

**Dépendance et mesure de liaison entre les composantes d'un
vecteur aléatoire basées sur la distance entre les matrices
d'information de Fisher**

par

Fouodjio René

mémoire présenté au Département de mathématiques et d'informatique
en vue de l'obtention du grade de maître ès sciences (M.Sc.)

FACULTÉ DES SCIENCES
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, février 2004

SOMMAIRE

L'association positive entre les composantes d'un vecteur aléatoire exprime que les variables tendent à prendre toutes des grandes valeurs ou toutes des petites valeurs en même temps tandis qu'une association négative révèle qu'un sous vecteur tend à prendre des grandes valeurs alors que le sous-vecteur complémentaire tend à prendre des petites valeurs et inversement. En cas d'indépendance entre ces composantes, le fait de savoir que certaines des variables prennent des valeurs prédéterminées ne modifie pas la probabilité de variation des autres; on n'a donc pas plus d'information sur la variation de ces dernières. Il importe donc de déterminer la force ou l'intensité d'une liaison entre ces variables lorsque celle-ci est avérée.

Ce mémoire porte sur la mesure de dépendance de K. Zografos. La première partie est consacrée à l'étude des formes de dépendance entre les composantes d'un vecteur aléatoire, la deuxième partie est une approche visant à détecter la dépendance ou l'indépendance à l'aide de la matrice d'information de Fisher, la troisième partie est une présentation des résultats obtenus par K. Zografos et les résultats des simulations constituent la quatrième partie.

REMERCIEMENTS

Je tiens à remercier vivement mon directeur, le professeur Bernard Colin et mon codirecteur, le professeur Ernest Monga pour leurs précieux conseils, leur inconditionnelle disponibilité et leur soutien financier.

J'aimerais remercier Messieurs Alain Boulanger et Luc-Désiré Adjengue d'avoir accepté de faire partie du jury. Ma gratitude va particulièrement au professeur Alain Boulanger dont les suggestions ont donné la clarté à ce mémoire.

J'exprime aussi ma gratitude à l'Université de Sherbrooke pour m'avoir octroyé une bourse d'étude. Mes vifs remerciements vont également à l'endroit de Diane et Éric pour l'aide qu'ils m'ont apportée en programmation.

Je remercie aussi mes parents et mes amis pour leur soutien tout au long de la maîtrise. Je remercie grandement tous ceux qui m'ont aidé de près ou de loin à réaliser ce mémoire.

TABLE DES MATIÈRES

SOMMAIRE	iii
REMERCIEMENTS	iv
TABLE DES MATIÈRES	v
LISTE DES TABLEAUX	ix
INTRODUCTION	1
CHAPITRE 1 — Structure de dépendance d'un vecteur aléatoire.	3
1.1 Notations et cadre probabiliste.	4
1.2 Formes de liaison entre les composantes d'un vecteur aléatoire.	5
1.2.1 Covariance et association entre les composantes d'un vecteur aléatoire.	5
1.2.2 Dépendance positive (négative) par orthant.	8
1.2.3 Quelques propriétés des dépendances positives (négatives).	12

1.2.4	Positivité et négativité totales multidimensionnelles d'ordre deux.	16
1.2.5	Dépendance séquentielle.	19
CHAPITRE 2 — Dépendance et matrice d'information de Fisher.		23
2.1	Notations.	24
2.2	Paramètre de tendance centrale.	24
2.3	Conditions de régularité.	25
2.4	Quantité d'information de Fisher.	26
2.4.1	Propriétés de la quantité d'information de Fisher.	26
2.4.2	Exemples.	27
2.5	Matrice d'information de Fisher.	29
2.5.1	Quelques rappels sur les matrices semi-définies et définies positives.	29
2.5.2	Quelques propriétés des matrices (S)DP.	29
2.5.3	Matrice d'information.	30
2.5.4	Matrice d'information dans le cas de l'indépendance.	32
2.5.5	Propriétés de la matrice d'information.	33
2.5.6	Lien avec la matrice de variance-covariance.	34
2.6	Caractérisation de l'indépendance à l'aide de la matrice d'information. . .	35
CHAPITRE 3 — Mesure de dépendance.		38
3.1	Axiomes de Rényi.	39

3.2	Mesure de dépendance de Zografos.	40
3.2.1	Borne inférieure de la mesure de Zografos.	41
3.2.2	Majorants de la mesure de Zografos.	42
3.2.3	Propriétés de la mesure de Zografos.	43
3.2.4	La mesure de Zografos, la ϕ -divergence et l'entropie relative.	46
3.2.5	Étude de quelques exemples.	51
3.2.6	Exemple 1 : Loi binormale.	51
3.2.7	Exemple 2 : Loi multinormale avec matrice d'équicorrélation.	51
3.2.8	Exemple 3 : Modèle multinormal interclasse.	54
3.2.9	Exemple 4 : Cas de la distribution de Dirichlet inverse multidimensionnelle.	57
3.2.10	Version empirique de la mesure de Zografos.	60
3.3	Ordre de dépendance basée sur la matrice d'information de Fisher.	62
3.3.1	Exemple d'ordre de dépendance multidimensionnelle : cas de la loi multinormale.	64
3.4	Autres mesures de dépendance multidimensionnelle.	65
3.4.1	Le tau de Kendall.	65
3.4.2	Le rhô de Spearman.	66
3.4.3	L'entropie relative.	66
3.4.4	La mesure de Harry Joe.	68

CHAPITRE 4 — Simulations.	69
CONCLUSION	73
<u>Annexe A</u>	75
BIBLIOGRAPHIE	88

LISTE DES TABLEAUX

3.1	Axiomes de Rényi pour la mesure de Zografos.	46
4.1	Résultats des simulations basés sur l'estimation de $\delta_{h,f}(X, Y), \rho =$ 0.94	70
4.2	Résultats des simulations basés sur l'estimation de $\delta_{h,f}(X, Y), \rho =$ 0.99	70
4.3	Résultats des simulations basés sur l'estimation de $\delta_{h,f}(X, Y), \rho = 0$	70
4.4	Résultats des simulations basés sur les matrices empiriques, $\rho =$ 0.94	72
4.5	Résultats des simulations basés sur les matrices empiriques, $\rho =$ 0.99	72
4.6	Résultats des simulations basés sur les matrices empiriques, $\rho =$ 0.50	72
4.7	Programme SAS pour générer les couples d'observations d'une loi binormale.	75
4.8	Programme SAS d'extraction des matrices de variance-covariance.	76

INTRODUCTION

La mesure la plus classique utilisée pour quantifier la force d'une liaison entre deux variables aléatoires X_1 et X_2 est sans doute le coefficient de corrélation de Pearson. Mais cette mesure de dépendance fait intervenir les lois marginales des variables, ce qui en complique l'interprétation et l'utilisation. Un moyen efficace de contourner ce problème consiste à utiliser les mesures ou coefficients de corrélation basés sur les rangs des observations tels que le rhô (ρ) de Spearman ou le tau (τ) de Kendall. En effet, les coefficients basés sur les rangs présentent l'avantage de ne pas être influencés par les lois marginales.

En présence de trois variables et plus, les calculs de ceux-ci deviennent très vite fastidieux les rendant presque inutilisables.

Nous présentons dans ce mémoire, la mesure de dépendance de Zografos [1] qui, à la différence du coefficient de corrélation partiel de Pearson, caractérise l'intensité de la dépendance entre toutes les composantes d'un vecteur aléatoire. Nous en préciserons notamment son comportement asymptotique.

Les formes de dépendance sous-jacentes aux calculs de l'intensité de la liaison qui unit les composantes d'un vecteur aléatoire constituent des détails pertinents et font l'objet du chapitre 1. Dans ce dernier, nous définissons les notions d'association positive et négative au sens de Esary, Proschan et Walkup [2] dont nous dériverons les dépendances positive et négative par orthant. Par la suite, les notions de positivité et de négativité

totales multidimensionnelles d'ordre 2 sont présentées telles qu'elles apparaissent dans Glaz [3]. Nous terminons ce chapitre par l'étude des dépendances conditionnelles. Au chapitre 2, nous définissons la notion de quantité d'information au sens de Fisher sur un paramètre contenue dans un n -échantillon et nous introduisons la matrice d'information d'un vecteur aléatoire. Nous montrons ensuite que la structure de cette matrice peut, sous certaines conditions, révéler la dépendance ou l'indépendance entre les variables du vecteur aléatoire concerné. Le chapitre 3 est une présentation de la mesure de K. Zografos et d'autres mesures de dépendance multidimensionnelle entre les composantes d'un vecteur aléatoire tandis que le chapitre 4 est consacré aux simulations en vue de déterminer la convergence de la mesure empirique de Zografos vers sa version théorique. Enfin, le mémoire soulève la question sur la performance de cette mesure par rapport aux autres mesures de dépendance et en tant que statistique pouvant permettre de mettre en oeuvre les tests d'indépendance.

CHAPITRE 1

Structure de dépendance d'un vecteur aléatoire.

Afin de mieux analyser l'information contenue dans un vecteur aléatoire, nous devons avoir une bonne connaissance de la structure de dépendance liant ses composantes. Le concept d'association est couramment utilisé pour mettre en évidence les différents types de liaisons entre les composantes d'un vecteur aléatoire. Mais ce concept offre une condition de dépendance qui n'est en général pas facile à satisfaire. Dans ce chapitre nous définissons d'abord cette notion d'association comme elle apparaît dans [2] dont nous dériverons les notions de dépendance positive et négative par orthant. Nous introduisons ensuite les notions de positivité et de négativité totales multidimensionnelles d'ordre deux telles que présentées par Glaz [3] ainsi que de dépendance stochastique séquentielle au sens de Harry Joe [5] qui, en plus de décrire la structure de dépendance entre les variables d'un vecteur aléatoire, donnent lieu à des conditions fortes de dépendance qui semblent plus faciles à vérifier et qui impliquent les conditions d'association.

1.1 Notations et cadre probabiliste.

Nous considérerons l'espace probabilisé $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, \nu, \mathcal{P}_\theta; \theta \in \Theta)$ où :

- Θ est un ouvert non vide de \mathbb{R}^n ;
- $\mathcal{B}_{\mathbb{R}^n}$ est la tribu des boréliens de \mathbb{R}^n ;
- ν est une mesure σ -finie telle que la famille \mathcal{P}_θ soit dominée par ν [c'est-à-dire : $\mathcal{P}_\theta \ll \nu$] et $f(x, \theta)d\nu = d\mathcal{P}_\theta$, pour tout $\theta \in \Theta$ et pour tout $X \in \mathbb{R}^n$, $f(x, \theta)$ étant la densité de \mathcal{P}_θ par rapport à la mesure ν . Nous prendrons en particulier la mesure produit de Lebesgue $\nu = \times_{i=1}^n \nu_i$ sur $\mathcal{B}_{\mathbb{R}^n}$ où ν_i est la mesure de Lebesgue sur l'ensemble des boréliens de \mathbb{R} . Dans ce cas, $f(x, \theta)$ désignera la densité de \mathcal{P}_θ par rapport à la mesure produit ν et ses marginales fixes $f_i(x_i, \theta_i)$, les densités respectives par rapport à ν_i , pour $i = 1, 2, \dots, n$. Pour alléger les écritures, nous écrirons $f(x)$ au lieu de $f(x, \theta)$, $P(\cdot)$ pour $\mathcal{P}_\theta(\cdot)$ et poserons alors $dP = f(x)d\nu = f(x)dx_1 \cdots dx_n = f(x)dx$.

Les vecteurs aléatoires considérés sont continus et à valeurs dans \mathbb{R}^n dont les éléments seront notés $x = (x_1, x_2, \dots, x_n)$ et $y = (y_1, y_2, \dots, y_n)$... L'ensemble \mathbb{R}^n est aussi muni de l'ordre partiel naturel \leq défini, pour tout $x, y \in \mathbb{R}^n$, par :

$x \leq y$ si et seulement si $x_i \leq y_i$ pour tout $i = 1, 2, \dots, n$.

Pour tout vecteur $X = (X_1, X_2, \dots, X_n)$ et $\mathcal{I} = \{i_1, i_2, \dots, i_k\} \subset \{1, 2, \dots, n\}$, le vecteur $X_{\mathcal{I}} = (X_{i_1}, X_{i_2}, \dots, X_{i_k})$ sera défini comme un sous-vecteur aléatoire de X . Nous dirons de toute fonction à valeurs réelles g dont le domaine de définition est \mathbb{R}^n , qu'elle est :

- croissante si pour tout $x, y \in \mathbb{R}^n$, $x \leq y$ entraîne $g(x) \leq g(y)$;

- décroissante si pour tout $x, y \in \mathbb{R}^n$, $x \leq y$ entraîne $g(x) \geq g(y)$. Les formulations analogues sont valables pour g strictement croissante ou strictement décroissante. On pourrait noter que si g est croissante, alors $-g$ est décroissante.
- Deux fonctions f et g définies sur \mathbb{R}^n et à valeurs réelles seront dites identiquement monotones si elles sont toutes croissantes ou toutes décroissantes.

Si U est un borélien de \mathbb{R}^n , \mathcal{I}_U désignera la fonction indicatrice de U définie comme suit : $\mathcal{I}_U(x) = 1$ si $x \in U$ et $\mathcal{I}_U(x) = 0$ sinon. Enfin, le symbole $cov(\cdot, \cdot)$ représentera la covariance entre deux variables aléatoires données et $\mathbb{E}(\cdot)$, l'espérance mathématique d'une variable aléatoire d'intérêt et sera prise par rapport à la mesure de Lebesgue ν , de sorte que nous puissions écrire par exemple : $\mathbb{E}(g(X)) = \int g(x)f(x) dx$ où g est une fonction réelle ν -intégrable et X un vecteur aléatoire de densité $f(x)$.

1.2 Formes de liaison entre les composantes d'un vecteur aléatoire.

1.2.1 Covariance et association entre les composantes d'un vecteur aléatoire.

Une mesure fréquemment utilisée pour quantifier la force de la liaison entre deux variables aléatoires réelles est la covariance définie comme suit :

Définition 1.1 Soient X_1 et X_2 deux variables aléatoires réelles telles que : $\mathbb{E}(X_1^2) < \infty$ et $\mathbb{E}(X_2^2) < \infty$. La covariance entre X_1 et X_2 est la quantité définie par : $cov(X_1, X_2) = \mathbb{E}(X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2)) = \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2)$.

Nous dirons que les variables X_1 et X_2 sont liées, respectivement non liées, si $\text{cov}(X_1, X_2) \neq 0$, respectivement $\text{cov}(X_1, X_2) = 0$. Notons que lorsque les variables X_1 et X_2 ont tendance à croître ou décroître simultanément, alors $\text{cov}(X_1, X_2) \geq 0$. Dans ce cas, les variables X_1 et X_2 sont dites positivement liées. Dans le cas contraire, c'est-à-dire $\text{cov}(X_1, X_2) \leq 0$, X_1 et X_2 sont négativement liées.

Nous introduisons à présent la notion d'association qui caractérise la force de la liaison entre les composantes d'un vecteur aléatoire.

Définition 1.2 Soit $X = (X_1, X_2, \dots, X_n)$ un vecteur aléatoire réel. Le vecteur X est dit à composantes positivement associées (PA), respectivement négativement associées (NA), si :

$$\text{cov}(f(X), g(X)) \geq 0, \quad (1.1)$$

respectivement

$$\text{cov}(f(X), g(X)) \leq 0, \quad (1.2)$$

pour toutes fonctions f et g identiquement monotones sur \mathbb{R}^n telles que $\text{cov}(f(X), g(X))$ existe.

Remarque 1.3 i.) Soient $X = (X_1, X_2, \dots, X_n)$ un vecteur aléatoire à composantes positivement associées et $X_{\mathcal{I}} = (X_{i_1}, X_{i_2}, \dots, X_{i_k})$, $1 \leq k < n$, $i_1 < i_2 < \dots < i_k$, un sous-vecteur de X . Soient g_1 et g_2 deux fonctions croissantes sur \mathbb{R}^k telles que $\text{cov}(g_1, g_2)$ existe. On peut prolonger g_1 et g_2 sur \mathbb{R}^n de sorte qu'elles restent encore des fonctions croissantes. On a alors :

$\text{cov}(g_2(X_{\mathcal{I}}), g_1(X_{\mathcal{I}})) = E(g_2(X_{\mathcal{I}})g_1(X_{\mathcal{I}})) - E(g_2(X_{\mathcal{I}}))E(g_1(X_{\mathcal{I}})) \geq 0$; ce qui montre que tout sous-vecteur d'un vecteur aléatoire réel à composantes positivement associées en est un à composantes positivement associées. Une remarque analogue est valable lorsque X est à composantes négativement associées.

ii :) Supposons que $X = (X_1, X_2, \dots, X_n)$ soit à composantes positivement associées et soient $x = (x_1, x_2, \dots, x_n)$ et $a = (a_1, a_2, \dots, a_n)$ deux éléments de \mathbb{R}^n . Posons ensuite $g_1(x) = \mathcal{I}_{(a_1, \infty) \times \dots \times (a_{n-1}, \infty)}(x_1, x_2, \dots, x_{n-1})$ et $g_2(x) = \mathcal{I}_{(a_n, \infty)}(x_n)$. Il est facile de vérifier que les fonctions g_1 et g_2 ainsi définies sont croissantes sur \mathbb{R}^n . Le vecteur X étant à composantes positivement associées, nous avons en vertu de la relation (1.1) :

$$P(X_i > a_i; i = 1, 2, \dots, n) \geq P(X_i > a_i; i = 1, 2, \dots, n-1)P(X_n > a_n).$$

Par la remarque i :) précédente et l'induction suivante :

$P(X_i > a_i; i = 1, 2, \dots, k) \geq P(X_i > a_i; i = 1, 2, \dots, k-1)P(X_k > a_k)$ pour $k = n-1, n-2, \dots, 2$, nous obtenons la relation :

$$P(X_i > a_i; i = 1, 2, \dots, n) \geq \prod_{i=1}^n P(X_i > a_i). \quad (1.3)$$

iii :) De la même façon qu'en ii :), nous avons pour les fonctions décroissantes :

$g_1(x) = \mathcal{I}_{(-\infty, a_1) \times \dots \times (-\infty, a_{n-1})}(x_1, x_2, \dots, x_{n-1})$ et $g_2(x) = \mathcal{I}_{(-\infty, a_n)}(x_n)$, l'inégalité ci-dessous :

$$P(X_i < a_i; i = 1, 2, \dots, n) \geq \prod_{i=1}^n P(X_i < a_i). \quad (1.4)$$

Remarquons que la définition 1.2 exprime que deux fonctions monotones du vecteur aléatoire X admettent une covariance positive ou négative sans toutefois préciser la manière dont varient ses composantes. De fait, nous devons être en mesure de dire comment les petites valeurs, les valeurs moyennes et les grandes valeurs des composantes du vecteur X sont liées. La section suivante lève le voile sur ces points.

1.2.2 Dépendance positive (négative) par orthant.

Définition 1.4 Soit $X = (X_1, X_2)$ un vecteur aléatoire réel bidimensionnel de fonction de répartition F . On dit que X ou F est positivement dépendant par quadrant (PDQ), respectivement négativement dépendant par quadrant (NDQ), si :

$$P(X_1 > a_1, X_2 > a_2) \geq P(X_1 > a_1)P(X_2 > a_2), \quad (1.5)$$

respectivement

$$P(X_1 > a_1, X_2 > a_2) \leq P(X_1 > a_1)P(X_2 > a_2), \quad (1.6)$$

pour tout $(a_1, a_2) \in \mathbb{R}^2$.

Remarque 1.5 a :) Notons que les équivalences suivantes ont lieu :

$$P(X_1 > a_1, X_2 > a_2) \geq P(X_1 > a_1)P(X_2 > a_2)$$

\iff

$$P(X_1 \leq a_1, X_2 \leq a_2) \geq P(X_1 \leq a_1)P(X_2 \leq a_2).$$

b :) La relation (1.5) indique que les variables X_1 et X_2 sont dépendantes et ont une forte tendance à prendre simultanément des grandes valeurs tandis que la relation (1.6), indique que les variables sont dépendantes et ont une faible tendance à prendre simultanément des grandes valeurs. Ainsi, compte tenu des définitions 1.1 et 1.2, nous pouvons résumer ces remarques dans le théorème suivant, dû à Lehmann[6], qui donne une caractérisation de la (PDQ) et de la (NDQ).

Théorème 1.6 Soit $X = (X_1, X_2)$ un vecteur aléatoire réel bidimensionnel de fonction de répartition F . Alors X ou F est (PDQ), respectivement (NDQ) si et seulement si :

$$\text{cov}(f(X_1), g(X_2)) \geq 0,$$

respectivement

$$\text{cov}(f(X_1), g(X_2)) \leq 0,$$

pour toutes fonctions f et g identiquement monotones sur \mathbb{R} dont la covariance existe.

Ce résultat est un cas particulier du théorème 1.10 subséquent.

Définition 1.7 Supposons que $X = (X_1, X_2, \dots, X_n)$ soit un vecteur aléatoire réel de fonction de répartition F . On dit que X ou F est :

g₁ :) positivement dépendant par orthant supérieur (PDOS) si l'inégalité suivante est vérifiée :

$$P(X_i > a_i; i = 1, 2, \dots, n) \geq \prod_{i=1}^n P(X_i > a_i), \quad (1.7)$$

pour tout $(a_1, a_2, \dots, a_n) \in \mathbb{R}^n$;

g₂ :) positivement dépendant par orthant inférieur (PDOI) si l'inégalité suivante est vérifiée :

$$P(X_i \leq a_i; i = 1, 2, \dots, n) \geq \prod_{i=1}^n P(X_i \leq a_i), \quad (1.8)$$

pour tout $(a_1, a_2, \dots, a_n) \in \mathbb{R}^n$,

g₃ :) positivement dépendant par orthant (PDO) si (1.7) et (1.8) sont satisfaites simultanément.

