal{L}) \quad \mathbf{M} \in \mathcal{F},$$

où  $\mathcal{F}$  est un sous-espace vectoriel *connu* de  $\mathbb{R}^n$ . Le but de ce paragraphe est d'étudier une certaine catégorie d'estimateurs de  $\mathbf{M}$  sous ces hypothèses. Nous considérerons aussi, par la même occasion, l'estimation de  $\sigma^2$ . Dans ce qui suit, si  $E$  et  $F$  sont des espaces vectoriels, nous noterons

$$E \rightleftharpoons F, \quad (5.7.1)$$

pour exprimer que les espaces vectoriels  $E$  et  $F$  sont *isomorphes*.

**Définition 5.7.1.** Soit  $\mathbf{Y} \in \mathbb{R}^n$  un vecteur aléatoire dont la loi dépend d'un paramètre vectoriel  $\theta \in \Theta \subseteq \mathbb{R}^p$  (où  $\Theta$  est un ouvert de  $\mathbb{R}^p$  et  $p \leq n$ ). On appelle *estimateur* de  $\theta$  toute statistique  $\tilde{\theta} = \tilde{\theta}(\mathbf{Y})$ , fonction de  $\mathbf{Y}$  et à valeurs dans  $\Theta$ . L'estimateur  $\tilde{\theta}$  de  $\theta$  est dit *sans biais* si l'identité suivante a lieu.

$$\mathbb{E}_\theta(\tilde{\theta}) = \theta, \quad \forall \theta \in \Theta. \quad (5.7.2)$$

Le *biais* de l'estimateur  $\tilde{\theta}$  de  $\theta$  est, lorsqu'elle existe, la différence  $\mathbb{E}(\tilde{\theta}) - \theta$ . On dit que l'estimateur  $\tilde{\theta} = \tilde{\theta}(\mathbf{Y})$  de  $\theta$  est *linéaire*, si  $\Theta \rightleftharpoons \mathbb{R}^p$  et si l'application  $y \in \mathbb{R}^n \rightarrow \tilde{\theta}(y) \in \Theta \rightleftharpoons \mathbb{R}^p$  est une application linéaire de  $\mathbb{R}^n$  dans  $\Theta \rightleftharpoons \mathbb{R}^p$ .

Dans (5.7.2), la notation  $\mathbb{E}_\theta$  est utilisée pour désigner l'espérance, sous l'hypothèse que la valeur véritable du paramètre dont dépend la loi de  $\mathbf{Y}$  est  $\theta$ . Dans la suite, nous noterons le plus souvent  $\mathbb{E}(\cdot)$  au lieu de  $\mathbb{E}_\theta(\cdot)$ , lorsque cette notation n'est pas ambiguë.

**Proposition 5.7.1.** *Supposons que  $\theta = \mathbb{E}_\theta(\mathbf{Y})$ . Alors, sous l'hypothèse que  $\theta \subseteq \mathbb{R}^n$  varie dans un espace vectoriel  $\Theta \simeq \mathbb{R}^p$ ,  $\Theta \subseteq \mathbb{R}^n$ , une condition nécessaire et suffisante pour que  $\tilde{\theta} = \tilde{\theta}(\mathbf{Y})$  soit un estimateur linéaire sans biais de  $\theta$  est que l'application  $y \in \mathbb{R}^n \rightarrow \tilde{\theta}(y) \in \Theta$  soit une projection de  $\mathbb{R}^n$  sur  $\Theta$ .*

**Preuve.** Ecrivons matriciellement  $\tilde{\theta} = \tilde{\theta}(\mathbf{Y}) = L\mathbf{Y}$ . On a  $\mathbb{E}_\theta(\tilde{\theta}) = \theta = L\mathbb{E}_\theta(\mathbf{Y}) = L\theta$ . L'identité  $\theta = L\theta$  devant avoir lieu pour tout  $\theta \in \Theta$ , on obtient bien ainsi une projection de  $\mathbb{R}^n$  sur  $\Theta$ .  $\square$

Nous rappelons qu'une projection  $P : E \rightarrow F$  de l'espace  $E = \mathbb{R}^n$  sur un sous-espace  $F$  de  $E$  est une application linéaire de  $E$  sur  $F$  vérifiant les propriétés équivalentes suivantes.

- (1) La restriction de  $P$  à  $F$  est l'identité de  $F$ ;
- (2)  $P^2 = P$ ;
- (3)  $F = \text{Im}(P)$  et  $\text{Im}(P) \oplus \text{Ker}(P) = E$ ;
- (4)  $P$  est diagonalisable et admet pour seules valeurs propres 1 et 0.

Pour construire une projection de  $E$  sur  $F$ , il suffit donc de se donner un sous-espace vectoriel  $G$  de  $E$  en somme directe avec  $F$ , tel que  $F \oplus G = E$ , et d'identifier  $G$  à  $\text{Ker}(P)$ . Sous les hypothèses de la proposition 5.7.1, il y a donc autant d'estimateurs sans biais de  $\theta \in \Theta$  qu'il y a de sous-espaces supplémentaires de  $\Theta$  dans  $\mathbb{R}^n$ .

**Définition 5.7.2.** *Soient deux estimateurs  $\tilde{\theta}_1$  et  $\tilde{\theta}_2$  de  $\theta \in \Theta \simeq \mathbb{R}^p$ . Nous dirons que la dispersion (ou la variance) de  $\tilde{\theta}_1$  est plus petite que la dispersion (ou la variance) de  $\tilde{\theta}_2$  si l'une des conditions équivalentes suivantes est satisfaite (indépendamment de  $\theta \in \Theta$ ).*

- (1) Pour tout  $u \in \mathbb{R}^p$ ,  $\text{Var}(u'\tilde{\theta}_1) \leq \text{Var}(u'\tilde{\theta}_2)$ ;
- (2) Pour tout  $u \in \mathbb{R}^p$ ,  $u'\text{Var}(\tilde{\theta}_1)u \leq u'\text{Var}(\tilde{\theta}_2)u$ .

Les propriétés (1) et (2) ci-dessus sont équivalentes, du fait que  $\text{Var}(u'\tilde{\theta}) = u'\text{Var}(\tilde{\theta})u$ . On note commodément ces relations en écrivant que

$$\text{Var}(\tilde{\theta}_1) \leq \text{Var}(\tilde{\theta}_2). \quad (5.7.3)$$

L'ordre défini entre les estimateurs par (5.7.3) n'est, en général, pas un ordre total. Si, toutefois, une classe d'estimateurs possède un estimateur de dispersion (ou de variance) plus petite que la dispersion de chaque autre estimateur de même type, ce dernier sera appelé *estimateur à dispersion (ou variance) minimale*.

**Théorème 5.7.1.** *1°) Soit  $\mathbf{X} \in \mathbb{R}^n$  un vecteur aléatoire vérifiant les hypothèses (C)–(L), c'est à dire, tel que  $\mathbb{E}(\mathbf{X}) = \mathbf{M} \in \mathcal{F}$  et  $\text{Var}(\mathbf{X}) = \sigma^2 \mathbb{I}_n$ , où  $\mathcal{F}$  est un sous-espace vectoriel spécifié, de dimension  $p$ , de  $\mathbb{R}^m$ . Alors, il existe un unique estimateur linéaire sans biais à dispersion minimale de  $\mathbf{M} \in \mathcal{F}$ , défini par*

$$\tilde{\mathbf{M}} = P_{\mathcal{F}}^\perp(\mathbf{X}), \quad (5.7.4)$$

où  $P_{\mathcal{F}}^\perp$  désigne la projection orthogonale (relativement au produit euclidien usuel sur  $\mathbb{R}^n$ ) de  $\mathbb{R}^n$  sur  $\mathcal{F}$ . De plus, si  $\|u\| = \{u'u\}^{1/2} = \langle u, u \rangle^{1/2}$  et  $\langle u, v \rangle$  désignent respectivement la norme euclidienne et le produit euclidien sur  $\mathbb{R}^n$ , alors, on a les propriétés suivantes.

$$(i) \quad \|\mathbf{X} - \tilde{\mathbf{M}}\|^2 = \inf_{\mathbf{Z} \in \mathcal{F}} \|\mathbf{X} - \mathbf{Z}\|^2; \quad (ii) \quad \mathbf{X} - \tilde{\mathbf{M}} \perp \mathbf{Z}, \quad \forall \mathbf{Z} \in \mathcal{F}; \quad (5.7.5)$$

$$(iii) \quad \|\mathbf{X} - \mathbf{M}\|^2 = \|\mathbf{X} - \tilde{\mathbf{M}}\|^2 + \|\tilde{\mathbf{M}} - \mathbf{M}\|^2; \quad (5.7.6)$$

$$(iv) \quad \text{Cov}(\mathbf{X} - \mathbf{M}, \tilde{\mathbf{M}} - \mathbf{M}) = \mathbb{E}(\{\mathbf{X} - \mathbf{M}\}\{\tilde{\mathbf{M}} - \mathbf{M}\}') = \mathbf{O}; \quad (5.7.7)$$

$$(v) \quad \mathbb{E}(\|\mathbf{X} - \tilde{\mathbf{M}}\|^2) = (m - p)\sigma^2, \quad \mathbb{E}(\|\tilde{\mathbf{M}} - \mathbf{M}\|^2) = p\sigma^2 \quad (5.7.8)$$

2°) Si, de plus, l'hypothèse  $(\mathcal{N})$  est satisfaite, alors  $\mathbf{X} - \widetilde{\mathbf{M}}$  et  $\widetilde{\mathbf{M}}$  sont indépendants, et tels que, pour tout  $\mathbf{R} \in \mathbb{R}^n$  et  $\mathbf{S} \in \mathbb{R}^n$ ,

$$\frac{1}{\sigma^2} \|\mathbf{X} - \widetilde{\mathbf{M}} - \mathbf{R}\|^2 \stackrel{d}{=} \chi_{n-p}^2 \left( \frac{1}{\sigma^2} \|\mathbf{R}\|^2 \right) \quad \text{et} \quad \frac{1}{\sigma^2} \|\widetilde{\mathbf{M}} - \mathbf{M} - \mathbf{S}\|^2 \stackrel{d}{=} \chi_p^2 \left( \frac{1}{\sigma^2} \|\mathbf{S}\|^2 \right). \quad (5.7.9)$$

En particulier, pour  $\mathbf{R} = \mathbf{S} = \mathbf{0}$ ,

$$\frac{1}{\sigma^2} \|\mathbf{X} - \widetilde{\mathbf{M}}\|^2 \stackrel{d}{=} \chi_{n-p}^2 \quad \text{et} \quad \frac{1}{\sigma^2} \|\widetilde{\mathbf{M}} - \mathbf{M}\|^2 \equiv \chi_p^2. \quad (5.7.10)$$

**Preuve.** Nous nous limitons à apporter la preuve du théorème sous les hypothèses  $(\mathcal{N})$ – $(\mathcal{L})$ . Le remplacement de  $(\mathcal{N})$  par  $(\mathcal{C})$  consistant essentiellement à remplacer la propriété d'indépendance par l'absence de corrélation, les méthodes de démonstration sont essentiellement les mêmes dans ce cas.

On choisit tout d'abord une base orthonormée de  $\mathbb{R}^n$  dont les  $p$  premiers vecteurs  $e_1, \dots, e_p$  engendrent  $\mathcal{F}$ , et les  $n - p$  suivants,  $e_{p+1}, \dots, e_n$ , l'orthogonal  $\mathcal{F}^\perp$  de  $\mathcal{F}$  dans  $\mathbb{R}^n$ . Les coordonnées  $Y_1, \dots, Y_n$  de  $\mathbf{X}$  dans cette base forment une matrice colonne  $\mathbf{Y}$  liée à  $\mathbf{X} \in \mathbb{R}^n$  par la relation  $\mathbf{Y} = H\mathbf{X}$ , où  $H$  est une matrice orthogonale convenable. On transforme alors le problème en posant  $\mathbf{N} = H\mathbf{M}$ ,  $\mathcal{G} = H\mathcal{F}$ , en vérifiant que  $\mathbf{Y} \equiv N_n(\mathbf{N}, \sigma^2\mathbb{I})$ . Cette dernière propriété implique que  $Y_1, \dots, Y_n$  sont des variables aléatoires normales indépendantes de variance 1. La projection orthogonale de

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_p \\ Y_{p+1} \\ \vdots \\ Y_n \end{bmatrix} \quad \text{sur } \mathcal{G} \text{ est } \widehat{\mathbf{N}} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_p \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathcal{G}, \quad \text{qui est indépendant de } \mathbf{y} - \widehat{\mathbf{N}} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ Y_{p+1} \\ \vdots \\ Y_n \end{bmatrix}.$$

Considérons maintenant un estimateur linéaire sans biais  $\widetilde{\mathbf{N}} = L\mathbf{Y}$  de  $\mathbf{N}$ . Comme, par la proposition 5.7.1,  $L$  est la matrice d'une projection, on a nécessairement

$$\widetilde{\mathbf{N}} = L\mathbf{Y} = L(\widehat{\mathbf{N}} + \{\mathbf{y} - \widehat{\mathbf{N}}\}) = \widehat{\mathbf{N}} + L(\mathbf{y} - \widehat{\mathbf{N}}) = \widehat{\mathbf{N}} + \mathbf{Z},$$

où  $\mathbf{Z}$  est indépendant de  $\widehat{\mathbf{N}}$ . On a donc, nécessairement

$$\text{Var}(\widetilde{\mathbf{N}}) = \text{Var}(\widehat{\mathbf{N}}) + \text{Var}(\mathbf{Z}), \quad (5.7.11)$$

ce qui implique (compte tenu du fait que  $\text{Var}(\mathbf{Z}) \geq 0$ ) que  $\widehat{\mathbf{N}}$  est un estimateur linéaire sans biais à dispersion minimale de  $\mathbf{N}$ . L'unicité est évidente, puisque  $\widehat{\mathbf{N}}$  ne peut être à dispersion minimale que si  $\text{Var}(\mathbf{Z}) = \mathbb{O}$ , ce qui implique l'égalité  $\widetilde{\mathbf{N}} = \widehat{\mathbf{N}}$ .

Supposons maintenant que  $\widetilde{\mathbf{M}}_1$  et  $\widetilde{\mathbf{M}}_2$  soient deux estimateurs sans biais de  $\mathbf{M}$  vérifiant l'inégalité  $\text{Var}(\widetilde{\mathbf{M}}_1) \leq \text{Var}(\widetilde{\mathbf{M}}_2)$ . Alors, il est facile de constater que  $\widetilde{\mathbf{N}}_1 = H\widetilde{\mathbf{M}}_1$  et  $\widetilde{\mathbf{N}}_2 = H\widetilde{\mathbf{M}}_2$  soient deux estimateurs sans biais de  $\mathbf{N} = H\mathbf{M}$  tels que  $\text{Var}(\widetilde{\mathbf{N}}_1) \leq \text{Var}(\widetilde{\mathbf{N}}_2)$ , et réciproquement. Cette équivalence vient du fait que, en général, pour une matrice inversible  $P$ ,

$$u' \text{Var}(P\mathbf{V}_1)u \leq u' \text{Var}(P\mathbf{V}_2)u, \quad \forall u \Leftrightarrow v' \text{Var}(\mathbf{V}_1)v \leq v' \text{Var}(\mathbf{V}_2)v, \quad \forall v = P'u.$$

Par application de ce principe, on constate que  $\widehat{\mathbf{M}} = H'\widehat{\mathbf{N}}$  est un estimateur linéaire sans biais de  $\mathbf{M}$  à dispersion minimale. Comme on a nécessairement  $\mathbf{y} - \widehat{\mathbf{N}} \perp \mathbf{u}, \forall \mathbf{u} \in \mathcal{G}$ , on a  $H'\mathbf{y} - H'\widehat{\mathbf{N}} \perp H'\mathbf{u}, \forall \mathbf{u} \in \mathcal{G}$ , ce qui équivaut à  $\mathbf{X} - \widehat{\mathbf{M}} \perp \mathbf{v}, \forall \mathbf{v} \in \mathcal{F}$ . On a donc établi que  $\widehat{\mathbf{M}}$  est la projection orthogonale de  $\mathbf{X}$  sur  $\mathcal{F}$ . Les autres propriétés s'établissent par le même procédé en faisant usage du corollaire 4.2.1. Nous omettons les détails de la fin de cette démonstration.  $\square$

**Corollaire 5.7.1.** 1°) Sous les hypothèses (C)–(L), si  $p < n$ , alors l'estimateur de  $\sigma^2$  défini par

$$s^2 = \frac{1}{n-p} \|\mathbf{X} - \widehat{\mathbf{M}}\|^2, \quad (5.7.12)$$

est sans biais (et vérifie donc  $\mathbb{E}(s^2) = \sigma^2$ ).

2°) Si, de plus, l'hypothèse (N) est satisfaite, alors

$$(n-p) \frac{s^2}{\sigma^2} \stackrel{d}{=} \chi_{n-p}^2. \quad (5.7.13)$$

**Preuve.** C'est une conséquence directe du théorème 5.7.1.□

**Remarque 5.7.1.** Il convient de ne pas confondre  $s^2$  défini par (5.7.12) ci-dessus avec l'estimateur du maximum de vraisemblance  $\widehat{\sigma}^2$  de  $\sigma^2$  défini, sous les hypothèses (N)–(L), par

$$\widehat{\sigma}^2 = \frac{1}{n} \|\mathbf{X} - \widehat{\mathbf{M}}\|^2 = \left\{ \frac{n-p}{n} \right\} s^2. \quad (5.7.14)$$

Sous les hypothèses (N)–(L), la vraisemblance de  $\mathbf{X} \stackrel{d}{=} N_n(\mathbf{M}, \sigma^2 \mathbb{I}_n)$  est définie par

$$L(\mathbf{M}, \sigma^2) = \frac{1}{(2\sigma^2\pi)^{n/2}} \exp\left(-\frac{\|\mathbf{X} - \mathbf{M}\|^2}{2\sigma^2}\right). \quad (5.7.15)$$

A l'évidence, pour tout  $\sigma^2 > 0$  fixé, la maximum de  $L(\mathbf{M}, \sigma^2)$  est atteint pour  $\mathbf{M} = \widehat{\mathbf{M}}$ . On constate alors que

$$\frac{\partial}{\partial \sigma^2} \log L(\widehat{\mathbf{M}}, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \|\mathbf{X} - \widehat{\mathbf{M}}\|^2, \quad (5.7.16)$$

qui s'annule pour  $\sigma^2 = \widehat{\sigma}^2$ , donné par (5.7.14). Les estimateurs du maximum de vraisemblance de  $\mathbf{M}$  et  $\sigma^2$  sont donc donnés par  $\widehat{\mathbf{M}}$  et  $\widehat{\sigma}^2$ . Toutefois, alors que  $\widehat{\mathbf{M}}$  est sans biais, lorsque  $p < n$ , il n'en est pas de même pour  $\widehat{\sigma}^2$ , qui, d'une part est défini dans tous les cas, et d'autre part, vérifie (cette propriété restant vraie sous (C)–(L))

$$\mathbb{E}(\widehat{\sigma}^2) = \left\{ \frac{n-p}{n} \right\} \sigma^2. \quad (5.7.17)$$

A partir de ces relations, on observe que

$$\sup_{\mathbf{M}, \sigma} L(\mathbf{M}, \sigma^2) = L(\widehat{\mathbf{M}}, \widehat{\sigma}^2) = (2\pi)^{-n/2} \left\{ \frac{1}{n} \|\mathbf{X} - \widehat{\mathbf{M}}\|^2 \right\}^{-n/2} e^{-n/2}. \quad (5.7.18)$$

Il est intéressant de comparer cette expression à celle obtenue sous l'hypothèse  $\{\mathbf{M} = \mathbf{0}\}$ . Dans ce cas, la vraisemblance (5.7.15) devient

$$L(\mathbf{0}, \sigma^2) = \frac{1}{(2\sigma^2\pi)^{n/2}} \exp\left(-\frac{\|\mathbf{X}\|^2}{2\sigma^2}\right), \quad (5.7.19)$$

dont le maximum est atteint pour

$$\widehat{\sigma}^2 = \frac{1}{n} \|\mathbf{X}\|^2, \quad (5.7.20)$$

de sorte que

$$\sup_{\sigma^2} L(\mathbf{0}, \sigma^2) = L(\mathbf{0}, \widehat{\sigma}^2) = (2\pi)^{-n/2} \left\{ \frac{1}{n} \|\mathbf{X}\|^2 \right\}^{-n/2} e^{-n/2}. \quad (5.7.21)$$

On déduit de (5.7.18)–(5.7.21), et de la relation  $\|\mathbf{X}\|^2 = \|\mathbf{X} - \widehat{\mathbf{M}}\|^2 + \|\widehat{\mathbf{M}}\|^2$  la formule

$$\frac{\sup_{\sigma^2} L(\mathbf{0}, \sigma^2)}{\sup_{\mathbf{M}, \sigma} L(\mathbf{M}, \sigma^2)} = \left\{ \frac{\|\mathbf{X}\|^2}{\|\mathbf{X} - \widehat{\mathbf{M}}\|^2} \right\}^{-n/2} = \left\{ 1 + \frac{\|\widehat{\mathbf{M}}\|^2}{\|\mathbf{X} - \widehat{\mathbf{M}}\|^2} \right\}^{-n/2}. \quad (5.7.22)$$

Le test du rapport de vraisemblance de l'hypothèse  $\{\mathcal{M} = \mathbb{O}\}$ , contre l'alternative où  $\{\mathcal{M} \in \mathcal{F}\}$  est quelconque, rejette donc  $\{\mathcal{M} = \mathbb{O}\}$  pour les valeurs élevées de la statistique

$$\frac{\|\widehat{\mathcal{M}}\|^2/p}{\|\mathcal{X} - \widehat{\mathcal{M}}\|^2/(n-p)}, \quad (5.7.23)$$

qui suit, sous l'hypothèse  $\{\mathcal{M} = \mathbb{O}\}$ , une loi de Fisher centrée  $F_{p,n-p}$ , et, sous l'alternative où  $\{\mathcal{M} \in \mathcal{F}\}$  est quelconque, une loi de Fisher non centrée  $F_{p,n-p}(\frac{1}{\sigma^2}\|\mathcal{M}\|^2)$ .

### 5.7.2 Plan complet à 0 ou 1 facteur - Echantillon d'une loi $N(\mu, \sigma^2)$ .

Un plan *factoriel* est défini par des données formant un tableau multi-indices  $\{X_{i_1, \dots, i_k}, \mathbf{i} = (i_1, \dots, i_k) \in \mathcal{I}\}$ , où  $\mathcal{I}$  désigne un ensemble fini d'indices. Quitte à renuméroter les indices de  $\mathcal{I}$ , on peut toujours se ramener au cas où  $\mathcal{I} \subset \mathbb{N}_*^k = \{1, 2, \dots\}^k$ , ce que nous ferons par la suite. Le modèle de base généralement adopté suppose que les observations composantes sont normales et mutuellement indépendantes, avec une variance constante, et une moyenne  $m_{\mathbf{i}} = m_{i_1, \dots, i_k} = \mathbb{E}(X_{\mathbf{i}})$ , dépendant de l'indice  $\mathbf{i}$ . On dit que c'est un plan à  $k$  facteurs, si les indices coordonnées  $i_1, \dots, i_k$  sont au nombre de  $k$ . Le plan est dit *complet*, si l'ensemble  $\mathcal{I}$  des indices possibles est un ensemble produit, et s'il n'y a qu'une seule observation par indice possible. Il est alors commode de supposer que l'indice  $i_j$  varie dans un ensemble de la forme  $\{1, \dots, m_j\}$ , pour chaque choix de  $j = 1, \dots, k$ . Certains indices d'un plan factoriel peuvent correspondre à des *répétitions*, sous-entendu de la même expérience. De tels indices n'ont pas d'influence sur la loi de l'expérience, et on préfère leur donner le statut d'*indice de répétition*, plutôt que de *facteur*, ce dernier terme étant réservé aux indices qui peuvent influencer le résultat de l'expérience. De facto, on peut, sans perte de généralité, supposer que l'indice de répétition éventuel est *unique*. Lorsqu'il n'y a pas de répétitions, on parle d'un plan *sans répétitions*.

**Exemple 5.7.1.** Une expérience, composée d'un tableau de variables aléatoires indépendantes de la forme

$$X_{i,j,k} = m_{i,j} + \varepsilon_{i,j,k},$$

pour  $1 \leq i \leq I$ ,  $1 \leq j \leq J$  et  $1 \leq k \leq K_{i,j}$ , où  $m_{i,j}$  est un tableau de constantes, et les  $\varepsilon_{i,j,k} \stackrel{d}{=} N(0, \sigma^2)$  sont des variables aléatoires réelles [v.a.r.] indépendantes et idem-distribuées [i.i.d.], est un plan à 2 facteurs (influentiels), correspondant aux indices  $i$  et  $j$ , et un indice de répétition, correspondant à l'indice  $k$ . Lorsque  $K_{i,j} = 1$  pour tout choix de  $i$  et  $j$ , on obtient un plan à 2 facteurs sans répétition. L'indice  $k$  est alors inutile, et donc, omis.

Le plan factoriel à 1 facteur sans répétition est défini par une suite

$$X_i \stackrel{d}{=} N(\mu + \alpha_i, \sigma^2), \quad i = 1, \dots, n.$$

Un tel plan est statistiquement inexploitable (autrement que par des formules triviales) sans hypothèse additionnelle sur les  $\alpha_i$ , puisqu'il comporte un nombre de paramètres libres égal à  $n + 1$ , et donc supérieur aux observations. On peut, par contre, en donner une analyse statistique non triviale, lorsque la moyenne  $\mathbb{E}(X_i) = \mu + \alpha_i$  ne dépend pas de l'indice  $i$ . Ceci revient à supposer que  $\alpha_1 = \dots = \alpha_n = 0$ . Stricto-sensu, on obtient alors un plan à 0 facteur, puisque l'indice  $i$  n'est associé qu'à des répétitions. L'analyse statistique du plan à 0 facteur est présentée ci-dessous.

Le plan à 0 facteur correspond à des données structurées sous forme d'un échantillon aléatoire  $X_1, \dots, X_n$  d'une loi  $N(\mu, \sigma^2)$ . On suppose donc que  $X_1, \dots, X_n$  sont des v.a. indépendantes et de même loi  $N(\mu, \sigma^2)$ , avec  $\sigma > 0$ . Il est équivalent d'écrire que

$$X_i = \mu + \varepsilon_i \quad \text{pour } i = 1, \dots, n, \quad (5.7.24)$$

où  $\mu \in \mathbb{R}$  est un paramètre inconnu, et  $\varepsilon_1, \dots, \varepsilon_n$  sont des "erreurs" indépendantes, de même loi normale centrée  $N(0, \sigma^2)$ ,  $\sigma > 0$  étant un paramètre inconnu. On adopte la représentation vectorielle

$$\mathcal{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \in \mathbb{R}^n \quad \text{avec} \quad \mathcal{X} \stackrel{d}{=} N_n(\mathcal{M}, \sigma^2 \mathbb{I}), \quad \text{où} \quad \mathcal{M} = \mathbf{1}\mu = \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix} \in \mathcal{F}, \quad (5.7.25)$$

où  $\mathcal{F}$  est le sous-espace vectoriel dimension 1 de  $\mathbb{R}^n$  défini par

$$\mathcal{F} = \{\mathbf{1}m : m \in \mathbb{R}\} \quad (5.7.26)$$

La projection orthogonale de  $\mathcal{X}$  sur  $\mathcal{M}$  est alors

$$\widehat{\mathcal{M}} = \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathcal{X} = \mathbf{1}\bar{X} \quad \text{où} \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i. \quad (5.7.27)$$

Comme on a toujours une somme directe  $\mathbb{R}^n = \mathcal{F} \oplus \mathcal{F}^\perp$  entre  $\mathcal{F}$  (de dimension 1) et son orthogonal  $\mathcal{F}^\perp$  (de dimension  $n-1$ ), on peut appliquer le théorème 4.1 pour obtenir que

$$\widehat{\mathcal{M}} = \begin{bmatrix} \bar{X} \\ \vdots \\ \bar{X} \end{bmatrix} \quad \text{et} \quad \mathcal{X} - \widehat{\mathcal{M}} = \begin{bmatrix} X_1 - \bar{X} \\ \vdots \\ X_n - \bar{X} \end{bmatrix}$$

sont indépendants, et tels que

$$\frac{1}{\sigma^2} \|\widehat{\mathcal{M}}\|^2 = \frac{n\bar{X}^2}{\sigma^2} \stackrel{d}{=} \chi_1^2 \quad \text{et} \quad \frac{1}{\sigma^2} \|\mathcal{X} - \widehat{\mathcal{M}}\|^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \stackrel{d}{=} \chi_{n-1}^2. \quad (5.7.28)$$

On obtient ainsi la proposition suivante, avec les notations habituelles

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, \quad \text{et} \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad \text{si } n \geq 2.$$

**Proposition 5.7.2.** *Sous les hypothèses ci-dessus,  $\bar{X}$  et  $\hat{\sigma}^2$  sont indépendants, de lois respectives*

$$\sqrt{n}\{\bar{X} - \mu\} \stackrel{d}{=} N(0, \sigma^2), \quad \frac{n\hat{\sigma}^2}{\sigma^2} \stackrel{d}{=} \chi_{n-1}^2, \quad (5.7.29)$$

et, pour  $n \geq 2$ , le rapport

$$\frac{\sqrt{n}\{\bar{X} - \mu\}}{\hat{\sigma}\sqrt{\frac{n}{n-1}}} = \frac{\sqrt{n-1}\{\bar{X} - \mu\}}{\hat{\sigma}} = \frac{\sqrt{n}\{\bar{X} - \mu\}}{s} \stackrel{d}{=} t_{n-1}, \quad (5.7.30)$$

suit une loi de Student à  $n-1$  degrés de liberté.

**Preuve.** Par application du théorème 5.7.1.  $\square$

Sous réserve que  $n \geq 2$ , on utilise donc, dans la pratique usuelle, les estimateurs centrés de  $\mu$  et  $\sigma^2$ , donnés respectivement par

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \text{et} \quad s^2 = \tilde{\sigma}^2 = \frac{n\hat{\sigma}^2}{n-1} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

On obtient alors un *intervalle de confiance exact* de niveau  $1 - \alpha$  (où  $0 < \alpha < 1$  est arbitraire) pour  $\mu$ , en écrivant que

$$\mathbb{P}\left(\mu \in \left[\bar{X} - \frac{st_{n_1;\alpha/2}}{\sqrt{n}}, \bar{X} + \frac{st_{n_1;\alpha/2}}{\sqrt{n}}\right]\right) = 1 - \alpha. \quad (5.7.31)$$

Ici,  $t_{n_1;\alpha/2}$  désigne le quantile supérieur d'ordre  $\alpha/2$  de la loi de Student à  $n-1$  degrés de liberté. Le *test de Student* de l'hypothèse que  $\mu = \mu_0$  rejette cette hypothèse au seuil  $\alpha$ , si  $\mu = \mu_0$  n'appartient pas à l'intervalle de confiance défini par (5.7.31).

D'une manière générale, un *intervalle de confiance* pour un paramètre réel inconnu est un intervalle  $[c, d]$ , aux extrémités aléatoires ( $c$  et  $d$  sont des statistiques des observations), tel que la probabilité  $\mathbb{P}(\theta \in [c, d])$  vérifie certaines propriétés. Pour un *seuil*  $0 < \alpha < 1$  spécifié, on dit que l'intervalle est *exact*, si  $\mathbb{P}(\theta \in [c, d]) = 1 - \alpha$ . L'intervalle est dit *asymptotique* ou *approximatif*, lorsque  $\mathbb{P}(\theta \in [c, d]) \rightarrow 1 - \alpha$ , lorsque le nombre d'observations augmente indéfiniment.

**Remarque 5.7.2.** Il est important de noter que la relation (5.7.31) est une identité (ce qui équivaut à dire que l'intervalle de confiance de Student est exact), si et seulement si l'échantillon  $X_1, \dots, X_n$  est issu d'une loi normale  $N(0, \sigma^2)$ . Si cela n'est pas le cas, on obtient néanmoins (par une démonstration évidente utilisant le lemme de Slutsky) un intervalle de confiance asymptotique, sous réserve que  $0 < \sigma^2 = \text{Var}(X_1) < \infty$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\mu \in \left[\bar{X} - \frac{st_{n_1; \alpha/2}}{\sqrt{n}}, \bar{X} + \frac{st_{n_1; \alpha/2}}{\sqrt{n}}\right]\right) = 1 - \alpha. \quad (5.7.32)$$

Il est possible de montrer que la relation (5.7.32) demeure satisfaite, même dans des cas où  $\sigma^2 = \infty$ , pourvu que  $X_1, \dots, X_n$  soit dans le domaine d'attraction de la loi normale, ceci exprimant l'existence de suites de constantes  $a_n$  et  $b_n$ , telles que, lorsque  $n \rightarrow \infty$ ,

$$\frac{1}{a_n} \sum_{i=1}^n (X_i - b_n) \xrightarrow{d} N(0, 1).$$

Cette dernière condition est nécessaire et suffisante pour que l'intervalle de Student soit asymptotiquement exact.

**Exercice 5.7.1.** On considère un échantillon de taille  $n = 50$  d'une loi normale  $N(\mu, \sigma^2)$ . On donne les valeurs numériques

$$\bar{X} = 3.92 \quad \text{et} \quad \hat{\sigma} = 2.96.$$

Tester (par un test de Student), au seuil  $\alpha = 5\%$ , l'hypothèse  $(H_0)$  que  $\mu = 0$ , contre l'alternative. Comparer l'intervalle de confiance de Student pour  $\mu$  avec les intervalles de confiance asymptotiques basés sur  $\hat{\sigma}$  ou  $s$ , et le théorème central limite.

### 5.7.3 Evaluation de la précision des intervalles de confiance pour la moyenne.

Soit  $X = X_1, X_2, \dots, X_n$  un échantillon aléatoire d'une v.a.r.  $X$  non dégénérée, possédant les caractéristiques suivantes. On suppose que

$$\limsup_{|u|+|v| \rightarrow \infty} |\mathbb{E}(\exp(iuX + ivX^2))| < 1.$$

Cette condition technique est satisfaite (voir Hall (1992), p.71), en particulier, si  $X$  possède une densité sur  $\mathbb{R}$  relativement à la mesure de Lebesgue. On suppose, de plus, que  $\mathbb{E}(X^4) < \infty$ , et on note les moments de  $X$  par

$$\mu = \mathbb{E}(X), \quad \sigma^2 = \text{Var}(X), \quad \gamma = \sqrt{\beta_1} = \sigma^{-3} \mathbb{E}((X - \mu)^3), \quad \beta_2 = \sigma^{-4} \mathbb{E}((X - \mu)^4), \quad \kappa = \beta_2 - 3.$$

Notons la densité et la fonction de répartition de la loi normale  $N(0, 1)$ , respectivement, par

$$\varphi(x) = \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} \quad \text{et} \quad \Phi(x) = \int_{-\infty}^x \varphi(t) dt.$$

Posons, avec les notations habituelles,

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, \quad s^2 = \tilde{\sigma}^2 = \frac{n}{n-1} \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Sous ces hypothèses, on a (voir Hall (1992), pp.70-74), pour tout  $x \in \mathbb{R}$  fixé, lorsque  $n \rightarrow \infty$ ,

$$\begin{aligned} \mathbb{P}\left(\frac{\sqrt{n}(\bar{X} - \mu)}{\hat{\sigma}} \leq x\right) &= \Phi(x) - n^{-1/2} \varphi(x) \left\{ \frac{1}{6} \gamma (x^2 - 1) \right\} \\ &+ n^{-1} \varphi(x) x \left\{ \frac{1}{12} \kappa (x^2 - 3) + \frac{1}{18} \gamma^2 (x^4 + 2x^2 - 3) - \frac{1}{4} (x^2 + 3) \right\} + o(n^{-1}). \end{aligned} \quad (5.7.33)$$

On obtient aisément la version de (5.7.33) obtenue avec le remplacement formel de  $\hat{\sigma}$  par  $s = \tilde{\sigma}$ . Il suffit pour cela d'écrire que

$$\mathbb{P}\left(\frac{\sqrt{n}(\bar{X} - \mu)}{s} \leq x\right) = \mathbb{P}\left(\frac{\sqrt{n}(\bar{X} - \mu)}{\hat{\sigma}} \leq x \frac{s}{\hat{\sigma}}\right) = \mathbb{P}\left(\frac{\sqrt{n}(\bar{X} - \mu)}{s} \leq x \left\{1 - \frac{1}{n}\right\}^{-1/2}\right).$$

Comme, lorsque  $n \rightarrow \infty$ ,

$$\left\{1 - \frac{1}{n}\right\}^{-1/2} = 1 + \frac{1 + o(1)}{2n},$$

on voit donc que

$$\Phi\left(x \frac{s}{\hat{\sigma}}\right) = \Phi(x) + \frac{x + o(1)}{2n} \varphi(x).$$

La contribution des autres termes de (5.7.33) obtenus en remplaçant  $x$  par  $xs/\hat{\sigma}$  étant en  $o(1/n)$ , on en déduit que, pour tout  $x \in \mathbb{R}$  fixé, lorsque  $n \rightarrow \infty$ ,

$$\begin{aligned} \mathbb{P}\left(\frac{\sqrt{n}(\bar{X} - \mu)}{s} \leq x\right) &= \Phi(x) - n^{-1/2} \varphi(x) \left\{ \frac{1}{6} \gamma (x^2 - 1) \right\} \\ &+ n^{-1} \varphi(x) x \left\{ \frac{1}{12} \kappa (x^2 - 3) + \frac{1}{18} \gamma^2 (x^4 + 2x^2 - 3) - \frac{1}{4} (x^2 + 1) \right\} + o(n^{-1}). \end{aligned} \quad (5.7.34)$$

Pour la loi normale standard, on a  $\gamma = \kappa = 0$ . Par conséquent, si  $T_{n-1} \stackrel{d}{=} t_{n-1}$  désigne une v.a. suivant une loi de Student à  $n - 1$  degrés de liberté, on déduit de (5.7.33) que, pour tout  $x \in \mathbb{R}$  fixé, lorsque  $n \rightarrow \infty$ ,

$$\mathbb{P}(T_{n-1} \leq x) = \Phi(x) - n^{-1} \varphi(x) x \left\{ \frac{1}{4} (x^2 + 1) \right\} + o(n^{-1}). \quad (5.7.35)$$

En combinant (5.7.34) et (5.7.35), on obtient l'évaluation suivante de la précision de l'*approximation de Student*, obtenue en approximant le membre de gauche de (5.7.34) par  $\mathbb{P}(T_{n-1} \leq x)$ . Pour tout  $x \in \mathbb{R}$  fixé, lorsque  $n \rightarrow \infty$ ,

$$\begin{aligned} \mathbb{P}\left(\frac{\sqrt{n}(\bar{X} - \mu)}{s} \leq x\right) &= \mathbb{P}(T_{n-1} \leq x) - n^{-1/2} \varphi(x) \left\{ \frac{1}{6} \gamma (x^2 - 1) \right\} \\ &+ n^{-1} \varphi(x) x \left\{ \frac{1}{12} \kappa (x^2 - 3) + \frac{1}{18} \gamma^2 (x^4 + 2x^2 - 3) \right\} + o(n^{-1}). \end{aligned} \quad (5.7.36)$$

Ce dernier résultat montre que, en général, l'approximation de Student introduit une erreur en  $O(n^{-1/2})$  pour les distributions *dissymétriques*, correspondant à des valeurs de  $\gamma \neq 0$ , ce qui est relativement médiocre. Par contre, lorsque  $\gamma = 0$  et  $\kappa \neq 0$ , l'ordre d'approximation devient en  $O(1/n)$ , ce qui peut être considéré comme tout à fait satisfaisant, dans la pratique.

Les relations (5.7.35) et (5.7.36) mettent en évidence une propriété remarquable des intervalles de confiance symétriques, obtenus aussi bien par la méthode de Student que par des formules asymptotiques. Leur précision est toujours en  $O(1/n)$ .

#### 5.7.4 Plan complet à 2 facteurs.

On appelle *plan complet à 2 facteurs* une expérience structurée sous la forme

$$X_{i,j} = \mu + \alpha_i + \beta_j + \gamma_{i,j} + \varepsilon_{i,j}, \quad (5.7.37)$$

où on suppose, par convention, que  $\mu$ ,  $\alpha_i$ ,  $1 \leq i \leq I$ ,  $\beta_j$ ,  $1 \leq j \leq J$ , et  $\gamma_{i,j}$ ,  $1 \leq i \leq I$ ,  $1 \leq j \leq J$  sont des paramètres vérifiant les identités

$$\alpha_{\bullet} = \beta_{\bullet} = \gamma_{i,\bullet} = \gamma_{\bullet,j} = 0, \quad \forall 1 \leq i \leq I, 1 \leq j \leq J, \quad (5.7.38)$$

et  $\{\varepsilon_{i,j} : 1 \leq i \leq I, 1 \leq j \leq J\}$  est un tableau de variables aléatoires indépendantes de loi  $N(0, \sigma^2)$ . On posera  $m_{i,j} = \mathbb{E}(X_{i,j})$ , de sorte que

$$m_{i,j} = \mathbb{E}(X_{i,j}) = \mu + \alpha_i + \beta_j + \gamma_{i,j} \quad \text{pour } 1 \leq i \leq I, 1 \leq j \leq J.$$

Nous adopterons ici la convention que l'expérience concerne l'expérimentation de  $J$  *traitements*, numérotés de 1 à  $J$  sur  $I$  *patients*, numérotés de 1 à  $I$ . Les effets (relatifs) des patients sont désignés par  $\alpha_1, \dots, \alpha_I$ , les



effets relatifs des traitements sont désignés par  $\beta_1, \dots, \beta_J$ . Les *effets croisés*, ou *interactions*, forment le tableau  $\{\gamma_{i,j} : 1 \leq i \leq I, 1 \leq j \leq J\}$ .

Il est d'usage, pour les tableaux multi-indices, d'adopter les conventions suivantes. Le remplacement d'un ou de plusieurs indices par un point correspond à l'opération consistant à faire la somme par rapport à toutes les valeurs possibles de ces mêmes indices. Si, de plus, la variable est surmontée d'une barre, on remplace la sommation par la moyenne. A titre d'exemple, dans le cas présent,

$$\alpha_{\bullet} = \sum_{i=1}^I \alpha_i, \quad \bar{\alpha}_{\bullet} = \frac{1}{I} \sum_{i=1}^I \alpha_i, \quad \beta_{\bullet} = \sum_{i=1}^I \beta_i, \quad \bar{\beta}_{\bullet} = \frac{1}{I} \sum_{i=1}^I \beta_i, \quad (5.7.39)$$

$$\gamma_{i,\bullet} = \sum_{j=1}^J \gamma_{i,j}, \quad \bar{\gamma}_{i,\bullet} = \frac{1}{J} \sum_{j=1}^J \gamma_{i,j}, \quad \gamma_{\bullet,j} = \sum_{i=1}^I \gamma_{i,j}, \quad \bar{\gamma}_{\bullet,j} = \frac{1}{I} \sum_{i=1}^I \gamma_{i,j}, \quad (5.7.40)$$

$$\gamma_{\bullet,\bullet} = \sum_{i=1}^I \sum_{j=1}^J \gamma_{i,j}, \quad \bar{\gamma}_{\bullet,\bullet} = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \gamma_{i,j}. \quad (5.7.41)$$

On travaille sur l'espace  $E$  à  $n = IJ$  dimensions, dont les vecteurs sont désignés par le tableau de coordonnées  $\{x_{i,j}\}$ , où  $x_{i,j}$  est la coordonnée d'indices  $i \in \{1, \dots, I\}$  et  $j \in \{1, \dots, J\}$ . L'espace  $E$  est muni du produit scalaire

$$\langle \{x_{i,j}\}, \{y_{i,j}\} \rangle = \frac{1}{\sigma^2} \sum_{i=1}^I \sum_{j=1}^J x_{i,j} y_{i,j}, \quad (5.7.42)$$

et de la norme euclidienne

$$\|\{x_{i,j}\}\| = \langle \{x_{i,j}\}, \{x_{i,j}\} \rangle^{1/2} = \left\{ \frac{1}{\sigma^2} \sum_{i=1}^I \sum_{j=1}^J x_{i,j}^2 \right\}^{1/2}. \quad (5.7.43)$$

On introduit alors les sous-espaces de  $E$  définis par

$$\begin{aligned} F_0 &= \left\{ \{x_{i,j}\} : x_{i,j} = m \quad \forall i, j \right\}; \\ F_1 &= \left\{ \{x_{i,j}\} : x_{i,j} = a_i \quad \forall i, j \quad \text{et} \quad a_{\bullet} = 0 \right\}; \\ F_2 &= \left\{ \{x_{i,j}\} : x_{i,j} = b_j \quad \forall i, j \quad \text{et} \quad b_{\bullet} = 0 \right\}; \\ F_{12} &= \left\{ \{x_{i,j}\} : x_{i,\bullet} = x_{\bullet,j} = 0 \quad \forall i, j \right\}. \end{aligned}$$

On constate que les sous-espaces  $F_0, F_1, F_2, F_{12}$  sont mutuellement orthogonaux relativement au produit scalaire  $\langle \cdot, \cdot \rangle$ , et forment une somme directe, telle que

$$E = F_0 \oplus F_1 \oplus F_2 \oplus F_{12}, \quad \dim(E) = IJ, \quad \dim(F_0) = 1, \quad (5.7.44)$$

$$\dim(F_1) = I - 1, \quad \dim(F_2) = J - 1, \quad \dim(F_{12}) = (I - 1)(J - 1). \quad (5.7.45)$$

Les projections orthogonales de  $\{X_{i,j}\}$  sur ces sous-espaces sont donc indépendantes et fournissent des estimateurs sans biais  $\hat{\mu}, \hat{\alpha}_i, \hat{\beta}_j, \hat{\gamma}_{i,j}$  des paramètres  $\mu, \alpha_i, \beta_j, \gamma_{i,j}$ , donnés par

$$\begin{aligned} \hat{\mu} &= \bar{X}_{\bullet,\bullet}, \\ \hat{\alpha}_i &= \bar{X}_{i,\bullet} - \bar{X}_{\bullet,\bullet}, \\ \hat{\beta}_j &= \bar{X}_{\bullet,j} - \bar{X}_{\bullet,\bullet}, \\ \hat{\gamma}_{i,j} &= X_{i,j} - \bar{X}_{i,\bullet} - \bar{X}_{\bullet,j} + \bar{X}_{\bullet,\bullet}. \end{aligned}$$

Les sommes de carrés ci-dessous sont indépendantes et suivent des lois du Khi-deux, avec les paramètres :

$$\begin{aligned}\frac{1}{\sigma^2}SS_0 &= \|\{\widehat{\mu}\}\|^2 = \frac{IJ}{\sigma^2}\{\widehat{\mu}\}^2 \stackrel{d}{=} \chi_1^2(\|\{\mu\}\|^2) = \chi_1^2\left(\frac{IJ}{\sigma^2}\mu^2\right), \\ \frac{1}{\sigma^2}SS_\alpha &= \|\{\widehat{\alpha}_i\}\|^2 = \frac{J}{\sigma^2}\sum_{i=1}^I\{\widehat{\alpha}_i\}^2 \stackrel{d}{=} \chi_{I-1}^2(\|\{\alpha_i\}\|^2) = \chi_{I-1}^2\left(\frac{J}{\sigma^2}\sum_{i=1}^I\alpha_i^2\right), \\ \frac{1}{\sigma^2}SS_\beta &= \|\{\widehat{\beta}_j\}\|^2 = \frac{I}{\sigma^2}\sum_{j=1}^J\{\widehat{\beta}_j\}^2 \stackrel{d}{=} \chi_{J-1}^2(\|\{\beta_j\}\|^2) = \chi_{J-1}^2\left(\frac{I}{\sigma^2}\sum_{j=1}^J\beta_j^2\right), \\ \frac{1}{\sigma^2}SS_\gamma &= \|\{\widehat{\gamma}_{i,j}\}\|^2 = \frac{1}{\sigma^2}\sum_{i=1}^I\sum_{j=1}^J\{\widehat{\gamma}_{i,j}\}^2 \stackrel{d}{=} \chi_{(I-1)(J-1)}^2(\|\{\gamma_{i,j}\}\|^2) = \chi_{(I-1)(J-1)}^2\left(\frac{1}{\sigma^2}\sum_{i=1}^I\sum_{j=1}^J\gamma_{i,j}^2\right), \\ \frac{1}{\sigma^2}SS_T &= \|\{X_{i,j}\}\|^2 = \frac{1}{\sigma^2}\sum_{i=1}^I\sum_{j=1}^J X_{i,j}^2 \stackrel{d}{=} \chi_{IJ}^2(\|\{m_{i,j}\}\|^2) = \chi_{IJ}^2\left(\frac{1}{\sigma^2}\sum_{i=1}^I\sum_{j=1}^J m_{i,j}^2\right),\end{aligned}$$

On notera l'identité

$$SS_0 + SS_\alpha + SS_\beta + SS_\gamma = SS_T.$$

On obtient la *table d'analyse de variance* (ou d'ANOVA (pour *Analysis of Variance*) suivante.

Origine	D° de liberté	Somme carrés	Moy. de carrés	Espérance moy. carrés
Moy. $\mu$	1	$SS_0$	$MS_0 = \frac{SS_0}{1}$	$\sigma^2 + IJ\mu^2$
Pat. $\alpha$	$I - 1$	$SS_\alpha$	$MS_\alpha = \frac{SS_\alpha}{(I-1)}$	$\sigma^2 + \frac{J}{I-1}\sum_{i=1}^I\alpha_i^2$
Trait. $\beta$	$J - 1$	$SS_\beta$	$MS_\beta = \frac{SS_\beta}{(J-1)}$	$\sigma^2 + \frac{I}{J-1}\sum_{j=1}^J\beta_j^2$
Int. $\alpha$ - $\beta$	$(I - 1)(J - 1)$	$SS_\gamma$	$MS_\gamma = \frac{SS_\gamma}{(I-1)(J-1)}$	$\sigma^2 + \frac{1}{(I-1)(J-1)}\sum_{i=1}^I\sum_{j=1}^J\gamma_{i,j}^2$
Total	$IJ$	$SS_T$	$MS_t = \frac{SS_T}{IJ}$	$\sigma^2 + \frac{1}{IJ}\sum_{i=1}^I\sum_{j=1}^J m_{i,j}^2$

Dans cette table d'ANOVA, les moyennes de carrés sont obtenues en divisant les sommes de carrés par les degrés de liberté correspondants.

La pratique usuelle de la statistique consiste à supposer que l'une des sommes de carrés est *centrée* (ou, en d'autres termes, correspond à une loi du  $\chi^2$  centrée). Faute de mieux, on suppose, le plus souvent, qu'il n'y a pas d'interactions, c'est à dire que

$$\gamma_{i,j} = 0 \quad \text{pour tout } 1 \leq i \leq I \quad \text{et } 1 \leq j \leq J. \quad (5.7.46)$$

L'hypothèse (5.7.46) n'est, cependant, pas toujours naturelle, et doit être remise en question au cours de l'analyse, sous réserve que ce soit possible.

Sous l'hypothèse admise (5.7.46), la quantité

$$s^2 = MS_\gamma = \frac{1}{(I-1)(J-1)} \sum_{i=1}^I \sum_{j=1}^J (X_{i,j} - \bar{X}_{i,\bullet} - \bar{X}_{\bullet,j} + \bar{X}_{\bullet,\bullet})^2,$$

est un estimateur sans biais de  $\sigma^2$ . On peut alors l'utiliser pour tester des hypothèses sur les autres paramètres par des tests de Fisher. Par exemple, pour tester l'hypothèse d'absence d'effets de traitements (ce qui, ici, équivaut au fait que  $\beta_1 = \dots = \beta_J = 0$ ) contre l'alternative, on utilise le fait que, sous (5.7.46), on a l'identité en loi

$$\frac{MS_\beta}{MS_\gamma} \stackrel{d}{=} F_{J-1, (I-1)(J-1)} \left( \frac{I}{\sigma^2} \sum_{j=1}^J \beta_j^2 \right).$$

On rejette donc, au seuil  $0 < \alpha < 1$ , l'hypothèse que  $\beta_1 = \dots = \beta_J = 0$ , lorsque le rapport  $MS_\beta/MS_\gamma$  excède le quantile supérieure d'ordre  $\alpha$  d'une loi de Fisher (centrée)  $F_{J-1, (I-1)(J-1)}$ .

### 5.7.5 Régression linéaire multiple classique.

On considère le modèle  $\mathbf{X} \stackrel{d}{=} N_m(\mathcal{M}, \sigma^2 \mathbb{I})$ , où

$$\mathcal{M} \in \mathcal{F} = \left\{ A\theta : \theta \in \mathbb{R}^p \right\}, \quad (5.7.47)$$

et  $A$  est une matrice ( $m \times p$ ) connue, de rang  $\text{rg}(A) = p \leq m$  (la matrice  $A$  est dite *de plein rang de colonnes*). Nous verrons plus loin comment traiter le cas où le rang de  $A$  est quelconque, par l'emploi de pseudo-inverses de matrices.

La relation (5.7.47) montre que  $\mathcal{F}$  est un sous-espace vectoriel de  $\mathbb{R}^m$  de dimension  $\text{rg}(A) = p$ . Ceci implique, en particulier, que l'application

$$\theta \in \mathbb{R}^p \quad \rightarrow \quad A\theta \in \mathcal{F}, \quad (5.7.48)$$

définit une bijection de  $\mathbb{R}^p$  sur  $\mathcal{F}$ . On en déduit que l'estimateur linéaire sans biais à dispersion minimale  $\widehat{\mathcal{M}}$  de  $\mathcal{M}$  est lié à l'estimateur linéaire sans biais à dispersion minimale  $\hat{\theta}$  de  $\theta$  par l'identité

$$\widehat{\mathcal{M}} = A\hat{\theta}. \quad (5.7.49)$$

**Théorème 5.7.2.** *Sous les hypothèses ci-dessus, l'estimateur linéaire sans biais à dispersion minimale  $\widehat{\mathcal{M}}$  de  $\mathcal{M} \in \mathcal{F}$  est défini par  $\widehat{\mathcal{M}} = A\hat{\theta}$ , où*

$$\hat{\theta} = (A'A)^{-1} A'\mathbf{X} \stackrel{d}{=} N_p(\theta, \sigma^2 (A'A)^{-1}). \quad (5.7.50)$$

De plus,  $\mathbf{X} - \widehat{\mathcal{M}} = \mathbf{X} - A\hat{\theta} = \{\mathbb{I} - A(A'A)^{-1}A'\}\mathbf{X}$  et  $\widehat{\mathcal{M}} = A\hat{\theta}$  sont indépendants et vérifient

$$\frac{1}{\sigma^2} \|\widehat{\mathcal{M}}\|^2 = \mathbf{X}' \left\{ A(A'A)^{-1}A' \right\} \mathbf{X} \stackrel{d}{=} \chi_p^2 \left( \frac{1}{\sigma^2} \|\mathcal{M}\|^2 \right), \quad (5.7.51)$$

$$\frac{1}{\sigma^2} \|\mathbf{X} - \widehat{\mathcal{M}}\|^2 = \mathbf{X}' \left\{ \mathbb{I} - A(A'A)^{-1}A' \right\} \mathbf{X} \stackrel{d}{=} \chi_{m-p}^2. \quad (5.7.52)$$

**Preuve.** Par application du théorème 4.1.□

Comme application de ce résultat et de (5.7.23), on utilise la procédure suivante pour tester l'hypothèse  $\{\mathcal{M} = \mathcal{M}_0\}$ . On rejette cette hypothèse au seuil  $\alpha$  si la statistique

$$\frac{\|\widehat{\mathcal{M}} - \mathcal{M}_0\|^2/p}{\|\mathbf{X} - \widehat{\mathcal{M}}\|^2/(m-p)}, \quad (5.7.53)$$

excède le quantile supérieur d'ordre  $\alpha$  d'une loi de Fisher  $F_{p, m-p}$ .

# Chapitre 6

## Inverses généralisés

### 6.1 Factorisations de plein rang.

Soit une matrice réelle  $A$ ,  $(p \times q)$  (à  $p$  lignes et  $q$  colonnes), associée à l'application  $x \in \mathbb{R}^q \rightarrow Ax \in \mathbb{R}^p$ , et de la forme

$$A = \begin{bmatrix} a_1 & \dots & a_q \end{bmatrix} = \begin{bmatrix} b'_1 \\ \vdots \\ b'_p \end{bmatrix} = \mathcal{B}' \Leftrightarrow \mathcal{B} = \begin{bmatrix} b_1 & \dots & b_p \end{bmatrix}, \quad (6.1.1)$$

où  $a_1, \dots, a_q \in \mathbb{R}^p$  désignent les vecteurs colonnes de  $A$ , et  $b_1, \dots, b_p \in \mathbb{R}^q$  les vecteurs lignes de  $A$  (ou colonnes de  $A' = \mathcal{B}$ ). On associe à  $A$  les sous-espaces  $\text{Im}(A) \subseteq \mathbb{R}^p$  et  $\text{Ker}(A) \subseteq \mathbb{R}^q$  (respectivement *image* et *noyau* de l'application  $x \in \mathbb{R}^q \rightarrow Ax \in \mathbb{R}^p$ ), définis par

$$\text{Ker}(A) = \{x \in \mathbb{R}^q : Ax = \mathbb{O}\} = \{x \in \mathbb{R}^q : b'_i x = 0 \ \forall i = 1, \dots, p\},$$

$$\text{Im}(A) = \{Ay : y \in \mathbb{R}^q\} = \{y_1 a_1 + \dots + y_q a_q : y_1, \dots, y_q \in \mathbb{R}\}.$$

Lorsque  $A' = \mathcal{B} = \begin{bmatrix} b_1 & \dots & b_p \end{bmatrix}$ , il est commode d'introduire la notation  $\text{Or}(\mathcal{B})$  pour désigner

$$\text{Or}(\mathcal{B}) = \text{Or}(\text{Im}(\mathcal{B})) = \{x \in \mathbb{R}^p : b'x = 0 \ \forall b \in \text{Im}(\mathcal{B})\} = \{x \in \mathbb{R}^p : b'_i x = 0 \ \forall i = 1, \dots, p\}.$$

Ceci permet d'écrire

$$\text{Ker}(A) = \text{Or}(A'). \quad (6.1.2)$$

Le rang  $r = \text{rg}(A)$  de  $A$  est, par définition, la dimension  $\dim(\text{Im}(A))$  de  $\text{Im}(A)$ , sous-espace vectoriel de  $\mathbb{R}^p$  engendré par les vecteurs colonnes  $a_1, \dots, a_q$  de  $A$ . De manière équivalente,  $r$  est défini comme la plus grande valeur de  $m \geq 1$  telle qu'on puisse extraire de  $A$  une matrice  $(m \times m)$  de déterminant non nul, si un tel  $m$  existe, et  $r = 0$  autrement. Les relations évidentes suivantes lient  $r = \text{rg}(A)$ ,  $\text{rg}(A')$ ,  $p$  et  $q$ .

$$r = \text{rg}(A) = \text{rg}(A') = \dim(\text{Im}(A)) = q - \dim(\text{Ker}(A)), \quad 0 \leq r \leq \min(p, q). \quad (6.1.3)$$

**Définition 6.1.1.** Une matrice  $A$  est dite

- de plein rang de lignes     si  $r = \text{rg}(A) = p \leq q$ ,
- de plein rang de colonnes     si  $r = \text{rg}(A) = q \leq p$ ,
- de plein rang                     si  $r = \text{rg}(A) = \min\{p, q\}$ .

En général, si  $A$  est  $(p \times q)$  et  $B$  est  $(q \times s)$ , on a

$$\text{rg}(AB) \leq \min(\text{rg}(A), \text{rg}(B)) \leq \min(p, q, s). \quad (6.1.4)$$

**Définition 6.1.2.** On dira qu'une factorisation  $B_1 \cdots B_k = A$  d'une matrice  $A$ , ( $p \times q$ ), est de plein rang, si chacune des matrices  $B_1, \dots, B_k$  est de plein rang.

Si  $r = \text{rg}(A)$ , cette propriété impose que  $B_1$  soit ( $p \times r$ ) et de plein rang de colonnes,  $B_k$  soit ( $r \times q$ ) et de plein rang de lignes, et  $B_j$  soit ( $r \times r$ ) et régulière, pour  $j = 2, \dots, k-1$ , lorsque  $k \geq 3$ .

**Lemme 6.1.1.** Toute matrice ( $p \times q$ )  $A \neq \mathbb{O}_{p,q}$ , de rang  $\text{rg}(A) = r \geq 1$ , admet une factorisation de plein rang de la forme  $A = BC$ , où  $B$  est ( $p \times r$ ) et  $A$  est ( $r \times q$ )

**Preuve.** Supposons que les vecteurs colonnes  $a_{j_1}, \dots, a_{j_r}$  de  $A = [a_1 \ \dots \ a_q]$  forment une famille libre de  $\mathbb{R}^p$  engendrant  $\text{Im}(A)$ , sous-espace vectoriel de  $\mathbb{R}^p$  engendré par les colonnes  $a_1, \dots, a_q$  de  $A$ . Posons  $B = [a_{j_1} \ \dots \ a_{j_r}]$ . Il existe alors une matrice ( $r \times q$ )  $C$  telle que  $A = BC$ . Comme alors  $\text{rg}(A) = \text{rg}(B) = r \leq \text{rg}(C) \leq r$ , on a nécessairement  $\text{rg}(C) = r$ , ce qui achève la démonstration.  $\square$

## 6.2 Pseudoinverses

**Définition 6.2.1.** Soit  $A$  une matrice ( $p \times q$ ). Une matrice  $X$  ( $q \times p$ ) est appelée  $\{i_1, \dots, i_k\}$ -pseudoinverse (ou  $\{i_1, \dots, i_k\}$ -inverse) de  $A$ , si les relations  $\{i_1, \dots, i_k\}$ , prises parmi les relations  $\{1\}$ – $\{5\}$  suivantes, sont vérifiées.

$$\begin{aligned} \{1\} \quad AXA &= A, \\ \{2\} \quad XAX &= X, \\ \{3\} \quad (AX)' &= AX, \\ \{4\} \quad (XA)' &= XA, \\ \{5\} \quad AX &= XA. \end{aligned}$$

On dit que  $X$  est :

- Un pseudoinverse, inverse généralisé ou  $g$ -inverse, de  $A$ , si  $X$  est un  $\{1\}$ -pseudoinverse de  $A$ , ce qui est noté  $A = A^-$  ;
- Un pseudoinverse de Moore-Penrose de  $A$ , si  $X$  est un  $\{1, 2, 3, 4\}$ -pseudoinverse de  $A$ , ce qui est noté  $A = A^\dagger$ .

**Remarque 6.2.1.** (i) La relation  $\{5\}$  de la définition 5.3 n'a de sens que si les matrices  $A$  et  $X$  sont carrées ( $A$  et  $X$  sont toutes deux ( $p \times p$ )).

(ii) Le nom d'inverse généralisé (ou  $g$ -inverse) est justifié par la propriété suivante. Une matrice  $G$  est un  $g$ -inverse de  $A$  si et seulement si  $x = Gy$  est une solution du système linéaire  $Ax = y$  chaque fois que le système est *consistant* (c'est à dire, admet au moins une solution particulière  $x_0$  telle que  $Ax_0 = y$ ).

En effet, le système  $Ax = y$  admet une solution particulière  $x_0$  si et seulement si  $y = Ax_0$ , auquel cas  $x = Gy$  vérifie  $Ax = AGy = AGAx_0 = Ax_0 = y$ . Cette relation implique que  $AGAx_0 = Ax_0$ , égalité ne pouvant être satisfaite pour tout  $x_0 \in \mathbb{R}^q$  que si  $AGA = A$ , c'est à dire, que si  $G$  est effectivement un  $g$ -inverse (ou  $\{1\}$ -inverse) de  $A$ .

(iii) Le système  $Ax = y$  est *consistant* si et seulement si  $y \in \text{Im}(A)$ . Il est dit *inconsistant* si cette condition n'est pas satisfaite.

La notion de  $\{1, 2, 3, 4\}$ -pseudoinverse a été découverte par Moore (Moore, E.H. (1920). On the reciprocal of the general algebraic matrix (abstract). *Bull. Amer. Math. Soc.* **26** 394–395), et redécouverte par Penrose (Penrose, R. (1955). A generalized inverse for matrices. *Proc. Cambridge Philos. Soc.* **62** 673–677).

