

N° d'ordre : 47/2008-M/MT

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE
UNIVERSITE DES SCIENCES ET DE LA TECHNOLOGIE
HOUARI BOUMEDIENE
FACULTE DE MATHEMATIQUES



MEMOIRE

*Présenté pour l'obtention du diplôme de MAGISTER
EN MATHEMATIQUE
Spécialité : Probabilités Statistiques*

Par : M^{elle}. YASSA MALIKA

THEME

Tests de comparabilité stochastique et applications

Soutenu le : 11 Septembre 2008 devant le jury composé de :

Mr-K. BOUKHETALA, Professeur, USTHB

Président

Mr-A. AISSANI, Professeur, USTHB

Directeur de thèse

Mme-H. GUERBYENNE, Maitre de Conférence, USTHB

Examinatrice

M^{elle}-O. SADKI, Chargée de Cours, USTHB

Invitée

Tests de comparabilité stochastique et applications

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Résumé

L'objet de cette étude est de construire un test statistique pour détecter l'appartenance d'une distribution à la classe de loi non paramétrique, la loi HNBUE.

Le test est basé sur une caractérisation de la loi HNBUE à l'aide de l'ordre stochastique convexe (ou dominance stochastique de second ordre).

Le test présente est une test application d'un test plus générale de Kaur-Rao-Singh pour la comparabilité au sens de l'ordre convexe de deux variables aléatoires représentant des gains (pertes) en finances, durées de bon fonctionnement en fiabilité, etc...

REMERCEMENTS

Je remercie ALLAH qui m'a orienté au chemin du savoir et les portes de la science.

- *Je remercie tout d'abord Monsieur A. Aissani de m'avoir proposé ce sujet de thèse, et de l'attention qu'il a porté à mon travail, et aussi pour ses orientations, ses conseils et ses encouragements.*
- *Je remercie Monsieur K. Boukhetala pour avoir accepté de présider le jury*
- *Je remercie Madame H Guerbyenne et M^{lle} O.Sadki d'avoir accepté de faire partir du jury et pour leurs orientations tout au long de mon travail*
- *M^{lle} -S. Taleb, pour les documents qu'elle a bien voulu à ma disposition*
- *Sans oublier de remercier tout les enseignants du département de Probabilité & Statistique.*
- *un grand merci à mes parents qui m'ont soutenue et encouragée tout au long de ces années*
 - *Sans oublier mon cousin Rachid.*
- *En fin, Je remercie toute la famille.*

SOMMAIRE

CHAPITRE 1: Introduction et généralité

1. Introduction et généralité	8
-------------------------------------	---

CHAPITRE 2: Comparaison stochastique et les lois non paramétriques

2.1 Ordre stochastique	12
2.2 La fonction d'utilité.....	12
Définition 2.1.....	13
Caractéristiques.....	14
2.3 Aversion au risque.....	14
2.4 Comparaison stochastique dans le domaine de la fiabilité.....	15
2.4.1 Caractéristiques de la fiabilité.....	15
a) La comparaison stochastique.....	15
b) La comparaison en moyenne.....	16
c) La comparaison en moyenne résiduelle.....	17
2.4.2 D'autres notions de comparaison stochastique	17
a) Comparaison convexe.....	17
b) Comparaison concave.....	19
c) Ordre de Lorenz.....	20
d) Ordre monotone convexe (concave).....	21
2.11 Lois non paramétriques.....	22
2.11.1 Propriétés de vieillissement.....	23
1) La distribution IFR.....	23
2) La distribution IFR(2).....	23
3) La distribution IFRA.....	24
4) La distribution NBU.....	24
5) La distribution NBUE.....	25
6) La distribution HNBUE.....	26
7) la distribution HNBUE(2).....	27
8) la distribution DMRL.....	27
Conclusion.....	28

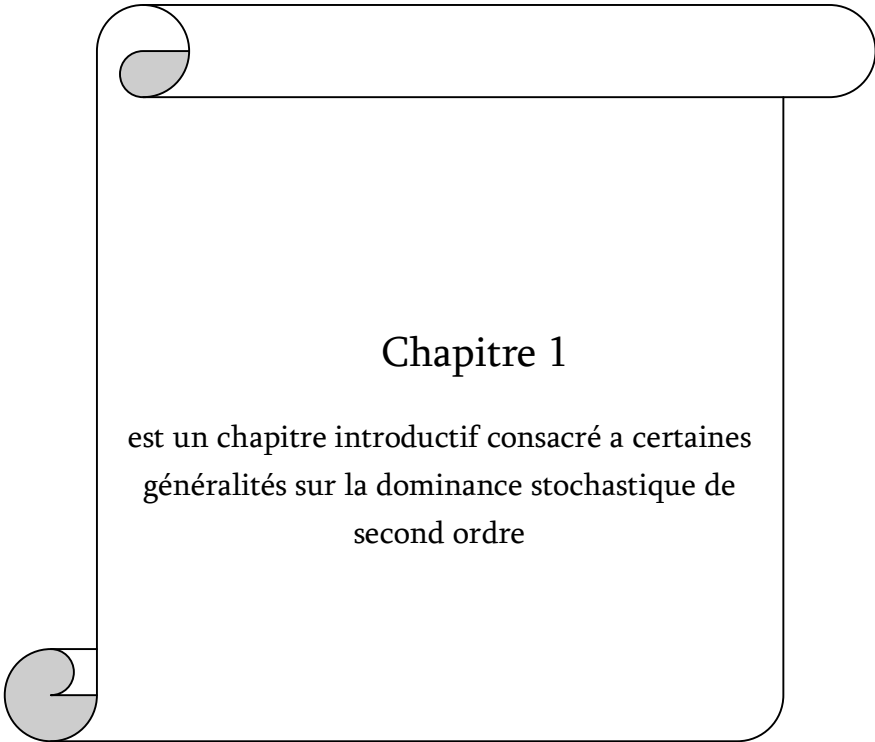
CHAPITRE 3 : Tests statistiques de comparaison stochastique

3.1 Tests paramétriques	30
3.2 Tests non paramétriques.....	31
3.2.1. Test Kolmogorov- Smirnov	31
3.2.2. Test de Mann Whitney.....	33
3.2.3 Test de HNBUE.....	34
3.3.2 Test de DMRL.....	34
3.3 Choix de l'hypothèse.....	35
3.4 Choix du test.....	32

3.5 Les avantages et les inconvénients	35
3.6 Test de comparaison convexe.....	36
3.6.1 Test d'hypothèse.....	36
3.6.2 La statistique du test.....	37
3.7 Application au test de la loi HNBUE.....	40
Conclusion.....	41

CHAPITRE 4 : Application sur le test de comparaison stochastique à l'ordre deux

4.1 Application de comparaison convexe.....	43
4.2 Test de comparaison convexe	43
4.2.1 Les hypothèses de test	43
4.2.2 La statistique de test.....	44
4.3 Exemple un.....	44
4.4 Exemple deux.....	48
Conclusion	52
Conclusion générale	54
ANNEXE.....	56
BIBLIOGRAPHIE.....	63



Chapitre 1

est un chapitre introductif consacré à certaines
généralités sur la dominance stochastique de
second ordre

Introduction

Pour comparer des variables aléatoires, on utilise le critère de la variance considérée comme risque, ou bien, on utilise le critère moyenne-variance, ces deux critères ne permettent pas toujours de comparer ces variables aléatoires, par exemple lorsqu'il s'agit de risque, de gain et de durée de vie ou de bon fonctionnement dans le domaine de la fiabilité. Par exemple, on a X, Y deux variables aléatoires, si on compare par le critère moyenne variance, on choisit X , et si on utilise comparaison stochastique, on choisit Y .

Pour améliorer cette comparaison, on considère une autre approche de comparabilité dans le sens stochastique (dominance stochastique dans les sciences économiques).

La comparabilité stochastique joue un rôle important dans le domaine de la fiabilité. Dans ce domaine, on s'intéresse en particulier, à la comparabilité convexe (qui correspond dans les sciences économiques et actuarielles à la dominance stochastique du second ordre).

Dans le domaine économique, plusieurs méthodes ont été proposées pour tester la dominance stochastique de second ordre. On cite en particulier :

- Deshpande / Singh (1985) ont proposé un test statistique d'un échantillon indépendant de type intégral pour la dominance stochastique de second ordre DS2. Ils ont supposé que la fonction de distribution F_0 , est connue. Ils ont testé $H_0 : F = F_0$ contre l'alternative de dominance avec F_0 connue, le test statistique est à double intégrale :

$$T_2 = \int_{-\infty}^{\infty} \int_{-\infty}^x (\hat{F}_X(u) - F_0(u)) du dF_0(x)$$

- Le test a de bonne propriété d'efficacité, mais il est limité d'utilisation car F_0 est rarement connu et les données doivent être indépendantes. Les économistes traitant l'hypothèse H_1 : « F_X domine à l'ordre deux F_Y » de deux échantillons en comparant

$$\int_{-\infty}^x F_{m,X}(u) du \text{ et } \int_{-\infty}^x F_{n,Y}(x) \text{ pour chaque } x, \text{ où } F_{m,X}(u) \text{ et } F_{n,X}(u) \text{ sont les fonctions de}$$

distribution estimée. Des algorithmes efficaces ont été développés pour comparer deux échantillons de grandes tailles.

- Kaur, Rao et Singh (1994) et Anderson (1996) et Davidson et Duclos (2000) ont proposé un test à l'ordre s . Leur test peut être étudié pour comparer les hypothèses suivantes:

$$H_0: X \succcurlyeq Y \text{ contre } H_1: X > Y \text{ ou } H'_0: Y \succcurlyeq X \text{ contre } H'_1: Y > X$$

On suppose que $[0,1]$ est le support de X et Y donc les statistique deviennent respectivement;

$$Z_M^s = \inf \{T^s(x)\} \quad \text{et} \quad Z_M^{s*} = \inf \{-T^s(x)\} \quad \text{avec} \quad T^s(x) = \frac{\hat{F}_X^s(x) - \hat{F}_Y^s(x)}{\sqrt{\hat{V}^s(x)}}$$

Kaur, Rao et Singh (1994) ont considéré le cas particulier de $s = 2$, c.-à-d., le cas de dominance stochastique de second ordre. La statistique inf proposée est asymptotiquement équivalente à $T^2(x)$. Ils ont proposé un algorithme pour trouver le minimum de cette statistique (le théorème 2.3) (KRS 1994). Cette statistique sera utilisée dans l'application que nous réalisons dans cette thèse.

- Dans le **chapitre 2** on s'intéresse aux notions d'ordre de comparabilité stochastique qui sont utilisables dans le domaine de la fiabilité. On a étudié aussi les lois non paramétriques (les classes de vieillissement de distributions de la durée de vie). Ces lois ont été introduites plutôt en théorie de la fiabilité.

- le **chapitre 3**, porte sur l'étude des tests non paramétriques. On décrit en particulier : le test Kolmogorov-Smirnov(KS) qui est basé sur la différence maximale entre les deux fonctions de distribution empiriques, le test de Wilcoxon-Mann Whitney qui est de meilleure performance que celui de KS, il n'utilise pas les valeurs observées mais, il utilise le rang des observations. Ces tests non paramétriques, ne donnent pas toujours des bons résultats. Pour améliorer ces résultats, on va passer à la comparabilité dans un sens stochastique, c'est à dire comparaison de la distance entre les deux distributions

empiriques, ces tests s'appellent tests de comparabilité stochastique. Plusieurs statistiques ont été proposées, pour cela, on s'est intéressé au test basé sur la statistique de KRS (1994).

En utilisant ce test statistique pour construire une autre statistique qui permet de détecter l'appartenance à la classe *HNBUE*.

- **Les chapitres 2 et 3** qui représentent l'apport essentiel de notre travail, où on se concentre sur la comparaison des tests qu'on a étudié précédemment. On trouve le meilleure test. Ces résultats, ont permît ensuite, dans le **chapitre 4** d'étudier l'efficacité de test de comparaison convexe pour un type le loi non paramétrique : la classe HNBUE.