La relation (1.7) exprime que des variables X_i (PDOS) sont dépendantes et ont toutes une forte tendance à prendre simultanément des grandes valeurs. Par une remarque analogue concernant la relation (1.8), nous en arrivons au fait que des X_i (PDOI) sont

dépendantes et ont toutes tendance à prendre simultanément de petites valeurs.

Notons que par la remarque **1.3** et compte tenu du fait que les fonctions $g_1(x) = \mathcal{I}_{(-\infty, a_1) \times \dots \times (-\infty, a_{n-1})}(x_1, x_2, \dots, x_{n-1})$ et $g_2(x) = \mathcal{I}_{(-\infty, a_n)}(x_n)$ sont décroissantes sur \mathbb{R}^n , nous pouvons conclure que si un vecteur X est à composantes positivement associées, alors X est positivement dépendant par orthant.

Définition 1.8 Soit $X = (X_1, X_2, \dots, X_n)$ un vecteur aléatoire réel de fonction de répartition F . On dit que X ou F est :

g₄ :) négativement dépendant par orthant supérieur (NDOS) si

$$P(X_i > a_i; i = 1, 2, \dots, n) \leq \prod_{i=1}^n P(X_i > a_i), \quad (1.9)$$

pour tout $(a_1, a_2, \dots, a_n) \in \mathbb{R}^n$,

g₅ :) négativement dépendant par orthant inférieur (NDOI) si

$$P(X_i \leq a_i; i = 1, 2, \dots, n) \leq \prod_{i=1}^n P(X_i \leq a_i), \quad (1.10)$$

$(a_1, a_2, \dots, a_n) \in \mathbb{R}^n$,

g₆ :) négativement dépendant par orthant (NDO) si (1.9) et (1.10) sont satisfaites simultanément.

Selon la condition (1.9), des variables X_i (NDOS) sont dépendantes et ont toutes peu tendance à prendre simultanément des grandes valeurs. La condition (1.10) nous montre que des variables (NDOI) sont dépendantes et ont toutes peu tendance à prendre simultanément des grandes valeurs. Par la remarque **1.3** et en vertu du fait que les fonctions $g_1(x) = \mathcal{I}_{(a_1, \infty) \times \dots \times (a_{n-1}, \infty)}(x_1, x_2, \dots, x_{n-1})$ et $g_2(x) = \mathcal{I}_{(a_n, \infty)}(x_n)$ sont décroissantes sur \mathbb{R}^n , nous en concluons que si un vecteur X est à composantes négativement associées, alors il est négativement dépendant par orthant.

Proposition 1.9 Soit $X = (X_1, X_2)$. On a les équivalences suivantes :

$$PDOS \iff PDOI \iff PDO$$

et

$$NDOS \iff NDOI \iff NDO.$$

Le théorème suivant donne une caractérisation des formes de dépendance associées aux définitions 1.7 et 1.8 précédentes :

Théorème 1.10 \mathbf{t}_1 :) Soit $X = (X_1, X_2, \dots, X_n)$ un vecteur aléatoire réel,

X est positivement dépendant par orthant supérieur si et seulement si

$$\mathbb{E}\left\{\prod_{i=1}^n g_i(X_i)\right\} \geq \prod_{i=1}^n \mathbb{E}\{g_i(X_i)\},$$

pour toutes fonctions g_1, \dots, g_n ν -intégrables, positives et croissantes sur \mathbb{R} ;

\mathbf{t}_2 :) X est positivement dépendant par orthant inférieur si et seulement si

$$\mathbb{E}\left\{\prod_{i=1}^n g_i(X_i)\right\} \geq \prod_{i=1}^n \mathbb{E}\{g_i(X_i)\}$$

pour toutes fonctions g_1, \dots, g_n ν -intégrables, positives et décroissantes sur \mathbb{R} .

Preuve : \mathbf{t}_1 :)

Considérons le vecteur aléatoire à composantes indépendantes $X^* = (X_1^*, X_2^*, \dots, X_n^*)$ tel que $X_i^* = X_i$ pour $i = 1, 2, \dots, n$ [égalité en distribution] et U le sous ensemble de \mathbb{R}^n défini par : $U = \{(x_1, x_2, \dots, x_n) \in \mathbb{R}^n \mid x_i > a_i\}$, pour a_i élément de \mathbb{R} , $i = 1, 2, \dots, n$.

Remarquons tout d'abord que X est positivement dépendant par orthant supérieur si et seulement si $\mathbb{E}(\mathcal{I}_U(X^*)) \leq \mathbb{E}(\mathcal{I}_U(X))$. Soient g_1, g_2, \dots, g_n , n fonctions de \mathbb{R} dans \mathbb{R} , ν -intégrables, positives et croissantes. Alors, la fonction g , définie par : $g(x_1, x_2, \dots, x_n) = \prod_{i=1}^n g_i(x_i)$, pour tout $x \in \mathbb{R}^n$, est ν -intégrable, positive et croissante sur \mathbb{R}^n . D'après

le théorème de Lebesgue, la fonction g est la limite d'une suite croissante $(f_m)_{m \geq 1}$ de fonctions étagées et mesurables, c'est-à-dire : $g = \lim_{m \rightarrow \infty} f_m$. Dans ce cas, la suite f_m peut toujours s'écrire : $f_m = \sum_{j=1}^m c_j \mathcal{I}_{U_j}$ où $c = (c_1, c_2, \dots, c_m) \in \mathbb{R}^m$. D'après la remarque précédente, il vient :

$$\mathbb{E}(\sum_{j=1}^m c_j \mathcal{I}_{U_j}(X^*)) \leq \mathbb{E}(\sum_{j=1}^m c_j \mathcal{I}_{U_j}(X)), \text{ et, par passage à la limite, nous obtenons :}$$

$$\mathbb{E}(\lim_{m \rightarrow \infty} f_m(X^*)) \leq \mathbb{E}(\lim_{m \rightarrow \infty} f_m)(X).$$

Donc, $\mathbb{E}(\prod_{i=1}^n g_i(X_i^*)) \leq \mathbb{E}(\prod_{i=1}^n g_i(X_i))$. Par l'indépendance entre les variables $g_i(X_i^*)$ et compte tenu du fait que $g_i(X_i^*) = g_i(X_i)$, nous avons : $\mathbb{E}(\prod_{i=1}^n g_i(X_i)) \geq \prod_{i=1}^n \mathbb{E}(g_i(X_i))$. La condition suffisante s'obtient en considérant le cas particulier des fonctions g_i , ν -intégrables, croissantes, positives et définies par : $g_i = \mathcal{I}_{(a_i, \infty)}(x_i)$, pour $i = 1, 2, \dots, n$. La preuve de la partie **t₂** :) du théorème s'obtient de façon similaire.

Remarque 1.11 *La preuve du théorème 1.6 s'en déduit aussitôt.*

1.2.3 Quelques propriétés des dépendances positives (négatives).

Nous étudions dans cette section les principaux résultats relatifs aux formes de dépendance que nous venons de définir.

Théorème 1.12 *Soient $X = (X_1, X_2, \dots, X_n)$ un vecteur aléatoire réel et f_1, f_2, \dots, f_n , n fonctions de \mathbb{R} dans \mathbb{R} strictement croissantes. Alors :*

p₁ :) *X est à composantes positivement, respectivement négativement associées entraîne que le vecteur $(f_1(X_1), f_2(X_2), \dots, f_n(X_n))$ est à composantes positivement, respectivement négativement associées.*

p₂ :) *X est positivement, respectivement négativement dépendant par orthant implique que $(f_1(X_1), f_2(X_2), \dots, f_n(X_n))$ est positivement, respectivement négativement dépendant par orthant.*

Preuve :

p₁ :) Considérons f et g , deux fonctions réelles ν -intégrables, positives et croissantes sur \mathbb{R}^n et f_1, f_2, \dots, f_n , n fonctions numériques strictement croissantes. Posons :

$$h(x_1, x_2, \dots, x_n) = (f_1(x_1), f_2(x_2), \dots, f_n(x_n)), \text{ pour tout } (x_1, x_2, \dots, x_n) \in \mathbb{R}^n,$$

$h_1 = f \circ h$ et $h_2 = g \circ h$. Alors les fonctions h_1 et h_2 ainsi définies sont ν -intégrables, positives et croissantes sur \mathbb{R}^n . Si le vecteur X est à composantes positivement associées, nous concluons à l'aide de la définition 1.2 que :

$$(f_1(X_1), f_2(X_2), \dots, f_n(X_n)) \text{ est aussi à composantes positivement associées.}$$

Le cas où X est à composantes négativement associées s'obtient de façon similaire.

p₂ :) Supposons que X soit positivement dépendant par orthant et que f_1, f_2, \dots, f_n soient n fonctions de \mathbb{R} dans \mathbb{R} strictement croissantes. Nous avons : $X_i \geq x_i \iff f_i(X_i) \geq f_i(x_i)$, et $X_i \leq x_i \iff f_i(X_i) \leq f_i(x_i)$, pour $i = 1, 2, \dots, n$ et le résultat s'en déduit à l'aide des inégalités (1.9) et (1.10).

Le cas où X est négativement dépendant par orthant se fait de façon similaire.

Théorème 1.13 Soit $X = (X_1, X_2, \dots, X_n)$ un vecteur aléatoire réel et f_1, f_2, \dots, f_n , n fonctions strictement croissantes sur \mathbb{R}^n . Alors :

p₃ :) Si X est à composantes positivement, respectivement négativement associées, alors $(f_1(X), f_2(X), \dots, f_n(X))$ est à composantes positivement, respectivement négativement associées.

Il s'ensuit que si X est à composantes positivement, respectivement négativement associées, alors $(f_1(X), f_2(X), \dots, f_n(X))$ est positivement, respectivement négativement dépendant par orthant.

p₄ :) Soit $Y = (Y_1, Y_2, \dots, Y_m)$ un autre vecteur aléatoire réel tel que X et Y soient indépendants. Si X et Y sont à composantes à positivement, respectivement négativement

tivement associées, alors $(X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m)$ est à composantes positivement, respectivement négativement associées.

Ainsi, lorsque X et Y sont positivement dépendants par orthant, respectivement négativement dépendants par orthant, il en est de même du vecteur $(X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m)$.

Preuve:

p₃ :) Supposons que X soit à composantes positivement associées. Soient f et g deux fonctions réelles ν -intégrables, positives et croissantes sur \mathbb{R}^n . Posons :

$$h(x_1, x_2, \dots, x_n) = (f_1(x_1, x_2, \dots, x_n), f_2(x_1, x_2, \dots, x_n), \dots, f_n(x_1, x_2, \dots, x_n)),$$

pour $x \in \mathbb{R}^n$, $h_1 = f \circ h$ et $h_2 = g \circ h$. Alors h_1 et h_2 sont positives et croissantes sur \mathbb{R}^n et le résultat s'en déduit à l'aide de la définition **1.2**.

Le cas où X est à composantes négativement associées s'obtient de façon analogue.

p₄ :) Soient $X = (X_1, X_2, \dots, X_n) \in \mathbb{R}^n$ et $Y = (Y_1, Y_2, \dots, Y_m) \in \mathbb{R}^m$ deux vecteurs aléatoires à composantes positivement associées et tels que X et Y soient indépendants, f et g deux fonctions ν -intégrables, positives et croissantes sur \mathbb{R}^{n+m} . Notons f pour $f(X, Y)$, g pour $g(X, Y)$ et désignons par $\mathbb{E}_{X,Y}$, \mathbb{E}_X et \mathbb{E}_Y les espérances respectives par rapport à la loi conjointe du couple (X, Y) et aux distributions marginales de X et de Y . Le vecteur Y étant (PA), $cov_Y(f, g) \geq 0$ et donc $\mathbb{E}_X(cov_Y(f, g)) \geq 0$. Par ailleurs, f et g étant croissantes, les fonctions $\mathbb{E}_Y(f)$ et $\mathbb{E}_Y(g)$ sont aussi croissantes et, X étant (PA), alors $cov_X\{\mathbb{E}_Y(f), \mathbb{E}_Y(g)\} \geq 0$, nous pouvons écrire :

$$cov(f, g) = \mathbb{E}_{X,Y}(fg) - \mathbb{E}_{X,Y}(f)\mathbb{E}_{X,Y}(g) \tag{1.11}$$

ou encore :

$$\begin{aligned} \text{cov}(f, g) &= \mathbb{E}_X \{ \mathbb{E}_Y (fg) \} - \mathbb{E}_X \{ \mathbb{E}_Y (f) \mathbb{E}_Y (g) \} + \\ &\quad \mathbb{E}_X \{ \mathbb{E}_Y (f) \mathbb{E}_Y (g) \} - \mathbb{E}_X (\mathbb{E}_Y (f)) \mathbb{E}_X (\mathbb{E}_Y (g)), \end{aligned}$$

en vertu du théorème de Fubini et de l'indépendance entre X et Y .

Ainsi,

$$\begin{aligned} \text{cov}(f, g) &= \mathbb{E}_X \{ \mathbb{E}_Y (fg) - \mathbb{E}_Y (f) \mathbb{E}_Y (g) \} + \text{cov}_X (\mathbb{E}_Y (f), \mathbb{E}_Y (g)) \\ &= \mathbb{E}_X \{ \text{cov}_Y (f, g) \} + \text{cov}_X \{ \mathbb{E}_Y (f), \mathbb{E}_Y (g) \} > 0. \end{aligned}$$

L'identité suivante, due à Hoeffding sera utile par la suite.

Lemme 1.14 *Soit $X = (X_1, X_2, \dots, X_n)$ un vecteur aléatoire réel. Pour toutes fonctions croissantes f et g sur \mathbb{R}^n dont la covariance existe, nous avons :*

$$\text{cov}(f(X), g(X)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{cov}(\mathcal{I}_{(s, \infty)}(f(x)), \mathcal{I}_{(t, \infty)}(g(x))) ds dt.$$

La preuve de ce lemme est présentée, entre autres, dans Lehmann [6].

Remarque 1.15 *Lorsque les fonctions f et g sont numériques, croissantes et positives, nous obtenons pour toute variable aléatoire réelle X , l'inégalité suivante connue sous le nom d'inégalité de Tchébychev :*

$$\text{cov}(f(X), g(X)) \geq 0.$$

En posant $f(X) = X_i$ et $g(X) = X_j$, nous obtenons la relation suivante :

$$\text{cov}(X_i, X_j) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (F(x_i, x_j) - F_i(x_i)F_j(x_j)) dx_i dx_j, \quad (1.12)$$

pour $1 \leq i < j \leq n$ où F , F_i et F_j désignent respectivement la fonction de répartition du couple (X_i, X_j) et les fonctions de répartition marginales de X_i et X_j .

Le théorème qui suit donne une condition suffisante pour que les composantes d'un vecteur soient associées.

Théorème 1.16 *Soient $X = (X_1, X_2, \dots, X_n)$ un vecteur aléatoire réel. Si l'inégalité $\text{cov}(u(X), v(X)) \geq 0$ est satisfaite pour toutes fonctions u et v binaires (dont les valeurs sont 0 et 1) et croissantes sur \mathbb{R}^n , alors X est à composantes positivement associées. La preuve de ce théorème est une application de l'identité de Hoeffding précédente. Nous en déduisons le résultat ci-après :*

Corollaire 1.17 *Si $X = (X_1, X_2, \dots, X_n)$ est un vecteur aléatoire réel à composantes positivement associées, alors toutes les variables de X sont indépendantes deux à deux si et seulement si $\text{cov}(X_i, X_j) = 0$, pour $1 \leq i < j \leq n$.*

La preuve de ce corollaire découle de la relation (1.12).

1.2.4 Positivité et négativité totales multidimensionnelles d'ordre deux.

Nous avons obtenu quelques caractérisations de la notion d'association qui, comme nous avons pu le remarquer, ne se manipulent pas sans difficultés. Nous introduisons dans cette section les notions de positivité et négativité totales multidimensionnelles d'ordre deux qui, comme Sarkar [7] l'a démontré, induisent les conditions d'association et offrent ainsi des conditions de dépendance plus faciles à vérifier une fois que la fonction de densité du vecteur aléatoire concerné est connue. L'autre intérêt de ces formes de dépendance est qu'elles nous permettent d'obtenir au chapitre suivant les conditions de dépendance

et d'indépendance des composantes d'un vecteur aléatoire continu basée sur les matrices d'information de Fisher.

Définition 1.18 Soit $Z = (X, Y)$ un couple de variables aléatoires réelles continues de densité conjointe f et admettant $A = A_1 \times A_2 \subset \mathbb{R}^2$ comme support. On dit que Z ou f jouit de la positivité totale bidimensionnelle d'ordre deux si la condition suivante est satisfaite :

$$f(x_1, y_1)f(x_2, y_2) \geq f(x_1, y_2)f(x_2, y_1), \quad (1.13)$$

pour tous les couples (x_1, y_1) et (x_2, y_2) dans A tels que $x_1 < x_2$ et $y_1 < y_2$.

La condition précédente (1.13) exprime que le vecteur aléatoire X a plus tendance à prendre les paires de couples dont l'un (des couples) est à composantes de grandes valeurs tandis que l'autre est à composantes de petites valeurs, que les paires de couples dont les composantes sont de grandeurs différentes. En changeant le sens de l'inégalité dans la relation précédente (1.13), nous obtenons la négativité totale bidimensionnelle d'ordre deux. Cette inégalité est équivalente à :

$$\frac{f(x_1, y_1)f(x_2, y_2)}{f(x_1, y_2)f(x_2, y_1)} \geq 1; \quad (1.14)$$

par passage au logarithme (1.14) équivaut à :

$$\log \frac{f(x_1, y_1)f(x_2, y_2)}{f(x_1, y_2)f(x_2, y_1)} \geq 0;$$

en divisant les deux membres de l'inégalité précédente par $\Delta x_1 = x_2 - x_1 > 0$ et $\Delta y_1 = y_2 - y_1 > 0$, on obtient de façon équivalente :

$$\log \left\{ \frac{f(x_1, y_1)f(x_1 + \Delta x_1, y_1 + \Delta y_1)}{f(x_1, y_1 + \Delta y_1)f(x_1 + \Delta x_1, y_1)} \right\} / \Delta x_1 \Delta y_1 \geq 0;$$

ce qui équivaut, par passage à la limite, à :

$$\lim_{\substack{\Delta x_1 \rightarrow 0 \\ \Delta y_1 \rightarrow 0}} \log \left\{ \frac{f(x_1, y_1) f(x_1 + \Delta x_1, y_1 + \Delta y_1)}{f(x_1, y_1 + \Delta y_1) f(x_1 + \Delta x_1, y_1)} \right\} / \Delta x_1 \Delta y_1 \geq 0.$$

Cette dernière inégalité est équivalente à :

$$\frac{\partial^2}{\partial x_1 \partial y_1} \log f(x_1, y_1) \geq 0.$$

En cas de négativité totale bidimensionnelle d'ordre deux nous avons :

$$\frac{\partial^2}{\partial x_1 \partial y_1} \log f(x_1, y_1) \leq 0.$$

Ces remarques nous permettent de caractériser ces deux formes de dépendance dans le théorème qui suit.

Théorème 1.19 *Considérons $Z = (X, Y)$ un couple de variables aléatoires réelles de densité f et admettant A pour support. Alors : Z vérifie la positivité, respectivement négativité totale multidimensionnelle d'ordre deux si et seulement si $\frac{\partial^2}{\partial x \partial y} \log f(x, y) \geq 0$, respectivement $\frac{\partial^2}{\partial x \partial y} \log f(x, y) \leq 0, \forall x, y \in \mathbb{R}$.*

Une généralisation multidimensionnelle de ces deux formes de dépendance est donnée à l'aide de la définition suivante :

Définition 1.20 *Soit $X = (X_1, X_2, \dots, X_n) \in \mathbb{R}^n$ un vecteur aléatoire de densité conjointe f . On dit que X jouit de la positivité, respectivement la négativité, totale multidimensionnelle d'ordre deux si pour toutes paires d'arguments x_i et x_j , la densité f , vue comme fonction de x_i et x_j lorsque les autres arguments restent fixés, vérifie la positivité,*

respectivement la négativité, totale bidimensionnelle d'ordre deux. On pourra consulter entre autres [3] pour une étude plus détaillée de cette forme de dépendance.

Une démarche en tout point semblable à la précédente conduit à une caractérisation de ces deux formes de dépendance comme énoncée dans le théorème ci-après.

Théorème 1.21 Soit $X = (X_1, X_2, \dots, X_n) \in \mathbb{R}^n$ un vecteur aléatoire de densité conjointe f . Alors :

X jouit de la positivité, respectivement négativité totale multidimensionnelle d'ordre deux si et seulement si $\frac{\partial^2}{\partial x_i \partial x_j} \log f(x_1, x_2, \dots, x_n) \geq 0$
respectivement $\frac{\partial^2}{\partial x_i \partial x_j} \log f(x_1, x_2, \dots, x_n) \leq 0, \forall i \neq j$.

Pour une étude plus détaillée de ces formes de dépendance l'on pourra consulter entre autres Karlin, S. [8].