**Lemme 6.2.1.** Pour que la matrice ( $q \times p$ ),  $X$  soit un inverse généralisé (ou  $g$ -inverse, ou  $\{1\}$ -pseudoinverse) de la matrice ( $p \times q$ ),  $A$ , il faut et il suffit que l'une des propriétés suivantes soit satisfaite :

- (i) La matrice  $H = XA$  est idempotente et  $\text{rg}(H) = \text{rg}(A) = \text{tr}(XA)$  ;
- (ii) La matrice  $F = AX$  est idempotente et  $\text{rg}(F) = \text{rg}(A) = \text{tr}(AX)$ .

**Preuve** Soit  $X$  un  $\{1\}$ -pseudoinverse de  $A$ , c'est à dire une matrice  $X$  telle que  $AXA = X$ . On a alors  $X(AXA) = XA$ , et  $(AXA)X = AX$ , ce qui signifie que  $XA$  et  $AX$  sont *idempotentes*. En général, une matrice idempotente  $Y$  est telle que  $Y^2 - Y = Y(Y - \mathbb{I}) = \mathbb{O}$ . Ceci implique qu'elle est de la forme

$$Y = P^{-1}DP = P^{-1}\text{diag}(1, \dots, 1, 0, \dots, 0)P,$$

où le rang de  $D = \text{diag}(1, \dots, 1, 0, \dots, 0)$  est égal au nombre de fois où la valeur propre 1 de  $Y$  apparaît sur sa diagonale. Ceci permet d'écrire l'identité

$$\text{rg}(Y) = \text{rg}(D) = \text{tr}(D) = \text{tr}(Y).$$

Cette dernière propriété, appliquée à  $Y = XA$ , puis à  $Y = AX$ , implique que

$$\text{rg}(XA) = \text{tr}(XA) = \text{tr}(AX) = \text{rg}(AX).$$

Enfin, du fait que  $AXA = A$ , les inégalités

$$\text{rg}(A) \geq \text{rg}(AX) \geq \text{rg}(AXA) = \text{rg}(A),$$

et, de même,

$$\text{rg}(A) \geq \text{rg}(XA) \geq \text{rg}(AXA) = \text{rg}(A),$$

achèvent la démonstration de (i) et (ii), lorsque  $X$  est un  $\{1\}$ -pseudoinverse de  $A$ .

Réciproquement, si la matrice ( $q \times q$ )

$$H = XA = \begin{bmatrix} h'_1 \\ \vdots \\ h'_q \end{bmatrix}, \quad \text{où} \quad A = \begin{bmatrix} b'_1 \\ \vdots \\ b'_q \end{bmatrix},$$

est telle que  $\text{rg}(H) = \text{rg}(A)$ , alors  $\text{Im}(H') = \text{Im}(A')$ , et donc  $\text{Or}(H') = \text{Or}(A')$ . Maintenant, le fait que  $H$  soit idempotente est équivalent à  $H(\mathbb{I} - H) = \mathbb{O}$ , ou, ce qui revient au même, par (6.1.2),

$$\text{Im}(\mathbb{I} - H) \subseteq \text{Ker}(H) = \text{Or}(H') = \text{Or}(A') = \text{Ker}(A).$$

Ceci implique que  $A(\mathbb{I} - H) = \mathbb{O}$ , soit encore  $A(\mathbb{I} - XA) = \mathbb{O}$ , et donc  $A = AXA$ . On a ainsi prouvé que  $X$  est un inverse généralisé de  $A$ . La démonstration est similaire lorsque  $AX$  est idempotente.  $\square$

**Théorème 6.2.1.** *Soit  $A$  une matrice ( $p \times q$ ) et soit  $A^-$  un inverse généralisé (ou  $g$ -inverse ou  $\{1\}$ -pseudoinverse) de  $A$ . Alors :*

(i) *La solution générale de l'équation  $Ax = \mathbb{O}$  est donnée par*

$$x = (\mathbb{I}_q - A^-A)z, \quad \text{où } z \in \mathbb{R}^q \text{ est arbitraire};$$

(ii) *La solution générale de l'équation  $Ax = y$ , lorsque cette équation est consistante, est donnée par*

$$x = A^-y + (\mathbb{I}_q - A^-A)z, \quad \text{où } z \in \mathbb{R}^q \text{ est arbitraire};$$

(iii) *Une condition nécessaire et suffisante pour que  $Ax = y$  soit consistante est que :*

$$AA^-y = y.$$

**Preuve** Procédons comme pour la démonstration du lemme 5.1, en posant  $H = A^-A$ . Le fait que  $A^-$  soit  $g$ -inverse de  $A$  équivaut alors à

$$A(\mathbb{I}_q - H) = A - AA^-A = \mathbb{O},$$

ce qui, par (6.1.2), implique l'inclusion  $\text{Im}(\mathbb{I}_q - H) \subseteq \text{Ker}(A) = \text{Or}(A')$ . D'autre part, par le lemme 5.2, l'hypothèse que  $A^-$  soit  $g$ -inverse de  $A$  implique que  $H = A^-A$  soit idempotente. En écrivant la matrice ( $q \times q$ )  $H$  sous la forme

$$H = P^{-1} \text{diag}(1, \dots, 1, 0, \dots, 0)P,$$

on voit que

$$\mathbb{I}_q - H = P^{-1}(\mathbb{I}_q - \text{diag}(1, \dots, 1, 0, \dots, 0))P = P^{-1} \text{diag}(0, \dots, 0, 1, \dots, 1)P,$$

ce qui implique l'égalité  $\text{rg}(\mathbb{I}_q - H) = q - \text{rg}(H) = q - \text{rg}(A) = \text{rg}(\text{Or}(A'))$ . Cette égalité, jointe au fait, établi plus haut, que  $\text{Im}(\mathbb{I}_q - H) \subseteq \text{Or}(A')$ , implique l'égalité  $\text{Im}(\mathbb{I}_q - H) = \text{Or}(A')$ , et donc la propriété (i).

On a vu (cf. remarque 5.1(ii)) que  $x_0 = A^-y$  fournit une solution particulière en  $x \in \mathbb{R}^q$  de l'équation  $Ax = y$  lorsque le système est consistant. Pour obtenir la solution générale de ce système, il suffit d'ajouter à la solution particulière  $x_0 = A^-y$  la solution générale de l'équation  $Ax = \mathbb{O}$ . Celle-ci étant de la forme  $(\mathbb{I}_q - H)z$  par le (i), on conclut donc ainsi à la validité de la propriété (ii).

Pour établir la propriété (iii), on remarque tout d'abord que lorsque  $AA^-y = y$ , le système  $Ax = y$  est bien consistant puisqu'il admet la solution particulière  $x_0 = A^-y$ . Inversement, s'il existe un  $x \in \mathbb{R}^q$  vérifiant  $y = Ax$ , alors, du fait que  $AA^-A = A$ , on a les égalités  $AA^-y = AA^-Ax = Ax = y$ , qui impliquent que  $AA^-y = y$ .  $\square$

Dans la suite de l'exposé, nous utiliserons les notations  $\|u\| = \langle u, u \rangle^{1/2} = (u'u)^{1/2}$  et  $\langle u, v \rangle = u'v$  pour désigner respectivement la norme et le produit scalaire euclidien dans l'espace  $\mathbb{R}^d$  (indépendamment de la valeur de  $d$ ).

**Définition 6.2.2.** On dit que l'inverse généralisé (ou  $g$ -inverse, ou  $\{1\}$ -pseudoinverse)  $G$  de  $A$  définit une solution de norme minimale  $\tilde{x} = Gy$  de l'équation consistante  $Ax = y$  en  $x \in \mathbb{R}^q$ , si, pour tout  $y \in \text{Im}(A)$ ,

$$\|Gy\| = \inf_{x \in \mathbb{R}^q : Ax=y} \|x\|.$$

**Théorème 6.2.2.** Soit  $A^-$  un  $\{1\}$ -pseudoinverse de  $A$ . Pour tout  $y \in \text{Im}(A)$ ,  $\tilde{x} = A^-y$  fournit une solution de norme minimale de l'équation  $Ax = y$  en  $x \in \mathbb{R}^q$ , si et seulement si la matrice  $A^-$  est un  $\{1, 4\}$ -pseudoinverse de  $A$ .

**Preuve.** Soit  $G$  un  $\{1\}$ -pseudoinverse de  $A$ . Par le théorème 5.1(ii), toutes les solutions en  $x \in \mathbb{R}^q$  de l'équation  $Ax = y$  sont de la forme  $x = Gy + (\mathbb{I}_q - GA)z$ , où  $z \in \mathbb{R}^q$  est arbitraire. Le fait que  $x = Gy$  soit de norme minimale parmi l'ensemble des solutions de cette équation revient donc à écrire que

$$\|Gy\| \leq \|Gy + (\mathbb{I}_q - GA)z\| \quad \text{pour tout } z \in \mathbb{R}^q \text{ et } y \in \text{Im}(A). \quad (6.2.1)$$

Lorsque  $x$  varie dans  $\mathbb{R}^q$ ,  $y = Ax$  parcourt  $\text{Im}(A)$ . On peut donc faire dans l'équation ci-dessus le changement de variable  $y = Ax$ . On constate donc que (6.2.1) équivaut à

$$\|GAx\|^2 \leq \|GAx + (\mathbb{I}_q - GA)z\|^2 \quad \text{pour tout } z \in \mathbb{R}^q \text{ et } x \in \mathbb{R}^q. \quad (6.2.2)$$

En développant (6.2.2) on obtient que

$$\begin{aligned} \|GAx\|^2 &\leq \|GAx + (\mathbb{I}_q - GA)z\|^2 = \|GAx\|^2 + \|(\mathbb{I}_q - GA)z\|^2 + 2\langle GAx, (\mathbb{I}_q - GA)z \rangle \\ &\text{pour tout } x \in \mathbb{R}^q \text{ et } z \in \mathbb{R}^q, \end{aligned}$$

ce qui n'est possible que si  $\langle GAx, (\mathbb{I}_q - GA)z \rangle = 0 \quad \forall z \in \mathbb{R}^q$ , ce qui équivaut à

$$(GA)'(\mathbb{I}_q - GA) = \mathbb{O} \quad \Leftrightarrow \quad (\mathbb{I}_q - GA)'(GA) = \mathbb{O}. \quad (6.2.3)$$

Comme le lemme 5.2(i) implique que  $GA$  est idempotente, il en est de même pour  $(GA)'$ . On a donc  $(GA)'(\mathbb{I}_q - (GA)') = \mathbb{O}$ , et, par conséquent,  $(GA)' = (GA)'(GA) = (GA)'(GA)'$ . De ce fait, si  $x \in \mathbb{R}^q$  est tel que  $(GA)x = \mathbb{O}$ , on a nécessairement  $(GA)'x = (GA)'(GA)x = \mathbb{O}$ . Ceci implique que  $\text{Ker}(GA) \subseteq \text{Ker}((GA)')$ . Comme  $\dim(\text{Ker}(GA)) = q - \text{tr}(GA) = q - \text{tr}((GA)') = \dim(\text{Ker}(GA)'),$  on a nécessairement  $\text{Ker}(GA) = \text{Ker}((GA)').$

On répète ce raisonnement à partir de (6.2.3), qui implique que  $(\mathbb{I}_q - GA)'(\mathbb{I}_q - (\mathbb{I}_q - GA)) = \mathbb{O}$ . Comme  $(\mathbb{I}_q - GA)'$  est idempotente, on aboutit aux égalités  $(\mathbb{I}_q - GA)' = (\mathbb{I}_q - GA)'(\mathbb{I}_q - GA) = (\mathbb{I}_q - GA)'(\mathbb{I}_q - GA)'$ , qui montrent que  $\text{Ker}(\mathbb{I}_q - GA) \subseteq \text{Ker}((\mathbb{I}_q - GA)') = \text{Ker}(\mathbb{I}_q - (GA)')$ . Ces espaces devant être de même dimension, sont identiques. Comme  $GA$  et  $(GA)'$  sont idempotents, on a  $\text{Im}(GA) = \text{Ker}(\mathbb{I}_p - GA)$  et  $\text{Im}((GA)') = \text{Ker}(\mathbb{I}_p - (GA)')$ . On obtient ainsi la relation  $\text{Im}(GA) = \text{Im}((GA)')$ . Compte tenu de l'égalité déjà établie  $\text{Ker}(GA) = \text{Ker}((GA)')$ , le fait que  $GA$  et  $(GA)'$  soient idempotents suffit à conclure que  $GA = (GA)'$ .  $G$  est donc bien un  $\{1, 4\}$ -pseudoinverse de  $A$ .

Réciproquement, si  $G$  est un  $\{1, 4\}$ -pseudoinverse de  $A$ , on a (6.2.3), et, par suite (6.2.1).  $\square$

**Définition 6.2.3.** On dit que l'inverse généralisé (ou  $\{1\}$ -pseudoinverse)  $G$  de  $A$  définit une solution  $\hat{x} = Gy$  en moindres carrés de l'équation (pas nécessairement consistante)  $Ax = y$  en  $x \in \mathbb{R}^q$ , si, pour tout  $y \in \mathbb{R}^p$ ,

$$\|A\hat{x} - y\| = \inf_{x \in \mathbb{R}^q} \|Ax - y\|.$$

**Théorème 6.2.3.** Soit  $A^-$  un  $\{1\}$ -pseudoinverse de  $A$ . Pour tout  $y \in \mathbb{R}^p$ ,  $\hat{x} = A^-y$  fournit une solution en moindres carrés de l'équation (pas nécessairement consistante)  $Ax = y$  en  $x \in \mathbb{R}^q$ , si et seulement si  $A^-$  est un  $\{1, 3\}$ -inverse de  $A$ . De plus, dans ce cas, toutes les solutions  $\hat{x}$  en moindres carrés de l'équation  $Ax = y$  en  $x \in \mathbb{R}^q$ , lorsque  $y$  varie dans  $\mathbb{R}^p$ , sont données par

$$\hat{x} = A^-y + (\mathbb{I} - A^-A)z, \quad \text{où } z \in \mathbb{R}^q \text{ est arbitraire.}$$

**Preuve** Soit  $G$  un  $\{1\}$ -pseudoinverse de  $A$ , vérifiant donc  $AGA = A$ . Le fait que  $\hat{x} = Gy$  définisse une solution en moindres carrés de  $Ax = y$  pour tout  $y \in \mathbb{R}^p$  équivaut au fait que

$$\|AGy - y\| \leq \|Ax - y\| \quad \text{pour tout } x \in \mathbb{R}^q \text{ et } y \in \mathbb{R}^p.$$

Faisons le changement de variable  $z = x - Gy$ . Comme alors  $\|Ax - y\| = \|AGy - y + Az\|$ , l'inéquation ci-dessus équivaut à

$$\begin{aligned} \|AGy - y\|^2 &\leq \|AGy - y + Az\|^2 = \|(AG - \mathbb{I}_p)y\|^2 + \|Az\|^2 + 2\langle (AG - \mathbb{I}_p)y, Az \rangle \\ &\text{pour tout } y \in \mathbb{R}^p \text{ et } z \in \mathbb{R}^q. \end{aligned}$$

Or, ceci n'est possible que si  $\langle Az, (AG - \mathbb{I}_p)y \rangle = 0$ , pour tout  $y \in \mathbb{R}^p$  et  $z \in \mathbb{R}^q$ . Cette dernière relation équivaut à  $A'AG = A'$ . On obtient, par conséquent, que  $G'A'AG = G'A'$ , soit  $(AG)'(AG) = (AG)'$ . Partant de ce point, la démonstration s'achève en répétant les mêmes arguments que ceux utilisés pour prouver le théorème 5.2.  $\square$

**Définition 6.2.4.** On dit que l'inverse généralisé (ou  $\{1\}$ -pseudoinverse)  $G$  de la matrice  $(p \times q)$   $A$  est un pseudoinverse de norme minimale en moindres carrés de  $A$ , si  $G$  est tel que :

- (i) Pour tout  $y \in \mathbb{R}^p$ ,  $\hat{x} = Gy$  est une solution en moindres carrés de l'équation  $Ax = y$  en  $x \in \mathbb{R}^q$  ;
- (ii) Pour tout  $y \in \mathbb{R}^p$ , on a

$$\|\hat{x}\| = \|Gy\| \leq \|\tilde{x}\| \quad \text{pour tout } \tilde{x} \in \mathbb{R}^q, \text{ solution en moindre carré de l'équation } Ax = y, \quad (6.2.4)$$

c'est à dire, telle que  $\|A\tilde{x} - y\| = \|A\hat{x} - y\| \leq \|Ax - y\|$  pour tout  $x \in \mathbb{R}^q$ .

**Remarque 6.2.2.** La propriété pour  $A^-$  d'être un pseudoinverse de norme minimale en moindres carrés (selon la définition 5.6) est *plus forte* que celle d'être à la fois un pseudoinverse donnant, par  $x = A^-y$ , une solution de norme minimale de  $Ax = y$  (selon la définition 5.4), et, par  $x = A^-y$ , une solution en moindres carrés de  $Ax = y$  (selon la définition 5.5). En effet, alors que la propriété (i) de la définition 5.6 équivaut aux conditions de la définition 5.4, la propriété (ii) de la définition 5.6 est plus restrictive que les conditions imposées par la définition 5.5. On constate, en effet, que la définition 5.5 requiert que l'inégalité

$$\|Gy\| \leq \|\tilde{x}\| \quad \text{pour tout } \tilde{x} \in \mathbb{R}^q, \text{ solution en moindres carrés de l'équation } Ax = y,$$

soit satisfaite seulement pour les  $y \in \text{Im}(A)$  (c'est à dire tels que  $Ax = y$  est consistante, auquel cas  $\tilde{x}$  vérifie  $A\tilde{x} = y$ ). La définition 5.6, quant à elle, impose cette inégalité pour tous les  $y \in \mathbb{R}^p$ .



**Théorème 6.2.4.** *Soit  $A^-$  un  $\{1\}$ -pseudoinverse de  $A$ . Pour que  $A^-$  soit un pseudoinverse de norme minimale en moindres carrés de  $A$ , il faut et il suffit que  $A^-$  soit un  $\{1, 2, 3, 4\}$ -inverse de  $A$ , c'est à dire, un pseudoinverse de Moore–Penrose de  $A$ .*

**Preuve** Soit  $G$  un  $\{1\}$ -pseudoinverse de  $A$ , (vérifiant donc  $AGA = A$ ). Compte tenu de la remarque 5.2, le fait que  $G$  soit un pseudoinverse de norme minimale en moindres carrés impose qu'il satisfait à la fois aux conditions imposées par les définitions 5.4 et 5.5. Par application directe des théorèmes 5.2 et 5.3, ceci implique donc que  $G$  est un  $\{1, 3, 4\}$ -pseudoinverse de  $A$ . Il reste à établir que, sous cette hypothèse,  $G$  est un  $\{2\}$ -pseudoinverse de  $A$ , c'est à dire, vérifie  $GAG = G$ .

Pour cela, on utilise le théorème 5.3 pour écrire que l'ensemble des solutions  $\tilde{x}$  en moindres carrés de l'équation  $Ax = y$  sont données par

$$\tilde{x} = Gy + (\mathbb{I} - GA)z \quad \text{où } z \text{ varie dans } \mathbb{R}^q.$$

Compte tenu de ceci, dans le contexte des hypothèses du théorème, la relation (6.2.4) peut donc se réécrire sous la forme

$$\begin{aligned} \|Gy\|^2 &\leq \|Gy + (\mathbb{I}_q - GA)z\|^2 = \|Gy\|^2 + \|(\mathbb{I} - GA)z\|^2 + 2\langle Gy, (\mathbb{I}_q - GA)z \rangle \\ &\text{pour tout } y \in \mathbb{R}^p \text{ et } z \in \mathbb{R}^q. \end{aligned}$$

Comme ceci équivaut au fait que  $\langle Ry, (\mathbb{I} - GA)z \rangle = 0$  pour tout  $y \in \mathbb{R}^p$  et  $z \in \mathbb{R}^q$ , on aboutit naturellement à

$$G'(\mathbb{I}_q - GA) = 0 \quad \Leftrightarrow \quad G' = G'GA.$$

Or,  $G$  est un  $\{4\}$ -pseudoinverse et vérifie  $GA = (GA)'$ . De ce fait, on a nécessairement

$$G' = G'GA = G'(GA)' = G'A'G' \quad \Leftrightarrow \quad G = GAG.$$

On établit ainsi que  $G$  est un  $\{1, 2, 3, 4\}$ -pseudoinverse de  $A$ . La réciproque est directe en inversant les arguments que nous avons utilisés ci-dessus.  $\square$

**Lemme 6.2.2.** *Pour toute matrice  $A$ , le pseudoinverse de Moore–Penrose  $A^\dagger$ , s'il existe, est unique.*

**Preuve** Soient  $X$  et  $Y$ , vérifiant les conditions  $\{1\}$ ,  $\{2\}$ ,  $\{3\}$  et  $\{4\}$  de la définition 5.3. Alors

$$\begin{aligned} X &= XAX = X(AX) = X(AX)' = XX'A' = XX'(AYA)' = XX'((AY)A)' = XX'A'(AY)' \\ &= X(AX)'(AY) = (XAX)(AY) = XAY = (XA)'Y = A'X'Y = (AYA)'X'Y \\ &= (A(YA))'X'Y = (YA)'A'X'Y = (YA)(XA)'Y = (YA)(XA)Y = Y(AXA)Y = YAY \\ &= Y, \end{aligned}$$

ce qui établit l'unicité de  $X = A^\dagger$  s'il existe.  $\square$

**Exemple 6.2.1.** Compte tenu du lemme 5.3, pour montrer qu'une matrice  $X$  est égale à  $A^\dagger$ , il suffit de vérifier que  $X$  satisfait les propriétés  $\{1\}$ ,  $\{2\}$ ,  $\{3\}$  et  $\{4\}$ . Par ce procédé, on obtient les exemples suivants :

(i)  $\mathbb{O}_{p,q}^\dagger = \mathbb{O}_{q,p}$  ;

(ii)  $\mathbb{1}_{p,q}^\dagger = (pq)^{-1} \mathbb{1}_{q,p}$  ;

(iii) Si  $A$  est une matrice symétrique ( $p \times p$ ) de rang  $r$ , décomposée sous la forme

$$A = H^{-1} \text{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0) H,$$

où  $H$  est une matrice orthogonale (telle que  $H' = H^{-1}$ ), alors

$$\begin{aligned} A^\dagger &= H^{-1} \text{diag}(\lambda_1^{-1}, \dots, \lambda_r^{-1}, 0, \dots, 0) H \\ &= H' \text{diag}(\lambda_1^{-1}, \dots, \lambda_r^{-1}, 0, \dots, 0) H \end{aligned}$$

est un  $\{1, 2, 3, 4, 5\}$ -inverse de  $A$ . De plus, toute matrice de la forme

$$\begin{aligned} A^- &= H^{-1} \operatorname{diag}(\lambda_1^{-1}, \dots, \lambda_r^{-1}, \rho_{r+1}, \dots, \rho_d) H \\ &= H' \operatorname{diag}(\lambda_1^{-1}, \dots, \lambda_r^{-1}, \rho_{r+1}, \dots, \rho_d) H \end{aligned}$$

est encore un  $\{1, 3, 4\}$ -inverse de  $A$ .

(iv) Si  $A$  est une matrice ( $p \times q$ ) de plein rang de lignes, i.e., telle que  $\operatorname{rg}(A) = p \leq q$ , alors,

$$A^\dagger = A'(AA')^{-1}, \quad (6.2.5)$$

auquel cas,  $A^\dagger = A'(AA')^{-1}$  est un inverse à droite de  $A$ , c'est à dire une matrice telle que

$$AA^\dagger = \mathbb{I}_p.$$

(v) Si  $A$  est une matrice ( $p \times q$ ) de plein rang de colonnes, i.e., telle que  $\operatorname{rg}(A) = q \leq p$ , alors,

$$A^\dagger = (A'A)^{-1}A', \quad (6.2.6)$$

auquel cas,  $A^\dagger = A'(AA')^{-1}$  est un inverse à gauche de  $A$ , c'est à dire une matrice telle que

$$A^\dagger A = \mathbb{I}_q.$$

(vi) Si  $A$  est une matrice carrée inversible, alors  $A^\dagger = A^{-1}$ . C'est le seul cas possible où il existe à la fois un inverse à droite  $A_d$  de  $A$  (ce qui n'est possible que si  $A$  est de plein rang de lignes) et un inverse à gauche  $A_g$  de  $A$  (ce qui n'est possible que si  $A$  est de plein rang de colonnes). Dans ce dernier cas,  $A_g A A_d = A_d = A_g = A^{-1}$ .

**Théorème 6.2.5.** *Pour toute matrice  $A$ , le pseudo-inverse de Moore–Penrose  $A^\dagger$  existe et est unique.*

**Preuve.** Soit  $A$  une matrice ( $p \times q$ ). L'unicité de  $A^\dagger$  lorsqu'elle existe ayant été établie dans le Lemme 5.3, il suffit de prouver l'existence de cette matrice. Si  $A = \mathbb{O}_{p,q}$ , alors, par le (i) de l'exemple 5.1, on a  $A^\dagger = \mathbb{O}_{q,p}$ . Dans le cas où  $A \neq \mathbb{O}_{p,q}$ , le lemme 4.1 montre l'existence d'une factorisation de plein rang de  $A$ , de la forme  $A = BC$ , où  $B$  est ( $p \times r$ ),  $C$  est ( $r \times q$ ), et  $\operatorname{rg}(A) = \operatorname{rg}(B) = \operatorname{rg}(C) = r$ . Nous allons maintenant établir un résultat, dû à C.C. MacDuffee (1959) (voir Theorem 5, p.23 de Ben–Israel et Greville (Ben–Israel, A. et Greville, T.N.E. (1974). *Generalized Inverses : Theory and Applications*. Wiley, New York), montrant que

$$A^\dagger = C^\dagger B^\dagger = C'(CC')^{-1}(B'B)^{-1}B' = C'(B'AC')^{-1}B'. \quad (6.2.7)$$

Remarquons que la dernière égalité de (6.2.7) s'obtient en inversant l'identité  $B'AC' = (B'B)(CC')$ . Pour établir la deuxième égalité de (6.2.7), on remarque, par les (iv) et (v) de l'exemple 5.1, que la matrice  $X$  définie par  $X = C^\dagger B^\dagger$  vérifie  $X = C^\dagger B^\dagger = C'(CC')^{-1}(B'B)^{-1}B'$ . On vérifie enfin que la matrice  $X$  ainsi définie est solution des relations  $\{1\}$ ,  $\{2\}$ ,  $\{3\}$  et  $\{4\}$  de la définition 5.3. On a, en effet :

- (i)  $AXA = BCC'(CC')^{-1}(B'B)^{-1}B'BC = BC = A$ ;
- (ii)  $XAX = C'(CC')^{-1}(B'B)^{-1}B'BCC'(CC')^{-1}(B'B)^{-1}B' = C'(CC')^{-1}(B'B)^{-1}B' = X$ ;
- (iii)  $AX = BCC'(CC')^{-1}(B'B)^{-1}B' = B(B'B)^{-1}B' = (AX)'$ ;
- (iv)  $XA = C'(CC')^{-1}(B'B)^{-1}B'BC = C'(CC')^{-1}C = (XA)'$ .

$X$  est donc bien un  $\{1, 2, 3, 4\}$ -inverse de  $A$ , ce qui conclut la preuve de l'existence et l'unicité de  $A^\dagger$ .  $\square$

**Théorème 6.2.6.** *Les propriétés suivantes sont satisfaites par les pseudo-inverses de Moore–Penrose.*

- (i)  $A^{\dagger\dagger} = A$ ;
- (ii)  $A^\dagger = A^{\dagger'}$ ;
- (iii)  $A^\dagger = (A'A)^\dagger A' = A'(AA')^\dagger$ ;
- (iv)  $(A'A)^\dagger = A^\dagger A^{\dagger'}$ ;

$$(v) \quad \text{Si } \alpha \in \mathbf{R}, \text{ alors } (\alpha A)^\dagger = \begin{cases} \alpha^{-1}A^\dagger & \text{si } \alpha \neq 0, \\ \mathbb{O} = 0 \times A^\dagger & \text{si } \alpha = 0. \end{cases}$$

$$(vi) \quad \text{rg}(A) = \text{rg}(A^\dagger) = \text{rg}(A^\dagger A) = \text{rg}(AA^\dagger) = \text{tr}(A^\dagger A);$$

$$(vii) \quad (A \otimes B)^\dagger = A^\dagger \otimes B^\dagger.$$

**Preuve.** Les relations (i) et (ii) sont évidentes du fait que, si le couple  $(A, X)$  vérifie les propriétés {1}, {2}, {3} et {4} de la définition 5.1, il en est de même des couples  $(X, A)$  et  $(A', X')$ .

Pour établir la première égalité de (iii), vérifions que  $X = (A'A)^\dagger A'$  vérifie les propriétés {1}, {2}, {3} et {4} de la définition 5.3. On obtient {3} et {4} en remarquant que les matrices  $XA = (A'A)^\dagger A'A$ , et (par le (ii))  $AX = A(A'A)^\dagger A'$  sont symétriques. La relation (1),  $XAX = X$ , est évidente, tandis que le {2} s'obtient en écrivant les identités

$$AXA = A(A'A)^\dagger A'A = A'^\dagger A'A(A'A)^\dagger A'A = A'^\dagger A'A = A.$$

La deuxième égalité de (iii) se démontre de manière analogue.

La relation (v) est évidente, compte tenu de (i) de l'exemple 5.1. Pour (vi), on observe tout d'abord que  $A^\dagger A$  est idempotente, i.e. telle que  $(A^\dagger A)(A^\dagger A) = A^\dagger A$ . Cette propriété est d'ailleurs vraie en général si on remplace  $A$  par un {1}-inverse ou un {2}-inverse de  $A$ . De ce fait,  $\text{rg}(A^\dagger A) = \text{tr}(A^\dagger A)$ . On constate ensuite que  $\text{rg}(AA^\dagger A) = \text{rg}(A) \leq \text{rg}(A^\dagger)$ , et de même,  $\text{rg}(A^\dagger AA^\dagger) = \text{rg}(A^\dagger) \leq \text{rg}(A)$ , d'où  $\text{rg}(A) = \text{rg}(A^\dagger)$ . Le même raisonnement montre que  $\text{rg}(A) = \text{rg}(AA^\dagger) = \text{rg}(A^\dagger A)$ , et permet de conclure.

Pour (vii), on observe, par le (iv) de la propriété 1.1, que  $(A \otimes B)(C \otimes D) = AC \otimes BD$ . On vérifie alors directement que  $A^\dagger \otimes B^\dagger$  est un {1, 2, 3, 4}-inverse de  $A \otimes B$ .  $\square$

**Remarque 6.2.3.** (1) En général, lorsque les matrices  $(p \times q)$   $B$  et  $(q \times s)$   $C$  sont arbitraires, on n'a pas nécessairement  $(BC)^\dagger = C^\dagger B^\dagger$ . L'exemple suivant (Exemple 3.1, p.31 de Cline, R.E. (1979). *Elements of the Theory of Generalized Inverses for Matrices*, U.M.A.P., Newton, Massachusetts, U.S.A.) est un contre-exemple de cette identité. Soient

$$B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad B^\dagger = (B'B)^{-1}B' = \frac{1}{2} \begin{bmatrix} 2 & -2 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 2 & 0 & 0 \\ -2 & 1 & 1 \end{bmatrix},$$

$$C = \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & -1 \end{bmatrix}, \quad C^\dagger = C'(CC')^{-1} = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 2 & -2 \\ -2 & 3 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 2 & -2 \\ 0 & 1 \\ 0 & -1 \end{bmatrix},$$

$$CB = (CB)^\dagger = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \neq B^\dagger C^\dagger = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

(2) On a, par contre, la relation  $(B_1 \cdots B_k)^\dagger = B_k^\dagger \cdots B_1^\dagger$  chaque fois que  $B_1 \cdots B_k$  est une factorisation de plein rang de  $A = B_1 \cdots B_k$ . En effet, si c'est le cas, on peut réécrire cette factorisation sous la forme  $A = BDC$ , où  $B = B_1$  est de plein rang de colonnes,  $D = B_2 \cdots B_{k-1}$  est carrée et régulière, et  $C = B_k$  est de plein rang de lignes. Dans ce cas, on peut constater que la matrice

$$X = C'(CC')^{-1}D^{-1}(B'B)^{-1}B' = C^\dagger D^\dagger B^\dagger = B_k^\dagger B_{k-1}^\dagger \cdots B_2^\dagger B_1^\dagger$$

vérifie les propriétés {1}, {2}, {3}, {4} de la définition 5.3, et coïncide donc avec  $A^\dagger$ .

Deuxième partie

# Modèle Linéaire Multivarié



# Chapitre 1

## Matrices aléatoires

### 1.1 Règles de changement de variables-I.

Pour introduire les notions qui vont suivre, nous considérons tout d'abord un changement de variable dans l'intégrale double

$$I = \iint_D f(\mathbf{X}) dx_1 dx_2 = \int_D f(\mathbf{X}) dx_1 dx_2,$$

où  $D \subseteq \mathbb{R}^2$  désigne un ouvert de  $\mathbb{R}^2$ , avec la notation

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in D \subseteq \mathbb{R}^2.$$

Supposons maintenant que  $\mathbf{X} = \mathbf{X}(\mathbf{Y})$ , soit une fonction continûment différentiable ainsi que son inverse (un difféomorphisme) de

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \in D' \subseteq \mathbb{R}^2,$$

dans lui-même, et que  $D = \mathbf{X}(D')$  où  $D' \subseteq \mathbb{R}^2$  est un ouvert (en bijection bicontinue avec  $D$ ). Sous ces hypothèses, il est possible d'effectuer un changement de variables dans le calcul de  $I$ , en écrivant cette intégrale sous la forme

$$I = \int_{D'} f(\mathbf{X}) dx_1 dx_2 = \int_{D'} f(\mathbf{X}(\mathbf{Y})) \left| \frac{d\mathbf{X}}{d\mathbf{Y}} \right| dy_1 dy_2, \quad (1.1.1)$$

où

$$\left| \frac{d\mathbf{X}}{d\mathbf{Y}} \right| = \left| \frac{d(x_1, x_2)}{d(y_1, y_2)} \right| = \left| \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix} \right| \quad (1.1.2)$$

désigne le *Jacobien* du changement de variable, c'est à dire, la valeur absolue du déterminant de la différentielle  $D(\mathbf{X}(\mathbf{Y}))$  de la fonction  $\mathbf{X}(\mathbf{Y})$  par rapport à  $\mathbf{Y}$ , celle-ci étant définie par l'identité

$$d\mathbf{X} = \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix} = \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix} \begin{bmatrix} dy_1 \\ dy_2 \end{bmatrix} = D(\mathbf{X}(\mathbf{Y})) d\mathbf{Y}. \quad (1.1.3)$$

Pour réaliser ce changement de variable de manière directe à partir des différentielles des coordonnées,

$$dx_1 = \frac{\partial x_1}{\partial y_1} dy_1 + \frac{\partial x_1}{\partial y_2} dy_2 \quad \text{et} \quad dx_2 = \frac{\partial x_2}{\partial y_1} dy_1 + \frac{\partial x_2}{\partial y_2} dy_2,$$

on introduit le *produit extérieur*  $dx_1 \wedge dx_2$  des formes différentielles  $dx_1$  et  $dx_2$ . Celui-ci est défini formellement comme la règle de calcul qui justifie l'identité

$$dx_1 \wedge dx_2 = \left( \frac{\partial x_1}{\partial y_1} dy_1 + \frac{\partial x_1}{\partial y_2} dy_2 \right) \wedge \left( \frac{\partial x_2}{\partial y_1} dy_1 + \frac{\partial x_2}{\partial y_2} dy_2 \right) = \det \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{bmatrix} \{ dy_1 \wedge dy_2 \}$$

$$= \left( \frac{\partial x_1}{\partial y_1} \frac{\partial x_2}{\partial y_2} - \frac{\partial x_1}{\partial y_2} \frac{\partial x_2}{\partial y_1} \right) \{dy_1 \wedge dy_2\}.$$

On a ainsi effectué le produit extérieur de 2 formes différentielles dans  $\mathbb{R}^2$ . Le procédé précédent se généralise sans peine, pour établir les règles de calcul permettant de définir le produit extérieur de  $r$  formes différentielles dans  $\mathbb{R}^m$  pour  $1 \leq r \leq m$ . On pose ainsi

$$\left( \sum_{i_1=1}^m h_{i_1,1}(\mathbf{y}) dy_{i_1} \right) \wedge \dots \wedge \left( \sum_{i_r=1}^m h_{i_r,r}(\mathbf{y}) dy_{i_r} \right) = \left\{ \sum_{1 \leq i_1 < \dots < i_r \leq m} h_{i_1,1}(\mathbf{y}) \dots h_{i_r,r}(\mathbf{y}) \right\} dy_{i_1} \wedge \dots \wedge dy_{i_r},$$

en utilisant les identités, valables pour tout couple d'entiers  $p, q$ , tels que  $1 \leq p, q \leq r$ ,

$$dy_{i_1} \wedge \dots \wedge dy_{i_p} \wedge \dots \wedge dy_{i_q} \wedge \dots \wedge dy_{i_r} = -dy_{i_1} \wedge \dots \wedge dy_{i_q} \wedge \dots \wedge dy_{i_p} \wedge \dots \wedge dy_{i_r},$$

ce qui donne, en particulier, pour  $p = q$ ,

$$dy_{i_1} \wedge \dots \wedge dy_{i_p} \wedge \dots \wedge dy_{i_p} \wedge \dots \wedge dy_{i_r} = 0.$$

Lorsque, à l'aide de ces règles de calcul, on effectue le produit extérieur de  $r$  formes différentielles sur  $\mathbb{R}^m$ , on obtient une forme différentielle extérieure de degré  $r$ , selon la définition ci-dessous.

**Définition 1.1.1.** Une forme différentielle extérieure (resp. extérieure continue) de degré  $r$  dans  $\mathbb{R}^m$  est une expression de la forme

$$\sum_{1 \leq i_1 < \dots < i_r \leq m} h_{i_1, \dots, i_r}(\mathbf{y}) dy_{i_1} \wedge \dots \wedge dy_{i_r},$$

où les fonctions  $h_{i_1, \dots, i_r}(\mathbf{y})$  sont des fonctions (respectivement, des fonctions continues), à valeurs réelles, de  $\mathbf{y} \in \mathbb{R}^m$ .

Dans le but de faciliter les calculs de changements de variables dans des intégrales multiples, nous considérerons ici uniquement des formes différentielles extérieures de degré  $m$  dans  $\mathbb{R}^m$ . Ces dernières sont réduites à une seule expression de la forme  $h(\mathbf{y}) dy_1 \wedge \dots \wedge dy_m$ . De plus, comme les Jacobiens de tels changements de variables sont toujours pris en compte *en valeur absolue*, nous adopterons la convention implicite qu'une forme différentielle extérieure de degré  $m$  intervenant dans une intégrale multiple n'est définie qu'*au signe près*. Cette convention permet de ne pas attacher une importance particulière à l'*ordre global* dans lequel est effectué un produit extérieur, et de noter indifféremment, par abus de langage

$$d(\mathbf{y}) = \bigwedge_{i=1}^m dy_i = \bigwedge_{1 \leq i \leq m} dy_i = dy_1 \wedge \dots \wedge dy_m = |dy_1 \wedge \dots \wedge dy_m|.$$

Compte tenu de ces remarques, si  $\mathbf{y}$  est un vecteur de  $\mathbb{R}^m$  variant dans  $D$ , soit

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \in D \subseteq \mathbb{R}^m,$$

où  $D$  désigne un ouvert de  $\mathbb{R}^m$ , nous écrirons, dans toute la suite, indifféremment

$$\int_D h(\mathbf{y}) d(\mathbf{y}) = \int_D h(\mathbf{y}) dy_1 \wedge \dots \wedge dy_m = \int_D h(\mathbf{y}) |dy_1 \wedge \dots \wedge dy_m| = \int_D h(\mathbf{y}) dy_1 \dots dy_m.$$

Nous considérerons dans la suite plusieurs cas particuliers, obtenus lorsque la variable d'intégration varie dans des espaces de matrices soumises à des restrictions spécifiques. Les notations suivantes seront alors adoptées.

(i)  $X = [x_{i,j}] \in M_{p,q}$  varie librement dans l'ensemble  $\mathcal{M}_{p,q}$  des matrices réelles ( $p \times q$ ). On pose

$$(dX) = \bigwedge_{i=1}^p \bigwedge_{j=1}^q dx_{i,j}.$$

(ii)  $Y = [y_{i,j}]$  varie librement dans l'ensemble des matrices réelles ( $p \times p$ ) symétriques (i.e., telles que  $y_{i,j} = y_{j,i}$  pour  $1 \leq i, j \leq p$ ). On pose

$$[dY] = \bigwedge_{1 \leq i \leq j \leq p} dy_{i,j}.$$

On adopte ici, ainsi que dans (iii-iv) ci-dessous, la notation  $[dY]$ , plutôt que  $(dY)$ , pour mettre en évidence le fait que le produit est distinct de celui, adopté dans (i), pour le cas d'une matrice non structurée. Par la suite, on utilisera, lorsqu'il n'y aura pas d'ambiguïté, indifféremment la notation  $(dY)$  ou  $[dY]$ .

(iii)  $T = [t_{i,j}]$  varie librement dans l'ensemble des matrices réelles ( $p \times p$ ) triangulaires supérieures (i.e., telles que  $t_{i,j} = 0$  pour  $i > j$ ). On pose

$$[dT] = \bigwedge_{1 \leq i \leq j \leq p} dt_{i,j}.$$

(iv)  $H$  varie librement dans l'ensemble  $\mathcal{O}_m$  des matrices orthogonales réelles ( $m \times m$ ). Nous renvoyons aux formules (1.2.1), (1.2.4) et (1.4.8) qui suivent, pour la définition de  $[dH]$  dans ce dernier cas.

On observera ici qu'il ne peut avoir d'ambiguïté entre les notations adoptées au (i), (ii), (iii) et (iv) ci-dessus, car les matrices  $X$ ,  $Y$ ,  $T$  et  $H$  sont soumises à des contraintes différentes dans chacun des quatre cas précédents.

Les propositions ci-dessous énoncent une suite de formules de changement de variables dont l'utilisation s'avèrera commode par la suite.

**Proposition 1.1.1.** Soient  $X$  et  $Y$  deux matrices ( $n \times m$ ), telles que  $X = BY$ , où  $B$  désigne une matrice régulière ( $n \times n$ ) constante. Alors

$$(dX) = (\det B)^m (dY). \quad (1.1.4)$$

**Preuve.** Posons  $X = [x_1, \dots, x_m] = [By_1, \dots, By_m]$ , où  $Y = [y_1, \dots, y_m]$ . De toute évidence,  $(dX) = \bigwedge_{i=1}^m (dx_i) = \bigwedge_{i=1}^m ((\det B)(dy_i))$ , d'où le résultat. On notera ici que (1.1.4) reste valable si  $B$  n'est pas une fonction de  $Y$ .  $\square$

**Proposition 1.1.2.** Soient  $X$  et  $Y$  deux matrices ( $n \times m$ ), telles que  $X = BYC$ , où  $B$  désigne une matrice régulière ( $n \times n$ ) constante, et  $C$  est une matrice régulière ( $m \times m$ ) constante. Alors

$$(dX) = (\det B)^m (\det C)^n (dY). \quad (1.1.5)$$

**Preuve.** Posons  $Z = YC$ , ou, ce qui revient au même,  $Z' = C'Y'$ . Par application de (1.1.4), on constate que  $(dZ) = (dZ') = (\det C')^n (dY') = (\det C)^n (dY)$ . Comme  $X = BZ$ , une nouvelle application de (1.1.4) montre que  $(dX) = (\det B)^m (dZ)$ , ce qui permet de conclure. On notera ici que (1.1.5) reste valable si  $B$  et  $C$  ne sont pas fonction de  $Y$ .  $\square$

**Proposition 1.1.3.** Soient  $X$  et  $Y$  deux matrices symétriques ( $m \times m$ ), liées par la relation  $X = BYB'$ , où  $B$  est une matrice régulière ( $m \times m$ ) constante. Alors

$$[dX] = (\det B)^{m+1} [dY]. \quad (1.1.6)$$

**Preuve.** Comme  $X = BYB'$ ,  $dX = B \cdot dY \cdot B'$ , et, par conséquent,  $[dX] = p(B)[dY]$ , où  $p(B)$  est une fonction (nécessairement) polynomiale des coefficients de  $B$ . On vérifie aisément que cette fonction polynomiale satisfait l'identité, pour tout choix de  $B_1, B_2 \in \mathcal{M}_{m,m}$ ,

$$p(B_1 B_2) = p(B_1) p(B_2).$$



On constate alors que les seuls polynômes possibles de ce type sont de la forme  $p(B) = (\det B)^k$ , pour une valeur convenable de  $k$ . Pour trouver la valeur de  $k$ , il suffit de considérer un cas particulier. Prenant  $B = \text{diag}(b, 1, \dots, 1)$ , et posant  $Y = (y_{i,j})$ , on obtient que

$$BYB' = \begin{bmatrix} b^2 y_{1,1} & by_{1,2} & \dots & by_{1,m} \\ by_{1,2} & y_{2,2} & \dots & y_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ by_{1,m} & y_{2,m} & \dots & y_{m,m} \end{bmatrix},$$

ce qui montre que  $[B \cdot dY \cdot B'] = b^{m+1}(dY)$ . On en déduit, dans ce cas particulier, que  $p(B) = b^{m+1} = (\det B)^{m+1}$ . Ceci montre que  $k = m + 1$ , et établit ainsi (1.1.6).  $\square$

**Proposition 1.1.4.** Soit  $A = [a_{i,j}] > 0$  une matrice symétrique définie positive ( $m \times m$ ), et soit la matrice triangulaire supérieure ( $m \times m$ ) unique

$$T = [t_{i,j}] = \begin{bmatrix} t_{1,1} & \dots & t_{1,m} \\ \vdots & \ddots & \vdots \\ 0 & \dots & t_{m,m} \end{bmatrix},$$

à éléments diagonaux (strictement) positifs (i.e., tels que  $t_{i,i} > 0 \forall 1 \leq i \leq m$ ) telle que  $A = T'T$ . Alors

$$[dA] = 2^m \prod_{i=1}^m t_{i,i}^{m-i+1} [dT]. \quad (1.1.7)$$

**Preuve.** L'existence et l'unicité de  $T$  ont été établies dans (2.4.2). On développe le produit  $A = T'T$ , en écrivant que

$$\begin{bmatrix} a_{1,1} & \dots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \dots & a_{m,m} \end{bmatrix} = \begin{bmatrix} t_{1,1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ t_{1,m} & \dots & t_{m,m} \end{bmatrix} \begin{bmatrix} t_{1,1} & \dots & t_{1,m} \\ \vdots & \ddots & \vdots \\ 0 & \dots & t_{m,m} \end{bmatrix},$$

ce qui fournit en effectuant le produit de matrices

$$\begin{aligned} a_{1,1} &= t_{1,1}^2 & a_{1,2} &= t_{1,1}t_{1,2} & \dots & a_{1,m} &= t_{1,1}t_{1,m} \\ a_{2,2} &= t_{1,2}^2 + t_{2,2}^2 & \dots & a_{2,m} &= t_{1,2}t_{2,m} + t_{2,2}t_{2,m} \\ & & & \ddots & & \vdots & \\ a_{m,m} &= t_{1,m}^2 + \dots + t_{m,m}^2. \end{aligned}$$

On différencie, dans l'ordre,  $a_{1,1}, \dots, a_{1,m}$ , puis  $a_{2,2}, \dots, a_{2,m}$ , et ainsi de suite, jusqu'à  $a_{m,m}$ . A chaque fois, on ne retient pas les termes faisant intervenir des différentielles  $dt_{i,j}$  qui ont déjà été observées dans les termes précédents. Les termes additionnels correspondants sont alors désignés par "+\*". Ils ont vocation à disparaître dans le produit extérieur final. On obtient ainsi le tableau

$$\begin{aligned} da_{1,1} &= 2t_{1,1}dt_{1,1}, & \dots, & da_{1,m} &= t_{1,1}dt_{1,m} + * \\ da_{2,2} &= 2t_{2,2}dt_{2,2} + *, & \dots, & da_{2,m} &= t_{2,2}dt_{2,m} + * \\ & \vdots & & & \\ da_{m,m} &= 2t_{m,m}dt_{m,m} + *, \end{aligned}$$

ce qui fournit (1.1.7), par produit extérieur de ces formes différentielles.  $\square$

## 1.2 Loi gamma multivariée et loi de Wishart.

Dans ce qui suit, nous allons appliquer ces règles de changement de variables au calcul de certaines intégrales liées à la loi normale. Nous évaluons tout d'abord la *fonction Gamma multivariée*.

**Définition 1.2.1.** Pour tout entier  $m \geq 1$ , et pour tout réel  $r \in \mathbb{R}$ , vérifiant  $r > \frac{1}{2}(m-1)$ , on définit la fonction Gamma multivariée  $\Gamma_m(r)$  d'ordre  $m$  par l'intégrale

$$\Gamma_m(r) = \int_{\mathcal{A}>0} (\det \mathcal{A})^{r-(m+1)/2} \text{etr}(-\mathcal{A}) [d\mathcal{A}], \quad (1.2.1)$$

où  $\mathcal{A}$  varie dans l'ensemble des matrices symétriques définies positives ( $m \times m$ ).

**Remarque 1.2.1.** (i) La fonction  $\text{etr}(X)$ , de la variable matricielle  $X$  qui apparaît dans (1.2.1) a été introduite, sous forme systématique, par Herz (Herz, C.S. (1955). Bessel functions of matrix argument. *Ann. Math.* **61** 474–523), et est définie par

$$\text{etr}(X) = \exp(\text{tr}(X)). \quad (1.2.2)$$

(ii) Nous verrons plus loin (voir (3.3.3)) que  $\text{etr}(X) = {}_0F_0(X)$  s'interprète également comme une fonction hypergéométrique matricielle de  $X$ .

(iii) Il est parfois utile de faire intervenir la notion de *transformée de Laplace matricielle*. D'une manière générale, si  $f(\mathcal{A})$  désigne une fonction de la matrice ( $m \times m$ ) définie positive  $\mathcal{A} > 0$ , la transformée de Laplace de  $f(\mathcal{A})$  est la fonction  $g(Z) = \tilde{f}(Z)$  de la matrice symétrique ( $m \times m$ ) complexe  $Z = U + iV$ , avec  $U = \text{Re}(Z)$  et  $V = \text{Im}(Z)$  (voir (2.1.2)), définie par

$$g(Z) = \tilde{f}(Z) = \int_{\mathcal{A}>0} f(\mathcal{A}) \text{etr}(-\mathcal{A}Z) [d\mathcal{A}], \quad (1.2.3)$$

lorsque cette intégrale a un sens.

En général, on peut toujours décomposer la matrice carrée  $M = [m_{i,j}]$  en  $M = M_S + M_A$ , où  $M_S = \frac{1}{2}(M + M')$  est *symétrique*, et  $M_A = \frac{1}{2}(M - M')$  est *antisymétrique*. Avec ces notations, on constate que

$$\text{tr}(M) = \sum_{i=1}^m m_{i,i} = \text{tr}(M_S).$$

A cause de ceci, dans l'intégrale (1.2.3), le fait que  $\mathcal{A}$  soit symétrique implique que

$$\text{etr}(-\mathcal{A}Z) = \text{etr}\left(-\frac{1}{2}\mathcal{A}(Z + Z')\right),$$

ce qui explique qu'on limite la définition de  $g(Z) = \tilde{f}(Z)$  aux *matrices complexes symétriques*  $Z$ , telles que  $Z = Z'$ , alors qu'on aurait pu s'attendre à travailler plutôt sur des matrices autoadjointes (c'est à dire, telles que  $Z' = \overline{Z}$ ).

Par ailleurs, on observera que, si  $Z = U + iV$ , les matrices  $U = \text{Re}(Z)$  et  $V = \text{Im}(Z)$  étant réelles,

$$|\text{etr}(-\mathcal{A}Z)| = |\exp(\text{tr}(-\mathcal{A}U)) \exp(\text{tr}(-i\mathcal{A}V))| = |\exp(\text{tr}(-\mathcal{A}U))| = |\exp(\text{tr}(-\mathcal{A} \cdot \text{Re}(Z)))|,$$

ce qui permet d'établir que

$$\int_{\mathcal{A}>0} |f(\mathcal{A}) \text{etr}(-\mathcal{A} \cdot \text{Re}(Z))| [d\mathcal{A}] < \infty \quad \Rightarrow \quad \tilde{f}(Z) = \int_{\mathcal{A}>0} f(\mathcal{A}) \text{etr}(-\mathcal{A}Z) [d\mathcal{A}] \in \mathbb{C} \quad \text{existe.}$$

De cette dernière formule, on déduit que si  $U_0 \geq 0$  est une matrice ( $m \times m$ ) positive telle que

$$\sup_{\mathcal{A}>0} |f(\mathcal{A}) \text{etr}(-\mathcal{A}U_0)| < \infty,$$

alors

$$\tilde{f}(Z) = \int_{\mathcal{A}>0} f(\mathcal{A}) \text{etr}(-\mathcal{A}Z) [d\mathcal{A}] \in \mathbb{C}$$

existe pour toute matrice symétrique  $Z$  telle que  $Re(Z) > U_0$ . Dans la pratique, toutes les transformées de Laplace du type (1.2.3) que nous considérerons par la suite seront de ce type, et convergentes lorsque  $Re(Z) > U_0$  pour une matrice  $U_0 \geq 0$  convenable. Lorsque tel est le cas, il est possible de vérifier que  $g(Z)$  est une fonction analytique de  $Z = [z_{ij}]$  (ou ce qui revient au même des  $z_{ij}$ ) dans le domaine des matrices symétriques complexes défini par  $re(Z) > U_0$ .

Le grand intérêt des transformées de Laplace est que la connaissance de la fonction transformée  $g(Z) = \tilde{f}(Z)$  détermine complètement la fonction initiale  $f(Z)$ . De plus, par l'unicité du prolongement analytique, la connaissance de  $g(Z)$  pour  $Z$  symétrique réelle telle que  $Z > U_0$  détermine complètement  $g(Z)$  pour les matrices  $Z$  telles que  $Re(Z) > U_0$ . Nous nous référons à Herz (1955) (voir Herz, C.S. (1955). Bessel functions of matrix arguments. *Ann. Math.* **61** 474–523) pour une description plus générale des propriétés de la transformation de Laplace matricielle, ainsi qu'à Muirhead (1982), pp.252–253 (voir Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York), pour des formules inverses exprimant  $f$  à partir de  $g$ . Nous donnerons dans (1.2.11) ci-dessous un premier exemple utile de transformée de Laplace  $\tilde{f}(Z)$  obtenu pour  $f(Z) = (\det Z)^{a-(m+1)/2}$ .

**Proposition 1.2.1.** *Pour tout entier  $m \geq 1$ , et pour tout réel  $r \in \mathbb{R}$  vérifiant  $r > \frac{1}{2}(m-1)$ , on a*

$$\Gamma_m(r) = \int_{\mathcal{A} > 0} (\det \mathcal{A})^{r-(m+1)/2} \text{etr}(-\mathcal{A}) [d\mathcal{A}] = \pi^{\frac{m(m-1)}{4}} \prod_{i=1}^m \Gamma\left(r - \frac{1}{2}(i-1)\right). \quad (1.2.4)$$

**Preuve.** On effectue le changement de variable  $\mathcal{A} \leftrightarrow T$ , où  $T = [t_{i,j}]$  est l'unique matrice triangulaire supérieure à éléments diagonaux positifs définie par  $\mathcal{A} = T'T$  (voir (2.4.2)). La condition  $\mathcal{A} > 0$  équivaut à faire varier  $t_{i,i}$  dans  $(0, \infty)$  pour  $i = 1, \dots, m$ , et  $t_{i,j}$  dans  $\mathbb{R} = (-\infty, \infty)$  pour  $1 \leq i < j \leq m$ . Un calcul direct montrant que

$$\text{tr } \mathcal{A} = \text{tr } T'T = \sum_{1 \leq i \leq j \leq m} t_{i,j}^2, \quad \text{et} \quad \det \mathcal{A} = (\det T)^2 = \prod_{i=1}^m t_{i,i}^2,$$

on en déduit donc, à partir de (1.2.4) et (1.2.1), que

$$\begin{aligned} \Gamma_m(r) &= \int_{t_{i,i} > 0, 1 \leq i \leq m} \int_{t_{i,j} \in \mathbb{R}, 1 \leq i < j \leq m} \exp\left(-\sum_{1 \leq i < j \leq m} t_{i,j}^2\right) \left\{ \prod_{i=1}^m t_{i,i}^{2r-(m+1)} \right\} \\ &\quad \times \left\{ 2^m \prod_{i=1}^m t_{i,i}^{(m-i+1)} \right\} \bigwedge_{1 \leq i < j \leq m} dt_{i,j} \\ &= \left\{ \prod_{i=1}^m \int_0^\infty (t_{i,i}^2)^{r-\frac{1}{2}(i-1)-1} \exp(-t_{i,i}^2) d(t_{i,i}^2) \right\} \prod_{1 \leq i < j \leq m} \left( \int_{-\infty}^\infty \exp(-t_{i,j}^2) dt_{i,j} \right) \\ &= \prod_{i=1}^m \Gamma\left(r - \frac{1}{2}(i-1)\right) \times (\pi^{\frac{1}{2}})^{\frac{m(m-1)}{2}}, \end{aligned} \quad (1.2.5)$$

ce qu'il fallait démontrer.  $\square$

Comme, de toute évidence, lorsque  $m = 1$ ,  $\Gamma_1(r) = \Gamma(r)$  se réduit à la fonction Gamma habituelle, définie pour  $r > 0$ , on doit s'attendre, pour  $m \geq 1$  arbitraire, à une généralisation de la formule  $\Gamma(r+1) = r\Gamma(r)$ . Cette généralisation est fournie par le corollaire suivant de la proposition 1.2.1.

**Corollaire 1.2.1.** *Pour tout entier  $m \geq 1$ , et pour tout réel  $r$ , tel que  $r > \frac{1}{2}(m-1)$ , on a*

$$\Gamma_m(r+1) = \left\{ \prod_{i=1}^m \left(r - \frac{1}{2}(i-1)\right) \right\} \Gamma_m(r). \quad (1.2.6)$$

**Preuve.** Par (1.2.4), on a

$$\begin{aligned}\Gamma_m(r+1) &= \pi^{\frac{m(m-1)}{4}} \prod_{i=1}^m \Gamma\left(r+1 - \frac{1}{2}(i-1)\right) \\ &= \left\{ \prod_{i=1}^m \left(r - \frac{1}{2}(i-1)\right) \right\} \pi^{\frac{m(m-1)}{4}} \prod_{i=1}^m \Gamma\left(r - \frac{1}{2}(i-1)\right) \\ &= \left\{ \prod_{i=1}^m \left(r - \frac{1}{2}(i-1)\right) \right\} \Gamma_m(r),\end{aligned}$$

ce qu'il fallait démontrer.  $\square$

**Définition 1.2.2.** Soit  $\Lambda > 0$  une  $(m \times m)$  matrice symétrique, définie positive. Une matrice aléatoire  $(m \times m)$  symétrique, définie positive, est dite suivre une loi Gamma multivariée  $\Gamma_m(r, \Lambda)$ , où  $r > \frac{1}{2}(m-1)$ , ce qui sera noté  $A \stackrel{d}{=} \Gamma_m(r, \Lambda)$ , si la densité de  $A$ , sur l'ensemble des matrices  $\mathcal{A} > 0$  définies positives, est définie par

$$f(\mathcal{A}) = \frac{(\det(\Lambda))^r}{\Gamma_m(r)} (\det \mathcal{A})^{r-\frac{1}{2}(m+1)} \text{etr}(-\Lambda \mathcal{A}). \quad (1.2.7)$$

On vérifie que la fonction  $f(\mathcal{A})$  donnée en (1.2.7) définit bien une densité. Pour ceci, on effectue le changement de variable  $\mathcal{A} = \Lambda^{-1/2} \mathcal{B} \Lambda^{-1/2}$ , de sorte que  $\det(\mathcal{A}) = (\det(\Lambda))^{-1} \det(\mathcal{B})$ ,  $\text{etr}(-\Lambda \mathcal{A}) = \text{etr}(-\mathcal{B})$ , et, par (1.1.5),  $[d\mathcal{A}] = (\det(\Lambda))^{-\frac{1}{2}(m+1)} [d\mathcal{B}]$ . Compte tenu de (1.2.1), on en déduit que

$$\begin{aligned}\int_{\mathcal{A} > 0} f(\mathcal{A}) [d\mathcal{A}] &= \int_{\mathcal{A} > 0} \frac{(\det(\Lambda))^r}{\Gamma_m(r)} (\det \mathcal{A})^{r-\frac{1}{2}(m+1)} \text{etr}(-\Lambda \mathcal{A}) [d\mathcal{A}] \\ &= \frac{(\det(\Lambda))^r}{\Gamma_m(r)} \int_{\mathcal{B} > 0} (\det \Lambda)^{-r+\frac{1}{2}(m+1)} (\det \mathcal{B})^{r-\frac{1}{2}(m+1)} \text{etr}(-\mathcal{B}) (\det \Lambda)^{-\frac{1}{2}(m+1)} [d\mathcal{B}] \\ &= \frac{1}{\Gamma_m(r)} \int_{\mathcal{B} > 0} (\det \mathcal{B})^{r-\frac{1}{2}(m+1)} \text{etr}(-\mathcal{B}) [d\mathcal{B}] = 1.\end{aligned}$$

La définition de la loi de Wishart (centrée) qui suit sera complétée plus loin par une construction de cette loi à partir de vecteurs normaux (voir (2.1.2) et (2.1.24)). Celle-ci autorisera des matrices  $\Sigma \geq 0$  positives, sans être définies positives.

**Définition 1.2.3.** Soit  $\Sigma > 0$  une matrice  $(m \times m)$ , symétrique, définie positive. On dit qu'une matrice aléatoire,  $(m \times m)$  symétrique,  $A$  suit une loi de Wishart (centrée)  $W_m(n, \Sigma)$  pour  $n \geq m$ , si elle suit une loi  $\Gamma_m\left(\frac{n}{2}, \frac{1}{2}\Sigma^{-1}\right)$ . Ceci est noté

$$A \stackrel{d}{=} W_m(n, \Sigma) \stackrel{d}{=} \Gamma_m\left(\frac{1}{2}n, \frac{1}{2}\Sigma^{-1}\right). \quad (1.2.8)$$

Compte tenu de (1.2.7),  $A \stackrel{d}{=} W_m(n, \Sigma)$ , si et seulement si sa densité est donnée par

$$f(\mathcal{A}) = \frac{2^{-mn/2}}{\Gamma_m\left(\frac{1}{2}n\right)} (\det \Sigma)^{-n/2} (\det \mathcal{A})^{\frac{n}{2}-\frac{1}{2}(m+1)} \text{etr}\left(-\frac{1}{2}\Sigma^{-1}\mathcal{A}\right) \quad \text{pour } \mathcal{A} > 0. \quad (1.2.9)$$

Le corollaire suivant de la proposition 1.2.1, établit une nouvelle vérification du fait que  $f(\mathcal{A})$ , définie par (1.2.9), définit bien une densité. De plus, il étend ce résultat au cas où  $\frac{n}{2} = r$  est un nombre complexe.