Chapitre 2

porte sur l'étude :

- des notions de comparaison stochastique qui sont utilisables dans le domaine fiabilité.
- des lois non paramétriques-

Chapitre 2 Comparaison stochastique et les lois non paramétriques

Dans ce chapitre on définit la notion de comparaison stochastique en général. On se limitera aux notions de comparaison convexe (dominance stochastique à l'ordre deux).

2.1 Comparaison stochastique

Bien que très populaire, le critère moyenne-variance ne suffit pas toujours à comparer deux variables aléatoires, et peut même conduire à des aberrations. Pour améliorer cette comparaison on suppose une autre approche de comparaison stochastique.

Cette approche de comparaison stochastique est non paramétrique et contient plusieurs avantages importants. Elle est basée sur les fonctions de répartition empirique, ou bien sur les fonctions de répartition de quantile empirique. En suite, on compare (maximise) la différence entre les deux fonctions de répartition empiriques, on fait des suppositions concernant la fonction de répartition et la fonction d'utilité. Une comparaison dans un sens stochastique peut être vue comme un ordre partiel sur un ensemble de fonctions de répartition cumulé. Il existe plusieurs types de comparaison stochastiques : nous commençons tout d'abord par présenter la comparaison stochastique, largement utilisée en fiabilité, pour comparer des variables aléatoires réelles représentant la durée de bon fonctionnement.

2.2 La fonction d'utilité

La théorie de Von-Neumann et de Marganstrem du comportement économique suppose que chaque individu possède une relation de préférence indépendante définie sur l'espace des variables aléatoires. Cette relation a une représentation d'utilité. Ceci signifie qu'étant données les variables aléatoires quelconques X et Y (non dégénérées) ayant les fonctions de répartition (F_X, F_Y) , un individu préfère X à Y si :

$$E(u(X)) \geq E(u(Y)) \quad (1)$$

Chapitre 2 Comparaison stochastique et les lois non paramétriques

Définition 2.1

Une fonction d'utilité est une fonction à valeurs réelles définie sur l'ensemble des nombres réels (représentant les résultats possibles de la variable aléatoire), et qui mesure la valeur relative des résultats.

La fonction d'utilité espérée d'une variable aléatoire X est définie par:

$$E[u(X)] = \int_{-\infty}^{\infty} u(\alpha) dF_X(\alpha),$$

avec F_X la fonction de répartition de X .

Dans le cas discret, si la variable aléatoire prend les valeurs $\alpha_1, \alpha_2, \dots, \alpha_T$, avec des probabilités $p_1, p_2, \dots, p_T$, alors l'utilité espérée est

$$E[u(X)] = p_1 u(\alpha_1) + p_2 u(\alpha_2) + \dots + p_T u(\alpha_T)$$

Il y a beaucoup de situations où la fonction d'utilité d'une décision appartient à la classe suivante de fonction d'utilité :

$$U = \left\{ u; \mathbb{R} \rightarrow \mathbb{R}, \frac{d}{dx} u(x) \geq 0, \frac{d^2}{dx^2} u(x) \leq 0, \forall x \right\}.$$

Dans la classe U , les fonctions d'utilité sont croissantes concaves (la dérivée impaire est positive et la dérivée paire est négative). Cette classe est connue comme classe des fonctions d'utilité d'aversion pour le risque (la seconde dérivée est négative). L'ensemble des fonctions exponentielles négatives constituent une base pour cette famille de fonctions d'utilité et aussi les fonctions logarithmes.

On peut écrire la fonction d'utilité comme suit :

La fonction d'utilité logarithmique: $u(x) = \ln(x)$, vérifie :

$$u'(x) = (1/x)$$

$$u''(x) = -(1/x^2)$$

Chapitre 2 Comparaison stochastique et les lois non paramétriques

La fonction d'utilité exponentielle $u(x) = -\exp(-\lambda x)$ pour $\lambda > 0$ vérifie :

$$u'(x) = \lambda \exp(-\lambda x)$$

$$u''(x) = -\lambda^2 \exp(-\lambda x)$$

La dérivée n^{ème} $u^{(n)}(x)$ est positive pour n impair, et négative pour n pair.

Remarque

- On suppose dans la suite que la fonction d'utilité est deux fois dérivable sur son support.

Caractéristiques

- La concavité est une condition nécessaire pour que l'inégalité $E(u(X)) \geq E(u(Y))$ (1) soit vérifiée.
- Le critère de l'utilité moyenne est plus général que le critère espérance-variance.

2.3 Aversion au risque :

Définition 2.2

Soit X une variable aléatoire réelle. Un individu averse au risque vérifie :

$$u(E(X)) > E(u(X)) \quad (2)$$

L'inégalité (2) porte le nom d'inégalité de Jensen.

Propriété 2.1

L'inégalité de Jensen est liée au concept de l'ordre convexe (lorsque l'inégalité est inversée).

L'inégalité de Jensen n'est pas vérifiée pour toute fonction d'utilité concave.

Chapitre 2 Comparaison stochastique et les lois non paramétriques

Remarque

- L'aversion pour le risque se traduit par la concavité de la fonction d'utilité.

2.4 Comparaison stochastique dans le domaine de la fiabilité

Les différentes notions d'ordre stochastique sont utilisées dans d'autres domaines que les sciences économiques, par exemple la théorie de fiabilité.

Dans le sens fiabilité on compare la durée de bon fonctionnement de deux systèmes S_1, S_2 de durées de vie X, Y , en connaissant les principales caractéristiques de fiabilité de chacun des systèmes. On dit qu'un système est plus fiable qu'un autre si sa caractéristique de fiabilité est meilleure.

2.4.1 Caractéristiques de fiabilité

a) La comparaison stochastique

Soient X et Y des variables aléatoires réelles,

Définition 2.3: La comparaison stochastique est définie par :

$$X \geq_{st} Y \Leftrightarrow E[u(x)] \leq E[u(y)] \quad \forall u: \mathcal{R} \rightarrow \mathcal{R}^+ \text{ Croissante continue.}$$

Théorème 2.1 [25]

Les variables aléatoires X, Y sont classées pour la relation de comparaison stochastique

$X \geq_{st} Y$ si et seulement si l'inégalité suivante est vérifiée pour toute fonction u croissante

$$E(u(X)) \leq E(u(Y)).$$

Chapitre 2 Comparaison stochastique et les lois non paramétriques

La comparaison stochastique peut également être caractérisée à l'aide des fonctions de répartition :

Propriété 2.2:

$$X \underset{st}{\geq} Y \Leftrightarrow F_X(x) \leq F_Y(x), \quad \forall x \in \mathcal{R}$$

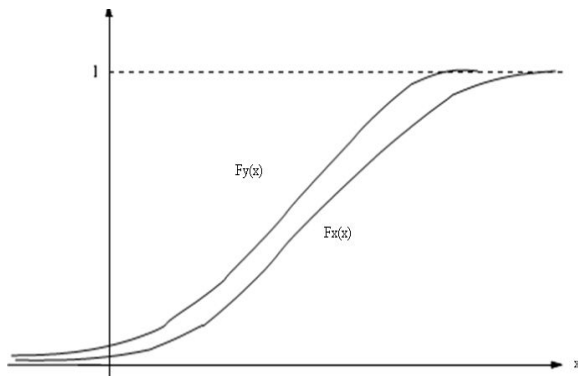


Figure 2.1 Exemples de comparaison stochastique

Propriété 2.3

On peut définir la comparaison stochastique à l'aide des fonctions quantiles.

Soient Q_X, Q_Y les quantiles de X, Y .

Théorème 2.8 [21] : $X \underset{st}{\geq} Y \Leftrightarrow Q_X(p) \geq Q_Y(p), \quad \forall p \in]0,1]$

b) la Comparaison en moyenne

Définition 2.4 Soient X, Y deux variables aléatoires, on dit que X est plus fiable en moyenne par rapport à Y ($X \geq_m Y$) si $E(X) \geq E(Y)$.

c) La comparaison en moyenne résiduelle

Chapitre 2 Comparaison stochastique et les lois non paramétriques

Définition 2.5 Soit X, Y deux variables aléatoires, on dit que X est plus fiable à l'usage par rapport à Y ($X \geq_{mr} Y$) si $m_1(t) \geq m_2(t)$ pour tout t positif.

On note $m(t) = E(X - t / X > t) = \int_t^\infty [\exp(-\int_0^u h(v)dv)] du$. Chacune de ces définitions ci-dessus est raisonnable pour les applications mais privilégie un aspect particulier du bon fonctionnement de l'élément étudié.

2.4.2 D'autres notions de comparaison stochastique

a) Comparaison convexe

Définition 2.6

Soient X, Y deux variables aléatoires tel que: $E[u(X)] \leq E[u(Y)]$ (1) $\forall u : \mathcal{R} \rightarrow \mathcal{R}$ convexe

On dit que X est inférieur à Y dans l'ordre convexe et on note $X \leq_{cx} Y$.

Théorème 2.2 [25]

Soient X, Y deux variables aléatoires telle que $E(X)=E(Y)$. On a que $X \leq_{cx} Y$, si et seulement si

$$E|X - a| \leq E|Y - a| \iff \int_{-\infty}^x F_X(u) du \leq \int_{-\infty}^x F_Y(u) du \quad \forall a \in \mathcal{R}$$

Démonstration

Soit X une variable aléatoire et u_1, u_2 deux fonctions convexes, définies par:

$$\begin{cases} u_1(x) = (x - a)^+ \\ u_2(x) = (a - x)^+ \end{cases}$$

$$\text{Si } Y \leq_{cx} X \text{ alors } \begin{cases} E[(X - a)^+] \leq E[(Y - a)^+] \\ E[(a - X)^+] \leq E[(a - Y)^+] \end{cases} \quad (1) \quad \forall a$$

Chapitre 2 Comparaison stochastique et les lois non paramétriques

On intègre les deux expressions quand l'expression (1) est réalisée et on trouve

$$\int_x^{+\infty} \bar{F}_X(u) du \leq \int_x^{+\infty} \bar{F}_Y(u) du \quad \forall x \in \mathfrak{R} \quad \text{et}$$

$$\int_{-\infty}^x F_X(u) du \leq \int_{-\infty}^x F_Y(u) du \quad \forall x \in \mathfrak{R}$$

On suppose que les deux intégrales sont calculables.

a 1) Caractéristique de comparaison convexe

Théorème 2.3 [25]

Soient X, Y deux variables aléatoires non négatives, avec l'égalité des moyennes.

1. $X \leq_{cx} Y$ si et seulement si $\int_x^{+\infty} \bar{F}_X(u) du \leq \int_x^{+\infty} \bar{F}_Y(u) du$
2. $X \leq_{cx} Y$ si et seulement si $\int_{-\infty}^x F_X(u) du \leq \int_{-\infty}^x F_Y(u) du$

Introduisons la fonction F_X^2 (fonction de comparaison) qui est la primitive de la fonction de répartition F_X

Définition 2.7 : On définit

$$F_X^2 : \mathfrak{R} \rightarrow \mathfrak{R}^+ \\ x \rightarrow \int_{-\infty}^x F_X(u) du$$

Chapitre 2 Comparaison stochastique et les lois non paramétriques

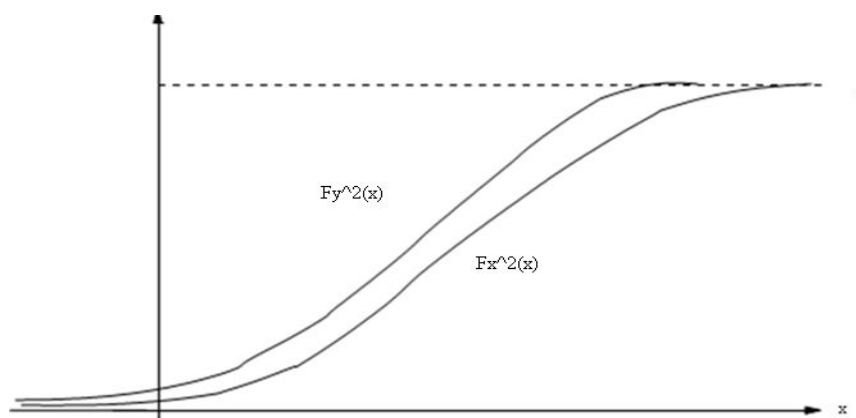


Figure 2.2. Exemples de comparaison convexe

- La comparaison convexe, permet de classer deux variables aléatoires X, Y . ceci se fait à l'aide d'un graphe de comparaison lorsque les courbes de $F_X^2(x)$ et $F_Y^2(x)$ ne se croisent pas.