1.2.5 Dépendance séquentielle.

Toujours dans le but de décrire la structure de dépendance d'un vecteur aléatoire et plus particulièrement d'obtenir les conditions de dépendance positive et de dépendance négative de ce dernier, nous introduisons ci-après quelques formes de dépendance conditionnelle.

Définition 1.22 Soit $X = (X_1, X_2) \in \mathbb{R}^2$. On dit que la variable X_1 est stochastiquement :

- s_1 :) croissante par rapport à la variable X_2 (SC) si : $P(X_2 > x_2 | X_1 = x_1)$ est croissante par rapport à x_1 , pour tout $x_2 \in \mathbb{R}$;
- s_2 :) décroissante par rapport à la variable X_2 (SD) si : $P(X_2 > x_2 | X_1 = x_1)$ est décroissante par rapport à x_1 , pour tout $x_2 \in \mathbb{R}$.

Notons que dans le cas où les variables sont continues, nous avons

$$P(X_2 > x_2 | X_1 = x_1) = \int_{x_2}^{\infty} f_{X_2|X_1}(u|x_1) du$$

où

$$f_{X_2|X_1}(x_2|x_1) = \frac{f_{(X_1, X_2)}(x_1, x_2)}{f_{X_1}(x_1)}.$$

Une extension multidimensionnelle de ces deux formes dépendance est donnée par :

Définition 1.23 Soit $X = (X_1, X_2, \dots, X_n) \in \mathbb{R}^n$.

On dit que X est conditionnellement :

s₃ :) *croissante en séquence (CCS) si : $P(X_i > x_i | X_1 = x_1, X_2 = x_2, \dots, X_{i-1} = x_{i-1})$ est croissante par rapport à x_1, x_2, \dots, x_{i-1} , pour tout $x_i \in \mathbb{R}$, et $i = 1, 2, \dots, n$;*

s₄ :) *décroissante en séquence (CDS) si : $P(X_i > x_i | X_1 = x_1, X_2 = x_2, \dots, X_{i-1} = x_{i-1})$ est décroissante par rapport à x_1, x_2, \dots, x_{i-1} , pour tout $x_i \in \mathbb{R}$, et $i = 1, 2, \dots, n$.*

Définition 1.24 Soit $X = (X_1, X_2, \dots, X_n) \in \mathbb{R}^n$ un vecteur aléatoire. On dit de X qu'il est :

s₅ :) *positivement dépendant en séquence (PDS) si $P(X_i > x_i; i \neq j | X_j = x)$ est croissante par rapport à x réel pour tout $i, j = 1, 2, \dots, n$.*

s₆ :) *négativement décroissant en séquence (NDS) si $P(X_i > x_i; i \neq j | X_j = x)$ est décroissante par rapport à x réel, pour tout $i, j = 1, 2, \dots, n$.*

Notons que tout sous-vecteur d'un vecteur aléatoire positivement, respectivement négativement dépendant en séquence est aussi positivement, respectivement négativement

dépendant. En outre, il existe un lien entre les formes de dépendance précédentes et les formes de dépendance de types (PDO) , (NDO) , (PA) et (NA) comme le montre le théorème suivant.

Théorème 1.25 Soit $X = (X_1, X_2, \dots, X_n) \in \mathbb{R}^n$.

a :) Si X est positivement, respectivement négativement dépendant en séquence, alors X est positivement, respectivement négativement dépendant par orthant;

b :) Si X est conditionnellement croissant, respectivement décroissant en séquence, alors X est à composantes positivement, respectivement négativement associées.

Preuve : a :) Si X est positivement dépendant en séquence, alors :

$$P(X_2 > x_2, X_3 > x_3, \dots, X_n > x_n | X_1 = x_1) \geq P(X_2 > x_2, X_3 > x_3, \dots, X_n > x_n | X_1 = x'_1),$$

pour $(x_1, x'_1) \in \mathbb{R}^2$ tel que $x_1 \geq x'_1$ et pour $(x_2, \dots, x_n) \in \mathbb{R}^{n-1}$. En remarquant que $x'_1 < x_1 \implies \int_{x'_1}^{\infty} dF_1(z) \geq \int_{x_1}^{\infty} dF_1(z)$,

on a :

$$\begin{aligned} P(X_2 > x_2, X_3 > x_3, \dots, X_n > x_n | X_1 \geq x_1) &= \\ &= \frac{\int_{x_1}^{\infty} P(X_2 > x_2, X_3 > x_3, \dots, X_n > x_n | X_1 = z) dF_1(z)}{\int_{x_1}^{\infty} dF_1(z)} \\ &\geq \frac{\int_{x'_1}^{\infty} P(X_2 > x_2, X_3 > x_3, \dots, X_n > x_n | X_1 = z) dF_1(z)}{\int_{x'_1}^{\infty} dF_1(z)} \\ &= P(X_2 > x_2, X_3 > x_3, \dots, X_n > x_n | X_1 \geq x'_1), \end{aligned}$$

pour tout $(x_1, x'_1) \in \mathbb{R}^2$ tel que $x_1 \geq x'_1$. En faisant tendre x'_1 vers $-\infty$ nous obtenons :

$P(X_i > x_i; i = 1, 2, \dots, n) \geq P(X_1 > x_1)P(X_2 > x_2, X_3 > x_3, \dots, X_n > x_n)$. Et, comme tout sous-vecteur d'un vecteur aléatoire positivement dépendant en séquence en est un, le sous-vecteur $(X_i, \dots, X_n), i = 1, 2, \dots, n - 1$ est positivement dépendant en séquence et par récurrence nous avons que X est positivement dépendant par orthant supérieur. Le cas où X est positivement dépendant par orthant inférieur s'obtient de façon similaire. Enfin, une preuve analogue peut-être faite pour X négativement dépendant en séquence.

b :) Pour démontrer cette partie du théorème, commençons par réécrire la relation (1.11) de la page 14 sous la forme suivante : si $(U, V) \in \mathbb{R}^2$ est un couple de variables aléatoires et Z une autre variable aléatoire réelle, alors la covariance entre U et V peut s'écrire :

$$\text{cov}(U, V) = \mathbb{E}\{\text{cov}(U, V)|Z\} + \text{cov}\{\mathbb{E}(U|Z), \mathbb{E}(V|Z)\}. \quad (1.15)$$

Supposons que $X = (X_1, X_2, \dots, X_n)$ soit conditionnellement croissant en séquence et soient f, g deux fonctions réelles positives et croissantes sur \mathbb{R}^n . D'après la relation (1.15) précédente, nous avons :

$$\text{cov}(f(X), g(X)) = \mathbb{E}\{\text{cov}(f(X), g(X)|X_1\} + \text{cov}\{\mathbb{E}(f(X)|X_1), \mathbb{E}(g(X)|X_1)\}. \quad (1.16)$$

Posons $f_1(X_1) = \mathbb{E}(f(X)|X_1)$ et $g_1(X_1) = \mathbb{E}(g(X)|X_1)$. Les fonctions f_1 et g_1 ainsi définies sont positives et croissantes et d'après l'inégalité de Tchébychev, le second terme de la relation (1.16) est positif. Pour x_1 fixé, $f(x_1, x_2, \dots, x_n)$ et $g(x_1, x_2, \dots, x_n)$ sont des fonctions positives et croissantes de x_2, \dots, x_n ; ce qui montre que le premier terme de (1.16) est positif, d'où $\text{cov}(f(X), g(X)) \geq 0$ et donc X est à composantes positivement associées. Une preuve analogue peut être faite lorsque X est conditionnellement décroissant en séquence.

Loin d'être exhaustive quant aux formes de dépendance pouvant être mises en évidence dans un vecteur aléatoire, l'étude qui vient d'être faite décrit pour nous les hypothèses de travail en même temps qu'elle sous-tend notre objectif de mesurer la force ou l'intensité d'une forme de dépendance donnée.

CHAPITRE 2

Dépendance et matrice d'information de Fisher.

Après l'étude au chapitre 1, des principales formes de dépendance entre les composantes d'un vecteur aléatoire réel, nous poursuivons notre objectif principal qui consiste à mesurer la force ou l'intensité de chacune d'elles. Mais avant cela, nous proposons dans ce chapitre une idée originale de K. Zografos qui nous permet de détecter la dépendance et l'indépendance (mutuelle) entre les composantes d'un vecteur aléatoire donné. Cette approche est essentiellement basée sur la matrice d'information de Fisher. Après avoir défini précisément ce que l'on entend par quantité d'information de Fisher sur un paramètre contenu dans la loi d'un vecteur aléatoire, nous introduisons la matrice d'information de Fisher, en étudions ses propriétés et nous caractérisons l'indépendance entre les composantes d'un vecteur aléatoire dont la densité vérifie la positivité ou la négativité totales multidimensionnelles d'ordre 2.

2.1 Notations.

Le cadre probabiliste et les notations restent globalement les mêmes qu'au chapitre 1. Nous considérons l'espace probabilisé $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, \nu, \mathcal{P}_\theta; \theta \in \Theta)$ où $\mathcal{B}_{\mathbb{R}^n} = \otimes_{i=1}^n \mathcal{B}_i$, \mathcal{B}_i étant la σ -algèbre des boréliens de \mathbb{R} . La fonction $f(x, \theta)$ désignera la densité d'un vecteur aléatoire réel continu $X = (X_1, X_2, \dots, X_n)$ dont les densités marginales seront notées $f_i(x_i, \theta)$, pour $i = 1, 2, \dots, n$. De plus, la quantité θ sera appelée paramètre de tendance centrale ou de centralité unidimensionnel ($\theta \in \mathbb{R}$) ou multidimensionnel ($\theta = (\theta_1, \theta_2, \dots, \theta_p) \in \mathbb{R}^p$) pour le vecteur X .

La fonction $f_0(x_1, x_2, \dots, x_n, \theta) = \prod_{i=1}^n f_i(x_i, \theta)$ désignera la densité de X en cas d'indépendance tandis que $E_\theta = \{x \in \mathbb{R}^n | f(x, \theta) > 0\}$ et $E_{\theta_i} = \{x_i \in \mathbb{R} | f_i(x_i, \theta) > 0\}$ seront les supports respectifs de X et X_i .

Notons qu'en général $\theta \in \Theta \subset \mathbb{R}^p$, avec $p \neq n$ et que la quantité d'information est définie par rapport aux familles de distributions admettant un paramètre. Dans tout le texte qui suit, nous prendrons $p = n$. Avant de définir la quantité d'information de Fisher, il est important de présenter les notions de paramètres de tendance centrale et les conditions de régularité que doit satisfaire une fonction de densité $f(x, \theta)$ dans le cadre de ce chapitre et du mémoire.

2.2 Paramètre de tendance centrale.

Soit Y une variable aléatoire réelle de densité $g(y, \theta)$, $\theta \in \mathbb{R}$. Posons $g_0(y) = g(y, 0)$. Nous dirons que θ est un paramètre de tendance centrale pour Y si la variable aléatoire $Y - \theta$ admet une fonction de densité qui ne dépend plus du paramètre θ . Autrement dit la fonction de densité de la variable aléatoire Y vérifie $g(y, \theta) = g_0(y - \theta)$.

Une généralisation multidimensionnelle est donnée par :

le vecteur aléatoire $X = (X_1, X_2, \dots, X_n)$ étant de densité $f(x, \theta)$, $\theta = (\theta_1, \theta_2, \dots, \theta_n) \in \mathbb{R}^n$ est un vecteur de paramètre de tendance centrale si le vecteur $X - \theta = (X_1 - \theta_1, X_2 - \theta_2, \dots, X_n - \theta_n)$ admet une fonction de densité qui ne dépend plus du paramètre θ . La densité vérifie alors l'égalité $f(x, \theta) = f_0(x - \theta)$. Pour simplifier la notation nous écrirons désormais $f(x)$ pour $f_0(x)$.

On peut trouver dans [9] plusieurs exemples de variables ou vecteurs aléatoires réels admettant un paramètre de centralité.

2.3 Conditions de régularité.

On dira qu'une fonction de densité $f(x, \theta)$, $\theta \in \mathbb{R}^n$ est régulière si les conditions suivantes sont satisfaites :

R₁ :) $E_\theta = \{x \in \mathbb{R}^n | f(x, \theta) > 0\}$ est indépendant de θ . Nous le noterons désormais E .

R₂ :) $\nabla f(x, \theta) = (\frac{\partial}{\partial \theta_1} f(x, \theta), \frac{\partial}{\partial \theta_2} f(x, \theta), \dots, \frac{\partial}{\partial \theta_n} f(x, \theta))$ existe pour tout θ_i ν -presque sûrement.

R₃ :) $\forall A \in \mathcal{B}_{\mathbb{R}^n}$, A ne dépend pas de θ et pour toute statistique T telle que $\mathbb{E}(|T(X)|) < \infty$,

nous avons :

$$\frac{\partial^2}{\partial \theta_i \partial \theta_j} \int_A T(x) f(x, \theta) dx = \int_A T(x) \frac{\partial^2}{\partial \theta_i \partial \theta_j} f(x, \theta) dx, \quad (2.1)$$

pour $1 \leq i, j, \leq n$ et telle que le membre de droite de (2.1) existe.

Nous présenterons deux cas de la quantité d'information sur un paramètre θ : le cas où θ est un paramètre unidimensionnel et celui où θ est multidimensionnel. Dans

le premier cas, nous parlerons de la quantité d'information de Fisher tandis que dans le deuxième cas, nous aurons la matrice d'information de Fisher.

2.4 Quantité d'information de Fisher.

Définition 2.1 Soit $X = (X_1, X_2, \dots, X_n)$ un vecteur aléatoire continu dont la densité $f(x, \theta)$, $\theta \in \Theta \subset \mathbb{R}$ satisfait aux conditions de régularité. On appelle quantité d'information de Fisher fournie par le vecteur aléatoire X sur le paramètre inconnu θ , la quantité si elle existe, définie par :

$$\begin{aligned} I(\theta) &= \mathbb{E}\left[\left(\frac{\partial}{\partial \theta} \log f(x, \theta)\right)^2\right] \\ &= \int_E \left[\frac{\partial}{\partial \theta} \log f(x, \theta)\right]^2 f(x, \theta) dx. \end{aligned}$$

Afin d'alléger les écritures, nous noterons désormais la quantité d'information $I(\theta)$ par I ou I_X .

2.4.1 Propriétés de la quantité d'information de Fisher.

Les propriétés de la quantité d'information de Fisher sont connues et nous en donnons ici deux dont nous nous servirons par la suite.

P₁ :) Sous les hypothèses de régularité, nous pouvons écrire la quantité d'information de Fisher sous l'une des deux formes équivalentes suivantes :

$$\begin{aligned} I &= \text{Var}\left(\left[\frac{\partial}{\partial \theta} \log f(x, \theta)\right]\right) \\ &= -\mathbb{E}\left[\frac{\partial^2}{\partial \theta^2} \log f(x, \theta)\right], \end{aligned}$$

pour tout $\theta \in \Theta$. Cette remarque résulte du fait que la variable aléatoire réelle $\frac{\partial}{\partial \theta} \log f(x, \theta)$ est centrée.

En outre, la quantité d'information est toujours positive.

P₂ : Soit $\tilde{X} = (X_1, X_2, \dots, X_n)$ un n -échantillon d'une variable aléatoire réelle continue X . Désignons par I_X la quantité d'information sur le paramètre θ . Sous les hypothèses de régularité, la quantité d'information fournie par \tilde{X} est donnée par :

$$I_{\tilde{X}} = nI_X, \text{ pour tout } \theta \in \Theta.$$

Autrement dit, un n -échantillon d'une variable aléatoire réelle contient n fois plus d'information sur le paramètre θ qu'une seule observation.

2.4.2 Exemples.

Supposons que la variable Y suive une normale $N(\mu, \sigma^2)$ et que nous disposions d'un échantillon indépendant $\tilde{Y} = (Y_1, Y_2, \dots, Y_n)$.

Déterminons la quantité d'information $I(\mu)$ apportée par cet échantillon sur la moyenne μ .

Calculs de $I(\mu)$ pour un échantillon.

Nous avons :

$$f(y_1, y_2, \dots, y_n, \mu) = \frac{1}{(2\pi)^{\frac{n}{2}}} \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right)$$

$$\log f(y_1, y_2, \dots, y_n, \mu) = -\frac{n}{2} \log 2\pi - n \log \sigma - \frac{\sum_{i=1}^n (y_i - \mu)^2}{2\sigma^2}$$

$$\frac{\partial}{\partial \mu} \log f(y_1, y_2, \dots, y_n, \mu) = \frac{1}{\sigma^2} \left(\sum_{i=1}^n y_i - n\mu \right)$$

$$\frac{\partial^2}{\partial \mu^2} \log f(y_1, y_2, \dots, y_n, \mu) = -\frac{n}{\sigma^2}$$

et en prenant l'espérance dans le cas de cette dernière relation, nous avons : $I(\mu) = \frac{n}{\sigma^2}$ et dans le cas où $n = 1$, $I(\mu) = \frac{1}{\sigma^2}$ [information apportée par une seule observation]. Nous pouvons dire que l'information apportée par un échantillon est d'autant importante que la variance est petite et la taille d'échantillon élevée. Cette information permet dans ce cas, de mieux discriminer entre la vraie valeur de la moyenne μ et les valeurs voisines. C'est le cas inverse quand la variance est très grande ou en présence d'un petit échantillon.

Déterminons la quantité d'information $I(\sigma^2)$ apportée par cet échantillon sur la variance σ^2 .

Quantité d'information sur σ^2 pour le cas d'une observation .

$$\log f(y, \sigma^2) = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 - \frac{(y-\mu)^2}{2\sigma^2}$$

$$\frac{\partial}{\partial \sigma^2} \log f(y, \sigma^2) = -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}(y - \mu)^2$$

$$\frac{\partial^2}{(\partial \sigma^2)^2} \log f(y, \sigma^2) = -\frac{1}{2\sigma^4} + \frac{1}{\sigma^6}(y - \mu)^2$$

$$\mathbb{E}\left(\frac{\partial^2}{(\partial \sigma^2)^2} \log f(y, \sigma^2)\right) = -\frac{1}{2\sigma^4} + \frac{1}{\sigma^4} \mathbb{E}\frac{(y - \mu)^2}{\sigma^2}.$$

Comme $\frac{(Y-\mu)}{\sigma}$ est une variable normale centrée réduite, $\frac{(Y-\mu)^2}{\sigma^2}$ est une *khi - deux* à un degré de liberté et donc d'espérance 1. Ainsi, $I(\sigma^2) = -\frac{1}{2\sigma^4} + \frac{1}{\sigma^4} = \frac{1}{2\sigma^4}$. Par **P₂** :), nous en déduisons la quantité d'information $I(\sigma^2)$ pour un échantillon de taille n : $I(\sigma^2) = \frac{n}{2\sigma^4}$. Comme précédemment, cette information est fonction de la grandeur de la variance de la variable d'intérêt Y et de la taille d'échantillon observé.

2.5 Matrice d'information de Fisher.

2.5.1 Quelques rappels sur les matrices semi-définies et définies positives.

Définition 2.2 Soit A une matrice carrée symétrique $n \times n$.

A est dite semi-définie positive (SDP) si : $x^t Ax \geq 0$ pour tout x dans \mathbb{R}^n .

Elle est dite définie-positive (DP) si : $x^t Ax > 0$ pour tout $x \neq 0$ dans \mathbb{R}^n .

Par la suite, nous écrirons indifféremment A SDP ou $A \geq 0$ et A DP ou $A > 0$.

Nous noterons aussi $A \geq B$, respectivement $A > B$ pour exprimer que $A - B$ est SDP, respectivement DP.

2.5.2 Quelques propriétés des matrices (S)DP.

M₁ :) A est SDP, respectivement DP si et seulement si toutes les valeurs propres de A sont positives ou nulles, respectivement strictement positives.

M₂ :) Soit A telle que A^{-1} existe. Alors A est DP si et seulement si A^{-1} est DP.

M₃ :) Soit A telle que A^{-1} existe. Alors A^{-1} est symétrique.

M₄ :) Soient A et B deux matrices $n \times n$. On a : A et B (S)DP si et seulement si AB est (S)DP.

M₅ :) Si A est une matrice symétrique alors

$$x^t Ax = tr(Axx^t),$$

où $tr(\cdot)$ désigne la trace d'une matrice donnée.

M₆ :) A est (S)DP si et seulement si toute sous-matrice principale de A est (S)DP.

Rappelons que A_S est une sous-matrice principale de $A = (a_{i,j})_{1 \leq i \leq j \leq n}$ si $A_S = (a_{i,j})$, $i, j \in J$ avec $J \subset \{1, 2, \dots, n\}$.

2.5.3 Matrice d'information.

Soit $X = (X_1, X_2, \dots, X_n)$ un vecteur aléatoire continu de densité régulière $f(x, \theta)$, $X \in \mathbb{R}^n$, et $\theta \in \Theta \subset \mathbb{R}^n$ un vecteur de paramètres de tendance centrale. On peut vérifier que le vecteur aléatoire réel

$\nabla \log f(x, \theta) = (\frac{\partial}{\partial \theta_1} \log f(x, \theta), \dots, \frac{\partial}{\partial \theta_n} \log f(x, \theta))$ est centré.