**Corollaire 1.2.2.** Pour tout  $r \in \mathbb{C}$  tel que  $\text{Re}(r) > \frac{1}{2}(m-1)$ , et pour toute matrice  $(m \times m)$  complexe symétrique  $\Sigma$  telle que  $\text{re}(\Sigma) > 0$ , on a

$$\int_{\mathcal{A} > 0} (\det \mathcal{A})^{r-(m+1)/2} \text{etr}\left(-\frac{1}{2}\Sigma^{-1}\mathcal{A}\right) [d\mathcal{A}] = 2^{mr} (\det \Sigma)^r \Gamma_m(r). \quad (1.2.10)$$

**Preuve.** Supposons, tout d'abord, que  $\Sigma > 0$  soit une matrice réelle symétrique et définie positive. On effectue alors le changement de variable  $\mathcal{A} = (2\Sigma)^{1/2}\mathcal{V}(2\Sigma)^{1/2}$  qui donne  $\text{tr}(\frac{1}{2}\Sigma^{-1}\mathcal{A}) = \text{tr}(\mathcal{V})$  et, par (1.1.6),

$$[d\mathcal{A}] = \{\det(2\Sigma)\}^{\frac{1}{2}(m+1)}[d\mathcal{V}],$$

et

$$\det(\mathcal{A})^{r-\frac{1}{2}(m+1)} = \{\det(2\Sigma)\}^{r-\frac{1}{2}(m+1)}\det(\mathcal{V})^{r-\frac{1}{2}(m+1)}.$$

Comme  $\det(2\Sigma) = 2^m \det(\Sigma)$ , ce changement de variables transforme donc l'intégrale (1.2.10) en

$$2^{mr}(\det \Sigma)^r \int_{\mathcal{V}>0} (\det \mathcal{V})^{r-\frac{1}{2}(m+1)} \text{etr}(-\mathcal{V})[d\mathcal{V}],$$

expression qui, compte tenu de la définition (1.2.1) de  $\Gamma_m(r)$ , équivaut à (1.2.10). Le fait que le résultat (1.2.10) subsiste lorsque  $\Sigma$  est une matrice complexe est une conséquence de l'unicité du prolongement analytique (voir la remarque (1.2.1)).□

**Corollaire 1.2.3.** *Pour tout  $r \in \mathbb{C}$  tel que  $\text{Re}(r) > \frac{1}{2}(m-1)$ , et pour toute matrice  $(m \times m)$  complexe symétrique  $Z$  telle que  $\text{Re}(Z) > 0$ , la transformée de Laplace  $\tilde{f}$  de la fonction  $f(\mathcal{A}) = (\det \mathcal{A})^{r-(m+1)/2}$  est donnée par*

$$\tilde{f}(Z) = \int_{\mathcal{A}>0} (\det \mathcal{A})^{r-\frac{1}{2}(m+1)} \text{etr}(-Z\mathcal{A})[d\mathcal{A}] = (\det Z)^{-r} \Gamma_m(r). \quad (1.2.11)$$

**Preuve.** On obtient (1.2.11) en posant  $\Sigma^{-1} = 2Z$  dans (1.2.10). Compte tenu de la remarque (1.2.1)-(iii), le fait que la formule (1.2.11) soit valable pour toute matrice  $(m \times m)$  symétrique réelle  $Z > 0$  implique que celle-ci soit vérifiée lorsque  $Z$  est symétrique complexe telle que  $\text{Re}(Z) > 0$ .□

### 1.3 Règles de changement de variables - II.

Le changement de variables suivant nécessite quelques explications préalables. On utilise le fait (voir (2.4.5) que toute matrice  $(n \times m)$   $Z = [z_{i,j}]$ , de plein rang de lignes, égal à  $m = \text{rg}(Z)$ , avec  $m \leq n$ , peut se factoriser de manière unique sous la forme

$$Z = H_1 T, \quad (1.3.1)$$

où  $H_1$  est une portion de matrice orthogonale, c'est à dire, une matrice  $(n \times m)$  vérifiant  $H_1' H_1 = \mathbb{I}_m$ , et  $T = [t_{ij}]$  est une matrice  $(m \times m)$ , triangulaire supérieure dont les éléments diagonaux  $t_{ii} > 0$ ,  $i = 1, \dots, m$ , sont (strictement) positifs. Il est alors toujours possible de compléter la matrice  $H_1$ ,  $(n \times m)$ , par une matrice  $H_2$ ,  $(n \times (n-m))$ , (éventuellement vide si  $m = n$ ), de telle sorte que la matrice  $H = [H_1 \ H_2]$  soit  $(n \times n)$  et orthogonale. Décomposons alors cette matrice en

$$H = [h_1 \ \dots \ h_m \ h_{m+1} \ \dots \ h_n], \quad \text{où } h_j = \begin{bmatrix} h_{1,j} \\ \vdots \\ h_{n,j} \end{bmatrix} \in \mathbb{R}^n \quad \forall 1 \leq j \leq n.$$

L'orthogonalité de  $H$  est caractérisée par le fait que ses vecteurs colonnes,  $h_1, \dots, h_n$ , sont orthonormés, et vérifient donc

$$h_j' h_i = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i \neq j. \end{cases}$$

Observant que, pour tout  $j \in \{1, \dots, n\}$ ,

$$h_j' dh_i = \sum_{r=1}^n h_{r,j} dh_{r,i}$$

est une forme différentielle, on constate qu'on peut définir une forme différentielle extérieure, avec des éléments différentiels ne faisant intervenir que des éléments de  $H_1$ , et des coefficients faisant intervenir à la fois des éléments de  $H_1$  et de  $H_2$ , par

$$[H'_1 dH_1] = \bigwedge_{i=1}^m \bigwedge_{j=i+1}^n h'_j dh_i. \quad (1.3.2)$$

On notera bien que  $[H'_1 dH_1]$  est une *notation* pour désigner le membre de droite de (1.3.2). L'intérêt de cette notation sera justifié dans la proposition 1.1.2 plus loin.

**Remarque 1.3.1.** La définition (1.3.2) de  $[H'_1 dH_1]$  présente la particularité *non évidente* de définir une forme différentielle extérieure des coordonnées de  $H_1$  qui (au signe près) *ne dépend pas du choix de  $H_2$*  (sous la seule réserve que la matrice  $\begin{bmatrix} H_1 & H_2 \end{bmatrix}$  soit orthogonale). Cette caractéristique justifie qu'on écrive cette expression en utilisant la notation  $[H'_1 dH_1]$ , qui ne mentionne pas  $H_2$ . Soit, en effet, une matrice orthogonale  $Q$  arbitraire, de dimensions  $(n \times n)$ . Si on pose

$$\begin{aligned} K &= [K_1 \ K_2] = [k_1 \ \dots \ k_m \ k_{m+1} \ \dots \ k_n] \\ &= QH = [Qh_1 \ \dots \ Qh_m \ Qh_{m+1} \ \dots \ Qh_n], \end{aligned}$$

on constate que

$$[K'_1 dK_1] = \bigwedge_{i=1}^m \bigwedge_{j=i+1}^n k'_j dk_i = \bigwedge_{i=1}^m \bigwedge_{j=i+1}^n h'_j Q' Q dh_i = [H'_1 dH_1]. \quad (1.3.3)$$

Or, pour deux choix différents  $H_{2a}$  et  $H_{2b}$  de  $H_2$  tels que  $\begin{bmatrix} H_1 & H_{2a} \end{bmatrix}$  et  $\begin{bmatrix} H_1 & H_{2b} \end{bmatrix}$  soient orthogonales avec  $H_1$  fixée, il existe toujours une matrice orthogonale  $Q$  telle que  $K := Q \begin{bmatrix} H_1 & H_{2a} \end{bmatrix} = \begin{bmatrix} H_1 & H_{2b} \end{bmatrix}$ , ce qui, compte tenu de (1.3.3), établit bien la propriété annoncée.

**Proposition 1.3.1.** Soit  $Z = H_1 T$  une matrice  $(n \times m)$ , de rang  $m \leq n$ , où  $H_1$  et  $T$  sont comme en (1.3.1), avec  $H'_1 H_1 = \mathbb{I}_m$ , et où  $T$  est une matrice triangulaire supérieure à diagonale positive. Alors

$$(dZ) = \prod_{i=1}^m t_{ii}^{n-i} [dT][H'_1 dH_1]. \quad (1.3.4)$$

**Preuve.** Complétons  $H_1$  par une matrice  $H_2$  de sorte que  $\begin{bmatrix} H_1 & H_2 \end{bmatrix}$  soit  $(n \times n)$  orthogonale. En faisant usage de (1.1.4), et du fait que  $H$  est orthogonale (et donc de déterminant égal à  $\pm 1$ ), on voit que (en faisant usage de la convention que  $(dZ)$  est défini au signe près), comme  $Z = H_1 T$ ,

$$(H' \cdot dZ) = (\det H')^m (dZ) = (dZ). \quad (1.3.5)$$

Par conséquent, (1.3.5) ramène le calcul de  $(dZ)$  à celui de

$$(dZ) = (H' \cdot dZ). \quad (1.3.6)$$

En vue du calcul du membre de droite de (1.3.6), développons la matrice (de formes différentielles)  $H' \cdot dZ$ . Comme  $Z = H_1 T$ , on a l'identité de différentielles

$$dZ = dH_1 \cdot T + H_1 \cdot dT.$$

Cette relation, prise en combinaison avec les identités  $H'_1 H_1 = \mathbb{I}_m$ , et  $H'_2 H_1 = \mathbb{O}_{n-m, m}$ , implique que

$$\begin{aligned} H' \cdot dZ &= \begin{bmatrix} H'_1 \\ H'_2 \end{bmatrix} dZ = \begin{bmatrix} H'_1 \\ H'_2 \end{bmatrix} (dH_1 \cdot T + H_1 \cdot dT) \\ &= \begin{bmatrix} H'_1 \cdot dH_1 \cdot T + H'_1 \cdot H_1 \cdot dT \\ H'_2 \cdot dH_1 \cdot T + H'_2 \cdot H_1 \cdot dT \end{bmatrix} = \begin{bmatrix} H'_1 \cdot dH_1 \cdot T + dT \\ H'_2 \cdot dH_1 \cdot T \end{bmatrix}. \end{aligned} \quad (1.3.7)$$

Par conséquent, en combinant (1.3.5) et (1.3.6), on ramène le calcul de  $(dZ)$  à celui de

$$(dZ) = (H' \cdot dZ) = \left( \begin{bmatrix} H'_1 \cdot dH_1 \cdot T + dT \\ H'_2 \cdot dH_1 T \end{bmatrix} \right). \quad (1.3.8)$$

Le membre de droite de (1.3.8) est obtenu en faisant le produit extérieur de l'ensemble des composantes de la matrice comprise entre les parenthèses. Pour effectuer ce calcul, nous allons expliciter cette matrice, soit

$$H' \cdot dZ = \begin{bmatrix} H'_1 \cdot dH_1 \cdot T + dT \\ H'_2 \cdot dH_1 T \end{bmatrix},$$

ce qui nous permettra, ensuite, d'effectuer le produit extérieur de ses composantes. Rappelons que

$$H_1 = [h_1 \quad \dots \quad h_m], \quad H_2 = [h_{m+1} \quad \dots \quad h_n] \quad \text{et} \quad dH_1 = [dh_1 \quad \dots \quad dh_m].$$

Le produit extérieur des formes différentielles qui composent la matrice  $H' \cdot dZ$  est composé de deux parties. En premier lieu, nous évaluons la contribution qui lui est apportée par les  $n - m$  lignes inférieures de  $H' \cdot dZ$ . Compte tenu de (1.3.7) et (1.3.8), ces  $n - m$  lignes composent la matrice

$$H'_2 \cdot dH_1 \cdot T = \begin{bmatrix} h'_{m+1} dh_1 & \dots & h'_{m+1} dh_m \\ \vdots & & \vdots \\ h'_n dh_1 & \dots & h'_n dh_m \end{bmatrix} T = \begin{bmatrix} h'_{m+1} \cdot dH_1 \cdot T \\ \vdots \\ h'_n \cdot dH_1 \cdot T \end{bmatrix}.$$

Le produit extérieur de l'ensemble des formes différentielles composant les éléments de la matrice  $H'_2 \cdot dH_1 \cdot T$  est égal à

$$(H'_2 \cdot dH_1 \cdot T) = \bigwedge_{j=m+1}^n (h'_j \cdot dH_1 \cdot T) = \bigwedge_{j=m+1}^n \left( (\det T) (h'_j \cdot dH_1) \right) \quad (1.3.9)$$

$$\begin{aligned} &= \bigwedge_{j=m+1}^n \left( (\det T) \bigwedge_{i=1}^m h'_j dh_i \right) = (\det T)^{n-m} \bigwedge_{j=m+1}^n \left( \bigwedge_{i=1}^m h'_j dh_i \right) \\ &= \left( \prod_{i=1}^m t_{i,i}^{n-m} \right) \bigwedge_{j=m+1}^n \left( \bigwedge_{i=1}^m h'_j dh_i \right), \end{aligned} \quad (1.3.10)$$

où nous avons fait usage de (1.1.4).

Pour évaluer la contribution des  $m$  lignes supérieures restantes de  $H' \cdot dZ$ , données dans l'expression (1.3.7), nous commençons par différencier le produit  $H'_1 H_1 = \mathbb{I}_m$ . Ceci nous fournit les identités

$$H'_1 \cdot dH_1 + dH'_1 \cdot H_1 = \mathbb{O} \quad \iff \quad H'_1 \cdot dH_1 = -\{H'_1 \cdot dH_1\}'.$$

La dernière de ces identités exprime que la matrice  $M = H'_1 \cdot dH_1$  est *antisymétrique* (la transposée de  $M$  est égale à  $-M$ ), ce qui permet d'écrire le produit  $M \cdot T + dT$  sous la forme

$$H'_1 \cdot dH_1 \cdot T + dT = \begin{bmatrix} 0 & -h'_2 dh_1 & -h'_3 dh_1 & \dots & -h'_m dh_1 \\ h'_2 dh_1 & 0 & -h'_3 dh_2 & \dots & -h'_m dh_2 \\ h'_3 dh_1 & h'_3 dh_2 & 0 & \dots & -h'_m dh_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h'_m dh_1 & h'_m dh_2 & h'_m dh_3 & \dots & 0 \end{bmatrix} \begin{bmatrix} t_{1,1} & t_{1,2} & \dots & t_{1,m} \\ 0 & t_{2,1} & \dots & t_{2,m} \\ 0 & 0 & \dots & t_{3,m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & t_{m,m} \end{bmatrix} + dT.$$

En effectuant le produit matriciel ci-dessus, sans tenir compte des termes en  $h'_i dh_j$  apparus dans les colonnes antérieures (puisque ceux-ci mèneront, par la suite, à des produits extérieurs nuls), et en se limitant aux termes

sous-diagonaux (les termes non pris en compte sont notés, ci-dessous, par le symbole " \* "), on obtient la matrice

$$H'_1 \cdot dH_1 \cdot T + dT = \begin{bmatrix} 0 + dt_{1,1} & * + dt_{1,2} & * + dt_{1,3} & \dots & * + dt_{1,m} \\ h'_2 dh_1 t_{1,1} & * + dt_{2,2} & * + dt_{2,3} & \dots & * + dt_{2,m} \\ h'_3 dh_1 t_{1,1} & h'_3 dh_2 t_{2,2} + * & * + dt_{3,3} & \dots & * + dt_{3,m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h'_m dh_1 t_{1,1} & h'_m dh_2 t_{2,2} + * & h'_m dh_3 t_{3,3} + * & \dots & * + dt_{m,m} \end{bmatrix}.$$

Le produit extérieur des termes sous-diagonaux de  $H'_1 \cdot dH_1 \cdot T + dT$  est donc égal à

$$(H'_1 \cdot dH_1 \cdot T) = \left( \prod_{i=1}^m t_{i,i}^{m-i} \right) \bigwedge_{i=1}^m \left( \bigwedge_{j=i+1}^m h'_j dh_i \right). \quad (1.3.11)$$

De même, le produit extérieur des termes sur la diagonale, et au dessus de la diagonale, de  $H'_1 \cdot dH_1 \cdot T + dT$  est égal à

$$(dT) + \left\{ \text{termes faisant intervenir les composantes de } (dH_1) \right\}. \quad (1.3.12)$$

Compte tenu de (1.3.6), le produit extérieur de l'ensemble des expressions trouvées en (1.3.10), (1.3.11) et (1.3.12), mène donc à la formule

$$(dZ) = \left\{ \left( \prod_{i=1}^m t_{i,i}^{n-m} \right) \bigwedge_{j=m+1}^n \left( \bigwedge_{i=1}^m h'_j dh_i \right) \right\} \wedge \left\{ \left( \prod_{i=1}^m t_{i,i}^{m-i} \right) \bigwedge_{i=1}^m \left( \bigwedge_{j=i+1}^m h'_j dh_i \right) \right\} \wedge \left\{ (dT) \right\},$$

qui coïncide ainsi avec (1.3.4).  $\square$

Le résultat énoncé dans la proposition 1.3.2 suivante joue un rôle essentiel dans le calcul de la densité de la loi de Wishart, lorsque cette distribution est construite à partir de la loi normale multivariée (voir (1.2.10), (2.1.2) et (2.1.24)). Il peut être établi comme conséquence simple des propositions 1.1.4 et 1.3.1.

**Proposition 1.3.2.** *Soit  $Z = H_1 T$  une matrice  $(n \times m)$ , de rang  $m \leq n$ , où  $H_1$  et  $T$  sont comme dans (1.3.1), et soit  $A = Z'Z = T'T > 0$ . Alors*

$$(dZ) = 2^{-m} (\det A)^{(n-m-1)/2} [dA][H'_1 dH_1]. \quad (1.3.13)$$

**Preuve.** Tout d'abord, par (1.3.4),

$$(dZ) = \prod_{i=1}^m t_{i,i}^{n-i} [dT][H'_1 dH_1].$$

D'autre part,  $A = Z'Z = T'T$ , et par (1.1.7),

$$[dA] = 2^m \prod_{i=1}^m t_{i,i}^{m-i+1} [dT] \iff [dT] = 2^{-m} \prod_{i=1}^m t_{i,i}^{-m+i-1} [dA]. \quad (1.3.14)$$

En remarquant que

$$\prod_{i=1}^m t_{i,i} = \det T = (\det A)^{1/2},$$

on en déduit que

$$\begin{aligned} (dZ) &= \left\{ \prod_{i=1}^m t_{ii}^{n-i} \right\} \left\{ 2^{-m} \prod_{i=1}^m t_{i,i}^{-m+i-1} \right\} [dA] \\ &= 2^{-m} \prod_{i=1}^m t_{ii}^{n-m-1} [dA][H'_1 dH_1] = 2^{-m} (\det A)^{(n-m-1)/2} [H'_1 dH_1], \end{aligned}$$

ce qui établit (1.3.13).  $\square$



## 1.4 Variétés de Stiefel, groupe orthogonal et mesure invariante.

Une partie  $\mathcal{E}$  de l'espace  $\mathbb{R}^N$  est appelée *variété* (sous-entendu *topologique*) de dimension  $p \leq N$ , si tout point  $x \in \mathcal{E}$  possède un voisinage  $V_x$  dans  $\mathbb{R}^N$ , tel que  $V_x \cap \mathcal{E}$  peut être mis en bijection par une application continue (une *carte locale*), ainsi que son inverse, avec un ouvert de  $\mathbb{R}^p$ . La variété est dite *différentiable* si les cartes locales sont différentiables ainsi que leurs inverses. Dans la suite de notre exposé, toutes les variétés considérées seront différentiables, et ce qualificatif sera sous-entendu. Les cartes locales seront, en général, fournies par des *représentations paramétriques* simples, définissant la structure de ces ensembles.

**Définition 1.4.1.** *Pour chaque choix possible des entiers  $m$  et  $n$  vérifiant  $1 \leq m \leq n$ , l'ensemble des matrices,  $(n \times m)$ ,  $H_1 = [h_1 \ \dots \ h_m]$ , avec  $h_j \in \mathbb{R}^n$  pour  $1 \leq j \leq m$ , telles que  $H_1' H_1 = \mathbb{I}_m$  est appelée variété de Stiefel de paramètres  $m$  et  $n$ , et noté  $\mathcal{V}_{m,n}$ .*

Les  $\frac{1}{2}m(m+1)$  relations  $h_i' h_i = 1$  pour  $1 \leq i \leq m$ , et  $h_i' h_j = 0$  pour  $1 \leq i < j \leq m$ , liant les  $mn$  coordonnées de  $h_1, \dots, h_m$ , permettent de considérer  $\mathcal{V}_{m,n}$  comme une *variété* de dimension  $mn - \frac{1}{2}m(m+1)$  de l'espace  $\mathcal{M}_{n,m}$  des matrices  $(n \times m)$ . Ce dernier espace, de dimension  $mn$ , est isomorphe à l'espace  $\mathbb{R}^{mn}$ . Les deux cas suivants de variétés de Stiefel sont particulièrement dignes d'intérêt.

– Lorsque  $m = n$ , la variété de Stiefel  $\mathcal{V}_{n,n}$ , de dimension  $\frac{1}{2}n(n-1)$ , coïncide avec le *groupe orthogonal*, noté  $\mathcal{O}_n = \mathcal{V}_{n,n}$ , composé de l'ensemble des matrices orthogonales  $(n \times n)$ , muni du produit de matrices usuel.

– Lorsque  $m = 1$ , la variété de Stiefel  $\mathcal{V}_{1,n}$ , de dimension  $n-1$ , s'identifie à la *surface de la sphère unité* dans  $\mathbb{R}^n$ , et est, ainsi, composée des vecteurs  $h_1 \in \mathbb{R}^n$  tels que  $h_1' h_1 = 1$ .

**Proposition 1.4.1.** *Pour tout  $1 \leq m \leq n$ , on a*

$$\int_{\mathcal{V}_{m,n}} [H_1' dH_1] = \frac{2^m \pi^{mn/2}}{\Gamma_m(\frac{1}{2}n)}. \quad (1.4.1)$$

**Preuve.** Soit  $\mathcal{Z} = [Z_{i,j}] \in \mathcal{M}_{n,m}$  une matrice  $(n \times m)$  dont les coordonnées sont des variables aléatoires indépendantes de même loi normale  $N(0, 1)$ . Leur densité jointe (relativement à la mesure de Lebesgue dans  $\mathbb{R}^{mn}$ ) est donc

$$(2\pi)^{-mn/2} \exp\left(-\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^m z_{i,j}^2\right) = (2\pi)^{-mn/2} \text{etr}\left(-\frac{1}{2} \mathcal{Z}' \mathcal{Z}\right),$$

où on a posé, pour simplifier les notations,  $\mathcal{Z} = [z_{i,j}] \in \mathcal{M}_{n,m}$ . Comme l'intégrale de cette densité sur  $\mathbb{R}^{mn}$  vaut 1, on a l'identité

$$\int_{\mathcal{Z} \in \mathcal{M}_{n,m}} \text{etr}\left(-\frac{1}{2} \mathcal{Z}' \mathcal{Z}\right) (d\mathcal{Z}) = (2\pi)^{mn/2}. \quad (1.4.2)$$

En posant, comme en (1.3.1),  $\mathcal{Z} = H_1 T$ , de sorte que  $\mathcal{Z}' \mathcal{Z} = T' T$ , et en remplaçant, grâce à (1.3.4),  $(d\mathcal{Z})$  par l'élément différentiel

$$(d\mathcal{Z}) = \prod_{i=1}^m t_{i,i}^{n-i} [dT] [H_1' dH_1],$$

on peut réécrire (1.4.2) en

$$\begin{aligned} & \left\{ \int_{t_{i,i} > 0: 1 \leq i \leq m} \int_{t_{i,j} \in \mathbb{R}: 1 \leq i < j \leq m} \exp\left(-\frac{1}{2} \sum_{1 \leq i < j \leq m} t_{i,j}^2\right) \prod_{i=1}^m t_{i,i}^{n-i} \bigwedge_{1 \leq i < j \leq m} dt_{i,j} \right\} \int_{\mathcal{V}_{m,n}} [H_1' dH_1] \\ &= \prod_{1 \leq i < j \leq m} \left( \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} t_{i,j}^2\right) dt_{i,j} \right) \times \prod_{i=1}^m \left( \int_0^{\infty} \exp\left(-\frac{1}{2} t_{i,i}^2\right) t_{i,i}^{n-i} dt_{i,i} \right) \times \int_{\mathcal{V}_{m,n}} [H_1' dH_1] \\ &= \prod_{1 \leq i < j \leq m} \left( (2\pi)^{1/2} \right) \times \prod_{i=1}^m \left( 2^{(n-i-1)/2} \Gamma\left(\frac{1}{2}(n-i+1)\right) \right) \times \int_{\mathcal{V}_{m,n}} [H_1' dH_1] \end{aligned}$$

$$\begin{aligned}
&= \left\{ \pi^{\frac{m(m-1)}{4}} \prod_{i=1}^m \Gamma\left(\frac{1}{2}(n-i+1)\right) \right\} \times 2^{\frac{1}{2}mn-m} \times \int_{\mathcal{V}_{m,n}} [H'_1 dH_1] \\
&= \Gamma_m\left(\frac{1}{2}n\right) 2^{\frac{1}{2}mn-m} \times \int_{\mathcal{V}_{m,n}} [H'_1 dH_1] = (2\pi)^{\frac{1}{2}mn},
\end{aligned} \tag{1.4.3}$$

où nous avons utilisé (1.2.4). Finalement, on déduit directement (1.4.1) de la relation (1.4.3).□

**Corollaire 1.4.1.** *On a, pour tout entier  $n \geq 1$ ,*

$$\int_{\mathcal{O}_n} [H'_1 dH_1] = \int_{\mathcal{V}_{n,n}} [H'_1 dH_1] = \int_{\mathcal{O}_n} [H' dH] = \frac{2^n \pi^{n^2/2}}{\Gamma_n\left(\frac{1}{2}n\right)}. \tag{1.4.4}$$

**Preuve.** On pose  $m = n$  dans (1.4.1). On remarquera ici que la matrice  $H_2$  est inexistante pour  $m = n$ , ce qui explique qu'on peut écrire, plus simplement  $H = H_1$ . Nous conservons néanmoins l'usage de l'une ou l'autre des notations  $H$  ou  $H_1$  pour  $m = n$ , par souci d'homogénéité de l'exposé.□

**Corollaire 1.4.2.** *1°) Pour tout  $R \geq 0$  et pour tout entier  $n \geq 1$ , la surface de la sphère de rayon  $R$ , dans  $\mathbb{R}^n$ , définie par  $R \times S_n$ , avec  $S_n = \mathcal{V}_{1,n} = \{h \in \mathbb{R}^n : h'h = 1\}$ , est donnée par*

$$\text{Surface}(R \times S_n) = R^{n-1} \int_{\mathcal{V}_{1,n}} [H'_1 dH_1] = R^{n-1} \frac{2\pi^{n/2}}{\Gamma\left(\frac{1}{2}n\right)} = \begin{cases} \frac{2\pi^k}{(k-1)!} R^{2k-1} & \text{si } n = 2k, \\ \frac{2^{2k+1} \pi^k k!}{(2k)!} R^{2k} & \text{si } n = 2k + 1. \end{cases} \tag{1.4.5}$$

*2°) Le volume de la boule  $R \times B_n$ , de rayon  $R$ , dans  $\mathbb{R}^n$ , où  $B_n = \{h \in \mathbb{R}^n : h'h \leq 1\}$ , est donné par*

$$\text{Volume}(R \times B_n) = R^n \frac{\pi^{n/2}}{\Gamma\left(\frac{1}{2}n + 1\right)} = \begin{cases} \frac{\pi^k}{k!} R^{2k} & \text{si } n = 2k, \\ \frac{2^{2k+1} \pi^k k!}{(2k+1)!} R^{2k} & \text{si } n = 2k + 1. \end{cases} \tag{1.4.6}$$

**Preuve.** On obtient (1.4.5) en posant  $m = 1$  dans (1.4.1). Les relations (1.4.5) et (1.4.6) se déduisent l'une de l'autre, respectivement, par dérivation et par intégration relativement à  $R$ .□

**Remarque 1.4.1.** Soit  $\mathcal{O}_n = \mathcal{V}_{n,n} = \{H \in \mathcal{M}_{n,n} : HH' = \mathbb{I}_n\}$ , le *groupe orthogonal*, composé des matrices orthogonales,  $(n \times n)$ , de  $\mathbb{R}^n$ , avec, comme loi de composition interne, le produit matriciel. Alors, la mesure  $\mu_n$  sur  $\mathcal{O}_n$ , définie par

$$\mu_n(\mathcal{B}) = \int_{\mathcal{B}} [H'_1 dH_1] = \int_{\mathcal{B}} [H' dH], \quad \text{pour } \mathcal{B} \text{ borélien, } \mathcal{B} \subseteq \mathcal{O}_n, \tag{1.4.7}$$

est *invariante* (relativement à la loi de composition interne du groupe), au sens que  $\mu_n(H\mathcal{B}) = \mu_n(\mathcal{B})$  pour tout  $H \in \mathcal{O}_n$  et  $\mathcal{B}$ , partie borélienne de  $\mathcal{O}_n$ . Cette dernière propriété caractérise  $\mu_n$  à une constante multiplicative près. En effet, il s'agit alors d'une *mesure de Haar* sur le groupe localement compact  $\mathcal{O}_n$ . Une telle mesure est toujours unique, à une constante multiplicative près).

On normalise cette mesure en posant (notation spécifique aux matrices orthogonales)

$$[dH] = \frac{\Gamma_n\left(\frac{1}{2}n\right)}{2^n \pi^{n^2/2}} [H'_1 dH_1] = \frac{\Gamma_n\left(\frac{1}{2}n\right)}{2^n \pi^{n^2/2}} [H' dH] = \frac{\Gamma_n\left(\frac{1}{2}n\right)}{2^n \pi^{n^2/2}} \bigwedge_{1 \leq i \leq j \leq n} h'_j dh_i, \tag{1.4.8}$$

où  $H = [h_1 \ \dots \ h_n]$ . On définit alors une mesure de probabilité  $\mu_n^*$  sur  $\mathcal{O}_n$ , dite *loi uniforme sur  $\mathcal{O}_n$* , en posant

$$\mu_n^*(\mathcal{B}) = \frac{\mu_n(\mathcal{B})}{\mu_n(\mathcal{O}_n)} = \frac{\Gamma_n\left(\frac{1}{2}n\right)}{2^n \pi^{n^2/2}} \mu_n(\mathcal{B}) = \int_{\mathcal{B}} [dH], \quad \forall \mathcal{B} \text{ borélien, } \mathcal{B} \subseteq \mathcal{O}_n. \tag{1.4.9}$$

La mesure  $\mu_n^* = [dH]$  vérifie, en particulier,

$$\mu_n^*(\mathcal{O}_n) = \int_{\mathcal{O}_n} [dH] = 1. \tag{1.4.10}$$

## 1.5 Règles de changement de variables-III.

La proposition suivante, due à Constantine et Muirhead (voir Constantine, A.G. et Muirhead, R.J. (1976). Asymptotic expansions for distributions of latent roots in multivariate analysis. *Journal of Multivariate Analysis*. **6** 369–391), est utile dans le calcul d'intégrales liées à la loi de Wishart non centrée. Dans son énoncé,  $f(H)$  désigne une fonction d'une matrice orthogonale  $H \in \mathcal{O}_n$ , et intégrable relativement à la mesure de Haar  $[dH]$  sur  $\mathcal{O}_n$ , telle qu'elle est définie en (1.4.10). On fera usage dans ce qui suit de la mesure  $[H'dH]$ , proportionnelle à  $[dH]$ , et définie comme en (1.4.7). Faisant usage de la décomposition

$$H = \begin{bmatrix} H_1 & H_2 \end{bmatrix} \quad \text{pour } H \in \mathcal{O}_n,$$

où  $H_1 \in \mathcal{V}_{m,n}$  est  $(n \times m)$  et  $H_2 \in \mathcal{V}_{n-m,n}$  est  $(n \times (n-m))$ , on pose  $f(H) = f(H_1, H_2)$ .

**Proposition 1.5.1.** *Pour tout  $H_1 \in \mathcal{V}_{m,n}$ , nous fixons une matrice  $G = G(H_1) \in \mathcal{V}_{n-m,n}$ , fonction de  $H_1 \in \mathcal{V}_{m,n}$ , et vérifiant les propriétés suivantes. Nous posons*

$$H_1 = [h_1 \ \dots \ h_m] \quad \text{et} \quad G = [g_1 \ \dots \ g_{n-m}],$$

et supposons que, indépendamment de  $H_1$ ,  $[H_1 \ G] = [h_1 \ \dots \ h_m \ g_1 \ \dots \ g_{n-m}]$  définit une matrice orthogonale. Par ailleurs, lorsque  $H_1$  varie dans  $\mathcal{V}_{m,n}$ , nous notons  $H_2 \in \mathcal{V}_{n-m,n}$  une matrice quelconque de  $\mathcal{V}_{n-m,n}$ , telle que  $H = [H_1 \ H_2] \in \mathcal{O}_n$  prenne ses valeurs dans l'ensemble des matrices  $(n \times n)$  orthogonales. Alors, on a l'identité

$$\begin{aligned} & \int_{H = \begin{bmatrix} H_1 & H_2 \end{bmatrix} \in \mathcal{O}_n} f(H_1, H_2) [H'dH] \\ &= \int_{H_1 \in \mathcal{V}_{m,n}} \left\{ \int_{K \in \mathcal{O}_{n-m}} f(H_1, GK) [K'dK] \right\} [H'_1 dH_1]. \end{aligned} \quad (1.5.1)$$

**Preuve.** A toute matrice  $H = [H_1 \ H_2]$ , associons une matrice  $K$ , de dimensions  $((n-m) \times (n-m))$ , définie par la relation  $H_2 = GK$ . Comme, à la fois  $G$  et  $H_2$  appartiennent à  $\mathcal{V}_{n-m,n}$ , cette identité impose que la matrice  $K = [\kappa_1 \ \dots \ \kappa_{n-m}] \in \mathcal{O}_{n-m}$  est orthogonale. De plus, il est clair que, lorsque  $H_1$  est fixé et  $H = [H_1 \ H_2]$  parcourt  $\mathcal{O}_n$ , la matrice  $K$  parcourt  $\mathcal{O}_{n-m}$ . Posons  $H_2 = [h_{m+1} \ \dots \ h_n]$ . L'égalité  $H_2 = GK$  implique que  $h_{m+j} = G\kappa_j$  pour  $j = 1, \dots, n-m$ . Par conséquent, on a  $dh_{m+j} = Gd\kappa_j$  pour  $j = 1, \dots, n-m$ . Par (1.3.2), ceci implique que

$$\begin{aligned} [H'dH] &= \bigwedge_{1 \leq i < j \leq n} h'_j dh_i \\ &= \left\{ \bigwedge_{1 \leq i < j \leq m} h'_j dh_i \right\} \wedge \left\{ \bigwedge_{j=1}^{n-m} \bigwedge_{i=1}^m h'_{m+j} dh_i \right\} \wedge \left\{ \bigwedge_{1 \leq i < j \leq n-m} h'_{m+j} dh_{m+i} \right\} \\ &= \left\{ \bigwedge_{1 \leq i < j \leq m} h'_j dh_i \right\} \wedge \left\{ \bigwedge_{j=1}^{n-m} \bigwedge_{i=1}^m \kappa'_j G' dh_i \right\} \wedge \left\{ \bigwedge_{1 \leq i < j \leq n-m} d\kappa'_j G' G d\kappa_i \right\} = [H'_1 dH_1] [K'dK], \end{aligned}$$

ce qui permet de conclure.  $\square$

## Chapitre 2

# Lois de Wishart et Applications

### 2.1 Loi de Wishart centrée.

#### 2.1.1 Définition de la loi de Wishart centrée.

Les définitions (1.2.2) et (1.2.3) ont défini la loi de Wishart  $W_m(n, \Sigma)$ , lorsque  $\Sigma > 0$  est définie positive, comme cas particulier de loi Gamma multivariée, à l'aide de la formule (1.2.8), qui s'écrit

$$A \stackrel{d}{=} W_m(n, \Sigma) \stackrel{d}{=} \Gamma_m\left(\frac{1}{2}n, \frac{1}{2}\Sigma^{-1}\right). \quad (2.1.1)$$

Dans ce qui suit, nous allons donner une construction alternative de la loi de Wishart, qui se trouve être un peu plus générale que la précédente, car elle autorise une matrice  $\Sigma \geq 0$ , positive, sans être définie positive. Nous établirons, ensuite, que la nouvelle relation (2.1.2), utilisée dans la définition 2.1.1, est équivalente à la relation (2.1.1) de la définition 1.2.3 antérieure, lorsque  $\Sigma > 0$ .

**Définition 2.1.1.** Soient  $Z_1, \dots, Z_n$ ,  $n$  vecteurs aléatoires indépendants de  $\mathbb{R}^m$ , de même loi normale centrée  $N_m(\mathbb{O}, \Sigma)$ , où  $\Sigma \geq 0$  désigne une matrice  $(m \times m)$  symétrique positive. La loi de Wishart (ou loi de Wishart centrée) de dimension  $m$ , à  $n$  degrés de liberté, est la loi de probabilité de la matrice aléatoire  $(m \times m)$  symétrique positive définie par

$$A = \sum_{i=1}^n Z_i Z_i'. \quad (2.1.2)$$

On note cette propriété par  $A \stackrel{d}{=} W_m(n, \Sigma)$ .

Compte tenu de la convention  $\sum_{i=1}^0 Z_i Z_i' = \mathbb{O}_{m,m}$ , on prolonge la définition (2.1.2) au cas où  $n = 0$ , en définissant la loi de Wishart  $W_m(0, \Sigma)$  comme la loi de la matrice carrée  $(m \times m)$  nulle  $\mathbb{O}_{m,m}$ .

Posons

$$\mathbb{Z} = [Z_1 \quad \dots \quad Z_n]' = \begin{bmatrix} Z_1' \\ \vdots \\ Z_n' \end{bmatrix} \in \mathcal{M}_{n,m} \iff \mathbb{Z}' = [Z_1 \quad \dots \quad Z_n] \in \mathcal{M}_{m,n}.$$

Conformément aux notations habituelles des lois normales matricielles (voir (3.4.7)), la matrice aléatoire  $(n \times m)$   $\mathbb{Z}$  suit une loi normale matricielle de paramètres

$$\mathbb{Z} \stackrel{d}{=} N_{n,m}(\mathbb{O}_{n,m}, \mathbb{I}_n \otimes \Sigma).$$

Rappelons que, avec ces notations, si  $\mathbb{Y} \in \mathcal{M}_{n,m}$  suit une loi normale matricielle, on a l'équivalence

$$\mathbb{Y} \stackrel{d}{=} N_{n,m}(M, \Gamma) \iff M = \mathbb{E}(\mathbb{Y}), \quad \Gamma = \text{var}(\text{vec}(\mathbb{Y}')).$$

On constate alors que la relation (2.1.2) s'écrit de manière équivalente sous la forme

$$A = \mathbb{Z}'\mathbb{Z} \stackrel{d}{=} W_m(n, \Sigma). \quad (2.1.3)$$

### 2.1.2 Loi de Wishart et échantillonnage de la loi normale.

La loi de Wishart apparaît naturellement dans l'échantillonnage de la loi normale multivariée. Il s'agit de l'analyse statistique d'un *échantillon aléatoire*, composé d'une suite finie  $X_1, \dots, X_n$  d'observations indépendantes de même loi normale  $N_m(\mu, \Sigma)$ . Ayant observé  $X_1, \dots, X_n$ , on cherche à en déduire des renseignements sur les paramètres, a priori inconnus,  $\mu \in \mathbb{R}^p$  et  $\Sigma$ . En particulier, le théorème suivant joue un rôle fondamental dans l'estimation de  $\mu$  et  $\Sigma$ .

**Théorème 2.1.1.** *Soit  $X_1, \dots, X_n$ , un échantillon de  $n \geq 1$  vecteurs aléatoires indépendants de même loi normale  $N_m(\mu, \Sigma)$ , où  $\Sigma \geq 0$  désigne une matrice ( $m \times m$ ) symétrique positive. Alors les statistiques*

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad A = \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})' = \sum_{i=1}^n X_i X_i' - n\bar{X}\bar{X}', \quad (2.1.4)$$

sont indépendantes et telles que

$$X \stackrel{d}{=} N_m(\mu, n^{-1}\Sigma) \quad \text{et} \quad A \stackrel{d}{=} W_m(n-1, \Sigma). \quad (2.1.5)$$

**Preuve.** Posons

$$\mathbb{X} = \begin{bmatrix} X_1' \\ \vdots \\ X_n' \end{bmatrix} \quad \text{de sorte que} \quad \mathbb{E}(\mathbb{X}) = \begin{bmatrix} \mu' \\ \vdots \\ \mu' \end{bmatrix} = \mathbb{1}_n \otimes \mu' = \mathbb{1}_n \mu'. \quad (2.1.6)$$

Notons que

$$A = \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})' = \sum_{i=1}^n X_i X_i' - n\bar{X}\bar{X}' = \mathbb{X}'\mathbb{X} - n\bar{X}\bar{X}'.$$

Construisons maintenant une matrice ( $n \times n$ ) orthogonale  $H$  de la forme

$$H = \begin{bmatrix} n^{-1/2}\mathbb{1}'_n \\ h_2' \\ \vdots \\ h_n' \end{bmatrix} = \begin{bmatrix} h_1' \\ h_2' \\ \vdots \\ h_n' \end{bmatrix},$$

où  $h_1 = n^{-1/2}\mathbb{1}_n$  et  $h_2, \dots, h_n$  sont des vecteurs orthonormés de  $\mathbb{R}^n$ , vérifiant  $h_j' h_\ell = 1$  ou  $0$ , suivant que  $j = \ell$  ou  $j \neq \ell$ . Une telle matrice existe nécessairement, puisque  $h_1 = n^{-1/2}\mathbb{1}_n$  est de norme euclidienne égale à 1. Pour construire  $h_2, \dots, h_n$ , il suffit donc de choisir une base orthonormée quelconque du sous-espace de  $\mathbb{R}^n$  orthogonal (pour le produit scalaire habituel) au vecteur  $\mathbb{1}_n \in \mathbb{R}^n$ .

On constate alors que

$$\mathbb{Y} = \begin{bmatrix} Y_1' \\ Y_2' \\ \vdots \\ Y_n' \end{bmatrix} = H\mathbb{X} = \begin{bmatrix} h_1' \mathbb{X} \\ h_2' \mathbb{X} \\ \vdots \\ h_n' \mathbb{X} \end{bmatrix} = \begin{bmatrix} n^{1/2}\bar{X} \\ h_2' \mathbb{X} \\ \vdots \\ h_n' \mathbb{X} \end{bmatrix} \quad \text{et} \quad \mathbb{E}\mathbb{Y} = \begin{bmatrix} h_1' \mathbb{E}(\mathbb{X}) \\ h_2' \mathbb{E}(\mathbb{X}) \\ \vdots \\ h_n' \mathbb{E}(\mathbb{X}) \end{bmatrix} = \begin{bmatrix} h_1' \mathbb{1}_n \mu' \\ h_2' \mathbb{1}_n \mu' \\ \vdots \\ h_n' \mathbb{1} \end{bmatrix} = \begin{bmatrix} n^{1/2}\mu' \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (2.1.7)$$

En effet, compte tenu de (2.1.6) (voir (2.5.2)), et de la linéarité de l'espérance, on a, pour  $j = 1, \dots, n$ ,

$$\mathbb{E}(Y_j') = \mathbb{E}(h_j' \mathbb{X}) = (h_j' \mathbb{1}_n) \mu' = \begin{cases} n^{1/2}\mu' & \text{pour } j = 1, \\ \mathbb{O}'_m & \text{pour } j \geq 2, \end{cases}$$

et donc, par transposition de ces formules,

$$\mathbb{E}(Y_j) = \begin{cases} n^{1/2}\mu & \text{pour } j = 1, \\ \mathbb{O}_m & \text{pour } j \geq 2. \end{cases}$$

D'autre part, en posant, pour  $j = 1, \dots, n$ ,

$$h_j = \begin{bmatrix} h_{j,1} \\ \vdots \\ h_{j,n} \end{bmatrix},$$

on peut écrire les relations

$$Y_j' = h_j' \mathbb{X} = [h_{j,1} \quad \dots \quad h_{j,n}] \begin{bmatrix} X_1' \\ \vdots \\ X_n' \end{bmatrix} = \sum_{r=1}^n h_{j,r} X_r' \quad \Leftrightarrow \quad Y_j = \sum_{r=1}^n h_{j,r} X_r.$$

Par conséquent, on a les relations, pour  $1 \leq j, \ell \leq n$ ,

$$\begin{aligned} \mathbb{E}((Y_j - \mathbb{E}(Y_j))(Y_\ell - \mathbb{E}(Y_\ell))') &= \mathbb{E}\left(\left\{\sum_{r=1}^n h_{j,r}(X_r - \mu)\right\}\left\{\sum_{s=1}^n h_{\ell,s}(X_s - \mu)'\right\}\right) \\ &= \mathbb{E}\left(\sum_{r=1}^n \sum_{s=1}^n h_{j,r} h_{\ell,s} (X_r - \mu)(X_s - \mu)'\right) \\ &= \sum_{r=1}^n h_{j,r} h_{\ell,r} \Sigma = (h_j' h_\ell) \Sigma = \begin{cases} \Sigma & \text{si } j = \ell, \\ \mathbb{O}_{m,m} & \text{si } j \neq \ell. \end{cases} \end{aligned}$$

On en déduit que les vecteurs aléatoires  $Y_1, Y_2, \dots, Y_n$  sont indépendants, et tels que

$$Y_1 = n^{1/2} \bar{X} \stackrel{d}{=} N_m(n^{1/2}\mu, \Sigma) \quad \text{et} \quad Y_j \stackrel{d}{=} N_m(\mathbb{O}, \Sigma) \quad \text{pour } j = 2, \dots, m.$$

De plus, par (2.1.7), l'orthogonalité de  $H$  permet d'écrire que

$$\sum_{i=1}^n X_i X_i' = \mathbb{X}' \mathbb{X} = \mathbb{X}' H' H \mathbb{X} = \mathbb{Y}' \mathbb{Y} = \sum_{i=1}^n Y_i Y_i' = n \bar{X} \bar{X}' + \sum_{i=2}^n Y_i Y_i',$$

ce qui permet d'établir que

$$A = \sum_{i=1}^n X_i X_i' - n \bar{X} \bar{X}' = \sum_{i=2}^n Y_i Y_i' \equiv W_m(n-1, \Sigma),$$

est indépendant de  $\bar{X} = n^{-1/2} Y_1 v N_m(\mu, n^{-1}\Sigma)$ .  $\square$

### 2.1.3 Estimation du Maximum de Vraisemblance de $\mu$ et $\Sigma$ .

Les statistiques  $\bar{X}$  et  $A$  définies en (2.1.4) sont directement liées aux *estimateurs du maximum de vraisemblance* des paramètres  $\mu$  et  $\Sigma$  de la loi normale  $N_m(\mu, \Sigma)$ , à partir d'un échantillon  $X_1, \dots, X_n$  de cette loi. On a les résultats suivants.

Soit

$$\mathbb{X} = \begin{bmatrix} X_1' \\ \vdots \\ X_n' \end{bmatrix} \stackrel{d}{=} N_{n,m}(\mathbb{1}_n \mu', \mathbb{I}_n \otimes \Sigma) = N_{n,m}(\mathbb{1}_n \otimes \mu', \mathbb{I}_n \otimes \Sigma),$$

la matrice des observations associée à un *échantillon*  $X_1, \dots, X_n$ , de *taille*  $n \geq 1$ , composé de variables aléatoires indépendantes de même loi [i.i.d.], normale  $N_m(\mu, \Sigma)$ . Sous l'hypothèse que  $\Sigma > 0$  est définie positive,  $\mathbb{X}$  possède une densité, donnée comme le cas particulier de (3.4.14) obtenu pour  $C = \mathbb{I}$ ,  $D = \Sigma$  et  $M = \mathbb{1}_n \mu'$ , et que nous réécrivons ci-dessous par commodité sous la forme

$$\begin{aligned} f(\mathbf{X}; \mu, \Sigma) &= (2\pi)^{-mn/2} (\det \Sigma)^{-n/2} \text{etr} \left( -\frac{1}{2} (\mathbf{X} - \mathbb{1}_n \mu') \Sigma^{-1} (\mathbf{X} - \mathbb{1}_n \mu')' \right) \\ &= (2\pi)^{-mn/2} (\det \Sigma)^{-n/2} \text{etr} \left( -\frac{1}{2} \Sigma^{-1} (\mathbf{X} - \mathbb{1}_n \mu')' (\mathbf{X} - \mathbb{1}_n \mu') \right) \\ &= (2\pi)^{-mn/2} (\det \Sigma)^{-n/2} \text{etr} \left( -\frac{1}{2} \Sigma^{-1} (\mathbf{X} - \mathbb{E}(\mathbf{X}))' (\mathbf{X} - \mathbb{E}(\mathbf{X})) \right). \end{aligned} \quad (2.1.8)$$

Il est utile d'exprimer la *vraisemblance*

$$L(\mu, \Sigma) := f(\mathbb{X}; \mu, \Sigma), \quad (2.1.9)$$

obtenue en remplaçant  $\mathbf{X}$  par  $\mathbb{X}$  dans (2.1.8), et en faisant intervenir les expressions

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad A = \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})'. \quad (2.1.10)$$

On obtient le lemme suivant.

**Lemme 2.1.1.** *Sous les hypothèses ci-dessus, les matrices indépendantes  $A$  et  $\bar{X}$  sont telles que*

$$L(\mu, \Sigma) = f(\mathbb{X}; \mu, \Sigma) = (2\pi)^{-mn/2} (\det \Sigma)^{-N/2} \text{etr} \left( -\frac{1}{2} \Sigma^{-1} \left\{ A + n(\bar{X} - \mu)(\bar{X} - \mu)' \right\} \right). \quad (2.1.11)$$

**Preuve.** L'indépendance de  $\bar{X}$  et  $A$  est conséquence du théorème 2.1.1. Pour établir (2.1.11), on écrit que

$$\begin{aligned} (\mathbb{X} - \mathbb{E}(\mathbb{X}))' (\mathbb{X} - \mathbb{E}(\mathbb{X})) &= \sum_{i=1}^n (X_i - \mu)(X_i - \mu)' \\ &= \sum_{i=1}^n (X_i - \bar{X} + \bar{X} - \mu)(X_i - \bar{X} + \bar{X} - \mu)' \\ &= n(\bar{X} - \mu)(\bar{X} - \mu)' + \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})' \\ &= A + n(\bar{X} - \mu)(\bar{X} - \mu)'. \end{aligned}$$

On conclut en remplaçant  $\mathbf{X}$  par  $\mathbb{X}$  dans (2.1.8).  $\square$

Compte tenu de (2.1.11), on obtient aisément l'expression des estimateurs  $\hat{\mu}$  et  $\hat{\Sigma}$  du maximum de vraisemblance de  $\mu$  et  $\Sigma$ . Ceux-ci sont, par définition, les valeurs (ici uniques) de  $\hat{\mu}$  et  $\hat{\Sigma}$ , telles que

$$L(\hat{\mu}, \hat{\Sigma}) = \sup_{\mu, \Sigma} L(\mu, \Sigma). \quad (2.1.12)$$

**Théorème 2.1.2.** *Sous l'hypothèse que  $\Sigma > 0$ , les estimateurs  $\hat{\mu}$  et  $\hat{\Sigma}$ , du maximum de vraisemblance de  $\mu$  et de  $\Sigma$ , basés sur l'observation de l'échantillon aléatoire  $X_1, \dots, X_n$  de la loi normale  $N_m(\mu, \Sigma)$ , sont donnés par*

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad \hat{\Sigma} = \frac{1}{n} A = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})'. \quad (2.1.13)$$

De plus, indépendamment de  $\Sigma > 0$ ,

$$L(\hat{\mu}, \Sigma) = \sup_{\mu} L(\mu, \Sigma) = L(\bar{X}, \Sigma) = (2\pi)^{-mn/2} (\det \Sigma)^{-n/2} \text{etr} \left( -\frac{1}{2} \Sigma^{-1} A \right), \quad (2.1.14)$$

et

$$L(\hat{\mu}, \hat{\Sigma}) = \sup_{\mu, \Sigma} L(\mu, \Sigma) = \sup_{\Sigma} L(\hat{\mu}, \Sigma) = (2\pi)^{-mn/2} n^{mn/2} (\det A)^{-n/2} e^{-mn/2}. \quad (2.1.15)$$

**Preuve.** De (2.1.11), on déduit aisément que

$$\begin{aligned} L(\mu, \Sigma) &= (2\pi)^{-mn/2} (\det \Sigma)^{-n/2} \text{etr} \left( -\frac{1}{2} \Sigma^{-1} A \right) \exp \left( -\frac{n}{2} (\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu) \right) \\ &\leq L(\bar{X}, \Sigma) = (2\pi)^{-mn/2} (\det \Sigma)^{-n/2} \text{etr} \left( -\frac{1}{2} \Sigma^{-1} A \right), \end{aligned}$$

avec égalité si et seulement si  $\mu = \bar{X}$ . Ceci établit (2.1.14). La preuve de (2.1.15) est plus délicate. On écrit que

$$\begin{aligned} \log L(\hat{\mu}, \Sigma) &= -\frac{mn}{2} \log(2\pi) - \frac{n}{2} \log \{ \det \Sigma \} - \frac{1}{2} \text{tr} \{ \Sigma^{-1} A \} \\ &= -\frac{mn}{2} \log(2\pi) + \frac{n}{2} \log \left\{ \det A^{1/2} \Sigma^{-1} A^{1/2} \right\} - \frac{1}{2} \text{tr} \left\{ A^{1/2} \Sigma^{-1} A^{1/2} \right\} \\ &\quad - \frac{n}{2} \log \det A \\ &= -\frac{mn}{2} \log(2\pi) + \frac{1}{2} \sum_{i=1}^m (n \log \lambda_i - \lambda_i) - \frac{n}{2} \log \det A, \end{aligned}$$

où  $\lambda_1, \dots, \lambda_m$  désignent les valeurs propres de la matrice symétrique positive  $A^{1/2} \Sigma^{-1} A^{1/2}$ . Or, la fonction

$$g(\lambda) = n \log \lambda - \lambda,$$

lorsque  $\lambda$  varie dans  $\mathbb{R}^+$ , atteint son maximum pour  $\lambda = n$ . On en déduit que

$$\log L(\hat{\mu}, \Sigma) \leq -\frac{mn}{2} \log(2\pi) + \frac{m}{2} (n \log n - n) - \frac{n}{2} \log \det A,$$

avec égalité si et seulement si  $\lambda_1 = \dots = \lambda_m = n$ . Or, cette dernière condition équivaut à  $A^{1/2} \Sigma^{-1} A^{1/2} = n \mathbb{I}_m$ , et donc, au fait que  $\Sigma = \hat{\Sigma} = (1/n)A$ . Faisant usage de (2.1.14), on en déduit que

$$\sup_{\mu, \Sigma} L(\mu, \Sigma) = \sup_{\Sigma} L(\hat{\mu}, \Sigma) = L(\hat{\mu}, \hat{\Sigma}) = L(\bar{X}, n^{-1}A) = (2\pi)^{-mn/2} (\det \{n^{-1}A\})^{-n/2} \text{etr} \left( -\frac{1}{2} n \mathbb{I} \right)$$

La conclusion (2.1.15) est alors immédiate.  $\square$

**Remarque 2.1.1.** La démonstration du théorème 2.1.2 montre que, d'une manière générale, pour toute matrice symétrique  $A > 0$  ( $m \times m$ ), et pour tout  $\nu > 0$  (non nécessairement entier),

$$\sup_{\Sigma > 0} \left\{ (\det \Sigma)^{-\nu/2} \text{etr} \left( -\frac{1}{2} \Sigma^{-1} A \right) \right\} = \nu^{m\nu/2} e^{-m\nu/2} (\det A)^{-\nu/2}, \quad (2.1.16)$$

le maximum étant atteint pour  $\Sigma = (1/\nu)A$ .

**Théorème 2.1.3.** *Sous l'hypothèse que  $\sigma > 0$ , les estimateurs  $\hat{\mu}$  et  $\hat{\sigma}^2$  du maximum de vraisemblance de  $\mu$  et de  $\sigma$ , basés sur l'observation de l'échantillon aléatoire  $X_1, \dots, X_n$  de la loi normale  $N_m(\mu, \sigma^2 \mathbb{I}_m)$ , sont donnés par*

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad \hat{\sigma}^2 = \frac{1}{m} \hat{\Sigma} = \frac{1}{mn} \text{tr} A = \frac{1}{mn} \text{tr} \left\{ \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})' \right\}. \quad (2.1.17)$$

De plus, indépendamment de  $\sigma > 0$ ,

$$L(\hat{\mu}, \lambda \mathbb{I}_m) = \sup_{\mu} L(\mu, \sigma^2 \mathbb{I}_m) = L(\bar{X}, \sigma^2 \mathbb{I}_m) = (2\pi)^{-mn/2} \sigma^{-mn} \text{etr} \left\{ -\frac{1}{2\sigma^2} A \right\}, \quad (2.1.18)$$

et

$$L(\hat{\mu}, \hat{\sigma}^2 \mathbb{I}_m) = \sup_{\mu, \sigma} L(\mu, \sigma^2 \mathbb{I}_m) = \sup_{\sigma} L(\hat{\mu}, \sigma^2 \mathbb{I}_m) = (2\pi)^{-mn/2} \left\{ \frac{1}{mn} \text{tr} A \right\}^{-mn/2} e^{-mn/2}. \quad (2.1.19)$$



**Preuve.** En remplaçant  $\Sigma$  par  $\sigma^2 \mathbb{I}_n$  dans (2.1.11), on obtient que

$$\begin{aligned} L(\mu, \sigma^2 \mathbb{I}_m) &= (2\pi)^{-mn/2} (\det \sigma^2 \mathbb{I}_m)^{-n/2} \text{etr} \left\{ -\frac{1}{2\sigma^2} \left\{ A + n(\bar{X} - \mu)(\bar{X} - \mu)' \right\} \right\} \\ &= (2\pi)^{-mn/2} \sigma^{-mn} \text{etr} \left\{ -\frac{1}{2\sigma^2} A \right\} \exp \left\{ -\frac{n}{2\sigma^2} (\bar{X} - \mu)' (\bar{X} - \mu) \right\} \\ &\leq (2\pi)^{-mn/2} \sigma^{-mn} \text{etr} \left\{ -\frac{1}{2\sigma^2} A \right\} = (2\pi)^{-mn/2} \sigma^{-mn} \exp \left\{ -\frac{1}{2\sigma^2} \text{tr} A \right\}, \end{aligned}$$

le maximum étant atteint, indépendamment de  $\sigma > 0$ , pour  $\mu = \hat{\mu} = \bar{X}$ . Ceci établit donc (2.1.18). On constate ensuite que

$$\frac{d}{d(\sigma^2)} \log L(\bar{X}, \sigma^2 \mathbb{I}_m) = -\frac{mn}{2\sigma^4} + \frac{1}{2\sigma^2} \text{tr} A = 0,$$

pour  $\sigma^2 = \hat{\sigma}^2 = \frac{1}{mn} \text{tr} A$ , ce qui établit (2.1.19).  $\square$

#### 2.1.4 Test du rapport de vraisemblance de l'hypothèse de sphéricité.

Les résultats de ce paragraphe sont des conséquences directes des développements des §2.1.2 & 2.1.3.

Par *hypothèse de sphéricité* sur  $X \stackrel{d}{=} N_m(\mu, \Sigma)$ , on entend l'hypothèse que la matrice de variances-covariances, de la forme  $\Sigma = \sigma^2 \mathbb{I}_m$ , est *diagonale*, pour une valeur convenable de  $\sigma > 0$ . Les résultats précédents permettent une description simple du test du *rapport de vraisemblance* de cette hypothèse, contre l'alternative que  $\Sigma > 0$  est quelconque.

**Corollaire 2.1.1.** *Etant donné un échantillon aléatoire  $X_1, \dots, X_n$ , issu d'une loi normale  $N_m(\mu, \Sigma)$ , la statistique du test du rapport de vraisemblance de l'hypothèse nulle*

$$(H.0) \Sigma = \sigma^2 \mathbb{I}_m \text{ pour une valeur de } \sigma > 0;$$

*contre l'hypothèse alternative*

$$(H.1) \Sigma > 0 \text{ est quelconque};$$

*est définie par*

$$\mathcal{T} = \frac{\sup_{\sigma > 0} L(\hat{\mu}, \sigma^2 \mathbb{I}_m)}{\sup_{\mu, \Sigma} L(\mu, \Sigma)} = \left\{ \frac{\det A}{\left( \frac{1}{mn} \text{tr} A \right)^m} \right\}^{n/2}. \quad (2.1.20)$$

**Preuve.** On déduit aisément (2.1.20) de (2.1.15) et (2.1.19).  $\square$

**Remarque 2.1.2.** Il est commode d'écrire  $\mathcal{T} = \mathcal{V}^{n/2}$ , où

$$\mathcal{V} = \frac{\det A}{\left( \frac{1}{mn} \text{tr} A \right)^m}. \quad (2.1.21)$$

Le rejet de l'hypothèse (H.0) est effectué au seuil  $\alpha$  lorsque  $\mathcal{V} \leq v_\alpha$ , où  $v_\alpha$  est choisi de sorte que

$$\mathbb{P}(\mathcal{V} \leq v_\alpha \mid (H.0)) = \alpha.$$

On obtient, par application de la théorie générale des tests du rapport de vraisemblance, que, sous (H.0), la loi limite de  $-2 \log \mathcal{T}$ , lorsque  $n \rightarrow \infty$ , est donnée par (voir (2.4.14))

$$-2 \log \mathcal{T} = -n \log \mathcal{V} \xrightarrow{d} \chi_f^2 \quad \text{où} \quad f = \frac{(m+2)(m-1)}{2} = \frac{m(m+1)}{2} - 1. \quad (2.1.22)$$

La valeur de  $f$  correspond au nombre  $\frac{m(m+1)}{2}$ , de paramètres libres caractérisant  $\Sigma$  sous (H.1), duquel on retranche 1, nombre de paramètres libres caractérisant  $\Sigma = \sigma^2 \mathbb{I}_m$  sous (H.0).

### 2.1.5 Densité de la loi de Wishart centrée.

Sous les hypothèses du §2.1.1, revenons à l'étude du *cas centré*, correspondant à un échantillon  $Z_1, \dots, Z_n$  de  $n \geq 1$  vecteurs aléatoires indépendants, de même loi normale centrée  $N_m(\mathbb{O}, \Sigma)$ . On pose alors

$$A = \mathbb{Z}'\mathbb{Z} = \sum_{i=1}^n Z_i Z_i' \stackrel{d}{=} W_m(n, \Sigma) \quad \text{avec} \quad \mathbb{Z} = [Z_1 \ \dots \ Z_n]' = \begin{bmatrix} Z_1' \\ \vdots \\ Z_n' \end{bmatrix}.$$

Comme précédemment,  $\Sigma \geq 0$  désigne une matrice  $(m \times m)$  symétrique positive.

D'une manière générale, comme  $A = \mathbb{Z}'\mathbb{Z}$ , le rang de la matrice  $(n \times n)$   $\mathbb{Z}$  vérifie les relations

$$\text{rg } A = \text{rg } \mathbb{Z} \leq \min\{m, n\}.$$

D'autre part, on a nécessairement  $\text{rg } \mathbb{Z} \leq \text{rg } \Sigma \leq m$ . Donc, si  $A \stackrel{d}{=} W_m(n, \Sigma)$  est comme en (2.1.2), on a nécessairement

$$\text{rg } A \leq \min(n, \text{rg } \Sigma). \quad (2.1.23)$$

De ce fait la loi de Wishart ne peut posséder de densité que si, d'une part,  $n \geq m$  (le nombre de degrés de liberté est au moins égal à la dimension de l'espace), et si, d'autre part, la matrice de variances-covariances  $\Sigma$  est définie positive.

Le théorème suivant donne l'expression de cette densité lorsque ces deux conditions sont satisfaites. Rappelons la définition 1.2.1 de la fonction gamma multivariée  $\Gamma_m(a)$  pour  $a > \frac{1}{2}(m-1)$ , qui, dans le cas particulier où  $a = n/2$  donne la relation

$$\Gamma_m\left(\frac{1}{2}n\right) = \pi^{\frac{m(m-1)}{4}} \prod_{i=1}^m \Gamma\left(\frac{1}{2}(n-i+1)\right).$$

**Théorème 2.1.4.** (*Densité de la loi de Wishart*) Si  $A \equiv W_m(n, \Sigma)$ , avec  $n \geq m$  et  $\Sigma > 0$ , alors la densité de  $A$  existe et est donnée par

$$f(A) = \frac{2^{-mn/2}}{\Gamma_m\left(\frac{1}{2}n\right)} (\det \Sigma)^{-n/2} (\det A)^{\frac{1}{2}n - \frac{1}{2}(m+1)} \text{etr} \left(-\frac{1}{2}\Sigma^{-1}A\right), \quad (2.1.24)$$

pour  $A > 0$ , et  $f(A) = 0$  autrement.