Théorème 2.4 [21] : Pour tout X tel que $E[X] < +\infty$ on a

$$i. \quad F_X^{-2} = Q_X^2$$

$$ii. \quad F_X^2 = Q_X^{-2}$$

Le Q_X^2 est l'intégrale de Q_X

Théorème 2.5 [21] : $X \leq_{cx} Y \Leftrightarrow Q_X^2(p) \geq Q_Y^2(p), \quad \forall p \in]0,1]$

b) Comparaison concave

Définition 2.8 [25]

Soient X, Y deux variables aléatoires. On dit que $X \leq_{cv} Y$ si $E[u(X)] \leq E[u(Y)]$

$\forall u : \mathcal{R} \rightarrow \mathcal{R}$ concave. On montre que $X \leq_{cv} Y$ si et seulement si $Y \leq_{cx} X$. L'ordre

concave et l'ordre convexe sont deux notions inséparables.

Chapitre 2 Comparaison stochastique et les lois non paramétriques

Remarque 1

$$3. \quad Y \leq_{cv} X \text{ si et seulement si } \int_x^{+\infty} \bar{F}_X(u) du \leq \int_x^{+\infty} \bar{F}_Y(u) du$$

$$4. \quad Y \leq_{cv} X \text{ si et seulement si } \int_{-\infty}^x F_X(u) du \leq \int_{-\infty}^x F_Y(u) du$$

c) Ordre de Lorenz

La comparaison convexe est également connue comme comparaison généralisée *de Lorenz*. La courbe de Lorenz est utilisée dans le domaine économique pour mesurer l'inégalité entre deux distributions (l'inégalité de revenu). Soient X, Y deux variables aléatoires non négatives et soient F_X, F_Y deux distributions associées. Soient les fonctions inverses F_X^{-1}, F_Y^{-1} respectivement, et les quantiles Q_X, Q_Y de X, Y , et μ_X la moyenne de la variable aléatoire X .

On note L_X, L_Y les courbes de Lorenz de X, Y .

Avec,

$$L_X(p) = \frac{\int_0^p F_X^{-1}(t) dt}{\int_0^1 F_X^{-1}(t) dt} = \frac{1}{\mu_X} \int_0^p Q_X(t) d(t) = Q_X^2(p)$$

En effet, on a

$$L_X(p) = \frac{1}{\mu_X} E(1_{\{X \leq Q_X(p)\}} X) = \frac{1}{\mu_X} \int_{(-\infty, Q_X(p)]} t dF(t) = \frac{1}{\mu_X} \int_0^p Q_X(t) d(t) \text{ avec } p \in [0, 1]$$

La courbe de Lorenz généralisée est par définition

$$GL_X(p) = \mu_X L_X(p) = \int_{(-\infty, Q_X(p)]} t dF(t) = \int_0^p Q_X(t) d(t) = Q_X^2(p)$$

$$\lim_{p \rightarrow 1} L_X(p) = 1, \quad \lim_{x \rightarrow 0} L_X(p) = 0$$

Chapitre 2 Comparaison stochastique et les lois non paramétriques

L'expression ci-dessus $GL_X(p) = Q_X^2(p)$ montre que la comparaison convexe et la comparaison de Lorenz généralisé sont équivalentes. Si X et Y ont la même moyenne, ces relations de comparaison sont également équivalentes à la comparaison de Lorenz.

Si $\frac{X}{\mu_X} \leq_{cx} \frac{Y}{\mu_Y}$, on dit que la comparaison convexe est équivalente à l'ordre de Lorenz.

Le théorème ci-dessous démontre cette équivalence.

Théorème 2.6 [25]

Soient X, Y deux variables aléatoires non négatives, avec égalité des moyennes. On a $X \leq_{Lorenz} Y$ (équivalent à $X \leq_{cx} Y$, avec $E(X)=E(Y)$, si et seulement si $L_X(x) \geq L_Y(x) \quad \forall x \in [0,1]$.

d) Comparaison monotone convexe (concave) [25]

Soient X, Y deux variables aléatoires telles que: $E[u(X)] \leq E[u(Y)] \quad \forall u: \mathbb{R} \rightarrow \mathbb{R}$ croissante convexe (concave) (2). On dit donc que X est inférieure à Y et « moins variable» (inférieur et «plus variable»), on note $X \leq_{icx} Y$ ($X \leq_{icv} Y$).

On peut définir aussi la comparaison convexe décroissante (concave décroissante) lorsque $E[u(X)] \leq E[u(Y)] \quad \forall u: \mathbb{R} \rightarrow \mathbb{R}$ décroissante convexe (concave). On note $X \leq_{dcx} Y$ ($X \leq_{dcv} Y$).

Théorème 2.7 [25]

Soient X, Y deux variables aléatoires de fonctions des distributions F_X, F_Y , et $\bar{F}_X, \bar{F}_Y$ leurs fonctions de fiabilité alors:

1. $X \leq_{icx} Y$ si et seulement si $\int_x^{+\infty} \bar{F}_X(u) du \leq \int_x^{+\infty} \bar{F}_Y(u) du \quad \forall x$

Chapitre 2 Comparaison stochastique et les lois non paramétriques

$$2. \quad X \leq_{icv} Y \text{ si et seulement si } \int_{-\infty}^x F_X(u) du \leq \int_{-\infty}^x F_Y(u) du \quad \forall x$$

Ces inégalités concernent toutes les fonctions convexes et pas seulement les fonctions croissantes convexes.

Remarque

Si deux variables aléatoires ont même moyenne, alors comparaison convexe est équivalent à la comparaison convexe croissante. Le théorème suivant démontre cette remarque.

Théorème 2.8 [25]

Soient X, Y deux variables aléatoires, avec des moyennes finies, et $E(X)=E(Y)$

a) Si $X \leq_{icx} Y$ alors $X \leq_{cx} Y$

b) Si $X \leq_{icv} Y$ alors $Y \leq_{cx} X$

2.5 Lois non paramétriques

Introduction

Un aspect important de l'analyse de fiabilité est donné par la distribution de durée de vie qui peut décrire le comportement de vieillissement d'élément (un moteur, une composante électronique, une ampoule, etc.)

Les notions de vieillissements sont utilisées pour expliquer comment fonctionne l'élément utilisé selon différents critères et pour classer ses propriétés de vieillissement positivement et négativement. Ces notions correspondent au taux de défaillance estimé décroissant et croissant, respectivement, quand il existe.

Les classes de vieillissement de distributions de la durée de vie sont souvent basées sur la comparaison entre les durées de vie de fonctionnement de l'élément nouveau et usagé. Les distributions qui ont ces propriétés s'appellent les lois non paramétriques. Ces lois ont été introduites plutôt en théorie de fiabilité. Pour des raisons de commodité, nous utiliserons

Chapitre 2 Comparaison stochastique et les lois non paramétriques

également la variable t plutôt que x , pour souligner le fait que cette variable est temporelle ; de plus nous travaillerons plutôt avec la fiabilité $\bar{F}_X(t) = 1 - F_X(t)$ au lieu de $F(t)$.

2.5.1 Propriété de vieillissement (les classes de fiabilité)

1) La distribution *IFR*

Notons : $h(t) = \frac{f_X(t)}{1 - F_X(t)}$ le risque instantané (le taux de défaillance).

Définition 2.9 :

La variable aléatoire X de fonction de répartition F est dite *IFR* (taux de défaillance croissante), si elle vérifie l'une des propriétés ci-dessous.

Propriété 2.4 :

- i. Le risque instantané est monotone croissant.
- ii. La fonction $\text{Log} \bar{F}_X(t)$ est concave sur $\{t : \bar{F}_X(t) > 0\}$.
- iii. Le risque cumulé $H(t) = -\text{Log} \bar{F}_X(t)$ est convexe sur $\{t : \bar{F}_X(t) > 0\}$.
- iv. $\bar{F}_X(t) = \frac{\bar{F}_X(t+x)}{\bar{F}_X(x)} = \bar{F}_X(x/t)$ est décroissante en $t \geq 0$.

2) La distribution *IFR (2)* [19](increasing failure rate of second order)

Définition 2.10

la fonction de répartition F est dite *IFR(2)* si et seulement si

$$\int_0^x \frac{\bar{F}_X(u+s)}{\bar{F}_X(s)} du \geq \int_0^x \frac{\bar{F}_X(u+t)}{\bar{F}_X(t)} du \text{ pour } x \geq 0, t \geq s$$

Définition 2.11 :

Chapitre 2 Comparaison stochastique et les lois non paramétriques

La variable aléatoire X de fonction de répartition F est dite *DFR* (défaillance décroissant) si $\bar{F}_X(x/t)$ est croissante en t , et si aussi $-\log(\bar{F}_X(x))$ est concave, et si aussi le risque instantané est décroissant.

3) La distribution *IFRA* :

Définition 2.12:

La variable aléatoire X de taux de défaillance h , est dite *IFRA*, ou à taux de défaillance croissant en moyenne si elle vérifie l'une des propriétés ci-dessous.

Propriétés 2.4:

Notons: $\bar{F}_X(t) = e^{-H(t)}$ avec $H(t) = \int_t^{+\infty} h(u)du$ est le risque cumulé sur l'intervalle $[0, 1]$.

- i. La fonction qui à $t > 0$ associe : $\frac{1}{t} \int_t^{+\infty} h(u)du = \frac{H(t)}{t} = -\frac{1}{t} \log(\bar{F}_X(t))$ est décroissante en t ;
- ii. La fonction $(\bar{F}_X(t))^{1/t}$ est croissante en t ;
- iii. $\forall t > 0, \forall a \in [0,1], \bar{F}_X(at) \geq \bar{F}_X(t)^a$

4) La distribution *NBU* :

Définition 2.13:

La variable aléatoire X de fiabilité $\bar{F}_X$ est dite *NBU* ou encore *neuve mieux qu'usagée* si elle vérifie l'une des propriétés ci-dessous.

Propriétés 2.5:

- i. $\bar{F}_X(t+x) \leq \bar{F}_X(t)\bar{F}_X(x), \quad \forall t, x \geq 0$
- ii. $\log(\bar{F}_X(t+x)) \leq \log(\bar{F}_X(t)) + \log(\bar{F}_X(x))$ pour $t, x \geq 0$.

Chapitre 2 Comparaison stochastique et les lois non paramétriques

iii. en notant $H(x) = \int_0^x h(u)du = \log\left(\frac{1}{\overline{F}_X(t)}\right) \quad \forall x > 0$, la fonction de risque cumulé:

$$H(t+x) \geq H(t) + H(x) \text{ (H est sur additive).}$$

En termes de fiabilité la propriété NBU signifie que la fiabilité d'un élément usagé d'âge x est plus petite que celle d'un élément neuf c'est-à-dire d'âge $x = 0$.

La propriété NBU correspond à la comparaison stochastique entre la durée de survie et la durée de vie : $X_t \leq_{st} X$.

5) La distribution NBUE :

Définition 2.14 :

La variable aléatoire X , de fiabilité $\overline{F}$, d'espérance $E(X) = m$ est dite : NBUE ou (neuve mieux qu'usage en moyenne, si elle vérifie l'une des propriétés ci-dessous.

Propriétés 2.6:

- i. $E(T_t) \leq E(T), \quad \forall t, 0 < t < \infty$.
- ii. La pseudo fiabilité est inférieure à la fiabilité, c'est à dire $F_Y(t) \leq \overline{F}_X(t)$ pour tout positif, où F_Y est définie par : $F_Y(t) = \frac{1}{m} \int_t^\infty \overline{F}(u)du$ et $m = E(X) = \int_0^\infty \overline{F}(x)dx$.
- iii. La fiabilité $\overline{F}$ vérifie l'inégalité : $\int_0^\infty \overline{F}(t+u)du \leq m\overline{F}(t)$
- iv. La propriété NBUE correspond à l'ordre « en moyen » : $X_t \leq_m X$

En d'autres termes, la durée de vie moyenne résiduelle d'un élément d'âge t est plus petite que celle d'un élément neuf (d'âge $t = 0$).

$$\frac{\int_t^\infty \overline{F}_X(u)du}{\overline{F}_Y(t)} = E[X - t > X] \leq E[X] \int_t^\infty \overline{F}_X(u)du$$

Chapitre 2 Comparaison stochastique et les lois non paramétriques

La variable aléatoire positive X est NWU si $F_X(x+y) \geq F_X(x)F_X(y)$, pour $x, y \geq 0$.