Définition 2.3 On appelle matrice d'information de Fisher concernant θ fournie par le vecteur aléatoire X , la matrice de variance-covariance du vecteur aléatoire réel $\nabla \log f(x, \theta)$. Elle sera notée \mathbf{I} et ses éléments sont donnés par :

$$\begin{aligned} \mathbf{I}_{i,j} &= \text{cov}[\frac{\partial}{\partial \theta_i} \log f(x, \theta), \frac{\partial}{\partial \theta_j} \log f(x, \theta)] \\ &= \mathbb{E}[\frac{\partial}{\partial \theta_i} \log f(x, \theta) \frac{\partial}{\partial \theta_j} \log f(x, \theta)] \\ &= \int_E [\frac{\partial}{\partial \theta_i} \log f(x, \theta) \frac{\partial}{\partial \theta_j} \log f(x, \theta)] f(x, \theta) dx, \end{aligned}$$

pour $i \leq i, j \leq n$.

Comme θ est un vecteur de paramètres de tendance centrale mutidimensionnel et $X \in \mathbb{R}^n$, nous pouvons écrire :

$\frac{\partial}{\partial \theta_i} \log f(x, \theta) = \frac{\partial}{\partial \theta_i} \log f(x - \theta, 0)$. En posant $u_i = x_i - \theta_i$, pour $i = 1, 2, \dots, n$, nous

avons :

$$\begin{aligned}
 \frac{\partial}{\partial \theta_i} \log f(x, \theta) &= \frac{\partial(x_i - \theta_i)}{\partial \theta_i} \frac{\partial}{\partial u_i} \log f(x - \theta, 0) \\
 &= -\frac{\partial}{\partial u_i} \log f(x - \theta, 0) \\
 &= -\frac{\partial}{\partial x_i} \log f(x - \theta, 0) \\
 &= -\frac{\partial}{\partial x_i} \log f(x).
 \end{aligned}$$

Cette paramétrisation nous permet d'écrire dorénavant la matrice d'information sous la forme suivante :

$$\begin{aligned}
 \mathbf{I} &= \text{cov}\left[-\frac{\partial}{\partial x_i} \log f(x), -\frac{\partial}{\partial x_j} \log f(x)\right] \\
 &= \mathbb{E}\left[\frac{\partial}{\partial x_i} \log f(x) \frac{\partial}{\partial x_j} \log f(x)\right],
 \end{aligned}$$

pour $1 \leq i, j \leq n$.

Remarque 2.4 *Sous les hypothèses de régularité et en prenant en indice les dérivées partielles comme suit :*

$$f_i(x) = \frac{\partial}{\partial x_i} f(x),$$

nous avons :

$$[\log f(x)]_i = \frac{f_i(x)}{f(x)}.$$

En dérivant membre à membre l'égalité précédente par rapport à x_j , nous avons :

$$\begin{aligned}
 [[\log f(x)]_i]_j &= \left[\frac{f_i(x)}{f(x)}\right]_j \\
 &= \frac{f_{ij}(x)f(x) - f_i(x)f_j(x)}{f^2(x)} \\
 &= \frac{f_{ij}(x)}{f(x)} - [\log f(x)]_i [\log f(x)]_j.
 \end{aligned}$$

En remarquant que $\mathbb{E}\left[\frac{f_{ij}(x)}{f(x)}\right] = 0$, et en prenant l'espérance mathématique des deux membres de la dernière ligne des inégalités ci-dessus, on a :

$$\begin{aligned}\mathbb{E}\{[\log f(x)]_i[\log f(x)]_j\} &= \mathbb{E}\left(\frac{\partial^2 \log f(x)}{\partial x_i \partial x_j}\right) \\ &= -\mathbb{E}\{[\log f(x)]_i[\log f(x)]_j\} \\ &= -\mathbb{E}\left(\frac{\partial \log f(x)}{\partial x_i} \frac{\partial \log f(x)}{\partial x_j}\right).\end{aligned}$$

Donc :

$$\mathbb{E}\left(\frac{\partial \log f(x)}{\partial x_i} \frac{\partial \log f(x)}{\partial x_j}\right) = -\mathbb{E}\left(\frac{\partial^2 \log f(x)}{\partial x_i \partial x_j}\right),$$

pour $i, j = 1, 2, \dots, n$.

Ainsi, la proposition suivante nous donne une nouvelle expression de la matrice d'information plus appropriée pour les calculs.

Proposition 2.5 *Sous les hypothèses de régularité, la matrice d'information est donnée par :*

$$\mathbf{I} = (\mathbf{I}_{i,j}) = -\mathbb{E}\left[\frac{\partial^2}{\partial x_i \partial x_j} \log f(x)\right], \quad (2.2)$$

pour $i, j = 1, 2, \dots, n$.

2.5.4 Matrice d'information dans le cas de l'indépendance.

Lorsque les composantes X_i du vecteur aléatoire X sont indépendantes, alors $f(x) = \prod_{i=1}^n f_i(x_i)$. Dans ce cas la fonction de densité sera notée $f_0(x)$. Il s'ensuit que :

$$\frac{\partial^2}{\partial x_i \partial x_j} \log f(x) = 0,$$

pour $1 \leq i < j \leq n$, et pour $i = j$,

$$\frac{\partial^2}{\partial x_i \partial x_j} \log f(x) = \frac{d^2 \log f_i(x_i)}{dx_i^2}.$$

Ainsi, sous l'hypothèse d'indépendance,

$$\mathbb{E}\left(\frac{\partial^2}{\partial x_i \partial x_j} \log f(x)\right) = \delta_{i,j} \int_{-\infty}^{\infty} \left[\frac{d^2}{dx_i^2} \log f_i(x_i)\right] f_i(x_i) dx_i,$$

pour $i, j = 1, 2, \dots, n$. Ainsi, lorsque les composantes X_i du vecteur aléatoire X sont indépendantes, la matrice d'information de Fisher est diagonale et sous les hypothèses de régularité, ses éléments sont donnés par :

$$\mathbf{I}_{ij} = -\delta_{i,j} \int_{-\infty}^{\infty} \left[\frac{d^2}{dx_i^2} \log f_i(x_i)\right] f_i(x_i) dx_i$$

pour $i, j = 1, 2, \dots, n$. Dans ce cas, elle sera notée \mathbf{I}_0 .

2.5.5 Propriétés de la matrice d'information.

Nous énonçons ci-après les propriétés de la matrice d'information qui seront utiles par la suite. L'on pourra consulter entre autres Edouardo Mayer-Wolf [10], Papathanasiou [11] pour obtenir une étude plus détaillée de ces propriétés.

P₁ :) \mathbf{I} existe toujours (sous les hypothèses de régularité).

P₂ :) \mathbf{I} est non singulière.

P₃ :) \mathbf{I} est symétrique et semi-définie positive.

P₄ :) Soient $X = (X_1, X_2, \dots, X_n)$ un vecteur aléatoire continu, A une matrice $n \times n$ non-singulière et $b = (b_1, b_2, \dots, b_n)$ un vecteur $1 \times n$.

La matrice d'information associée au vecteur aléatoire $Y = XA + b$, notée \mathbf{J} , est

donnée par : $\mathbf{J} = A^{-1}\mathbf{I}(A^{-1})^t$.

Preuve :

En conservant les notations précédentes (de la propriété \mathbf{P}_4 :)) et en désignant par $f(x)$ la fonction de densité du vecteur aléatoire X . Rappelons l'expression suivante qui sera utile par la suite : $\nabla \log f(x, \theta) = (\frac{\partial}{\partial \theta_1} \log f(x, \theta), \dots, \frac{\partial}{\partial \theta_n} \log f(x, \theta))$. La fonction de densité du vecteur aléatoire Y est donnée par :

$g(y) = f[(y_1 - b_1, y_2 - b_2, \dots, y_n - b_n)A^{-1}] \det J_a$, où $\det J_a$ est le jacobien de la transformation inverse de $Y = XA + b$.

Nous en déduisons :

$$\begin{aligned} \nabla \log g(y) &= [(\nabla \log f(yA^{-1} - bA^{-1}))(A^{-1})^t] \\ &= [(\nabla \log f(x))(A^{-1})^t] \end{aligned}$$

La matrice d'information \mathbf{J} ($n \times n$) sur le vecteur aléatoire Y est donnée par :

$$\begin{aligned} \mathbf{J} &= \mathbb{E}[(\nabla \log g(y))^t(\nabla \log g(y))], \text{ par définition} \\ &= \mathbb{E}\{[A^{-1}[\nabla \log f(x)]^t[\nabla \log f(x)](A^{-1})^t]\} \\ &= A^{-1}\mathbb{E}[(\nabla \log f(x))^t(\nabla \log f(x))](A^{-1})^t \\ &= A^{-1}\mathbf{I}(A^{-1})^t, \end{aligned}$$

d'où le résultat.

2.5.6 Lien avec la matrice de variance-covariance.

En désignant par I_n la matrice identité, O_n la matrice nulle, \mathbf{I} la matrice d'information et Σ la matrice de variance-covariance du vecteur X , toutes de format $n \times n$, nous pouvons définir les matrices de format $2n \times 2n$, A et B suivantes :

$$A = \begin{pmatrix} I_n & -\mathbf{I}^{-1} \\ O_n & \mathbf{I}^{-1} \end{pmatrix};$$

et

$$B = \begin{pmatrix} \Sigma & I_n \\ I_n & \mathbf{I} \end{pmatrix}.$$

Ces deux matrices sont *SDP*, car toutes leurs sous-matrices principales le sont. Leur produit est donné par :

$$AB = \begin{pmatrix} \Sigma - \mathbf{I}^{-1} & O_n \\ \mathbf{I}^{-1} & I_n \end{pmatrix},$$

et est *SDP*, car A et B le sont. Ainsi, en vertu de la propriété \mathbf{M}_6 :) précédente, la matrice $\Sigma - \mathbf{I}^{-1}$ est *SDP*, car sous-matrice principale de AB . Le théorème qui suit établit le lien entre la matrice de variance-covariance Σ d'un vecteur aléatoire X et sa matrice d'information de Fisher \mathbf{I} .

Théorème 2.6 Soit $X = (X_1, X_2, \dots, X_n)$ un vecteur aléatoire continu admettant Σ comme matrice de variance-covariance et \mathbf{I} comme matrice d'information. Alors :

$$\Sigma \geq \mathbf{I}^{-1}.$$

Dans le cas de la loi normale multidimensionnelle $N(\mu, \Sigma)$, où μ est le paramètre de tendance centrale, nous avons :

$$\left[-\frac{\partial^2}{\partial x_i \partial x_j} \log f(x) \right] = \Sigma^{-1}, \text{ pour } i, j = 1, 2, \dots, n.$$

Il s'ensuit que :

$$\mathbf{I} = (\mathbf{I}_{ij}) = (\mathbb{E}[-\frac{\partial^2}{\partial x_i \partial x_j} \log f(x)]) = \Sigma^{-1} \int f(x) dx = \Sigma^{-1}.$$

2.6 Caractérisation de l'indépendance à l'aide de la matrice d'information.

Tout comme la matrice des corrélations résume la structure des liaisons linéaires entre les composantes d'un vecteur aléatoire, la matrice d'information synthétise la struc-

ture de variance-covariance entre les composantes du vecteur aléatoire $\nabla \log f(x) = (\frac{\partial}{\partial x_1} \log f(x), \dots, \frac{\partial}{\partial x_n} \log f(x))$.

Le théorème qui suit montre que lorsque les conditions de régularité sont vérifiées, la structure de dépendance des composantes du vecteur aléatoire $\nabla \log f(x)$ révèle celle de X .

Théorème 2.7 *Soit $X = (X_1, X_2, \dots, X_n)$ un vecteur aléatoire continu admettant un vecteur de paramètre de tendance centrale $\theta \in \mathbb{R}^n$ et pour fonctions de densité conjointe $f(x)$ et en cas d'indépendance $f_0(x)$. On suppose que $f(x)$ et $f_0(x)$ sont régulières et satisfont la positivité totale multidimensionnelle d'ordre deux.*

Alors, si la matrice d'information \mathbf{I} est diagonale, les composantes du vecteur X sont indépendantes deux à deux et $\mathbf{I} = \mathbf{I}_0$.

Preuve : Comme $f(x)$ vérifie la positivité totale multidimensionnelle d'ordre deux, on a $\frac{\partial^2}{\partial x_i \partial x_j} \log f(x_1, x_2, \dots, x_n) \geq 0$. La matrice d'information \mathbf{I} étant diagonale, on a :

$$\int_E [\frac{\partial^2}{\partial x_i \partial x_j} \log f(x)] f(x) dx = 0,$$

pour tout $1 \leq i < j \leq n$. On peut donc conclure que :

$$\frac{\partial^2}{\partial x_i \partial x_j} \log f(x) = 0,$$

ν -presque sûrement, $1 \leq i < j \leq n$ et pour tout $(X_1, X_2, \dots, X_n) \in \mathbb{R}^n$. Cette dernière relation est équivalente à l'indépendance deux à deux.

Remarquons aussi que la double égalité suivante :

$$\frac{\partial}{\partial x_i} \log f(x_1, x_2, \dots, x_n) = \frac{d}{dx_i} \log f_i(x_i) = \frac{d}{dx_i} \log \prod_{i=1}^n f_i(x_i)$$

est équivalente à $f(x) = \prod_{i=1}^n f_i(x_i)$, qui équivaut à son tour à l'égalité entre les matrices d'information : $\mathbf{I} = \mathbf{I}_0$.

En résumé :

l'indépendance $\iff \mathbf{I} = \mathbf{I}_0 \iff f(x) = \prod_{i=1}^n f_i(x_i)$.

L'égalité entre les matrices d'information $\mathbf{I} = \mathbf{I}_0$ telle qu'énoncée sous les hypothèses du théorème précédent induit l'indépendance entre les composantes du vecteur aléatoire X donnant lieu ainsi à une caractérisation de l'indépendance des composantes du vecteur aléatoire X dont la fonction de densité vérifie la positivité ou la négativité totales multidimensionnelles d'ordre 2. Toutefois, l'indépendance entre les composantes de X conduit à l'égalité $\mathbf{I} = \mathbf{I}_0$ sans d'autres conditions supplémentaires sur les fonctions de densité $f(x)$ et $f_0(x)$.

CHAPITRE 3

Mesure de dépendance.

Au chapitre 1, nous avons décrit les principales formes de dépendance entre les composantes d'un vecteur aléatoire. Nous avons obtenu notamment les conditions de dépendance positive et de dépendance négative. Le chapitre 2 a offert une caractérisation de l'indépendance [entre ces variables] basée sur la matrice d'information. Lorsqu'une forme de dépendance est révélée, il est important de mesurer sa force. Dans ce chapitre, nous désirons quantifier la force de la liaison entre les composantes d'un vecteur aléatoire. En règle générale, une mesure de liaison ou de dépendance entre les composantes d'un vecteur aléatoire est un nombre réel qui quantifie celle-ci; plus ce nombre est élevé plus la liaison est forte et réciproquement. Il existe plusieurs mesures pour déterminer le niveau de liaison entre les variables : la divergence, la ϕ -divergence, l'entropie relative, le coefficient de corrélation de Pearson, le coefficient de Spearman et bien d'autres. Cependant, le choix d'une mesure de dépendance peut dépendre des facteurs tels que la facilité de son calcul et de son interprétation de même que du type des variables à l'étude. Les conditions énoncées ci-après et appelées axiomes de Rényi [12] forment un ensemble de propriétés que toute mesure de dépendance doit idéalement satisfaire. Elles ont été enrichies par

Schweizer et B., Wolf [13].

3.1 Axiomes de Rényi.

Soit $R(X_1, X_2, \dots, X_n)$ une mesure de dépendance entre les variables aléatoires continues X_i , pour $i = 1, 2, \dots, n$. Alors $R(X_1, X_2, \dots, X_n)$ doit satisfaire les axiomes ci-après.

- a :) La quantité $R(X_1, X_2, \dots, X_n)$ existe pour toutes variables aléatoires X_1, X_2, \dots, X_n ;
- b :) La mesure $R(X_1, X_2, \dots, X_n)$ est invariante par permutation des variables $X_i, i = 1, 2, \dots, n$; c'est-à-dire que si $(\sigma(1), \sigma(2), \dots, \sigma(n))$ est une permutation de l'ensemble $\{1, 2, \dots, n\}$, alors : $R(X_{\sigma(1)}, X_{\sigma(2)}, \dots, X_{\sigma(n)}) = R(X_1, X_2, \dots, X_n)$;
- c :) $0 \leq R(X_1, X_2, \dots, X_n) \leq 1$;
- d :) Si les variables sont indépendantes alors $R(X_1, X_2, \dots, X_n) = 0$;
- e :) Dans le cas où $n = 2$ et que l'une des variables est une fonction strictement croissante de l'autre, la quantité $R(X_1, X_2) = 1$;
- f :) Si les fonctions f_i , pour $i = 1, \dots, n$ sont numériques et identiquement monotones sur les ensembles des valeurs des variables X_i , pour $i = 1, \dots, n$ respectivement, alors l'égalité suivante est vérifiée : $R(f_1(X_1), f_2(X_2), \dots, f_n(X_n)) = R(X_1, X_2, \dots, X_n)$;
- g :) Lorsque $n = 2$ et que le couple (X_1, X_2) est binormal et admet pour coefficient de corrélation r alors la quantité $R(X_1, X_2)$ est une fonction croissante de $|r|$;

Remarquons que dans les axiomes de Rényi originels, l'axiome e :) exigeait que l'une des deux variables soit une fonction \mathcal{B} -mesurable de l'autre, pour l'axiome g :), l'égalité

$R(X_1, X_2) = |r|$ devait être satisfaite tandis que $\mathbf{h} :)$ n'était pas inclus. En outre, plusieurs mesures de dépendance proposées dans la littérature appartiennent à l'intervalle $[0, \infty]$. Mais il est toujours possible de les ramener par une transformation appropriée dans l'intervalle $[0, 1]$. C'est le cas de l'entropie, de la ϕ -divergence ou de la mesure de Zografos comme nous le verrons plus tard dans ce chapitre. Une mesure ne saurait être considérée comme mesure de dépendance si elle ne satisfait pas l'axiome $\mathbf{d} :)$. L'axiome $\mathbf{c} :)$ porte essentiellement sur l'interprétabilité de la mesure de dépendance concernée. Notons enfin qu'un moyen efficace de construire une mesure de dépendance satisfaisant à l'axiome $\mathbf{d} :)$ consiste à la définir en termes de semi-distance entre la distribution conjointe et la distribution représentant l'indépendance.

Le vaste champ d'action de ce chapitre couvre la mesure de dépendance de K. Zografos, l'étude de ses propriétés et des exemples ainsi que sa version empirique. D'autres mesures multidimensionnelles sont aussi introduites.

En 2000, K. Zografos a proposé une mesure de dépendance multidimensionnelle qui, plutôt que de se baser sur la matrice de variance-covariance comme la plupart des mesures d'association rencontrées, exploite d'une façon particulière l'écart entre les matrices d'information \mathbf{I} et \mathbf{I}_0 telles que définies par les relations (2.3) et (2.6) du chapitre précédent. Nous présentons cette mesure dans la définition qui suit.

3.2 Mesure de dépendance de Zografos.

Définition 3.1 Soient $X = (X_1, X_2, \dots, X_n)$ un vecteur aléatoire admettant $f(x)$ pour fonction de densité et $h : [0, \infty) \rightarrow [0, \infty)$ une fonction convexe vérifiant :

$$\mathbf{i}_1 :) \quad h(1) = h'(1) = 0;$$

\mathbf{i}_2 :) $h''(1) > 0$.

En désignant par $\lambda_1, \dots, \lambda_n$ les valeurs propres de la matrice $\mathbf{I}^{-1}\mathbf{I}_0$, une mesure de dépendance multidimensionnelle entre les composantes du vecteur X est donnée par :

$$\delta_{h,f}(X) = \delta_{h,f}(X_1, X_2, \dots, X_n) = \sum_{i=1}^n h(\lambda_i).$$

La quantité $\delta_{h,f}(X)$ est la mesure multidimensionnelle proposée par K. Zografos [1]. Avant d'étudier les propriétés de cette mesure, il est important de donner sa réelle motivation. Rappelons que l'égalité $\mathbf{I} = \mathbf{I}_0$ entre les matrices d'information de Fisher indiquait une absence de corrélation entre les composantes du vecteur aléatoire correspondant dont la densité vérifiait la positivité ou la négativité multidimensionnelles d'ordre deux; ce qui amène à penser que la force de la liaison entre ces variables peut être approximée par la valeur $\delta_{h,f}(X)$ exprimant l'écart entre les matrices \mathbf{I}^{-1} et \mathbf{I}_0 . Notons que la quantité $\delta_{h,f}(X)$ est finie et que l'indépendance entre les variables est d'autant réalisée que ses valeurs sont voisines de 0 tandis que l'extrême dépendance est satisfaite pour ses grandes valeurs comme le montrent les deux sections suivantes.

3.2.1 Borne inférieure de la mesure de Zografos.

Notons tout d'abord que lorsque les variables sont indépendantes, la mesure $\delta_{h,f}(X)$ vaut zéro. Bien que condition nécessaire à l'indépendance des variables, la condition $\delta_{h,f}(X) = 0$ n'en constitue pas une condition suffisante. La positivité ou la négativité totales multidimensionnelles d'ordre deux sont des conditions supplémentaires que doit satisfaire la fonction de densité $f(x)$ afin qu'il en soit ainsi. Nous le montrerons plus tard dans ce chapitre. En outre, nous savons que l'indépendance entraîne l'égalité $\mathbf{I} = \mathbf{I}_0$. Il s'ensuit que $\mathbf{I}^{-1}\mathbf{I}_0 = I_n$ où I_n désigne la matrice identité $n \times n$. La mesure $\delta_{h,f}(X)$ dans

ce cas vaut :

$$\delta_{h,f}(X) = \sum_{i=1}^n h(\lambda_i) = nh(1) = 0, \text{ car } \lambda_i = 1 \text{ pour tout } i = 1, 2, \dots, n.$$

La condition \mathbf{i}_1 :) précédente est donc une exigence permettant de caractériser l'indépendance par la valeur nulle qui représente la borne inférieure de la mesure $\delta_{h,f}(X)$.