**Preuve.** Posons  $M = \mathbb{O}$ ,  $C = \mathbb{I}_n$ ,  $D = \Sigma$  et  $\mathbb{X} = \mathbb{Z}$  dans (3.4.14). En utilisant le fait que

$$\text{tr}(\mathbb{Z}\Sigma^{-1}\mathbb{Z}') = \text{tr}(\Sigma^{-1}\mathbb{Z}'\mathbb{Z}),$$

on constate que l'élément différentiel correspondant à la densité de  $\mathbb{Z}$  est donné par l'expression

$$\begin{aligned} & \left\{ \prod_{i=1}^n (2\pi)^{-m/2} (\det \Sigma)^{-1/2} \exp\left(-\frac{1}{2}Z_i' \Sigma^{-1} Z_i\right) \right\} (d\mathbb{Z}) \\ &= (2\pi)^{-mn/2} (\det \Sigma)^{-n/2} \text{etr} \left(-\frac{1}{2}\Sigma^{-1}\mathbb{Z}'\mathbb{Z}\right) (d\mathbb{Z}) \\ &= (2\pi)^{-mn/2} (\det \Sigma)^{-n/2} \text{etr} \left(-\frac{1}{2}\Sigma^{-1}A\right) (d\mathbb{Z}). \end{aligned}$$

Effectuons maintenant le changement de variable  $\mathbb{Z} = H_1 T$ , comme en (1.3.1). On obtient alors le jacobien de ce changement de variable par (1.3.13), ce qui permet d'écrire

$$(d\mathbb{Z}) = 2^{-m} (\det A)^{\frac{1}{2}n - \frac{1}{2}(m+1)} [dA][H_1' dH_1].$$

L'élément différentiel correspondant à la densité jointe de  $A$  et de  $H_1$  est donc donné par

$$(2\pi)^{-mn/2} (\det \Sigma)^{-n/2} \text{etr} \left(-\frac{1}{2}\Sigma^{-1}A\right) 2^{-m} (\det A)^{\frac{1}{2}n - \frac{1}{2}(m+1)} [dA][H_1' dH_1].$$

Comme, par (1.4.1),

$$\int_{\mathcal{V}_{m,n}} [H'_1 dH_1] = \frac{2^m \pi^{mn/2}}{\Gamma_m(\frac{1}{2}n)},$$

on obtient alors directement (2.1.24) après intégration relativement à  $H_1$  sur  $\mathcal{V}_{m,n}$ .  $\square$

La forme de la densité donnée en (2.1.24) a été obtenue pour la première fois, dans le cas général où  $m \geq 1$  est quelconque, en 1928 par Wishart (Wishart, J. (1928). The generalized product moment distribution in samples from a normal multivariate population. *Biometrika* **20A** 32–43). Lorsque  $m = 1$  et  $\Sigma = \sigma^2 \mathbb{I} = [\sigma^2]$ , la loi de Wishart  $W_1(n, \sigma^2 \mathbb{I})$  coïncide avec celle de la loi  $\sigma^2 \chi_n^2$ .

### 2.1.6 Décomposition de Bartlett associée à la loi de Wishart centrée.

Lorsque  $\Sigma = \mathbb{I}_m$ , la loi de Wishart  $W_m(n, \Sigma) = W_m(n, \mathbb{I})$  admet une décomposition particulièrement simple, décrite dans le théorème suivant.

**Théorème 2.1.5.** (*Décomposition de Bartlett*) Soit  $A \stackrel{d}{=} W_m(n, \mathbb{I}_m)$ , avec  $n \geq m$ , et soit  $A = T'T$  la décomposition unique de  $A$ , obtenue lorsque  $T = [t_{i,j}]$  est une matrice triangulaire supérieure à éléments diagonaux  $t_{i,i}$  positifs, pour  $1 \leq i \leq m$ . Alors, les composantes  $t_{i,j}$ ,  $1 \leq i \leq j \leq m$ , de  $T$  sont des variables aléatoires indépendantes, telles que  $t_{i,i}^2 \stackrel{d}{=} \chi_{n-i+1}^2$ , pour  $1 \leq i \leq m$ , et  $t_{i,j} \stackrel{d}{=} N(0, 1)$  pour  $1 \leq i < j \leq m$ .

**Preuve.** Remarquons qu'une matrice triangulaire supérieure est, par définition, de la forme

$$T = \begin{bmatrix} t_{1,1} & \cdots & t_{1,m} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & t_{m,m} \end{bmatrix}.$$

L'élément différentiel de la distribution de la matrice  $A$  est donné par (2.1.24), soit

$$\frac{2^{-mn/2}}{\Gamma_m(\frac{1}{2}n)} (\det A)^{\frac{1}{2}n - \frac{1}{2}(m+1)} \text{etr}\left(-\frac{1}{2}A\right)(dA).$$

Comme  $A = T'T$ , on a

$$\text{tr } A = \text{tr } T'T = \sum_{1 \leq i \leq j \leq m} t_{i,j}^2, \quad \text{et} \quad \det A = \det T'T = (\det T)^2 = \prod_{i=1}^m t_{i,i}^2.$$

D'autre part, par (1.1.7), on a

$$(dA) = 2^m \prod_{i=1}^m t_{i,i}^{m-i+1} (dT) = 2^m \prod_{i=1}^m t_{i,i}^{m-i+1} \bigwedge_{1 \leq i \leq j \leq m} dt_{i,j}, \quad (2.1.25)$$

tandis que, par (1.2.4), on a

$$\Gamma_m(\frac{1}{2}n) = \pi^{\frac{1}{4}m(m-1)} \prod_{i=1}^m \Gamma\left(\frac{1}{2}(n-i+1)\right).$$

En effectuant les remplacements successifs dans l'élément différentiel de  $A$  de  $\text{tr } A$ ,  $\det A$ ,  $(dA)$  et  $\Gamma_m(\frac{1}{2}n)$ , par leurs expressions respectives, données ci-dessus, on obtient que l'élément différentiel associé à la densité jointe des  $t_{i,j}$ , pour  $1 \leq i \leq j \leq m$ , se factorise sous la forme

$$\left\{ \prod_{1 \leq i < j \leq m} \left( \frac{\exp(-\frac{1}{2}t_{i,j}^2)}{\sqrt{2\pi}} dt_{i,j} \right) \right\} \left\{ \prod_{i=1}^m \left( \frac{2^{-\frac{1}{2}(n-i+1)}}{\Gamma(\frac{1}{2}(n-i+1))} (t_{i,i}^2)^{\frac{1}{2}(n-i+1)-1} \exp(-\frac{1}{2}t_{i,i}^2) d(t_{i,i}^2) \right) \right\}.$$

Ceci permet de conclure, par (3.1.1) et (4.1.9).  $\square$

### 2.1.7 Variance Généralisée.

Les résultats principaux de ce paragraphe sont des conséquences naturelles de la décomposition de Bartlett (théorème 2.1.5 du §2.1.6).

La notion de variance se généralise de plusieurs manières dans  $\mathbb{R}^m$  pour  $m \geq 2$ . On utilise souvent la définition suivante, due à Wilks (1932) (Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, **24** 471-494).

**Définition 2.1.2.** Soit  $X \in \mathbb{R}^m$  un vecteur aléatoire de matrice de variances-covariances  $\Sigma = \text{Var}(X)$ . On appelle variance généralisée de  $X$ , le déterminant  $\det \Sigma$  de  $\Sigma$ .

Si  $X_1, \dots, X_n$  désigne un échantillon de taille  $n = N + 1$  d'une loi normale  $N_m(\mu, \Sigma)$ , on estime la variance généralisée :

- Soit par la variance généralisée d'échantillonnage  $\det S$
- Soit par la variance généralisée empirique  $\det \hat{\Sigma}$ .

Ici, on pose

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad (2.1.26)$$

$$A = \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})' \stackrel{d}{=} W_m(n-1, \Sigma), \quad \hat{\Sigma} = \frac{1}{n} A, \quad (2.1.27)$$

$$S = \frac{n}{n-1} \hat{\Sigma} = \frac{1}{n-1} A = \frac{1}{N} A \quad \text{lorsque } n = N + 1 \geq 2. \quad (2.1.28)$$

Le théorème suivant est une conséquence simple de la décomposition de Bartlett (Théorème 2.1.5).

**Théorème 2.1.6.** Soit  $A \stackrel{d}{=} W_m(N, \Sigma)$  une matrice de Wishart, où  $m \geq N \geq 1$ , et où la matrice  $\Sigma > 0$  est supposée définie positive. On pose  $S = N^{-1}A$ . Alors,

$$N^m \frac{\det S}{\det \Sigma} = \frac{\det A}{\det \Sigma} \stackrel{d}{=} \prod_{i=1}^m \chi_{N-i+1}^2, \quad (2.1.29)$$

où  $\prod_{i=1}^m \chi_{N-i+1}^2$  désigne symboliquement le produit de  $m$  variables aléatoires indépendantes de lois respectives  $\chi_{N-i+1}^2$ , pour  $i = 1, \dots, m$ .

**Preuve.** Si  $A \stackrel{d}{=} W_m(N, \Sigma)$ , alors  $B = \Sigma^{-1/2} A \Sigma^{-1/2} \stackrel{d}{=} W_m(N, \mathbb{I}_m)$ . Posons alors  $B = T'T$ , où  $T = [t_{i,j}]$  est une matrice triangulaire supérieure, à éléments diagonaux positifs. Le résultat est alors une conséquence directe, d'une part du théorème 2.1.5, qui exprime que les  $t_{i,i}^2$  sont des variables aléatoires indépendantes, suivant, respectivement, des lois du  $\chi_{n-i+1}^2$ , pour  $i = 1, \dots, m$ , et d'autre part, de l'observation que

$$\frac{\det A}{\det \Sigma} = \det B = (\det T)^2 = \prod_{i=1}^m t_{i,i}^2. \quad (2.1.30)$$

Cette dernière relation est, de toute évidence, équivalente à (2.1.29).□

Hormis les cas où l'expression de la loi du produit de variables aléatoires suivant des lois du  $\chi^2$  est simple, il est commode d'utiliser l'approximation suivante pour la distribution de la statistique donnée par (2.1.29).

**Corollaire 2.1.2.** Soit  $A \stackrel{d}{=} W_m(N, \Sigma)$  une matrice de Wishart dont le nombre  $N$  de degrés de liberté est supérieur ou égal à la dimension  $m$  de l'espace,  $N \geq m$ , et correspondant à une matrice  $\Sigma > 0$  définie positive. On pose  $S = N^{-1}A$ . Alors, lorsque  $N \rightarrow \infty$ ,

$$\sqrt{\frac{N}{2m}} \log \left\{ \frac{\det S}{\det \Sigma} \right\} \stackrel{d}{\rightarrow} N(0, 1). \quad (2.1.31)$$

**Preuve.** On constate tout d'abord que si  $\eta_N \stackrel{d}{=} \chi_N^2$ , on peut poser  $\eta_N = \sum_{i=1}^N \zeta_i$ , où  $\zeta_1, \dots, \zeta_n$  désignent des variables aléatoires, indépendantes, et de même loi du  $\chi_1^2$ . Comme alors  $\mathbb{E}(\zeta_i) = 1$  et  $\text{Var}(\zeta_i) = 2$ , pour  $i = 1, \dots, N$ , on a, par le théorème central limite, la convergence en loi, lorsque  $N \rightarrow \infty$ ,

$$\sqrt{\frac{N}{2}} (N^{-1}\eta_N - 1) = \sqrt{\frac{1}{2N}} (\eta_N - N) \xrightarrow{d} N(0, 1).$$

Il s'ensuit que, lorsque  $N \rightarrow \infty$ , la variable  $\theta_N = N^{-1}\eta_N - 1$  vérifie  $\theta_N = N^{-1}\eta_N - 1 = O_{\mathbb{P}}(N^{-1/2})$ . On a donc, en conséquence, lorsque  $N \rightarrow \infty$ ,

$$\log \eta_N - \log N = \log(1 + \theta_N) = \theta_N + O_{\mathbb{P}}(N^{-1}).$$

Par conséquent, on constate que, lorsque  $N \rightarrow \infty$ ,

$$\sqrt{\frac{N}{2}} (\log \eta_N - \log N) = \sqrt{\frac{N}{2}} \theta_N + O_{\mathbb{P}}(N^{-1/2}) \xrightarrow{d} N(0, 1). \quad (2.1.32)$$

Nous nous sommes, ici, servis du développement asymptotique  $\log(1+z) = z + O(z^2)$  lorsque  $z \rightarrow 0$ . Supposons maintenant que  $t_{i,i}^2 \stackrel{d}{=} \chi_{N-i+1}^2$  soit comme en (2.3.11). Alors, on déduit directement de la convergence (2.1.32), que, indépendamment de  $i = 1, \dots, m$ , lorsque  $N \rightarrow \infty$ ,

$$\sqrt{\frac{N}{2}} (\log t_{i,i}^2 - \log N) \xrightarrow{d} N(0, 1).$$

Comme les  $t_{i,i}^2$ ,  $i = 1, \dots, m$ , sont indépendants, il s'ensuit que, lorsque  $N \rightarrow \infty$ ,

$$\sqrt{\frac{N}{2}} \sum_{i=1}^m (\log t_{i,i}^2 - \log N) \xrightarrow{d} N(0, m).$$

Cette propriété, jointe au fait que, par (2.1.30), on a

$$\log \left\{ \frac{\det S}{\det \Sigma} \right\} = \log \left\{ \frac{\det A}{\det \Sigma} \right\} - m \log N = \sum_{i=1}^m (\log t_{i,i}^2 - \log N),$$

relation qui mène directement à (2.1.31).  $\square$

**Remarque 2.1.3.** D'une manière générale, on écrit, pour une suite  $u_n > 0$ , que  $Y_n = O_{\mathbb{P}}(u_n)$  lorsque  $n \rightarrow \infty$ , si on a

$$\lim_{r \rightarrow \infty} \left\{ \limsup_{n \rightarrow \infty} \mathbb{P}(|Y_n| \geq r u_n) \right\} = 0.$$

De manière analogue, on écrit que  $Y_n = o_{\mathbb{P}}(u_n)$ , lorsque  $n \rightarrow \infty$ , si, pour tout  $r > 0$  fixé,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|Y_n| \geq r u_n) = 0.$$

## 2.2 Propriétés de la loi de Wishart centrée.

### 2.2.1 Fonction caractéristique de la loi de Wishart centrée.

Soit  $A = [a_{\ell,j}] \stackrel{d}{=} W_m(n, \Sigma)$  une matrice aléatoire ( $m \times m$ ) symétrique. On désigne par  $\Theta = [\theta_{\ell,j}]$  est une matrice (non-aléatoire) symétrique ( $m \times m$ ). Observons que

$$\text{tr}(\Theta A) = \sum_{\ell=1}^m \sum_{j=1}^m \theta_{\ell,j} a_{\ell,j} = \sum_{\ell=1}^m \theta_{\ell,\ell} a_{\ell,\ell} + 2 \sum_{1 < \ell < j \leq m} \theta_{\ell,j} a_{\ell,j}.$$

Désignons par  $U = [u_{\ell,j}]$ , une matrice symétrique (non-aléatoire)  $(m \times m)$ . Si on définit  $\Theta = [\theta_{\ell,j}]$ , en fonction de  $U$  par les relations

$$\theta_{i,j} = \begin{cases} \frac{1}{2}u_{\ell,j} & \text{si } i \neq j, \\ u_{\ell,\ell} & \text{si } \ell = j, \end{cases}$$

on constate que

$$\sum_{1 \leq j \leq \ell \leq m} u_{\ell,j} a_{\ell,j} = \sum_{\ell=1}^m \sum_{j=1}^m \theta_{\ell,j} a_{\ell,j} = \text{tr}(\Theta A).$$

Ces préliminaires nous amènent à définir la *fonction caractéristique* de  $A$ , par l'une ou l'autre des fonctions

$$\phi_A(U) = \mathbb{E} \left\{ \exp \left( i \sum_{1 \leq \ell \leq j \leq m} u_{\ell,j} a_{\ell,j} \right) \right\} = \phi_A^*(\Theta) = \mathbb{E} \left\{ \exp (i \text{tr}(A\Theta)) \right\}.$$

Comme dans les calculs la forme  $\phi_A^*(\Theta)$  de la fonction caractéristique est généralement plus commode à exprimer que  $\phi_A(U)$ , nous nous limiterons, dans ce qui suit, à l'évaluation de  $\phi_A^*(\Theta)$ , lorsque  $A$  suit une loi de Wishart. On obtient alors le théorème suivant.

**Théorème 2.2.1.** *Soit  $A \stackrel{d}{=} W_m(n, \Sigma)$ . Alors,*

$$\phi_A^*(\Theta) = \mathbb{E} \left\{ \exp (i \text{tr}(A\Theta)) \right\} = \det(\mathbb{I}_m - 2i\Theta\Sigma)^{-m/2}. \quad (2.2.1)$$

**Preuve.** Soient  $X = X_1, \dots, X_n$  des vecteurs aléatoires i.i.d., de loi comune  $N_m(0, \Sigma)$ . On peut alors écrire, sans perte de généralité,  $A$  sous la forme

$$A = \sum_{j=1}^n X_j X_j'.$$

On en déduit que

$$\phi_A^*(\Theta) = \mathbb{E} \left\{ \prod_{j=1}^n \text{etr}(iX_j X_j' \Theta) \right\} = \mathbb{E} \left\{ \exp (iX' \Theta X) \right\}^n.$$

Posons maintenant  $X = \Sigma^{1/2}Y$ , où  $Y \stackrel{d}{=} N_m(0, \mathbb{I}_m)$ . Compte tenu de ce qui précède, l'évaluation de  $\phi_A^*(\Theta)$  se ramène à l'évaluation de

$$J(\Theta) = \mathbb{E} \left\{ \exp (iX' \Theta X) \right\} = \mathbb{E} \left\{ \exp (iY' \Sigma^{1/2} \Theta \Sigma^{1/2} Y) \right\}.$$

Comme  $\Lambda = \Sigma^{1/2} \Theta \Sigma^{1/2}$  est symétrique  $(m \times m)$ , il existe une matrice orthogonale  $H$  telle que

$$\Lambda = H \Sigma^{1/2} \Theta \Sigma^{1/2} H' = \text{diag}(\lambda_1, \dots, \lambda_m),$$

où  $\lambda_1, \dots, \lambda_m$  sont les valeurs propres de  $\Sigma^{1/2} \Theta \Sigma^{1/2}$ . Posons  $Z = H'Y = [Z_1 \ \dots \ Z_m]'$  ou, de manière équivalente,  $Y = HZ$ . De toute évidence,  $Z \stackrel{d}{=} N_m(0, \mathbb{I}_m)$ , et les composantes  $Z_1, \dots, Z_m$ , de  $Z$ , sont i.i.d.  $N(0, 1)$ . On a donc

$$\begin{aligned} J(\Theta) &= \mathbb{E} \left\{ \exp (iZ' \Lambda Z) \right\} = \mathbb{E} \left\{ \exp \left( i \sum_{j=1}^m \lambda_j Z_j^2 \right) \right\} \\ &= \prod_{j=1}^m \left\{ 1 - 2i\lambda_j \right\}^{-1/2} = \det \left( \mathbb{I}_m - 2i\Lambda \right)^{-1/2} = \det \left( \mathbb{I}_m - 2i\{H\Sigma^{1/2}\} \Theta \Sigma^{1/2} H' \right)^{-1/2} \\ &= \det \left( \mathbb{I}_m - 2i\Theta \Sigma^{1/2} H' \{H\Sigma^{1/2}\} \right)^{-1/2} = \det \left( \mathbb{I}_m - 2i\Theta \Sigma \right)^{-1/2}, \end{aligned}$$

où nous nous sommes servis de (2.6.11). La conclusion (2.2.1) est alors immédiate.  $\square$

### 2.2.2 Loi de Wishart et régression linéaire.

Le théorème suivant a une importance fondamentale dans une série d'applications. Nous considérons le cas où  $\Sigma > 0$  est une matrice  $(m \times m)$  définie positive, partitionnée en blocs, de même que  $A \stackrel{d}{=} W_m(n, \Sigma)$ , sous la forme suivante. On pose

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} \quad \text{et} \quad \Sigma = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix}. \quad (2.2.2)$$

On suppose que  $\Sigma_{1,1}$  et  $A_{1,1}$  sont  $(k \times k)$ , et que  $\Sigma_{2,2}$  et  $A_{2,2}$  sont  $((m-k) \times (m-k))$ . De plus, on raisonne dans le cas où  $1 \leq k < m$ .

**Théorème 2.2.2.** *Supposons  $n \geq m$ . On pose*

$$A_{11.2} = A_{1,1} - A_{1,2}A_{2,2}^{-1}A_{2,1} \quad \text{et} \quad \Sigma_{1,1.2} = \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1}. \quad (2.2.3)$$

Alors, sous les hypothèses ci-dessus,

- (1)  $A_{1,1} \stackrel{d}{=} W_k(n, \Sigma_{1,1})$  et  $A_{2,2} \stackrel{d}{=} W_{m-k}(n, \Sigma_{2,2})$ ;
- (2)  $A_{1,1.2} \stackrel{d}{=} W_k(n-m+k, \Sigma_{1,1.2})$  est indépendant de  $A_{1,2}$  et  $A_{2,2}$ .
- (3) Conditionnellement à  $A_{2,2} = a_{2,2}$ ,  $A_{1,2}$  suit une loi normale matricielle

$$\mathcal{L}(A_{1,2} | A_{2,2} = a_{2,2}) \equiv N_{k, m-k}(\Sigma_{1,2}\Sigma_{2,2}^{-1}a_{2,2}, \Sigma_{1,1.2} \otimes a_{2,2}). \quad (2.2.4)$$

**Preuve.** Pour démontrer (1), il suffit de revenir à la définition (2.1.2) de  $W_m(n, \Sigma)$ . Désignons par

$$Y_i = \begin{bmatrix} Y_{i,1} \\ Y_{i,2} \end{bmatrix} \stackrel{d}{=} N_m \left( \mathbb{O}_m, \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix} \right), \quad i = 1, \dots, n,$$

une suite de vecteurs aléatoires indépendants de même loi. On observe que, respectivement,

$$Y_{i,1} \stackrel{d}{=} N_k(\mathbb{O}_k, \Sigma_{1,1}), \quad i = 1, \dots, n \quad \text{et} \quad Y_{i,2} \stackrel{d}{=} N_{m-k}(\mathbb{O}_{m-k}, \Sigma_{2,2}), \quad i = 1, \dots, n,$$

forment deux suites, composées, l'une comme l'autre, de vecteurs aléatoires indépendants de même loi. Par conséquent, si nous posons, par (2.1.2),

$$A = \sum_{i=1}^n Y_i Y_i' = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix}$$

alors, toujours par (2.1.2), nous avons les relations

$$A_{1,1} = \sum_{i=1}^n Y_{i,1} Y_{i,1}' \stackrel{d}{=} W_k(n, \Sigma_{1,1}) \quad \text{et} \quad A_{2,2} = \sum_{i=1}^n Y_{i,2} Y_{i,2}' \stackrel{d}{=} W_{m-k}(n, \Sigma_{2,2}).$$

Pour démontrer les relations (2) et (3) du théorème 2.2.2, observons tout d'abord que l'élément différentiel de la loi de  $A \stackrel{d}{=} W_m(n, \Sigma)$  est donné par (2.1.24), soit

$$f(A)[dA] = \frac{2^{-mn/2}}{\Gamma_m(\frac{1}{2}n)} (\det \Sigma)^{-n/2} (\det A)^{\frac{1}{2}n - \frac{1}{2}(m+1)} \text{etr} \left( -\frac{1}{2}\Sigma^{-1}A \right) [dA]. \quad (2.2.5)$$

Nous allons effectuer, dans (2.2.5), le changement de variables

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} \longleftrightarrow \left( B_{1,1} = A_{1,1.2} = A_{1,1} - A_{1,2}A_{2,2}^{-1}A_{2,1}, B_{1,2} = A_{1,2}, B_{2,2} = A_{2,2} \right).$$

En posant  $B_{2,1} = B'_{1,2}$ , ce changement de variables équivaut à poser

$$\begin{aligned} A_{1,1} &= A_{1,1,2} + A_{1,2}A_{2,2}^{-1}A_{2,1} = B_{1,1} + B_{1,2}B_{2,2}^{-1}B_{2,1}, \\ A_{1,2} &= B_{1,2}, \quad A_{2,1} = B_{2,1} = B'_{1,2}, \\ A_{2,2} &= B_{2,2}. \end{aligned}$$

En tenant compte des définitions spécifiques de  $[dA]$ ,  $[dA_{1,1}]$ ,  $(dA_{1,2})$ ,  $[dA_{2,2}]$ ,  $[dB_{1,1}]$  ( $dB_{1,2}$ ) et  $[dB_{2,2}]$  dues au fait que les matrices  $A$ ,  $A_{1,1}$ ,  $A_{2,2}$ ,  $B_{1,1}$  et  $B_{2,2}$  sont symétriques, on constate que

$$[dA] = [dA_{1,1}] \wedge (dA_{1,2}) \wedge [dA_{2,2}] = [dB_{1,1}] \wedge (dB_{1,2}) \wedge [dB_{2,2}]. \quad (2.2.6)$$

Pour évaluer  $\det A$  et  $\det \Sigma$  dans (2.2.5), on utilise l'identité (2.6.10) de la proposition 2.6.2 pour écrire que

$$\begin{aligned} \det(A) &= \det \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} = \det(A_{2,2}) \times \det\{A_{1,1} - A_{1,2}A_{2,2}^{-1}A_{2,1}\} \\ &= \det(A_{2,2}) \times \det(A_{1,1,2}) = \det(B_{2,2}) \times \det(B_{1,1}), \end{aligned} \quad (2.2.7)$$

et

$$\begin{aligned} \det(\Sigma) &= \det \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix} = \det(\Sigma_{2,2}) \times \det\{\Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1}\} \\ &= \det(\Sigma_{2,2}) \times \det(\Sigma_{1,1,2}). \end{aligned} \quad (2.2.8)$$

Dans le but d'obtenir une expression simple de  $\text{etr}(-\frac{1}{2}\Sigma^{-1}A)$  en fonction de  $B_{1,1}$ ,  $B_{1,2}$  et  $B_{2,2}$ , posons

$$\begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix}^{-1} = \begin{bmatrix} \Sigma^{1,1} & \Sigma^{1,2} \\ \Sigma^{2,1} & \Sigma^{2,2} \end{bmatrix}, \quad (2.2.9)$$

pour écrire, compte tenu de  $B_{2,1} = B'_{1,2}$ ,

$$\begin{aligned} \text{tr}(\Sigma^{-1}A) &= \text{tr} \begin{bmatrix} \Sigma^{1,1} & \Sigma^{1,2} \\ \Sigma^{2,1} & \Sigma^{2,2} \end{bmatrix} \begin{bmatrix} B_{1,1} + B_{1,2}B_{2,2}^{-1}B_{2,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{bmatrix} \\ &= \text{tr}(\Sigma^{1,1}B_{1,1}) + \text{tr}(\Sigma^{1,1}B_{1,2}B_{2,2}^{-1}B_{2,1}) \\ &\quad + 2\text{tr}(\Sigma^{1,2}B_{2,1}) + \text{tr}(\Sigma^{2,2}B_{2,2}) \\ &= \text{tr}(\Sigma^{1,1}B_{1,1}) + \text{tr}(\{\Sigma^{2,2} - \Sigma^{2,1}(\Sigma^{1,1})^{-1}\Sigma^{1,2}\}B_{2,2}) \\ &\quad + \text{tr}(\Sigma^{1,1}\{B_{1,2} + (\Sigma^{1,1})^{-1}\Sigma^{1,2}B_{2,2}\}B_{2,2}^{-1}\{B_{1,2} + (\Sigma^{1,1})^{-1}\Sigma^{1,2}B_{2,2}\}'). \end{aligned} \quad (2.2.10)$$

Pour évaluer cette expression en fonction des  $\Sigma_{i,j}$ ,  $1 \leq i, j \leq 2$ , on utilise les identités (2.6.3) et (2.6.4) de la proposition 1.2.1. Celles-ci permettent d'écrire les relations suivantes liant les  $\Sigma^{i,j}$  aux  $\Sigma_{i,j}$  pour  $1 \leq i, j \leq 2$ . Nous avons les égalités

$$\Sigma^{1,1} = \{\Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1}\}^{-1} = \Sigma_{1,1,2}^{-1}, \quad (\Sigma^{1,1})^{-1}\Sigma^{1,2} = -\Sigma_{1,2}\Sigma_{2,2}^{-1}, \quad (2.2.11)$$

et

$$\Sigma_{2,2}^{-1} = \Sigma^{2,2} - \Sigma^{2,1}(\Sigma^{1,1})^{-1}\Sigma^{1,2}. \quad (2.2.12)$$

En combinant (2.2.10) avec (2.2.11) et (2.2.12) on obtient que

$$\begin{aligned} \text{tr}(\Sigma^{-1}A) &= \text{tr}(\Sigma_{1,1,2}^{-1}B_{1,1}) + \text{tr}(\Sigma_{2,2}^{-1}B_{2,2}) \\ &\quad + \text{tr}(\Sigma_{1,1,2}^{-1}\{B_{1,2} - \Sigma_{1,2}\Sigma_{2,2}^{-1}B_{2,2}\}B_{2,2}^{-1}\{B_{1,2} - \Sigma_{1,2}\Sigma_{2,2}^{-1}B_{2,2}\}') \\ &= \text{tr}(\Sigma_{1,1,2}^{-1}A_{1,1,1}) + \text{tr}(\Sigma_{2,2}^{-1}A_{2,2}) \\ &\quad + \text{tr}(\Sigma_{1,1,2}^{-1}\{A_{1,2} - \Sigma_{1,2}\Sigma_{2,2}^{-1}A_{2,2}\}A_{2,2}^{-1}\{A_{1,2} - \Sigma_{1,2}\Sigma_{2,2}^{-1}A_{2,2}\}'). \end{aligned} \quad (2.2.13)$$



En combinant les formules (2.2.5), (2.2.6), (2.2.7), (2.2.8) et (2.2.13) on constate que la densité jointe de  $B_{1,1} = A_{1,1,2}$ ,  $B_{1,2} = A_{1,2}$  et  $B_{2,2} = A_{2,2}$  est donnée par

$$\begin{aligned}
& f(A_{1,1,2}, A_{1,2}, A_{2,2}) \\
&= \frac{2^{-mn/2}}{\Gamma_m(\frac{1}{2}n)} (\det \Sigma)^{-n/2} \operatorname{etr} \left( -\frac{1}{2} \Sigma^{-1} A \right) (\det A)^{(n-m-1)/2} (dA_{1,1,2})(dA_{1,2})(dA_{2,2}) \\
&= \frac{2^{-mn/2}}{\Gamma_m(\frac{1}{2}n)} \left\{ \frac{\det(A_{2,2})^{(n-m-1)/2} \times \det(A_{1,1,2})^{(n-m-1)/2}}{\det(\Sigma_{2,2})^{n/2} \times \det(\Sigma_{1,1,2})^{n/2}} \right\} \operatorname{etr} \left( -\frac{1}{2} \Sigma_{1,1,2}^{-1} A_{1,1,1} \right) \\
&\quad \times \operatorname{etr} \left( -\frac{1}{2} \Sigma_{2,2}^{-1} A_{2,2} \right) \times \operatorname{etr} \left( -\frac{1}{2} \Sigma_{1,1,2}^{-1} \{A_{1,2} - \Sigma_{1,2} \Sigma_{2,2}^{-1} A_{2,2}\} A_{2,2}^{-1} \{A_{1,2} - \Sigma_{1,2} \Sigma_{2,2}^{-1} A_{2,2}\}' \right) \\
&= \left\{ \frac{2^{-k(n-m+k)/2}}{\Gamma_k(\frac{1}{2}(n-m+k))} \frac{(\det A_{1,1,2})^{(n-m+k-k-1)/2}}{(\det \Sigma_{1,1,2})^{(n-m+k)/2}} \operatorname{etr} \left( -\frac{1}{2} \Sigma_{1,1,2}^{-1} A_{1,1,1} \right) \right\} \\
&\quad \times \left\{ \frac{2^{-(m-k)n/2}}{\Gamma_{m-k}(\frac{1}{2}n)} \frac{(\det A_{2,2})^{(n-m+k-1)/2}}{(\det \Sigma_{2,2})^{n/2}} \operatorname{etr} \left( -\frac{1}{2} \Sigma_{2,2}^{-1} A_{2,2} \right) \right\} \\
&\quad \times \left\{ \frac{\Gamma_k(\frac{1}{2}(n-m+k)) \Gamma_{m-k}(\frac{1}{2}n)}{\Gamma_m(\frac{1}{2}n)} \right\} \frac{2^{-k(m-k)/2}}{(\det \Sigma_{1,1,2})^{(m-k)/2} (\det A_{2,2})^{k/2}} \\
&\quad \times \operatorname{etr} \left( -\frac{1}{2} \Sigma_{1,1,2}^{-1} \{A_{1,2} - \Sigma_{1,2} \Sigma_{2,2}^{-1} A_{2,2}\} A_{2,2}^{-1} \{A_{1,2} - \Sigma_{1,2} \Sigma_{2,2}^{-1} A_{2,2}\}' \right). \tag{2.2.14}
\end{aligned}$$

On peut simplifier les derniers facteurs de (2.2.14) en observant que

$$\begin{aligned}
\Gamma_m \left( \frac{1}{2}n \right) &= \pi^{m(m-1)/4} \prod_{i=1}^m \Gamma \left( \frac{1}{2}(n-i+1) \right) \\
&= \pi^{k(k-1)/4} \prod_{i=1}^k \Gamma \left( \frac{1}{2}(n-m+k-i+1) \right) \\
&\quad \times \pi^{(m-k)(m-k-1)/4} \prod_{i=1}^{m-k} \Gamma \left( \frac{1}{2}(n-i+1) \right) \times \pi^{k(m-k)/2} \\
&= \pi^{k(m-k)/2} \Gamma_k \left( \frac{1}{2}(n-m+k) \right) \Gamma_{m-k} \left( \frac{1}{2}n \right). \tag{2.2.15}
\end{aligned}$$

On en déduit l'expression de  $f(A_{1,1,2}, A_{1,2}, A_{2,2}) = f(B_{1,1}, B_{1,2}, B_{2,2})$ , donnée par

$$\begin{aligned}
& f(A_{1,1,2}, A_{1,2}, A_{2,2}) \\
&= \left\{ \frac{2^{-k(n-m+k)/2}}{\Gamma_k(\frac{1}{2}(n-m+k))} \frac{(\det A_{1,1,2})^{(n-m+k-k-1)/2}}{(\det \Sigma_{1,1,2})^{(n-m+k)/2}} \operatorname{etr} \left( -\frac{1}{2} \Sigma_{1,1,2}^{-1} A_{1,1,1} \right) \right\} \\
&\quad \times \left\{ \frac{2^{-(m-k)n/2}}{\Gamma_{m-k}(\frac{1}{2}n)} \frac{(\det A_{2,2})^{(n-m+k-1)/2}}{(\det \Sigma_{2,2})^{n/2}} \operatorname{etr} \left( -\frac{1}{2} \Sigma_{2,2}^{-1} A_{2,2} \right) \right\} \\
&\quad \times \frac{\operatorname{etr} \left( -\frac{1}{2} \Sigma_{1,1,2}^{-1} \{A_{1,2} - \Sigma_{1,2} \Sigma_{2,2}^{-1} A_{2,2}\} A_{2,2}^{-1} \{A_{1,2} - \Sigma_{1,2} \Sigma_{2,2}^{-1} A_{2,2}\}' \right)}{(2\pi)^{k(m-k)/2} (\det \Sigma_{1,1,2})^{(m-k)/2} (\det A_{2,2})^{k/2}}.
\end{aligned}$$

Les deux premiers facteurs du produit ci-dessus sont les densités des lois de Wishart  $W_k(n-m+k, \Sigma_{1,1,2})$  et  $W_{m-k}(n, \Sigma_{2,2})$ . Comme l'expression est un produit, on obtient bien (2) avec l'indépendance de  $A_{1,1,2}$  et de  $(A_{1,2}, A_{2,2})$ . Comme, par le (1), on sait que  $A_{2,2} \stackrel{d}{=} W_{m-k}(n, \Sigma_{2,2})$ , le troisième facteur donne la densité conditionnelle de  $A_{1,2}$  sachant  $A_{2,2}$ , soit

$$f(A_{2,2} | A_{2,2} = a_{2,2}) = \frac{\operatorname{etr} \left( -\frac{1}{2} \Sigma_{1,1,2}^{-1} \{A_{1,2} - \Sigma_{1,2} \Sigma_{2,2}^{-1} a_{2,2}\} a_{2,2}^{-1} \{A_{1,2} - \Sigma_{1,2} \Sigma_{2,2}^{-1} a_{2,2}\}' \right)}{(2\pi)^{k(m-k)/2} (\det \Sigma_{1,1,2})^{(m-k)/2} (\det a_{2,2})^{k/2}},$$

qui, compte tenu de (3.4.14), s'identifie avec la densité de la loi normale matricielle

$$N_{k,m-k}(\Sigma_{1,2}\Sigma_{2,2}^{-1}a_{2,2}, \Sigma_{1,1\cdot 2} \otimes a_{2,2}).$$

On a donc ainsi établi (3).□

**Remarque 2.2.1.** Soit, sous les hypothèses du théorème 2.2.2,

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \stackrel{d}{=} N_m \left( \mathbb{O}_m, \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix} \right).$$

On constate alors que

$$\mathcal{L}(Y_1|Y_2 = y_2) \stackrel{d}{=} N_k(\Sigma_{1,2}\Sigma_{2,2}^{-1}y_2, \Sigma_{1,1\cdot 2}).$$

Ceci mène à l'expression de la *régression linéaire* de  $Y_1$  sachant  $Y_2$ , soit

$$R(y_2) = \mathbb{E}(Y_1|Y_2 = y_2) = \Sigma_{1,2}\Sigma_{2,2}^{-1}y_2.$$

De plus,  $Y_2$  et  $Y_1 - R(Y_2)$  sont indépendants, la loi de  $Y_1 - R(Y_2)$  étant donnée par

$$Y_1 - R(Y_2) = Y_1 - \Sigma_{1,2}\Sigma_{2,2}^{-1}Y_2 \equiv N_k(\mathbb{O}, \Sigma_{1,1\cdot 2}).$$

A partir de la matrice de variances covariances empirique  $\widehat{\Sigma} = n^{-1}A$ , qu'on partitionne en  $\widehat{\Sigma}_{i,j}$  pour  $1 \leq i, j \leq 2$  comme ci-dessus, on définit la *régression linéaire empirique* de  $Y_1$  sachant  $Y_2$  par

$$\widehat{R}(y_2) = \widehat{\mathbb{E}}(Y_1|Y_2 = y_2) = \widehat{\Sigma}_{12}\widehat{\Sigma}_{2,2}^{-1}y_2 = A_{1,2}A_{2,2}^{-1}y_2.$$

Le théorème 2.2.2 montre, en particulier, que la fonction de régression empirique  $\widehat{R}(\cdot)$  est indépendante de  $\widehat{\Sigma}_{1,1\cdot 2} = \widehat{\Sigma}_{1,1} - \widehat{\Sigma}_{1,2}\widehat{\Sigma}_{2,2}^{-1}\widehat{\Sigma}_{2,1} = n^{-1}A_{1,1\cdot 2}$ .

### 2.2.3 Loi de Wishart inversée.

Si  $A \stackrel{d}{=} W_m(n, \Sigma)$  suit une loi de Wishart (avec  $n \geq m$  et  $\Sigma > 0$ ), l'inverse  $A^{-1}$  de la matrice  $A$  a des propriétés remarquables que nous exposons dans ce qui suit.

En particulier, le théorème suivant, qui se déduit sans grande difficulté du théorème 2.2.2, sera essentiel pour la construction du test du  $T^2$  de Hotelling, étudié au §2.2.4.

**Théorème 2.2.3.** Soit  $A \stackrel{d}{=} W_m(n, \Sigma)$  avec  $\Sigma > 0$  et  $n \geq m$ , et soit  $M$ , une matrice (non aléatoire) ( $k \times m$ ) de rang  $k = \text{rg}(M)$ . Alors,

$$(MA^{-1}M')^{-1} \stackrel{d}{=} W_k(n - m + k, (M\Sigma^{-1}M')^{-1}). \quad (2.2.16)$$

**Preuve.** Dans une première étape, nous allons ramener la démonstration de (2.2.16) au cas particulier où  $\Sigma = \mathbb{I}_m$  est la matrice identité. Pour cela, nous effectuons le changement de variables  $B = \Sigma^{-1/2}A\Sigma^{-1/2} \stackrel{d}{=} W_m(n, \mathbb{I}_m)$  et  $Q = M\Sigma^{-1/2}$ . Nous en déduisons les égalités

$$(MA^{-1}M')^{-1} = (Q\Sigma^{1/2}A^{-1}\Sigma^{1/2}Q')^{-1} = (QB^{-1}Q')^{-1},$$

et

$$(M\Sigma^{-1}M')^{-1} = (Q\Sigma^{1/2}\Sigma^{-1}\Sigma^{1/2}Q')^{-1} = (QQ')^{-1}.$$

En posant  $B = \Sigma^{-1/2}A\Sigma^{-1/2} \stackrel{d}{=} W_m(n, \mathbb{I}_m)$ , ces relations justifient l'équivalence cherchée, soit

$$(MA^{-1}M')^{-1} \stackrel{d}{=} W_k(n - m + k, (M\Sigma^{-1}M')^{-1}) \iff (QB^{-1}Q')^{-1} \stackrel{d}{=} W_k(n - m + k, (QQ')^{-1}).$$

Compte tenu de ce qui précède, dans une deuxième étape, nous établissons la propriété (2.2.16) dans le cas où  $\Sigma = \mathbb{I}_m$ . En conséquence, nous supposons, désormais, que  $A \stackrel{d}{=} W_m(n, \mathbb{I}_m)$ . Nous faisons usage de l'hypothèse que la matrice  $M$ , de dimensions  $(k \times m)$  est de rang  $\text{rg}(M) = k$ , pour la factoriser en

$$M = C \begin{bmatrix} \mathbb{I}_k & \mathbb{O}_{k, m-k} \end{bmatrix} H, \quad (2.2.17)$$

où  $C$  désigne une matrice régulière  $(k \times k)$ , et  $H$ , une matrice orthogonale  $(m \times m)$ . Une telle factorisation est toujours possible. En effet, le fait que la matrice

$$M = \begin{bmatrix} w'_1 \\ \vdots \\ w'_k \end{bmatrix},$$

de dimensions  $(k \times m)$ , soit de rang  $k$  impose que ses vecteurs lignes  $w_1, \dots, w_k$  soient linéairement indépendants dans  $\mathbb{R}^m$ . Soit  $h_1, \dots, h_k$  une base orthonormée du sous-espace vectoriel engendré par  $w_1, \dots, w_k$  dans  $\mathbb{R}^m$ . Désignons, de même, par  $h_{k+1}, \dots, h_m$  des vecteurs orthonormés de  $\mathbb{R}^m$ , tels que la matrice

$$H' = [ h_1 \quad \dots \quad h_k \quad h_{k+1} \quad \dots \quad h_m ],$$

soit  $(m \times m)$  orthogonale. Il existe alors une matrice régulière  $P$ , de dimensions  $(k \times k)$ , telle que

$$M' = [ w_1 \quad \dots \quad w_k ] = [ h_1 \quad \dots \quad h_k ] P = H' \begin{bmatrix} \mathbb{I}_k & \\ & \mathbb{O}_{m-k, k} \end{bmatrix},$$

ce qui s'écrit, sous forme équivalente,

$$M = \begin{bmatrix} w'_1 \\ \vdots \\ w'_k \end{bmatrix} = P' \begin{bmatrix} h'_1 \\ \vdots \\ h'_k \end{bmatrix} = P' \begin{bmatrix} \mathbb{I}_k & \mathbb{O}_{k, m-k} \end{bmatrix} H,$$

ce qui donne (2.2.17), en posant  $C = P'$ . Nous écrivons, ensuite, compte tenu de (2.2.17), que

$$\begin{aligned} (MM')^{-1} &= \left( C \begin{bmatrix} \mathbb{I}_k & \mathbb{O}_{k, m-k} \end{bmatrix} H H' \begin{bmatrix} \mathbb{I}_k & \\ & \mathbb{O}_{m-k, k} \end{bmatrix} C' \right)^{-1} \\ &= (C')^{-1} \left( \begin{bmatrix} \mathbb{I}_k & \mathbb{O}_{k, m-k} \end{bmatrix} \begin{bmatrix} \mathbb{I}_k & \\ & \mathbb{O}_{m-k, k} \end{bmatrix} \right)^{-1} C^{-1} \\ &= (C')^{-1} C^{-1}. \end{aligned}$$

Par ailleurs, le fait que  $H$  soit orthogonale implique que la matrice  $D$  définie par  $D = H A H'$  est telle que  $D = H A H' \stackrel{d}{=} W_m(n, \mathbb{I}_m)$ . Décomposons cette matrice  $D$  en

$$D = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \quad \text{et posons} \quad D^{-1} = \begin{bmatrix} D^{11} & D^{12} \\ D^{21} & D^{22} \end{bmatrix}.$$

Compte tenu de (2.6.3) et (2.6.4), le théorème 2.2.2 montre que

$$(D^{11})^{-1} = D_{11} - D_{12} D_{22}^{-1} D_{21} = D_{11.2} \equiv W_k(n - m + k, \mathbb{I}_k).$$

On en déduit que

$$\begin{aligned} (M A^{-1} M')^{-1} &= \left( C \begin{bmatrix} \mathbb{I}_k & \mathbb{O}_{k, m-k} \end{bmatrix} H A^{-1} H' \begin{bmatrix} \mathbb{I}_k & \\ & \mathbb{O}_{m-k, k} \end{bmatrix} C' \right)^{-1} \\ &= (C')^{-1} \left( \begin{bmatrix} \mathbb{I}_k & \mathbb{O}_{k, m-k} \end{bmatrix} D^{-1} \begin{bmatrix} \mathbb{I}_k & \\ & \mathbb{O}_{m-k, k} \end{bmatrix} \right)^{-1} C^{-1} \\ &= (C')^{-1} (D^{11})^{-1} C^{-1} \\ &\stackrel{d}{=} W_k(n - m + k, (C')^{-1} C^{-1}) = W_k(n - m + k, (MM')^{-1}), \end{aligned}$$

ce qu'il fallait démontrer.  $\square$

**Théorème 2.2.4.** Soit  $A \stackrel{d}{=} W_m(n, \Sigma)$ , où  $\Sigma > 0$  et  $n \geq m$ . Soit  $Y \in \mathbb{R}^m$  un vecteur aléatoire indépendant de  $A$ , et tel que  $\mathbb{P}(Y = \mathbb{O}_m) = 0$ . Alors la statistique  $(Y'\Sigma^{-1}Y)/(Y'A^{-1}Y)$  est indépendante de  $Y$  et de loi donnée par

$$\frac{Y'\Sigma^{-1}Y}{Y'A^{-1}Y} \stackrel{d}{=} \chi_{n-m+1}^2. \quad (2.2.18)$$

**Preuve.** On pose  $M = y'$  et  $k = 1$  dans (2.2.16) pour obtenir que, si  $y \neq 0$ ,

$$\mathcal{L}\left((Y'A^{-1}Y)^{-1} \mid Y = y\right) = \mathcal{L}\left((y'A^{-1}y)^{-1}\right) = W_1\left(n - m - 1, (y'\Sigma^{-1}y)^{-1}\right),$$

ce qui équivaut à écrire que

$$\mathcal{L}\left(\frac{Y'\Sigma^{-1}Y}{Y'A^{-1}Y} \mid Y = y\right) = \mathcal{L}\left(\frac{y'\Sigma^{-1}y}{y'A^{-1}y}\right) = \chi_{n-m+1}^2.$$

La conclusion est alors évidente.  $\square$

## 2.2.4 La statistique du $T^2$ de Hotelling.

La statistique du  $T^2$  de Hotelling (Hotelling, H. (1931). The generalization of Student's ratio. *Ann. Math. Statist.* **2** 360–378) apparaît comme la généralisation naturelle de la statistique univariée de Student. Sa forme de base est décrite par le théorème suivant.

**Théorème 2.2.5.** Supposons que  $\mathbf{X} \stackrel{d}{=} N_m(\mu, \Sigma)$  et  $A = nS \stackrel{d}{=} W_m(n, \Sigma)$  soient indépendants, et que les conditions  $n \geq m$  et  $\Sigma > 0$  soient satisfaites. Alors, la statistique  $T^2 = \mathbf{X}'S^{-1}\mathbf{X}$  a sa loi donnée par

$$\left\{\frac{n - m + 1}{mn}\right\} T^2 \stackrel{d}{=} F_{m, n-m+1}(\mu'\Sigma^{-1}\mu). \quad (2.2.19)$$

**Preuve.** On écrit

$$\frac{T^2}{n} = \mathbf{x}'A^{-1}\mathbf{x} = \left\{\frac{\mathbf{x}'\Sigma^{-1}\mathbf{x}}{\mathbf{x}'A^{-1}\mathbf{x}}\right\}^{-1} \mathbf{x}'\Sigma^{-1}\mathbf{x}.$$

Or, par (2.2.18), le théorème 2.2.4 montre que

$$\frac{\mathbf{x}'\Sigma^{-1}\mathbf{x}}{\mathbf{x}'A^{-1}\mathbf{x}} \stackrel{d}{=} \chi_{n-m+1}^2$$

est indépendant de  $\mathbf{x}'\Sigma^{-1}\mathbf{x} \stackrel{d}{=} \chi_m^2(\mu'\Sigma^{-1}\mu)$ . Avec des notations évidentes, on en déduit que

$$\frac{T^2}{n} = \left\{\frac{\mathbf{x}'\Sigma^{-1}\mathbf{x}}{\mathbf{x}'A^{-1}\mathbf{x}}\right\}^{-1} \mathbf{x}'\Sigma^{-1}\mathbf{x} \stackrel{d}{=} \frac{\chi_m^2(\mu'\Sigma^{-1}\mu)}{\chi_{n-m+1}^2},$$

ce qui implique que

$$\frac{T^2}{n} \times \left\{\frac{1/m}{1/(n-m+1)}\right\} = \frac{T^2}{n} \times \frac{n-m+1}{m} \equiv F_{m, n-m+1}(\mu'\Sigma^{-1}\mu),$$

et achève la démonstration de (2.2.19).  $\square$

Nous allons maintenant établir que, sous les hypothèses et notations du théorème 2.2.5, le test du rapport de vraisemblance de l'hypothèse

$$(H.0) \mu = \mathbb{O};$$

contre l'alternative

$$(H.1) \mu \in \mathbb{R}^m \text{ est quelconque};$$

est une fonction monotone simple de  $T^2$ . Rappelons tout d'abord, par (2.1.24), que si  $\mathbf{X} \stackrel{d}{=} N_m(\mu, \Sigma)$  et  $A = nS \stackrel{d}{=} W_m(n, \Sigma)$  soient indépendants, la vraisemblance de  $\mathbf{X}$  et  $A$  est donnée, en posant  $N = n + 1$ , par

$$\begin{aligned} L(\mu, \Sigma) &= (2\pi)^{-m/2} (\det \Sigma)^{-1/2} \text{etr} \left( -\frac{1}{2} (\mathbf{X} - \mu)' \Sigma^{-1} (\mathbf{X} - \mu) \right) \\ &\quad \times \frac{2^{-mn/2}}{\Gamma_m(\frac{1}{2}n)} (\det \Sigma)^{-n/2} \text{etr} \left( -\frac{1}{2} \Sigma^{-1} A \right) (\det A)^{(n-m-1)/2} \\ &= \mathcal{C} (\det \Sigma)^{-N/2} \text{etr} \left( -\frac{1}{2} \Sigma^{-1} A \right) \exp \left( -\frac{1}{2} (\mathbf{X} - \mu)' \Sigma^{-1} (\mathbf{X} - \mu) \right) \\ &= \mathcal{C} (\det \Sigma)^{-N/2} \text{etr} \left( -\frac{1}{2} \Sigma^{-1} A \right) \text{etr} \left( -\frac{1}{2} \Sigma^{-1} (\mathbf{X} - \mu) (\mathbf{X} - \mu)' \right), \end{aligned}$$

où  $\mathcal{C}$  ne dépend pas de  $\mu$  et  $\Sigma$ .

**Théorème 2.2.6.** *Supposons que  $\mathbf{X} \stackrel{d}{=} N_m(\mu, \Sigma)$  et  $A = nS \stackrel{d}{=} W_m(n, \Sigma)$  soient indépendants, et que les conditions  $n \geq m$  et  $\Sigma > 0$  soient satisfaites. Alors, la statistique  $T^2 = \mathbf{X}' S^{-1} \mathbf{X}$  est telle que, pour  $N = n + 1$ ,*

$$\frac{\sup_{\Sigma > 0} L(\mathbb{O}, \Sigma)}{\sup_{\mu \in \mathbb{R}^m, \Sigma > 0} L(\mu, \Sigma)} = \left\{ \frac{1}{1 + T^2/n} \right\}^{N/2}. \quad (2.2.20)$$

**Preuve.** On constate tout d'abord aisément que

$$L(\mu, \Sigma) \leq L(\mathbf{X}, \Sigma) = \mathcal{C} (\det \Sigma)^{-N/2} \text{etr} \left( -\frac{1}{2} \Sigma^{-1} A \right),$$

avec égalité lorsque  $\mu = \mathbf{X}$ . On constate ensuite, par (2.1.16), que

$$L(\mathbf{X}, \Sigma) \leq L(\mathbf{X}, \widehat{\Sigma}) = \mathcal{C} N^{mN/2} e^{-mN/2} (\det A)^{-N/2},$$

avec égalité lorsque  $\Sigma = \widehat{\Sigma} = (1/N)A$ . Ceci suffit à montrer que

$$\sup_{\mu \in \mathbb{R}^m, \Sigma > 0} L(\mu, \Sigma) = \mathcal{C} N^{mN/2} e^{-mN/2} (\det A)^{-N/2}. \quad (2.2.21)$$

De même, on voit que

$$\begin{aligned} L(\mathbb{O}, \Sigma) &= \mathcal{C} (\det \Sigma)^{-N/2} \text{etr} \left( -\frac{1}{2} \Sigma^{-1} \left\{ A + (\mathbf{X} - \mu) (\mathbf{X} - \mu)' \right\} \right) \\ &= \mathcal{C} (\det \Sigma)^{-N/2} \text{etr} \left( -\frac{1}{2} \Sigma^{-1} A^* \right), \end{aligned}$$

où  $A^* = A + \mathbf{X}\mathbf{X}'$ . Par application de (2.1.16), on constate de même que

$$L(\mathbb{O}, \Sigma) \leq L(\mathbb{O}, \widetilde{\Sigma}) = \mathcal{C} N^{mN/2} e^{-mN/2} (\det A^*)^{-N/2},$$

avec égalité lorsque  $\Sigma = \widetilde{\Sigma} = (1/N)A^*$ . Ceci suffit à montrer que

$$\sup_{\Sigma > 0} L(\mathbb{O}, \Sigma) = \mathcal{C} N^{mN/2} e^{-mN/2} (\det A^*)^{-N/2}. \quad (2.2.22)$$

On déduit de (2.2.21) et (2.2.22) que

$$\begin{aligned} \frac{\sup_{\Sigma > 0} L(\mathbb{O}, \Sigma)}{\sup_{\mu \in \mathbb{R}^m, \Sigma > 0} L(\mu, \Sigma)} &= \left\{ \frac{\det A}{\det \{A + \mathbf{X}\mathbf{X}'\}} \right\}^{N/2} \\ &= \left\{ \frac{\det \mathbb{I}}{\det \{\mathbb{I} + A^{-1} \mathbf{X}\mathbf{X}'\}} \right\}^{N/2} \\ &= \left\{ \frac{1}{1 + \mathbf{X}' A^{-1} \mathbf{X}} \right\}^{N/2}, \end{aligned}$$

où nous avons fait usage de (2.6.11). Comme  $S = n^{-1}A$  et  $T^2 = \mathbf{X}'S^{-1}\mathbf{X} = n\mathbf{X}'A^{-1}\mathbf{X}$ , la conclusion (2.2.20) est directe.  $\square$

### 2.3 Le critère de Wilks et ses applications.

Le critère  $\Lambda$  de Wilks (Wilks, S.S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, **24**, 471-494) joue, pour des données vectorielles, le même rôle que la loi de Fisher, pour des observations à valeurs réelles. Avant de décrire dans le détail ce critère et son utilisation, nous allons introduire la *loi Béta multivariée*, à l'aide du théorème suivant.

**Théorème 2.3.1.** Soient  $A \stackrel{d}{=} W_m(n_1, \Sigma)$ , et  $B \stackrel{d}{=} W_m(n_2, \Sigma)$ , deux matrices de Wishart indépendantes, telles que  $n_1 \geq m$ ,  $n_2 \geq m$ , et  $\Sigma > 0$ . On pose  $A + B = T'T$ , où  $T$  est une matrice  $m \times m$ , triangulaire supérieure à éléments diagonaux positifs, et on définit une matrice symétrique ( $m \times m$ ),  $U$ , par  $A = T'UT$ . Alors,  $T$  et  $U$  sont indépendantes, et la densité de  $U$  est donnée par

$$\frac{\Gamma_m(\frac{1}{2}(n_1 + n_2))}{\Gamma_m(\frac{1}{2}n_1)\Gamma_m(\frac{1}{2}n_2)} (\det U)^{(n_1 - m - 1)/2} (\det(\mathbb{I}_m - U))^{(n_2 - m - 1)/2} \quad \text{pour } U > 0 \text{ et } \mathbb{I}_m - U > 0. \quad (2.3.1)$$

**Preuve.** L'élément différentiel donnant la densité jointe de  $A$  et  $B$  est égal à

$$\frac{1}{\Gamma_m(\frac{1}{2}n_1)\Gamma_m(\frac{1}{2}n_2)} 2^{-m(n_1+n_2)/2} (\det \Sigma)^{-(n_1+n_2)/2} \text{etr}\left(-\frac{1}{2}\Sigma^{-1}(A+B)\right) \\ \times (\det A)^{(n_1 - m - 1)/2} (\det B)^{(n_2 - m - 1)/2} [dA] \wedge [dB].$$

On effectue tout d'abord le changement de variables  $(A, B) \rightarrow (A, C)$ , où  $C = A + B$ . Comme, de toute évidence,  $[dA] \wedge [dB] = [dA] \wedge [dC]$ , l'élément différentiel donnant la densité jointe de  $A$  et  $C$  est égal à

$$\frac{1}{\Gamma_m(\frac{1}{2}n_1)\Gamma_m(\frac{1}{2}n_2)} 2^{-m(n_1+n_2)/2} (\det \Sigma)^{-(n_1+n_2)/2} \text{etr}\left(-\frac{1}{2}\Sigma^{-1}C\right) \\ \times (\det A)^{(n_1 - m - 1)/2} (\det(C - A))^{(n_2 - m - 1)/2} [dA] \wedge [dC]. \quad (2.3.2)$$

On effectue ensuite le changement de variables  $(A, C) \rightarrow (U, T)$ , où  $C = T'T$ ,  $T = (t_{i,j})$  est triangulaire supérieure à éléments diagonaux positifs, et  $A = T'UT$ . Par (1.1.6),

$$\begin{aligned} [dA] \wedge [dC] &= (T'dUT) \wedge (d(T'T)) \\ &= (\det T)^{m+1} (dU) \wedge (d(T'T)) \\ &= (\det T'T)^{(m+1)/2} (dU) \wedge (d(T'T)). \end{aligned}$$

Comme, de plus,

$$\det A = (\det T'T) \det U, \quad \text{et} \quad \det(C - A) = (\det T'T) \det(\mathbb{I}_m - U),$$

on en déduit que l'élément différentiel donnant la densité jointe de  $U$  et  $C = T'T$  est égal à

$$\frac{2^{-m(n_1+n_2)/2} (\det \Sigma)^{-(n_1+n_2)/2}}{\Gamma_m(\frac{1}{2}(n_1 + n_2))} (\det C)^{(n_1+n_2-m-1)/2} \text{etr}\left(-\frac{1}{2}\Sigma^{-1}C\right) (dC) \\ \times \frac{\Gamma_m(\frac{1}{2}(n_1 + n_2))}{\Gamma_m(\frac{1}{2}n_1)\Gamma_m(\frac{1}{2}n_2)} (\det U)^{(n_1 - m - 1)/2} (\det(\mathbb{I}_m - U))^{(n_2 - m - 1)/2} (dU).$$

On reconnaît dans le premier facteur la densité de la loi de Wishart de

$$C = A + B \equiv W_m(n_1 + n_2, \Sigma),$$

(voir (2.1.24)), ce qui permet de conclure. En passant, nous avons établi la généralisation suivante de l'identité d'Euler (obtenue pour  $m = 1$ ), pour  $n_1 \geq m$  et  $n_2 \geq m$ ,

$$\begin{aligned} \beta_m(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2) &= \int_{0 < U < \mathbb{I}_m} (\det U)^{(n_1 - m - 1)/2} (\det(\mathbb{I}_m - U))^{(n_2 - m - 1)/2} (dU) \\ &= \frac{\Gamma_m(\tfrac{1}{2}n_1)\Gamma_m(\tfrac{1}{2}n_2)}{\Gamma_m(\tfrac{1}{2}(n_1 + n_2))}, \end{aligned} \quad (2.3.3)$$

où la notation  $0 < U < \mathbb{I}_m$  signifie que les matrices  $U$  et  $\mathbb{I}_m - U$  sont définies positives.  $\square$

**Définition 2.3.1.** Une matrice symétrique  $(m \times m)$  aléatoire  $U$  est dite suivre une loi bêta multivariée de paramètres  $r > \frac{1}{2}(m - 1)$  et  $s > \frac{1}{2}(m - 1)$ , ce qui sera noté symboliquement par  $U \equiv \beta_m(r, s)$ , si elle a une densité donnée par

$$\frac{\Gamma_m(r + s)}{\Gamma_m(r)\Gamma_m(s)} (\det U)^{r - (m - 1)/2} (\det(\mathbb{I}_m - U))^{s - (m - 1)/2} \quad \text{pour } 0 < U < \mathbb{I}_m. \quad (2.3.4)$$

Le théorème suivant est dû à Kshirsagar A.M. (1961). The noncentral multivariate beta distribution. *Ann. Math. Statist.* **32** 104-111.

**Théorème 2.3.2.** Soient  $n_1$  et  $n_2$  deux entiers tels que  $n_1 \geq m$  et  $n_2 \geq m$ . Si  $U \equiv \beta_m(\frac{1}{2}n_1, \frac{1}{2}n_2)$  est factorisée sous la forme  $U = T'T$ , où  $T = [t_{ij}]$  est une matrice triangulaire supérieure à éléments diagonaux positifs, alors les  $t_{11}, \dots, t_{mm}$  sont indépendants et  $t_{ii}^2 \equiv \beta(\frac{1}{2}(n_1 - i + 1), \frac{1}{2}n_2)$  pour  $i = 1, \dots, m$ .