La variable aléatoire X non négative avec une moyenne finie est appelé NBUE si $E[X-t/X > T] \leq E[X]$ pour tout $t > 0$ est appelé NWUE si $E[X-t/X > T] \geq E[X]$ pour tout $t > 0$, si les variables aléatoires concerne la théorie de la fiabilité. Dans le prochain résultat on montre que les variables aléatoires NBUE (NWUE) sont plus petites (plus grandes) que les variables aléatoires exponentielles par l'ordre convexe.

Théorème 2.9 [25]

Si X une variable aléatoire de NBUE et $Exp(\mu)$ une loi exponentielle de moyenne μ alors :

$$X \leq_{cx} Exp(\mu) \Leftrightarrow X \geq_{cv} Exp(\mu)$$

Soit X la durée de vie de fonction de fiabilité $\bar{F}$, alors

$$\int_x^{+\infty} \bar{F}(u) du \leq \mu Exp[-x/\mu],$$

Soit $X_1, X_2, \dots, X_n$ un processus de renouvellement, et soit $N(t)$ le nombre de renouvellement dans l'intervalle $[0, t]$, on a $E(N(t)) = M(t)$.

Théorème 2.10 [19]

Soit $E(X) = \mu$, la moyenne de durée de vie d'un composant.

$$\text{si } F \text{ est NBUE, alors } \frac{t}{\mu} - 1 \leq M(t) \leq \frac{t}{\mu}$$

$$\text{si } F \text{ est NWUE, alors } M(t) \leq \frac{t}{\mu}$$

Chapitre 2 Comparaison stochastique et les lois non paramétriques

6) La distribution HNBUE

Definition 2.15

La classe des variables *HNBUE* est la classe des variables inférieures à la comparaison convexe à la loi exponentielle de même moyenne. On peut récrire cette définition à l'aide des pseudo variables.

$$F \text{ est HNBUE} \Leftrightarrow \int_x^{+\infty} \bar{F}_X(u) du \leq \int_x^{+\infty} \bar{F}_Y(u) du$$

On a aussi : la loi *HNBUE* vérifie la propriété de l'ordre convexe.

Propriété de HNBUE 2.7

▪ Soit X une variable aléatoire qui suit une loi HNBUE, et Y une variable aléatoire qui suit une loi exponentielle de paramètre $1/\lambda$. On a $E(X) = E(Y)$, donc on écrit:

$$X \leq_{cx} Y \Leftrightarrow \int_x^{+\infty} \bar{F}_X(u) du \leq \int_x^{+\infty} \exp(-\lambda x) du$$

$$X \leq_{cx} Y \Leftrightarrow \int_{-\infty}^x F_X(u) du \leq \int_{-\infty}^x (1 - \exp(-\lambda x)) du$$

$$X \leq_{cx} \exp(1/\lambda)$$

7) La distributon *HNBUE*(3) [19] (harmonic new batter than used of thrid order)

Définition 2.16

La fonction de répartition F est dite *HNBUE*(3) si et seulement si

$$\int_x^{\infty} \int_t^{\infty} \bar{F}_X(u) du \leq \mu^2 \exp\left(-\frac{x}{\mu}\right) \text{ pour tout } t \geq 0$$

Chapitre 2 Comparaison stochastique et les lois non paramétriques

8) La distribution DMRL

Définition 2.17:

Une variable aléatoire suit une loi DMRL (Decreasing Mean Residual Life = Temps résiduel moyen décroissant) (IMRL (Increasing Mean Residual Life = Temps résiduel moyen croissant)) si et seulement si : $v(t) = \int_0^\infty \bar{F}_X(x/t) dx$ est décroissante (croissante) en t , tel que : $v(t)$ est la fonction de vie moyenne résiduelle de la variable aléatoire c-à-dire $u(s) \geq v(t) (v(s) \leq v(t))$ pour $0 < s < t$.

Propriété 2.8 :

Si X est DMRL et Y IFR alors : $X+Y$ est DMRL.

$$IFR \Rightarrow IFRA \Rightarrow NBU \Rightarrow NBUE \Rightarrow HNBUE$$

$$IFR \Rightarrow IFR(2) \Rightarrow DMRL$$

$$DFR \Rightarrow DFRA \Rightarrow NWU \Rightarrow NWUE$$

$$HNBUE \Rightarrow HNBUE(3)$$

Cette chaîne d'implications doit être comprise au sens suivant. Par exemple, $IFR \Rightarrow IFRA$ signifie que la classe des distributions de type IFR est contenue dans la classe des distributions de type IFRA. Dans la littérature et du point de vue de l'interprétation, l'implication est comprise au sens où si F possède la propriété (de vieillissement) IFR, alors elle possède également la propriété IFRA.

Conclusion

Toutes les variables aléatoires sont comparables en moyenne, alors qu'il est facile de trouver des variables aléatoires incomparables au sens comparaison stochastique.

La loi HNBUE vérifie une des propriétés de comparaison convexe.

Chapitre 2 Comparaison stochastique et les lois non paramétriques

Dans le chapitre 3 on va appliquer le test de KRS(1994), pour construire un test de *HNBUE*.

Chapitre 3

Porte sur l'étude :

- de quelques tests non paramétriques
- de tests de comparaison stochastique
- tests sur une classe de loi non paramétrique : la classe HNBUE

3.1 Tests paramétriques

Définition 3.1

Un test paramétrique requiert un modèle à fortes contraintes (normalité des distributions, égalité des variances) pour lequel les mesures doivent avoir été réalisées dans une échelle au moins d'intervalle. Ces hypothèses sont d'autant plus difficiles à vérifier que les effectifs étudiés sont plus réduits.

Le modèle statistique est paramétrique c'est-à-dire que la loi p_θ de variable aléatoire X est connue, dépendant d'un paramètre θ inconnu sur lequel on définit l'hypothèse.

Exemple

Soit X_i , $i=1, \dots, n$ un échantillon indépendant de taille n de loi p_θ avec θ un paramètre multidimensionnel appartenant à $H = H_0 + H_1$.

On suppose $X \rightarrow N(m, \sigma)$. On peut tester les deux paramètres par l'hypothèse suivant :

$$H_0 : \begin{cases} m = 0 \\ \sigma \leq 0,1 \end{cases} \quad H_1 : \begin{cases} m \neq 0 \\ \sigma > 0,1 \end{cases}$$

▪ On dit paramétrique parce que on met l'hypothèse sur les paramètres de la loi des données X testées. Généralement cette loi suit une loi normale qui est garantie par le théorème centrale limite quand n tend vers l'infini. Selon $\hat{E}(X)$ suit une loi normale, et

$\frac{\sqrt{n}\hat{E}(X)}{\hat{\sigma}(X)}$ suit une loi normale $N(0,1)$.

▪ On rejette l'hypothèse nulle quand on a trop peu de données (augmentation du risque α), ou bien avec une grande taille (augmentation du risque β).

▪ On accepte l'hypothèse nulle quand on a trop de données (augmentation du risque β), ou bien quand le test est trop puissant.

3.2 Test non paramétrique

Définition 3.2

Un test non paramétrique est un test dont le modèle ne précise pas les conditions que doivent remplir les paramètres de la population dont a été extrait l'échantillon. Cependant certaines conditions d'application doivent être vérifiées. Les échantillons considérés doivent être aléatoires [lorsque tous les individus ont la même probabilité de faire partie de l'échantillon] et simples [tous les individus qui doivent former l'échantillon sont prélevés indépendamment les uns des autres], et éventuellement indépendants les uns des autres [emploi de tables de nombres aléatoires].

On va présenter quelques exemples de tests non paramétriques.

3.2.1 Test Kolmogorov -Smirnov (K.S)

C'est un test d'adéquation qui est fondé sur les statistiques de F_n , et F assimilable à des distances ou des « pseudo distances » entre lois de probabilité.

Soit la statistique $K_n = \sup_x |F_n(x) - F_0(x)| = \max(K_n^+, K_n^-)$

$$K_n^+ = \sup_x (F_n(x) - F_0(x))$$

$$K_n^- = \sup_x (F_0(x) - F_n(x))$$

On peut aussi d'après le type de ce test comparer deux distributions (comparaison de deux échantillons).

Le test de Kolmogorov Smirnov est utilisée pour tester l'hypothèse nulle de distributions égales contre l'alternative très générale que les distributions sont inégales (comparer la distribution du premier échantillon avec celle du deuxième). Ce test est basé sur la différence maximale entre les deux fonctions de distribution empiriques

Définition 3.3

Soient $X_1, X_2, \dots, X_n$, $Y_1, Y_2, \dots, Y_m$ deux échantillons indépendants des lois F_n, F_m respectivement, lois supposées toutes deux continues. On note $\hat{F}_n, \hat{F}_m$ les fonctions de répartition estimées associées à ces échantillons. Le test de Kolmogorov Smirnov de l'hypothèse $H_0 : \langle F_n = F_m \rangle$ vs $H_1 : \langle F_n \neq F_m \rangle$ est construit à partir de la statistique :

$D_{m,n} = \max_{X_i} |\hat{F}_n(x) - \hat{F}_m(x)| = \max_{X_i, Y_j} |\hat{F}_n(x) - \hat{F}_m(x)|$. La région critique de seuil α étant, pour tout α appartient à $]0,1[$, tel que $P_0(D_{m,n} \geq \lambda_\alpha(F_n)) = \alpha$

Un des intérêts du test de Kolmogorov Smirnov réside dans le fait que la valeur de $\lambda_\alpha(F_n)$ est indépendante de F .

Si $D_{m,n}$ « grand », les distributions F_n, F_m sont significativement différentes.

Si $D_{m,n}$ “petit”, on ne peut pas les distinguer à partir des échantillons $X_1, X_2, \dots, X_n$ $Y_1, Y_2, \dots, Y_m$, donc on “accepte” $H_0 : F_n = F_m$

Théorème 3.1 [5] :

Soient $X_1, X_2, \dots, X_n$, $Y_1, Y_2, \dots, Y_m$ deux échantillons indépendants des lois F_n, F_m respectivement, lois supposées toutes deux continues. Soient $\hat{F}_n, \hat{F}_m$ les fonctions de répartition empiriques associées à ces échantillons et soit $D_{m,n}$ la statistique de Kolmogorov-Smirnov correspondante. Alors, pour tout entier k strictement positif et en

$$\text{posant } d=k/n \text{ on a } P_0(D_{n,n} \geq d) = 2 \sum_{j=1}^{\lfloor n/K \rfloor} (-1)^{j+1} \frac{(n!)^2}{(n-jk)!}$$

où P_0 désigne la loi de probabilité de $D_{m,n}$ sous H_0 , pour $n=m$

Pour $n \neq m$ l'expression de cette loi est beaucoup plus complexe. Pour les grandes valeurs de n et m , on peut utiliser la loi asymptotique de $D_{m,n}$. Cette loi est due à Kolmogorov(1933) et Smirnov(1939). On a

$$\lim_{m,n \rightarrow \infty} P_0(mn/N)^{1/2} D_{m,n} \leq d = 1 - 2 \sum_{k=1}^{\lfloor n/K \rfloor} (-1)^{k+1} \exp(-2k^2 d^2)$$

Cette loi asymptotique ne peut servir d'approximation pour la loi exacte que pour les valeurs m et n assez élevées car la convergence est lente (Hodges (1958)).

Lorsque l'alternative est de type $H_1: \langle F_n \geq F_m \rangle$, on peut spécialiser le test de Kolmogorov-Smirnov en considérant la statistique

$$D_{m,n}^+ = \max_{x \in \mathcal{R}_i} (\hat{F}_n(x) - \hat{F}_m(x)).$$

On rejette l'hypothèse $H_0: \langle F_n = F_m \rangle$ la statistique $D_{m,n}^+$ prend des valeurs trop grandes.