3.2.2 Majorants de la mesure de Zografos.

Lorsque la matrice $\mathbf{I} - \mathbf{I}_0$ est *SDP* ce qui n'est pas en général le cas comme nous le montrerons plus loin, la famille de mesures $\delta_{h,f}(X)$ admet le nombre $nh(0)$ comme un majorant qui est une valeur atteinte en cas d'extrême dépendance entre les variables. En effet, si la matrice $\mathbf{I} - \mathbf{I}_0$ est *SDP*, alors la matrice $I_n - \mathbf{I}^{-1}\mathbf{I}_0$ l'est aussi. Ainsi, les valeurs propres de la matrice $\mathbf{I}^{-1}\mathbf{I}_0$ sont $1 - \lambda'_i$ où les $\lambda'_i, i = 1, 2, \dots, n$ sont les valeurs propres de $I_n - \mathbf{I}^{-1}\mathbf{I}_0$. En outre, nous avons l'inégalité $0 < \lambda'_i < 1$, pour tout $i = 1, 2, \dots, n$, car la matrice $\mathbf{I}^{-1}\mathbf{I}_0$ est *DP*. Nous en déduisons que les valeurs propres de $\mathbf{I}^{-1}\mathbf{I}_0$ appartiennent à l'intervalle $[0, 1]$. Par définition, la fonction h vérifie $h(1) = h'(1) = 0$ et admet un minimum en $x = 1$ tel que exprimé par \mathbf{i}_2 :) à la page 40; elle doit donc être décroissante sur l'intervalle $(0, 1]$, car sinon elle prendrait des valeurs négatives. Désignant par $\lambda_1, \lambda_2, \dots, \lambda_n$ ces valeurs propres de la matrice $\mathbf{I}^{-1}\mathbf{I}_0$ et en posant $h(0) = \lim_{\epsilon \rightarrow 0^+} h(\epsilon)$, ϵ étant positif et aussi petit que l'on veut, il vient : $h(\lambda_i) \leq h(0)$, pour tout $i = 1, 2, \dots, n$. D'où :

$$\delta_{h,f}(X) = \sum_{i=1}^n h(\lambda_i) \leq nh(0).$$

Dans le cas où $\mathbf{I} - \mathbf{I}_0$ n'est pas *SDP*, un autre majorant de la mesure de dépendance $\delta_{h,f}(X)$ peut être obtenu à l'aide de la proposition 3.6, P.59 de Vадja [14], qui stipule que toute fonction convexe

$f : [0, \infty) \rightarrow [0, \infty)$ vérifiant les propriétés suivantes :

- $\lim_{x \rightarrow 0^+} \frac{f(x)}{x} = \alpha$ existe,

- $\lim_{x \rightarrow \infty} \frac{f(x)}{x} = \beta$ existe,

satisfait :

$$f(x) \leq f(0) + x\beta.$$

Nous en déduisons pour toutes valeurs propres λ_i de la matrice $\mathbf{I}^{-1}\mathbf{I}_0$, l'inégalité suivante :

$$h(\lambda_i) \leq h(0) + \lambda_i \lim_{x \rightarrow \infty} \frac{h(x)}{x},$$

pour $0 \leq x < \infty$.

D'où :

$$\delta_{h,f}(X) \leq nh(0) + \sum_{i=1}^n \lambda_i \lim_{x \rightarrow \infty} \frac{h(x)}{x},$$

ou encore :

$$\delta_{h,f}(X) \leq nh(0) + \beta \text{tr}(\mathbf{I}^{-1}\mathbf{I}_0).$$

Dans certains cas d'extrême dépendance et selon que $\mathbf{I} - \mathbf{I}_0$ est *SDP* ou non , les bornes $nh(0)$ et $nh(0) + \beta \text{tr}(\mathbf{I}^{-1}\mathbf{I}_0)$ peuvent être atteintes.

3.2.3 Propriétés de la mesure de Zografos.

P₁ :) La quantité $\delta_{h,f}(X)$ est définie pour tout vecteur aléatoire continu.

Preuve :

L'existence de $\delta_{h,f}(X)$ est due à l'existence des matrices d'information définies-positives \mathbf{I} et \mathbf{I}_0 données par les définitions (2.3) et (2.6) du chapitre précédent.

P₂ :) La mesure $\delta_{h,f}(X)$ est invariante par permutation des variables $X_i, i = 1, 2, \dots, n$: c'est-à-dire que si $(\sigma(1), \sigma(2), \dots, \sigma(n))$ est une permutation de l'ensemble $\{1, 2, \dots, n\}$, alors :

$$\delta_{h,f}(X_{\sigma(1)}, X_{\sigma(2)}, \dots, X_{\sigma(n)}) = \delta_{h,f}(X).$$

Preuve :

Désignons par $v(i)$, la i^{eme} colonne de la matrice identité I_n et posons $P = [v(1), v(2), \dots, v(n)]$. Alors P est une matrice non-singulière et il est facile de vérifier que P est une matrice orthogonale c'est-à-dire $P^{-1} = P^t$. Toute permutation $Y = (X_{\sigma(1)}, X_{\sigma(2)}, \dots, X_{\sigma(n)})$ des variables X_i , pour $i = 1, 2, \dots, n$, est alors obtenue à l'aide d'une transformation de la forme $Y = XP$. Si f^* est la densité conjointe de Y et f_0^* sa densité en cas d'indépendance, nous savons, en vertu de la propriété **P₄** :) de la page 33, que les matrices d'information respectives associées \mathbf{I}^* et \mathbf{I}_0^* sont données par :

$\mathbf{I}^* = P^t \mathbf{I} P$ et $\mathbf{I}_0^* = P^t \mathbf{I}_0 P$. Nous en déduisons alors que :

$$\begin{aligned} \mathbf{I}^{*-1} \mathbf{I}_0^* &= P^{-1} \mathbf{I}^{-1} (P^t)^{-1} P^t \mathbf{I}_0 P \\ &= P^{-1} \mathbf{I}^{-1} \mathbf{I}_0 P. \end{aligned}$$

Les matrices $\mathbf{I}^{*-1} \mathbf{I}_0^*$ et $\mathbf{I}^{-1} \mathbf{I}_0$ sont semblables et admettent donc des valeurs propres identiques.

La matrice $\mathbf{I}^{-1} \mathbf{I}_0$ est DP , donc ses valeurs propres λ_i , $i = 1, 2, \dots, n$ sont strictement positives, d'où la propriété suivante :

P₃ :) $0 \leq \delta_{h,f}(X) < \infty$.

P₄ :) Lorsque les composantes du vecteur X sont indépendantes alors $\delta_{h,f}(X) = 0$.

Preuve :

Lorsque les variables composantes sont indépendantes, alors $f(x) = f_0(x)$ et la matrice $\mathbf{I}^{-1} \mathbf{I}_0 = I_n$ où I_n désigne la matrice identité $n \times n$, admet 1 pour valeur propre multiple d'ordre n . Comme $h(1) = 0$ selon **i₁** :), nous avons :

$$\delta_{h,f}(X) = \delta_{h,f}(X_1, X_2, \dots, X_n) = \sum_{i=1}^n h(\lambda_i) = 0.$$

D'où le résultat.

P₅ :) Dans le cas où la fonction de densité $f(x)$ vérifie la positivité ou la négativité totales multidimensionnelles d'ordre deux et si de plus, $f(x)$ et $f_0(x)$ sont régulières, alors $\delta_{h,f}(X) = 0$ entraîne l'indépendance deux à deux entre les composantes du vecteur X .

Preuve :

Rappelons (voir page 36) que si l'égalité $\mathbf{I} = \mathbf{I}_0$ est vérifiée alors :

$$\frac{\partial}{\partial x_i} \log f(x_1, x_2, \dots, x_n) = \frac{d}{dx_i} \log f_i(x_i),$$

pour tout $i = 1, 2, \dots, n$ et $(X_1, X_2, \dots, X_n) \in \mathbb{R}^n$. Ainsi, s'il existe un indice i_0 tel que $\frac{\partial}{\partial x_{i_0}} \log f(x) \neq \frac{d}{dx_{i_0}} \log f_0(x_{i_0})$, alors la matrice $\mathbf{I}^{-1}\mathbf{I}_0 - I_n \neq O$ où O désigne la matrice nulle $n \times n$. Donc il existe au moins une valeur propre de $\mathbf{I}^{-1}\mathbf{I}_0$ qui est différente de 1. D'où $\delta_{h,f}(X) \neq 0$.

P₆ :) Soient A une matrice $n \times n$ non singulière et b un vecteur réel $n \times 1$. Désignons par $f(x)$ et $f^*(x)$ les fonctions de densité respectives des vecteurs X et $Y = AX + b$. Alors :

$$\begin{aligned} \delta_{h,f^*}(Y) &= \delta_{h,f^*}(AX + b) \\ &= \delta_{h,f}(X). \end{aligned}$$

Autrement dit, la mesure $\delta_{h,f}(X)$ est invariante par transformation linéaire.

Preuve :

Notons par \mathbf{I} et \mathbf{J} les matrices d'information respectives des vecteurs X et $Y = AX + b$. Selon la propriété **P₄ :**) du chapitre précédent, nous pouvons écrire :

$\mathbf{J} = (A^{-1})^t \mathbf{I} A^{-1}$ et $\mathbf{J}_0 = (A^{-1})^t \mathbf{I}_0 A^{-1}$. D'où :

$$\begin{aligned} \mathbf{J}^{-1} \mathbf{J}_0 &= A \mathbf{I}^{-1} [(A^{-1})^t]^{-1} (A^{-1})^t \mathbf{I}_0 A^{-1} \\ &= A \mathbf{I}^{-1} \mathbf{I}_0 A^{-1}. \end{aligned}$$

Ainsi, les matrices $\mathbf{J}^{-1}\mathbf{J}_0$ et $\mathbf{I}^{-1}\mathbf{I}_0$ sont semblables et donc admettent les mêmes valeurs propres, d'où le résultat.

Le tableau ci-après résume les axiomes de Rényi satisfaits par cette mesure de dépendance multidimensionnelle :

Axiomes de Rényi :							
a	b	c	d	e	f	g	h
	•		•			•	

Tableau 3.1: Axiomes de Rényi pour la mesure de Zografos.

3.2.4 La mesure de Zografos, la ϕ -divergence et l'entropie relative.

Soient $(\mathbb{R}^n, \mathcal{B}_i, \nu_i)$, $i = 1, 2$ deux espaces probabilisés tels que les mesures ν_i , $i = 1, 2$ soient absolument continues par rapport à une mesure de référence λ . On note $\nu_i \ll \lambda$. Nous savons d'après le théorème de Radon-Nykodym que :

$\nu_i \ll \lambda \implies \exists f_i$ appelée densité de probabilité généralisée telle que $f_i = \frac{d\nu_i}{d\lambda}$, $i = 1, 2$.

Nous prenons λ égale à la mesure de Lebesgue sur \mathcal{B}_i .

Soit $\phi : [0, \infty) \rightarrow [0, \infty)$ une fonction convexe. On appelle ϕ -divergence entre deux fonctions de densité f_1 et f_2 définies sur \mathbb{R}^n la quantité définie par :

$$D_\phi(f_1, f_2) = \int f_1 \phi\left(\frac{f_2}{f_1}\right) dx$$

Cette quantité $D_\phi(f_1, f_2)$ est une mesure d'écart entre les densités f_1 et f_2 . Notons que cette expression n'est pas symétrique par rapport à f_1 et f_2 . Pour $\phi(w) = w - 1 - \log w$,

nous avons :

$$\begin{aligned} D_\phi(f_1, f_2) &= \int f_1 \left[-\log\left(\frac{f_2}{f_1}\right) + \frac{f_2}{f_1} - 1 \right] dx \\ &= \int f_1 \log\left(\frac{f_1}{f_2}\right) dx = \mathcal{E}(f_1, f_2). \end{aligned}$$

On appelle entropie relative entre deux fonctions de densité f_1 et f_2 définies sur \mathbb{R}^n la quantité donnée par :

$$\mathcal{E}(f_1, f_2) = \int f_1 \log\left(\frac{f_1}{f_2}\right) dx$$

La quantité $\mathcal{E}(\cdot, \cdot)$ représente le niveau d'incertitude à discriminer entre les densités f_1 et f_2 . Notons qu'ainsi définie, cette quantité n'est pas symétrique en f_1 et f_2 .

Dans les deux prochaines sections, nous présentons la mesure de Zografos, la ϕ -divergence et l'entropie relative dans deux cas importants : gaussien et discret.

Cas gaussien.

Considérons l'ensemble des distributions centrées et gaussiennes sur \mathbb{R}^n définies par :

$\mathcal{F} = \{f_\Sigma(x) : f_\Sigma(x) = (2\pi)^{-n/2} \det \Sigma^{-1/2} \exp[-1/2x^t \Sigma^{-1}x], \text{ où la matrice } \Sigma \text{ est } DP \text{ et } x \in \mathbb{R}^n\}$. Si f_{Σ_1} et f_{Σ_2} sont deux fonctions de densité appartenant à \mathcal{F} , alors, pour $\phi(\omega) = \omega - 1 - \log(\omega)$, leur ϕ -divergence est donnée par :

$$D_\phi(f_{\Sigma_1}, f_{\Sigma_2}) = \int f_{\Sigma_1} \log\left(\frac{f_{\Sigma_1}}{f_{\Sigma_2}}\right) dx.$$

Dans le cas où $\Sigma_2 = \Sigma_0$, Σ_0 représentant la matrice de variance-covariance du vecteur aléatoire X en cas d'indépendance, la quantité $D_\phi(f_{\Sigma_1}, f_{\Sigma_0})$ est égale à l'entropie relative.

En désignant par $\lambda_i, i = 1, 2, \dots, n$ les valeurs propres de la matrice $\mathbf{I}^{-1}\mathbf{I}_0$ et par h la fonction définie par $h(\omega) = \phi(\omega)$, les conditions de convexité sur la fonction h sont satisfaites et la mesure de Zografos est :

$$\begin{aligned}
\delta_{h,f}(X) &= \sum_{i=1}^n \phi(\lambda_i) \\
&= -\sum_{i=1}^n \log \lambda_i + \sum_{i=1}^n \lambda_i - n = -\log \prod_{i=1}^n \lambda_i + \sum_{i=1}^n \lambda_i - n.
\end{aligned}$$

Comme $\det(\mathbf{I}^{-1}\mathbf{I}_0) = \frac{\det \mathbf{I}_0}{\det \mathbf{I}} = \prod_{i=1}^n \lambda_i$ et que $tr(\mathbf{I}^{-1}\mathbf{I}_0) = \sum_{i=1}^n \lambda_i$, nous obtenons :

$$\delta_{h,f}(X) = -\log \frac{\det \mathbf{I}_0}{\det \mathbf{I}} + tr(\mathbf{I}^{-1}\mathbf{I}_0) - n.$$

En vertu du théorème (2.6) du chapitre 2, nous avons dans ce cas : $\mathbf{I} = \Sigma^{-1}$ et $\mathbf{I}_0 = \Sigma_0^{-1}$.

Donc : $\mathbf{I}^{-1}\mathbf{I}_0 = \Sigma\Sigma_0^{-1}$.

Ainsi :

$$\delta_{h,f}(X) = -\log \frac{\det \Sigma}{\det \Sigma_0} + tr(\Sigma\Sigma_0^{-1}) - n.$$

Remarque 3.2 *Etant donné un vecteur X de matrice de variance-covariance Σ et de vecteur moyen $\mathbb{E}(X) = \mu$ et une matrice Q de format $n \times n$, l'espérance mathématique de la forme quadratique $X^t Q X$ est donnée par : $tr(Q\Sigma) + \mu^t Q \mu$.*

En effet :

$$\mathbb{E}(X^t Q X) = \mathbb{E}(tr(X^t Q X)) = \mathbb{E}(tr(Q X X^t)) = tr(Q \mathbb{E}(X X^t)).$$

Or, $\mathbb{E}(X X^t) = \Sigma + \mathbb{E}(X)\mathbb{E}(X^t)$, d'où :

$$\begin{aligned}
\mathbb{E}(X^t Q X) &= tr[Q(\Sigma + \mathbb{E}(X)\mathbb{E}(X^t))] \\
&= tr[Q(\Sigma + \mu\mu^t)] = tr(Q\Sigma) + tr(Q\mu\mu^t) = tr(Q\Sigma) + \mu^t Q \mu.
\end{aligned}$$

Par ailleurs, on peut faire le calcul explicite de $D_\phi(f_{\Sigma_1}, f_{\Sigma_0}) = \mathcal{E}(f_{\Sigma_1}, f_{\Sigma_0})$, de la façon suivante :

$$\frac{f_{\Sigma_1}}{f_{\Sigma_0}} = \det\left(\frac{\Sigma_0}{\Sigma_1}\right)^{\frac{1}{2}} \exp\left[-\frac{1}{2}x^t \Sigma_1^{-1}x + \frac{1}{2}x^t \Sigma_0^{-1}x\right],$$

d'où :

$$\log\left[\frac{f_{\Sigma_1}}{f_{\Sigma_0}}\right] = \log\left(\det\left(\frac{\Sigma_0}{\Sigma_1}\right)^{\frac{1}{2}}\right) - \frac{1}{2}x^t \Sigma_1^{-1}x + \frac{1}{2}x^t \Sigma_0^{-1}x.$$

Ainsi,

$$\begin{aligned} \mathcal{E}(f_{\Sigma_1}, f_{\Sigma_0}) &= \int f_{\Sigma_1} \left[\log\left(\frac{f_{\Sigma_1}}{f_{\Sigma_0}}\right)\right] dx \\ &= \log\left[\det\left(\frac{\Sigma_0}{\Sigma_1}\right)^{\frac{1}{2}}\right] \int f_{\Sigma_1} dx - \int \frac{1}{2}x^t \Sigma_1^{-1}x f_{\Sigma_1} dx + \int \frac{1}{2}x^t \Sigma_0^{-1}x f_{\Sigma_1} dx. \end{aligned}$$

D'après la remarque précédente, la quantité $\int \frac{1}{2}x^t \Sigma_1^{-1}x f_{\Sigma_1} dx$ représente l'espérance mathématique de la forme quadratique $\frac{1}{2}x^t \Sigma_1^{-1}x$ par rapport à la densité f_{Σ_1} et s'écrit alors :

$$\int \frac{1}{2}x^t \Sigma_1^{-1}x f_{\Sigma_1} dx = \text{tr}(\Sigma_1^{-1} \Sigma_1) = \text{tr}(I_n) = n.$$

De même, en vertu du fait que la diagonale principale de la matrice $\Sigma_0^{-1} \Sigma_1$ n'a que des 1, il vient :

$$\int \frac{1}{2}x^t \Sigma_0^{-1}x f_{\Sigma_1} dx = \text{tr}(\Sigma_0^{-1} \Sigma_1) = n.$$

Enfin,

$$\begin{aligned} \delta_{h,f}(X) &= 2D_\phi(f_{\Sigma_1}, f_{\Sigma_0}) \\ &= 2\mathcal{E}(f_{\Sigma_1}, f_{\Sigma_0}) \\ &= \log\left(\det\left(\frac{\Sigma_0}{\Sigma_1}\right)\right). \end{aligned}$$

Cas Discret.

Soient X et Y deux vecteurs aléatoires discrets de densités de probabilité \mathbb{P} et \mathbb{Q} données
données respectivement par :

$$\mathbb{P} = \left(\frac{\lambda_1}{\sum_{i=1}^n \lambda_i}, \frac{\lambda_2}{\sum_{i=1}^n \lambda_i}, \dots, \frac{\lambda_n}{\sum_{i=1}^n \lambda_i} \right).$$

et par :

$$\mathbb{Q} = \frac{1}{n} \sum_{i=1}^n \delta_i,$$

où δ_i désigne la masse de Dirac. Soit h définissant $\delta_{h,f}(X)$. Considérons la fonction
convexe $\varphi(\cdot)$ définie par :

$$\varphi(\omega) = h \left(\frac{\omega}{n} \sum_{i=1}^n \lambda_i \right).$$

La φ -divergence entre les mesures de probabilité \mathbb{P} et \mathbb{Q} s'exprime alors sous la forme :

$$\begin{aligned} D_\varphi(\mathbb{P}, \mathbb{Q}) &= \sum_{i=1}^n \varphi \left(\frac{d\mathbb{P}}{d\mathbb{Q}} \right) d\mathbb{Q} \\ &= \sum_{i=1}^n h \left(\frac{n\lambda_i}{\sum_{i=1}^n \lambda_i} \frac{\sum_{i=1}^n \lambda_i}{n} \right) \frac{1}{n} \\ &= \frac{1}{n} \sum_{i=1}^n h(\lambda_i) \end{aligned}$$

et donc $\delta_{h,f}(X) = nD_\varphi(\mathbb{P}, \mathbb{Q})$.

3.2.5 Étude de quelques exemples.

3.2.6 Exemple 1 : Loi binormale.

Soit $X = (X_1, X_2)$ un vecteur normal bidimensionnel de matrice de variance-covariance donnée par :

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

et

$$\Sigma_0 = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

en cas d'indépendance. Dans ce cas, nous savons que : $\mathbf{I} = \Sigma^{-1}$ et $\mathbf{I}_0 = \Sigma_0^{-1}$.