**Preuve.** On effectue le changement de variable  $U = T'T$ , avec  $T$  matrice triangulaire supérieure à diagonale positive, ce qui donne

$$\det U = \det T'T = \prod_{i=1}^m t_{ii}^2,$$

et, par (2.1.25),

$$(dU) = 2^m \prod_{i=1}^m t_{ii}^{m-i+1} \bigwedge_{1 \leq i \leq j \leq m} dt_{ij}.$$

L'élément différentiel donnant la densité de  $T$  est donc égal à

$$f(T; m, n_1, n_2)(dT) = \frac{\Gamma_m(\frac{1}{2}(n_1 + n_2))}{\Gamma_m(\frac{1}{2}n_1)\Gamma_m(\frac{1}{2}n_2)} 2^m \prod_{i=1}^m t_{ii}^{n_1 - i} (\det(\mathbb{I}_m - T'T))^{(n_2 - m - 1)/2} (dT). \quad (2.3.5)$$

Ecrivons maintenant  $T$  sous la forme d'une matrice partitionnée

$$T = \begin{bmatrix} t_{11} & \tau' \\ \mathbb{O} & T_{22} \end{bmatrix},$$

ce qui donne

$$T'T = \begin{bmatrix} t_{11} & \mathbb{O} \\ \tau & T'_{22} \end{bmatrix} \begin{bmatrix} t_{11} & \tau' \\ \mathbb{O} & T_{22} \end{bmatrix} = \begin{bmatrix} t_{11}^2 & t_{11}\tau' \\ t_{11}\tau & \tau\tau' + T'_{22}T_{22} \end{bmatrix}$$

où  $T_{22}$  est  $((m - 1) \times (m - 1))$ . On utilise alors la suite d'égalités suivante. On a

$$\begin{aligned} \det(\mathbb{I}_m - T'T) &= \det \begin{bmatrix} 1 - t_{11}^2 & -t_{11}\tau' \\ -t_{11}\tau & \mathbb{I}_{m-1} - \tau\tau' - T'_{22}T_{22} \end{bmatrix} \\ &= \det \left\{ \begin{bmatrix} 1 & \mathbb{O} \\ 0 & \mathbb{I}_{m-1} - T'_{22}T_{22} \end{bmatrix} - \begin{bmatrix} t_{11} \\ \tau \end{bmatrix} \begin{bmatrix} t_{11} \\ \tau \end{bmatrix}' \right\} \end{aligned}$$

$$\begin{aligned}
&= \det \begin{bmatrix} 1 & 0 \\ 0 & \mathbb{I}_{m-1} - T'_{22}T_{22} \end{bmatrix} \\
&\quad \times \det \left\{ \mathbb{I}_m - \begin{bmatrix} 1 & 0 \\ 0 & (\mathbb{I}_{m-1} - T'_{22}T_{22})^{-1/2} \end{bmatrix} \begin{bmatrix} t_{11} \\ \tau \end{bmatrix} \begin{bmatrix} t_{11} \\ \tau \end{bmatrix}' \begin{bmatrix} 1 & 0 \\ 0 & (\mathbb{I}_{m-1} - T'_{22}T_{22})^{-1/2} \end{bmatrix} \right\} \\
&= \det(\mathbb{I}_{m-1} - T'_{22}T_{22}) \det \left\{ \mathbb{I}_m - \begin{bmatrix} t_{11} \\ (\mathbb{I}_{m-1} - T'_{22}T_{22})^{-1/2}\tau \end{bmatrix} \begin{bmatrix} t_{11} \\ (\mathbb{I}_{m-1} - T'_{22}T_{22})^{-1/2}\tau \end{bmatrix}' \right\}.
\end{aligned}$$

Pour évaluer ce dernier déterminant, on utilise le fait que, si  $H$  est une matrice orthogonale telle que

$$Hu = \begin{bmatrix} (u'u)^{1/2} \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

alors

$$\mathbb{I}_m - uu' = H(\mathbb{I}_m - uu')H' = \begin{bmatrix} 1 - u'u & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{bmatrix} \Rightarrow \det(\mathbb{I}_m - uu') = 1 - u'u.$$

On applique ces formules au choix de  $u$  défini par

$$u = \begin{bmatrix} t_{11} \\ (\mathbb{I}_{m-1} - T'_{22}T_{22})^{-1/2}\tau \end{bmatrix}, \quad u'u = t_{11}^2 + \tau'(\mathbb{I}_{m-1} - T'_{22}T_{22})\tau,$$

et on obtient donc, à partir des expressions précédentes, que

$$\begin{aligned}
\det(\mathbb{I}_m - T'T) &= \det(\mathbb{I}_{m-1} - T'_{22}T_2) \left\{ 1 - t_{11}^2 - \tau'(\mathbb{I}_{m-1} - T'_2T_2)\tau \right\} \\
&= \det(\mathbb{I}_{m-1} - T'_{22}T_{22}) (1 - t_{11}^2) \left\{ 1 - \frac{\tau'(\mathbb{I}_{m-1} - T'_{22}T_{22})^{-1}\tau}{1 - t_{11}^2} \right\} \\
&= \det(\mathbb{I}_{m-1} - T'_{22}T_2) (1 - t_{11}^2) (1 - v'v).
\end{aligned} \tag{2.3.6}$$

Nous avons effectué ici le changement de variables  $(t_{11}, \tau, T_{22}) \rightarrow (t_{11}, v, T_{22})$ , où  $v$  est défini par

$$v = (1 - t_{11})^{-1/2}(\mathbb{I}_{m-1} - T'_{22}T_{22})^{-1/2}\tau.$$

Par (1.1.4), on constate que

$$\begin{aligned}
(dT) &= \bigwedge_{1 \leq i \leq j \leq m} dt_{ij} = dt_{11} \wedge (dT_{22}) \wedge (d\tau) \\
&= (1 - t_{11})^{(m-1)/2} (\mathbb{I} - T'_{22}T_{22})^{1/2} dt_{11} \wedge (dT_{22}) \wedge (dv).
\end{aligned} \tag{2.3.7}$$

En combinant (2.3.5), (2.3.6) et (2.3.7), on constate que l'élément différentiel donnant la densité jointe de  $t_{11}$ ,  $T_{22}$  et  $v$ , est égal à

$$\begin{aligned}
&\frac{\Gamma_m(\frac{1}{2}(n_1 + n_2))}{\Gamma_m(\frac{1}{2}n_1)\Gamma_m(\frac{1}{2}n_2)} \times \frac{\Gamma_{m-1}(\frac{1}{2}(n_1 - 1))\Gamma_{m-1}(\frac{1}{2}n_2)}{\Gamma_{m-1}(\frac{1}{2}(n_1 + n_2 - 1))} \times \frac{\Gamma(\frac{1}{2}n_1)\Gamma(\frac{1}{2}n_2)}{\Gamma(\frac{1}{2}(n_1 + n_2))} \\
&\times \left\{ (1 - v'v)^{(n_2 - m - 1)/2} (dv) \right\} \\
&\times \frac{1}{\beta(\frac{1}{2}n_1, \frac{1}{2}n_2)} \left\{ (t_{11}^2)^{n_1/2 - 1} (1 - t_{11}^2)^{n_2/2 - 1} d(t_{11}^2) \right\} \\
&\left\{ \frac{\Gamma_{m-1}(\frac{1}{2}(n_1 + n_2 - 1))}{\Gamma_{m-1}(\frac{1}{2}(n_1 - 1))\Gamma_{m-1}(\frac{1}{2}n_2)} 2^{(m-1)} \prod_{i=2}^m t_{11}^{n_1 - i} (\det(\mathbb{I}_{m-1} - T'_{22}T_{22}))^{(n_2 - m)/2} (dT_{22}) \right\}.
\end{aligned} \tag{2.3.8}$$



Ceci suffit pour établir l'indépendance de  $t_{11}$ ,  $T_{22}$  et  $v$ . De plus, la densité de  $t_{11}^2$  est proportionnelle à  $(t_{11}^2)^{n_1/2-1}(1-t_{11}^2)^{n_2/2-1}$ , ce qui montre que  $t_{11}^2$  suit une loi  $\beta(\frac{1}{2}n_1, \frac{1}{2}n_2)$ . De même, la densité de  $T_{22}$  est, avec la notation introduite en (2.3.5), proportionnelle à  $f(T_{22}; m-1, n_1-1, n_2)$ , et donc égale à cette densité. On peut donc appliquer le même raisonnement à  $T_{22}$  pour établir que  $t_{22}^2$  suit une loi  $\beta(\frac{1}{2}(n_1-1), \frac{1}{2}n_2)$ , et la conclusion du théorème s'impose par une récurrence évidente.  $\square$

**Définition 2.3.2.** Soient  $A \equiv W_m(n_1, \Sigma)$  et  $B \equiv W_m(n_2, \Sigma)$  deux matrices de Wishart indépendantes, et telles que  $\Sigma > 0$ . Le critère  $\Lambda$  de Wilks associé, de paramètres  $n_1$  et  $n_2$  est alors défini par

$$\Lambda^{2/(n_1+n_2)} = \Lambda(n_1 + n_2, m, n_2) = \frac{\det A}{\det(A+B)} \quad (2.3.9)$$

Le critère de Wilks est utilisé (sous l'hypothèse admise que  $A$  suit une loi de Wishart centrée) pour vérifier l'hypothèse que la matrice  $B$  suit une loi de Wishart centrée. Plus précisément, si

$$B = \sum_{i=1}^{n_2} Y_i Y_i',$$

où  $Y_i \equiv N_m(\mu_i, \Sigma)$  pour  $i = 1, \dots, n_2$ , on teste l'hypothèse

$$(H.0) \mu_1 = \dots = \mu_{n_2} = \mathbb{O};$$

contre l'alternative

$$(H.1) \text{ Les } \mu_1, \dots, \mu_{n_2} \text{ sont quelconques.}$$

On rejette donc l'hypothèse (H.0) lorsque  $\Lambda \leq c_\alpha$ , où  $c_\alpha$  est un seuil critique choisi de telle sorte que

$$\mathbb{P}(\Lambda \leq c_\alpha \mid (H.0)) = \alpha.$$

Pour déterminer la valeur explicite de ce seuil, on utilise en général l'*approximation de Bartlett* (Bartlett, M.S. (1938). Further aspects of the theory of multiple regression. *Proc. Cambridge Philos. Soc.*, **34**, 33-40), résumée dans l'énoncé suivant.

**Propriété 2.3.1.** Sous (H.0), lorsque  $n \rightarrow \infty$ ,  $q$  et  $m$  étant fixés,

$$- \left\{ n - \frac{1}{2}(m+q+1) \right\} \log \Lambda(n, m, q) \xrightarrow{d} \chi_{mq}^2. \quad (2.3.10)$$

**Preuve.** Omise.  $\square$

Bartlett (1938) a montré que l'approximation était valable à trois décimales près entre les fonctions de répartition correspondantes, pourvu que la condition suivante soit satisfaite.

$$m^2 + q^2 \leq n - \frac{1}{2}(m+q+1). \quad (2.3.11)$$

Le coefficient  $n - \frac{1}{2}(m+q+1)$  est appelé dans la littérature le *coefficient d'ajustement (ou de correction) de Bartlett*. L'approximation résumée en (2.3.10) et (2.3.11) permet de faire une analyse de variance dans  $\mathbb{R}^m$  à l'aide seulement d'une table donnant la loi du  $\chi^2$ , sous réserve, bien entendu que les degrés de liberté correspondants aux données soient compatibles avec les conditions de l'approximation.

Lorsque tel n'est pas le cas, il faut utiliser la loi exacte du critère de Wilks. Celle ci est relativement complexe et s'obtient à partir du théorème 2.3.3 ci-dessous. On observera que la distribution du critère de Wilks a été tabulée par Wall, F. J. (1968). *The General Variance Ratio of the U-statistic*. Albuquerque, NM. The Dikewood Corporation. La loi exacte du critère de Wilks a été récemment explicitée par Coelho, C. A. (1998). The generalized integer gamma distribution - A basis for distributions in multivariate statistics. *J. Multivariate Analysis*. **64** 86-102.

**Théorème 2.3.3.** *Lorsque  $m \leq q < n$ , on a*

$$\Lambda(n, m, q) \stackrel{d}{=} \prod_{i=1}^m \beta\left(\frac{1}{2}(n - q - i + 1), \frac{1}{2}q\right), \quad (2.3.12)$$

où  $\prod_{i=1}^m \beta\left(\frac{1}{2}(n - q - i + 1), \frac{1}{2}q\right)$  désigne symboliquement le produit de  $m$  variables aléatoires indépendantes de lois respectives  $\beta\left(\frac{1}{2}(n - q - i + 1), \frac{1}{2}q\right)$ , pour  $i = 1, \dots, m$ . Lorsque  $q < m < n$ , on a

$$\Lambda(n, m, q) \stackrel{d}{=} \Lambda(n, q, m). \quad (2.3.13)$$

**Preuve.** La preuve de (2.3.12) est une conséquence directe du Théorème 2.3.2. La preuve de (2.3.13) est omise.  $\square$

L'usage de la statistique  $\Lambda$  de Wilks est justifié par le fait qu'elle constitue la statistique du rapport de vraisemblance de l'hypothèse que  $B$  est contrée contre l'alternative. Plus précisément, si

$$B = \mathbb{Y}'\mathbb{Y} = \sum_{i=1}^q Y_i Y_i',$$

où  $\mathbb{Y} = [Y_1 \ \cdots \ Y_q]'$   $\equiv N_{m,q}(M, \mathbb{I}_q \otimes \Sigma)$ , et

$$A = \mathbb{X}'\mathbb{X} = \sum_{i=1}^{n-q} X_i X_i',$$

où  $\mathbb{X} = [X_1 \ \cdots \ X_{n-q}]'$   $\equiv N_{m,n-q}(\mathbb{O}, \mathbb{I}_q \otimes \Sigma)$ , la vraisemblance associée à  $[\mathbb{X} \ \mathbb{Y}]$ , donnée par (3.4.14), est égale à

$$L(M, \Sigma) = (2\pi)^{-nm/2} (\det \Sigma)^{-n/2} \text{etr} \left\{ -\frac{1}{2} \Sigma^{-1} \left( \mathbb{X}\mathbb{X}' + (\mathbb{Y} - M)(\mathbb{Y} - M)' \right) \right\}.$$

On a alors le théorème suivant.

**Théorème 2.3.4.** *Si  $\Sigma > 0$ , et  $n \geq m$ , on a*

$$\Lambda = \frac{\sup_{\Sigma} L(\mathbb{O}, \Sigma)}{\sup_{M, \Sigma} L(M, \Sigma)} = \left( \frac{\det A}{\det (A + B)} \right)^{n/2}. \quad (2.3.14)$$

**Preuve.** Faisant usage de (2.1.16), on constate que  $L(\mathbb{O}, \Sigma)$  est maximal pour  $\Sigma = \frac{1}{n}(A + B)$ , ce qui donne

$$L(\mathbb{O}, \frac{1}{n}(A + B)) = (2\pi)^{-nm/2} \det \left( \frac{1}{n}(A + B) \right)^{-n/2} e^{-mn/2}.$$

De même,  $L(M, \Sigma)$  est maximal pour  $M = Y$  et  $\Sigma = \frac{1}{n}A$ , ce qui donne

$$L(M, \frac{1}{n}A) = (2\pi)^{-nm/2} \det \left( \frac{1}{n}A \right)^{-n/2} e^{-mn/2}.$$

La conclusion est immédiate.  $\square$

## 2.4 Analyse en composantes principales.

### 2.4.1 Valeurs propres de matrices aléatoires définies positives.

#### Composantes principales.

Soit  $\Sigma \geq 0$  une matrice ( $m \times m$ ) symétrique et positive. Si l'on désigne par  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$  ses valeurs propres rangées par ordre décroissant, il est toujours possible de construire une matrice ( $m \times m$ ) orthogonale

$H = [h_1 \ \dots \ h_m]$  telle que, de manière équivalente,

$$H'\Sigma H = \Lambda := \text{diag}(\lambda_1, \dots, \lambda_m) \Leftrightarrow \Sigma = H\Lambda H' = \sum_{j=1}^m \lambda_j h_j h_j'. \quad (2.4.1)$$

Bien entendu, les vecteurs colonnes de  $H$  composent une base orthonormée dans  $\mathbb{R}^m$ , de vecteurs propres de  $\Sigma$ . Ces vecteurs vérifient donc  $\Sigma h_j = \lambda_j h_j$  pour  $j = 1, \dots, m$ , et  $h_i' h_j = 1$  si  $i = j$  et 0 si  $i \neq j$ .

Considérons maintenant un vecteur aléatoire  $X \stackrel{d}{=} N_m(\mu, \Sigma)$ , et posons

$$U = H'X = \begin{bmatrix} U_1 \\ \vdots \\ U_m \end{bmatrix} = \begin{bmatrix} h_1' X \\ \vdots \\ h_m' X \end{bmatrix} \stackrel{d}{=} N_m(H'\mu, \Lambda). \quad (2.4.2)$$

**Définition 2.4.1.** Les coordonnées  $U_1, \dots, U_m$  de  $U$  sont appelés composantes principales de  $X$ .

On notera que les  $m$  composantes principales de  $X \in \mathbb{R}^m$  sont des variables aléatoires normales indépendantes. La  $j^{\text{ème}}$  composante principale  $U_j$  de  $X$  est telle que

$$U_j = h_j' X \stackrel{d}{=} N(h_j' \mu, \lambda_j).$$

Il est toujours possible de définir  $m$  variables aléatoires  $\omega_1, \dots, \omega_m$ , indépendantes de même loi normale  $N(0, 1)$ , de sorte que pour  $j = 1, \dots, m$ ,

$$U_j = \sqrt{\lambda_j} \omega_j.$$

Ceci permet d'écrire la *décomposition de Karhunen-Loève* de  $X$ , sous la forme

$$X = HU = [h_1 \ \dots \ h_m] \begin{bmatrix} U_1 \\ \vdots \\ U_m \end{bmatrix} = \sum_{j=1}^m \sqrt{\lambda_j} \omega_j h_j.$$

A partir d'un échantillon  $X_1, \dots, X_N$  de  $X \stackrel{d}{=} N_m(\mu, \Sigma)$ , pour  $N \geq 2$ , on construit l'estimation sans biais  $S$  de  $\Sigma$  fournie par

$$S = n^{-1}A \quad \text{où} \quad A = \sum_{i=1}^N (X_i - \bar{X})(X_i - \bar{X})' \quad \text{avec} \quad n = N - 1.$$

On construit alors, selon le même schéma, les valeurs propres  $\ell_1 \geq \dots \geq \ell_m \geq 0$  de  $S$ , ainsi qu'une matrice orthogonale  $Q = [q_1 \ \dots \ q_m]$ , telle que

$$Q'SQ = L := \text{diag}(\ell_1, \dots, \ell_m) \Leftrightarrow S = \sum_{j=1}^m \ell_j q_j q_j'. \quad (2.4.3)$$

On pose alors, pour  $i = 1, \dots, N$ ,

$$\hat{U}(i) = Q'X_i = \begin{bmatrix} \hat{U}_1(i) \\ \vdots \\ \hat{U}_m(i) \end{bmatrix}. \quad (2.4.4)$$

**Définition 2.4.2.** Pour  $i = 1, \dots, N$ , les coordonnées  $\hat{U}_1(i), \dots, \hat{U}_m(i)$  de  $\hat{U}(i)$  sont appelés composantes principales empiriques de  $X_i$ .

### Principe de l'analyse en composantes principales.

L'analyse en composantes principale [ACP] des vecteurs d'un échantillon se résume, la plupart du temps, à l'étude d'échantillons de couples de composantes principales des vecteurs  $X_1, \dots, X_n$ , relatives aux paires de valeurs propres  $\ell_r, \ell_s$  prenant les valeurs numériques les plus élevées (typiquement  $\ell_1, \ell_2, \ell_1, \ell_3$  et  $\ell_2, \ell_3$ ). On projette ainsi les vecteurs  $X_1, \dots, X_n$  sur les plans définis par les couples de vecteurs  $(q_r, q_s)$ , correspondant à certains choix spécifiés de  $r, s$  parmi  $\{1, \dots, m\}$  (typiquement  $(1, 2)$ ,  $(1, 3)$  et  $(2, 3)$ ). On se limite donc, le plus souvent, aux axes de coordonnées définis par les paires  $(q_r, q_s)$  composées par  $q_1, q_2$  et  $q_3$ . Très souvent, l'examen visuel des nuages de points ainsi constitués apporte des renseignements utiles sur la structure des observations.

L'étude théorique de l'ACP constituée à partir d'échantillons  $X_1, \dots, X_n$  gaussiens (i.e., issus de lois normales) requiert l'analyse des propriétés des valeurs propres et vecteurs propres des matrices de Wishart. Tel sera le but poursuivi dans ce qui suit. Nous aborderons ce problème dans un cadre sensiblement élargi.

#### 2.4.2 Densité jointe des valeurs propres.

Nous considérons tout d'abord le problème suivant. Etant donné une matrice symétrique  $(m \times m)$  aléatoire  $A$ , définie positive et de densité  $f(A)$ , relativement à la mesure  $[dA]$ . Comment caractériser la loi jointe des valeurs propres  $\ell_1 \geq \dots \geq \ell_m$  de  $A$  sur  $\mathbb{R}^m$ ? Le théorème ci-dessous donne une réponse appropriée à cette question.

**Théorème 2.4.1.** *Si  $A$  est  $(m \times m)$ , symétrique, positive et de densité  $f(A)$  sur l'ensemble des matrices  $(m \times m)$  définies positives, alors la densité jointe des valeurs propres  $\ell_1 > \dots > \ell_m$  de  $A$ , rangées par ordre décroissant, est donnée par*

$$\frac{\pi^{m^2/2}}{\Gamma_m(\frac{1}{2}m)} \prod_{1 \leq i < j \leq m} (\ell_i - \ell_j) \int_{\mathcal{O}(m)} f(HLH') [dH] \quad \text{pour } \ell_1 > \dots > \ell_m, \quad (2.4.5)$$

où  $L = \text{diag}(\ell_1, \dots, \ell_m)$ , et où  $[dH]$  désigne la mesure invariante normalisée sur le groupe orthogonal  $\mathcal{O}_m$ , définie par (1.4.8).

**Preuve.** Tout d'abord, on vérifie que l'existence d'une densité  $f(A)$  pour  $A$  implique que la probabilité que deux valeurs propres  $\ell_i$  et  $\ell_j$  de  $A$  soient égales, pour  $1 \leq i \neq j \leq m$ , est nulle. Cette propriété peut être établie en constatant que  $A$  a deux valeurs propres égales si et seulement si le polynôme caractéristique  $P(z) = \det(A - z\mathbb{I})$  de  $A$  a une racine commune avec  $P'(z)$ . Ceci implique que  $A$  varie dans une variété de dimension strictement inférieure à  $m(m+1)/2$  et donc de mesure nulle relativement à  $[dA]$ .

Avec probabilité 1, on peut donc ranger les valeurs propres de  $A$  dans un ordre strictement décroissant, pour obtenir  $\ell_1 > \dots > \ell_m > 0$ . Ceci permet de leur associer, de manière unique à un coefficient multiplicateur près, des vecteurs propres de  $A$ . Ces derniers peuvent donc être choisis de façon à composer une base orthonormée de  $\mathbb{R}^m$ , qui sera notée  $h_1, \dots, h_m \in \mathbb{R}^m$ . Les vecteurs de cette base sont définis de façon unique à un coefficient multiplicateur  $\pm 1$  près. En posant  $H = [\pm h_1 \ \dots \ \pm h_m]$ , et  $L = \text{diag}(\ell_1, \dots, \ell_m)$ , on peut donc écrire de  $2^m$  manières différentes la décomposition

$$A = HLH', \quad (2.4.6)$$

où  $H \in \mathcal{O}_m$  est une matrice orthogonale  $(m \times m)$ . L'application  $A \rightarrow (H, L)$  peut être rendue bijective en imposant une condition supplémentaire, comme celle de rendre positive la première coordonnée non nulle de chacun des vecteurs  $h_1, \dots, h_m$ . Evaluons maintenant le jacobien de cette transformation.

Tout d'abord, en différenciant les deux membres de (2.4.6), on constate que

$$dA = dH \times L \times H' + H \times dL \times H' + H \times L \times dH'.$$

En multipliant à gauche cette expression par  $H'$  et à droite par  $H$ , et en faisant état du fait que la relation  $H'H = \mathbb{I}_m$  implique que  $dH'H = -H'dH$ , on obtient

$$H' dA H = H' dH L - L H' dH + dL. \quad (2.4.7)$$

En vertu de (1.1.6), le produit extérieur des éléments distincts de la matrice symétrique  $H' dA H$  du membre de gauche de (2.4.7) est (au signe près)

$$|\det H|^{m+1} [dA] = [dA], \quad (2.4.8)$$

car  $H$ , orthonormale, a un déterminant égal à  $\pm 1$ . Effectuons maintenant le même calcul sur le membre de droite de (2.4.7). Tout d'abord, on constate que

$$\begin{aligned} H' dH L - L H' dH &= \begin{bmatrix} h'_1 \\ \vdots \\ h'_m \end{bmatrix} [\ell_1 dh_1 \quad \dots \quad \ell_m dh_m] - \begin{bmatrix} \ell_1 h'_1 \\ \vdots \\ \ell_m h'_m \end{bmatrix} [dh_1 \quad \dots \quad dh_m] \\ &= \begin{bmatrix} 0 & h'_1 dh_2 (\ell_2 - \ell_1) & \dots & h'_1 dh_m (\ell_m - \ell_1) \\ h'_2 dh_1 (\ell_1 - \ell_2) & 0 & \dots & h'_2 dh_m (\ell_m - \ell_2) \\ \vdots & \vdots & \ddots & \vdots \\ h'_m dh_1 (\ell_1 - \ell_m) & h'_m dh_2 (\ell_2 - \ell_m) & \dots & 0 \end{bmatrix}. \end{aligned}$$

Le produit extérieur (au signe près) des formes différentielles non liées intervenant dans  $H' dH L - L H' dH$  est donc égal à

$$\bigwedge_{1 \leq i < j \leq m} h'_j dh_i \prod_{1 \leq i < j \leq m} (\ell_i - \ell_j).$$

En rajoutant les termes intervenant dans  $dL$ , puis en faisant usage de (2.4.8) ainsi que de la définition (1.3.2), appliquée pour  $m = n$ , on en déduit que

$$\begin{aligned} [dA] &= \bigwedge_{1 \leq i < j \leq m} h'_j dh_i \prod_{1 \leq i < j \leq m} (\ell_i - \ell_j) \bigwedge_{i=1}^m d\ell_i = (H' dH) \prod_{1 \leq i < j \leq m} (\ell_i - \ell_j) \bigwedge_{i=1}^m d\ell_i \\ &= \frac{2^m \pi^{m^2/2}}{\Gamma_m(\frac{1}{2}m)} (dH) \prod_{1 \leq i < j \leq m} (\ell_i - \ell_j) \bigwedge_{i=1}^m d\ell_i, \end{aligned} \quad (2.4.9)$$

où nous avons fait usage de (1.4.8). En divisant les deux membres de (2.4.9) par  $2^m$ , on obtient (2.4.5), ce qui achève la démonstration, après intégration sur  $\mathcal{O}_m$  relativement à  $[dH]$ .  $\square$

### 2.4.3 Valeurs propres d'une matrice de Wishart.

Nous faisons usage de la forme particulière de la loi de Wishart (voir le §2.1.5 ci-dessus), pour appliquer les formules du paragraphe précédent au cas où  $A \stackrel{d}{=} W_m(n, \Sigma)$  suit une loi de Wishart, associée à une matrice de variances-covariances  $\Sigma > 0$ , avec  $n \geq m$ . On a alors

$$f(A) = \frac{2^{-mn/2} (\det \Sigma)^{-n/2}}{\Gamma_m(\frac{1}{2}n)} (\det A)^{\frac{n}{2} - \frac{m+1}{2}} \operatorname{etr} \left( -\frac{1}{2} \Sigma^{-1} A \right) \quad \text{pour } A > 0. \quad (2.4.10)$$

On obtient le corollaire suivant du théorème 2.4.1.

**Corollaire 2.4.1.** *Si  $A \stackrel{d}{=} W_m(n, \Sigma)$  suit une loi de Wishart de paramètres  $\Sigma > 0$  et  $n \geq m$ , la loi jointe des valeurs propres  $\ell_1 > \dots > \ell_m > 0$  de  $A$  a une densité sur  $\mathbb{R}^m$ , donnée par*

$$\frac{\pi^{m^2/2} 2^{-mn/2} (\det \Sigma)^{-n/2}}{\Gamma_m(\frac{1}{2}m) \Gamma_m(\frac{1}{2}n)} \prod_{i=1}^m \ell_i^{(n-m-1)/2} \prod_{1 \leq i < j \leq m} (\ell_i - \ell_j) \int_{\mathcal{O}_m} \operatorname{etr} \left( -\frac{1}{2} \Sigma^{-1} H L H' \right) [dH], \quad (2.4.11)$$

pour  $\ell_1 > \dots > \ell_m > 0$ ,

avec  $L = \operatorname{diag}(\ell_1, \dots, \ell_m)$ .

**Preuve.** Il suffit de combiner (2.4.5), (2.4.10) et l'observation que  $\det A = \det H L H' = \prod_{i=1}^m \ell_i$ .  $\square$

L'intégrale apparaissant dans (2.4.11) dépend des valeurs propres  $\ell_1 > \dots > \ell_m$  de  $A$  par l'intermédiaire de  $L = \text{diag}(\ell_1, \dots, \ell_m)$ . Cette dépendance est en général difficile à expliciter, et sera précisée plus loin à l'aide de *polynômes zonaux*. On obtient toutefois un résultat particulièrement simple dans le cas où  $\Sigma$  est proportionnelle à la matrice identité. On obtient alors le corollaire suivant.

**Corollaire 2.4.2.** *Si  $A \stackrel{d}{=} W_m(n, \sigma^2 \mathbb{I}_m)$  avec  $n \geq m$ , la loi jointe des valeurs propres  $\ell_1 > \dots > \ell_m > 0$  de  $A$  a une densité donnée par*

$$\frac{\pi^{m^2/2}}{(2\sigma^2)^{mn/2} \Gamma_m(\frac{1}{2}m) \Gamma_m(\frac{1}{2}n)} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^m \ell_i\right) \prod_{i=1}^m \ell_i^{(n-m-1)/2} \prod_{1 \leq i < j \leq m} (\ell_i - \ell_j) \quad (2.4.12)$$

pour  $\ell_1 > \dots > \ell_m > 0$ ,

avec  $L = \text{diag}(\ell_1, \dots, \ell_m)$ .

**Preuve.** Il suffit de remplacer  $\Sigma$  par  $\sigma^2 \mathbb{I}_m$  dans (2.4.11), puis de constater que  $\det \Sigma = \lambda^m$ , et enfin d'écrire que  $\text{tr}(\Sigma^{-1} H L H') = \sigma^{-2} \text{tr}(L H' H) = \lambda^{-1} \text{tr}(L)$ . On en déduit que

$$\int_{\mathcal{O}_m} \text{etr}\left(-\frac{1}{2}\Sigma^{-1} H L H'\right) [dH] = \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^m \ell_i\right) \int_{\mathcal{O}_m} (dH) = \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^m \ell_i\right),$$

ce qui permet d'obtenir (2.4.12).  $\square$

#### 2.4.4 Application au test de sphéricité.

Le théorème suivant est d'une grande importance pratique, compte tenu du corollaire 2.4.2 et de la remarque 2.1.2 (voir le §2.1.4). La statistique

$$\mathcal{U} = \frac{\det A}{\left\{\frac{1}{m} \text{tr} A\right\}^m} \quad (2.4.13)$$

est la statistique du rapport de vraisemblance, utilisée pour tester l'hypothèse, dite *de sphéricité*, ( $H.0$ ) que  $\Sigma$  est de la forme  $\sigma^2 \mathbb{I}_m$  pour une valeur convenable de  $\sigma > 0$ , contre l'hypothèse alternative ( $H.1$ ) que  $\Sigma > 0$  est quelconque. Il a été introduit par Mauchly, J. W. (1940). Significant test of sphericity of  $n$ -variate normal populations. *Ann. Math. Statist.* **11** 204-209, et est souvent appelé, de ce fait, le *test de Mauchly*.

**Théorème 2.4.2.** *Si  $A \stackrel{d}{=} W_m(n, \sigma^2 \mathbb{I}_m)$  suit une loi de Wishart de paramètres vérifiant  $n \geq m$  et  $\sigma > 0$ , alors les statistiques  $\mathcal{U} = (\det A) / (\frac{1}{m} \text{tr} A)^m$  et  $\text{tr} A$  sont indépendantes. De plus,  $(1/\sigma^2) \text{tr} A$  suit une loi du  $\chi_{mn}^2$  centrée.*

**Preuve.** On se sert tout d'abord du fait que, sous les hypothèses du théorème,  $(1/\sigma^2)A \equiv W_m(n, \mathbb{I}_m)$ . On constate alors directement que les éléments diagonaux de la matrice  $W_m(n, \mathbb{I}_m)$  sont indépendants et suivent des lois du  $\chi_m^2$ , d'où le fait que  $(1/\sigma^2) \text{tr} A \equiv \chi_{mn}^2$ .

Pour établir l'indépendance de  $\mathcal{U}$  et  $\text{tr} A$ , on utilise (2.4.13) pour établir que la densité jointe des valeurs propres  $\ell_1 > \dots > \ell_m$  de  $A$ , est égale à

$$\frac{\pi^{m^2/2}}{(2\sigma^2)^{mn/2} \Gamma_m(\frac{1}{2}m) \Gamma_m(\frac{1}{2}n)} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^m \ell_i\right) \left(\prod_{i=1}^m \ell_i\right)^{(n-m-1)/2} \prod_{1 \leq i < j \leq m} (\ell_i - \ell_j).$$

On effectue alors le changement de variables

$$\left\{ \ell_1, \dots, \ell_m \right\} \rightarrow \left\{ \bar{\ell} = \frac{1}{n} \sum_{i=1}^n \ell_i, y_1 = \frac{\ell_1}{\bar{\ell}}, \dots, y_{m-1} = \frac{\ell_{m-1}}{\bar{\ell}} \right\},$$

qui permet d'établir, en posant  $y_m = \ell_m/\bar{\ell} = m - y_1 - \dots - y_{m-1}$ , que la densité jointe de  $\bar{\ell}, y_1, \dots, y_{m-1}$  est donnée par

$$\frac{\pi^{m^2/2}(\det \Sigma)^{-n/2}}{(2\sigma^2)^{mn/2}\Gamma_m(\frac{1}{2}m)\Gamma_m(\frac{1}{2}n)} \left\{ \bar{\ell} \right\}^{mn/2-1} \exp\left(-\frac{m\bar{\ell}}{2\lambda}\right) \left( \prod_{i=1}^m y_i \right)^{(n-m-1)/2} \prod_{1 \leq i < j \leq m} (y_i - y_j).$$

L'expression précédente étant factorisée, on constate à l'évidence que  $\bar{\ell}$  est indépendante de  $y_1, \dots, y_m$ . On conclut en observant que  $\mathcal{U} = \prod_{i=1}^m y_i$ , tandis que  $\text{tr}(A) = m\bar{\ell}$ .  $\square$

**Remarque 2.4.1.** Sous l'hypothèse  $(H.0)$  que  $\Sigma = \sigma^2 \mathbb{I}_m$  pour une valeur de  $\sigma > 0$ , le test de sphéricité de Mauchly, décrit en (2.1.21) (voir le §2.1.4 et le corollaire 2.1.1), est un test du rapport de vraisemblance, qui rejette  $(H.0)$  lorsque

$$\mathcal{U} = \frac{\det A}{\left\{ \frac{1}{m} \text{tr} A \right\}^m} \leq u_\alpha, \quad (2.4.14)$$

où  $u_\alpha$  est une valeur critique pouvant être évaluée, lorsque  $n \rightarrow \infty$ , grâce à (2.1.22). Une meilleure approximation (voir Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York, p.344, et Anderson, T.W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York, §10.7.4) est donnée par les formules suivantes.

$$\mathbb{P}(-\rho n \log \mathcal{U} \leq x) = \mathbb{P}(\chi_f^2 \leq x) + \frac{\gamma}{M^2} \left\{ \mathbb{P}(\chi_{f+4}^2 \leq x) - \mathbb{P}(\chi_f^2 \leq x) \right\} + O(M^{-3}), \quad (2.4.15)$$

où

$$\rho = 1 - \frac{2m^2 + m + 2}{6mn}, \quad f = \frac{(m+2)(m-1)}{2},$$

$$\gamma = \frac{(m-1)(m-2)(m+2)(2m^3 + 6m^2 + 3m + 2)}{288m^2}.$$

On notera que, pour  $m \geq 1$  fixé,  $\rho \rightarrow 1$  et  $M = \rho n \rightarrow \infty$ , de sorte que l'approximation ci-dessus peut s'écrire, sous forme plus grossière,

$$\mathbb{P}(-n \log \mathcal{U} \leq x) = \mathbb{P}(\chi_f^2 \leq x) + o(1).$$

## 2.4.5 Lois limites des valeurs propres de matrices de Wishart

Ce paragraphe est à éviter en première lecture, et nécessite une connaissance préalable des propriétés essentielles des polynômes zonaux, ainsi que des lois hypergéométriques matricielles. Le théorème suivant est dû à James, A.T. (1960). The distribution of the latent roots of the covariance matrix. *Ann. Math. Statist.* **31** 151–158.

**Théorème 2.4.3.** Soit  $A = nS \stackrel{d}{=} W_m(n, \Sigma)$  une matrice de Wishart, dont les paramètres vérifient  $n \geq m$  et  $\Sigma > 0$ . Alors la densité jointe des valeurs propres  $\ell_1 > \dots > \ell_m > 0$  de  $S$  est donnée par

$$f(\ell_1, \dots, \ell_m) = \left(\frac{n}{2}\right)^{nm/2} \frac{\pi^{m^2/2}(\det \Sigma)^{-n/2}}{\Gamma_m(\frac{1}{2}m)\Gamma_m(\frac{1}{2}n)} \prod_{i=1}^m \ell_i^{(n-m-1)/2} \prod_{1 \leq i < j \leq m} (\ell_i - \ell_j)$$

$$\times {}_0F_0^{(m)}\left(-\frac{1}{2}nL, \Sigma^{-1}\right) \quad \text{pour } \ell_1 > \dots > \ell_m > 0, \quad (2.4.16)$$

où  $L = \text{diag}(\ell_1, \dots, \ell_m)$  et

$${}_0F_0^{(m)}\left(-\frac{1}{2}nL, \Sigma^{-1}\right) = \sum_{k=0}^{\infty} \sum_{\kappa \in \mathcal{P}_m(k)} \frac{C_\kappa(-\frac{1}{2}nL)C_\kappa(\Sigma^{-1})}{k!C_\kappa(\mathbb{I}_m)}. \quad (2.4.17)$$

**Preuve.** Par application de (2.4.11), la loi jointe de  $\ell_1, \dots, \ell_m$  est donnée par

$$\begin{aligned} & \left(\frac{n}{2}\right)^{mn/2} \frac{\pi^{m^2/2} 2^{-mn/2} (\det \Sigma)^{-n/2}}{\Gamma_m(\frac{1}{2}m) \Gamma_m(\frac{1}{2}n)} \prod_{i=1}^m \ell_i^{(n-m-1)/2} \prod_{1 \leq i < j \leq m} (\ell_i - \ell_j) \\ & \int_{\mathcal{O}_m} \text{etr} \left( -\frac{1}{2} \Sigma^{-1} H L H' \right) [dH], \end{aligned} \quad (2.4.18)$$

pour  $\ell_1 > \dots > \ell_m > 0$ . Il suffit alors d'écrire, par application directe de (3.2.47) et (3.3.4), que

$$\begin{aligned} \int_{\mathcal{O}_m} \text{etr} \left( -\frac{1}{2} \Sigma^{-1} H L H' \right) [dH] &= \sum_{k=0}^{\infty} \sum_{\kappa} \frac{1}{k!} \int_{\mathcal{O}_m} C_{\kappa} \left( -\frac{1}{2} \Sigma^{-1} H L H' \right) [dH] \\ &= \sum_{k=0}^{\infty} \sum_{\kappa} \frac{C_{\kappa} \left( -\frac{1}{2} n L \right) C_{\kappa} \left( \Sigma^{-1} \right)}{k! C_{\kappa}(\mathbb{I}_m)} = {}_0F_0^{(m)} \left( -\frac{1}{2} n L, \Sigma^{-1} \right), \end{aligned}$$

ce qu'il fallait démontrer.  $\square$

**Théorème 2.4.4.** Soient  $\ell_1 > \dots > \ell_m > 0$  les valeurs propres de  $S$ , où  $nS \stackrel{d}{=} W_m(n, \Sigma)$  avec  $n \geq m$  et  $\Sigma > 0$ . Sous l'hypothèse que les valeurs propres  $\lambda_1, \dots, \lambda_m$  de  $\Sigma$  vérifient  $\lambda_1 > \dots > \lambda_k > \lambda_{k+1} = \dots = \lambda_m > 0$ , la loi limite jointe lorsque  $n \rightarrow \infty$  de  $X_1, \dots, X_m$ , où, pour  $i = 1, \dots, m$ ,

$$X_i = \left(\frac{n}{2}\right)^{1/2} \left(\frac{\ell_i - \lambda_i}{\lambda_i}\right), \quad (2.4.19)$$

a une densité donnée par

$$\begin{aligned} f(x_1, \dots, x_m) &= \left\{ \prod_{i=1}^k \varphi(x_i) \right\} \frac{\pi^{q(q-1)/4}}{2^{q/2} \Gamma_q(\frac{1}{2}q)} \exp \left( -\frac{1}{2} \sum_{j=k+1}^m x_j^2 \right) \prod_{k+1 \leq i < j \leq m} (x_i - x_j) \\ &\text{pour } x_1, \dots, x_m \in \mathbb{R} \text{ et } x_{k+1} > \dots > x_m. \end{aligned} \quad (2.4.20)$$

où  $q = m - k$  et  $\varphi(t) = (2\pi)^{-1/2} \exp(-t^2/2)$  est la densité de la loi normale  $N(0, 1)$  standard.

**Preuve.** Voir Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York, pp.392–403. La démonstration, dont nous omettons les détails, utilise essentiellement un développement asymptotique basé sur (2.4.18). Les formules se simplifient très notablement lorsque les valeurs propres  $\lambda_1, \dots, \lambda_m$  de  $\Sigma$  sont distinctes. On peut alors utiliser la formule suivante, due à Anderson, T.W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Statist.* **34** 122–148. Lorsque  $n \rightarrow \infty$ ,

$${}_0F_0^{(m)} \left( -\frac{1}{2} n L, \Sigma^{-1} \right) = (1 + o(1)) \frac{\Gamma_m(\frac{1}{2}m)}{\pi^{m^2/2}} \exp \left( -\frac{n}{2} \sum_{i=1}^m \frac{\ell_i}{\lambda_i} \right) \prod_{1 \leq i < j \leq m} \left( \frac{2\pi \lambda_i \lambda_j}{n(\ell_i - \ell_j)(\lambda_i - \lambda_j)} \right). \quad (2.4.21)$$

En combinant (2.4.16) avec (2.4.19) et (2.4.21), on obtient aisément la loi limite (2.4.20) de  $X_1, \dots, X_m$ , dans ce cas, un produit de lois  $N(0, 1)$  indépendantes.  $\square$

Le théorème suivant fournit la loi limite des vecteurs propres orthonormés  $q_1, \dots, q_m$  associés à des valeurs propres  $\ell_1 > \dots > \ell_m$  de  $S$  lorsque les valeurs propres  $\lambda_1 > \dots > \lambda_m$  de  $\Sigma$  sont *distinctes*.

**Théorème 2.4.5.** Supposons que les valeurs propres  $\lambda_1 > \dots > \lambda_m > 0$  de  $\Sigma > 0$  soient *distinctes*, et soient  $h_1, \dots, h_m$  les vecteurs propres correspondants. Soit  $nS \stackrel{d}{=} W_m(n, \Sigma)$  pour  $n \geq m$ , et soient  $\ell_1 > \dots > \ell_m > 0$  les valeurs propres de  $S$ , et  $q_1, \dots, q_m$  les vecteurs propres associés (avec la convention, pour chaque  $i = 1, \dots, m$ , que  $q_i$  est le vecteur propre, de norme  $(q_i' q_i)^{1/2} = 1$ , associé à la valeur propre  $\lambda_i$ , et qui maximise  $q_i' h_i$ ). Alors, pour chaque  $i = 1, \dots, m$ , la loi limite de  $n^{1/2}(q_i - h_i)$  est asymptotiquement indépendante de  $\ell_i$  est normale  $N_m(\mathbb{O}, \Gamma)$ , avec

$$\Gamma = \lambda_i \sum_{\substack{j=1 \\ j \neq i}}^m \frac{\lambda_j}{(\lambda_i - \lambda_j)^2} h_j h_j'. \quad (2.4.22)$$



**Preuve.** Nous omettons la démonstration de ce théorème, en renvoyons à Anderson, T.W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Statist.* **34** 122–148, pour plus de détails.□

## Chapitre 3

# Polynômes Zonaux et Applications.

### 3.1 Polynômes Zonaux.

#### 3.1.1 Introduction.

Les polynômes zonaux ont été introduits par James dans une série d'articles à partir de 1961 (voir James, A.T. (1961). Zonal polynomials of the real positive definite symmetric matrices. *Annals of Mathematics*. **74** 456–469, James, A.T. (1968). Calculation of zonal polynomial coefficients by use of the Laplace–Beltrami operator. *Ann. Math. Statist.* **39** 1711–1718, et les références citées dans ce même article). Ils jouent un rôle fondamental, entre autres, dans le calcul d'intégrales exprimant de la densité des valeurs propres d'une matrice de Wishart (voir (2.4.11)), comme la suivante

$$\int_{\mathcal{O}_m} \text{etr}\left(-\frac{1}{2}\Sigma^{-1}HLH'\right)[dH].$$

Par ailleurs les polynômes zonaux sont indispensables à la définition des *fonctions hypergéométriques matricielles*, dont l'emploi est rendu nécessaire par l'étude des lois de Wishart non centrées.

Cependant, leur définition, comme la description de leurs propriétés, font usage de techniques délicates, et on évitera donc d'aborder ce paragraphe en première lecture.

#### 3.1.2 Partitions d'un entier.

Dans ce qui suit, pour tout entier positif  $m \geq 1$ , et pour tout entier positif  $k \geq 0$ , on appellera  $m$ -partition de  $k$  tout  $m$ -uplet  $\kappa = (k_1, \dots, k_m)$  d'entiers tels que

$$k_1 \geq \dots \geq k_m \geq 0, \quad \text{et} \quad \sum_{i=1}^m k_i = k. \quad (3.1.1)$$

Pour toute  $m$ -partition  $\kappa = (k_1, \dots, k_m)$  de l'entier  $k \geq 1$ , on appelle :

- hauteur de  $\kappa$  le nombre  $h(\kappa) = k_1 = \max_{j \geq 1} k_j$  ;
- longueur de  $\kappa$  la plus petite valeur  $L(\kappa)$  de  $q \geq 0$  telle que  $k_m = 0$  pour  $m \geq q + 1$  ;
- degré de  $\kappa$  le nombre  $k$ , noté  $|\kappa| = k$ .

Ces formules s'appliquent dans le cas particulier de la  $m$ -partition nulle  $\mathbb{O} = (0, \dots, 0)$  de  $k = 0$ . On définit ainsi

- la hauteur de  $\mathbb{O}$  par  $h(\mathbb{O}) = 0$  ;
- la longueur de  $\mathbb{O}$  par  $L(\mathbb{O}) = 0$  ;
- le degré de  $\mathbb{O}$  par  $|\mathbb{O}| = 0$ .

Pour tout  $m \geq 1$ , l'ensemble des  $m$ -partitions de  $k \geq 0$  sera noté  $\mathcal{P}_m(k)$ . Lorsque  $k = 0$ , l'ensemble  $\mathcal{P}_m(0) = \{\mathbb{O}\}$  est réduit au seul élément  $\mathbb{O} = (0, \dots, 0)$ . Pour toute  $m$ -partition  $\kappa = (k_1, \dots, k_m)$  de  $k \geq 0$ , et pour chaque valeur de  $j = 1, 2, \dots$ , on appelle *multiplicité*  $\nu_j$  de  $j$ , le nombre de fois que  $k_i = j$  prend la valeur  $j$  lorsque l'indice  $i$  varie de 1 à  $m$  ( $1 \leq i \leq m$ ).

**Exemple 3.1.1.** Par exemple,  $\kappa = (2, 2, 1, 1, 1, 0)$  est une 6-partition de degré  $|\kappa| = 7$ , de hauteur  $h(\kappa) = 2$  et de longueur  $L(\kappa) = 5$ . Les multiplicités correspondantes sont  $\nu_1 = 3$ ,  $\nu_2 = 2$  et  $\nu_j = 0$  pour  $j \geq 3$ .

On déduit de ces définitions l'égalité, pour toute  $m$ -partition  $\kappa = (k_1, \dots, k_m)$  de  $k \geq 0$ ,

$$|\kappa| = k = \nu_1 + 2\nu_2 + \dots + k\nu_k, \quad (3.1.2)$$

La forme explicite  $\kappa = (k_1, \dots, k_m)$  est appelée *forme développée* de la partition. Il est commode d'écrire une telle partition sous *forme résumée*, en posant, conformément aux notations ci-dessus, lorsque  $|\kappa| \geq 1$ ,

$$\kappa = (1^{\nu_1} 2^{\nu_2} \dots k^{\nu_k}), \quad (3.1.3)$$

avec la convention que  $j^\nu$  n'apparaît pas dans (3.1.3) si  $\nu = \nu_j = 0$ . On notera de même  $j^\nu$  par  $j$ , au lieu de  $j^1$  lorsque  $\nu = \nu_j = 1$ . Dans le cas particulier de la partition nulle  $\mathbb{O} = (0, \dots, 0)$ , on adoptera la forme résumée

$$\mathbb{O} = (0, \dots, 0) = (0). \quad (3.1.4)$$

**Exemple 3.1.2.** La 6-partition  $\kappa = (2, 2, 1, 1, 1, 0)$  de 7 est notée, sous forme résumée  $\kappa = (1^3 2^2)$ .

Les conventions ci-dessous seront adoptées par la suite. Pour tout choix de  $r \geq 1$ , la  $m$ -partition  $\kappa = (k_1, \dots, k_m)$  de  $k$  sera également considérée comme une  $(m+r)$ -partition de  $k$ , en identifiant  $(k_1, \dots, k_m)$  à la  $(m+r)$ -partition  $(k_1, \dots, k_m, 0, \dots, 0)$ , obtenue en rajoutant  $r$  zéros à  $\kappa$  au delà du dernier terme. De la sorte, on a l'inclusion naturelle

$$\mathcal{P}_1(k) \subseteq \mathcal{P}_2(k) \subseteq \dots \subseteq \mathcal{P}_m(k) \subseteq \dots \subseteq \mathcal{P}(k), \quad (3.1.5)$$

où  $\mathcal{P}(k)$  désigne l'ensemble des suites infinies  $\kappa = (k_1, k_2, \dots)$  d'entiers vérifiant

$$k_1 \geq k_2 \geq \dots \geq 0 \quad \text{et} \quad \sum_{i=1}^{\infty} k_i = k,$$

et où nous identifions  $\kappa = (k_1, \dots, k_m) \in \mathcal{P}_m(k)$  à  $(k_1, \dots, k_m, 0, \dots) \in \mathcal{P}(k)$ .

Ceci mène à définir en général une *partition* de  $k \geq 0$  comme toute, suite infinie, décroissante au sens large,  $\kappa = (k_1, k_2, \dots)$  d'entiers positifs ou nuls de somme égale à  $k$ . Une telle suite devant nécessairement être nulle à partir du rang  $k+1$ , il existe donc une valeur de  $m$ , vérifiant  $0 \leq m \leq k$ , pour laquelle, d'une part  $(k_1, \dots, k_m)$  soit une  $m$ -partition de  $k$  lorsque  $k \geq 1$ , et d'autre part,  $k_{m+1} = k_{m+2} = \dots = 0$ . La plus petite valeur de  $m$  vérifiant cette propriété est la longueur de  $\kappa$ . On peut donc écrire, pour  $k \geq 1$ ,

$$\mathcal{P}(k) = \bigcup_{m=1}^{\infty} \mathcal{P}_m(k),$$

et définir la longueur d'une partition  $\kappa \in \mathcal{P}(k)$  de  $k \geq 0$  par

$$\begin{aligned} L(\kappa) &= \inf\{m \geq 1 : \kappa \in \mathcal{P}_m(k)\} \quad \text{lorsque} \quad |\kappa| \geq 1, \\ L(\mathbb{O}) &= 0. \end{aligned}$$

À l'exception de la formule précédente, qui distingue le cas où  $\kappa \neq \mathbb{O}$  de celui où  $\kappa = \mathbb{O}$ , les formules obtenues pour les partitions d'un entier  $k \geq 1$  restent valables pour le cas où  $k = 0$ . Nous explicitons ceci plus en détail. Pour  $m \geq 1$  et  $k = 0$ , on adoptera la convention que l'unique  $m$ -partition de 0 est la suite  $\mathbb{O} = (0, \dots, 0)$ , de  $m$  zéros consécutifs, elle-même identifiée à la suite infinie  $\mathbb{O} = (0, 0, \dots)$ . En notant  $\mathcal{P}_m(0)$  l'ensemble composé de cette suite, soit  $\mathcal{P}(0) = \{\mathbb{O} = (0, 0, \dots)\}$ , on peut conserver un sens à (3.1.5) pour tout  $k \in \mathbb{N}$ .

On adoptera l'ordre *lexicographique* pour classer les partitions d'un même entier  $k \in \mathbb{N}$ . Cet ordre est défini en posant, lorsque  $\kappa = (k_1, k_2, \dots) \in \mathcal{P}(k)$  et  $\lambda = (\ell_1, \ell_2, \dots) \in \mathcal{P}(k)$  sont deux partitions distinctes de l'entier  $k$ ,

$$\kappa > \lambda \Leftrightarrow \lambda < \kappa \Leftrightarrow \exists r \geq 1 : k_i = \ell_i \quad \forall 1 \leq i < r \quad \text{et} \quad k_r > \ell_r. \quad (3.1.6)$$

par extension de (3.1.6), on posera

$$\kappa \geq \lambda \Leftrightarrow \lambda \leq \kappa \Leftrightarrow \text{ou bien } \kappa > \lambda, \quad \text{ou bien } \kappa = \lambda. \quad (3.1.7)$$

Nous noterons que l'ordre lexicographique défini par  $\lambda \leq \kappa$  ou  $\lambda \geq \kappa$  est un *ordre total* sur l'ensemble  $\mathcal{P}(k)$  des partitions de  $k \in \mathbb{N}$ . En d'autres termes, si  $\kappa$  et  $\lambda$  sont deux partitions quelconques de  $k$ , on a nécessairement, ou bien  $\kappa < \lambda$ , ou bien  $\kappa > \lambda$ , ou bien  $\kappa = \lambda$ .

**Exemple 3.1.3.** La liste ci-dessous donne la *forme développée* de  $\kappa = (k_1, k_2, \dots)$  ainsi que la *forme résumée*  $\kappa = (1^{\nu_1} 2^{\nu_2} \dots)$  des partitions de  $k = 5$ , rangées dans l'ordre lexicographique, et exprimées sous chacune des formes, développée ou résumée.

$$\begin{aligned} (5) &> (4, 1) > (3, 2) > (3, 1, 1) > (2, 2, 1) > (2, 1, 1, 1) > (1, 1, 1, 1, 1), \\ (5) &> (41) > (32) > (31^2) > (2^2 1) > (21^3) > (1^5). \end{aligned}$$

### 3.1.3 Polynômes symétriques.

Un *polynôme*  $P(y) = P(y_1, \dots, y_m)$  des variables  $y_1, \dots, y_m$  est une fonction de  $y = \{y_1, \dots, y_m\}$  de la forme

$$P(y) = P(y_1, \dots, y_m) = \sum_{0 \leq k_1, \dots, k_m \leq N} c_{k_1, \dots, k_m} y_1^{k_1} \dots y_m^{k_m}, \quad (3.1.8)$$

où les  $c_{k_1, \dots, k_m} \in \mathbb{R}$  sont des constantes. Le polynôme  $P(\cdot)$  s'exprime ainsi comme une combinaison linéaire finie des *monômes simples*

$$y_1^{k_1} \dots y_m^{k_m}. \quad (3.1.9)$$

On appellera *monôme symétrique* associé au *monôme simple*  $y_1^{k_1} \dots y_m^{k_m}$ , le polynôme défini par

$$M(y) = M(y_1, \dots, y_m) = \sum_{\{\pi_1, \dots, \pi_m\} \in \mathcal{P}_m} y_{\pi_1}^{k_1} \dots y_{\pi_m}^{k_m}, \quad (3.1.10)$$

où  $\{\pi_1, \dots, \pi_m\}$  parcourt l'ensemble  $\mathcal{P}_m$  des permutations de  $\{1, \dots, m\}$ . Comme une telle expression ne dépend pas de l'ordre dans lequel sont donnés  $k_1, \dots, k_m$ , on peut supposer, sans perte de généralité que ces entiers sont rangés par ordre décroissant, de telle sorte que  $k_1 \geq \dots \geq k_m$ . On constate alors que le monôme symétrique  $M(\cdot)$  de (3.1.10) est défini de manière unique par la  $m$ -partition  $\kappa = (k_1, \dots, k_m) \in \mathcal{P}_m(k)$  de  $k = k_1 + \dots + k_m$ . Pour cette raison, on désignera ce monôme symétrique par  $M_\kappa(\cdot)$ , en adoptant la notation

$$M_\kappa(y) = M_\kappa(y_1, \dots, y_m) = \sum_{\{\pi_1, \dots, \pi_m\} \in \mathcal{P}_m} y_{\pi_1}^{k_1} \dots y_{\pi_m}^{k_m}. \quad (3.1.11)$$

On dira alors que  $y_1^{k_1} \dots y_m^{k_m}$  est le monôme simple *dominant* ou *générateur* de  $M_\kappa$ .

Considérons deux monômes simples de degré  $k$ , soit  $y_1^{k_1} \dots y_m^{k_m}$  et  $y_{\pi_1}^{k_1} \dots y_{\pi_m}^{k_m} = y_1^{r_1} \dots y_m^{r_m}$ , associés à la même  $m$ -partition  $\kappa = (k_1, k_2, \dots)$  de  $k$ ,  $\pi = \{\pi_1, \dots, \pi_m\}$  étant une permutation de  $\{1, \dots, m\}$ .

**Définition 3.1.1.** Nous dirons que l'ordre de  $y_1^{k_1} \dots y_m^{k_m}$  est supérieur (respectivement égal) à celui de  $y_{\pi_1}^{k_1} \dots y_{\pi_m}^{k_m} = y_1^{r_1} \dots y_m^{r_m}$  si  $\{r_1, \dots, r_m\} \neq \{k_1, \dots, k_m\}$  (resp.  $\{r_1, \dots, r_m\} = \{k_1, \dots, k_m\}$ ).

**Exemple 3.1.4.** Pour  $\kappa = (1^2) = (1, 1)$  on a  $M_\kappa(y_1, y_2) = 2y_1y_2$ , et le monôme générateur  $y_1y_2$  est de même ordre que  $y_2y_1$ . Pour  $\kappa = (2) = (2, 0)$  on a  $M_\kappa(y_1, y_2) = y_1^2 + y_2^2$ . Dans ce dernier polynôme, le monôme dominant est  $y_1^2$ , et l'ordre de  $y_2^2$  est inférieur à celui de  $y_1^2$ .

Aussi bien dans (3.1.8)–(3.1.9) que dans les expressions qui suivent, on adopte systématiquement la convention que  $y^0 = 1$  ( $y$  compris lorsque  $y = 0$ , en posant donc  $0^0 = 1$ ). De plus, sauf mention du contraire, tous les polynômes considérés ici sont à coefficients réels. Le *degré* d'un polynôme de ce type, est, par définition, le degré du polynôme d'une seule variable obtenu en remplaçant chacune des variables  $y_1, \dots, y_k$  par la même variable réelle  $y$ . La définition du degré d'un polynôme d'une seule variable que nous adoptons ici est la suivante. Si

$$R(y) = \sum_{k=0}^N c_k y^k \quad \text{avec } c_N \neq 0,$$

on pose  $\deg(R) = N$ . Par convention, si  $\mathbb{O}$  désigne le polynôme nul, on pose  $\deg(\mathbb{O}) = -\infty$ . Ayant ainsi défini par  $\deg(R)$  le degré d'un polynôme  $R \neq \mathbb{O}$  non identiquement nul, il importe de distinguer, lorsque  $R(y) = c$  est une constante, le cas où  $c \neq 0$  (pour lequel  $\deg(R) = 0$ ), du cas où  $c = 0$  (pour lequel  $\deg(R) = -\infty$ ).

Un *polynôme homogène* de degré  $k$  est, par définition, un polynôme de degré  $k$  s'exprimant comme une combinaison linéaire finie non nulle de monômes simples, chacun de degré égal à  $k$ . Un tel polynôme est donc nécessairement de la forme

$$P(y_1, \dots, y_m) = \sum_{\substack{k_1, \dots, k_m \geq 0 \\ k_1 + \dots + k_m = k}} c_{k_1, \dots, k_m} y_1^{k_1} \dots y_m^{k_m}. \quad (3.1.12)$$

L'hypothèse que  $P(y) \neq \mathbb{O}$  implique, à l'évidence, que  $\deg(P) = k$ .

Considérons deux monômes simples de degré  $k$ , soit  $y_{\pi'_1}^{k_1} \dots y_{\pi'_m}^{k_m}$  et  $y_{\pi''_1}^{\ell_1} \dots y_{\pi''_m}^{\ell_m}$ , associés respectivement aux  $m$ -partitions  $\kappa = (k_1, k_2, \dots)$  et  $\lambda = (\ell_1, \ell_2, \dots)$  de  $k$ ,  $\pi' = \{\pi'_1, \dots, \pi'_m\}$  et  $\pi'' = \{\pi''_1, \dots, \pi''_m\}$  étant deux permutations de  $\{1, \dots, m\}$ .

**Définition 3.1.2.** *Nous dirons que le poids de  $y_{\pi'_1}^{k_1} \dots y_{\pi'_m}^{k_m}$  est supérieur (resp. égal) au poids de  $y_{\pi''_1}^{\ell_1} \dots y_{\pi''_m}^{\ell_m}$ , si  $\kappa > \lambda$  (resp.  $\kappa = \lambda$ ).*

Un *polynôme symétrique*  $P(y_1, \dots, y_m)$  des variables  $y_1, \dots, y_m$  est un polynôme vérifiant l'identité

$$P(y_1, \dots, y_m) = P(y_{\pi_1}, \dots, y_{\pi_m}),$$

pour toute permutation  $\pi = \{\pi_1, \dots, \pi_m\}$  de l'ensemble  $\mathcal{P}_m$  des permutations de  $\{1, \dots, m\}$ . On constate qu'un polynôme symétrique et homogène de degré  $k$  s'écrit de manière unique sous la forme

$$\begin{aligned} P(y_1, \dots, y_m) &= \sum_{\kappa \in \mathcal{P}_m(k)} d_\kappa M_\kappa(y_1, \dots, y_m) \\ &= \sum_{\kappa \in \mathcal{P}_m(k)} d_\kappa \sum_{\{\pi_1, \dots, \pi_m\} \in \mathcal{P}_m} y_{\pi_1}^{k_1} \dots y_{\pi_m}^{k_m}. \end{aligned}$$

Dans une telle expression, il est toujours possible de classer les  $m$ -partitions  $\kappa$  de  $k$  pour lesquelles  $d_\kappa \neq 0$  par poids décroissants. Si  $\kappa_0$  désigne la partition de plus grands poids correspondante, il sera commode par la suite de décomposer  $P(\cdot)$  sous la forme

$$P(y_1, \dots, y_m) = d_{\kappa_0} M_{\kappa_0}(y_1, \dots, y_m) + (\text{termes de poids inférieur}), \quad (3.1.13)$$

expression qu'on résumera plus simplement en remplaçant  $M_{\kappa_0}(y_1, \dots, y_m)$  par son *terme dominant*. Si  $\kappa_0 = (k_1, \dots, k_m)$ , ce dernier se réduit à  $y_1^{k_1} \dots y_m^{k_m}$ . On rajoute alors, par convention, toutes les composantes monomiales d'ordre inférieur, et donc de la forme

$$y_{\pi_1}^{k_1} \dots y_{\pi_m}^{k_m} = y_1^{\ell_1} \dots y_m^{\ell_m},$$

correspondant à des permutations  $\{\pi_1, \dots, \pi_m\}$  de  $\{1, \dots, n\}$  pour lesquelles  $\{k_1, \dots, k_m\} \neq \{\ell_1, \dots, \ell_m\}$  à l'ensemble des termes de poids inférieur. On écrira ainsi

$$\begin{aligned} P(y_1, \dots, y_m) &= d_{\kappa_0} M_{\kappa_0}(y_1, \dots, y_m) + (\text{termes de poids inférieur}) \\ &= d_{\kappa_0} r_{\kappa_0} y_1^{k_1} \dots y_m^{k_m} + (\text{termes d'ordre ou de poids inférieur}). \end{aligned} \quad (3.1.14)$$

Ici,  $r_{\kappa_0}$  désigne le nombre de permutations distinctes  $\{\pi_1, \dots, \pi_m\}$  de  $\{1, \dots, m\}$  telles que

$$y_{\pi_1}^{k_1} \dots y_{\pi_m}^{k_m} = y_1^{k_1} \dots y_m^{k_m}.$$

Il est à noter que les termes, d'ordre ou de poids inférieur, dans l'expression ci-dessus comprennent, d'une part des combinaisons linéaires de monômes symétriques associés à des  $m$ -partitions  $\lambda$  de  $k$  telles que  $\lambda < \kappa_0$ , et d'autre part, des combinaisons linéaires de monômes simples  $y_{\pi_1}^{k_1} \dots y_{\pi_m}^{k_m} = y_1^{\ell_1} \dots y_m^{\ell_m}$ , où  $\{\ell_1, \dots, \ell_m\} \neq \{k_1, \dots, k_m\}$ .