Ce test est également un test convergent pour l'hypothèse alternative $H_1: \langle F_n \geq F_m \rangle$. On construit un test analogue pour l'hypothèse alternative $\langle F_n \leq F_m \rangle$.

Lorsque $m=n$, on obtient, pour tout entier k strictement positif, en posant $d=k/n$,

$$P_0(D_{n,n}^+ \geq d) = \frac{(n!)^2}{(n-k)!(n+k)!}.$$

Les tests de Wilcoxon et Mann Whitney testent la dominance de premier ordre stochastique.

Le test de Wilcoxon tient compte de la direction de la différence, mais pas de l'importance de la différence entre chaque paire de points.

3.2.2 Test de Wilcoxon et Mann Whitney

Définition 3.4

Soient X_i, Y_j $i=1 \dots m, j=1, \dots, n$ deux échantillons aléatoires ordonnés, la statistique de test

utilisée est $U_N = W_N - \frac{m(m+1)}{2} = \sum_{i=1}^m R_i - \frac{m(m+1)}{2} = \sum_{i=1}^m \sum_{j=1}^n I_R(X_i - Y_j)$; U_N varie donc de 0 à mn

et $E(U_N) = \frac{mn}{2}$, $V(U_N) = \frac{mn}{12}(m+n+1)$

U_N représente le nombre de fois où un Y est plus grand qu'un X (ou le nombre de fois où un X précède un Y dans l'échantillon global ordonné. C'est la statistique du signe calculée sur les mn différences $Y_j - X_i$.

Pour les petites valeurs de m et n ($m \leq n \leq 8$) on utilise la loi exacte de U_N et au de-là, l'approximation normale.

On utilise cette statistique pour tester l'hypothèse $H_0: \mu = 0$ contre l'hypothèse alternative $H_1: \mu \neq 0$.

Pour tester l'hypothèse H_0 contre l'hypothèse H_1 , ils ont proposé de compter le nombre de couples (X_i, Y_j) pour lesquels Y_i est plus grand que X_j . En effet, la loi F_Y est éloignée de la loi F_X , au sens de cette alternative générale, plus on doit s'attendre à avoir un grand nombre de couple (X_i, Y_j) vérifiant l'inégalité $Y_j > X_i$.

Plusieurs tests ont été proposés pour tester les classes des lois non paramétrique. On va proposer un test pour la classe HNBUE, et un autre pour la classe DMRL.

3.2.3 Test de HNBUE [19]

Singh et Kochar (1986) ont prouvé que le test basé sur la statistique :

$$T_n = \frac{1}{n} \sum_{i=1}^n \exp\left(-\frac{x_i}{\hat{x}}\right)$$

consiste à tester l'hypothèse H_0 : « F exponentiel » contre H_1 : « F HNBUE ». Une petite valeur de T_n indique que F est HNBUE.

Klar (2000) a proposé une statistique qui teste l'hypothèse H_0 : « F exponentiel » contre

H_1 : « F HNBUE », définie par

$$T_{n,a} = \frac{1}{n a^2} \sum_{j=1}^n \exp(-a y_j) - \frac{1}{a^2(1+a)}, a > 0, \quad \text{avec}$$

$$y_i = \frac{x_i}{\hat{x}_n}, \text{ pour } i = 1 \dots n$$

Pour $a=1$, $T_{n,1}$ est équivalente la statistique T_n

3.2.4 Test de DMRL [19]

Hollander et Proschan ont proposé un test V statistique de l'hypothèse nulle

« F exponentiel » contre « F DMRL », ce test est basé sur une fonction linéaire de statistique d'ordre de l'échantillon $x_{(1)} < x_{(2)} < \dots x_{(n)}$, définit par

$$V^* = \frac{V}{\hat{x}}$$

$$\text{où } V = n^{-4} \sum_{i=1}^n c_{i,n} x_i \text{ avec } c_{i,n} = \frac{3}{4}i^3 - 4ni^2 + 3n^2i - \frac{1}{3}n^3 + \frac{1}{2}n^2 - \frac{1}{2}i^2 + \frac{1}{6}i$$

Les grandes valeurs significatives de V^* indiquent que la moyenne résiduelle de vie alternative décroît, les petites valeurs significatives de V^* suggèrent que la moyenne résiduelle de vie alternative croît.

Langenberg, P et Srinivasan, R (1979) ont obtenu la distribution exacte de la statistique de test V sous l'hypothèse nulle.

3.3 Le choix de l'hypothèse

Le choix de l'hypothèse nulle de comparaison stochastique (resp. de non comparaison stochastique) maximise la probabilité de ne pas rejeter la dominance (resp. la non dominance), et donc minimise la probabilité de ne pas rejeter la non dominance (resp. la dominance).

3.4 Choix de test

Avant de choisir un test il faut tenir compte de :

- L'analyse des données (caractéristiques des données) ;
- La taille de l'échantillon ;
- L'hypothèse nulle et l'alternative qu'on veut considérer ;
- Le niveau de mesure : tout dépend des données (intervalles de confiance, statistiques de test...).

3.5 Les avantages et les inconvénients

Si dans les conditions où les tests paramétriques peuvent s'appliquer, les tests non paramétriques sont moins efficaces. Ils ont, par contre, par rapport à ces tests paramétriques de nombreux avantages et une plus grande généralité d'applications.

- Si les fonctions de distributions sont connues, on fait donc usage de cette information utilisant les tests paramétriques.
- Pour les petits échantillons, les statistiques sont en général plus faciles à calculer et l'application plus aisée que dans le cas des méthodes paramétriques (classiques).
- Les méthodes non paramétriques ne permettent pas toujours de déterminer l'intervalle de confiance.
- Les méthodes paramétriques ne peuvent pas tester plusieurs hypothèses, par contre les méthodes non paramétriques peuvent l'étudier.

3.6 Test de comparaison convexe

Notre étude est basée sur le test de Kaur-Rao-Singh (1994). La statistique de ce test est une distance (différence) entre les distributions empiriques qui est normalisée par l'écart type de deux distributions.

3.6.1 Test d'hypothèse

Pour déterminer la région critique on a besoin de calculer la loi de probabilité de la statistique sous l'hypothèse H_0 .

Soient $(X_1, X_2, \dots, X_m), (Y_1, Y_2, \dots, Y_n)$ deux échantillons aléatoires issus respectivement de F_X, F_Y . $F_{X,m}, F_{Y,n}$ sont distributions empiriques. On note $X_{(i)}, i = 1 \dots m$ et $Y_{(j)}, j = 1 \dots n$ les statistiques d'ordres des deux échantillons. On suppose que F_X, F_Y sont continues et leurs moments d'ordre deux, sont finis.

On a l'hypothèse de test de comparaison convexe :

$$\begin{aligned} H_0 : F_X^2(x) &\leq F_Y^2(x) \quad \text{pour certaines valeurs de l'intervalle} & [a, b] \\ H_1 : F_X^2(x) &> F_Y^2(x) & \forall x \in [a, b] \end{aligned}$$

où a et b sont des valeurs réelles.

Pour faire ce test on va déterminer la statistique du test.

3.6.2 La statistique du test

Kaur, Rao, Singh (1994) ont proposé la statistique :

$$Z_{m,n} = \inf_{a < x < b} Z_{m,n}(x) = \inf_{a < x < b} \frac{\hat{F}_{X,n}^2(x) - \hat{F}_{Y,n}^2(x)}{\sqrt{\frac{1}{m} s_{m,U}^2(x) + \frac{1}{n} s_{n,V}^2(x)}} = \inf_{a < x < b} \frac{\bar{U}(x) - \bar{V}(x)}{\sqrt{\frac{1}{m} s_{m,U}^2(x) + \frac{1}{n} s_{n,V}^2(x)}}$$

$$Z_{m,n}(x) = \frac{A_{m,n}(x)}{D_{m,n}(x)} + \frac{B_{m,n}(x)}{D_{m,n}(x)},$$

$$A_{m,n}(x) = \int_{-\infty}^x \sqrt{m} [F_{X,m}(y) - F_{Y,n}(y)] dy - \frac{\sqrt{m}}{\sqrt{n}} \int_{-\infty}^x \sqrt{n} [F_{X,m}(y) - F_{Y,n}(y)] dy$$

$$= \sqrt{m} [\bar{U}(x) - E(\bar{U}_1(x))] - \frac{\sqrt{m}}{\sqrt{n}} [\sqrt{n} (\bar{V}(x) - E(V_1(x)))]$$

$$B_{m,n}(x) = \sqrt{m} \int_{-\infty}^x (F_{X,m}(y) - F_{Y,n}(y)) dy$$

$$D_{m,n}(x) = m \left[\frac{1}{m} S_{m,U}^2(x) + \frac{1}{n} S_{n,V}^2(x) \right].$$

Proposition 1[13]

Pour x fixé et $x \in [a, b]$, et pour $i = 1, \dots, n$ et $j = 1, \dots, m$ on a l'hypothèse suivante :

$$H_{0x} : \int_{-\infty}^x F_X(u) du \leq \int_{-\infty}^x F_Y(u) du$$

$$H_{1x} : \int_{-\infty}^x F_X(u) du > \int_{-\infty}^x F_Y(u) du$$

Pour toute valeur de x fixé, nous définirons d'abord le niveau α de test asymptotique de l'hypothèse simple H_{0x} contre H_{1x} . Puis, on rejette H_{0x} au niveau α pour toute valeur de x.

$$U_i(x) = \begin{cases} x - X_i & \text{si, } X_i < x \\ 0 & \text{si, non} \end{cases}$$

$$V_j(x) = \begin{cases} x - Y_j & \text{si, } Y_j < x \\ 0 & \text{si, non} \end{cases}$$

soit

$$\overline{U}(x) = \frac{1}{m} \sum_{i=1}^n U_i(x) = \int_{-\infty}^x F_{X,n}(u) du = \hat{F}_{X,n}^2(x)$$

et

$$\overline{V}(x) = \frac{1}{n} \sum_{j=1}^m V_j(x) = \int_{-\infty}^x F_{Y,m}(u) du = \hat{F}_{Y,n}^2(x)$$

$$Z_{n,m}(x) = \frac{\overline{U}(x) - \overline{V}(x)}{\sqrt{\frac{1}{m} s_{n,U}^2(x) + \frac{1}{n} s_{m,V}^2(x)}} = \frac{\hat{F}_{X,n}^2(x) - \hat{F}_{Y,n}^2(x)}{\sqrt{\frac{1}{m} s_{m,U}^2(x) + \frac{1}{n} s_{n,V}^2(x)}}$$

$$s_{n,U}^2(x) = \sum_{i=1}^n [U_i(x) - \overline{U}(x)]^2,$$

$$s_{m,V}^2(x) = \sum_{j=1}^m [V_j(x) - \overline{V}(x)]^2$$

Pour les grandes dimensions de l'échantillon, ce test rejette H_0 si et seulement si

$$Z_{n,m} = \inf_{a \leq x \leq b} Z_{n,m}(x) > z_\alpha \text{ où } z_\alpha \text{ est la quantile d'ordre } (1-\alpha) \text{ de la loi normale.}$$

Pour calculer cette statistique on utilise l'algorithme du théorème 3.3.

Théorème 3.1 [13]

Il permet de calculer la distribution de la probabilité de la statistique du test $Z_{n,m}$.