D'où :

$$\begin{aligned} \mathbf{I}^{-1}\mathbf{I}_0 &= \Sigma\Sigma_0^{-1} \\ &= \begin{pmatrix} 1 & \rho\frac{\sigma_1}{\sigma_2} \\ \rho\frac{\sigma_2}{\sigma_1} & 1 \end{pmatrix}. \end{aligned}$$

Le polynôme caractéristique de $\mathbf{I}^{-1}\mathbf{I}_0$, noté $p(\lambda)$, est donné par : $p(\lambda) = \det(\mathbf{I}^{-1}\mathbf{I}_0 - \lambda I_2) = (1 - \lambda)^2 - \rho^2$. Ainsi, les valeurs propres de la matrice $\mathbf{I}^{-1}\mathbf{I}_0$ sont $1 - \rho$ et $1 + \rho$. La mesure de dépendance s'écrit alors :

$$\delta_{h,f}(X_1, X_2) = h(1 - \rho) + h(1 + \rho).$$

La quantité $\delta_{h,f}(X_1, X_2)$ est une fonction du coefficient de corrélation ρ . Pour $h(w) = \frac{(w-1)^2}{2}$, nous avons $\delta_{h,f}(X_1, X_2) = \rho^2$ et pour $h(w) = -1 + w - \log w$, nous obtenons $\delta_{h,f}(X_1, X_2) = -\log(1 - \rho^2)$.

3.2.7 Exemple 2 : Loi multinormale avec matrice d'équicorrélation.

Soit $X = (X_1, X_2, \dots, X_n)$ un vecteur n -dimensionnel suivant une loi multinormale $N(\mu, \Sigma)$ telle que : $\Sigma = (\sigma_{ij})$, $Var(X_i) = \sigma_{ii}^2$ et $\sigma_{ij} = \rho\sigma_i\sigma_j$, pour $i \neq j$ et $\Sigma_0 = (\sigma_i^2)$,

$i, j = 1, 2, \dots, n$ est la matrice de variance-covariance en cas d'indépendance. Dans ce cas, la matrice de corrélation est donnée par :

$$\mathbf{R} = \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \vdots \\ \vdots & \vdots & \ddots & \rho \\ \rho & \dots & \rho & 1 \end{pmatrix}$$

et peut s'écrire :

$$\mathbf{R} = (1 - \rho)I_n + \rho J_n$$

où I_n et J_n sont respectivement des matrices identité et unité $n \times n$. Rappelons que la matrice unité est celle dont toutes les colonnes sont formées de 1. Le polynôme caractéristique de la matrice \mathbf{R} , noté $P(\lambda)$, est alors donné par :

$$P(\lambda) = \det(\mathbf{R} - \lambda I_n) = \det[(1 - \rho - \lambda)I_n + \rho J_n].$$

Comme la matrice J_n est de rang 1, elle admet une seule valeur propre $n = \text{tr}(J_n)$.

Ainsi,

$$P(1 + (n - 1)\rho) = \rho^n \det(J_n - nI_n) = 0.$$

Nous avons aussi que :

$P(1 - \rho) = \rho^n \det J_n = 0$. Ainsi, la valeur $1 + (n - 1)\rho$ est une valeur propre simple de \mathbf{R} tandis que $1 - \rho$ en est une valeur propre multiple d'ordre $n - 1$. Nous en déduisons que :

$$\det \mathbf{R} = (1 - \rho)^{n-1} (1 + (n - 1)\rho).$$

En remarquant que :

$$J_n^2 = nJ_n \text{ et que :}$$

$$\mathbf{R} \left[\frac{1}{1-\rho} I_n - \frac{\rho}{(1-\rho)(1+(n-1)\rho)} J_n \right] = I_n,$$

nous obtenons l'inverse de \mathbf{R} donnée par la relation :

$$\mathbf{R}^{-1} = \frac{1}{1-\rho} I_n - \frac{\rho}{(1-\rho)(1+(n-1)\rho)} J_n.$$

Or,

$$\mathbf{R} = D_{\frac{1}{s}} \Sigma D_{\frac{1}{s}}$$

où

$$D_{\frac{1}{s}} = \text{diag}(1/\sigma_1, 1/\sigma_2, \dots, 1/\sigma_n), \text{ d'où :}$$

$$\Sigma^{-1} = D_{\frac{1}{s}} \mathbf{R}^{-1} D_{\frac{1}{s}}.$$

Lorsque les inégalités suivantes sont satisfaites $-\frac{1}{n-1} < \rho < 1$ et

$$\mathbf{R} = (1-\rho)I_n + \rho J_n,$$

\mathbf{R} est dite matrice d'équicorrélation. Nous savons que : $\mathbf{I} = \Sigma^{-1}$ et $\mathbf{I}_0 = \Sigma_0^{-1}$, donc

$$\Sigma_0^{-1} = D_{\frac{1}{s}} D_{\frac{1}{s}} = \text{diag}(1/\sigma_1^2, 1/\sigma_2^2, \dots, 1/\sigma_n^2).$$

D'après ce qui précède, nous obtenons:

$$\mathbf{I}^{-1} \mathbf{I}_0 = \Sigma \Sigma_0^{-1} = D_s \mathbf{R} D_s^{-1} \text{ où}$$

$D_s = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$, ce qui montre que les matrices $\mathbf{I}^{-1} \mathbf{I}_0$ et \mathbf{R} ont exactement les mêmes valeurs propres $1-\rho$ d'ordre $n-1$ et $1+(n-1)\rho$ d'ordre 1. Par définition, la

mesure de dépendance est alors donnée par :

$$\delta_{h,f}(X) = (n-1)h(1-\rho) + h(1+(n-1)\rho)$$

et pour $h(w) = \frac{(w-1)^2}{2}$, nous avons :

$$\delta_{h,f}(X) = \frac{n(n-1)\rho^2}{2}.$$

Lorsque les variances des variables sont égales, c'est-à-dire :

$\text{var}(X_i) = \sigma^2, i = 1, 2, \dots, n$, nous avons le modèle multinormal intraclasse. Dans ce cas, $\sigma_{ij} = \rho\sigma^2, i, j = 1, 2, \dots, n$. En supposant que $-\frac{1}{n-1} < \rho < 1$, nous obtenons comme précédemment que les valeurs $1-\rho$ et $1+(n-1)\rho$ sont des valeurs propres de la matrice $\mathbf{I}^{-1}\mathbf{I}_0$ d'ordre de multiplicité $n-1$ et 1 respectivement. Pour le choix de la fonction $h(w) = -1 + w - \log w$, la mesure de dépendance $\delta_{h,f}(X)$ est donnée par :

$$\begin{aligned} \delta_{h,f}(X) &= \sum_{i=1}^{n-1} [-\log(1-\rho) + (1-\rho) - 1] - \log(1+(n-1)\rho) + 1 + (n-1)\rho - 1 \\ &= -(n-1)\log(1-\rho) - \log(1+(n-1)\rho) \\ &= -\log[(1-\rho)^{(n-1)}(1+(n-1)\rho)]. \end{aligned}$$

Pour $n = 2$, nous trouvons :

$$\delta_{h,f}(X) = -\log(1-\rho^2).$$

3.2.8 Exemple 3 : Modèle multinormal interclasse.

Considérons le vecteur aléatoire $T = (X, Y)$ suivant une loi multinormale $N_{n+1}(\mu_1, \Sigma_1)$ et tel que le vecteur $X = (X_1, X_2, \dots, X_n)$ suive une multinormale $N_n(\mu, \Sigma)$ où $\Sigma = (\sigma_{ij})$, $\sigma_{ii} = \sigma^2$ et $\sigma_{ij} = \rho\sigma^2$, pour $i = 1, 2, \dots, n$ et où Y est variable aléatoire normale $N(\mu_Y, \sigma_Y^2)$. Le modèle ainsi obtenu est appelé modèle normal interclasse. Posons

$Cov(X_i, Y) = \psi\sigma\sigma_Y$, pour $i = 1, 2, \dots, n$. La matrice de variance-covariance Σ_1 de T est alors donnée par :

$$\Sigma_1 = \begin{pmatrix} \Sigma & \sigma_{XY} \\ \sigma_{XY}^t & \sigma_Y^2 \end{pmatrix}$$

où $\sigma_{XY}^t = (\psi\sigma\sigma_Y, \dots, \psi\sigma\sigma_Y)$ est un vecteur $1 \times n$. Désignons par f la fonction de densité du vecteur T et f_0 en cas d'indépendance de X et Y . Nous savons que les matrices d'information associées sont données respectivement par :

$\mathbf{I} = \Sigma_1^{-1}$ et

$$\mathbf{I}_0 = \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & \frac{1}{\sigma_Y^2} \end{pmatrix}.$$

D'où :

$$\mathbf{I}^{-1}\mathbf{I}_0 = \begin{pmatrix} I_n & \frac{\sigma_{XY}}{\sigma_Y^2} \\ \sigma_{XY}^t \Sigma^{-1} & 1 \end{pmatrix}$$

et son polynôme caractéristique $K(\lambda)$ est donné par :

$$\begin{aligned} K(\lambda) &= \det(\mathbf{I}^{-1}\mathbf{I}_0 - \lambda I_{n+1}) \\ &= \det \begin{pmatrix} (1 - \lambda)I_n & \frac{\sigma_{XY}}{\sigma_Y^2} \\ \sigma_{XY}^t \Sigma^{-1} & 1 - \lambda \end{pmatrix}. \end{aligned}$$

D'après la propriété suivante sur le déterminant des matrices :

$$\det \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \det A_{11} \det[A_{22} - A_{21}A_{11}^{-1}A_{12}],$$

nous obtenons :

$$K(\lambda) = (1 - \lambda)^{n-1} [(1 - \lambda)^2 - \frac{1}{\sigma_Y^2} \sigma_{XY}^t \Sigma^{-1} \sigma_{XY}].$$

En réécrivant la matrice $\Sigma^{-1} = (\alpha_{ij})$ où

$$\alpha_{ii} = \frac{1}{\sigma^2(1 - \rho)} - \frac{\rho}{\sigma^2(1 - \rho)(1 + (n - 1)\rho)}$$

et

$$\alpha_{ij} = -\frac{\rho}{\sigma^2(1-\rho)(1+(n-1)\rho)} \text{ si } i \neq j,$$

nous obtenons après quelques calculs l'égalité suivante :

$$\frac{1}{\sigma_Y^2} \sigma_{XY}^t \Sigma^{-1} \sigma_{XY} = \frac{n\psi^2}{1+(n-1)\rho}.$$

Le polynôme caractéristique s'écrit alors :

$$K(\lambda) = (1-\lambda)^{n-1} \left[(1-\lambda)^2 - \frac{n\psi^2}{1+(n-1)\rho} \right].$$

Ainsi, 1 est une valeur propre d'ordre de multiplicité $n-1$ tandis que $1 - \sqrt{\frac{n\psi^2}{1+(n-1)\rho}}$ et $1 + \sqrt{\frac{n\psi^2}{1+(n-1)\rho}}$ sont des valeurs propres simples de la matrice $\mathbf{I}^{-1}\mathbf{I}_0$.

Dans ce cas, la mesure de dépendance est donnée par :

$$\delta_{h,f}(X) = h\left(1 - \sqrt{\frac{n\psi^2}{1+(n-1)\rho}}\right) + h\left(1 + \sqrt{\frac{n\psi^2}{1+(n-1)\rho}}\right).$$

Notons que $\delta_{h,f}(X) = 0$ si et seulement si $\psi = 0$. De plus, pour $h(w) = \frac{(w-1)^2}{2}$, nous avons :

$$\delta_{h,f}(X) = \frac{n\psi^2}{1+(n-1)\rho}.$$

3.2.9 Exemple 4 : Cas de la distribution de Dirichlet inverse multidimensionnelle.

Soient p_1, \dots, p_{n+1} des nombres réels et $X = (X_1, X_2, \dots, X_n)$ un vecteur aléatoire réel de fonction de densité $f(x)$. On dit que la loi de X est une distribution de Dirichlet inverse si $f(x)$ s'écrit :

$$f(x) = \frac{\Gamma(p_1 + \dots + p_{n+1})}{\Gamma(p_1) \dots \Gamma(p_{n+1})} \cdot \frac{x_1^{p_1-1} \dots x_n^{p_n-1}}{(1 + x_1 + \dots + x_n)^{p_1 + \dots + p_{n+1}}}$$

pour $x_i \geq 0$ et $p_i > 0$. Les fonctions de densité marginales sont alors données par :

$$f_i(x_i) = \frac{\Gamma(p_i + p_{n+1})}{\Gamma(p_i)\Gamma(p_{n+1})} \cdot \frac{x_i^{p_i-1}}{(1 + x_i)^{p_i + p_{n+1}}},$$

avec $x_i \geq 0$, $p_i > 0$ et $i = 1, 2, \dots, n$.

Considérons le cas particulier où $p_1 = p_2 = \dots = p_n = M$ et $p_{n+1} = \alpha$.

Il vient :

$$f(x) = \frac{\Gamma(nM + \alpha)}{(\Gamma(M))^n \Gamma(\alpha)} \cdot \frac{(\prod_{i=1}^n x_i)^{M-1}}{(1 + x_1 + \dots + x_n)^{nM + \alpha}} \quad (3.1)$$

et

$$f_i(x_i) = \frac{\Gamma(M + \alpha)}{\Gamma(M)\Gamma(\alpha)} \cdot \frac{x_i^{M-1}}{(1 + x_i)^{M + \alpha}}. \quad (3.2)$$

La fonction de densité $f_0(x)$ en cas d'indépendance est alors donnée par :

$$f_0(x) = \prod_{i=1}^n f_i(x_i) \quad (3.3)$$

$$= \left(\frac{\Gamma(M + \alpha)}{\Gamma(M)\Gamma(\alpha)} \right)^n \cdot \prod_{i=1}^n \frac{x_i^{M-1}}{(1 + x_i)^{M + \alpha}}. \quad (3.4)$$

Des relations (3.1) et (3.4) précédentes, nous déduisons respectivement, les formules suivantes :

$$\int \frac{(\prod_{i=1}^n x_i)^{M-1}}{(1+x_1+\dots+x_n)^{nM+\alpha}} dx = \frac{(\Gamma(M))^n \Gamma(\alpha)}{\Gamma(nM+\alpha)}, \quad (3.5)$$

et

$$\int \prod_{i=1}^n \frac{x_i^{M-1}}{(1+x_i)^{M+\alpha}} dx = \left[\frac{\Gamma(M)\Gamma(\alpha)}{\Gamma(M+\alpha)} \right]^n. \quad (3.6)$$

A l'aide de la relation (3.5) et de la définition (2.3) de la matrice d'information \mathbf{I} introduite précédemment, nous obtenons que l'expression de la matrice d'information $\mathbf{I} = (a_{ij})$ associée à $f(x)$ est donnée par :

$$a_{ii} = \frac{\alpha(\alpha+1)(nM-M+\alpha+3)}{(M-2)(nM+\alpha+1)}$$

et

$$a_{ij} = -\frac{\alpha(\alpha+1)}{nM+\alpha+1},$$

pour $i, j = 1, 2, \dots, n$ et $M > 2$. De même, en combinant la relation (3.6) et la définition (2.6) donnant la matrice d'information \mathbf{I}_0 en cas d'indépendance, il vient :

$$\mathbf{I}_0 = \frac{\alpha(\alpha+1)(\alpha+3)}{(M-2)(M+\alpha+1)} I_n,$$

pour $i = 1, 2, \dots, n$ et $M > 2$. Notons que comme précédemment, la matrice \mathbf{I} peut s'écrire :

$$\mathbf{I} = \frac{\alpha(\alpha+1)}{M-2} I_n - \frac{\alpha(\alpha+1)}{nM+\alpha+1} J_n.$$

Sachant que $J_n^2 = nJ_n$, quelques calculs permettent d'obtenir :

$$\left(\frac{\alpha(\alpha+1)}{M-2} I_n - \frac{\alpha(\alpha+1)}{nM+\alpha+1} J_n \right) \cdot \left(\frac{M-2}{\alpha(\alpha+1)} I_n + \frac{(M-2)^2}{\alpha(\alpha+1)(2n+\alpha+1)} J_n \right) = I_n;$$

ce qui donne l'inverse de la matrice \mathbf{I}^{-1} exprimée par :

$$\mathbf{I}^{-1} = \frac{M-2}{\alpha(\alpha+1)}I_n + \frac{(M-2)^2}{\alpha(\alpha+1)(2n+\alpha+1)}J_n.$$

D'où :

$$\begin{aligned}\mathbf{I}^{-1}\mathbf{I}_0 &= \left[\frac{M-2}{\alpha(\alpha+1)}I_n + \frac{(M-2)^2}{\alpha(\alpha+1)(2n+\alpha+1)}J_n \right] \cdot \left[\frac{\alpha(\alpha+1)(\alpha+3)}{(M-2)(M+\alpha+1)}I_n \right] \\ &= \frac{\alpha+3}{M+\alpha+1} \left[I_n + \frac{M-2}{2n+\alpha+1}J_n \right].\end{aligned}$$

Le polynôme caractéristique $Q(\lambda)$ de la matrice $\mathbf{I}^{-1}\mathbf{I}_0$ est alors donné par :

$$\begin{aligned}Q(\lambda) &= \det(\mathbf{I}^{-1}\mathbf{I}_0 - \lambda I_n) \\ &= \det\left[\left(\frac{\alpha+3}{M+\alpha+1} - \lambda \right) I_n + \frac{(\alpha+3)(M-2)}{(M+\alpha+1)(2n+\alpha+1)}J_n \right].\end{aligned}$$

Comme la matrice J_n est de rang 1 et admet pour unique valeur propre $n = tr(J_n)$, nous avons :

$$\begin{aligned}Q\left(\frac{(\alpha+3)(nM+\alpha+1)}{(M+\alpha+1)(2n+\alpha+1)}\right) &= \left[\frac{(\alpha+3)(M-2)}{(M+\alpha+1)(2n+\alpha+1)} \right]^n \det(J_n - nI_n) \\ &= 0,\end{aligned}$$

ainsi, la valeur

$$\lambda = \frac{(\alpha+3)(nM+\alpha+1)}{(M+\alpha+1)(2n+\alpha+1)}$$

est une valeur propre simple de $\mathbf{I}^{-1}\mathbf{I}_0$.

De même, nous avons :

$$\begin{aligned}Q\left(\frac{\alpha+3}{M+\alpha+1}\right) &= \left[\frac{(\alpha+3)(M-2)}{(M+\alpha+1)(2n+\alpha+1)} \right]^n \det(J_n) \\ &= 0;\end{aligned}$$

donc,

$$\lambda = \frac{\alpha + 3}{M + \alpha + 1},$$

est une valeur propre d'ordre $n - 1$ de $\mathbf{I}^{-1}\mathbf{I}_0$. La mesure de dépendance est alors donnée par :

$$\delta_{h,f}(X) = (n - 1)h\left(\frac{\alpha + 3}{M + \alpha + 1}\right) + h\left(\frac{(\alpha + 3)(nM + \alpha + 1)}{(M + \alpha + 1)(2n + \alpha + 1)}\right),$$

pour $M > 2$ et $\alpha > 0$.

3.2.10 Version empirique de la mesure de Zografos.

Supposons que le paramètre θ ($\theta \in \mathbb{R}^n$) de la loi $f(x, \theta)$ du vecteur aléatoire $X = (X_1, X_2, \dots, X_n)$ soit inconnu et désignons par $\hat{X}_1, \hat{X}_2, \dots, \hat{X}_m$ un échantillon de taille m du vecteur X et supposons par ailleurs que l'on dispose d'un estimateur $\hat{\theta}$ de θ qui soit asymptotiquement sans biais et convergent. Notons par \hat{f} et \hat{f}_0 les densités $f(x, \hat{\theta})$ et $f_0(x, \hat{\theta}) = \prod_{i=1}^n f_i(x_i, \hat{\theta})$ et par :

$$\hat{\delta}_{h,f}(X) = \hat{\delta}_{h,f}(X_1, X_2, \dots, X_n) = \sum_{i=1}^n h(\hat{\lambda}_i)$$

la mesure de Zografos empirique associée à (\hat{f}, \hat{f}_0) où $\hat{\lambda}_i$ $i = 1, 2, \dots, n$ sont les valeurs propres de la matrice $\hat{\mathbf{I}}^{-1}\hat{\mathbf{I}}_0$. Remarquons que les matrices $\hat{\mathbf{I}}^{-1}\hat{\mathbf{I}}_0$ et $\hat{\mathbf{I}}_0^{\frac{1}{2}}\hat{\mathbf{I}}^{-1}\hat{\mathbf{I}}_0^{\frac{1}{2}}$ possèdent les mêmes valeurs propres $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_n$ et comme la matrice $\hat{\mathbf{I}}_0^{\frac{1}{2}}\hat{\mathbf{I}}^{-1}\hat{\mathbf{I}}_0^{\frac{1}{2}}$ est presque sûrement définie positive, les valeurs propres $\hat{\lambda}_i$ sont, pour tout $i = 1, 2, \dots, n$, presque sûrement strictement positives. Nous pouvons facilement les déterminer à partir de la matrice :

$$\hat{\Phi} = \left[\mathbb{E}_{\hat{f}_0} \left(\hat{S}_0^t \times \hat{S}_0 \right) \right]^{\frac{1}{2}} \left[\mathbb{E}_{\hat{f}} \left(\hat{S}^t \times \hat{S} \right) \right]^{-1} \left[\mathbb{E}_{\hat{f}_0} \left(\hat{S}_0^t \times \hat{S}_0 \right) \right]^{\frac{1}{2}},$$

où les vecteurs \hat{S} et \hat{S}_0 sont donnés respectivement par :

$$\hat{S} = \nabla \text{Log} \hat{f}(x) \quad \text{et par} \quad \hat{S}_0 = \nabla \text{Log} \hat{f}_0(x).$$

Par ailleurs les valeurs propres $\hat{\lambda}_i$ de $\hat{\Phi}$ et $\hat{\mu}_i$ de $\hat{\Phi}^{-1}$ sont, pour tout $i = 1, 2, \dots, n$, liées entre elles par les relations $\hat{\mu}_i = \frac{1}{\hat{\lambda}_i}$ de sorte que la mesure empirique de Zografos peut également s'écrire sous la forme :

$$\hat{\delta}_{h,f}(X) = \hat{\delta}_{h,f}(X_1, X_2, \dots, X_n) = \sum_{i=1}^n h\left(\frac{1}{\hat{\mu}_i}\right)$$

où $\hat{\mu}_i$, $i = 1, 2, \dots, n$, sont les valeurs propres de la matrice :

$$\hat{\Phi}^{-1} = \left[\mathbb{E}_{\hat{f}_0} \left(\hat{S}_0^t \times \hat{S}_0 \right) \right]^{-\frac{1}{2}} \mathbb{E}_{\hat{f}} \left(\hat{S}^t \times \hat{S} \right) \left[\mathbb{E}_{\hat{f}_0} \left(\hat{S}_0^t \times \hat{S}_0 \right) \right]^{-\frac{1}{2}}$$

qui n'est autre, à un changement d'échelle près, que la matrice de variance-covariance du vecteur aléatoire $\nabla \text{Log} \hat{f}(x)$.