**Exemple 3.1.5.** Pour  $\kappa_0 = (1, 1, 0)$ , le monôme générateur est  $y_1 y_2$ , d'ordre égal à  $y_2 y_1$ . Les monômes  $y_1 y_3$  et  $y_2 y_3$  sont d'ordre inférieur à  $y_1 y_2$ . On a donc  $r_{\kappa_0} = 2$  et

$$M_{\kappa}(y_1, y_2, y_3) = 2y_1 y_2 + 2y_1 y_3 + 2y_2 y_3 = 2y_1 y_2 + (\text{termes d'ordre inférieur}).$$

**Proposition 3.1.1.** Soient deux polynômes  $P_1(\cdot)$  et  $P_2(\cdot)$  de  $y_1, \dots, y_m$ , homogènes et symétriques, de degrés respectifs  $k \geq 0$  et  $\ell \geq 0$ . On suppose que

$$\begin{aligned} P_1(y_1, \dots, y_m) &= c_1 y_1^{k_1} \dots y_m^{k_m} + (\text{termes d'ordre ou de poids inférieur}), \\ P_2(y_1, \dots, y_m) &= c_2 y_1^{\ell_1} \dots y_m^{\ell_m} + (\text{termes d'ordre ou de poids inférieur}), \end{aligned}$$

où  $(k_1, \dots, k_m)$  est une  $m$ -partition de  $k$ , et  $(\ell_1, \dots, \ell_m)$  une  $m$ -partition de  $\ell$ . Alors

$$P_1(y_1, \dots, y_m) P_2(y_1, \dots, y_m) = c_1 c_2 y_1^{k_1 + \ell_1} \dots y_m^{k_m + \ell_m} + (\text{termes d'ordre ou de poids inférieur}). \quad (3.1.15)$$

**Preuve.** Evidente.  $\square$

**Définition 3.1.3.** On appelle fonctions symétriques élémentaires de  $y_1, \dots, y_m$  les polynômes définis, pour  $r = 1, \dots, m$ , par

$$a_r(y_1, \dots, y_m) = \sum_{1 \leq i_1 < \dots < i_r \leq m} y_{i_1} \dots y_{i_r}, \quad (3.1.16)$$

et, pour  $r \geq 1$ , par

$$s_r(y_1, \dots, y_m) = \sum_{i=1}^m y_i^r. \quad (3.1.17)$$

**Proposition 3.1.2.** Tout polynôme  $P(\cdot)$ , homogène symétrique de  $y_1, \dots, y_m$ , de degré  $k \geq 0$ , est un polynôme des fonctions symétriques élémentaires  $a_1, \dots, a_m$  de  $y_1, \dots, y_m$ .

**Preuve.** Mettons le polynôme  $P(\cdot)$  sous la forme

$$P_0(y_1, \dots, y_m) = P(y_1, \dots, y_m) = c_1 y_1^{k_1} \dots y_m^{k_m} + (\text{termes d'ordre ou de poids inférieur}),$$

et construisons le polynôme

$$R_1(y_1, \dots, y_m) = c_1 a_1^{k_1 - k_2} a_2^{k_2 - k_3} \dots a_{m-1}^{k_{m-1} - k_m} a_m^{k_m}.$$

Comme

$$a_r = a_r(y_1, \dots, y_m) = y_1 \dots y_r + (\text{termes d'ordre ou de poids inférieur}),$$

on déduit de (3.1.15) que

$$R_1(y_1, \dots, y_m) = c_1 y_1^{k_1} \dots y_m^{k_m} + (\text{termes d'ordre ou de poids inférieur}),$$

et donc que

$$P_0(y_1, \dots, y_m) = R_1(y_1, \dots, y_m) + (\text{termes de poids inférieur}).$$

On répète le même raisonnement pour  $P_1 = P_0 - R_1$  dont le poids est inférieur à celui de  $P_0$ . On obtient par récurrence une suite de polynômes  $P_1, P_2, \dots$  de poids décroissants. Comme cette suite est nécessairement finie, il existe un indice  $r$  tel que  $P_r = \mathbb{O}$ , ce qui montre que  $P = R_1 + \dots + R_r$ .  $\square$

**Exemple 3.1.6.** Soit  $p = P_0 = s_3(y_1, y_2, y_3)$ . On voit que

$$\begin{aligned} P_0 &= s_3(y_1, y_2, y_3) = y_1^3 + y_2^3 + y_3^3 = y_1^3 + (\text{termes d'ordre ou de poids inférieur}) \\ &= a_1^3 + (\text{termes d'ordre ou de poids inférieur}) \\ &= (y_1 + y_2 + y_3)^3 + (\text{termes d'ordre ou de poids inférieur}). \end{aligned}$$

On obtient donc, en posant  $R_1 = a_1^3$ ,

$$\begin{aligned} P_1 &= P_0 - R_1 = y_1^3 + y_2^3 + y_3^3 - (y_1 + y_2 + y_3)^3 \\ &= -3y_1^2y_2 - 3y_2^2y_1 - 3y_1^2y_3 - 3y_3^2y_1 - 3y_2^2y_3 - 3y_3^2y_2 - 6y_1y_2y_3 \\ &= -3y_1^2y_2 + (\text{termes d'ordre ou de poids inférieur}) \\ &= -3a_1a_2 + (\text{termes d'ordre ou de poids inférieur}) \\ &= -3(y_1 + y_2 + y_3)(y_1y_2 + y_1y_3 + y_2y_3) + (\text{termes d'ordre ou de poids inférieur}). \end{aligned} \quad (3.1.18)$$

On obtient donc, en posant  $R_2 = -3a_1a_2$ ,

$$P_2 = P_1 - R_1 = 9y_1y_2y_3 - 6y_1y_2y_3 = 3y_1y_2y_3 = 3a_3.$$

On conclut en vérifiant que

$$s_3(y_1, y_2, y_3) = a_1^3 - 3a_1a_3 + 3a_3.$$

## 3.2 Polynômes zonaux

### 3.2.1 Polynômes matriciels.

Afin d'introduire la notion de polynôme zonal, nous allons tout d'abord étudier les polynômes des coefficients de matrices carrées quelconques ( $m \times m$ ). Soit  $Z = [z_{ij}]$  une telle matrice. Nous considérons donc les fonctions polynômiales de la forme

$$P(Z) = P_1(\{z_{ij} : 1 \leq i, j \leq m\}). \quad (3.2.1)$$

Lorsque  $Z = X = [x_{ij}]$  est symétrique, c'est à dire telle que  $x_{ij} = x_{ji}$  pour tout couple d'entiers  $(i, j)$  tels que  $1 \leq i, j \leq m$ , on peut simplifier cette écriture en

$$P(X) = P_2(\{x_{ij} : 1 \leq i \leq j \leq m\}). \quad (3.2.2)$$

Naturellement, les polynômes  $P_1$  et  $P_2$  de (3.2.1) et (3.2.2) ne sont pas les mêmes (l'un est à  $m^2$  variables tandis que le second est à  $m(m+1)/2$  variables).

**Définition 3.2.1.** Nous dirons que le polynôme  $P(X)$  de la matrice symétrique  $(m \times m)$   $X$  est invariant, si pour toute matrice orthogonale  $(m \times m)$   $H \in \mathcal{O}_m$ ,

$$P(X) = P(HXH'). \quad (3.2.3)$$

**Proposition 3.2.1.** Pour que le polynôme  $P(X)$  de la matrice symétrique  $X$  soit invariant, il faut et il suffit qu'il soit un polynôme symétrique des valeurs propres  $y_1, \dots, y_m$  de  $X$ .

**Preuve.** Supposons tout d'abord  $P(X)$  invariant. Comme  $X$  est  $(m \times m)$  symétrique, il existe toujours une matrice orthogonale  $H$  telle que  $H'XH = \text{diag}(y_1, \dots, y_m)$ . Par (3.2.3), on a donc

$$P(X) = P(\text{diag}(y_1, \dots, y_m)),$$

expression qu'on notera, par abus de langage  $P(y_1, \dots, y_m)$ . En notant  $H = [h_1 \ \dots \ h_m]$  on constate que les colonnes  $h_1, \dots, h_m$  de  $H$  sont vecteurs propres de  $X$  associés aux valeurs propres  $y_1, \dots, y_m$  de  $X$ . Pour

toute permutation  $\pi = \{\pi_1, \dots, \pi_m\}$  de  $\{1, \dots, n\}$ , la matrice  $H_\pi = [h_{\pi_1} \ \dots \ h_{\pi_m}] \in \mathcal{O}_m$  est donc telle que  $H'_\pi X H_\pi = \text{diag}(y_{\pi_1}, \dots, y_{\pi_m})$ . Comme, par (3.2.3),

$$P(y_1, \dots, y_m) = P(X) = P(H'_\pi X H_\pi) = P(y_{\pi_1}, \dots, y_{\pi_m}),$$

on en déduit que  $P$  ne peut être qu'un polynôme symétrique des variables  $y_1, \dots, y_m$ .

Inversement, si le polynôme  $P(X)$  est un polynôme symétrique des valeurs propres  $y_1, \dots, y_m$  de  $X$ , il est nécessairement invariant, puisque les valeurs propres  $y_1, \dots, y_m$  de  $X$  sont aussi celles de  $H'XH$  pour tout  $H \in \mathcal{O}_m$ .  $\square$

**Définition 3.2.2.** Nous dirons que le polynôme  $P(Z)$  de la matrice carrée (pas nécessairement symétrique)  $(m \times m)$   $Z$  est invariant, si pour toute matrice régulière  $(m \times m)$   $L \in \mathfrak{GL}(m, \mathbb{C})$ ,

$$P(Z) = P(LZL^{-1}). \quad (3.2.4)$$

**Proposition 3.2.2.** Pour que le polynôme  $P(Z)$  de la matrice carrée  $Z$  soit invariant, il faut et il suffit qu'il soit un polynôme symétrique des valeurs propres  $z_1, \dots, z_m$  de  $Z$ .

**Preuve.** Considérons un polynôme  $P(Z)$  des coefficients de la matrice  $(m \times m)$  (non nécessairement symétrique)  $Z$ . La matrice  $Z$  étant donnée, il existe toujours une matrice régulière  $L$  à coefficients dans  $\mathbb{C}$  telle que

$$LXL^{-1} = \begin{bmatrix} D_1 & \dots & \mathbb{O} \\ \vdots & \ddots & \vdots \\ \mathbb{O} & \dots & D_r \end{bmatrix},$$

se mette sous la forme d'une matrice réduite de Jordan, où, pour  $1 \leq j \leq r$ ,  $D_j$  est de la forme

$$D_j = \begin{bmatrix} z & 1 & 0 & \dots & 0 \\ 0 & z & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & z & 1 \\ 0 & 0 & \dots & 0 & z \end{bmatrix},$$

$z \in \{z_1, \dots, z_m\} \in \mathbb{C}^m$  étant une valeur propre de  $Z$ . Par (3.2.4), il s'ensuit que  $P(Z)$  est alors nécessairement un polynôme des valeurs propres  $z_1, \dots, z_m$  de  $X$ , de la forme

$$P(Z) = P(\text{diag}(D_1, \dots, D_r)).$$

Comme  $Z$  n'est pas nécessairement diagonalisable, on ne peut pas répéter sans modification le raisonnement utilisé plus haut (dans la démonstration de la proposition (3.2.1)) pour montrer que  $P(z_1, \dots, z_m)$  doit être symétrique en  $z_1, \dots, z_m$ . Toutefois, on obtient que ce résultat demeure valide en perturbant  $Z$  en

$$Z_\varepsilon = Z + \varepsilon \text{diag}(\lambda_1, \dots, \lambda_m),$$

où  $\lambda_1, \dots, \lambda_m$  sont choisis de manière à ce que les valeurs propres  $z_1 + \varepsilon\lambda_1, \dots, z_m + \varepsilon\lambda_m$  de  $Z_\varepsilon$  soient distinctes (auquel cas  $Z_\varepsilon$  est diagonalisable) pour tout  $|\varepsilon| \leq \varepsilon_0$  suffisamment petit. On obtient alors que  $P(Z_\varepsilon)$  doit être un polynôme symétrique en  $z_1 + \varepsilon\lambda_1, \dots, z_m + \varepsilon\lambda_m$  pour tout  $|\varepsilon| \leq \varepsilon_0$ . Comme  $Z_\varepsilon \rightarrow Z$  lorsque  $\varepsilon \rightarrow 0$ , le fait que  $P(Z_\varepsilon)$  tende vers  $P(Z)$  lorsque  $\varepsilon \rightarrow 0$  permet alors de conclure que  $P(z)$  est symétrique en  $z_1, \dots, z_m$ .

Inversement, si  $P(Z)$  est un polynôme symétrique des valeurs propres de  $Z$ , on a nécessairement (3.2.4) pour toute matrice  $L \in \mathfrak{GL}(m, \mathbb{C})$ .  $\square$

**Exemple 3.2.1.** Si  $X$  est une matrice carrée  $(m \times m)$  de valeurs propres  $y_1, \dots, y_m$ , pour tout couple d'entiers  $k \geq 0$  et  $r \geq 0$ , les fonctions suivantes de  $y_1, \dots, y_m$  sont des polynômes invariants de la matrice  $X$ .

$$\text{tr}(X)^k = (y_1 + \dots + y_m)^k \quad \text{et} \quad \det(X)^r = y_1^r \dots y_m^r. \quad (3.2.5)$$



D'une manière générale, si  $Z$  est une matrice  $(m \times m)$  quelconque (et non plus symétrique), les fonctions symétriques élémentaires des valeurs propres  $z_1, \dots, z_r$  de  $Z$ , définies pour  $0 \leq r \leq m$  par

$$a_r(z_1, \dots, z_r) = \sum_{1 \leq i_1 < \dots < i_r \leq m} z_{i_1} \dots z_{i_r} \quad \text{et} \quad a_0 = 1, \quad (3.2.6)$$

s'expriment de manière simple en fonction des coefficients du polynôme caractéristique de  $z$ . On a, en effet,

$$P_Z(z) = \det(Z - z\mathbb{I}_m) = (-1)^m \prod_{j=1}^m (z - z_j) = (-1)^m \sum_{r=0}^m (-1)^r a_r z^{m-r}. \quad (3.2.7)$$

Comme le polynôme caractéristique de  $Z$  est invariant par changement de base, ou sens que

$$P_Z(z) = P_{L^{-1}ZL}(z),$$

pour toute matrice  $L \in \mathcal{GL}(m, \mathbb{C})$ ,  $(m \times m)$  inversible, on vérifie que  $a_0, \dots, a_m$  sont, d'une part, des polynômes (à coefficients réels) des coefficients  $z_{ij}$  de  $Z = [z_{ij}]$ , et d'autre part, des fonctions symétriques des valeurs propres  $z_1, \dots, z_m$  de  $Z$ . On obtient ainsi des exemples simples de polynômes invariants de  $Z$ . Cette construction est générale, et est résumée dans la proposition ci-dessous.

**Proposition 3.2.3.** *A tout polynôme invariant  $P(X)$  de la matrice  $(m \times m)$  symétrique  $X$  est associé une unique polynôme invariant  $\mathcal{P}(Z)$  de la matrice  $(m \times m)$  carrée  $Z$  tel que  $\mathcal{P}(X) = P(X)$  soit une identité en  $X$ .*

**Preuve.** A l'aide des propositions 3.2.1 et 3.2.4, on se ramène au cas où  $\mathcal{P}(Z)$  et  $P(X)$  sont des polynômes homogènes symétriques des valeurs propres  $z_1, \dots, z_m$  de  $Z$  et  $y_1, \dots, y_m$  de  $X$ , respectivement. Posons  $\alpha_r = a_r(z_1, \dots, z_m)$  et  $a_r = a_r(x_1, \dots, x_m)$  pour  $r = 1, \dots, m$ . Par la proposition 8.1, il existe des polynômes  $R(a_1, \dots, a_m)$  et  $\mathcal{R}(\alpha_1, \dots, \alpha_m)$  tels que  $\mathcal{P}(Z) = \mathcal{R}(\alpha_1, \dots, \alpha_m)$  et  $P(X) = R(a_1, \dots, a_m)$ . Par l'identité  $\mathcal{P}(X) = P(X)$ , valable pour toute matrice  $(m \times m)$  symétrique  $X$ , on a nécessairement l'identité de polynômes  $\mathcal{R}(a_1, \dots, a_m) = R(a_1, \dots, a_m)$ . Celle-ci implique que  $\mathcal{R}(\cdot) = R(\cdot)$ , et donc que  $\mathcal{P}(\cdot) = P(\cdot)$ .  $\square$

Compte tenu de la proposition 3.2.3, dans ce qui suit, nous pourrions supposer indifféremment que les polynômes symétriques invariants des valeurs propres de matrices  $P(\cdot)$  considérés sont définis pour des matrices  $X$  symétriques  $(m \times m)$  ou pour des matrices  $Z$  carrées  $(m \times m)$  quelconques.

**Exemple 3.2.2.** Considérons les matrices  $(2 \times 2)$

$$X = \begin{bmatrix} a & b \\ b & d \end{bmatrix} \quad \text{et} \quad Z = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Le polynôme

$$P(X) = P(a, b, d) = ad - b^2 = \det(X) = y_1 y_2 = a_2(y_1, y_2),$$

de la matrice symétrique  $X$ , est, bien évidemment invariant. Sa version définie sur la matrice carrée  $Z$  est donnée par

$$\mathcal{P}(Z) = \mathcal{P}(a, b, c, d) = ad - bc = \det(Z) = z_1 z_2 = a_2(z_1, z_2).$$

Dans la suite, on utilisera souvent, par abus de langage, la même notation pour  $\mathcal{P}$  et  $P$ .

Nous considérons maintenant l'ensemble  $V_k$  des polynômes homogènes symétriques des valeurs propres  $y_1, \dots, y_m$  de la matrice  $(m \times m)$  symétrique  $X$ , et de degré égal soit à  $k$ , soit à  $-\infty$  (dans le cas du polynôme nul  $\mathbb{O}$ ). Rappelant la définition (3.1.11) du monôme homogène symétrique  $M_\kappa(y_1, \dots, y_m)$  associé à la  $m$ -partition  $\kappa = (k_1, \dots, k_m) \in \mathcal{P}_m(k)$  de  $k$ , pour toute matrice symétrique  $X$  de valeurs propres  $y_1, \dots, y_m$ , on posera

$$M_\kappa(X) = M_\kappa(y_1, \dots, y_m) = \sum_{\{\pi_1, \dots, \pi_m\} \in \mathcal{P}_m} y_{\pi_1}^{k_1} \dots y_{\pi_m}^{k_m}. \quad (3.2.8)$$

**Proposition 3.2.4.** *Pour tout entier  $k \geq 0$ ,  $V_k$  est un espace vectoriel, engendré par les monômes symétriques  $M_\kappa(X)$ , lorsque  $\kappa$  varie dans l'ensemble  $\mathcal{P}_m(k)$  des  $m$ -partitions de  $k$ . De plus, les  $M_\kappa(X)$ , pour  $\kappa \in \mathcal{P}_m(k)$ , forment une base de  $V_k$ .*

**Preuve.** Un polynôme  $P(X) = P(y_1, \dots, y_m)$  appartient à  $V_k$  s'il vérifie l'identité

$$P(\lambda y_1, \dots, \lambda y_m) = \lambda^k P(y_1, \dots, y_m). \tag{3.2.9}$$

Il est clair que si  $P_1$  et  $P_2$  vérifient (3.2.9), il en est de même pour  $\lambda_1 P_1 + \lambda_2 P_2$  pour tout choix de  $\lambda_1, \lambda_2 \in \mathbb{R}$ . Il est clair que les  $M_\kappa$  engendrent  $V_k$  lorsque  $\kappa$  varie dans l'ensemble des  $m$ -partitions de  $k$ . Comme les  $M_\kappa$  forment une famille libre, on a donc le résultat annoncé.  $\square$

**Définition 3.2.3.** *Soit  $X$  une matrice symétrique ( $m \times m$ ), de valeurs propres  $y_1, \dots, y_m$ . On appelle polynômes zonaux associés à  $X$  les polynômes  $C_\kappa(X)$  des variables  $y_1, \dots, y_m$  vérifiant les propriétés suivantes.*

(i) *Pour tout entier  $k \geq 0$  et  $\kappa = (k_1, \dots, k_m) \in \mathcal{P}_m(k)$ ,  $m$ -partition de  $k$ ,  $C_\kappa(X)$  est un polynôme homogène symétrique de degré  $k$  en  $y_1, \dots, y_m$  dont le monôme de plus grand poids est proportionnel à  $y_1^{k_1} \dots y_m^{k_m}$ .*

(ii) *Pour tout entier  $k \geq 0$  et  $\kappa = (k_1, \dots, k_m)$ ,  $m$ -partition de  $k$ ,  $C_\kappa(X)$  est une fonction propre de l'opérateur différentiel  $\Delta_X$  défini par*

$$\Delta_X = \sum_{i=1}^m y_i^2 \frac{\partial^2}{\partial y_i^2} + \sum_{i=1}^m \sum_{\substack{j=1 \\ j \neq i}}^m \frac{y_i^2}{y_i - y_j} \frac{\partial}{\partial y_i}. \tag{3.2.10}$$

(iii) *Pour tout entier  $k \geq 0$ , on a l'identité*

$$(\text{tr } X)^k = (y_1 + \dots + y_m)^k = \sum_{\kappa \in \mathcal{P}_m(k)} C_\kappa(X), \tag{3.2.11}$$

où  $\kappa \in \mathcal{P}_m(k)$  varie dans toutes les  $m$ -partitions de  $k$ .

**Remarque 3.2.1.** La définition du polynôme zonal  $C_\kappa(X)$  ci-dessus est due à Constantine, A.G. (1963). Some noncentral distribution problems in multivariate analysis. *Ann. Math. Statist.* **34** 1270–1285. Elle diffère de la définition originale, notée  $Z_\kappa(X)$ , due à James, A.T. (1961). Zonal polynomials of the real positive definite symmetric matrices. *Annals of Mathematics.* **74** 456–469, par une constante multiplicative. On notera, pour mémoire (voir (4.3.39) et (4.3.53) dans Mathai, A.M., Provost, S.B. et Hayakawa, T. (1995). *Bilinear Forms and Zonal Polynomials*. Lecture Notes in Statistics **102** Springer-Verlag, New York.), que, si  $\kappa = (k_1, \dots, k_m)$  est une  $m$ -partition de  $k$  avec  $k_m \geq 1$ ,

$$C_\kappa(X) = c_\kappa Z_\kappa(X), \tag{3.2.12}$$

où

$$c_\kappa = \frac{2^k k! \prod_{1 \leq i < j \leq m} (2k_i - 2k_j - i + j)!}{\prod_{i=1}^m (2k_i + m - i)!} \tag{3.2.13}$$

**Remarque 3.2.2.** Nous admettons que les propriétés énoncées dans la définition 3.2.3 ci-dessus sont compatibles, dans la mesure où elles impliquent, par le biais d'une construction par récurrence, l'existence et l'unicité des polynômes  $C_\kappa(X)$ . Nous nous référons à Saw, J.G. (1977). Zonal polynomials : An alternative approach. *Journal of Multivariate Analysis.* **7** 461–467, pour la description d'algorithmes permettant de calculer les coefficients de ces polynômes, et nous limitons, pour l'essentiel à l'étude de quelques exemples.

**Exemple 3.2.3.** Le cas le plus simple de polynôme zonal est obtenu pour  $m = 1$ , auquel cas la matrice  $X$  est de la forme  $X = [y]$ , de valeur propre unique  $y \in \mathbb{R}$ . La seule  $m$ -partition de  $k$  pour  $m = 1$  est  $\kappa = (k)$  unique élément de  $\mathcal{P}_1(k)$ , et la formule (3.2.11) mène trivialement à

$$C_{(k)}(X) = y^k \quad \text{pour tout } k \geq 0. \tag{3.2.14}$$

On vérifie sans peine que, pour tout entier  $k \in \mathbb{N}$ ,  $C_{(k)}(X)$  est valeur propre, associée à la valeur propre  $k(k-1)$ , de l'opérateur différentiel  $\Delta_X$  de (3.2.10), qui se réduit ici à  $\Delta_X = y^2 \frac{d^2}{dy^2}$ .

**Exemple 3.2.4.** Pour  $k = 0$ , il existe une  $m$ -partition unique de 0 donnée par  $\mathbb{0} = (0, \dots, 0) = (0^m)$ . De ce fait, avec la convention habituelle que  $y^0 = 1 \forall y$ , (3.2.11) définit de manière unique

$$C_{(0^m)}(X) = 1 \quad \text{pour} \quad \kappa = (0^m) = (0, \dots, 0). \quad (3.2.15)$$

On vérifie aisément que les propriétés (i) et (ii) de la définition 3.2.3 sont satisfaites par  $C_{(0^m)}(X)$  tel qu'il est défini en (3.2.15).

Dans le cas général, les valeurs propres de l'opérateur différentiel  $\Delta_X$  associées aux polynômes zonaux  $C_\kappa(X)$ , sont données par le théorème suivant.

**Théorème 3.2.1.** *Pour tout entier  $k \geq 1$ , et pour toute matrice  $X$ , symétrique et  $(m \times m)$ , le polynôme zonal  $C_\kappa(X)$  associé à la  $m$ -partition  $\kappa = (k_1, \dots, k_m)$  de  $k$  satisfait l'identité*

$$\Delta_X C_\kappa(X) = \{\rho_\kappa + k(m-1)\} C_\kappa(X), \quad (3.2.16)$$

où  $\Delta_X$  est comme en (3.2.10), et

$$\rho_\kappa = \sum_{i=1}^m k_i(k_i - i). \quad (3.2.17)$$

**Preuve.** Compte tenu de la définition 3.2.3, il suffit de vérifier que

$$\Delta_X (y_1^{k_1} \dots y_m^{k_m}) = \{\rho_\kappa + k(m-1)\} y_1^{k_1} \dots y_m^{k_m} + (\text{termes d'ordre ou de poids inférieur}).$$

Or, un calcul direct basé sur (3.2.10) montre que, comme  $k_1 + \dots + k_m = k$ ,

$$\begin{aligned} \Delta_X (y_1^{k_1} \dots y_m^{k_m}) &= y_1^{k_1} \dots y_m^{k_m} \left\{ \sum_{i=1}^m k_i(k_i - 1) + \sum_{i=1}^m \sum_{\substack{j=1 \\ j \neq i}}^m \frac{y_i k_i}{y_i - y_j} \right\} \\ &= y_1^{k_1} \dots y_m^{k_m} \left\{ \sum_{i=1}^{m-1} k_i^2 - k + \sum_{i=1}^{m-1} \sum_{j=i+1}^m \left\{ \frac{y_i k_i}{y_i - y_j} + \frac{y_j k_j}{y_j - y_i} \right\} \right\} \\ &= y_1^{k_1} \dots y_m^{k_m} \left\{ \sum_{i=1}^m k_i^2 - k + \sum_{i=1}^{m-1} \sum_{j=i+1}^m \left\{ k_i + \frac{y_j}{y_j - y_i} (k_j - k_i) \right\} \right\} \\ &= y_1^{k_1} \dots y_m^{k_m} \left\{ \sum_{i=1}^m k_i^2 - k + \sum_{i=1}^m k_i(m-i) \right\} + (\text{termes d'ordre ou de poids inférieur}) \\ &= y_1^{k_1} \dots y_m^{k_m} \left\{ \sum_{i=1}^m k_i(k_i - i) + k(m-1) \right\} + (\text{termes d'ordre ou de poids inférieur}). \end{aligned}$$

Cette identité permet de conclure.  $\square$

La nature de l'opérateur  $\Delta_X$  intervenant en (3.2.10), dans la définition 3.2.3 des polynômes zonaux, peut sembler assez mystérieuse à première vue. Son intérêt deviendra évident par la suite, grâce à l'introduction de l'opérateur différentiel de Laplace–Beltrami  $\Delta_X^*$ , qui lui est très voisin et que nous décrivons maintenant.

Soit  $X = [x_{i,j}] > 0$  une matrice  $(m \times m)$ , symétrique et définie positive. On lui associe sa différentielle  $dX = [dx_{i,j}]$ , ainsi que la forme quadratique différentielle, notée  $(ds)^2$  et définie par

$$(ds)^2 = \text{tr} \left( X^{-1} dX \cdot X^{-1} dX \right) = \text{tr} \left( \{ X^{-1} dX \}^2 \right). \quad (3.2.18)$$

La forme quadratique différentielle  $(ds)^2$  est, à l'évidence, une forme quadratique en  $d\mathcal{X}$ , où le vecteur  $\mathcal{X} = V(X) \in \mathbb{R}^n$ , avec  $n = \frac{1}{2}m(m+1)$ , est défini par

$$\mathcal{X} = V(X) = [x_{11} \ \cdots \ x_{1,m} \ x_{22} \ \cdots \ x_{2,m} \ \cdots \ x_{m-1,m-1} \ x_{m-1,m} \ x_{m,m}]' \in \mathbb{R}^n. \quad (3.2.19)$$

On peut donc écrire  $(ds)^2$  sous la forme

$$(ds)^2 = \text{tr}\left(\{X^{-1}dX\}^2\right) = d\mathcal{X}'G(\mathcal{X})d\mathcal{X}. \quad (3.2.20)$$

où  $G(\mathcal{X})$  désigne une matrice  $(n \times n)$ .

**Lemme 3.2.1.** *La matrice  $G(\mathcal{X})$  définie par l'identité  $(ds)^2 = \text{tr}\left(\{X^{-1}dX\}^2\right) = d\mathcal{X}'G(\mathcal{X})d\mathcal{X}$ , est symétrique et définie positive.*

**Preuve.** A l'évidence,  $(ds)^2$  est une forme quadratique en  $\{dx_{ij} : 1 \leq i \leq j \leq m\}$ . La matrice  $G(\mathcal{X})$  définie par (3.2.20) est donc nécessairement symétrique. Considérons maintenant une matrice  $(m \times m)$  symétrique  $X$  quelconque. Il existe toujours une matrice orthogonale  $H$  et une matrice diagonale  $C$  telles que  $X = HCH'$ . Lorsque, de plus,  $X$  est régulière, on a  $X^{-1} = HC^{-1}H'$ . Comme, par ailleurs,

$$dX = H.dC.H' + dH.C.H' + H.C.dH',$$

on constate que

$$\begin{aligned} X^{-1}dX &= HC^{-1}.dC.H' + HC^{-1}H'.dH.C.H' + HC^{-1}H'HC.dH' \\ &= HC^{-1}.dC.H', \end{aligned}$$

où l'on a fait usage des égalités  $H'.dH = H.dH' = \mathbb{O}$ , obtenues en différentiant  $HH' = H'H = \mathbb{I}$ . On en déduit, en posant  $C = \text{diag}(c_1, \dots, c_m)$ , que

$$\begin{aligned} \text{tr}\left(\{X^{-1}dX\}^2\right) &= \text{tr}\left(\{HC^{-1}.dC.H'\}^2\right) = \text{tr}\left(H\{C^{-1}.dC\}^2H'\right) \\ &= \text{tr}\left(\{C^{-1}.dC\}^2H'H\right) = \text{tr}\left(\{C^{-1}.dC\}^2\right) \\ &= \sum_{i=1}^m \{c_i^{-1}dc_i\}^2. \end{aligned} \quad (3.2.21)$$

On déduit de cette dernière expression que  $G(\mathcal{X}) \geq 0$  est une matrice symétrique positive. On déduit un autre résultat utile de (3.2.21). Il ressort en effet de cette expression que la transformation de  $X$  en  $C$  via  $X = HCH'$ , et la transformation de  $dX$  en  $dC$  via  $H.dC.H'$ , transforment  $\text{tr}\left(\{X^{-1}dX\}^2\right)$  en  $\text{tr}\left(\{C^{-1}.dC\}^2\right)$ . Pour la construction de  $G(\mathcal{X})$ , tout se passe donc comme si  $X$  et  $dX$  étaient simultanément diagonalisables par une transformation orthogonale. Pour montrer que  $G(\mathcal{X})$  est définie positive, on se sert précisément de cette propriété, dans un cadre sensiblement plus général.

On procède en considérant deux matrices symétriques définies positives  $Z$  et  $Y$  quelconques pouvant être simultanément diagonalisées dans une même base orthonormée (cette propriété n'est en général pas satisfaite pour  $Z$  et  $Y$  arbitraires). Ceci signifie qu'il existe une matrice orthogonale  $H$  telle que, pour des matrices diagonales convenables  $C = \text{diag}(c_1, \dots, c_m)$  et  $D = \text{diag}(d_1, \dots, d_m)$ , avec  $c_1, \dots, c_m > 0$  et  $d_1, \dots, d_m > 0$ , on ait  $Z = HCH'$  et  $Y = HDH'$ . Comme  $HH' = H'H = \mathbb{I}$ , ceci implique que

$$Z(Y - Z)Z(Y - Z) = H\text{diag}\left(c_1^2(d_1 - c_1)^2, \dots, c_m^2(d_m - c_m)^2\right)H'.$$

Par conséquent,  $\text{tr}(Z(Y - Z)Z(Y - Z)) = \sum_{i=1}^m c_i^2(d_i - c_i)^2 = 0$  si et seulement si  $d_i = c_i$  pour  $i = 1, \dots, m$ , c'est à dire, si  $Z = Y$ . En posant  $Y = Z + dX$ , et  $Z = X^{-1}$ , on obtient que  $\text{tr}(X^{-1}dX \cdot X^{-1}dX) = 0$  n'est possible que si  $X^{-1} = X^{-1} + dX$ , ce qui est, bien évidemment, exclu. On obtient ainsi le résultat annoncé.  $\square$

**Exemple 3.2.5.** Il est intéressant de donner la forme explicite de  $G(\mathbf{X})$  dans les cas particuliers  $m = 1$  et  $m = 2$ . Pour  $m = 1$ , on a  $X = [x] = \mathbf{X}$  avec  $x > 0$ . Par conséquent, on déduit de (3.2.18) que

$$(ds)^2 = x^{-2}(dx)^2 \quad \text{et} \quad G(\mathbf{X}) = [x^{-2}]. \quad (3.2.22)$$

Pour  $m = 2$ , on constate de même que

$$\begin{aligned} (ds)^2 &= \frac{1}{(x_{11}x_{22} - x_{12}^2)^2} \operatorname{tr} \left\{ \begin{bmatrix} x_{22} & -x_{12} \\ -x_{12} & x_{11} \end{bmatrix} \begin{bmatrix} dx_{11} & dx_{12} \\ dx_{12} & dx_{22} \end{bmatrix} \begin{bmatrix} x_{22} & -x_{12} \\ -x_{12} & x_{11} \end{bmatrix} \begin{bmatrix} dx_{11} & dx_{12} \\ dx_{12} & dx_{22} \end{bmatrix} \right\} \\ &= \frac{1}{(x_{11}x_{22} - x_{12}^2)^2} \left( x_{11}^2 dx_{22}^2 + x_{22}^2 dx_{11}^2 - 2x_{12}^2 dx_{11} dx_{22} \right. \\ &\quad \left. + 2(x_{12}^2 + x_{11}x_{22}) dx_{12}^2 - 4x_{12}x_{22} dx_{12} dx_{11} - 4x_{12}x_{11} dx_{12} dx_{22} \right) \\ &= [dx_{11} \quad dx_{12} \quad dx_{22}] G(\mathbf{X}) \begin{bmatrix} dx_{11} \\ dx_{12} \\ dx_{22} \end{bmatrix}, \end{aligned} \quad (3.2.23)$$

avec

$$G(\mathbf{X}) = \frac{1}{(x_{11}x_{22} - x_{12}^2)^2} \begin{bmatrix} x_{22}^2 & -2x_{12}x_{22} & x_{12}^2 \\ -2x_{12}x_{22} & 2(x_{12}^2 + x_{11}x_{22}) & -2x_{12}x_{11} \\ x_{12}^2 & -2x_{12}x_{11} & x_{11}^2 \end{bmatrix}.$$

Le lemme suivant établit une propriété d'invariance de  $(ds)^2$ , en tant que fonction de  $X$ . Nous notons, conformément à l'usage,  $\mathcal{GL}(m) = \mathcal{GL}(m, \mathbb{R})$  le *groupe linéaire d'ordre  $m$* , ensemble des matrices réelles, régulières,  $(m \times m)$ , muni du produit matriciel usuel. Nous faisons opérer ce groupe sur l'ensemble  $\mathcal{S}_m^+$  des matrices symétriques  $(m \times m)$  définies positives, par la transformation

$$L \in \mathcal{GL}(m, \mathbb{R}) : X \in \mathcal{S}_m^+ \rightarrow Z = LXL' \in \mathcal{S}_m^+. \quad (3.2.24)$$

**Lemme 3.2.2.** *Pour toute matrice  $(m \times m)$  régulière  $L \in \mathcal{GL}(m, \mathbb{R})$ , la transformation*

$$X \in \mathcal{S}_m^+ \rightarrow Z = LXL',$$

*laisse invariante la forme différentielle  $(ds)^2 = \operatorname{tr}(\{X^{-1} dX\}^2)$ .*

**Preuve.** Compte tenu de la définition (3.2.18) de  $(ds)^2$ , il suffit de vérifier l'identité (triviale par le fait que  $L \in \mathcal{GL}(m)$  est inversible)

$$(ds)^2 = \operatorname{tr}(X^{-1} \cdot dX \cdot X^{-1} \cdot dX) = \operatorname{tr}(\{LXL'\}^{-1} d\{LXL'\} \{LXL'\}^{-1} d\{LXL'\}),$$

pour toute matrice  $L \in \mathcal{GL}(m, \mathbb{R})$ .  $\square$

Nous avons maintenant à notre disposition tous les outils nécessaires pour définir l'*opérateur différentiel de Laplace-Beltrami*. Il sera commode de faire usage à cet effet de la définition (3.2.18) de la matrice  $G(\mathbf{X})$  associée à  $X$ , par l'intermédiaire de  $\mathbf{X} = V(X)$ . Nous adopterons également la notation matricielle

$$\frac{\partial}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix},$$

où  $n = \frac{1}{2}m(m+1)$  et  $\mathbf{X}$  est comme en (3.2.19). Rappelons la définition (3.2.20) de  $G(\mathbf{X})$ .

**Définition 3.2.4.** Pour toute matrice  $X = [x_{ij}] > 0$ , ( $m \times m$ ), symétrique et définie positive, l'opérateur différentiel de Laplace–Beltrami est défini par

$$\begin{aligned}\Delta_X^* &= \{\det G(\mathbf{X})\}^{-1/2} \sum_{j=1}^n \frac{\partial}{\partial x_j} \left\{ \{\det G(\mathbf{X})\}^{1/2} \sum_{i=1}^n g(\mathbf{X})^{ij} \frac{\partial}{\partial x_i} \right\} \\ &= \{\det G(\mathbf{X})\}^{-1/2} \left[ \frac{\partial}{\partial \mathbf{X}} \right]' \left\{ \{\det G(\mathbf{X})\}^{1/2} G(\mathbf{X})^{-1} \frac{\partial}{\partial \mathbf{X}} \right\},\end{aligned}\quad (3.2.25)$$

où  $G(\mathbf{X})^{-1} = [g(\mathbf{X})^{ij}]$ , et  $n = \frac{1}{2}m(m+1)$ .

On remarquera que cette définition est rendue possible par le lemme 3.2.1, qui établit que  $G(\mathbf{X}) > 0$ . Cette dernière propriété donne un sens, aussi bien à  $G(\mathbf{X})^{-1}$  qu'à  $\{\det G(\mathbf{X})\}^{-1/2} > 0$ .

L'intérêt principal de l'opérateur  $\Delta_X^*$  est d'être, à l'instar de  $(ds)^2$ , *invariant* par la transformation (3.2.24). Cette propriété est établie dans le lemme suivant.

**Lemme 3.2.3.** Pour toute matrice ( $m \times m$ ) régulière  $L \in \mathfrak{GL}(m, \mathbb{R})$ , la transformation

$$X \in \mathcal{S}_m^+ \rightarrow Z = LXL' \in \mathcal{S}_m^+,$$

laisse invariant l'opérateur  $\Delta_X^*$ . On a donc

$$\Delta_X^* = \Delta_Z^* = \Delta_{LXL'}^*. \quad (3.2.26)$$

**Preuve.** Comme  $L \in \mathfrak{GL}(m, \mathbb{R})$ , la transformation  $X \rightarrow Z = LXL'$  est inversible et linéaire en  $X$ . Par conséquent, il existe une matrice ( $n \times n$ ) régulière  $T_L$  telle que, si  $\mathbf{X} = V(X)$  et  $\mathbf{Z} = V(Z)$  sont comme en (3.2.19), on ait

$$\mathbf{Z} = T_L \mathbf{X}.$$

Compte tenu du Lemme 7.2, cette dernière propriété, jointe à la définition (3.2.18) de  $G(\mathbf{X})$ , implique que

$$(ds)^2 = d\mathbf{X}' G(\mathbf{X}) d\mathbf{X} = d\mathbf{Z}' G(\mathbf{Z}) d\mathbf{Z} = d(T_L \mathbf{X})' G(T_L \mathbf{X}) d(T_L \mathbf{X}) = d\mathbf{X}' \left( T_L' G(T_L \mathbf{X}) T_L \right) d\mathbf{X}.$$

On en déduit les identités

$$G(\mathbf{X}) = T_L' G(T_L \mathbf{X}) T_L \quad \text{et} \quad G(\mathbf{Z}) = G(T_L \mathbf{X}) = T_L'^{-1} G(\mathbf{X}) T_L^{-1}.$$

On remplace maintenant  $\mathbf{X}$  par  $\mathbf{Z}$  dans (3.2.25), et on utilise la relation  $G(\mathbf{Z}) = G(T_L \mathbf{X}) = T_L'^{-1} G(\mathbf{X}) T_L^{-1}$  ci-dessus. En faisant usage des relations évidentes

$$\frac{\partial}{\partial \mathbf{X}} = T_L' \frac{\partial}{\partial \mathbf{Z}} \quad \text{et} \quad \frac{\partial}{\partial \mathbf{Z}} = T_L'^{-1} \frac{\partial}{\partial \mathbf{X}},$$

et de  $Z = LXL'$ , on en déduit les égalités

$$\begin{aligned}\Delta_Z^* &= \Delta_{LXL'}^* = \{\det G(\mathbf{Z})\}^{-1/2} \left[ \frac{\partial}{\partial \mathbf{Z}} \right]' \left\{ \{\det G(\mathbf{Z})\}^{1/2} G(\mathbf{Z})^{-1} \frac{\partial}{\partial \mathbf{Z}} \right\} \\ &= \{\det G(T_L \mathbf{X})\}^{-1/2} \left[ \frac{\partial}{\partial \mathbf{X}} \right]' T_L^{-1} \left\{ \{\det G(T_L \mathbf{X})\}^{1/2} G(T_L \mathbf{X})^{-1} T_L'^{-1} \frac{\partial}{\partial \mathbf{X}} \right\} \\ &= \{\det T_L\} \{\det G(\mathbf{X})\}^{-1/2} \left[ \frac{\partial}{\partial \mathbf{X}} \right]' T_L^{-1} \\ &\quad \times \left\{ \{\det T_L\}^{-1} \{\det G(\mathbf{X})\}^{1/2} T_L G(\mathbf{X})^{-1} T_L' T_L^{-1} \frac{\partial}{\partial \mathbf{X}} \right\} \\ &= \{\det G(\mathbf{X})\}^{-1/2} \left[ \frac{\partial}{\partial \mathbf{X}} \right]' \left\{ \{\det G(\mathbf{X})\}^{1/2} G(\mathbf{X})^{-1} \frac{\partial}{\partial \mathbf{X}} \right\} = \Delta_X^*,\end{aligned}\quad (3.2.27)$$

ce qu'il fallait démontrer.  $\square$

**Remarque 3.2.3.** Dans la démonstration du lemme 3.2.3, la matrice  $L \in \mathfrak{GL}(m, \mathbb{R})$  intervenant dans l'identité  $Z = LXL'$  est supposée constante, au sens qu'elle ne dépend pas de  $X$ . Ceci implique que la matrice  $T_L$  intervenant dans l'identité  $\mathfrak{Z} = T_L \mathfrak{X}$  est constante, et donc que  $d\mathfrak{Z} = T_L d\mathfrak{X}$ . Le lemme suivant donne une version du lemme 8.3 permettant de traiter le cas plus général où  $\mathfrak{Z} = V(Z)$  dépend de  $\mathfrak{X} = V(X)$  par une relation non linéaire.

**Lemme 3.2.4.** *Supposons que  $\mathfrak{Z}_1 \in \mathbb{R}^n$  soit une fonction de  $\mathfrak{X} = V(X)$  dont la différentielle  $d\mathfrak{Z}_1 = T_1 d\mathfrak{X}$  soit définie par une matrice  $(n \times n)$   $T_1 = T_1(\mathfrak{X})$  régulière. Supposons de plus que  $\mathfrak{Z}_1 = [\mathfrak{Y} \ \alpha]'$ ,  $\mathfrak{Y} \in \mathbb{R}^{n-q}$ ,  $\alpha \in \mathbb{R}^q$ , et notons  $d\mathfrak{Z} = [d\mathfrak{Y} \ d\beta] = [d\mathfrak{Y} \ R(\alpha)d\alpha]$ , où  $R = R(\alpha)$  est une matrice  $(q \times q)$  régulière. Notons enfin par  $T = T(\mathfrak{X})$  la matrice  $(n \times n)$  régulière définie par  $d\mathfrak{Z} = T(d\mathfrak{X})$ , de sorte que*

$$T = \begin{bmatrix} \mathbb{I} & \mathbb{O} \\ \mathbb{O} & R(\alpha) \end{bmatrix} T_1.$$

Supposons enfin qu'un puisse définir une matrice  $(n \times n)$  symétrique  $\mathcal{G}(\mathfrak{Y})$  par l'identité

$$\mathcal{G}(\mathfrak{Y}) = T'^{-1} G(\mathfrak{X}) T^{-1},$$

ou, de manière équivalente

$$(ds)^2 = d\mathfrak{X}' G(\mathfrak{X}) d\mathfrak{X} = d\mathfrak{Z}' \mathcal{G}(\mathfrak{Y}) d\mathfrak{Z}. \quad (3.2.28)$$

Alors, l'opérateur différentiel

$$\Delta_X^* = \{\det G(\mathfrak{X})\}^{-1/2} \left[ \frac{\partial}{\partial \mathfrak{X}} \right]' \left\{ \{\det G(\mathfrak{X})\}^{1/2} G(\mathfrak{X})^{-1} \frac{\partial}{\partial \mathfrak{X}} \right\},$$

est identique à

$$\Delta_X^* = \Delta_Z^* = \{\det \mathcal{G}(\mathfrak{Y})\}^{-1/2} \left[ \frac{\partial}{\partial \mathfrak{Z}} \right]' \left\{ \{\det \mathcal{G}(\mathfrak{Y})\}^{1/2} \mathcal{G}(\mathfrak{Y})^{-1} \frac{\partial}{\partial \mathfrak{Z}} \right\}. \quad (3.2.29)$$

**Preuve.** La démonstration est essentiellement identique à celle du lemme 3.2.3, en remplaçant  $T_L$  par  $T$ . Nous en omettons les détails. On observera dans l'énoncé de ce lemme la subtilité que la variable  $\mathfrak{Z}$  (par l'intermédiaire de sa composante  $\beta$ ) n'est pas nécessairement définie, alors que  $d\mathfrak{Z}$  et  $\frac{\partial}{\partial \mathfrak{Z}}$  ont un sens. On peut, en effet, avoir affaire ici à des différentielles  $d\beta$  qui ne sont pas nécessairement des différentielles totales.  $\square$

Nous allons maintenant montrer que l'opérateur de Laplace–Beltrami  $\Delta_X^*$ , défini en (3.2.25), est lié par une relation simple à l'opérateur  $\Delta_X$ , défini en (3.2.10). Ceci est établi dans la proposition suivante.

**Proposition 3.2.5.** *Soit  $X > 0$  une matrice symétrique  $(m \times m)$  définie positive, de valeurs propres  $y_1, \dots, y_m$ . Alors, pour toute fonction  $h(y_1, \dots, y_m)$  de  $y_1, \dots, y_m$ , admettant des dérivées partielles d'ordre 2, on a*

$$\begin{aligned} \Delta_X^*(h) &= \left\{ \Delta_X - \frac{m-3}{2} \sum_{i=1}^m y_i \frac{\partial}{\partial y_i} \right\} (h) \\ &= \left\{ \sum_{i=1}^m y_i^2 \frac{\partial^2}{\partial y_i^2} + \sum_{\substack{j=1 \\ j \neq i}}^m \frac{y_i^2}{y_i - y_j} \frac{\partial}{\partial y_i} - \frac{m-3}{2} \sum_{i=1}^m y_i \frac{\partial}{\partial y_i} \right\} (h). \end{aligned} \quad (3.2.30)$$

**Preuve.** Nous commençons par mettre  $X$  sous la forme  $X = HYH'$ , où  $H \in \mathcal{O}_m$  désigne une matrice orthogonale, et  $Y = \text{diag}(y_1, \dots, y_m)$  désigne la matrice diagonale des valeurs propres de  $X$ . Nous remplaçons ensuite, dans la définition (3.2.18) de  $(ds)^2$ ,  $dX$  par

$$dX = dH \cdot Y \cdot H' + H \cdot dY \cdot H' + H \cdot Y \cdot dH',$$

puis, respectivement,  $H'X^{-1}H$  par  $Y^{-1}$ , et  $H' \cdot dX \cdot H$  par

$$\begin{aligned} H' \cdot dX \cdot H &= H' \{ dH \cdot Y \cdot H' + H \cdot dY \cdot H' + H \cdot Y \cdot dH' \} H \\ &= d\Theta \cdot Y - Y \cdot d\Theta + dY, \end{aligned}$$

où  $d\Theta$ , est définie par l'une ou l'autre des expressions  $d\Theta = H' \cdot dH$  ou  $d\Theta = -dH' \cdot H$ , ce qui est justifié l'identité  $d(H'H) = d\mathbb{1} = \mathbb{0}$ . On obtient ainsi que

$$\begin{aligned}
(ds)^2 &= \operatorname{tr}\left(X^{-1}dX \cdot X^{-1}dX\right) \\
&= \operatorname{tr}\left(\{H'X^{-1}H\}\{H' \cdot dX \cdot H\}\{H'X^{-1}H\}\{H' \cdot dX \cdot H\}\right) \\
&= \operatorname{tr}\left(\{Y^{-1}d\Theta \cdot Y - d\Theta + Y^{-1}dY\}\{Y^{-1}d\Theta \cdot Y - d\Theta + Y^{-1}dY\}\right) \\
&= \operatorname{tr}\left(Y^{-1} \cdot dY \cdot Y^{-1} \cdot dY\right) - 2\operatorname{tr}\left(d\Theta \cdot Y^{-1} \cdot d\Theta \cdot Y\right) + 2\operatorname{tr}\left(d\Theta \cdot d\Theta\right). \tag{3.2.31}
\end{aligned}$$

Remarquons que la matrice  $d\Theta = [d\theta_{i,j}]$  définie par  $d\Theta = H' \cdot dH = -dH' \cdot H$  vérifie  $d\Theta' = -d\Theta$ , et est donc *antisymétrique*, c'est à dire telle que  $d\theta_{i,j} = -d\theta_{j,i}$  pour  $1 \leq i < j \leq m$ , et  $d\theta_{i,i} = 0$  pour  $1 \leq i \leq m$ . Cette remarque, jointe au fait que  $Y = \operatorname{diag}(y_1, \dots, y_m)$  permet d'établir les formules

$$\begin{aligned}
\operatorname{tr}\left(Y^{-1} \cdot dY \cdot Y^{-1} \cdot dY\right) &= \sum_{i=1}^m \frac{(dy_i)^2}{y_i^2}, \\
\operatorname{tr}\left(d\Theta \cdot Y^{-1} \cdot d\Theta \cdot Y\right) &= -\sum_{i=1}^m \sum_{j=1}^m \frac{y_i}{y_j} (d\theta_{i,j})^2 = -\sum_{1 \leq i < j \leq m} \left\{ \frac{y_i}{y_j} + \frac{y_j}{y_i} \right\} (d\theta_{i,j})^2 \\
\operatorname{tr}\left(d\Theta \cdot d\Theta\right) &= -\sum_{i=1}^m \sum_{j=1}^m (d\theta_{i,j})^2 = -2 \sum_{1 \leq i < j \leq m} (d\theta_{i,j})^2.
\end{aligned}$$

Ces formules permet de réécrire sous forme explicite (3.2.31) pour obtenir

$$(ds)^2 = \sum_{i=1}^m \frac{(dy_i)^2}{y_i^2} + 2 \sum_{1 \leq i < j \leq m} \frac{(y_i - y_j)^2}{y_i y_j} (d\theta_{i,j})^2. \tag{3.2.32}$$

La comparaison de (3.2.32) avec l'expression de  $(ds)^2$  qu'on obtient à partir de (3.2.28), soit

$$(ds)^2 = [d\mathbf{y}' \quad d\theta'] \mathcal{G}(\mathbf{y}) \begin{bmatrix} d\mathbf{y} \\ d\theta \end{bmatrix}, \tag{3.2.33}$$

où

$$\mathbf{y} = [y_1 \quad \dots \quad y_m]'$$

les différentielles  $d\mathbf{y}$  et  $d\theta$  étant définies par

$$\begin{aligned}
d\mathbf{y} &= [dy_1 \quad \dots \quad dy_m]', \\
d\theta &= [d\theta_{1,2} \quad \dots \quad d\theta_{1,m} \quad d\theta_{2,3} \quad \dots \quad d\theta_{2,m} \quad \dots \quad d\theta_{m-1,m}]',
\end{aligned}$$

montre que  $\mathcal{G}(\mathbf{y})$  ne dépend que de  $\mathbf{y}$ , et est donnée par

$$\mathcal{G}(\mathbf{y}) = \operatorname{diag}\left(\frac{1}{y_1^2}, \dots, \frac{1}{y_m^2}, \frac{2(y_1 - y_2)^2}{y_1 y_2}, \dots, \frac{2(y_1 - y_m)^2}{y_1 y_m}, \dots, \frac{2(y_{m-1} - y_m)^2}{y_{m-1} y_m}\right). \tag{3.2.34}$$



On applique maintenant le lemme 3.2.4, qui permet d'écrire, compte tenu de (3.2.32), (3.2.33) et (3.2.34),

$$\begin{aligned}
\Delta_X^* &= \{\det G(\mathbf{X})\}^{-1/2} \left[ \frac{\partial}{\partial \mathbf{X}} \right]' \left\{ \{\det G(\mathbf{X})\}^{1/2} G(\mathbf{X})^{-1} \frac{\partial}{\partial \mathbf{X}} \right\} \\
&= \{\det \mathcal{G}(\mathbf{Y})\}^{-1/2} \left( \frac{\partial}{\partial \theta} \right)' \left[ \{\det \mathcal{G}(\mathbf{Y})\}^{1/2} \mathcal{G}(\mathbf{Y})^{-1} \left( \frac{\partial}{\partial \theta} \right) \right] \\
&= \sum_{i=1}^m y_i^2 \frac{\partial^2}{\partial y_i^2} + \sum_{i=1}^m \sum_{\substack{j=1 \\ j \neq i}}^m \frac{y_i^2}{y_i - y_j} \frac{\partial}{\partial y_i} - \frac{m-3}{2} \sum_{i=1}^m y_i \frac{\partial}{\partial y_i} \\
&\quad + \frac{1}{4} \sum_{1 \leq i < j \leq m} \frac{y_i y_j}{(y_i - y_j)^2} \frac{\partial^2}{\partial \theta_{i,j}^2}.
\end{aligned} \tag{3.2.35}$$

Compte tenu de (3.2.10), on a donc

$$\Delta_X^* = \Delta_Y - \frac{m-3}{2} \sum_{i=1}^m y_i \frac{\partial}{\partial y_i} + \frac{1}{4} \sum_{1 \leq i < j \leq m} \frac{y_i y_j}{(y_i - y_j)^2} \frac{\partial^2}{\partial \theta_{i,j}^2}. \tag{3.2.36}$$

Rappelons que  $X = HYH'$  et que  $Y$  ne dépend pas de  $\theta$ . Il s'ensuit que  $h(y_1, \dots, y_m)$  fonction des valeurs propres de  $Y$  ne dépend pas de  $\theta$ . On a donc

$$\Delta_X^*(h) = \left\{ \Delta_X - \frac{m-3}{2} \sum_{i=1}^m y_i \frac{\partial}{\partial y_i} \right\} (h),$$

ce qui permet de conclure.  $\square$

Le résultat suivant est une conséquence facile du théorème 8.1 et de la proposition 3.2.5.

**Théorème 3.2.2.** *Pour tout entier  $k \geq 1$ , et pour toute matrice  $X$ , symétrique et  $(m \times m)$ , le polynôme zonal  $C_\kappa(X)$  associé à la  $m$ -partition  $\kappa = (k_1, \dots, k_m)$  de  $k$  est un vecteur propre de l'opérateur de Laplace-Beltrami  $\Delta_X^*$ , et vérifie*

$$\Delta_X^* C_\kappa(X) = \{\rho_\kappa + \frac{1}{2}k(m+1)\} C_\kappa(X), \tag{3.2.37}$$

où

$$\rho_\kappa = \sum_{i=1}^m k_i(k_i - i). \tag{3.2.38}$$

**Preuve.** Comme, par définition, le polynôme zonal  $C_\kappa(X)$  ne dépend que des valeurs propres  $y_1, \dots, y_m$  de  $X$ , on peut appliquer l'identité (3.2.30), qui s'écrit

$$\Delta_X^* C_\kappa(X) = \Delta_X C_\kappa(X) + \frac{1}{2}(m-3)E_X C_\kappa(X) = \{\rho_\kappa + k(m-1)\} C_\kappa(X) + \frac{1}{2}(m-3)E_X C_\kappa(X),$$

où  $E_X$  désigne l'opérateur d'Euler, défini, à partir des valeurs propres  $y_1, \dots, y_m$  de  $X$  par la relation

$$E_X = \sum_{i=1}^m y_i \frac{\partial}{\partial y_i}. \tag{3.2.39}$$

Comme, par définition,  $C_\kappa(X)$  est un polynôme homogène de degré  $k$  de  $y_1, \dots, y_m$ , on constate aisément que

$$E_X C_\kappa(X) = k C_\kappa(X), \tag{3.2.40}$$

ce qui permet de conclure.  $\square$

Nous pouvons maintenant donner une définition, dont l'équivalence à la définition 8.6 découle d'une application du théorème 8.2.

**Définition 3.2.5.** Soit  $X$  une matrice symétrique ( $m \times m$ ), de valeurs propres  $y_1, \dots, y_m$ . On appelle polynômes zonaux associés à  $X$  les polynômes  $C_\kappa(X)$  des variables  $y_1, \dots, y_m$  vérifiant les propriétés suivantes.

(i) Pour tout entier  $k \geq 0$  et  $\kappa = (k_1, \dots, k_m)$ ,  $m$ -partition de  $k$ ,  $C_\kappa(X)$  est un polynôme homogène symétrique de degré  $k$  en  $y_1, \dots, y_m$  dont le monôme de plus grand poids est  $y_1^{k_1} \dots y_m^{k_m}$ .

(ii) Pour tout entier  $k \geq 0$  et  $\kappa = (k_1, \dots, k_m)$ ,  $m$ -partition de  $k$ ,  $C_\kappa(X)$  est une fonction propre, associée à la valeur propre  $\rho_\kappa + \frac{1}{2}k(m+1)$ , de l'opérateur différentiel  $\Delta_X^*$ , où  $\rho_\kappa$  et  $\Delta_X^*$  sont définis respectivement par (3.2.38) et (3.2.37).

(iii) Pour tout entier  $k \geq 0$ , on a l'identité

$$(\operatorname{tr} X)^k = (y_1 + \dots + y_m)^k = \sum_{\kappa \in \mathcal{P}_m(k)} C_\kappa(X), \quad (3.2.41)$$

où  $\kappa$  varie dans l'ensemble  $\mathcal{P}_m(k)$  de toutes les  $m$ -partitions de  $k$ .

Le théorème 8.3 ci-dessous est d'une importance capitale dans les applications des polynômes zonaux. Il justifie par lui-même l'intérêt de faire intervenir l'opérateur de Laplace-Beltrami  $\Delta_X^*$ , introduit dans la définition 8.7. Le caractère invariant (établi dans les lemmes 8.3 et 8.4) de cet opérateur par la transformation

$$X \rightarrow LXL', \quad \forall L \in \mathcal{GL}(m, \mathbb{R}),$$

est, bien entendu, la propriété essentielle qui va être utilisée. Quelques préliminaires seront utiles pour comprendre son énoncé et sa démonstration. Nous admettrons sans démonstration que la constante  $\rho_\kappa$  définie en (3.2.38) détermine uniquement la partition  $\kappa$  de  $k \geq 1$ . Par ailleurs, nous admettrons la formule suivante, due à Constantine, A.G. (1963). Some noncentral distribution problems in multivariate analysis. *Ann. Math. Statist.* **34** 1270–1285. Celle-ci fournit la valeur de  $C_\kappa(\mathbb{I}_m)$  lorsque la longueur  $L(\kappa)$  de  $\kappa$  est égale à  $p$ . Pour  $p = L(\kappa) \geq 1$ , ceci correspond au cas où  $\kappa = (k_1, \dots, k_p, 0, \dots, 0, \dots)$  comprend exactement  $p$  termes  $k_1 \neq 0, \dots, k_p \neq 0$  non nuls. Lorsque  $\kappa = \mathbb{O} = (0, 0, \dots)$ , la longueur de  $\kappa$  est  $p = L(\kappa) = L(\mathbb{O}) = 0$ . Sous l'hypothèse que  $p = L(\kappa)$ , on a ainsi

$$C_\kappa(\mathbb{I}_m) = 2^{2k} k! (\frac{1}{2}m)_\kappa \frac{\prod_{1 \leq i < j \leq p} (2k_i - 2k_j - i + j)}{p \prod_{i=1}^p (2k_i + p - i)!}. \quad (3.2.42)$$

Dans cette formule, on utilise le *symbole de Pochhammer multivarié* défini, pour toute  $m$ -partition  $\kappa = (k_1, k_2, \dots)$  de  $k \geq 0$ , de longueur  $p = L(\kappa) \leq m$ , par

$$\begin{aligned} (a)_\kappa &= \prod_{i=1}^m \left(a - \frac{1}{2}(i-1)\right)_{k_i} = \prod_{i=1}^m \left\{ \frac{\Gamma(a + k_i - \frac{1}{2}(i-1))}{\Gamma(a - \frac{1}{2}(i-1))} \right\} \\ &= \prod_{i=1}^p \left(a - \frac{1}{2}(i-1)\right)_{k_i} = \prod_{i=1}^p \left\{ \frac{\Gamma(a + k_i - \frac{1}{2}(i-1))}{\Gamma(a - \frac{1}{2}(i-1))} \right\} \\ &= \prod_{i=1}^\infty \left(a - \frac{1}{2}(i-1)\right)_{k_i} = \prod_{i=1}^\infty \left\{ \frac{\Gamma(a + k_i - \frac{1}{2}(i-1))}{\Gamma(a - \frac{1}{2}(i-1))} \right\}, \end{aligned} \quad (3.2.43)$$

où  $(a)_k = a(a+1)\dots(a+k-1) = \Gamma(a+k)/\Gamma(a)$  et  $(a)_0 = 1$ . Rappelons la définition (1.2.1) et l'expression développée (1.2.4) de la fonction gamma multivariée. On a, pour  $\operatorname{re}(a) > \frac{1}{2}(m-1)$ ,

$$\Gamma_m(a) = \int_{A>0} \operatorname{etr}\left(-\frac{1}{2}\Sigma^{-1}A\right) (\det A)^{a-(m+1)/2} (dA) = \pi^{m(m-1)/4} \prod_{i=1}^m \Gamma\left(a - \frac{1}{2}(i-1)\right).$$

On pose, dans le même esprit, pour  $\kappa = (k_1, \dots, k_m)$ ,  $m$ -partition de  $k$ ,

$$\Gamma_m(a; \kappa) = \pi^{m(m-1)/4} \prod_{i=1}^m \Gamma\left(a + k_i - \frac{1}{2}(i-1)\right). \quad (3.2.44)$$

de sorte que (3.2.43) peut être réécrit sous la forme

$$(a)_\kappa = \frac{\Gamma_m(a; \kappa)}{\Gamma_m(a)}. \quad (3.2.45)$$

On remarquera que les définitions (3.2.43) et (3.2.45) de  $(a)_\kappa$  lorsque  $\kappa \in \mathcal{P}$ , ne dépendent pas du choix de  $m \geq p = L(\kappa)$ .