Soient $m, n \rightarrow \infty$, tel que $\frac{m}{m+n} \rightarrow \lambda$, avec $0 \leq \lambda < 1$. et notons

$$D^2(x) = \text{var}(U_1(x)) + \frac{\lambda}{1-\lambda} \text{var}(V_1(x)),$$

si $D(x) > 0$, alors

$$p(Z_{m,n}(x) > z_\alpha) = \begin{cases} 0 & \text{si } \int_{-\infty}^x F_X(u) du < \int_{-\infty}^x F_Y(u) du \\ \alpha & \text{si } \int_{-\infty}^x F_X(u) du = \int_{-\infty}^x F_Y(u) du \\ 1 & \text{si } \int_{-\infty}^x F_X(u) du > \int_{-\infty}^x F_Y(u) du \end{cases}$$

Si $D(x) = 0$, alors $Z_{m,n}(x) = 0$, ps, pour toute valeur de m, n et $P(Z_{m,n}(x) > z_\alpha) = 0$

Théorème 3.2 [13]

Soient F_X, F_Y deux fonctions de distributions. Soient $m, n \rightarrow \infty$, tel que

$$\frac{m}{m+n} \rightarrow \lambda, \text{ avec } 0 \leq \lambda < 1; \text{ alors,}$$

- i) Pour $F_X, F_Y \in H_0$,

$$\lim_{m, n \rightarrow \infty} \sup P(\inf_{x \in [a, b]} Z_{n, m}(x) > Z_\alpha) \leq \alpha$$

- ii) S'il existe $x_0 \in [a, b]$ tel que : $\int_{-\infty}^{x_0} F_{X, m}(u) du - \int_{-\infty}^{x_0} F_{Y, n}(u) du \neq 0$, et si $\forall x \in [a, b] \setminus \{x_0\}$

$$\int_{-\infty}^x F_{X, m}(u) du > \int_{-\infty}^x F_{Y, n}(u) du, \text{ alors } \lim_{\substack{m \rightarrow \infty \\ n \rightarrow \infty}} P(\inf_{x \in [a, b]} Z_{n, m}(x) > Z_\alpha) = \alpha.$$

- iii) Pour $F_X, F_Y \in H_1$, $P(\inf_{x \in [a, b]} Z_{n, m}(x) > Z_\alpha) \rightarrow 1$. ps quand $m, n \rightarrow \infty$.

Théorème 3.3 [13]

Soient $X_i, i = 1, \dots, m$ et $Y_j, j = 1, \dots, n$ deux échantillons et soient $W_1, W_2, \dots, W_{m+n}$ des observations ordonnées de couple (X_i, Y_j)

On a pour $W_p \leq X < W_{p+1}, p = 1, 2, \dots, m+n-1$,

$$\bar{U}(x) = \frac{1}{m} \sum_{i=1}^{\ell_p} (x - X_{(i)}) \quad 0 \leq \ell_p \leq m$$

$$\bar{V}(x) = \frac{1}{n} \sum_{j=1}^{k_p} (x - Y_{(j)}) \quad 0 \leq k_p \leq n$$

ℓ_p le nombre de X_i inférieur à W_p et k_p le nombre de Y_j inférieur à W_p

$$S_{m, U}^2 = \frac{1}{m} \left[\sum_{i=1}^{\ell_p} (x - X_{(i)})^2 - m(\bar{U}(x))^2 \right]$$

$$S_{n, V}^2 = \frac{1}{n} \left[\sum_{j=1}^{k_p} (x - Y_{(j)})^2 - n(\bar{V}(x))^2 \right]$$

et

$$[D_{m,n}^2]^{\frac{1}{2}} \frac{d}{dx} Z_{m,n}(x) = b_p x + c_p$$

où

$$b_p = \left[\frac{m - \ell p}{n m^3} + \frac{n - kp}{m n^3} \right] \left[kp \sum_{i=1}^{\ell p} (X_{(i)})^2 - \ell p \sum_{j=1}^{kp} (Y_{(j)})^2 \right]$$

$$\begin{aligned} c_p = & \left[\frac{\ell p}{m} - \frac{kp}{n} \right] \left[\frac{1}{m^2} \sum_{i=1}^{\ell p} (X_{(i)})^2 + \frac{1}{n^2} \sum_{j=1}^{kp} (Y_{(j)})^2 \right] \\ & + \frac{1}{mn} \left[\frac{m - \ell p}{m^3} - \frac{n - kp}{n^3} \right] \left(\sum_{i=1}^{\ell p} X_{(i)} \right) \left(\sum_{j=1}^{kp} Y_{(j)} \right) + \frac{m - \ell p}{n m^3} \left(\sum_{i=1}^{\ell p} X_{(i)} \right)^2 \\ & - \frac{n - kp}{m n^3} \left(\sum_{j=1}^{kp} Y_{(j)} \right)^2 \end{aligned}$$

3.7 Application au test de la loi HNBUE

La loi HNBUE est caractérisée par l'ordre convexe.

Soit $Y \rightarrow \exp(\lambda)$

X est HNBUE si et seulement si

$$X \leq_{cx} Y \Leftrightarrow \int_0^x F_X(u) du \leq \int_0^x (1 - \exp(-\lambda u)) du$$

pour détecter la classe HNBUE, on va construire le test de la loi HNBUE à partir de statistique de KRS (1994).

On suppose l'hypothèse suivante :

H_0 : « F est HNBUE » contre H_1 : « F n'est pas HNBUE »

La statistique de cette hypothèse :

$$Z_m(x) = \frac{\int_0^x F_{X,m}(u) du - \int_0^x (1 - \exp(-\lambda)) du}{\sqrt{\text{var}(\hat{F}_X(u)) + \text{var}(Y)}}$$

$$\int_0^x (1 - \exp(-\lambda u)) du = x + \frac{1}{\lambda} \exp(-\lambda x) - \frac{1}{\lambda}$$

Donc on peut écrire Z_m sous la forme :

$$Z_m(x) = \frac{\int_0^x F_{X,m}(u) du - x - \frac{1}{\lambda} \exp(-\lambda x) + \frac{1}{\lambda}}{\sqrt{\text{var}(\hat{F}_X(u)) + \frac{1}{\lambda^2}}}$$

D'où la statistique du test de HNBUE est

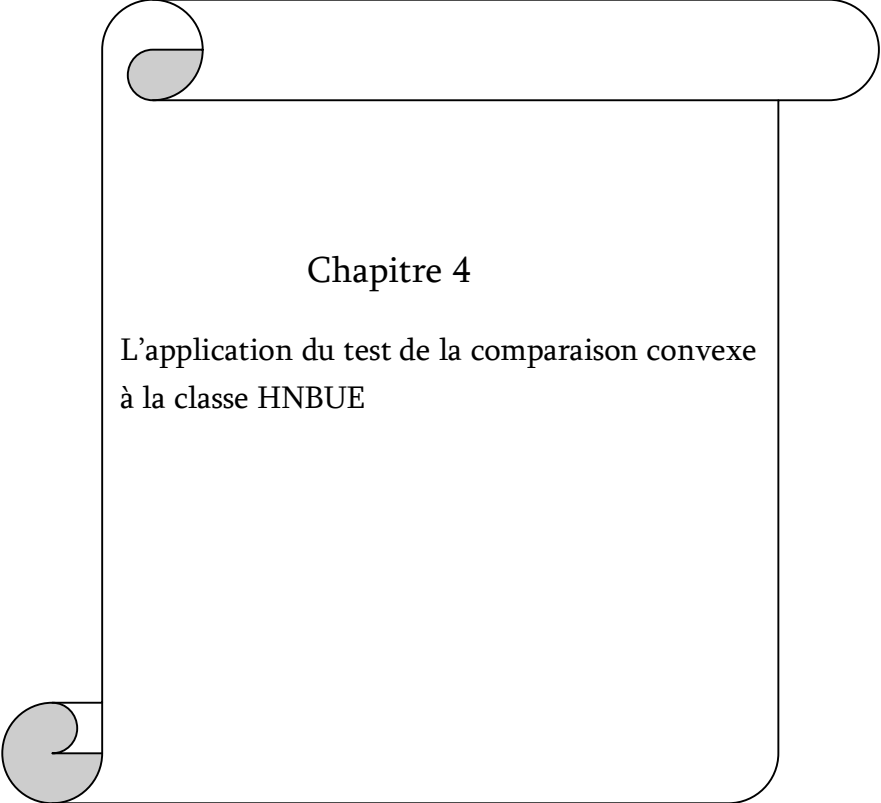
$$Z_m(x) = \frac{\int_0^x F_{X,m}(u) du - x - \frac{1}{\lambda} \exp(-\lambda x) + \frac{1}{\lambda}}{\sqrt{\text{var}(\hat{F}_X(u)) + \frac{1}{\lambda^2}}} = \frac{\int_0^x F_{X,m}(u) du - x - \frac{1}{\lambda} \exp(-\lambda x) + \frac{1}{\lambda}}{\sqrt{\frac{1}{m} [\sum_{i=1}^{\ell p} (x - X_{(i)})^2 - m(\bar{U}(x))^2] + \frac{1}{\lambda^2}}}$$

Lorsque les deux variables aléatoires ont même moyenne alors la statistique devient

$$Z_m = \frac{\sum_{i=1}^m (x - X_i) - x - \hat{E}(X) \exp\left(-\frac{1}{\hat{E}(X)} x\right) + \hat{E}(X)}{\sqrt{\frac{1}{m} [\sum_{i=1}^m (x - X_i) - (\sum_{i=1}^m (x - X_i))]^2 + (\hat{E}(X))^2}}$$

Conclusion

Le test de Kaur, Rao, Singh est un bon point de départ pour détecter l'appartenance d'une distribution à la classe de loi HNBUE où la HNBUE est l'hypothèse nulle.



Chapitre 4

L'application du test de la comparaison convexe
à la classe HNBUE

4.1 Application de comparaison convexe

Introduction

On présente dans ce chapitre un algorithme basé sur le (théorème 3.2) permettant de tester la comparaison convexe. On applique cette méthode pour détecter l'appartenance à la classe des lois non paramétriques HNBUE. L'idée est que la loi HNBUE est caractérisée par l'ordre convexe. Cela veut dire une alternative nulle correspondant à la loi exponentielle et qui est donc simple. Mais pour la valider, on va comparer ces résultats avec d'autres tests des lois non paramétrique par exemple, le test de Kolmogorov Smirnov, et le test de Wilcoxon-Mann-Whitney (comparaison entre deux distributions).

4.2 Test de comparaison convexe

Soit X une variable aléatoire de moyenne $\lambda = E(X)$, et soit Y une variable aléatoire de fonction de répartition $Exp\left(\frac{1}{\lambda}\right)$, alors X est HNBUE si et seulement si $X \leq_{cx} \exp(1/\lambda)$.

Par la suite on a :

$$X \text{ est HNBUE} \Leftrightarrow \int_0^x F(u) du \leq \int_0^x (1 - \exp(-\lambda u)) du.$$

On suppose que la variable aléatoire X suit une loi de Weibull de paramètres (α, β) et la variable aléatoire Y suit une loi exponentielle de même moyenne.

4.2.1 Les hypothèses de test :

Le test de comparaison convexe se fait par les l'hypothèse suivant

$$H_0: \int_0^x F_X(u) du \leq \int_0^x F_Y(u) du \Leftrightarrow F_X^2(x) \leq F_Y^2(x)$$

$$H_1: \int_0^x F_X(u) du > \int_0^x F_Y(u) du \Leftrightarrow F_X^2(x) > F_Y^2(x)$$

Comme les classes de HNBUE est caractérisée par l'ordre convexe, donc, on propose les l'hypothèse suivante :

$$H_0: F \text{ est HNBUE}$$

$H_1: F \text{ n'est pas HNBUE}$

Pour l'hypothèse H_1 , on peut dire qu'il n'existe pas une relation de comparaison stochastique.

4.2.2 La statistique de test

On calcule la statistique de test défini dans le chapitre 3

On va tirer deux échantillons, de loi exponentielle de taille n , et de loi HNBUE de taille m . Par la méthode de simulation, et on va étudier le comportement de la statistique sur l'intervalle choisi. Le calcul de la statistique utilisée, est dans le théorème 3.3

4.3 Exemple un

4.3.1 Etude graphique de deux fonctions

Soient X, Y , deux variables aléatoires des lois HNBUE et exponentielle de même moyenne définies comme suit :

$$F_X(x) = 1 - \exp(-x^2)$$

$$F_Y(x) = 1 - \exp\left(-\frac{7x}{2\pi}\right)$$

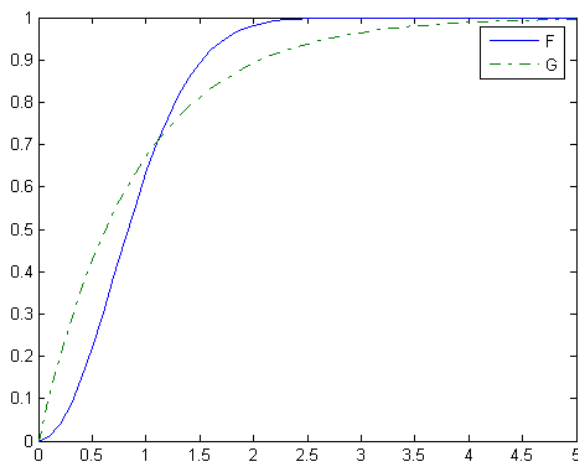


Figure 4.1: graphique de deux distributions

La figure 4.1 indique que les deux distributions se croisent en deux points :

$$x = 0, x = 1.11$$

Sur l'intervalle $[0, 1.11]$, on remarque que l'aire de la distribution F_X est au dessous de l'aire de la distribution F_Y et après le deuxième point de l'intersection, F_X est au dessus de F_Y .