Remarque : Rappelons que dans le cas de la loi normale multidimensionnelle $N(\mu, \Sigma)$, où μ est le paramètre de tendance centrale, les matrices \mathbf{I} et \mathbf{I}_0 ont pour expressions :

$$\mathbf{I} = \Sigma^{-1} \quad \text{et} \quad \mathbf{I}_0 = \Sigma_0^{-1} = \text{diag}(1/\sigma_1^2, 1/\sigma_2^2, \dots, 1/\sigma_n^2)$$

où σ_i^2 désigne la variance de la variable X_i pour $i = 1, 2, \dots, n$. Dans ce cas les matrices Φ et $\hat{\Phi}$ ne sont autres, pour les lois f et \hat{f} , que les matrices des corrélations du vecteur aléatoire X et où Φ est donnée, en posant $S = \nabla \text{Log} f(x)$ et $S_0 = \nabla \text{Log} f_0(x)$ par :

$$\Phi = \left[\mathbb{E}_{f_0} \left(S_0^t \times S_0 \right) \right]^{\frac{1}{2}} \left[\mathbb{E}_f \left(S^t \times S \right) \right]^{-1} \left[\mathbb{E}_{f_0} \left(S_0^t \times S_0 \right) \right]^{\frac{1}{2}}$$

Ainsi, en ce qui concerne en particulier la loi normale, Φ est une matrice de variance-covariance. Dans ce cas et selon Pi-Erh Lin [21], la matrice $\hat{\Phi}$, converge en probabilité vers Φ lorsque m tend vers $+\infty$. De plus, ses valeurs propres empiriques $\hat{\lambda}_i$ convergent également en probabilité vers les valeurs propres de Φ , λ_i , pour tout $i = 1, 2, \dots, n$. Il s'ensuit que pour toute fonction numérique et différentiable h , nous avons la convergence de $h(\hat{\lambda}_i)$ vers $h(\lambda_i)$ pour tout $i = 1, 2, \dots, n$; ce qui nous amène à conclure que la mesure de Zografos empirique $\hat{\delta}_{f,h}$ converge en probabilité, vers sa version théorique $\delta_{f,h}$.

3.3 Ordre de dépendance basée sur la matrice d'information de Fisher.

Dans cette section, nous présentons la notion d'ordre de dépendance basée sur la matrice d'information. Les situations sont nombreuses pour lesquelles, en présence de deux ou plusieurs vecteurs aléatoires, nous souhaitons connaître le vecteur qui prend les grandes ou les petites valeurs avec de fortes ou de faibles probabilités. Par exemple, la fiabilité d'un système à n composants en série fonctionnant simultanément au-delà d'un temps t et dont le temps d'arrêt de chaque composant est X_i est donnée par : $R = P(X_i > t; i = 1, 2, \dots, n)$. Lorsque ces composants fonctionnent indépendamment les uns des autres, la fiabilité du système est donnée par : $R' = \prod_{i=1}^n P(X_i > t)$. Nous savons que l'inégalité $R \geq R'$ a lieu lorsque les variables X_i sont positivement associées et que l'inégalité $R \leq R'$ a lieu au cas où ces variables sont négativement associées. Ces deux formes de dépendance nous permettent de comparer le vecteur X au vecteur $Y = (Y_1, Y_2, \dots, Y_n)$ où les variables Y_i sont toutes indépendantes et vérifient $X_i = Y_i, i = 1, 2, \dots, n$. En général, un ordre de dépendance peut être défini comme une comparaison de deux ou plusieurs distributions multidimensionnelles dans le but de déterminer laquelle est la plus dépendante. On pourra consulter [15] pour obtenir une étude détaillée sur les ordres stochastiques de dépendance.

Les matrices d'information \mathbf{I} et \mathbf{I}_0 ne satisfont pas en général la relation $\mathbf{I} - \mathbf{I}_0 \geq 0$. En effet, en prenant $\alpha = 4$ et $M = 7$ dans l'exemple 4 précédent, les matrices d'information correspondantes sont données par :

$$\mathbf{I} = \begin{pmatrix} \frac{56}{19} & \frac{-20}{19} \\ \frac{-20}{19} & \frac{56}{19} \end{pmatrix}$$

et

$$\mathbf{I}_0 = \begin{pmatrix} \frac{7}{3} & 0 \\ 0 & \frac{7}{3} \end{pmatrix}.$$

Pour $X = (1, 1)^t$, nous avons :

$X^t(\mathbf{I} - \mathbf{I}_0)X = -0.88 \leq 0$; ce qui montre que $\mathbf{I} - \mathbf{I}_0$ n'est pas *SDP* dans ce cas. Nous considérons, la famille \mathcal{F}^* de distributions n -dimensionnelles à marginales fixes dont les matrices d'information correspondantes \mathbf{I} et \mathbf{I}_0 vérifient la relation $\mathbf{I} - \mathbf{I}_0 \geq 0$.

Nous définissons un ordre partiel noté " \ll " sur \mathcal{F}^* par :

pour toutes fonctions $f, g \in \mathcal{F}^*$, $f \ll g$ si et seulement si $\mathbf{I}_g - \mathbf{I}_f \geq 0$ où \mathbf{I}_g et \mathbf{I}_f désignent respectivement les matrices d'information associées aux densités f et g .

Notons que " \ll " est un ordre partiel strict et que bien que restrictive, la condition $\mathbf{I} - \mathbf{I}_0 \geq 0$ qui apparaît dans la définition de l'ensemble \mathcal{F}^* est nécessaire à la preuve du résultat suivant qui établit le pont reliant l'ordre " \ll " entre deux fonctions de densité f et g appartenant à \mathcal{F}^* à l'ordre naturel entre leur mesure de dépendance correspondante $\delta_{h,f}(X)$ et $\delta_{h,g}(X)$.

Théorème 3.3

Soient $f, g \in \mathcal{F}^$. Alors $f \ll g$ entraîne $\delta_{h,f}(X) \leq \delta_{h,g}(X)$.*

Preuve:

Remarquons tout d'abord que les trois relations $\mathbf{I}_0 \leq \mathbf{I}_g$, $\mathbf{I}_0 \leq \mathbf{I}_f$ et $\mathbf{I}_f \leq \mathbf{I}_g$ impliquent que : $\mathbf{I}_0 \leq \mathbf{I}_f \leq \mathbf{I}_g$ et $\mathbf{I}_g^{-1}\mathbf{I}_0 \leq \mathbf{I}_f^{-1}\mathbf{I}_0$. De plus, en désignant par λ_i^f et λ_i^g , $i = 1, 2, \dots, n$, les valeurs propres $\mathbf{I}_f^{-1}\mathbf{I}_0$ et $\mathbf{I}_g^{-1}\mathbf{I}_0$, nous pouvons établir, grâce à une transformation orthogonale, que les valeurs propres λ'_i de la matrice $\mathbf{I}_f^{-1}\mathbf{I}_0 - \mathbf{I}_g^{-1}\mathbf{I}_0$ sont données par la relation :

$$\lambda'_i = \lambda_i^f - \lambda_i^g \text{ pour } i = 1, 2, \dots, n.$$

En effet, pour toute matrice (*S*)*DP* A , il existe toujours une transformation de la forme $y = Px$ telle que :

$$x^t A x = y^t \Lambda y = \sum_{i=1}^n \lambda_i y_i^2$$

où Λ est la matrice diagonale des $\lambda_i \geq 0$.

On peut alors écrire les deux égalités suivantes : $x^t(\mathbf{I}_f^{-1}\mathbf{I}_0 - \mathbf{I}_g^{-1}\mathbf{I}_0)x = \sum_{i=1}^n \lambda_i^f y_i^2 = \sum_{i=1}^n (\lambda_i^f - \lambda_i^g) y_i^2$.

Comme la matrice $\mathbf{I}_f^{-1}\mathbf{I}_0 - \mathbf{I}_g^{-1}\mathbf{I}_0$ est *SDP*, toutes les valeurs propres λ_i^f sont positives ou nulles; ce qui entraîne : $\lambda_i^f \geq \lambda_i^g$, pour $i = 1, 2, \dots, n$. Nous savons que $0 \leq \lambda_i^f \leq 1$ et que la fonction h est décroissante dans l'intervalle $(0, 1]$, donc : $h(\lambda_i^f) \leq h(\lambda_i^g)$, pour $i = 1, 2, \dots, n$. D'où :

$$\begin{aligned} \delta_{h,f}(X) &= \sum_{i=1}^n h(\lambda_i^f) \\ &\leq \sum_{i=1}^n h(\lambda_i^g) = \delta_{h,g}(X). \end{aligned}$$

Au regard de ce théorème, l'ordre " \ll " peut être considéré comme un ordre de dépendance pour les distributions multidimensionnelles et dans ce cas $f \ll g$ signifie que la loi g contient plus de dépendance que la loi f .

3.3.1 Exemple d'ordre de dépendance multidimensionnelle : cas de la loi multinormale.

Considérons la famille des distributions gaussiennes définie plus haut par : $\mathcal{F} = \{f_\Sigma(x) : f_\Sigma(x) = (2\pi)^{-\frac{n}{2}} (\det \Sigma)^{-\frac{1}{2}} \exp[-\frac{1}{2}(x - \mu)^t \Sigma^{-1}(x - \mu)]\}$, où la matrice Σ est DP, $x \in \mathbb{R}^n$.

Les densités marginales $f_i(x_i)$ du vecteur X sont fixes, de moyennes μ_i et de variances σ_i^2 . Soient $f_{\Sigma_1}, f_{\Sigma_2} \in \mathcal{F}$. Rappelons que dans ce cas, la mesure de dépendance $\delta_{h,f}$ est donnée par : $\delta_{h,f} = \log(\det \frac{\Sigma_1}{\Sigma_2})$. L'ordre " \ll " défini sur \mathcal{F} par : $f_{\Sigma_1} \ll f_{\Sigma_2}$ si et seulement si $\Sigma_1 \geq \Sigma_2$, est un ordre de dépendance multidimensionnelle et pour cet ordre, nous avons :

$$\delta_{h,f_{\Sigma_1}}(X) \leq \delta_{h,f_{\Sigma_2}}(X),$$

pour $h(w) = -1 + w - \log w$. En effet, la relation $\Sigma_1 \geq \Sigma_2$ implique que :

$$\det \Sigma_1 \geq \det \Sigma_2.$$

D'où :

$$\frac{\det \Sigma_1}{\det \Sigma_0} \geq \frac{\det \Sigma_2}{\det \Sigma_0} \geq 1.$$

Soit :

$$0 \leq \log\left[\frac{\det \Sigma_0}{\det \Sigma_1}\right] \leq \log\left[\frac{\det \Sigma_0}{\det \Sigma_2}\right],$$

d'où le résultat.

3.4 Autres mesures de dépendance multidimensionnelle.

3.4.1 Le tau de Kendall.

Le tau de Kendall, noté, τ est une mesure d'association entre deux ou plusieurs variables.

Selon Harry Joe [5], la définition dans le cas général est :

$$\tau(X) = \frac{2^n \mathbb{E}\{F(x)\} - 1}{2^{n-1} - 1},$$

où X est un vecteur aléatoire n -dimensionnel de fonction de répartition F . Notons que de façon générale, τ est compris entre -1 et 1 ; de plus, $\tau = 0$ si les variables sont indépendantes, mais le contraire n'est pas vrai en général. Soit

$$\tilde{X} = (X_{ij}), 1 \leq i \leq n \text{ et } 1 \leq j \leq m$$

un m -échantillon du vecteur aléatoire X . D'après Genest, Quessy et Remillard [16], la quantité $\mathbb{E}\{F(x)\}$ peut être estimée par :

$$\hat{P} = \frac{1}{n^2} \sum_{i=1}^n \sum_{k=1}^n \prod_{j=1}^m I_{(x_{kj} \leq x_{ij})}.$$

Une estimation du tau de Kendall, notée, $\hat{\tau}(X)$ est alors donnée par :

$$\hat{\tau}(X) = \frac{2^n \hat{P} - 1}{2^{n-1} - 1}.$$

3.4.2 Le rhô de Spearman.

Le rhô de Spearman, noté, ρ est fréquemment utilisé pour quantifier la force de liaison entre deux ou plusieurs variables aléatoires. Sa formulation dans le cas général est donnée selon [16] par :

$$\rho(X) = \frac{n+1}{2^{n-1}} \cdot \frac{2^n \mathbb{E}\{F_1(x_1)F_2(x_2)\dots F_n(x_n)\} - 1}{2^n}.$$

Remarquons que $\rho(X)$ est compris entre -1 et 1, et que $\rho(X) = 0$ en cas d'indépendance. Soit $X_i = (X_{i1}, X_{i2}, \dots, X_{in}), i = 1, 2, \dots, m$, un m -échantillon du vecteur aléatoire X . Désignons par R_{ij} le rang de la valeur X_{ij} dans l'observation $(X_{i1}, X_{i2}, \dots, X_{in})$. Alors une estimation du rhô de Spearman est donnée selon [16] par :

$$\hat{\rho}(X) = \frac{\frac{1}{m} \sum_{i=1}^m R_{i1} \dots R_{im} - \left(\frac{m+1}{2}\right)^n}{\frac{1}{m} \sum_{i=1}^m i^n - \left(\frac{m+1}{2}\right)^m}.$$

3.4.3 L'entropie relative.

Soit $X = (X_1, X_2, \dots, X_n)$ un vecteur aléatoire réel de fonction de densité conjointe $f(x)$ et $f_0(x)$ en cas d'indépendance. L'entropie relative du vecteur X , que nous noterons $E(X)$, peut être définie par :

$$E(X) = \int f(x) \log\left[\frac{f(x)}{\prod_{i=1}^n f_i(x_i)}\right] dx$$

Notons que dans le cas des variables discrètes, l'entropie relative est donnée par :

$$E(X) = \sum_{x_1, x_2, \dots, x_n} P(x_1, x_2, \dots, x_n) \log \left[\frac{P(x_1, x_2, \dots, x_n)}{\prod_{i=1}^n P_i(x_i)} \right],$$

où $P(\cdot)$ et $P_i(\cdot)$ désignent respectivement la loi conjointe du vecteur X et la loi marginale de la variable X_i .

Selon Csiszär et Körner [17], l'entropie relative peut être vue comme une mesure de dépendance entre plusieurs variables aléatoires et ses grandes valeurs indiquent une forte dépendance entre les variables. D'après Harry Joe [5], cette quantité est toujours positive et est nulle si et seulement si toutes les variables sont indépendantes. Soit $X_i = (X_{i1}, X_{i2}, \dots, X_{in}), i = 1, 2, \dots, m$, un m -échantillon du vecteur X . En désignant par $P(x_1, x_2, \dots, x_n)$ la proportion théorique des modalités x_1, x_2, \dots, x_n et par $\hat{P}(x_1, x_2, \dots, x_n)$ les proportions issues d'un échantillon de taille m . L'entropie relative peut être estimée dans le cas discret [4] par :

$$\hat{E}(X) = \sum_{x_1, x_2, \dots, x_n} \hat{P}(x_1, x_2, \dots, x_n) \log \left[\frac{\hat{P}(x_1, x_2, \dots, x_n)}{\prod_{i=1}^n \hat{P}_i(x_i)} \right].$$

Dans le cas continu, on obtient une estimation de la fonction de densité par :

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^m K\left(\frac{x - x_i}{h}\right),$$

pour $x \in \mathbb{R}^n$ et où K est une fonction de densité n -dimensionnelle vérifiant :

$$K(-u) = K(u) \text{ et}$$

$$\hat{f}_j(x) = \frac{1}{m} \sum_{i=1}^m K_j\left(\frac{x_j - x_{ij}}{h}\right),$$

où K_j est la j^{eme} marginale unidimensionnelle de K , les estimations respectives des fonctions $f(x)$ et f_j , une estimation de l'entropie relative est obtenue dans le cas continu (voir entre autres Silverman [18]) :

$$\hat{E}(X) = \frac{1}{m} \sum_{i=1}^m [\log \hat{f}(x_{i1}, x_{i2}, \dots, x_{in}) - \sum_{j=1}^m \log \hat{f}_j(x_{ij})].$$

3.4.4 La mesure de Harry Joe.

Partant de la mesure d'entropie relative $E(X)$, Harry Joe [4] a défini une mesure de dépendance, notée $\delta^*(X)$ entre les composantes d'un vecteur aléatoire par :

$$\delta^*(X) = 1 - \exp\{-2E(X)\}.$$

Relevons que $\delta^*(X)$ est plus facile à interpréter que $\mathcal{E}(\cdot)$, car comprise entre 0 et 1. La valeur 0 est atteinte en cas d'indépendance et les valeurs proches de 1 expriment une forte dépendance entre les composantes du vecteur X . Pour un m -échantillon, une estimation naturelle de $\delta^*(X)$ est donnée par :

$$\hat{\delta}^*(X) = 1 - \exp\{-2\hat{E}(X)\}.$$

CHAPITRE 4

Simulations.

Les simulations ont été réalisées dans le but de déterminer la convergence de la mesure de Zografos empirique vers sa version théorique. Ces simulations ont été faites dans le cas de la loi normale bidimensionnelle. Celle-ci comporte cinq paramètres, à savoir les moyennes et les variances des deux variables ainsi que le coefficient ρ caractérisant la force de la corrélation qui existe entre elles. Notons que la valeur zéro de ρ correspond à l'indépendance tandis que les valeurs -1 et 1 correspondent à une parfaite dépendance entre les deux variables. Puisque la mesure de Zografos dans ce cas ne dépend que de ρ et en vertu du fait qu'elle est invariante par transformation linéaire, des observations $(X_i, Y_i), i = 1, 2, \dots, n$ ont été générées d'une loi binormale $N_2(\mu, \Sigma)$, où $\mu = (0, 0)$ et

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_Y\sigma_X & \sigma_Y^2 \end{pmatrix},$$

$\rho \in [-1, 1]$. Un programme SAS [19] (voir premier tableau de l'annexe A) a été utilisé pour générer ces valeurs. Par la suite et compte tenu de la complexité SAS/IML, un programme C++ (voir l'annexe A) a été écrit pour calculer les mesures empiriques de Zografos sur des échantillons précédemment obtenus. Les résultats obtenus pour chaque taille d'échantillon sont consignés dans les tableaux 4.1, 4.2 et 4.3 donnés ci-dessous.