**Remarque 3.2.4.** Compte tenu de la proposition 8.5, il est possible d'étendre la définition des polynômes zonaux  $C_\kappa(X)$ , considérés comme polynômes des valeurs propres des matrices  $(m \times m)$  symétriques  $X$ , au cas de matrices  $(m \times m)$  carrées quelconques. Dans ce qui suit, on adoptera donc cette définition généralisée. Le cas suivant illustre l'intérêt de cette extension.

**Lemme 3.2.5.** *Si  $X$  est une matrice  $(m \times m)$  symétrique et si  $Y$  est une matrice  $(m \times m)$  symétrique positive, alors, les valeurs propres de  $XY$  et de  $X^{1/2}YX^{1/2}$  sont identiques, et, pour toute  $m$ -partition  $\kappa$  de  $k \geq 0$ ,*

$$C_\kappa(XY) = C_\kappa(X^{1/2}YX^{1/2}). \quad (3.2.46)$$

**Preuve.** On sait que les valeurs propres de  $CD$  et  $DC$  sont identiques pour deux matrices carrées  $C$  et  $D$  quelconques (voir le corollaire 1.2). Les valeurs propres de  $X^{1/2}X^{1/2}Y = XY$  et de  $X^{1/2}YX^{1/2}$  sont donc égales. La définition de  $X^{1/2}$  est rendue possible par l'hypothèse que  $X \geq 0$ . La conclusion (3.2.46) est alors une conséquence directe de la proposition 8.5, jointe au fait que  $C_\kappa(Z)$  n'est fonction que des valeurs propres de  $Z$ . On observera ici que  $X^{1/2}YX^{1/2}$  est symétrique, alors que  $XY$  ne l'est pas nécessairement.  $\square$

**Théorème 3.2.3.** *Soit  $X_1 > 0$  une matrice  $(m \times m)$  définie positive, et  $X_2$  une matrice symétrique  $(m \times m)$ . Alors, on a l'égalité*

$$\int_{\mathcal{O}_m} C_\kappa(X_1 H X_2 H') (dH) = \frac{C_\kappa(X_1) C_\kappa(X_2)}{C_\kappa(\mathbb{I}_m)}. \quad (3.2.47)$$

**Preuve.** Tout d'abord, on observera que la matrice  $X_1 H X_2 H'$  n'est, en général, pas symétrique. La définition de  $C_\kappa(X_1 H X_2 H')$  dans (3.2.47) est toutefois justifiée grâce à la remarque 8.4, et au lemme 8.5 qui montrent que

$$C_\kappa(X_1 H X_2 H') = C_\kappa(X_1^{1/2} H X_2 H' X_1^{1/2}).$$

Désignons par  $f_\kappa(X_2)$  la valeur du deuxième membre de (3.2.47), considéré comme fonction de  $X_2$ . Compte tenu du fait que la mesure  $(dH)$  est invariante sur  $\mathcal{O}_m$ , on constate que  $f_\kappa(X_2) = f_\kappa(PX_2P')$  pour toute matrice orthogonale  $P \in \mathcal{O}_m$ . Ceci implique que  $f_\kappa(X_2)$  ne dépend de  $X_2$  que par l'intermédiaire des valeurs propres  $y_1, \dots, y_m$  de cette matrice. Il est alors aisé de constater que  $f_\kappa(X_2)$  ne peut être qu'un polynôme homogène et symétrique de degré  $k$  de  $y_1, \dots, y_m$ .

Supposons maintenant que  $X_2$  soit définie positive. On peut alors appliquer l'opérateur  $\Delta_{X_2}^*$  de Laplace–Beltrami à la fonction  $f_\kappa(X_2)$ , pour obtenir

$$\begin{aligned} \Delta_{X_2}^* f_\kappa(X_2) &= \int_{\mathcal{O}_m} \Delta_{X_2}^* C_\kappa(X_1 H X_2 H') (dH) \\ &= \int_{\mathcal{O}_m} \Delta_{X_2}^* C_\kappa(X_1^{1/2} H X_2 H' X_1^{1/2}) (dH) \\ &= \int_{\mathcal{O}_m} \Delta_{X_2}^* C_\kappa(LX_2L') (dH), \end{aligned}$$

où nous avons posé  $L = X^{1/2}H$  et fait usage de (3.2.46). Par application des lemmes 8.3 et 8.4, on voit que  $\Delta_{X_2}^* = \Delta_{LX_2L'}^*$ . Par conséquent, on déduit de (1.4.10), (3.2.37), et (3.2.38) que

$$\begin{aligned} \Delta_{X_2}^* f_\kappa(X_2) &= \int_{\mathcal{O}_m} \Delta_{LX_2L'}^* C_\kappa(LX_2L')(dH), \\ &= \left\{ \rho_\kappa + \frac{1}{2}k(m+1) \right\} \int_{\mathcal{O}_m} C_\kappa(LX_2L')(dH), \\ &= \left\{ \rho_\kappa + \frac{1}{2}k(m+1) \right\} f_\kappa(X_2). \end{aligned}$$

Ceci implique que  $f_\kappa(X_2) = \lambda C_\kappa(X_2)$  pour un coefficient de proportionnalité  $\lambda$  convenable. Il suffit donc d'évaluer  $\lambda$  en prenant la valeur particulière  $X_2 = \mathbb{I}_m$ , qui fournit  $f_\kappa(\mathbb{I}_m) = C_\kappa(X_1)$ , et donc  $\lambda = C_\kappa(X_1)/C_\kappa(\mathbb{I}_m)$ , d'où la conclusion. Le fait que ce résultat subsiste pour  $X_2$  symétrique arbitraire peut être déduit de la proposition 8.5.□

Le lemme technique suivant sera particulièrement utile.

**Lemme 3.2.6.** *Soit  $Y = \text{diag}(y_1, \dots, y_m)$  et soit  $X = [x_{ij}] > 0$  une matrice définie positive ( $m \times m$ ). On pose*

$$X_1 = [x_{11}], X_2 = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}, \dots, X_r = \begin{bmatrix} x_{11} & \dots & x_{1r} \\ \vdots & \ddots & \vdots \\ x_{r1} & \dots & x_{rr} \end{bmatrix}, \dots, X_m = X. \quad (3.2.48)$$

Alors, pour toute  $m$ -partition  $\kappa = (k_1, \dots, k_m)$  de  $k \geq 0$ , on a

$$C_\kappa(XY) = d_\kappa y_1^{k_1} \dots y_m^{k_m} \prod_{r=1}^m (\det X_r)^{k_r - k_{r+1}} + (\text{termes d'ordre ou de poids inférieur}), \quad (3.2.49)$$

où l'on pose  $k_{m+1} = 0$  et où  $d_\kappa$  désigne le coefficient du terme de plus haut poids de  $C_\kappa(\cdot)$ .

**Preuve.** Omise.□

Le théorème suivant, qui sera fort utile par la suite, est de démonstration plus délicate. Rappelons la notation (2.1.2), définissant  $Re(A) = [Re(a_{ij})]$  et  $Im(A) = [Im(a_{ij})]$  comme matrices composées des parties réelle et imaginaire des éléments de la matrice  $A = [a_{ij}]$ .

**Théorème 3.2.4.** *Soit  $Y$  une matrice ( $m \times m$ ) symétrique et positive, et  $Z$  une matrice symétrique ( $m \times m$ ) à coefficients complexes telle que  $Re(Z) > 0$ . Alors, pour tout  $a \in \mathbb{C}$  vérifiant  $Re(a) > \frac{1}{2}(m-1)$ , on a*

$$\int_{X>0} \text{etr}(-XZ) (\det X)^{a-(m+1)/2} C_\kappa(XY) (dX) = (a)_\kappa \Gamma_m(a) C_\kappa(Y). \quad (3.2.50)$$

**Preuve.** Omise.□

### 3.3 Fonctions hypergéométriques matricielles

**Définition 3.3.1.** *Soit  $X$  une matrice symétrique ( $m \times m$ ). On définit la fonction hypergéométrique matricielle de première espèce d'ordre  $p, q \geq 0$  associée à  $X$  et aux réels  $a_1, \dots, a_p, b_1, \dots, b_q$  par la relation*

$${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; X) = \sum_{k=0}^{\infty} \sum_{\kappa \in \mathcal{P}_m(k)} \frac{(a_1)_\kappa \dots (a_p)_\kappa}{(b_1)_\kappa \dots (b_q)_\kappa} \frac{C_\kappa(X)}{k!}, \quad (3.3.1)$$

où la sommation est effectuée lorsque  $\kappa = (k_1, \dots, k_m)$  avec  $k_1 \geq \dots \geq k_m \geq 0$  parcourt l'ensemble  $\mathcal{P}_m(k)$  des  $m$ -partitions de  $k = k_1 + \dots + k_m$ . Ici,  $C_\kappa(X)$  désigne le polynôme zonal associé à  $\kappa$  et  $X$ , et  $(a)_\kappa$  est le symbole de Pochhammer multivarié, défini (cf. (3.2.43)) par

$$(a)_\kappa = \prod_{i=1}^m \left( a - \frac{1}{2}(i-1) \right)_{k_i} \quad \text{pour } \kappa = (k_1, \dots, k_m). \quad (3.3.2)$$

L'exemple le plus simple de fonction hyperg m trique matricielle est obtenu pour  $p = q = 0$ . On obtient alors, par (3.2.11),

$${}_0F_0(X) = \sum_{k=0}^{\infty} \left\{ \sum_{\kappa \in \mathcal{P}_m(k)} \frac{C_{\kappa}(X)}{k!} \right\} = \sum_{k=0}^{\infty} \frac{\text{tr}(X)^k}{k!} = \text{etr}(X). \quad (3.3.3)$$

En utilisant les m mes conventions que dans la d finition ci-dessus, on peut introduire, de m me, les fonctions hyperg m triques matricielles de seconde esp ce.

**D finition 3.3.2.** Soient  $X$  et  $Y$  deux matrices sym triques ( $m \times m$ ). On d finit la fonction hyperg m trique matricielle de deuxi me esp ce d'ordre  $p, q \geq 0$  associ e    $X, Y$  et aux r els  $a_1, \dots, a_p, b_1, \dots, b_q$  par la relation

$${}_pF_q^{(m)}(a_1, \dots, a_p; b_1, \dots, b_q; X) = \sum_{k=0}^{\infty} \sum_{\kappa \in \mathcal{P}_m(k)} \frac{(a_1)_{\kappa} \dots (a_p)_{\kappa}}{(b_1)_{\kappa} \dots (b_q)_{\kappa}} \frac{C_{\kappa}(X)C_{\kappa}(Y)}{k!C_{\kappa}(\mathbb{I}_m)}, \quad (3.3.4)$$

o  la sommation est effectu e sur toutes les  $m$ -partitions  $\kappa = (k_1, \dots, k_m)$  avec  $k_1 \geq \dots \geq k_m \geq 0$  de  $k = k_1 + \dots + k_m$ .

Les r sultats suivants seront tr s utiles par la suite.

**Th or me 3.3.1.** Soit  $Z$  une matrice complexe ( $m \times m$ ) v rifiant  $\text{Re}(Z) > 0$  et  $Y$  une matrice sym trique r elle ( $m \times m$ ). Alors, pour  $\text{Re}(a) > (m-1)/2$ , et, ou bien  $p < q$ , ou bien  $p = q$  et  $\|Z^{-1}\| < 1$ ,

$$\begin{aligned} & \int_{X>0} \text{etr}(-XZ)(\det X)^{a-(m+1)/2} {}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; X)(dX) \\ &= \Gamma_m(a)(\det Z)^{-a} {}_{p+1}F_q(a_1, \dots, a_p; b_1, \dots, b_q; Z^{-1}). \end{aligned} \quad (3.3.5)$$

De m me, pour  $\text{Re}(a) > (m-1)/2$ , et, ou bien  $p < q$ , ou bien  $p = q$ ,  $\|Z^{-1}\| < 1$  et  $\|Y\| < 1$ ,

$$\begin{aligned} & \int_{X>0} \text{etr}(-XZ)(\det X)^{a-(m+1)/2} {}_pF_q^{(m)}(a_1, \dots, a_p; b_1, \dots, b_q; X, Y)(dX) \\ &= \Gamma_m(a)(\det Z)^{-a} {}_{p+1}F_q^{(m)}(a_1, \dots, a_p; b_1, \dots, b_q; Z^{-1}, Y), \end{aligned} \quad (3.3.6)$$

**Preuve.** Omise.  $\square$

**Th or me 3.3.2.** Soit  $X$  une matrice ( $m \times n$ ) avec  $m \geq n$ . Si  $H$  est une matrice orthogonale ( $n \times n$ ), on pose  $H = [H_1 \ H_2]$ , o   $H_1 \in \mathcal{V}_{m,n}$  est ( $n \times m$ ) et  $H_2 \in \mathcal{V}_{n-m,n}$  est ( $n \times (n-m)$ ). Alors

$$\int_{\mathcal{O}_n} \text{etr}(XH_1)(dH) = {}_0F_1\left(\frac{1}{2}n; \frac{1}{4}XX'\right). \quad (3.3.7)$$

**Preuve.** Remarquons tout d'abord que l'int grale

$$\Psi(X) = \int_{\mathcal{O}_n} \text{etr}(XH_1)(dH)$$

est telle que, pour toute matrice ( $n \times n$ ) orthogonale  $K$ ,

$$\Psi(X) = \Psi(XK).$$

En effet, le changement de variable  $H \in \mathcal{O}_m \rightarrow L \in \mathcal{O}_m$  d fini par  $H = KL$  pour  $K \in \mathcal{O}_m$  fix e, transforme ( $dH$ ) en ( $dL$ ). En posant  $L = [L_1 \ L_2]$ , on voit que  $H_1 = KL_1$ , de sorte que

$$\Psi(X) = \int_{\mathcal{O}_m} \text{etr}(XKL_1)(dL) = \Psi(XK).$$

Par la proposition 1.5,  $X'$  se factorise sous la forme  $X' = \mathcal{H}_1 T$  où  $\mathcal{H} = [\mathcal{H}_1 \quad \mathcal{H}_2]$  est une matrice  $(n \times n)$  orthogonale, et  $T$ , une matrice  $(m \times m)$ , triangulaire supérieure à diagonale positive, et solution de l'équation  $T'T = XX'$ . Comme  $\Psi(X) = \Psi(XK)$  pour toute matrice  $K \in \mathcal{O}_m$ , nécessairement,  $\Psi(X)$  est fonction de  $T$  et donc de  $XX'$ . Pour établir (3.3.7), on se ramène au cas où  $\text{rg}(X) = m$ , de sorte que la matrice  $(m \times m)$   $X'X$  est définie positive. Dans ce cas, (3.3.7) est équivalent à l'identité des fonctions  $g_1$  et  $g_2$  définies comme suit.

$$\begin{aligned} g_1(X'X) &= \det(XX')^{(n-m-1)/2} \int_{\mathcal{O}_n} \text{etr}(XH_1)(dH) \\ &= g_2(X'X) = \det(XX')^{(n-m-1)/2} {}_0F_1\left(\frac{1}{2}n; \frac{1}{4}XX'\right). \end{aligned} \quad (3.3.8)$$

Nous allons établir (3.3.8) en montrant que les transformées de Laplace des fonctions  $g_1$  et  $g_2$  sont identiques. Ceci revient à établir que

$$\begin{aligned} \tilde{g}_1(Z) &= \int_{XX' > 0} g_1(XX') \text{etr}(-XX'Z)(d\{XX'\}) \\ &= \tilde{g}_2(Z) = \int_{XX' > 0} g_2(XX') \text{etr}(-XX'Z)(d\{XX'\}). \end{aligned} \quad (3.3.9)$$

Considérons tout d'abord  $\tilde{g}_1(Z)$ . Posons  $X' = \mathcal{H}_1 T$  et  $A = T'T$ , où  $\mathcal{H} = [\mathcal{H}_1 \quad \mathcal{H}_2]$  est une matrice orthogonale  $(n \times n)$ ,  $\mathcal{H}_1 \in \mathcal{V}_{m,n}$ , et  $T$  une matrice  $(m \times m)$ , triangulaire supérieure à diagonale positive. Comme, par (1.1.7),

$$(d\{XX'\}) = (dA) = 2^m \prod_{i=1}^m t_{ii}^{m-i+1} (dT), \quad (3.3.10)$$

par (1.3.4),

$$(dX) = \prod_{i=1}^m t_{ii}^{n-i} (dT)(\mathcal{H}'_1 d\mathcal{H}_1), \quad (3.3.11)$$

et par (1.4.1),

$$\int_{\mathcal{V}_{m,n}} (\mathcal{H}'_1 d\mathcal{H}_1) = \frac{2^m \pi^{mn/2}}{\Gamma_m(\frac{1}{2}n)}, \quad (3.3.12)$$

on a

$$2^{-m} \prod_{i=1}^m t_{ii}^{-m+i-1+n-i} (dA)(\mathcal{H}'_1 d\mathcal{H}_1) = 2^{-m} \det(A)^{(n-m-1)/2} (dA)(\mathcal{H}'_1 d\mathcal{H}_1) = (dX),$$

et donc

$$(dA)(\mathcal{H}'_1 d\mathcal{H}_1) = 2^m \det(A)^{-(n-m-1)/2} (dX).$$

Comme  $A = XX'$ , on constate donc que

$$\begin{aligned} \tilde{g}_1(Z) &= \frac{\Gamma_m(\frac{1}{2}n)}{2^m \pi^{mn/2}} \int_{K_1 \in \mathcal{V}_{m,n}} \tilde{g}(Z)(\mathcal{H}'_1 d\mathcal{H}_1) \\ &= \frac{\Gamma_m(\frac{1}{2}n)}{2^m \pi^{mn/2}} \int_{\mathcal{H}_1 \in \mathcal{V}_{m,n}} \int_{A > 0} g_1(A) \text{etr}(-AZ)(dA)(\mathcal{H}'_1 d\mathcal{H}_1) \\ &= \frac{\Gamma_m(\frac{1}{2}n)}{2^m \pi^{mn/2}} \int_{X \in \mathcal{M}_{n,m}} g_1(A) \text{etr}(-AZ) 2^m \det(A)^{-(n-m-1)/2} (dX) \\ &= \frac{\Gamma_m(\frac{1}{2}n)}{\pi^{mn/2}} \int_{X \in \mathcal{M}_{n,m}} \int_{H \in \mathcal{O}_n} \text{etr}(XH_1 - AZ)(dX)(dH) \\ &= \frac{\Gamma_m(\frac{1}{2}n)}{\pi^{mn/2}} \int_{H \in \mathcal{O}_n} \left\{ \int_{X \in \mathcal{M}_{n,m}} \text{etr}(XH_1 - XX'Z)(dX) \right\} (dH). \end{aligned}$$

Fixons maintenant la matrice  $Z$ , et supposons que celle-ci soit réelle, symétrique, et définie positive. Fixons également  $H_1$ , et effectuons le changement de variable

$$Y = Z^{1/2}X - \frac{1}{2}Z^{-1/2}H_1' = Z^{-1/2}\left(ZX - \frac{1}{2}H_1'\right).$$

De toute évidence, comme  $Z = Z'$  et  $H_1'H_1 = \mathbb{I}_m$ ,

$$\begin{aligned} YY' &= Z^{-1/2}\left(ZX - \frac{1}{2}H_1'\right)\left(X'Z' - \frac{1}{2}H_1\right)Z^{-1/2} \\ &= Z^{-1/2}\left(ZXX'Z - \frac{1}{2}H_1'X'Z - \frac{1}{2}H_1ZX + \frac{1}{4}H_1'H_1\right)Z^{-1/2}. \end{aligned}$$

En prenant la trace des deux membres de cette relation, on obtient que

$$\text{tr}(YY') = \text{tr}\left(XX'Z - XH_1\right) + \text{tr}\left(\frac{1}{4}Z^{-1}\right),$$

ce qui nous permet d'écrire, le jacobien du changement de variable donnant

$$(dY) = \det(Z^{1/2})^n(dX) \Leftrightarrow (dX) = \det(Z)^{-n/2}(dY),$$

$$\int_{X \in \mathcal{M}_{n,m}} \text{etr}(XH_1 - XX'Z)(dX) = (\det Z)^{-n/2} \text{etr}\left(\frac{1}{4}Z^{-1}\right) \int_{Y \in \mathcal{M}_{n,m}} \text{etr}(-YY')(dY).$$

On vérifie aisément que si  $Y = [y_{ij}]$  est une matrice  $(m \times n)$ , alors

$$\int_{Y \in \mathcal{M}_{n,m}} \text{etr}(-YY')(dY) = \prod_{i=1}^m \prod_{j=1}^n \left\{ \int_{-\infty}^{\infty} \exp(-y_{ij}^2) dy_{ij} \right\} = \pi^{mn/2}.$$

On en déduit donc que

$$\begin{aligned} \tilde{g}_1(Z) &= \frac{\Gamma_m(\frac{1}{2}n)}{\pi^{mn/2}} \times (\det Z)^{-n/2} \text{etr}\left(\frac{1}{4}Z^{-1}\right) \times \pi^{mn/2} \\ &= \Gamma_m(\frac{1}{2}n) (\det Z)^{-n/2} \text{etr}\left(\frac{1}{4}Z^{-1}\right). \end{aligned} \tag{3.3.13}$$

Il nous reste à faire usage de(3.3.5) pour constater que

$$\begin{aligned} \tilde{g}_2(Z) &= \int_{A>0} g_2(A) \text{etr}(-AZ)(dA) = \int_{A>0} (\det A)^{(n-m-1)/2} \text{etr}(-AZ) {}_0F_1\left(\frac{1}{2}n; \frac{1}{4}A\right)(dA) \\ &= \Gamma_m\left(\frac{1}{2}n\right) (\det Z)^{-n/2} {}_1F_1\left(\frac{1}{2}n; \frac{1}{2}n; Z\right) \\ &= \Gamma_m\left(\frac{1}{2}n\right) (\det Z)^{-n/2} \text{etr}\left(\frac{1}{4}Z^{-1}\right). \end{aligned}$$

Ceci établit que  $\tilde{g}_1(Z) = \tilde{g}_2(Z)$  pour toute matrice  $(m \times m)$  symétrique  $Z$  réelle et définie positive. L'identité pour  $Z$  quelconque vérifiant  $\text{Re}(Z) > 0$  s'obtient par prolongement analytique. Ceci impliquant que  $g_1 = g_2$ , on a ainsi montré (3.3.7).□

### 3.4 Loi de Wishart non centrée

La loi de Wishart non centrée a été étudiée pour la première fois de manière systématique par Anderson, G.A. (1946). The noncentral Wishart distribution and certain problems of multivariate analysis. *Ann. Math. Statist.* **17** 409–431. L'expression développée de sa densité à l'aide de polynômes zonaux est due à James, A.T. (1961). The distribution of noncentral means with known covariance. *Ann. Math. Statist.* **32** 874–882, et Constantine, A.G. (1963). Some noncentral distribution problems in multivariate analysis. *Ann. Math. Statist.* **34** 1270–1285.

**Définition 3.4.1.** Soit  $\mathbb{Z} \stackrel{d}{=} N_{n,m}(M, \mathbb{I}_n \otimes \Sigma)$  une matrice aléatoire ( $n \times m$ ), suivant une loi normale matricielle, avec  $n \geq m$  et  $\Sigma > 0$ . Alors la loi de  $A = \mathbb{Z}'\mathbb{Z}$  est appelée loi de Wishart matricielle à  $n$  degrés de liberté associée à la matrice de variances-covariances  $\Sigma$  et au paramètre de décentrement  $\Omega = \Sigma^{-1}M'M$ . On désigne cette loi par  $A \stackrel{d}{=} W_m(n, \Sigma, \Omega)$ .

**Remarque 3.4.1.** Posons

$$\mathbb{Z} = \begin{bmatrix} Z'_1 \\ \vdots \\ Z'_n \end{bmatrix} \quad \text{et} \quad M = \begin{bmatrix} \mu'_1 \\ \vdots \\ \mu'_n \end{bmatrix}.$$

Sous les hypothèses de la définition 3.4.1, les vecteurs aléatoires  $Z_1, \dots, Z_n$  sont indépendants et tels que

$$Z_i \stackrel{d}{=} N_m(\mu_i, \Sigma) \quad \text{pour} \quad i = 1, \dots, n.$$

On a alors, dans la définition 3.4.1,

$$A = \mathbb{Z}'\mathbb{Z} = \sum_{i=1}^n Z_i Z_i' \stackrel{d}{=} W_m\left(n, \Sigma, \Sigma^{-1} \sum_{i=1}^n \mu_i \mu_i'\right) \quad \text{et} \quad \Omega = \Sigma^{-1}M'M = \Sigma^{-1} \sum_{i=1}^n \mu_i \mu_i'.$$

Nous commençons par le plus simple, le calcul de la fonction caractéristique de  $A \stackrel{d}{=} W_m(n, \Sigma, \Omega)$ . Celle-ci nous montrera au passage la propriété non évidente et implicitement admise dans la définition ci-dessus que  $\mathcal{L}(A)$  ne dépend, pour  $m \geq 1$ ,  $n \geq 1$  et  $\Sigma > 0$  donnés, que de la matrice ( $m \times m$ )  $\Omega = \Sigma^{-1}M'M$ .

**Théorème 3.4.1.** Si  $A \stackrel{d}{=} W_m(n, \Sigma, \Omega)$ , alors, la fonction caractéristique de  $A$  est donnée, en fonction de  $\Theta$ , matrice symétrique ( $m \times m$ ), par

$$\phi(\Theta) = \mathbb{E}\left(\text{etr}(i\Theta A)\right) = \det\left(\mathbb{I}_m - 2i\Theta\Sigma\right)^{-n/2} \text{etr}\left(-\frac{1}{2}\Omega\right) \text{etr}\left(\frac{1}{2}\Omega\left\{\mathbb{I}_m - 2i\Theta\Sigma\right\}^{-1}\right). \quad (3.4.1)$$

**Preuve.** Soit  $\mathbb{Z} \stackrel{d}{=} N_{n,m}(M, \mathbb{I}_n \otimes \Sigma)$ . Par définition,  $A = \mathbb{Z}'\mathbb{Z} \stackrel{d}{=} W_m(n, \Sigma, \Omega)$ . Posons  $\mathbb{Z}' = [Z_1 \ \dots \ Z_n]$ , et  $M' = \mathbb{E}(\mathbb{Z}') = [\mu_1 \ \dots \ \mu_n]$ . Les vecteurs aléatoires  $Z_1, \dots, Z_n$  sont indépendants, de lois respectives  $Z_i \equiv N_m(\mu_i, \Sigma)$ , pour  $i = 1, \dots, n$ . Par conséquent,

$$\phi(\Theta) = \mathbb{E}(\text{etr}(i\Theta A)) = \mathbb{E}\left(\text{etr}\left(i\Theta \sum_{i=1}^n Z_i Z_i'\right)\right) = \prod_{i=1}^n \mathbb{E}\left(\exp(iZ_i' \Theta Z_i)\right). \quad (3.4.2)$$

On effectue maintenant le changement de variable  $Z_i = \Sigma^{1/2} H Y_i$ , où  $Y_i \stackrel{d}{=} N_m(\nu_i, \mathbb{I}_m)$ ,  $\nu_i = H' \Sigma^{-1/2} \mu_i$ , pour  $i = 1, \dots, n$ , et où  $H$  désigne une matrice ( $m \times m$ ) orthogonale que l'on précisera par la suite. On obtient alors que, pour  $1 \leq i \leq m$  fixé,

$$\mathbb{E}\left(\exp(iZ_i' \Theta Z_i)\right) = \mathbb{E}\left(\exp\left(iY_i' H' \{\Sigma^{1/2} \Theta \Sigma^{1/2}\} H Y_i\right)\right). \quad (3.4.3)$$

Pour toute matrice  $\Theta$  symétrique *fixée*,  $\Sigma^{1/2} \Theta \Sigma^{1/2}$  est une matrice symétrique, et on peut donc choisir une matrice orthogonale  $H$ , telle que la matrice

$$H' \{\Sigma^{1/2} \Theta \Sigma^{1/2}\} H = \text{diag}(\lambda_1, \dots, \lambda_m), \quad (3.4.4)$$

soit diagonale. Posons

$$Y_i = \begin{bmatrix} Y_{1,i} \\ \vdots \\ Y_{m,i} \end{bmatrix} \quad \text{et} \quad \nu_i = \begin{bmatrix} \nu_{1,i} \\ \vdots \\ \nu_{m,i} \end{bmatrix}.$$



On déduit de ce qui précède que

$$\mathbb{E}\left(\exp\left(iY_i' H' \{\Sigma^{1/2} \Theta \Sigma^{1/2}\} H Y_i\right)\right) = \prod_{k=1}^m \mathbb{E}\left(\exp\left(i\lambda_k Y_{k,i}^2\right)\right). \quad (3.4.5)$$

Fixons  $i = 1, \dots, n$ . Comme  $Y_i \stackrel{d}{=} N_m(\nu_i, \mathbb{I}_m)$ , les v.a.  $Y_{1,i}, \dots, Y_{m,i}$  sont indépendantes, de lois respectives données par  $Y_{k,i} \stackrel{d}{=} N(\nu_{k,i}, 1)$  pour  $1 \leq k \leq m$ . On déduit de (4.2.9) que, pour  $k = 1, \dots, m$ ,  $Y_{k,i}^2 \stackrel{d}{=} \chi_1^2(\nu_{k,i}^2)$ . Par conséquent, pour  $k = 1, \dots, m$ ,

$$\begin{aligned} \mathbb{E}\left(\exp\left(i\lambda_k Y_{k,i}^2\right)\right) &= (1 - 2i\lambda_k)^{-1/2} \exp\left(\frac{i\lambda_k \nu_{k,i}^2}{1 - 2i\lambda_k}\right) \\ &= (1 - 2i\lambda_k)^{-1/2} \exp\left(\frac{\nu_{k,i}^2 (2i\lambda_k - 1 + 1)}{2(1 - 2i\lambda_k)}\right) \\ &= (1 - 2i\lambda_k)^{-1/2} \exp\left(-\frac{1}{2}\nu_{k,i}^2\right) \exp\left(\frac{\nu_{k,i}^2}{2(1 - 2i\lambda_k)}\right). \end{aligned}$$

En combinant (3.4.2), (3.4.3) et (3.4.5) à ce dernier résultat, on obtient que

$$\phi(\Theta) = \left\{ \prod_{i=1}^n \prod_{k=1}^m (1 - 2i\lambda_k)^{1/2} \right\} \times \left\{ \prod_{i=1}^n \prod_{k=1}^m \exp\left(-\frac{1}{2}\nu_{k,i}^2\right) \right\} \times \left\{ \prod_{i=1}^n \prod_{k=1}^m \exp\left(\frac{\nu_{k,i}^2}{2(1 - 2i\lambda_k)}\right) \right\}. \quad (3.4.6)$$

Nous évaluons successivement chacun des termes de ce produit. Tout d'abord, faisant usage de l'identité (2.6.11) qui stipule que  $\det(\mathbb{I} + CD) = \det(\mathbb{I} + DC)$ , on constate aisément par (3.4.4) que

$$\prod_{i=1}^n \prod_{k=1}^m (1 - 2i\lambda_k)^{1/2} = \det(\mathbb{I}_m - 2iH' \{\Sigma^{1/2} \Theta \Sigma^{1/2}\} H)^{-n/2} = \det(\mathbb{I}_m - 2i\Theta \Sigma)^{-n/2}. \quad (3.4.7)$$

Ensuite, comme

$$\begin{aligned} \prod_{k=1}^m \exp\left(-\frac{1}{2}\nu_{k,i}^2\right) &= \exp\left(-\frac{1}{2} \sum_{k=1}^m \nu_{k,i}^2\right) = \exp\left(-\frac{1}{2}\nu_i' \nu_i\right) = \exp\left(-\frac{1}{2}\mu_i' \Sigma^{-1/2} H H' \Sigma^{-1/2} \mu_i\right) \\ &= \exp\left(-\frac{1}{2}\mu_i' \Sigma^{-1} \mu_i\right) = \text{etr}\left(-\frac{1}{2}\Sigma^{-1} \mu_i \mu_i'\right), \end{aligned}$$

on constate que

$$\prod_{i=1}^n \prod_{k=1}^m \exp\left(-\frac{1}{2}\nu_{k,i}^2\right) = \text{etr}\left(-\frac{1}{2}\Sigma^{-1} \sum_{i=1}^n \mu_i \mu_i'\right) = \text{etr}\left(-\frac{1}{2}\Sigma^{-1} M' M\right) = \text{etr}\left(-\frac{1}{2}\Omega\right). \quad (3.4.8)$$

Pour le dernier terme de (3.4.6), on fait usage de l'égalité  $\nu_j = H'\Sigma^{-1/2}\mu_j$  pour écrire que

$$\begin{aligned}
\prod_{k=1}^m \exp\left(\frac{\nu_{kj}^2}{2(1-2i\lambda_k)}\right) &= \exp\left(\frac{1}{2}\nu_j' \left\{ \mathbb{I}_m - 2i \operatorname{diag}(\lambda_1, \dots, \lambda_m) \right\}^{-1} \nu_j\right) \\
&= \exp\left(\frac{1}{2}\mu_j' \Sigma^{-1/2} H \left\{ \mathbb{I}_m - 2i \operatorname{diag}(\lambda_1, \dots, \lambda_m) \right\}^{-1} H' \Sigma^{-1/2} \mu_j\right) \\
&= \exp\left(\frac{1}{2}\mu_j' \left\{ \Sigma^{1/2} H \left( \mathbb{I}_m - 2i \operatorname{diag}(\lambda_1, \dots, \lambda_m) \right) H' \Sigma^{1/2} \right\}^{-1} \mu_j\right) \\
&= \exp\left(\frac{1}{2}\mu_j' \left\{ \Sigma^{1/2} H \left( \mathbb{I}_m - 2i H' \left\{ \Sigma^{1/2} \Theta \Sigma^{1/2} \right\} H \right) H' \Sigma^{1/2} \right\}^{-1} \mu_j\right) \\
&= \exp\left(\frac{1}{2}\mu_j' \left\{ \Sigma^{1/2} \left( \mathbb{I}_m - 2i \left\{ \Sigma^{1/2} \Theta \Sigma^{1/2} \right\} \right) \Sigma^{1/2} \right\}^{-1} \mu_j\right) \\
&= \exp\left(\frac{1}{2}\mu_j' \left\{ \Sigma \left( \Sigma^{-1} - 2i\Theta \right) \Sigma \right\}^{-1} \mu_j\right) \\
&= \exp\left(\frac{1}{2}\mu_j' \Sigma^{-1} \left( \Sigma^{-1} - 2i\Theta \right)^{-1} \Sigma^{-1} \mu_j\right) \\
&= \operatorname{etr}\left(\frac{1}{2}\Sigma^{-1} \mu_j \mu_j' \left\{ \left( \Sigma^{-1} - 2i\Theta \right) \Sigma \right\}^{-1}\right) \\
&= \operatorname{etr}\left(\frac{1}{2}\Sigma^{-1} \mu_j \mu_j' \left\{ \mathbb{I}_m - 2i\Theta \Sigma \right\}^{-1}\right).
\end{aligned}$$

Comme

$$M' = \mathbb{E}(Z') = [\mu_1 \quad \dots \quad \mu_n],$$

on en déduit que

$$\begin{aligned}
\prod_{j=1}^n \prod_{k=1}^m \exp\left(\frac{\nu_{kj}^2}{2(1-2i\lambda_k)}\right) &= \operatorname{etr}\left(\frac{1}{2}\Sigma^{-1} \sum_{j=1}^n \mu_j \mu_j' \left\{ \mathbb{I}_m - 2i\Theta \Sigma \right\}^{-1}\right) \\
&= \operatorname{etr}\left(\frac{1}{2}\Sigma^{-1} M' M \left\{ \mathbb{I}_m - 2i\Theta \Sigma \right\}^{-1}\right) \\
&= \operatorname{etr}\left(\frac{1}{2}\Omega \left\{ \mathbb{I}_m - 2i\Theta \Sigma \right\}^{-1}\right). \tag{3.4.9}
\end{aligned}$$

En combinant (3.4.6) (3.4.7) et (3.4.8) avec (3.4.9), on conclut que

$$\phi(\Theta) = \det\left(\mathbb{I}_m - 2i\Theta \Sigma\right)^{-n/2} \operatorname{etr}\left(-\frac{1}{2}\Omega\right) \operatorname{etr}\left(\frac{1}{2}\Omega \left\{ \mathbb{I}_m - 2i\Theta \Sigma \right\}^{-1}\right),$$

ce qu'il fallait démontrer.  $\square$

**Théorème 3.4.2.** Lorsque  $A \stackrel{d}{=} W_m(n, \Sigma, \Omega)$  avec  $n \geq m$  et  $\Sigma > 0$ , la densité de  $A$  existe et est donnée par la formule

$$\begin{aligned}
f(A) &= \frac{1}{\Gamma_m\left(\frac{1}{2}n\right)} 2^{-mn/2} (\det \Sigma)^{-n/2} \operatorname{etr}\left(-\frac{1}{2}\Sigma^{-1}A\right) (\det A)^{(n-m-1)/2} \\
&\quad \times \operatorname{etr}\left(-\frac{1}{2}\Omega\right) {}_0F_1\left(\frac{1}{2}n; \frac{1}{4}\Omega \Sigma^{-1}A\right). \tag{3.4.10}
\end{aligned}$$

**Preuve.** Par le (3.4.14), la densité de  $Z$  est donnée par

$$\begin{aligned}
f(Z) &= (2\pi)^{-mn/2} (\det \Sigma)^{-n/2} \operatorname{etr}\left(-\frac{1}{2}\Sigma^{-1}(Z-M)'(Z-M)\right) \\
&= (2\pi)^{-mn/2} (\det \Sigma)^{-n/2} \operatorname{etr}\left(-\frac{1}{2}\Sigma^{-1}Z'Z\right) \operatorname{etr}\left(-\frac{1}{2}\Sigma^{-1}M'M\right) \operatorname{etr}\left(\Sigma^{-1}M'Z\right) \\
&= (2\pi)^{-mn/2} (\det \Sigma)^{-n/2} \operatorname{etr}\left(-\frac{1}{2}\Sigma^{-1}Z'Z\right) \operatorname{etr}\left(-\frac{1}{2}\Omega\right) \operatorname{etr}\left(\Sigma^{-1}M'Z\right).
\end{aligned}$$

Posons maintenant  $\mathbb{Z} = H_1 T$ , où  $H = \begin{bmatrix} H_1 & H_2 \end{bmatrix}$  est une matrice  $(n \times n)$  orthogonale,  $H_1 \in \mathcal{V}_{m,n}$ ,  $H_2 \in \mathcal{V}_{n-m,n}$ , et  $T$  une matrice  $(m \times m)$ , triangulaire supérieure à diagonale positive. Ce changement de variable fournit un élément différentiel de la forme (voir (1.3.4))

$$(d\mathbb{Z}) = \prod_{i=1}^m t_{ii}^{n-i} (dT) (H'_1 dH_1). \quad (3.4.11)$$

De plus, compte tenu de (1.1.7), on peut écrire

$$(d\{T'T\}) = 2^m \prod_{i=1}^m t_{ii}^{m-i+1} (dT) \Leftrightarrow (dT) = 2^{-m} \prod_{i=1}^m t_{ii}^{-m+i-1} (d\{T'T\}); \quad (3.4.12)$$

de sorte que, en combinant (3.4.11), (3.4.12), et la remarque que  $A = \mathbb{Z}'\mathbb{Z} = T'T$ ,

$$\begin{aligned} (d\mathbb{Z}) &= 2^{-m} \prod_{i=1}^m t_{ii}^{n-i-m+i-1} (d\{T'T\}) (H'_1 dH_1) \\ &= 2^{-m} \prod_{i=1}^m t_{ii}^{n-i-m+i-1} (d\{T'T\}) (H'_1 dH_1) \\ &= 2^{-m} \{\det T'T\}^{(n-m-1)/2} (d\{T'T\}) (H'_1 dH_1) \\ &= 2^{-m} \{\det A\}^{(n-m-1)/2} (dA) (H'_1 dH_1). \end{aligned}$$

On a donc la densité jointe de  $A = \mathbb{Z}'\mathbb{Z}$  et  $H_1 \in \mathcal{V}_{m,n}$ , obtenue par l'identité

$$\begin{aligned} f(\mathbb{Z})(d\mathbb{Z}) &= (2\pi)^{-mn/2} (\det \Sigma)^{-n/2} \operatorname{etr} \left( -\frac{1}{2} \Sigma^{-1} A \right) \operatorname{etr} \left( -\frac{1}{2} \Omega \right) \operatorname{etr} \left( \Sigma^{-1} M' H_1 T \right) (d\mathbb{Z}) \\ &= 2^{-m} (2\pi)^{-mn/2} (\det \Sigma)^{-n/2} \{\det A\}^{(n-m-1)/2} \operatorname{etr} \left( -\frac{1}{2} \Sigma^{-1} A \right) \operatorname{etr} \left( -\frac{1}{2} \Omega \right) (dA) \\ &\quad \times \operatorname{etr} \left( \Sigma^{-1} M' H_1 T \right) (H'_1 dH_1) \\ &= f(A, H_1) (dA) (H'_1 dH_1). \end{aligned}$$

On en déduit que la densité cherchée de  $A$  est donnée par la formule

$$\begin{aligned} f(A) &= 2^{-m} (2\pi)^{-mn/2} (\det \Sigma)^{-n/2} \{\det A\}^{(n-m-1)/2} \operatorname{etr} \left( -\frac{1}{2} \Sigma^{-1} A \right) \operatorname{etr} \left( -\frac{1}{2} \Omega \right) (dA) \\ &\quad \times \int_{H_1 \in \mathcal{V}_{m,n}} \operatorname{etr} \left( \Sigma^{-1} M' H_1 T \right) (H'_1 dH_1). \end{aligned}$$

Pour évaluer cette dernière intégrale, nous nous servons d'un cas particulier de (1.5.1) que nous réécrivons ci-dessous par commodité, avec les dimensions qui correspondent au problème qui nous intéresse. On a, en général, pour une fonction  $f(H_1, H_2)$  intégrable sur  $\mathcal{O}_n$ ,

$$\begin{aligned} &\int_{H = \begin{bmatrix} H_1 & H_2 \end{bmatrix} \in \mathcal{O}_n} f(H_1, H_2) (H' dH) \\ &= \int_{H_1 \in \mathcal{V}_{m,n}} \left\{ \int_{K \in \mathcal{O}_{n-m}} f(H_1, GK) (K' dK) \right\} (H'_1 dH_1). \end{aligned} \quad (3.4.13)$$

Comme, par ailleurs, on a, par (1.4.1),

$$\int_{K \in \mathcal{O}_{n-m}} (K' dK) = \frac{2^{n-m} \pi^{(m-n)^2/2}}{\Gamma_{m-n}(\frac{1}{2}(n-m))}, \quad (3.4.14)$$

On peut écrire

$$\begin{aligned}
& \int_{H_1 \in \mathcal{V}_{m,n}} \text{etr}(\Sigma^{-1} M' H_1 T) (H_1' dH_1) \\
&= \frac{\Gamma_{n-m}(\frac{1}{2}(n-m))}{2^{n-m} \pi^{(n-m)^2/2}} \left\{ \int_{K \in \mathcal{O}_{n-m}} (K' dK) \right\} \int_{H_1 \in \mathcal{V}_{m,n}} \text{etr}(\Sigma^{-1} M' H_1 T) (H_1' dH_1) \\
&= \frac{\Gamma_{n-m}(\frac{1}{2}(n-m))}{2^{n-m} \pi^{(n-m)^2/2}} \int_{H_1 \in \mathcal{V}_{m,n}} \left\{ \int_{K \in \mathcal{O}_{n-m}} \text{etr}(\Sigma^{-1} M' H_1 T) (K' dK) \right\} (H_1' dH_1) \\
&= \frac{\Gamma_{n-m}(\frac{1}{2}(n-m))}{2^{n-m} \pi^{(n-m)^2/2}} \int_{H=[H_1 \quad H_2] \in \mathcal{O}_n} \text{etr}(\Sigma^{-1} M' H_1 T) (H' dH).
\end{aligned}$$

Il est maintenant commode de faire intervenir la mesure de Haar invariante ( $dH$ ) sur le groupe orthogonal  $\mathcal{O}_n$ . Compte tenu de (1.4.8), que nous écrivons sous la forme

$$(dH) = \frac{\Gamma_n(\frac{1}{2}n)}{2^n \pi^{n^2/2}} (H' dH) \quad \Leftrightarrow \quad (dH) = \frac{2^n \pi^{n^2/2}}{\Gamma_n(\frac{1}{2}n)} (H' dH), \quad (3.4.15)$$

$$\begin{aligned}
\int_{H_1 \in \mathcal{V}_{m,n}} \text{etr}(\Sigma^{-1} M' H_1 T) (H_1' dH_1) &= \left\{ \frac{\Gamma_{n-m}(\frac{1}{2}(n-m))}{\Gamma_n(\frac{1}{2}n)} \right\} \left\{ \frac{2^n \pi^{n^2/2}}{2^{n-m} \pi^{(n-m)^2/2}} \right\} \\
&\quad \times \int_{H=[H_1 \quad H_2] \in \mathcal{O}_n} \text{etr}(\Sigma^{-1} M' H_1 T) (dH).
\end{aligned}$$

On se sert maintenant de (1.2.4), que nous réécrivons par commodité,

$$\Gamma_m(a) = \pi^{\frac{m(m-1)}{4}} \prod_{i=1}^m \Gamma\left(a - \frac{1}{2}(i-1)\right), \quad (3.4.16)$$

pour vérifier que

$$\begin{aligned}
& \left\{ \frac{\Gamma_{n-m}(\frac{1}{2}(n-m))}{\Gamma_n(\frac{1}{2}n)} \right\} \left\{ \frac{2^n \pi^{n^2/2}}{2^{n-m} \pi^{(n-m)^2/2}} \right\} \left\{ \frac{\Gamma_m(\frac{1}{2}n)}{2^m \pi^{mn/2}} \right\} \\
&= \left\{ \frac{\pi^{\frac{(n-m)(n-m-1)}{4}} \prod_{i=1}^{n-m} \Gamma\left(\frac{1}{2}(n-m-i+1)\right) \pi^{\frac{m(m-1)}{4}} \prod_{i=1}^m \Gamma\left(\frac{1}{2}(n-i+1)\right)}{\pi^{\frac{n(n-1)}{4}} \prod_{i=1}^n \Gamma\left(\frac{1}{2}(n-i+1)\right)} \right\} \\
&\quad \times \left\{ \frac{\pi^{n^2/2}}{\pi^{(n-m)^2/2} \pi^{mn/2}} \right\} = 1.
\end{aligned}$$

Ceci mène à la relation

$$\int_{H_1 \in \mathcal{V}_{m,n}} \text{etr}(\Sigma^{-1} M' H_1 T) (H_1' dH_1) = \frac{2^m \pi^{mn/2}}{\Gamma_m(\frac{1}{2}n)} \int_{H \in \mathcal{O}_n} \text{etr}(\Sigma^{-1} M' H_1 T) (dH).$$

Par (3.3.7), on montre ensuite que

$$\begin{aligned}
& \frac{2^m \pi^{mn/2}}{\Gamma_m(\frac{1}{2}n)} \int_{H \in \mathcal{O}_n} \text{etr}(\Sigma^{-1} M' H_1 T) (dH) = \frac{2^m \pi^{mn/2}}{\Gamma_m(\frac{1}{2}n)} {}_0F_1\left(\frac{1}{2}n; \frac{1}{4} T \Sigma^{-1} M' M \Sigma^{-1} T'\right) \\
&= \frac{2^m \pi^{mn/2}}{\Gamma_m(\frac{1}{2}n)} {}_0F_1\left(\frac{1}{2}n; \frac{1}{4} T \Omega \Sigma^{-1} T'\right) = \frac{2^m \pi^{mn/2}}{\Gamma_m(\frac{1}{2}n)} {}_0F_1\left(\frac{1}{2}n; \frac{1}{4} \Omega \Sigma^{-1} A\right),
\end{aligned}$$

avec  $A = T' T$  et  $\Omega = \Sigma^{-1} M' M$ , ce qui établit finalement (3.4.10).  $\square$

### 3.5 Propriétés de la loi de Wishart non centrée

Soit  $\mathbf{X} \stackrel{d}{=} N_{n,d}(\mathcal{M}, \mathbb{I}_n \otimes \Sigma)$  une variable normale matricielle associée à une matrice  $\Sigma > 0$ ,  $d \times d$ , définie positive. Par convention,  $\mathcal{M} = \mathbb{E}(\mathbf{X})$ , et  $\mathbb{I}_n \otimes \Sigma = \text{Var}(\text{Vec}(\mathbf{X}))$ , où

$$\mathbf{X} = \begin{bmatrix} X'_1 \\ \vdots \\ X'_n \end{bmatrix}, \quad \mathcal{M} = \mathbb{E}(\mathbf{X}) = \begin{bmatrix} \mu'_1 \\ \vdots \\ \mu'_n \end{bmatrix}, \quad \text{Vec}(\mathbf{X}) = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}.$$

L'identité en loi  $\mathbf{X} \stackrel{d}{=} N_{n,d}(\mathcal{M}, \mathbb{I}_n \otimes \Sigma)$  est équivalente au fait que les vecteurs colonnes de  $\mathbf{X}'$  dans  $\mathbb{R}^d$  sont indépendants, de lois respectives  $X_1 \stackrel{d}{=} N_d(\mu_1, \Sigma), \dots, X_n \stackrel{d}{=} N_d(\mu_n, \Sigma)$ . Par la définition 3.4.1,

$$A := \mathbf{X}'\mathbf{X} = \sum_{i=1}^n X_i X'_i \stackrel{d}{=} W_d(n, \Sigma, \Omega) \stackrel{d}{=} W_d(n, \Sigma, \Sigma^{-1} \mathcal{M}'\mathcal{M}), \quad (3.5.17)$$

suit une loi de Wishart non centrée, de paramètre de décentrement  $\Omega = \Sigma^{-1} \mathcal{M}'\mathcal{M}$ . On notera que le paramètre de décentrement  $\Omega$  est une matrice  $d \times d$ . Cette matrice n'est, en général, pas symétrique. Elle n'est pas non plus une matrice  $d \times d$  quelconque, puisque définie comme le produit des deux matrices  $\Sigma > 0$  et  $\mathcal{M}'\mathcal{M} \geq 0$ . En particulier, elle est telle que, si

$$\mathcal{M} = \begin{bmatrix} \mu'_1 \\ \vdots \\ \mu'_n \end{bmatrix},$$

alors, on a l'équivalence

$$\text{tr}(\Omega) = \text{tr}(\mathcal{M}\Sigma^{-1}\mathcal{M}') = \sum_{i=1}^n \mu'_i \Sigma^{-1} \mu = 0 \quad \Leftrightarrow \quad \mu_1 = \dots = \mu_n = \mathbf{0}_d.$$

La relation  $\mathcal{M} = \mathbb{O}_{n,d}$  est donc équivalente à l'égalité  $\Omega = \mathbb{O}_{d,d}$ . Lorsque cette propriété est vérifiée, i.e., lorsque  $\Omega = \mathbb{O}_{d,d}$ , la loi de Wishart correspondante devient la loi de Wishart centrée usuelle, ce qui se traduit par l'identité en loi

$$W_d(n, \Sigma, \mathbb{O}_{n,d}) \stackrel{d}{=} W_d(n, \Sigma).$$

**Remarque 3.5.2.** On notera que la loi de Wishart centrée peut être définie pour une matrice  $\Sigma \geq 0$ , pas nécessairement définie positive. Par contre, la définition que nous avons adoptée pour la loi de Wishart non centrée  $A \stackrel{d}{=} W_d(n, \Sigma, \Omega)$ , requiert que  $\Sigma > 0$  soit définie positive. Cette condition est nécessaire pour donner un sens à  $\Omega$ . Il est, bien entendu, possible d'étendre la notion de loi de Wishart non centrée à toutes les expressions de la forme  $A := \mathbf{X}'\mathbf{X}$ , lorsque  $\mathbf{X} \stackrel{d}{=} N_{n,d}(\mathcal{M}, \mathbb{I}_n \otimes \Sigma)$ , et pour une matrice de variances-covariances  $\Sigma \geq 0$  pas nécessairement définie positive. Le paramétrage de la loi de  $A$  devient alors un peu plus délicat. Le fait que  $\Sigma \geq 0$  n'est pas définie positive est équivalent à la condition que les vecteurs composants de  $(\mathbf{X} - \mathcal{M})'$  prennent leurs valeurs dans  $\mathcal{F}$ , où  $\mathcal{F} \subset \mathbb{R}^d$  est un sous-espace vectoriel de  $\mathbb{R}^d$  (caractérisé par  $\Sigma$ ), de dimension inférieure à  $d$ . La difficulté vient alors du fait que les vecteurs composants de  $\mathcal{M}' = \mathbb{E}(\mathbf{X}')$  ne prennent pas nécessairement leurs valeurs dans  $\mathcal{F}$ .

**Proposition 3.5.1.** Lorsque  $d = 1$ ,  $\Sigma = [\sigma^2]$ , et  $\Omega = [\delta]$ , la loi de Wishart  $W_1(n, \sigma^2, \delta)$  se ramène à une loi du  $\chi^2$  non centrée, par la formule

$$W_1(n, \sigma^2, \delta) \stackrel{d}{=} \sigma^2 \chi_n^2(\delta). \quad (3.5.18)$$

**Preuve.** La loi de Wishart  $W_1(n, \sigma^2, \delta)$  est la loi de  $X_1^2 + \dots + X_n^2$ , où les variables aléatoires  $X_1 \stackrel{d}{=} N(\sigma\sqrt{\delta}, \sigma^2)$ , et  $X_2 \stackrel{d}{=} \dots \stackrel{d}{=} X_n \stackrel{d}{=} N(0, \sigma^2)$ , sont indépendantes. On peut alors poser  $X_i = \sigma Y_i$ , pour  $i = 1, \dots, n$ . Ces relations impliquent que  $X_1^2 + \dots + X_n^2 = \sigma^2(Y_1^2 + \dots + Y_n^2)$ , où  $Y_1 \stackrel{d}{=} N(\sqrt{\delta}, 1)$ , et  $Y_2 \stackrel{d}{=} \dots \stackrel{d}{=} Y_n \stackrel{d}{=} N(0, 1)$ , sont indépendantes. Par conséquent, on a bien  $Y_1^2 + \dots + Y_n^2 \stackrel{d}{=} \chi_n^2(\delta)$ , et donc, (3.5.18).  $\square$

**Proposition 3.5.2.** *L'espérance de  $A \stackrel{d}{=} W_d(n, \Sigma, \Omega)$  est donnée par*

$$\mathbb{E}(A) = n\Sigma + \mathbf{M}'\mathbf{M} = n\Sigma + \Sigma\Omega. \quad (3.5.19)$$

**Preuve.** Soit  $\mathbf{X} \stackrel{d}{=} N_{n,d}(\mathbf{M}, \mathbb{I}_n \otimes \Sigma)$ . On a alors  $\mathbf{Y} = \mathbf{X} - \mathbf{M} \stackrel{d}{=} N_{n,d}(\mathbb{O}_{n,d}, \mathbb{I}_n \otimes \Sigma)$ . Posons alors

$$\mathbf{Y} = \begin{bmatrix} Y'_1 \\ \vdots \\ Y'_n \end{bmatrix}.$$

La propriété  $\mathbf{Y} = \mathbf{X} - \mathbf{M} \stackrel{d}{=} N_{n,d}(\mathbb{O}_{n,d}, \mathbb{I}_n \otimes \Sigma)$  est équivalente au fait que  $Y_1 \stackrel{d}{=} \dots \stackrel{d}{=} Y_n \stackrel{d}{=} N_d(\mathbb{O}_{d,1}, \Sigma)$ . On en déduit que

$$\mathbb{E}(\mathbf{Y}'\mathbf{Y}) = \sum_{i=1}^n \mathbb{E}(Y_i Y'_i) = n\Sigma.$$

On écrit ensuite que  $A = \mathbf{X}'\mathbf{X} \stackrel{d}{=} W_d(n, \Sigma, \Sigma^{-1}\mathbf{M}'\mathbf{M}) \stackrel{d}{=} W_d(n, \Sigma, \Omega)$ , avec  $\Omega = \Sigma^{-1}\mathbf{M}'\mathbf{M}$ . On en déduit que

$$\begin{aligned} n\Sigma &= \mathbb{E}(\mathbf{Y}'\mathbf{Y}) = \mathbb{E}((\mathbf{X} - \mathbf{M})'(\mathbf{X} - \mathbf{M})) = \mathbb{E}(\mathbf{X}'\mathbf{X}) - \mathbb{E}(\mathbf{X}'\mathbf{M}) - \mathbb{E}(\mathbf{M}'\mathbf{X}) + \mathbf{M}'\mathbf{M} \\ &= \mathbb{E}(A) - \mathbf{M}'\mathbf{M} = \mathbb{E}(A) - \Sigma\Omega, \end{aligned}$$

d'où la conclusion (3.5.19).  $\square$

Les propositions qui suivent sont des conséquences plus ou moins directes des théorèmes 3.4.2 et 3.4.1, dont nous rappelons les conclusions. Sous l'hypothèse que  $\Sigma > 0$ , lorsque  $n < d$ , la matrice  $A \stackrel{d}{=} W_d(n, \Sigma, \Omega)$  n'est pas de plein rang. Par contre, c'est le cas pour tout  $n \geq d$ , et, par le théorème 3.4.2, la matrice aléatoire,  $d \times d$ ,  $A$  a alors une densité sur l'espace des matrices  $d \times d$  définies positives. Cette densité est donnée, pour  $A > 0$ , par

$$f(A) = \frac{2^{-dn/2} (\det \Sigma)^{-n/2}}{\Gamma_d(\frac{1}{2}n)} (\det A)^{\frac{n}{2} - \frac{d+1}{2}} \text{etr}(-\frac{1}{2}\Sigma^{-1}A) \text{etr}(-\frac{1}{2}\Omega) F_1(\frac{1}{2}n; \frac{1}{4}\Omega\Sigma^{-1}A). \quad (3.5.20)$$

Par le théorème 3.4.1, la fonction caractéristique de  $A \stackrel{d}{=} W_d(n, \Sigma, \Omega)$  est donnée, pour toute matrice  $d \times d$  symétrique  $\Theta$ , par

$$\phi(\Theta) = \mathbb{E}(\text{etr}(iA\Theta)) = \det(\mathbb{I}_d - 2i\Theta\Sigma)^{-n/2} \text{etr}(-\frac{1}{2}\Omega) \text{etr}(\frac{1}{2}\Omega(\mathbb{I}_d - 2i\Theta\Sigma)^{-1}). \quad (3.5.21)$$

On obtient alors les propositions suivantes.

**Proposition 3.5.3.** *Si  $A_i \stackrel{d}{=} W_d(n_i, \Sigma, \Omega_i)$ ,  $i = 1, \dots, m$  sont des matrices de Wishart indépendantes, alors,*

$$\sum_{i=1}^m A_i \stackrel{d}{=} W_d\left(\sum_{i=1}^m n_i, \Sigma, \sum_{i=1}^m \Omega_i\right). \quad (3.5.22)$$

**Preuve.** La fonction caractéristique de  $S := \sum_{i=1}^m A_i$  est égale au produit des fonctions caractéristiques de  $A_i$ , soit, par (3.5.21),

$$\begin{aligned} \phi_S(\Theta) &= \prod_{i=1}^m \phi_{A_i}(\Theta) = \prod_{i=1}^m \left\{ \det(\mathbb{I}_d - 2i\Theta\Sigma)^{-n_i/2} \text{etr}(-\frac{1}{2}\Omega_i) \text{etr}(\frac{1}{2}\Omega_i(\mathbb{I}_d - 2i\Theta\Sigma)^{-1}) \right\} \\ &= \det(\mathbb{I}_d - 2i\Theta\Sigma)^{-\frac{1}{2}\sum_{i=1}^m n_i} \text{etr}\left(-\frac{1}{2}\sum_{i=1}^m \Omega_i\right) \text{etr}\left(\frac{1}{2}\left\{\sum_{i=1}^m \Omega_i\right\}(\mathbb{I}_d - 2i\Theta\Sigma)^{-1}\right), \end{aligned}$$

ce qui implique (3.5.22), puisque la fonction caractéristique ci-dessus est celle d'une loi de Wishart, et du fait que cette fonction caractéristique détermine complètement la loi de  $S$ .  $\square$

**Proposition 3.5.4.** Si  $\mathbf{X} \stackrel{d}{=} N_{n,d}(\mathcal{M}, \mathbb{I}_n \otimes \Sigma)$ , et  $A = \mathbf{X}'\mathbf{X} \stackrel{d}{=} W_d(n, \Sigma, \Omega)$ , avec  $\Omega = \Sigma^{-1}\mathcal{M}'\mathcal{M}$ , et si  $P$  est une matrice  $m \times d$  de rang  $m \leq d$ , alors,

$$PAP' = \mathbf{X}'\mathbf{X}P' \stackrel{d}{=} W_m(n, P\Sigma P', (P\Sigma P')^{-1}P\mathcal{M}'\mathcal{M}P') \stackrel{d}{=} W_m(n, P\Sigma P', (P\Sigma P')^{-1}P\Sigma\Omega P'). \quad (3.5.23)$$

**Preuve.** On remarquera que, sous les hypothèses de la proposition 3.5.4,

$$\mathbf{X}P' = \begin{bmatrix} X_1'P' \\ \vdots \\ X_n'P' \end{bmatrix} \stackrel{d}{=} N_{n,m}(\mathcal{M}P', \mathbb{I}_n \otimes P\Sigma P').$$

La conclusion (3.5.23) est donc une conséquence directe de (3.5.17), après y avoir remplacé  $\Sigma$  par  $P\Sigma P'$ , et  $\mathcal{M}$ , par  $\mathcal{M}P'$ .  $\square$

**Corollaire 3.5.1.** Si  $A \stackrel{d}{=} W_d(n, \Sigma, \Omega)$ , et si  $\alpha \in \mathbb{R}^d \neq \mathbb{O}_d$ , alors

$$\frac{\alpha'A\alpha}{\alpha'\alpha} \stackrel{d}{=} \chi_n^2 \left( \frac{\alpha'\Sigma\Omega\alpha}{\alpha'\Sigma\alpha} \right).$$

**Preuve.** On applique la proposition 3.5.4 au cas où  $P = \alpha'$ , avec  $m = 1$ . On déduit alors de (3.5.23) que

$$\alpha'A\alpha \stackrel{d}{=} W_1 \left( n, \alpha'\Sigma\alpha, \frac{\alpha'\Sigma\Omega\alpha}{\alpha'\Sigma\alpha} \right),$$

ce qui implique le résultat, par (3.5.18).  $\square$

Le théorème suivant généralise la décomposition de Bartlett aux lois de Wishart non centrées.