On conclure, que F_X est HNBUE sur l'intervalle $[0, 1.11]$,

Pour confirmer ces résultats, on va tracer les courbes $F_X^2(x)$ et $F_Y^2(x)$

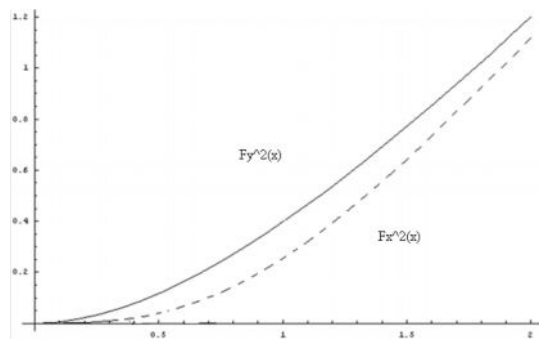


Figure 4.2: graphique de deux comparaisons stochastiques

On remarque d'après le figure 4.2 que $\int_0^x F_X(u) du \leq \int_0^x F_Y(x) du$. Donc F_X est HNBUE sur un intervalle $[0, 1.11]$.

4.3.2 Tests d'égalité de deux distributions

a 1) Test de Kolmogorov-Smirnov

Le test de Kolmogorov-Smirnov teste l'égalité de deux fonctions de distributions de deux échantillons empiriques

On a comparé deux distributions F_X est HNBUE et F_Y est exponentielle, et on a obtenu les résultats suivants

	m=27, n=34	m=54, n=64	m=120, n=130	m=230, n=350	m=350, n=430
p-value	0,002	0,0019	0,0007	-	-
k_{obs}	0,33	0,43	0,47	0,31	0,34
$k_{0.05}$	0,35	0,25	0,17	0,12	0,09

Tableau 4.1

On a d'après le tableau 4.1 la p-value, pour le seul 0.05 est inférieur à 0.05 quand les tailles échantillons augmentent, et aussi $Z > 1.67$, donc on rejette l'hypothèse $H_0 \ll F_X = F_Y \gg$, on conclut donc qu'il y a une différence entre les deux distributions

Il existe d'autre test qu'on appelle, tests de rang qui ont meilleur performance que celle de Kolmogorov-Smirnov.

Parmi les tests de rang on utilise le test Wilcoxon - Mann Whitney qui compare deux distributions.

a.2) Test de Wilcoxon - Mann Whitney

Le p-value et la statistique de test Wilcoxon Mann-Whitney

	m=27, n=34	m=54, n=64	m=120 n=130	m=230 n=350	m=350 n=430
P(0,05)	0,00013	0,000012	-	-	-
Z	5,24	5,24	7.87	9.75	10,62

Tableau4.2

D'après le tableau 2, la p-value, pour le seul $\alpha = 0.05$, est inférieur à 0.05, quand les tailles des échantillons augmentent, et aussi $Z > 1.67$, donc on rejette l'hypothèse $H_0 \ll F_X = F_Y \gg$. On conclut donc qu'il y a une différence entre les deux distributions $\ll F_X \neq F_Y \gg$.

Conclusion

D'après les tableaux ci-dessus, on conclut que la distribution F_X différente de F_Y . donc, on rejette l'hypothèse nulle (l'égalité de deux distributions), alors, on va tester la comparaison stochastique.

a.3 Test de HNBUE:

Ce test de niveau α est basé sur la statistique de KRS (1994), avec l'hypothèse nulle est HNBUE .

$H_0: F \text{ est HNBUE}$

$H_1: F \text{ n'est pas HNBUE}$

La statistique de KRS est donnée par la formule suivante :

on a calculé la statistique $\inf Z_{mn}$, et on a trouvé les résultats pour différentes tailles des échantillons.

	m=27, n=34	m=54, n=64	m=120, n=130	m=230, n=250	m=350, n=430	m=1000, n=980
Z	0, 24	0,37	0,26	0,42	0,047	0,5

Tableau 4.3

On remarque d'après le tableau 4.3 que la statistique $Z_{m,n}$ est inférieur au quantile de la loi normale au niveau 0.05. On a $Z < c = 1.67$ pour les différente taille, donc on accepte H_0 , d'où F est HNBUE

a.3.1) validation du test

La validation du test veut ce faire en montrant que la distribution F_X estimée à l'aide de l'échantillon est bien HNBUE car elle vérifiée le deuxième critère de comparabilité convexe.

D'après la définition de comparaison convexe, on dit que $X \leq_{cx} Y$ si seulement si : $E_X(u(x)) \leq E_Y(u(x))$ avec u une fonction croissante convexe. On va choisir la fonction exponentielle. Donc la fonction d'utilité est $u(x) = \exp(-2x)$, la dérivée de ce dernier est : $u'(x) = -2\exp(-2x)$, la deuxième dérivée est : $u''(x) = 4\exp(-2x)$

On a

$$E_X(u(x)) = \int_0^1 u(x) dF(x) = \int_0^1 [\exp(-2x) * 2x * \exp(-x^2)] dx = 0.21$$

$$E_Y(u(x)) = \int_0^1 u(x) dG(x) = \int_0^1 \left[\frac{7}{2\pi} \exp(-2x) * \exp\left(-\frac{7x}{2\pi}\right) \right] dx = 0.34$$

On conclut que $E_X(u(x)) \leq E_Y(u(x))$, avec $u(x)$ croissante convexe.²

Un individu averse au risque vérifie que $u(E(x)) < E_X(u(x))$, quant u est convexe

$$u(E(x)) = u(0.88) = \exp(-2 * 0.88) = 0.17$$

$$E_X(u(x)) = 0.21$$

$$\text{Donc } u(E(x)) = 0.17 < E_X(u(x)) = 0.2$$

Remarque

Ceci n'est pas une validation complète, car il faudra vérifier ceci par toutes les fonctions convexes.

4.3.2 Exemple deux

$$F_X(x) = 1 - \exp(-2x^3)$$

$$F_Y(x) = 1 - \exp(-\pi^2 x/7)$$

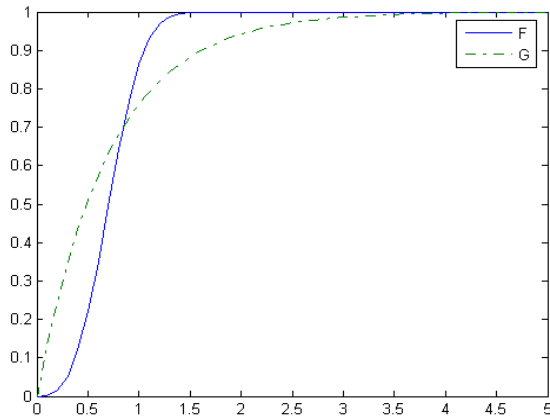


Figure 4.2: graphique de deux distributions

La figure 4.2 indique que les deux distributions se croisent en deux points :

$$x = 0, x = 0.85$$

Sur l'intervalle $[0, 0.85]$,

Sur l'intervalle $[0, 0.85]$, on remarque que l'aire de la distribution F_X est au dessous de l'aire de la distribution F_Y et après le deuxième point de l'intersection, F_X est au dessus de F_Y .

On conclure, que F_X est HNBUE sur l'intervalle $[0, 0.85]$,

Pour confirmer ces résultats, on va tracer les courbes $F_X^2(x)$ et $F_Y^2(x)$

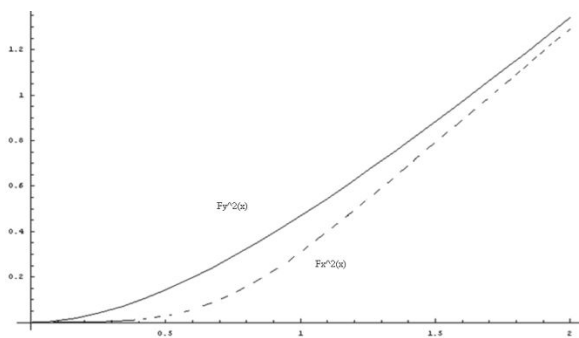


Figure 4.2: graphique de deux comparaisons stochastiques

Chapitre4 applications sur la comparaison des tests

On remarque d'après le figure 2 que $\int_0^x F_X(u)du \leq \int_0^x F_Y(x)du$.

F_X est HNBUE sur un intervalle $[0,0.85]$.

4.4.1) Tests d'égalité de deux distributions

b 1) Test de Kolmogorov-Smirnov

	m=27 n=34	m=54 n=64	m=120 n=130	m=230 n=250	m=350 n=430	m=1000 n=890
k(0,05)	0,49	0,58	0,62	0,48	0,50	0,75

Tableau 4.4

On a k_{obs} est grand, donc les deux distributions sont différents.

On rejette l'hypothèse H_0 « $F_X=F_Y$ », pour les différentes tailles, donc « $F_X \neq F_Y$ », il y a une différence entre les deux distributions.

b 2) Test de wilcoxon –Mann-Whitney

	m=27, n=34	m=54, n=64	m=120, n=130	m=230, n=250	m=350, n=430	m=1000 n=980
Z	4.20	7.01	9.911	11.47	15.32	19.98

Tableau 4.5

D'après le tableau 5, la statistique est supérieure à la valeur critique « $Z>1.67$ » de la loi normale, et aussi le p-value, pour le seuil $\alpha = 0.05$, est plus petit que 0.05, quand les tailles des échantillons augmentes, donc on rejette l'hypothèse H_0 « $F_X=F_Y$ » (l'égalité de deux distributions). On passe à tester la comparaison stochastique.

b 3) Test de HNBUE

D'après les calculs on a trouvé les résultats suivantes

	m=27, n=34	m=54, n=64	m=120, n=130	m=230, n=250	m=350, n=430	m=1000 n=980
Z	0,045	0,19	0,25	0, 43	0,49	0.50

Tableau 4.6

D'après ce tableau 6 on voit que, F_X est HNBUE.

b.3.1) validation du test

On va choisir dans cet exemple la loi exponentielle: $u(x) = \exp(-3x)$ la dérivée de cette dernière est :

$u'(x) = -3\exp(-3x)$, la deuxième dérivée est : $u''(x) = 9\exp(-3x)$

On a

$$E_X(u(x)) = \int_0^{0.8} u(x) dF(x) = \int_0^{0.8} [\exp(-3x) * 6x^2 * \exp(-2x^3)] dx = 0.13$$

$$E_Y(u(x)) = \int_0^{0.8} u(x) dG(x) = \int_0^{0.8} \left[\exp(-3x) * \frac{\pi^2}{7} * \exp\left(-\frac{\pi^2 x}{7}\right) \right] dx = 0.31$$

On conclut que $E_X(u(x)) \leq E_Y(u(x))$, avec $u(x)$ croissante convexe.

on dit donc que F est HNBUE.

Un individu averse au risque vérifie que $u(E(x)) < E_X(u(x))$, quant u est convexe

$$u(E(x)) = u(0.70) = \exp(-3 * 0.70) = 0.12$$

$$E_X(u(x)) = 0.13$$

Chapitre4 applications sur la comparaison des tests

Donc $u(E(x)) = 0.12 < E_x(u(x)) = 0.13$

Remarque

Ceci n'est pas une validation complète, car il faudra vérifier ceci par toutes les fonctions convexes.