$\rho = 0.94, h(w) = -1 + w - \log w$ et $\delta_{h,f}(X, Y) = -\log(1 - \rho^2) = 2.1507$		
taille d'échantillon	n=25	n=50
$\hat{\delta}_{h,f}(X, Y)$	2.2575	2.4460
nombre des répliques	m=200	m=200
moyenne des répliques	2.3398	2.3063

Tableau 4.1: Résultats des simulations basés sur l'estimation de $\delta_{h,f}(X, Y), \rho = 0.94$

$\rho = 0.99, h(w) = -1 + w - \log w$ et $\delta_{h,f}(X, Y) = -\log(1 - \rho^2) = 3.917$		
taille d'échantillon	n=25	n=50
$\hat{\delta}_{h,f}(X, Y)$	3.7754	4.2155
nombre des répliques	m=200	m=200
moyenne des répliques	3.9118	3.9359

Tableau 4.2: Résultats des simulations basés sur l'estimation de $\delta_{h,f}(X, Y), \rho = 0.99$

$\rho = 0, h(w) = -1 + w - \log w$ et $\delta_{h,f}(X, Y) = -\log(1 - \rho^2) = 0$		
taille d'échantillon	n=25	n=50
$\hat{\delta}_{h,f}(X, Y)$	0.00345	0.0032
nombre des répliques	m=200	m=200
moyenne des répliques	0.01187	0.00984

Tableau 4.3: Résultats des simulations basés sur l'estimation de $\delta_{h,f}(X, Y), \rho = 0$

Ces résultats semblent indiquer que la mesure empirique de Zografos converge (en probabilité) vers sa version théorique. Les répliques convergent aussi en moyenne vers cette même valeur. Toutefois l'approche utilisée pour ces simulations n'est pas basée sur les matrices d'information empiriques. En effet, dans ce cas particulier de la loi normale bidimensionnelle, rappelons que les matrices d'information sont données par : $\mathbf{I} = \Sigma^{-1}$, $\mathbf{I}_0 = \Sigma_0^{-1}$, et $\mathbf{I}^{-1}\mathbf{I}_0 = \Sigma\Sigma_0^{-1}$. Les matrices empiriques s'en déduisent comme suit :

$$\hat{\mathbf{I}} = \hat{\Sigma}^{-1} = \frac{1}{1 - R^2} \begin{pmatrix} \frac{1}{S_X^2} & \frac{-R}{S_X S_Y} \\ \frac{-R}{S_Y S_X} & \frac{1}{S_Y^2} \end{pmatrix}$$

et

$$\hat{\mathbf{I}}_0 = \hat{\Sigma}_0^{-1} = \frac{1}{1 - R'^2} \begin{pmatrix} \frac{1}{S_X'^2} & \frac{-R'}{S_X' S_Y'} \\ \frac{-R'}{S_Y' S_X'} & \frac{1}{S_Y'^2} \end{pmatrix}$$

où S_X^2 et S_Y^2 et R désignent respectivement les variances et le coefficient de corrélation empiriques des variables X et Y en cas de dépendance; $S_X'^2$, $S_Y'^2$ et R' désignant les mêmes quantités en cas d'indépendance. La matrice empirique $\hat{\mathbf{I}}^{-1}\hat{\mathbf{I}}_0$ est alors donnée par :

$$\hat{\mathbf{I}}^{-1}\hat{\mathbf{I}}_0 = \frac{1}{1 - R'^2} \begin{pmatrix} \frac{S_X^2}{S_X'^2} - RR' \frac{S_Y S_X}{S_Y' S_X'} & R \frac{S_Y S_X}{S_Y'^2} - R' \frac{S_X^2}{S_Y' S_X'} \\ R \frac{S_Y S_X}{S_X'^2} - R' \frac{S_Y^2}{S_Y' S_X'} & \frac{S_Y^2}{S_Y'^2} + RR' \frac{S_Y S_X}{S_Y' S_X'} \end{pmatrix}.$$

La mesure empirique de Zografos est donnée dans ce cas par : $\hat{\delta}_{h,f}(X, Y)^t = \sum_{i=1}^n h(\hat{\lambda}_i)$ où $\hat{\lambda}_i, i = 1, 2, \dots, n$ sont les valeurs propres de $\hat{\mathbf{I}}^{-1}\hat{\mathbf{I}}_0$. Les matrices de variance-covariance empiriques ont été obtenues en activant la ligne de code [`*proc corr cov noprob; *run;`] du programme de l'annexe A. Les calculs des mesures empiriques ont été faits à l'aide du dernier programme C++ de la même annexe. Cette approche alternative donne des résultats qui figurent dans les tableaux 4.4, 4.5 et 4.6 de la page suivante. Ces trois

tableaux montrent qu'en basant le calcul de la mesure empirique de Zografos sur la matrice $\hat{\mathbf{I}}^{-1}\hat{\mathbf{I}}_0$, les valeurs obtenues sont très proches de celles que nous avons eues selon la première approche; ce qui soutient l'idée d'après laquelle cette statistique converge en probabilité vers sa version théorique. Toutefois, plusieurs répliques sont nécessaires avant de tirer définitivement une telle conclusion pour ce cas et une éventuelle généralisation pour la loi multinormale et bien d'autres lois multidimensionnelles d'intérêt.

$\rho = 0.94, h(w) = -1 + w - \log w$ et $\delta_{h,f}(X, Y) = -\log(1 - \rho^2) = 2.1507$		
taille d'échantillon	n=25	n=50
$\hat{\delta}_{h,f}(X, Y)$	2.2433	2.8896
nombre des répliques	m=200	m=200
moyenne des répliques	2.4902	2.2170

Tableau 4.4: Résultats des simulations basés sur les matrices empiriques, $\rho = 0.94$

$\rho = 0.99, h(w) = -1 + w - \log w$ et $\delta_{h,f}(X, Y) = -\log(1 - \rho^2) = 3.9170$		
taille d'échantillon	n=25	n=50
$\hat{\delta}_{h,f}(X, Y)$	5.5848	3.3935
nombre des répliques	m=200	m=200
moyenne des répliques	4.04972	4.1272

Tableau 4.5: Résultats des simulations basés sur les matrices empiriques, $\rho = 0.99$

$\rho = 0.50, h(w) = -1 + w - \log w$ et $\delta_{h,f}(X, Y) = -\log(1 - \rho^2) = 0.2877$		
taille d'échantillon	n=25	n=50
$\hat{\delta}_{h,f}(X, Y)$	0.1008	0.1452
nombre des répliques	m=200	m=200
moyenne des répliques	0.7354	0.3923

Tableau 4.6: Résultats des simulations basés sur les matrices empiriques, $\rho = 0.50$

CONCLUSION

L'étude des principales formes de dépendance entre les composantes d'un vecteur aléatoire et de leurs propriétés respectives nous permettent de mieux comprendre la façon dont les variables d'un vecteur aléatoire varient en même temps qu'elle nous offrent la possibilité de construire d'autres vecteurs aléatoires dont les composantes satisfont les dépendances de même type. La dépendance totale multidimensionnelle d'ordre deux s'est révélée particulièrement intéressante, car elle offre des conditions de dépendance faciles à vérifier et qui impliquent celles de toutes les autres formes de liaison que nous avons étudiées. Cette forme de dépendance a également permis de définir la matrice d'information de Fisher qui a offert à son tour, une condition suffisante de l'indépendance entre les variables d'un vecteur aléatoire continu. A l'aide de cette matrice, la quantité $\delta_{h,f}(\cdot)$ est définie [1] comme mesure de dépendance multidimensionnelle entre les composantes d'un vecteur aléatoire. Cette mesure satisfait assez bien les axiomes de Rényi et peut donc être classée parmi les mesures de liaison entre les composantes d'un vecteur aléatoire rencontrées dans la littérature. Les simulations semblent indiquer que dans le cas de la loi normale bidimensionnelle, la version empirique de cette mesure converge vers sa version théorique. Cependant, une étude plus poussée est nécessaire afin de dégager une telle conclusion dans le cas multidimensionnel ($n \geq 3$) et juger de l'efficacité de la mesure $\hat{\delta}_{h,f}(\cdot)$ par rapport aux mesures empiriques de dépendance de même nature comme par exemple l'entropie relative et la mesure de Harry Joe telles que introduites à la fin du chapitre 3.

Une telle étude donnerait une idée précise de la vitesse de convergence de cette mesure vers sa version théorique. En outre, en tant que statistique obtenue sur un échantillon, il serait intéressant d'étudier dans le cadre plus général des lois multinormales à faibles dimensions, la loi de $\hat{\delta}_{h,f}(\cdot)$ dans le but de confronter les tests d'hypothèses, notamment les tests d'indépendance. Une telle piste serait prometteuse, car si elle est basée sur l'approche énoncée en fin du chapitre précédent, elle permettrait d'obtenir des résultats qui traduiraient l'impact ou le comportement réel de la matrice d'information empirique. Dans cette optique, une attention pourrait être portée à la fonction h afin de trouver un lien éventuel entre celle-ci et la performance de la statistique $\hat{\delta}_{h,f}(\cdot)$.

Annexe A

```
OPTIONS NODATE;
OPTIONS NONUMBER;
% macro sample;
%do i = 1 %to 5;
data sample&i;
retain observ 0;
keep X Y ;
mu1=0; mu2=0; var1=1; var2=1; rho=0.95;
c=sqrt(1-rho**2);
do i = 1 to 10;
obser + 1;
x=rannor(0);
y=rho*x+c*rannor(0);
x=mu1+sqrt(var1)*x;
y=mu2+sqrt(var2)*y;
output;
end ;
run;
proc print; run;
% end;
% mendl;
% sample;
```

Tableau 4.7: Programme SAS pour générer les couples d'observations d'une loi binormale.

Programme C++ pour calculer les mesures empiriques de Zografos : approche 1.

```

OPTIONS NODATE;
OPTIONS NONUMBER;
% macro sample;
%do i = 1 %to 5;
data sample&i;
retain observ 0;
keep X Y ;
mu1=0; mu2=0; var1=1; var2=1; rho=0.95;
c=sqrt(1-rho**2);
do i = 1 to 10;
obser + 1;
x=rannor(0);
y=rho*x+c*rannor(0);
x=mu1+sqrt(var1)*x;
y=mu2+sqrt(var2)*y;
output;
end ;
run;
proc corr cov noprob; run;
*proc print; run;
% end;
% mend;
% sample;

```

Tableau 4.8: Programme SAS d'extraction des matrices de variance-covariance.

```

*****
#include <iostream>
# include <fstream>
#include <cstdlib>
#include <iomanip>
#include <deque>
#include <string>
#include <math.h>
using namespace std;
void Erreur(const char* p, const char* q=""){
cerr<<p<< ' ' <<q<<endl;
exit(1);
}
void main(int argc, char* argv[]){
if(argc != 3) Erreur("Il vous faut 3 arguments");
ifstream ifs(argv[1]);
if(!ifs) Erreur("Ouverture incorrecte du fichier", argv[1]);
ofstream ofs(argv[2]);
if(!ofs) Erreur("Ouverture incorrecte du fichier", argv[2]);
char tab[20];
char *ta=tab; //Déclaration du pointeur.
string tab2;
deque<double> Data,Coeffs;
int n,q,p(0);
double coeffs_mean(0);
double temp[3],t[]={0.,0.,0.,0.,0.};
ifs>>tab;//Entrer le fichier source dans le tableau tab.

```

```

while(lifs.eof()){ // ouverture du fichier source.
tab2=tab;//fixe le pointeur sur le tableau.
if(tab2=="The"){
if(p!=0){
//calculs
n=(q-6)/3; //n est le nombre de lignes lues.
t[0]=0;t[1]=0;t[2]=0;t[3]=0;t[4]=0;//Initialisation.
for(q=0;q<2*n;q+=2){
t[0]+=Data[q]; //sum(x).
t[1]+=Data[q+1]; //sum(y).
t[2]+=Data[q]*Data[q+1]; //sum(x*y).
} printf("n=%d, sum(x)=%f, sum(y)=%f, sum(x*y)=%fendl",n,t[0],t[1],t[2]);
//Impression de sum(x), sum(y) et su(xy). temp[0]=(double)t[0]/n; //temp[0]=xbar.
temp[1]=(double)t[1]/n; //temp[1]=ybar.
temp[2]=0;
for(q=0;q<2*n;q+=2){
t[3]+=pow((Data[q]-temp[0]),2);
t[4]+=pow((Data[q+1]-temp[1]),2);
temp[2]+=(Data[q]-temp[0])*(Data[q+1]-temp[1]); // sum[(x-xbar)*(y-ybar)].
} t[3]=(double)t[3]/n; //t[3]=Var(x).
t[4]=(double)t[4]/n; //t[4]=Var(y).
printf("num=%f, denom=%f endl",temp[2],(n*sqrt(t[3]*t[4]))); // Impression des étapes
de calculs.
temp[2]=(double)temp[2]/(n*sqrt(t[3]*t[4])); //Calcul du coefficient de correlation.
temp[2]=-log(1-pow(temp[2],2)); //temp[2] est la mesure de Zografos pour  $h(w) = -1 + w - \log w$ .
//temp[2]=pow(temp[2],2); temp[2] est la mesure de Zografos pour  $h(w) = (w - 1)^2/2$ .

```

```

//temp[2]=fabs(temp[2]+log(1-pow(-0.85,2))); //Epsilon.
printf("var(x)=%f, var(y)=%f endl coeff=%f endl endl",t[3],t[4],temp[2]);
Coeffs.push_back(temp[2]); //
coeffs_mean+=temp[2]; // moyenne des répliques.
} Data.clear();
q=0;p++;
} else if((q>6)&&((q%3)!=0))
Data.push_back(atof(ta));
ifs>>tab;q++;
} ifs.close();
coeffs_mean =(double) coeffs_mean/Coeffs.size();
q=0;
ofs<<"COEFFICIENTS DE CORRELATION" <<endl;
ofs<<" ECH COEFFICIENTS" <<endl;
while(Coeffs.size() != 0){
ofs<<" " <<setw(3)<<q+1<<setw(12)<<setprecision(7)
<<Coeffs.front()<<endl;
Coeffs.pop_front();q++;
} ofs<<endl<<"MOYENNE DES COEFFICIENTS: " <<coeffs_mean<<endl;
ofs.close(); }
*****FIN DU PROGRAMME*****
*****FIN DU PROGRAMME*****

```

Programme C++ pour calculer les mesures empiriques de Zografos : approche 1.

```
*****
*****
*****
```

Programme C++ pour calculer les mesures empiriques de Zografos : approche 2.

```
*****
```

```
#include <iostream>
#include <fstream>
#include <cstdlib>
#include <deque>
#include <string>
#include <math.h>

using namespace std;

//structure de la matrice
struct matrice {
double m00;
double m01;
double m10;
double m11;
};

void Erreur(const char* p, const char* q=""){
cerr<<p<<' '<<q<<endl;
exit(1);
```



```
}
```

```
int main(int argc, char* argv[]){  
if(argc != 4) Erreur("Il vous faut 4 arguments"); //affichage du message d'erreur pour  
lecture incorrecte  
ifstream fic_dépendant(argv[1]);  
//affichage du message d'erreur si le fichier source dépendant n'est pas ouvert  
if(!fic_dépendant) Erreur("Ouverture incorrecte du fichier", argv[1]);  
  
ifstream fic_indépendant(argv[2]);  
//affichage du message d'erreur si le fichier source indépendant n'est pas ouvert  
if(!fic_indépendant) Erreur("Ouverture incorrecte du fichier", argv[2]);  
ofstream sortie(argv[3]);  
//affichage du message d'erreur si le fichier de sortie n'est pas ouvert  
if(!sortie) Erreur("Ouverture incorrecte du fichier", argv[3]);  
  
//définition de variables de lecture du fichier  
char ligne[80]; //variable de lecture des lignes du fichier  
char val_xy[5]; //variable de lecture des caractères dans le tableau XY  
matrice covD; //matrice de covariance_dépendant  
matrice covI; //matrice de covariance_indépendant  
matrice produit; //matrice de produit de covD par inverse de covI  
  
//variable de lecture des éléments d'une matrice  
double a,b,c,d,det;
```

```

double lamda1, lamda2, delta;
double delta_chapeau;
double moyen_repliques = 0;

//lecture du fichier source_dépendant
//lecture des 11 premières lignes du fichier source_dépendant
for(int i=1;i<=11;i++){
fic_dépendant.getline(ligne,80);
} //lecture des 5 premières lignes du fichier source_indépendant
for(int j=1;j<=11;j++){
fic_indépendant.getline(ligne,80);
}
while(!fic_dépendant.eof()){
//lecture des caractères des lignes X et Y
fic_dépendant»val_xy;
fic_indépendant»val_xy;
//lecture des éléments de la matrice XY
fic_dépendant»covD.m00;
fic_dépendant»covD.m01;
fic_indépendant»covI.m00;
fic_indépendant»covI.m01;
//lecture du caractère Y
fic_dépendant»val_xy;
fic_indépendant»val_xy;
//lecture des éléments de la matrice XY
fic_dépendant»covD.m10;

```

```

fic_dépendant»covD.m11;
fic_indépendant»covI.m10;
fic_indépendant»covI.m11;
a=covI.m00; b=covI.m01; c=covI.m10; d=covI.m11;

//calcul de l'inverse de la matrice_independante
det= 1/(a*d-c*b);
covI.m00=det*d;
covI.m01=-det*b;
covI.m10=-det*c;
covI.m11=det*a;
//calcul du produit de covD par l'inverse de covI
produit.m00 = covD.m00*covI.m00 + covD.m01*covI.m10;
produit.m01 = covD.m00*covI.m01 + covD.m01*covI.m11;
produit.m10 = covD.m10*covI.m00 + covD.m11*covI.m10;
produit.m11 = covD.m10*covI.m01 + covD.m11*covI.m11;
//calcul des valeurs propres
delta = pow((produit.m00 + produit.m11),2) - 4*(produit.m00*produit.m11 -
produit.m01*produit.m10);
lamda1 = (produit.m00 + produit.m11 + sqrt(delta))/2;
lamda2 = (produit.m00 + produit.m11 - sqrt(delta))/2;
//calcul de h(lamda1)+h(lamda2)
delta_chapeau = -2 + lamda1 +lamda2 - log(lamda1*lamda2);
sortie«delta_chapeau«endl;

cout«covD.m00«" "«covD.m01«" "«covI.m00«" "«covI.m01«" "«produit.m00«" "«pro-

```

```

duit.m01«" "«lamda1«endl;
cout«covD.m10«" "«covD.m11«" "«covI.m10«" "«covI.m11«" "«produit.m10«" "«pro-
duit.m11«" "«lamda2«endl;
cout«delta_chapeau«endl«endl;
//calcul de la somme des repliques
moyen_repliques = moyen_repliques + delta_chapeau;

//lecture des lignes du fichier
for(int i=1;i<=21;i++)
fic_dependant.getline(ligne,80);
if(strcmp(ligne,"fin du fichier")==0) return 0; //fin du traitement
}
for(int j=1;j<=21;j++)
fic_independant.getline(ligne,80);
if(strcmp(ligne,"fin du fichier")==0) return 0; //fin du traitement
}
}

//calcul de la moyenne des repliques
moyen_repliques = moyen_repliques/200;
sortie«endl«endl«"moyenne des repliques = "«moyen_repliques;
cout«endl«endl«"moyenne des repliques = "«moyen_repliques«endl;

//fermeture des fichiers
fic_dependant.close();
fic_independant.close();

```

```
sortie.close();
```

```
return 0;
```

```
}
```

```
*****FIN DU PROGRAMME*****
```

```
*****FIN DU PROGRAMME*****
```

Programme C++ pour calculer les mesures empiriques de Zografos : approche 2.

BIBLIOGRAPHIE

- [1] Zografos, K. Measures of multivariate Dependence based on a Distance Between Fisher Information Matrices. *Journal of Statistical Planning and Inference* 89:91–107, 2000.
- [2] Esary, J. D., Proschan and Walkup, D. W. Association of Random Variables with Applications. *Ann. Math. Statist.* 38:1466–1474, 1976.
- [3] Glaz, J. A Comparison of Bonferroni-type and Product-type Inequations in Presence of Dependence. *Topics in Statistical Dependence, IMS Lecture.* 1990.
- [4] Joe, H. Relative Entropy Measures of Multivariate Dependence. *J. Amer. Statist. Assoc.* 84:157–164, 1989.
- [5] Joe, H. *Multivariate Models and Dependence Concepts.* Chapman & Hall, London, 1997.
- [6] Lehmann, E. L. Some Concepts of Dependence. *Ann. Math. Statist.* 37:1137–1153, 1966.
- [7] Sarkar, T.K. Some Lower Bounds of Reliability. Tech. Report. n. 124, Department of Operations Research and Statistics, Stanford University, Stanford, CA, 1969.

- [8] Karlin, S., Rinott, Y. Classes of Orderings of Measures and Related Correlations Inequalities. I. Multivariate Totally Positive Distributions. *J. Multivariate Anal.* 10:467–498, 1980.
- [9] Rohatgi, V. K. *An Introduction to Probability and Mathematical Statistics.* Wiley Series in Probability and Mathematical Statistics. 1976.
- [10] Edouardo Mayer-Wolf. The Cramér-Rao Functionnal and Limiting Laws. *The Annals of Probability*, vol. 18, n. 2:840–850, 1988.
- [11] Papathanasiou, V. Some Characteristic Properties of the Fisher Information Matrix via Cacoullos-type Inequalities. *J. Multivariate Anal.* 44:256–265, 1993.
- [12] Renyi, A. On Measures of Dependence. *Acta Math. Hungar.* 10:441–451, 1959.
- [13] Schweizer, B. and Wolff, E.F. On nonparametric Measures of Dependence for Random Variables. *Ann. Statist.* 9:879–885, 1981.
- [14] Vadja, I. *Theory of Statistical Inference and Information.* Kluwer Academic Publishers, Dordrecht.
- [15] Moshe Shaked, J. G., Shanthikumar. *Stochastic Orders and their Applications.* Academic Press, New York, 1994.
- [16] Genest, C., Quessy, J-F., and Rémillard, B. Test of Serial Independence Based on Kendall’s Process. *The Canadian Journal of Statistics*, vol. 30 n. 3:271–287, 2002.
- [17] Janos Korner and Imre Csiszar. *Information Theory : Coding Theorems for Discrete Memoryless Systems.* Academic Press, New York, 1981.
- [18] Silverman, B. W. *Density Estimation for Statistics and Data Analysis.* Chapman & Hall.

- [19] Bernier, N. Utilisation d'un bootstrap amélioré dans l'estimation du biais. Mémoire de Maîtrise, 2000.
- [20] Frank L. Friedman, Elliot B. Koffman. Problem Solving, Abstraction, Design Using C++; Third Edition. Addison-Wesley, Longman, 2000.
- [21] Pi-Erh Lin. Measures of Association Between Vectors; Commun. Statist.-Theory Meth., 16(2), 321-338, 1987.