**Théorème 3.5.3.** Soient  $X_1 \stackrel{d}{=} N_d(\boldsymbol{\mu}_1, \mathbb{I}_d), \dots, X_n \stackrel{d}{=} N_d(\boldsymbol{\mu}_n, \mathbb{I}_d)$ , des vecteurs indépendants, tels que

$$\boldsymbol{\mu}_i = \begin{bmatrix} \mu_{1,i} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{pour } i = 1, \dots, n.$$

Alors, pour tout  $n \geq d$ , on a l'identité

$$A \stackrel{d}{=} W_d \left( n, \mathbb{I}, \sum_{i=1}^n \boldsymbol{\mu}_i \boldsymbol{\mu}_i' \right) = T'T \quad \text{où } T = \begin{bmatrix} t_{1,1} & \dots & t_{1,d} \\ \vdots & \ddots & \vdots \\ 0 & \dots & t_{d,d} \end{bmatrix},$$

est une matrice triangulaire à diagonale  $t_{j,j} > 0$  positive, pour  $j = 1, \dots, d$ , telle que tous les  $t_{i,j}$ ,  $1 \leq i \leq j \leq d$  soient indépendants, avec les lois

$$t_{1,1}^2 \stackrel{d}{=} \chi_n^2(\delta), \quad \delta = \sum_{i=1}^n \boldsymbol{\mu}_i' \boldsymbol{\mu}_i = \sum_{i=1}^n \mu_{1,i}^2; \quad t_{j,j} \stackrel{d}{=} \chi_{n-j+1}^2, \quad 2 \leq j \leq d; \quad t_{j,\ell} \stackrel{d}{=} N(0, 1), \quad 1 \leq j < \ell \leq d.$$

**Preuve.** Posons, dans l'expression (3.5.20) de la densité de  $A$ ,  $\Sigma = \mathbb{I}_d$ , et

$$\Omega = \text{diag}(\delta, 0, \dots, 0) = \begin{bmatrix} \delta & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix} = \mathcal{M}'\mathcal{M},$$

avec

$$\mathcal{M} = [ \mathbf{m} \quad \mathbb{O}_n \quad \dots \quad \mathbb{O}_n ],$$

avec  $\mathbf{m} \in \mathbb{R}^n$  et  $\delta = \mathbf{m}'\mathbf{m}$ . Observons que

$$\frac{1}{4}\Omega\Sigma^{-1}A = \begin{bmatrix} \delta & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} a_{1,1} & \dots & a_{1,d} \\ \vdots & \ddots & \vdots \\ a_{d,1} & \dots & a_{d,d} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} \delta a_{1,1} & \dots & \delta a_{1,d} \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix},$$

$a$ , comme valeurs propres,  $y_1 = \frac{1}{4}\delta a_{1,1}$ , et  $y_2 = \dots = y_d = 0$ , la valeur propre 0 étant, ici, de multiplicité  $d-1$ . Considérons alors un polynôme zonal  $C_\kappa(\frac{1}{4}\Omega\Sigma^{-1}A)$ , de degré  $|\kappa| = k$ , et associé à la  $d$ -partition  $\kappa = (k_1, \dots, k_d)$  de  $k = k_1 + \dots + k_d$ , avec  $k_1 \geq \dots \geq k_d \geq 0$ . Les monômes de ce polynôme, engendrés par des termes de la forme  $c_{r_1, \dots, r_d} y_1^{r_1} \dots y_d^{r_d}$ , avec  $r_1 \geq \dots \geq r_d \geq 0$ , sont nuls, chaque fois qu'il existe deux indices distincts  $r_i \geq 1$  et  $r_j \geq 1$ ,  $1 \leq i < j \leq d$  dans la partition  $r = (r_1, \dots, r_d)$  de  $k = |r|$ . La seule partition  $r$  possible qui n'annule pas ce monome est donc  $\kappa_0 := (k, 0, \dots, 0)$ . Comme elle est de poids maximal, elle ne peut figurer que dans le seul polynôme zonal  $C_{\kappa_0}$ . Maintenant, observons qu'il existe un coefficient  $\lambda \neq 0$ , tel que, pour toute matrice  $Z$  de valeurs propres  $z_1, \dots, z_d$ ,

$$C_{\kappa_0}(Z) = \lambda(z_1^k + \dots + z_d^k) + \text{termes de poids inférieur.}$$

Comme, par ailleurs,

$$\sum_{|\kappa|=k} C_\kappa(Z) = (\text{tr } Z)^k,$$

le coefficient  $\lambda$  est nécessairement égal à 1, puisque les termes  $z_1^k + \dots + z_d^k$  sont affectés de ce coefficient dans le développement de  $(\text{tr } Z)^k$ , et du fait que ces derniers n'apparaissent pas dans les polynômes  $C_\kappa$  de poids inférieur. On en déduit que  $C_\kappa(\frac{1}{4}\Omega\Sigma^{-1}A) = 0$ , sauf si  $\kappa = (k, 0, \dots, 0)$ , de sorte que

$$C_{\kappa_0}(\frac{1}{4}\Omega\Sigma^{-1}A) = \left(\frac{1}{4}\delta a_{1,1}\right)^k.$$

Par (3.3.2), la forme spécifique de  $\kappa_0 := (k, 0, \dots, 0) = (k_1, \dots, k_d)$  implique que, pour tout  $a$ ,

$$(a)_{\kappa_0} = \prod_{i=1}^d \left(a - \frac{1}{2}(i-1)\right)_{k_i} = (a)_k.$$

La fonction hypergéométrique matricielle de la formule (3.5.20) se réduit alors à une loi hypergéométrique univariée, grâce aux identités

$$\begin{aligned} {}_0F_1\left(\frac{1}{2}n; \frac{1}{4}\Omega\Sigma^{-1}A\right) &= \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{|\kappa|=k} \frac{C_\kappa(\frac{1}{4}\Omega\Sigma^{-1}A)}{\left(\frac{1}{2}n\right)_\kappa} = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{C_{\kappa_0}(\frac{1}{4}\Omega\Sigma^{-1}A)}{\left(\frac{1}{2}n\right)_{\kappa_0}} \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} \frac{\left(\frac{1}{4}\delta a_{1,1}\right)^k}{\left(\frac{1}{2}n\right)_k} = {}_0F_1\left(\frac{1}{2}n; \frac{1}{4}\delta a_{1,1}\right). \end{aligned}$$

Compte tenu de ces relations, la densité (3.5.20) de  $A$  peut être réécrite sous la forme

$$f(A) = \frac{2^{-dn/2}}{\Gamma_d(\frac{1}{2}n)} (\det A)^{\frac{n}{2} - \frac{d+1}{2}} \text{etr}\left(-\frac{1}{2}A\right) \exp\left(-\frac{1}{2}\delta\right) {}_0F_1\left(\frac{1}{2}n; \frac{1}{4}\delta a_{1,1}\right).$$

Réalisons maintenant le changement de variable  $A = T'T$ , où  $T = [t_{i,j}]$  est une matrice triangulaire supérieure à diagonale positive. On a la formule de changement de variables (1.1.7)

$$[dA] = 2^d \prod_{j=1}^d t_{j,j}^{d-j+1} [dT].$$



Compte tenu de la formule (1.2.4), on a les identités

$$\Gamma_d(r) = \pi^{\frac{d(d-1)}{4}} \prod_{j=1}^d \Gamma\left(r - \frac{1}{2}(j-1)\right), \quad \det A = \prod_{j=1}^d t_{j,j}^2, \quad \text{tr}(A) = \text{tr}(T'T) = \sum_{1 \leq i \leq j \leq d} t_{i,j}^2, \quad a_{1,1} = t_{1,1}^2.$$

On en déduit la densité jointe de  $z_1 = t_{1,1}^2, \dots, z_d = t_{d,d}^2$  et des  $t_{i,j}$  pour  $1 \leq i < j \leq d$ , qui est égale à

$$\left\{ \prod_{1 \leq i < j \leq d} \frac{e^{-\frac{1}{2}t_{i,j}^2}}{\sqrt{2\pi}} \right\} \times \left\{ \prod_{j=2}^d \frac{\left(\frac{1}{2}\right)^{\frac{n-j+1}{2}}}{\Gamma\left(\frac{n-j+1}{2}\right)} z_j^{\frac{n-j+1}{2}-1} e^{-\frac{1}{2}z_j} \right\} \times \left\{ \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} z_1^{\frac{n}{2}-1} e^{-\frac{1}{2}z_1} {}_0F_1\left(\frac{1}{2}n; \frac{1}{4}\delta z_1\right) \right\},$$

expression dans laquelle on reconnaît des lois indépendantes,  $N(0, 1)$  pour les  $t_{i,j}$ ,  $1 \leq i < j \leq d$ ,  $\chi_{n_j+1}^2$  pour les  $z_j = t_{j,j}^2$ ,  $2 \leq j \leq d$ , et  $\chi_n^2(\delta)$ , avec  $\delta = \mathbf{m}'\mathbf{m}$ , pour  $z_1 = t_{1,1}^2$ .  $\square$

Le théorème suivant s'obtient par développement asymptotique de la densité de la loi de Wishart non centrée. Son résultat sera admis.

**Théorème 3.5.4.** *On suppose que  $A \stackrel{d}{=} W_d(n_1, \Sigma)$  et  $B \stackrel{d}{=} W_d(n_2, \Sigma, \Omega)$  sont deux matrices de Wishart indépendantes. On pose*

$$\Lambda(n_1 + n_2, d, n_2; \Omega) = \frac{\det A}{\det(A + B)}.$$

Alors, sous l'hypothèse que  $d$ ,  $n_2$  et  $\Omega$  sont fixés, et que  $n_1 \rightarrow \infty$ , on a le développement asymptotique

$$\begin{aligned} \mathbb{P}\left(-\left\{n_1 - \frac{1}{2}(d + n_2 + 1)\right\} \log \Lambda(n_1 + n_2, d, n_2; \Omega) \leq x\right) &= \mathbb{P}\left(\chi_{dn_2}^2(\text{tr}(\Omega)) \leq x\right) \\ &+ \frac{1}{4\left\{n_1 - \frac{1}{2}(d + n_2 + 1)\right\}} \left\{ (n_2 + d + 1)(\text{tr} \Omega) \mathbb{P}\left(\chi_{dn_2+2}^2(\text{tr}(\Omega)) \leq x\right) \right. \\ &\quad - \left. \left( (d + n_2 + 1)\text{tr}(\Omega) - \text{tr}(\Omega^2) \right) \mathbb{P}\left(\chi_{dn_2+4}^2(\text{tr}(\Omega)) \leq x\right) \right. \\ &\quad \left. - (\text{tr} \Omega^2) \mathbb{P}\left(\chi_{dn_2+4}^2(\text{tr}(\Omega)) \leq x\right) \right\} + O(1/n_1^2). \end{aligned} \quad (3.5.24)$$

**Remarque 3.5.3.** *En pratique, on se limite le plus souvent à l'une ou l'autre des relations suivantes, approximativement satisfaites, lorsque  $n_1 \rightarrow \infty$ ,*

$$\begin{aligned} \mathbb{P}\left(-\left\{n - \frac{1}{2}(d + n_2 + 1)\right\} \log \Lambda(n_1 + n_2, d, n_2; \Omega) \leq x\right) &= \mathbb{P}\left(\chi_{dn_2}^2(\text{tr} \Omega) \leq x\right) + O\left(\frac{1}{n_1}\right), \\ \mathbb{P}\left(-n \log \Lambda(n_1 + n_2, d, n_2; \Omega) \leq x\right) &= \mathbb{P}\left(\chi_{dn_2}^2(\text{tr} \Omega) \leq x\right) + o(1). \end{aligned}$$

## Chapitre 4

# Applications statistiques du modèle linéaire multivarié.

### 4.1 Analyse du modèle linéaire multivarié classique.

#### 4.1.1 Le modèle linéaire univarié classique.

##### A. Généralités.

Le modèle linéaire univarié classique (ou standard) est décrit par la relation

$$\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} = A\boldsymbol{\theta} + \boldsymbol{\varepsilon} \stackrel{d}{=} N_n(A\boldsymbol{\theta}, \sigma^2 \mathbb{I}_n), \quad (4.1.1)$$

où  $A$  est une matrice  $(n \times p)$  connue, de plein rang de colonnes, vérifiant donc  $\text{rg}(A) = p \leq n$ ,  $\boldsymbol{\theta} \in \mathbb{R}^p$  un paramètre vectoriel inconnu, et  $\sigma^2 > 0$  une variance inconnue. Le "vecteur d'erreurs"  $\boldsymbol{\varepsilon} \in \mathbb{R}^n$  suit une loi normale centrée de paramètres  $\boldsymbol{\varepsilon} \stackrel{d}{=} N_n(\mathbb{O}_n, \sigma^2 \mathbb{I}_n)$ . Dans ce modèle, les composantes  $X_1, \dots, X_n$  du "vecteur de données"  $\mathbf{X} \in \mathbb{R}^n$  suivent des lois normales indépendantes de même variance  $\sigma^2$ .

Les estimations du maximum de vraisemblance  $\hat{\boldsymbol{\theta}}$  et  $\hat{\sigma}^2$  de  $\boldsymbol{\theta}$  et  $\sigma^2$  sont alors mutuellement indépendantes, et données par

$$\hat{\boldsymbol{\theta}} = (A'A)^{-1}A'\mathbf{X} \stackrel{d}{=} N_p(\boldsymbol{\theta}, \sigma^2(A'A)^{-1}) \quad \text{et} \quad \hat{\sigma}^2 = \frac{1}{n} \left\{ (\mathbf{X} - A\hat{\boldsymbol{\theta}})'(\mathbf{X} - A\hat{\boldsymbol{\theta}}) \right\} \stackrel{d}{=} \frac{\sigma^2 \chi_{n-p}^2}{n}. \quad (4.1.2)$$

**Remarque 4.1.1.** 1°) *Le fait que  $\hat{\boldsymbol{\theta}}$  et  $\hat{\sigma}^2$  sont les estimateurs du maximum de vraisemblance de  $\boldsymbol{\theta}$  et  $\sigma^2$  sera établi plus loin (voir la démonstration du théorème 4.1.1), dans le contexte, plus général, du modèle linéaire multivarié classique.*

2°) *Il y a coïncidence entre l'estimateur linéaire sans biais à variance minimale [ELSBVM] et l'estimateur du maximum de vraisemblance pour  $\boldsymbol{\theta}$ . Tous deux sont donnés par  $\hat{\boldsymbol{\theta}}$ . En particulier, dans ce modèle,*

*$A\hat{\boldsymbol{\theta}}$  est la projection orthogonale (relativement au produit scalaire usuel  $\langle u, v \rangle_{\mathbb{I}} = u'v$  de  $\mathbb{R}^n$ ) de  $\mathbf{X}$  sur le sous-espace vectoriel  $\mathcal{F} = \{A\boldsymbol{\theta} : \boldsymbol{\theta} \in \mathbb{R}^p\}$  de  $\mathbb{R}^n$  ;*

*$\mathbf{X} - A\hat{\boldsymbol{\theta}} = (\mathbb{I} - A(A'A)^{-1}A')\mathbf{X}$  est la projection orthogonale (relativement au produit scalaire usuel  $\langle u, v \rangle_{\mathbb{I}} = u'v$  de  $\mathbb{R}^n$ ) de  $\mathbf{X}$  sur le sous-espace vectoriel  $\mathcal{F}^\perp$ , orthogonal à  $\mathcal{F}$  dans  $\mathbb{R}^n$ .*

*Compte tenu de ces propriétés,  $A\hat{\boldsymbol{\theta}}$  et  $\mathbf{X} - A\hat{\boldsymbol{\theta}}$  sont indépendants, et orthogonaux (relativement au produit scalaire usuel  $\langle u, v \rangle_{\mathbb{I}} = u'v$  de  $\mathbb{R}^n$ ) dans  $\mathbb{R}^n$ . On en déduit l'indépendance de  $\hat{\boldsymbol{\theta}}$  et  $\hat{\sigma}^2$ .*

3°) On notera les expressions alternatives suivantes à la définition (4.1.2) de  $\hat{\sigma}^2$ . On a, de par l'orthogonalité de  $A\hat{\boldsymbol{\theta}}$  et  $\mathbf{X} - A\hat{\boldsymbol{\theta}}$ ,

$$\begin{aligned}\hat{\sigma}^2 &= n^{-1}\{\mathbf{X}'\mathbf{X} - (A\hat{\boldsymbol{\theta}})'(A\hat{\boldsymbol{\theta}})\} = n^{-1}\mathbf{X}'(\mathbb{I}_n - A(A'A)^{-1}A')\mathbf{X} \\ &= n^{-1}\{\mathbf{X}'\mathbf{X} - (A\hat{\boldsymbol{\theta}})'(A\hat{\boldsymbol{\theta}} - \mathbf{X} + \mathbf{X})\} = n^{-1}\{\mathbf{X}'\mathbf{X} - \hat{\boldsymbol{\theta}}'A'\mathbf{X}\}.\end{aligned}\quad (4.1.3)$$

### B. Test de l'hypothèse que $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ .

Soit  $\boldsymbol{\theta}_0 \in \mathbb{R}^p$  un vecteur donné. Nous désirons tester l'hypothèse

$$(H.0) \quad \boldsymbol{\theta} = \boldsymbol{\theta}_0,$$

contre l'alternative générale

$$(H.1) \quad \boldsymbol{\theta} \in \mathbb{R}^p.$$

Pour cela, on utilise le fait que la statistique

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'(A'A)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \stackrel{d}{=} \sigma^2 \chi_p^2\left(\frac{1}{\sigma^2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'(A'A)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right),$$

est indépendante de  $\hat{\sigma}^2$ . On utilise alors la statistique

$$\mathbf{S} = \left\{\frac{n-p}{pn\hat{\sigma}^2}\right\}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'(A'A)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \stackrel{d}{=} F_{p,n-p}\left(\frac{1}{\sigma^2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'(A'A)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right).$$

On rejette (H.0) lorsque  $\mathbf{S}$  excède la quantile supérieure d'ordre  $\alpha$ , désignée par  $F_{\alpha;p,n-p}$ , de la loi de Fisher  $F_{p,n-p}$ . La puissance  $1 - \beta$  du test basé sur  $\mathbf{S}$  s'évalue par la formule

$$1 - \beta = \mathbb{P}\left(F_{r,n-p}\left(\frac{1}{\sigma^2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'(A'A)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right) \geq F_{\alpha;p,n-p}\right).$$

### C. Test de l'hypothèse que $C\boldsymbol{\theta} = \mathbb{O}_r$ .

Soit  $C$  une matrice ( $r \times p$ ), connue et de rang  $r \leq p$ . Nous désirons tester l'hypothèse

$$(H.0) \quad C\boldsymbol{\theta} = \mathbb{O}_r,$$

contre l'alternative générale

$$(H.1) \quad C\boldsymbol{\theta} \in \mathbb{R}^r.$$

Pour cela, considérons, dans l'espace  $\mathbb{R}^p$  où  $\boldsymbol{\theta}$  prend ses valeurs, les deux sous-espaces vectoriels suivants de  $\mathbb{R}^p$  :

$$\mathcal{G} = \{\boldsymbol{\alpha} \in \mathbb{R}^p : C\boldsymbol{\alpha} = \mathbb{O}_r\};$$

$$\mathcal{G}^\perp = \{(A'A)^{-1}C'\boldsymbol{\beta} : \boldsymbol{\beta} \in \mathbb{R}^r\} = \{D\boldsymbol{\beta} : \boldsymbol{\beta} \in \mathbb{R}^r\}, \text{ où } D = (A'A)^{-1}C'.$$

On constate que les sous-espaces vectoriels  $\mathcal{G}$  et  $\mathcal{G}^\perp$ , de  $\mathbb{R}^p$ , de dimensions respectives  $\dim \mathcal{G} = p - r$  et  $\dim \mathcal{G}^\perp = r$ , sont orthogonaux relativement au produit scalaire  $\langle u, v \rangle_{A'A} = u'(A'A)v$  dans  $\mathbb{R}^p$ , et de somme directe égale à  $\mathbb{R}^p$ . La projection orthogonale (relativement au produit scalaire  $\langle u, v \rangle_{A'A} = u'(A'A)v$ ) de  $\hat{\boldsymbol{\theta}}$  sur  $\mathcal{G}^\perp$  est alors donnée par

$$\begin{aligned}\hat{\boldsymbol{\theta}} - \hat{\hat{\boldsymbol{\theta}}} &= D\left(D'(A'A)D\right)^{-1}D'(A'A)\hat{\boldsymbol{\theta}} \\ &= (A'A)^{-1}C'\left(\{(A'A)^{-1}C'\}'(A'A)\{(A'A)^{-1}C'\}\right)^{-1}\{(A'A)^{-1}C'\}'(A'A)\hat{\boldsymbol{\theta}} \\ &= (A'A)^{-1}C'\left(C(A'A)^{-1}C'\right)^{-1}C\hat{\boldsymbol{\theta}},\end{aligned}$$

et la projection orthogonale (relativement au produit scalaire  $\langle u, v \rangle_{A'A} = u'(A'A)v$ ) de  $\widehat{\boldsymbol{\theta}}$  sur  $\mathcal{G}$  est donnée par

$$\widehat{\widehat{\boldsymbol{\theta}}} = \widehat{\boldsymbol{\theta}} - \left\{ \widehat{\boldsymbol{\theta}} - \widehat{\widehat{\boldsymbol{\theta}}} \right\} = \left\{ \mathbb{I} - (A'A)^{-1}C'(C(A'A)^{-1}C')^{-1}C \right\} \widehat{\boldsymbol{\theta}}.$$

Comme, par (4.1.2),  $\widehat{\boldsymbol{\theta}} \stackrel{d}{=} N_p(\boldsymbol{\theta}, \sigma^2(A'A)^{-1})$ , les statistiques  $\widehat{\widehat{\boldsymbol{\theta}}}$  et  $\widehat{\boldsymbol{\theta}} - \widehat{\widehat{\boldsymbol{\theta}}}$  sont indépendantes (comme projections de  $\widehat{\boldsymbol{\theta}}$  sur des sous-espaces orthogonaux, relativement au produit scalaire  $\langle u, v \rangle_{A'A} = u'(A'A)v$ ). De plus, on a l'identité en loi

$$\begin{aligned} (\widehat{\boldsymbol{\theta}} - \widehat{\widehat{\boldsymbol{\theta}}})' A'A (\widehat{\boldsymbol{\theta}} - \widehat{\widehat{\boldsymbol{\theta}}}) &= \widehat{\boldsymbol{\theta}}' C' (C(A'A)^{-1}C')^{-1} C(A'A)^{-1} (A'A) (A'A)^{-1} C' (C(A'A)^{-1}C')^{-1} C \widehat{\boldsymbol{\theta}} \\ &= \widehat{\boldsymbol{\theta}}' C' (C(A'A)^{-1}C')^{-1} C \widehat{\boldsymbol{\theta}} \stackrel{d}{=} \sigma^2 \chi_r^2 \left( \frac{1}{\sigma^2} \boldsymbol{\theta}' C' (C(A'A)^{-1}C')^{-1} C \boldsymbol{\theta} \right). \end{aligned}$$

Cette dernière statistique sur une loi du  $\sigma^2 \chi_r^2$  centrée sous l'hypothèse (H.0) que  $C\boldsymbol{\theta} = \mathbb{O}_r$ . De plus, elle est indépendante de  $n\widehat{\sigma}^2 \stackrel{d}{=} \sigma^2 \chi_{n-p}^2$ . Le test du rapport de vraisemblance rejette donc (H.0) pour des valeurs de la statistique

$$\mathcal{J} = \left\{ \frac{n-p}{rn\widehat{\sigma}^2} \right\} \widehat{\boldsymbol{\theta}}' C' (C(A'A)^{-1}C')^{-1} C \widehat{\boldsymbol{\theta}},$$

dépassant un niveau critique. Pour un test de seuil  $\alpha \in (0, 1)$  (le seuil est la probabilité que le test rejette (H.0) alors que cette hypothèse est vraie), on rejette (H.0) lorsque  $\mathcal{J}$  excède la quantile supérieure d'ordre  $\alpha$ , désignée par  $F_{\alpha; r, n-p}$ , de la loi de Fisher  $F_{r, n-p}$ . En effet, d'une manière générale, on a l'identité en loi

$$\mathcal{J} = \left\{ \frac{n-p}{rn\widehat{\sigma}^2} \right\} \widehat{\boldsymbol{\theta}}' C' (C(A'A)^{-1}C')^{-1} C \widehat{\boldsymbol{\theta}} \stackrel{d}{=} F_{r, n-p} \left( \frac{1}{\sigma^2} \boldsymbol{\theta}' C' (C(A'A)^{-1}C')^{-1} C \boldsymbol{\theta} \right).$$

La puissance  $1 - \beta$  du test basé sur  $\mathcal{J}$  s'évalue par la formule

$$1 - \beta = \mathbb{P} \left( F_{r, n-p} \left( \frac{1}{\sigma^2} \boldsymbol{\theta}' C' (C(A'A)^{-1}C')^{-1} C \boldsymbol{\theta} \right) \geq F_{\alpha; r, n-p} \right).$$

**Exemple 4.1.1.** Un exemple simple est fourni par un échantillon  $X_1, \dots, X_n$  de la loi normale  $N(\mu, \sigma^2)$ . Dans ce cas, on peut poser  $p = 1$ ,

$$A = \mathbb{I}_m = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \boldsymbol{\theta} = [\mu], \quad \widehat{\boldsymbol{\theta}} = [\widehat{\mu}], \quad \widehat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

En choisissant dans les formules précédentes  $p = 1$ ,  $r = 1$  et  $C = [1]$ , on teste l'hypothèse (H.0)  $\mu = 0$ , contre l'hypothèse alternative, (H.1) que  $\mu \in \mathbb{R}$  est quelconque, en observant que

$$\mathcal{J} = \left\{ \frac{n-1}{n\widehat{\sigma}^2} \right\} n\widehat{\mu}^2 = \frac{(n-1)\widehat{\mu}^2}{\widehat{\sigma}^2} \stackrel{d}{=} F_{1, n-1} \left( \frac{n\mu^2}{\sigma^2} \right).$$

On rejette donc (H.0) au seuil  $\alpha$  lorsque

$$\mathcal{J} \geq F_{\alpha; 1, n-p}.$$

La puissance  $1 - \beta$  du test correspondant s'évalue par la formule

$$1 - \beta = \mathbb{P} \left( F_{1, n-1} \left( \frac{n\mu^2}{\sigma^2} \right) \geq F_{\alpha; 1, n-1} \right).$$

### 4.1.2 Le modèle linéaire multivarié classique.

#### A. Généralités.

Le modèle multivarié classique se présente comme une extension naturelle du modèle décrit au §4.1.1, à partir de la relation

$$\mathbf{X} = \begin{bmatrix} X'_1 \\ \vdots \\ X'_n \end{bmatrix} = A\boldsymbol{\Theta} + \boldsymbol{\varepsilon} \stackrel{d}{=} N_{n,d}(A\boldsymbol{\Theta}, \mathbb{I}_n \otimes \Sigma), \quad (4.1.4)$$

où

$$A \in \mathcal{M}_{n,p} \quad \text{et} \quad \boldsymbol{\Theta} \in \mathcal{M}_{p,d}.$$

L'hypothèse que  $\mathbf{X} \stackrel{d}{=} N_{n,d}(A\boldsymbol{\Theta}, \mathbb{I}_n \otimes \Sigma)$  implique, dans ce modèle, que  $X_1, \dots, X_n$  sont des vecteurs aléatoires normaux indépendants de  $\mathbb{R}^d$ , de même matrice de variances-covariances  $\Sigma > 0$ . Il est, parfois commode, de noter les vecteurs colonnes de la matrice aléatoire  $(n \times d)$  d'observations  $\mathbf{X}$ , les vecteurs colonnes de la matrice  $(p \times d)$  de paramètres  $\boldsymbol{\Theta}$ , et les vecteurs colonnes de la "matrice d'erreurs"  $(p \times d)$ ,  $\boldsymbol{\varepsilon}$ , respectivement, par

$$\mathbf{X} = [ \mathbf{X}_1 \quad \cdots \quad \mathbf{X}_d ], \quad \boldsymbol{\Theta} = [ \boldsymbol{\theta}_1 \quad \cdots \quad \boldsymbol{\theta}_d ] \quad \text{et} \quad \boldsymbol{\varepsilon} = [ \boldsymbol{\varepsilon}_1 \quad \cdots \quad \boldsymbol{\varepsilon}_d ].$$

Dans ce modèle,  $A$  désigne une matrice  $(n \times p)$  connue, de plein rang de colonnes, de sorte que  $\text{rg}(A) = p \leq n$ . La matrice  $(p \times d)$ , notée  $\boldsymbol{\Theta} \in \mathcal{M}_{p,d}$ , est un paramètre matriciel inconnu. Enfin, la matrice de variances covariances  $(d \times d)$  commune aux vecteurs aléatoires indépendants  $X_1, \dots, X_n \in \mathbb{R}^d$  est notée  $\Sigma > 0$  est supposée définie positive, et inconnue. Sous ces hypothèses, la "matrice d'erreurs"  $\boldsymbol{\varepsilon} \in \mathcal{M}_{n,d}$  suit une loi normale matricielle centrée, de paramètres  $\boldsymbol{\varepsilon} \stackrel{d}{=} N_{n,d}(\mathbb{O}_{n,d}, \mathbb{I}_n \otimes \Sigma)$ . Ceci signifie que cette matrice est de la forme

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}'_1 \\ \vdots \\ \boldsymbol{\varepsilon}'_n \end{bmatrix},$$

où  $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n$  sont des vecteurs aléatoires indépendants de même loi  $N_d(\mathbb{O}_d, \Sigma)$ . On retrouve le modèle linéaire univarié du §4.1.1 pour  $d = 1$  et  $\Sigma = [ \sigma^2 ]$ .

#### B. Estimation par le maximum de vraisemblance.

**Théorème 4.1.1.** *Sous l'hypothèse que  $\mathbf{X} \stackrel{d}{=} N_{n,d}(A\boldsymbol{\Theta}, \mathbb{I}_n \otimes \Sigma)$ , avec  $\text{rg} A = p$ ,  $\Sigma > 0$ , et  $n \geq d + p$ , les estimateurs du maximum de vraisemblance,  $\widehat{\boldsymbol{\Theta}}$  et  $\widehat{\Sigma}$ , de  $\boldsymbol{\Theta}$  et  $\Sigma$  sont donnés par*

$$\widehat{\boldsymbol{\Theta}} = (A'A)^{-1}A'\mathbf{X}, \quad (4.1.5)$$

et

$$\widehat{\Sigma} = \frac{1}{n} (\mathbf{X} - A\widehat{\boldsymbol{\Theta}})'(\mathbf{X} - A\widehat{\boldsymbol{\Theta}}) = \frac{1}{n} \mathbf{X}'(\mathbb{I}_n - A(A'A)^{-1}A')\mathbf{X}. \quad (4.1.6)$$

**Preuve.** Tout d'abord, on peut appliquer (3.4.14), pour écrire que la densité de  $\mathbf{X} \stackrel{d}{=} N_{n,d}(A\boldsymbol{\Theta}, \mathbb{I}_n \otimes \Sigma)$  est

donnée par

$$\begin{aligned}
f(\mathbf{x}) &= (2\pi)^{-dn/2}(\det \Sigma)^{-n/2} \text{etr} \left( -\frac{1}{2}(\mathbf{x} - A\boldsymbol{\Theta})\Sigma^{-1}(\mathbf{x} - A\boldsymbol{\Theta})' \right) \\
&= (2\pi)^{-dn/2}(\det \Sigma)^{-n/2} \text{etr} \left( -\frac{1}{2}(\mathbf{x} - A\widehat{\boldsymbol{\Theta}})\Sigma^{-1}(\mathbf{x} - A\widehat{\boldsymbol{\Theta}})' \right. \\
&\quad \left. -\frac{1}{2}(A\boldsymbol{\Theta} - A\widehat{\boldsymbol{\Theta}})\Sigma^{-1}(A\boldsymbol{\Theta} - A\widehat{\boldsymbol{\Theta}})' \right. \\
&\quad \left. +\frac{1}{2}(A\boldsymbol{\Theta} - A\widehat{\boldsymbol{\Theta}})\Sigma^{-1}(\mathbf{x} - A\widehat{\boldsymbol{\Theta}})' + \frac{1}{2}(\mathbf{x} - A\widehat{\boldsymbol{\Theta}})\Sigma^{-1}(A\boldsymbol{\Theta} - A\widehat{\boldsymbol{\Theta}})' \right) \\
&= (2\pi)^{-dn/2}(\det \Sigma)^{-n/2} \text{etr} \left( -\frac{1}{2}\Sigma^{-1}(\mathbf{x} - A\widehat{\boldsymbol{\Theta}})'(\mathbf{x} - A\widehat{\boldsymbol{\Theta}}) \right. \\
&\quad \left. -\frac{1}{2}\Sigma^{-1}(\boldsymbol{\Theta} - \widehat{\boldsymbol{\Theta}})'A'A(\boldsymbol{\Theta} - \widehat{\boldsymbol{\Theta}}) \right. \\
&\quad \left. +\frac{1}{2}\Sigma^{-1}\{A'(\mathbf{x} - A\widehat{\boldsymbol{\Theta}})\}'(\boldsymbol{\Theta} - \widehat{\boldsymbol{\Theta}}) + \frac{1}{2}\Sigma^{-1}(\boldsymbol{\Theta} - \widehat{\boldsymbol{\Theta}})\{A'(\mathbf{x} - A\widehat{\boldsymbol{\Theta}})\} \right). \tag{4.1.7}
\end{aligned}$$

Observons, maintenant, par (4.1.5), que

$$A'(\mathbf{X} - A\widehat{\boldsymbol{\Theta}}) = A'\mathbf{X} - A'A(A'A)^{-1}A'\mathbf{X} = \mathbb{O}_{p,d}.$$

En utilisant cette relation, après avoir posé  $\mathbf{x} = \mathbf{X}$  dans la formule (4.1.7) ci-dessus, et en faisant usage de la définition (4.1.6) de  $\widehat{\Sigma} = n^{-1}(\mathbf{X} - A\widehat{\boldsymbol{\Theta}})'(\mathbf{X} - A\widehat{\boldsymbol{\Theta}})$ , on en déduit que la vraisemblance des observations est donnée par

$$\begin{aligned}
L(\boldsymbol{\Theta}, \Sigma) &= f(\mathbf{X}) = (2\pi)^{-dn/2}(\det \Sigma)^{-n/2} \text{etr} \left( -\frac{n}{2}\Sigma^{-1}\widehat{\Sigma} - \frac{1}{2}\Sigma^{-1}(\boldsymbol{\Theta} - \widehat{\boldsymbol{\Theta}})'A'A(\boldsymbol{\Theta} - \widehat{\boldsymbol{\Theta}}) \right) \\
&= (2\pi)^{-dn/2}(\det \Sigma)^{-n/2} \text{etr} \left( -\frac{1}{2}\Sigma^{-1} \left\{ n\widehat{\Sigma} + (\boldsymbol{\Theta} - \widehat{\boldsymbol{\Theta}})'A'A(\boldsymbol{\Theta} - \widehat{\boldsymbol{\Theta}}) \right\} \right).
\end{aligned}$$

Observons que  $A'A$  est une matrice ( $d \times d$ ) inversible. Par conséquent, indépendamment de  $\Sigma > 0$ , on a

$$\sup_{\boldsymbol{\Theta} \in \mathcal{M}_{p,d}} L(\boldsymbol{\Theta}, \Sigma) = L(\widehat{\boldsymbol{\Theta}}, \Sigma) = (2\pi)^{-dn/2}(\det \Sigma)^{-n/2} \text{etr} \left( -\frac{n}{2}\Sigma^{-1}\widehat{\Sigma} \right).$$

Une fois atteint cette étape, on utilise les mêmes arguments que dans la démonstration du théorème 2.1.2, pour montrer que

$$\sup_{\Sigma > 0} L(\widehat{\boldsymbol{\Theta}}, \Sigma) = L(\widehat{\boldsymbol{\Theta}}, \widehat{\Sigma}) = (2\pi)^{-dn/2}(\det \widehat{\Sigma})^{-n/2} e^{-dn/2}.$$

Nous rappelons ici cette démonstration. En utilisant les propriétés de la trace et du déterminant, on écrit

$$L(\widehat{\boldsymbol{\Theta}}, \Sigma) = (2\pi)^{-dn/2}(\det \widehat{\Sigma})^{-n/2}(\det \{\Sigma^{-1/2}\widehat{\Sigma}\Sigma^{-1/2}\})^{n/2} \text{etr} \left( -\frac{n}{2}\Sigma^{-1/2}\widehat{\Sigma}\Sigma^{-1/2} \right).$$

On note alors  $\lambda_1, \dots, \lambda_d$  les valeurs propres (positives) de la matrice définie positive  $\Sigma^{-1/2}\widehat{\Sigma}\Sigma^{-1/2}$ . Ceci permet d'écrire la relation

$$L(\widehat{\boldsymbol{\Theta}}, \Sigma) = (2\pi)^{-dn/2}(\det \widehat{\Sigma})^{-n/2} \exp \left( -\frac{n}{2} \sum_{j=1}^d \{\lambda_j - \log \lambda_j\} \right)^{n/2}.$$

Or, le maximum de la fonction  $\psi(\lambda) = \lambda - \log \lambda$  de  $\lambda > 0$  est égal à 1, et est atteint pour  $\lambda = 1$ . Par conséquent, le supremum de  $L(\widehat{\boldsymbol{\Theta}}, \Sigma)$  est atteint lorsque  $\lambda_1 = \dots = \lambda_d = 1$ , ce qui impose que  $\Sigma^{-1/2}\widehat{\Sigma}\Sigma^{-1/2} = \mathbb{I}_d$ , et donc, pour  $\Sigma = \widehat{\Sigma}$ . On en déduit que

$$\sup_{\Sigma > 0} L(\widehat{\boldsymbol{\Theta}}, \Sigma) = L(\widehat{\boldsymbol{\Theta}}, \widehat{\Sigma}) = (2\pi)^{-dn/2}(\det \widehat{\Sigma})^{-n/2} e^{-nd/2}. \tag{4.1.8}$$

On obtient ainsi que  $\widehat{\Theta}$  et  $\widehat{\Sigma}$  sont les estimateurs du maximum de vraisemblance de  $\Theta$  et de  $\Sigma$ . Il reste à vérifier la deuxième égalité de (4.1.6). Pour cela, on fait usage de (4.1.5), pour écrire

$$(\mathbf{X} - A\widehat{\Theta})'(\mathbf{X} - A\widehat{\Theta}) = \mathbf{X}'(\mathbb{I}_n - A(A'A)^{-1}A')(\mathbb{I}_n - A(A'A)^{-1}A')\mathbf{X} = \mathbf{X}'(\mathbb{I}_n - A(A'A)^{-1}A')\mathbf{X}.$$

Ceci complète la démonstration du théorème.  $\square$

**Remarque 4.1.2.** Le modèle multivarié classique se ramène, pour le calcul de l'estimateur

$$\widehat{\Theta} = \begin{bmatrix} \widehat{\theta}_1 & \dots & \widehat{\theta}_d \end{bmatrix} \quad \text{en fonction de} \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \dots & \mathbf{X}_d \end{bmatrix},$$

à une répétition de  $d$  fois le calcul du modèle univarié, dans le cas d'une matrice de variances-covariances identité. En effet, on a, pour  $j = 1, \dots, d$ ,

$$\widehat{\theta}_j = (A'A)^{-1}A'\mathbf{X}_j. \quad (4.1.9)$$

On notera, au passage, que, dans ces formules, pour  $j = 1, \dots, d$ ,

$$\widehat{\theta}_j \in \mathbb{R}^p \quad \text{et} \quad \mathbf{X}_j \in \mathbb{R}^n.$$

On notera, au passage, que

$$\text{Var}(\text{Vec } \mathbf{X}) = \text{Var} \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_d \end{bmatrix} = \Sigma \otimes \mathbb{I}_n \quad \text{et} \quad \text{Var}(\text{Vec } \mathbf{X}') = \text{Var} \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} = \mathbb{I}_n \otimes \Sigma. \quad (4.1.10)$$

De ce fait, en général,  $\mathbf{X}_j$  et  $\mathbf{X}_\ell$  ne sont pas indépendants pour  $1 \leq j \neq \ell \leq d$ , puisque si  $\Sigma = [\sigma_{j,\ell}]$ , alors  $\text{Cov}(\mathbf{X}_j, \mathbf{X}_\ell) = \sigma_{j,\ell}\mathbb{I}_n$  pour  $1 \leq j, \ell \leq d$ .

**Théorème 4.1.2.** *Sous l'hypothèse que  $\mathbf{X} \stackrel{d}{=} N_{n,d}(A\Theta, \mathbb{I}_n \otimes \Sigma)$ , avec  $\text{rg } A = p$ ,  $\Sigma > 0$ , et  $n \geq d + p$ , les estimateurs du maximum de vraisemblance,  $\widehat{\Theta}$  et  $\widehat{\Sigma}$ , de  $\Theta$  et  $\Sigma$  sont mutuellement indépendants, et de lois respectives, données par*

$$\widehat{\Theta} \stackrel{d}{=} N_{p,d}(\Theta, (A'A)^{-1} \otimes \Sigma) \quad \text{et} \quad n\widehat{\Sigma} \stackrel{d}{=} W_d(n-p, \Sigma). \quad (4.1.11)$$

**Preuve.** Décomposons la matrice  $(p \times n)$   $A'$  sous la forme

$$A' = \begin{bmatrix} a'_1 \\ \vdots \\ a'_p \end{bmatrix},$$

où  $a_1, \dots, a_p$  sont des vecteurs de  $\mathbb{R}^n$ . Notons que l'hypothèse que  $A$  est de rang  $p$  implique que ces vecteurs forment une famille libre. Soit  $H = [h_{p+1} \ \dots \ h_n]$ , une matrice  $(n \times (n-p))$ , composée de colonnes  $h_{p+1}, \dots, h_n$ , orthonormales dans  $\mathbb{R}^n$ , telles que

$$A'H = \begin{bmatrix} a'_1 \\ \vdots \\ a'_p \end{bmatrix} [h_{p+1} \ \dots \ h_n] = \begin{bmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix} = \mathbb{O}_{p,n-p} \quad \Leftrightarrow \quad H'A = \mathbb{O}_{n-p,p}. \quad (4.1.12)$$

Observons ici que  $h_{n-p}, \dots, h_n$  forment une base orthonormale du sous-espace vectoriel  $\mathcal{F}^\perp$  de  $\mathbb{R}^n$ , orthogonal au sous-espace  $\mathcal{F}$ , de dimension  $p$ , engendré par  $a_1, \dots, a_p$ . La matrice  $H$  vérifie l'identité

$$H'H = \begin{bmatrix} h'_{p+1} \\ \vdots \\ h'_n \end{bmatrix} [h_{p+1} \ \dots \ h_n] = \mathbb{I}_{n-p}. \quad (4.1.13)$$

On constate que tout vecteur de  $\mathcal{F}^\perp$  s'écrit sous la forme

$$\sum_{j=n-p}^n h_j z_j = H\mathbf{z} \quad \text{où} \quad \mathbf{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_{n-p} \end{bmatrix} \in \mathbb{R}^{n-p}.$$

Maintenant, comme constaté au (4.1.9), pour tout  $j = 1, \dots, d$ ,  $A\hat{\boldsymbol{\theta}}_j = A(A'A)^{-1}A\mathbf{X}_j$  est la projection orthogonale de  $\mathbf{X}_j$  sur  $\mathcal{F}$ . En combinant cette propriété avec (4.1.13), on obtient l'identité

$$H(H'H)^{-1}H'\mathbf{X}_j = HH'\mathbf{X}_j = (\mathbb{I} - A(A'A)^{-1}A)\mathbf{X}_j.$$

On en déduit que

$$HH' = \mathbb{I} - A(A'A)^{-1}A, \quad (4.1.14)$$

et, par (4.1.6),

$$\mathbf{X}'HH'\mathbf{X} = \mathbf{X}'(\mathbb{I} - A(A'A)^{-1}A)\mathbf{X} = n\hat{\Sigma}. \quad (4.1.15)$$

Posons, maintenant

$$\mathbf{y} = \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{z} \end{bmatrix} = \begin{bmatrix} (A'A)^{-1}A' \\ H' \end{bmatrix} \mathbf{X} = B\mathbf{X}\mathbb{I}_n \quad \text{où} \quad B = \begin{bmatrix} (A'A)^{-1}A' \\ H' \end{bmatrix}.$$

Rappelons la relation (4.1.10). Celle-ci montre que si  $\mathbf{X} \stackrel{d}{=} N_{n,d}(A\boldsymbol{\theta}, \mathbb{I}_n \otimes \Sigma)$ , alors

$$\mathbb{E}(\mathbf{X}) = A\boldsymbol{\theta} \quad \text{et} \quad \text{Var}(\mathbf{X}) = \text{Var}(\text{Vec}(\mathbf{X}')) = \text{Var} \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} = I_n \otimes \Sigma.$$

Maintenant, si  $B$  et  $C$  sont des matrices de constantes, de dimensions appropriées, on constate que

$$\mathbf{y} = B\mathbf{X}C \quad \Leftrightarrow \quad \mathbf{y}' = C'\mathbf{X}'B' \quad \Rightarrow \quad \text{Vec}(\mathbf{y}') = (B \otimes C')\text{Vec}(\mathbf{X}'). \quad (4.1.16)$$

On en déduit, d'une part, en faisant usage de (4.1.12), que

$$\mathbb{E}(\mathbf{y}) = \begin{bmatrix} (A'A)^{-1}A' \\ H \end{bmatrix} A\boldsymbol{\theta} = \begin{bmatrix} (A'A)^{-1}A'A\boldsymbol{\theta} \\ H'A\boldsymbol{\theta} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta} \\ \mathbb{O}_{n-p,d} \end{bmatrix},$$

et, d'autre part, en appliquant la relation (4.1.16), que

$$\text{Vec}(\mathbf{y}') = (B \otimes \mathbb{I}_n)\text{Vec}(\mathbf{X}').$$

Cette dernière relation implique que

$$\begin{aligned} \text{Var}(\text{Vec}(\mathbf{y}')) &= (B \otimes \mathbb{I}_n)\text{Var}(\text{Vec}(\mathbf{X}'))(B' \otimes \mathbb{I}_n) = (B \otimes \mathbb{I}_n)(\mathbb{I}_n \otimes \Sigma)(B' \otimes \mathbb{I}_n) \\ &= BB' \otimes \Sigma. \end{aligned}$$

Il reste à calculer

$$BB' = \begin{bmatrix} (A'A)^{-1}A' \\ H' \end{bmatrix} \begin{bmatrix} A(A'A)^{-1} & H \end{bmatrix} = \begin{bmatrix} (A'A)^{-1} & \mathbb{O}_{p,n-p} \\ \mathbb{O}_{n-p,p} & \mathbb{I}_{n-p} \end{bmatrix}.$$

On en déduit que

$$\mathbf{y} = \begin{bmatrix} \hat{\boldsymbol{\theta}} \\ H'\mathbf{X} \end{bmatrix} \stackrel{d}{=} N_{n,d} \left( \begin{bmatrix} \boldsymbol{\theta} \\ \mathbb{O}_{n-p,d} \end{bmatrix}, \begin{bmatrix} (A'A)^{-1} & \mathbb{O}_{p,n-p} \\ \mathbb{O}_{n-p,p} & \mathbb{I}_{n-p} \end{bmatrix} \otimes \Sigma \right).$$

Ceci implique que  $\hat{\boldsymbol{\theta}} \stackrel{d}{=} N_{p,d}(\boldsymbol{\theta}, (A'A)^{-1} \otimes \Sigma)$  et  $\mathbf{z} = H'\mathbf{X} \stackrel{d}{=} N_{n-p,d}(\mathbb{O}_{n-p,d}, \mathbb{I}_{n-p} \otimes \Sigma)$  sont indépendants. Compte tenu de la relation (4.1.15), on vérifie que

$$\mathbf{z}'\mathbf{z} = n\hat{\Sigma} \stackrel{d}{=} W_d(n-p, \Sigma).$$

La démonstration du théorème 4.1.2 est ainsi achevée.  $\square$



### 4.1.3 Analyse de variance multivariée.

Nous nous plaçons sous les hypothèses du §4.1.2. Nous supposons donc que

$$\mathbf{X} = \begin{bmatrix} X'_1 \\ \vdots \\ X'_n \end{bmatrix} = A\boldsymbol{\Theta} + \boldsymbol{\varepsilon} \stackrel{d}{=} N_{n,d}(A\boldsymbol{\Theta}, \mathbb{I}_n \otimes \Sigma),$$

où

$$A \in \mathcal{M}_{n,p} \quad \text{et} \quad \boldsymbol{\Theta} \in \mathcal{M}_{p,d}.$$

Nous supposons donnée une matrice  $C \in \mathcal{M}_{r,p}$ , de plein rang de lignes, égal à  $\text{rg } C = r$ . Notre but est de tester l'hypothèse

$$(H.0) \quad C\boldsymbol{\Theta} = \mathbb{O}_{r,d},$$

contre l'alternative générale

$$(H.1) \quad C\boldsymbol{\Theta} \in \mathcal{M}_{r,d}.$$

Pour cela, nous calculons les quantités :

$$\begin{aligned} \widehat{\boldsymbol{\Theta}} &= (A'A)^{-1}A'\mathbf{X}, \\ \mathcal{A} &= \widehat{\boldsymbol{\Theta}}'C' \left\{ C(A'A)^{-1}C' \right\}^{-1} C\widehat{\boldsymbol{\Theta}}, \\ \mathcal{B} &= n\widehat{\Sigma} = \mathbf{X}'\mathbf{X} - \widehat{\boldsymbol{\Theta}}'A'\mathbf{X}. \end{aligned}$$

**Théorème 4.1.3.** *Sous l'hypothèse (H.0), les matrices  $\mathcal{A}$  et  $\mathcal{B}$  sont indépendantes, de lois respectives  $\mathcal{A} \stackrel{d}{=} W_d(r, \Sigma, \Omega)$  et  $\mathcal{B} \stackrel{d}{=} W_d(n-p, \Sigma)$ , où*

$$\Omega = \Sigma^{-1}\boldsymbol{\Theta}'C' \left\{ C(A'A)^{-1}C' \right\}^{-1} C\boldsymbol{\Theta}.$$

Le test du rapport de vraisemblance de (H.0) contre (H.1) rejette (H.0) pour les valeurs faibles de la statistique

$$\Lambda^{2/(n-p+r)} = \Lambda(n-p+r, d, r) = \frac{\det \mathcal{B}}{\det(\mathcal{A} + \mathcal{B})}.$$

On a la convergence en loi suivante, lorsque  $n \rightarrow \infty$ ,  $p$ ,  $r$  et  $d$  étant fixés,

$$-\left\{ n-p+r - \frac{1}{2}(d+r+1) \right\} \log \Lambda(n-p+r, d, r) \xrightarrow{d} \chi_{dr}^2(\text{tr } \Omega).$$

On rejette donc (H.0) au seuil (asymptotique lorsque  $n \rightarrow \infty$ )  $\alpha \in (0, 1)$  lorsque

$$-\left\{ n-p+r - \frac{1}{2}(d+r+1) \right\} \log \Lambda(n-p+r, d, r) \geq \chi_{dr; \alpha}^2.$$

La puissance du test correspondant est (asymptotiquement lorsque  $n \rightarrow \infty$ ) égale à

$$\mathbb{P}\left(\chi_{dr}^2(\text{tr } \Omega) \geq \chi_{dr; \alpha}^2\right).$$

## 4.2 Test d'homogénéité des moyennes (en cours de rédaction).

On suppose disposer de  $k$  échantillons indépendants de vecteurs aléatoires en dimension  $d$ . Ces échantillons sont résumés par les  $k$  matrices

$$\mathbf{X}_1 \in \mathcal{M}_{n_1, d}, \dots, \mathbf{X}_k \in \mathcal{M}_{n_k, d},$$

qui sont telles que, pour  $i = 1, \dots, k$ ,

$$\mathbf{X}_i = \begin{bmatrix} X'_{1,i} \\ \vdots \\ X'_{n_i,i} \end{bmatrix} \stackrel{d}{=} N_{n_i, d}(\mathbb{I} \otimes \mu_i, I_{n_i} \otimes \Sigma_i).$$

Ceci est équivalent à supposer que, pour  $i = 1, \dots, k$ , les vecteurs

$$X_{1,i}, \dots, X_{n_i,i} \stackrel{d}{=} N_d(\mu_i, \Sigma_i),$$

sont indépendants et de même loi.

On pose

$$n = \sum_{i=1}^k n_i,$$

et, pour  $i = 1, \dots, k$ ,

$$\hat{\mu}_i = \bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{j,i} \quad \text{et} \quad \hat{\Sigma}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} (X_{j,i} - \bar{X}_i)(X_{j,i} - \bar{X}_i)'$$

La vraisemblance des observations est donnée par

$$L(\mu_1, \dots, \mu_k; \Sigma_1, \dots, \Sigma_k) = (2\pi)^{-nd/2} \text{etr} \left( -\frac{1}{2} \sum_{i=1}^k n_i \left\{ \log \det \Sigma_i + \text{tr} \left( \Sigma_i^{-1} \{ \hat{\Sigma}_i + (\bar{X}_i - \mu_i)(\bar{X}_i - \mu_i)' \} \right) \right\} \right).$$

Cette vraisemblance est maximisée pour

$$\mu_i = \bar{X}_i = \hat{\mu}_i \quad \text{pour} \quad i = 1, \dots, k,$$

indépendamment du choix de  $\Sigma_1, \dots, \Sigma_k$ . On obtient

$$L(\hat{\mu}_1, \dots, \hat{\mu}_k; \Sigma_1, \dots, \Sigma_k) = (2\pi)^{-nd/2} \text{etr} \left( -\frac{1}{2} \sum_{i=1}^k n_i \left\{ \log \det \Sigma_i + \text{tr} \left( \Sigma_i^{-1} \hat{\Sigma}_i \right) \right\} \right).$$

La maximisation de cette dernière expression relativement à  $\Sigma_1, \dots, \Sigma_k$  montre que les estimateurs du maximum de vraisemblance de ces quantités sont donnés par  $\hat{\Sigma}_1, \dots, \hat{\Sigma}_k$ , de sorte que le maximum de la vraisemblance soit égal à

$$L(\hat{\mu}_1, \dots, \hat{\mu}_k; \hat{\Sigma}_1, \dots, \hat{\Sigma}_k) = (2\pi)^{-nd/2} e^{-nd/2} \prod_{i=1}^k (\det \hat{\Sigma}_i)^{n_i/2}.$$

Nous allons maintenant considérer des tests du rapport de vraisemblance pour trois types d'hypothèses, notées (H.a), (H.b) et (H.c), contre l'alternative.

- (H.a) :  $\mu_1 = \dots = \mu_k$  (homogénéité des moyennes);
- (H.b) :  $\Sigma_1 = \dots = \Sigma_k$  (homogénéité des variances-covariances)
- (H.c) :  $\mu_1 = \dots = \mu_k$  et  $\Sigma_1 = \dots = \Sigma_k$  (homogénéité totale).

## 4.3 Test d'homogénéité des matrices de variances-covariances (en cours de rédaction).



## Chapitre 5

# Distributions Elliptiques et Sphériques (en cours de rédaction).

Dans ce paragraphe, on supposera systématiquement que les vecteurs aléatoires considérés possèdent des densités relativement à la mesure de Lebesgue.

**Définition 5.0.1.** Soient  $\mu \in \mathbb{R}^m$  et  $\Sigma > 0$  une matrice  $(m \times m)$  définie positive.

1) Un vecteur aléatoire  $X \in \mathbb{R}^m$  est dit suivre une distribution elliptique de paramètres  $\mu$  et  $\Sigma$ . de paramètres  $\mu \in \mathbb{R}^m$  et  $\Sigma$ .

2) La distribution de  $X$  est dite sphérique si elle est elliptique de paramètres  $\mathbb{O}$  et  $\mathbb{I}$ .

**Théorème 5.0.1.** Si  $X \in \mathbb{R}^m$  suit une distribution sphérique telle que  $\mathbb{P}(X = \mathbb{O}) = 0$ , alors, si on pose

$$r(X) = \|X\| = (X'X)^{1/2} \quad \text{et} \quad T(X) = \|X\|^{-1}X, \quad (5.0.1)$$

$T(X)$  est uniformément distribué sur la sphère unité  $\mathcal{S}_m = \{x \in \mathbb{R}^m : \|x\| = 1\}$ . De plus,  $r(X)$  et  $T(X)$  sont indépendants.

**Preuve.** Soit  $C \in \mathbb{R}$  une partie borélienne de  $\mathbb{R}$  telle que  $\mathbb{P}(r \in C) > 0$ , et considérons la mesure de probabilité définie sur  $\mathcal{S}_m$  par

$$\mu_C(B) = \mathbb{P}(T(X) \in B | r \in C),$$

où  $B$  varie dans l'ensemble des parties boréliennes de  $\mathcal{S}_m$ . Soit  $H$  une matrice orthogonale quelconque. Il est clair que

$$T(HX) = \|HX\|^{-1}HX = \|X\|^{-1}HX = HT(X), \quad \text{et} \quad r(HX) = \|HX\| = \|X\| = r(X)$$

ce qui, compte tenu de la Définition 4.5, montre que

$$\mu_C(B) = \mathbb{P}(T(X) \in B | r \in C) = \mathbb{P}(T(HX) \in B | r(HX) \in C) = \mathbb{P}(T(X) \in H^{-1}B | r(X) \in C) = \mu_C(H^{-1}B).$$

Cette propriété étant satisfaite indépendamment de  $H \in \mathcal{O}_m$ , on constate ainsi que  $\mu_C$  est une mesure invariante par translation sur l'espace localement compact  $\mathcal{S}_m$ , sur lequel opère le groupe orthogonal  $\mathcal{O}_m$ . La mesure  $\mu_C$  coïncide donc avec la loi uniforme sur  $\mathcal{S}_m$ , indépendamment de  $C$ . On en déduit le résultat annoncé.  $\square$

Pour introduire le résultat suivant, nous rappelons les définitions des lois de Student et de Fisher.

**Définition 5.0.2.** Pour  $m \geq 2$ , si  $X_1, \dots, X_m$  sont des variables  $N(0, 1)$  indépendantes, la variable

$$Y = X_1 / \left( \frac{1}{m-1} \sum_{i=2}^m X_i^2 \right)^{1/2} \quad (5.0.2)$$

suit une loi de Student centrée  $t_{m-1}$  à  $m-1$  degrés de liberté, et la variable

$$Z = \left( \frac{1}{k} \sum_{i=1}^k X_i^2 \right) / \left( \frac{1}{m-k} \sum_{i=k+1}^m X_i^2 \right) \quad (5.0.3)$$

suit une loi de Fisher  $F_{k, m-k}$  à  $k$  et  $m-k$  degrés de liberté.

**Théorème 5.0.2.** Soit  $\mathbb{X} \in \mathbb{R}^m$  un vecteur aléatoire, suivant une loi sphérique dans  $\mathbb{R}^m$ , et tel que  $\mathbb{P}(\mathbb{X} = \mathbb{O}) = 0$ , et soit  $\alpha \in \mathbb{R}^m$  un vecteur constant tel que  $\|\alpha\| = (\alpha' \alpha)^{1/2} = 1$ . Alors, si  $U = \alpha' \mathbb{X} / \|\mathbb{X}\|$ , la variable aléatoire

$$Y = \frac{(m-1)^{1/2} U}{\sqrt{1-U^2}},$$

suit une loi de Student  $t_{m-1}$  à  $m-1$  degrés de liberté. De plus, si  $B$  est une matrice idempotente de rang  $k \geq 1$ , la variable aléatoire

$$Z = (\mathbb{X}' B \mathbb{X}) / \|\mathbb{X}\|^2,$$

suit une loi de béta de paramètres  $\beta(\frac{1}{2}k, \frac{1}{2}(m-k))$ .

**Preuve.** Il suffit de vérifier que le résultat est vrai lorsque les coordonnées de  $\mathbb{X}$  sont  $N(0,1)$  indépendantes, puis de constater, à l'aide du Théorème 5.8, que la conclusion subsiste lorsque  $\mathbb{X}$  suit une loi sphérique.  $\square$

### 5.0.1 Estimation des Coefficients de Corrélation.

Soit  $X = [X(1) \ \cdots \ X(m)]' \in \mathbb{R}^m$  un vecteur aléatoire d'espérance  $\mu = \mathbb{E}(X) = [\mu_1 \ \cdots \ \mu_m]'$  et de matrice de covariance  $\Sigma = \mathbb{E}((X-\mu)(X-\mu)') = [\sigma_{i,j}]$ . On pose ainsi, pour  $1 \leq i, j \leq m$ ,  $\sigma_{i,j} = \text{Cov}(X(i), X(j))$  et  $\sigma_i^2 = \sigma_{i,i} = \text{Var}(X(i))$ . Lorsque  $\sigma_i \neq 0$  et  $\sigma_j \neq 0$ , le coefficient de corrélation entre les coordonnées  $X(i)$  et  $X(j)$  de  $X$  est défini par

$$\rho_{i,j} = \rho_{j,i} = \frac{\sigma_{i,j}}{\sigma_i \sigma_j} = \frac{\text{Cov}(X(i), X(j))}{\sqrt{\text{Var}(X(i)) \text{Var}(X(j))}}. \quad (5.0.4)$$

Dans toute la suite de ce paragraphe, nous supposons que les coordonnées de  $X$  sont *non dégénérées*, c'est à dire non constantes presque sûrement. Cette condition (sous réserve de supposer l'existence de moments d'ordre 2 pour  $X(i)$ ) est équivalente à  $\sigma_i \neq 0$  pour  $i = 1, \dots, m$ . L'existence de  $\rho_{i,j}$  est alors assurée pour tout  $1 \leq i, j \leq m$ . Le cas où  $i = j$  donne trivialement  $\rho_{i,i} = 1$ .

On note

$$\Sigma_\rho = (\rho_{i,j}) = \begin{bmatrix} 1 & \cdots & \rho_{1,m} \\ \vdots & \ddots & \vdots \\ \rho_{m,1} & \cdots & 1 \end{bmatrix},$$

la matrice de corrélation de  $X$ .

**Lemme 5.0.1.** La matrice  $\Sigma_\rho$  est symétrique et positive. De plus,  $\Sigma_\rho > 0$  si et seulement si  $\Sigma > 0$ . Pour tout  $1 \leq i, j \leq m$ , on a l'inégalité  $-1 \leq \rho_{i,j} \leq 1$ , avec  $\rho_{i,j} = \pm 1$  si et seulement si  $X(i)$  et  $X(j)$  sont liés par une relation linéaire affine presque sûrement.

**Preuve.**  $\Sigma_\rho$  est la matrice de variances-covariances de  $\text{diag}(\sigma_1, \dots, \sigma_m)^{-1} X = [X(1)/\sigma_1 \ \cdots \ X(m)/\sigma_m]'$ , d'où les deux premières propriétés. On déduit de celles-ci que, pour tout  $1 \leq k \leq m$ ,  $\det[\rho_{i,j} : 1 \leq i, j \leq k] \geq 0$ . Pour  $k = 2$ , ceci implique que  $1 - \rho_{1,2}^2 \geq 0$  avec égalité si et seulement si  $X(1) - \mu_1$  et  $X(2) - \mu_2$  sont liés par une relation linéaire, et le résultat analogue pour  $\rho_{i,j}$  lorsque  $1 \leq i \neq j \leq m$  s'en déduit par permutation des coordonnées. Enfin, pour  $1 \leq i = j \leq m$ ,  $\rho_{i,j} = \rho_{i,i} = 1$ , ce qui achève la démonstration.  $\square$

Etant donné un échantillon  $X_1, \dots, X_N$  de taille  $N$  de  $X$ , on pose  $\bar{X} = N^{-1} \sum_{i=1}^N X_i = [\bar{X}(1) \ \cdots \ \bar{X}(m)]$ ,  $n = N - 1$  et

$$A = nS = N\hat{\Sigma} = \sum_{i=1}^N (X_i - \bar{X})(X_i - \bar{X})'. \quad (5.0.5)$$

On remarquera que  $\hat{\mu} = \bar{X}$  et  $\hat{\Sigma}$  sont respectivement l'espérance et la matrice de covariance de la *distribution empirique* qui affecte une probabilité de  $1/N$  à chacune des observations  $X_1, \dots, X_N$  de l'échantillon. Ceci implique que  $\hat{\Sigma} \geq 0$ , et, par conséquent,  $S = (n/(n-1))\hat{\Sigma} \geq 0$ . La matrice  $\hat{\Sigma} = [\hat{\sigma}_{i,j}]$  est appelée *matrice de variances-covariances empirique de l'échantillon*. On pose ainsi, pour  $1 \leq i, j \leq m$ ,

$$\hat{\sigma}_{i,j} = \frac{1}{N} \sum_{k=1}^N (X_k(i) - \bar{X}(i))(X_k(j) - \bar{X}(j)), \quad \text{et} \quad \hat{\sigma}_{i,i} = \hat{\sigma}_i^2 = \frac{1}{N} \sum_{k=1}^N (X_k(i) - \bar{X}(i))^2.$$

Sous réserve que les variances empiriques  $\hat{\sigma}_i^2$  soient non-nulles pour  $1 \leq i \leq m$ , on peut définir la matrice, notée  $R$  des coefficients de corrélation empiriques, définis respectivement par

$$R = \hat{\Sigma}_\rho = \begin{bmatrix} 1 & \cdots & r_{1,m} \\ \vdots & \ddots & \vdots \\ r_{m,1} & \cdots & 1 \end{bmatrix},$$

et, pour  $1 \leq i, j \leq m$ ,

$$r_{i,j} = \hat{\rho}_{i,j} = \frac{\sum_{k=1}^N (X_k(i) - \bar{X}(i))(X_k(j) - \bar{X}(j))}{\left( \sum_{k=1}^N (X_k(i) - \bar{X}(i))^2 \sum_{k=1}^N (X_k(j) - \bar{X}(j))^2 \right)^{1/2}}.$$