4.4 Comparaison entre les deux exemples

On calcule la distance entre la loi HNBUE et la loi exponentielle au point de l'intersection de deux graphiques de distribution par la formule suivante :

$$\int_0^x F_{X,m}(u)du - x - \frac{1}{\lambda} \exp(-\lambda x) + \frac{1}{\lambda}$$

	Exemple1	Exemple2
Différence entre deux intégrales estimées	0.30	0.46

Tableau 4.7

D'après le tableau ci-dessus, on voit que la distance entre la loi HNBUE et l'exponentielle du premier exemple est supérieure à la distance entre la loi HNBUE et exponentielle du deuxième exemple.

Tableau de la puissance estimée d'exemples précédents (estimation de deux lois)

	m=27, n=34	m=64, n=54	m=120, n=130	m=250, n=350
W (1; 2) Expo (0.88)	0,98	0,999	1	1
W (2; 3) Expo (0.70)	0,918	0,926	0,945	0,964

Tableau 4.8

Chapitre4 applications sur la comparaison des tests

On voit que la puissance augmente avec l'augmentation de la taille de l'échantillon. Mais elle est moins puissante par rapport aux autres exemples

On a pris d'autres lois et on a calculé la puissance estimée

On a utilisé la statistique construite

La statistique construite :

$$Z_m = \frac{\sum_{i=1}^m (x - X_i) - x - \hat{E}(X) \exp\left(-\frac{1}{\hat{E}(X)}x\right) + \hat{E}(X)}{\sqrt{\frac{1}{m} \left[\frac{1}{m} \sum_{i=1}^m (x - X_i)^2 - m \left(\sum_{i=1}^m (x - X_i) \right)^2 \right] + \frac{1}{m} \left(\hat{E}(X) \right)^2}}$$

H_0 : «F est HNBUE » contre H_1 : «F n'est pas HNBUE »

	20	100	300	500
W(1,2) HNBUE	0,912	0,995	1	1
W (2; 0,3) (IMRL)	0,363	0.211	0.088	0,036
Gamma (0.5, 2) DFR	0.089	0.001	0	0
Gamma(2,5) HNBUE	0.818	0.922	0.979	0.996

Tableau 4.9

Pour la W (1.2) et Gamma (2.5), la puissance s'améliore avec l'augmentation de la taille de l'échantillon, et pour W (2, 0.3) et Gamma (0.5, 2), la puissance diminue avec l'augmentation de la taille de l'échantillon.

On conclut que le test n'est pas performant pour le W (2, 0.3) et Gamma (0.5, 2), car la première loi est une IMRL et la deuxième est une DFR.

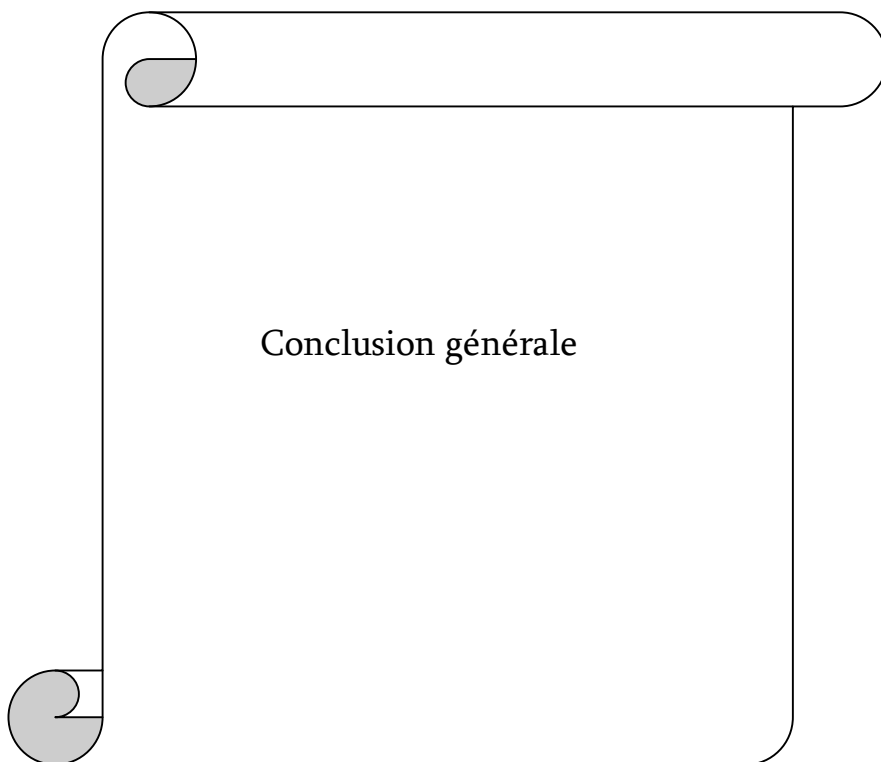
Conclusion

Le test de KS et le test Mann-Whitney, comparent l'égalité de deux distributions.

D'après ces tableaux, on conclut que le test de KRS(1994), est plus puissant par rapport aux autres tests, car le test de comparaison stochastique d'ordre deux compare les moyenne de deux où plusieurs variables aléatoires, quand l'ordre un, on ne peut pas appliquer.

Dans le domaine fiabilité le test de KRS(1994) est un bon point de départ, pour déterminer les classes HNBUE où l'hypothèse nulle est HNBUE

Ce test est valable pour comparer deux graphiques bien séparé, des fonctions de comparaison stochastique de second ordre, car si il y a un croisement entre les deux graphiques de dominance stochastique de second ordre, on ne peut pas appliquer ce test . Dans ce cas là on passe la dominance stochastique d'ordre trois.



Conclusion générale

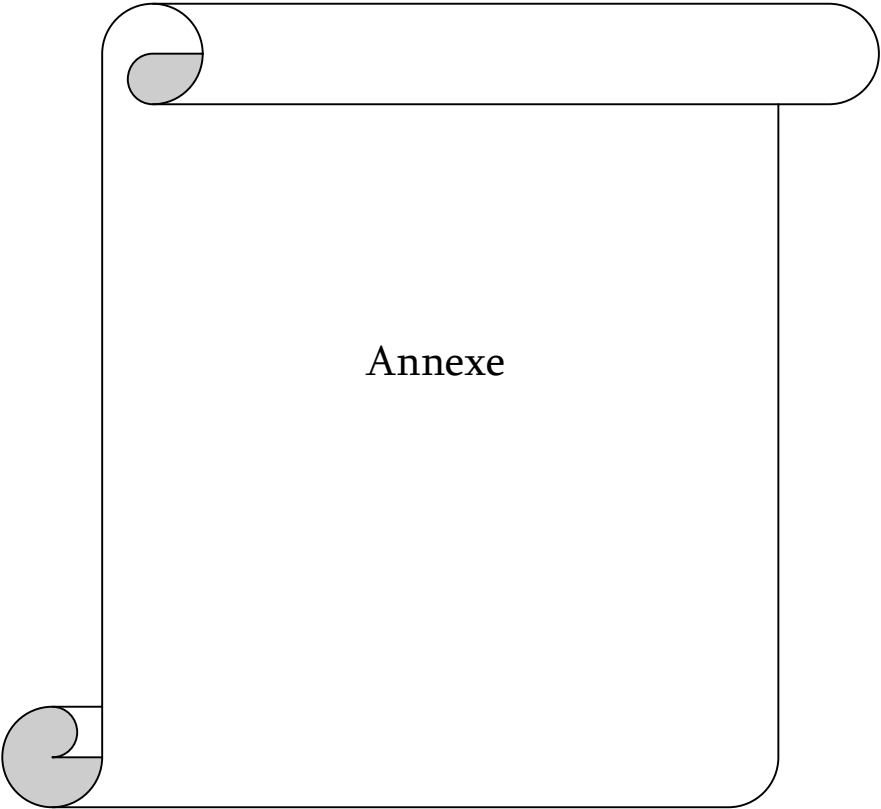
Conclusion générale

Dans cette thèse on a étudié les propriétés et les caractéristiques des lois non paramétriques. Particulièrement la classe de loi HNBUE qui caractérise par la comparabilité convexe. Pour tester l'appartenance à cette classe, on a utilisé un test de Kaur, Rao, Singh (KRS(1994)). Ce test est un bon point de départ pour tester la classe HNBUE.

Dans notre application, du test de HNBUE, on a trouvé de bons résultats pour les différentes tailles des échantillons. On conclut que ce test a un inconvénient, comme dans beaucoup d'autres tests, les variables aléatoires doivent être indépendantes.

Dans le domaine d'assurance, on peut utiliser ce test pour comparer deux portefeuilles risqués (la sélection de portefeuilles).

Comme les données des séries temporelles sont dépendantes, dans la suite de notre travail, on va essayer d'appliquer le test de comparaison stochastique, sur ces séries (pour comparer deux séries temporelles) et on fera une modification à cette approche.



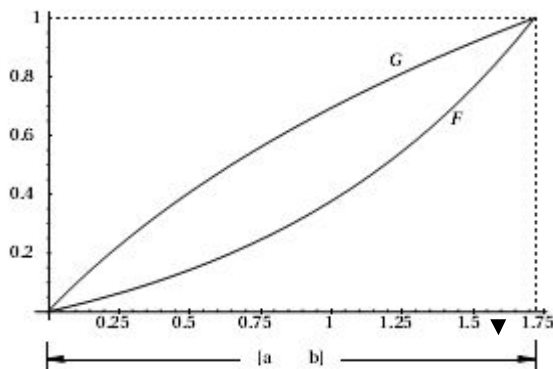
Dominance stochastique à l'ordre un (comparaison stochastique)

Soient F et G deux fonctions de distributions

On rappelle la condition de premier ordre, qui s'exprime sous la forme

$$X \geq_{DS1} Y \Leftrightarrow F(x) \leq G(x) \forall x \in [xa, xb]$$

Ou bien sous la forme $E_X(u(x)) \geq E_Y(u(x))$ avec u croissante sur $[xa, xb]$



$$E_X(u(x)) \geq E_Y(u(x)) \Leftrightarrow \int_{xa}^{xb} u(x)f(x)dx \geq \int_{xa}^{xb} u(x)g(x)dx \Leftrightarrow \int_{xa}^{xb} u(x)(f(x) - g(x))dx \geq 0$$

Par intégration par partie on a :

$$u = u(x), \quad u' = u'(x)$$

$$v' = f(x) - g(x), \quad v = F(x) - G(x)$$

Par la suite on a :

$$\int_{xa}^{xb} uv'dx = [u(x)(F(x) - G(x))]_{xa}^{xb} - \int_{xa}^{xb} u'(x)(F(x) - G(x))dx \geq 0$$

On a:

$$[u(xb)(F(xb) - G(xb))] - [u(xa)(F(xa) - G(xa))]$$

$$[u(xb)(1 - 1)] - [u(xa)(0 - 0)] = 0$$

Donc $E_X(u(x)) \geq E_Y(u(x)) \Leftrightarrow -\int_{ax}^{bx} u'(x)(F(x) - G(x))dx \geq 0$ (1)

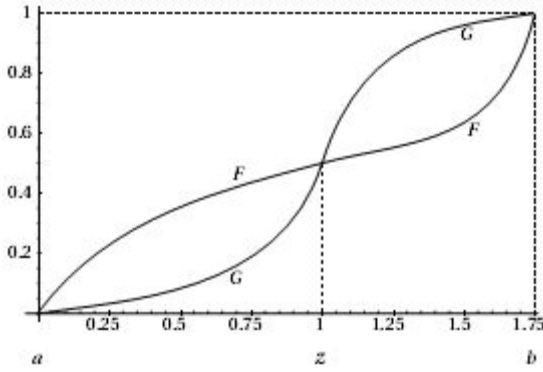
Si $u'(x) \geq 0$, donc l'inégalité (1) s'écrit $\int_{xa}^{xb} (F(x) - G(x))dx \leq 0$

Par la suite, la condition nécessaire de dominance stochastique s'écrit :

$X \geq_{DS1} Y \Leftrightarrow F(x) \leq G(x) \forall x \in [xa, xb]$ avec u croissante sur cet intervalle

Dominance stochastique du second ordre (comparaison convexe)

Si on a la différence $F(x)-G(x)$ positive sur l'intervalle $[xa, xz]$ et négative sur $[xz, xb]$ (les deux graphiques de distributions se croisent, donc on parle de dominance de second ordre).



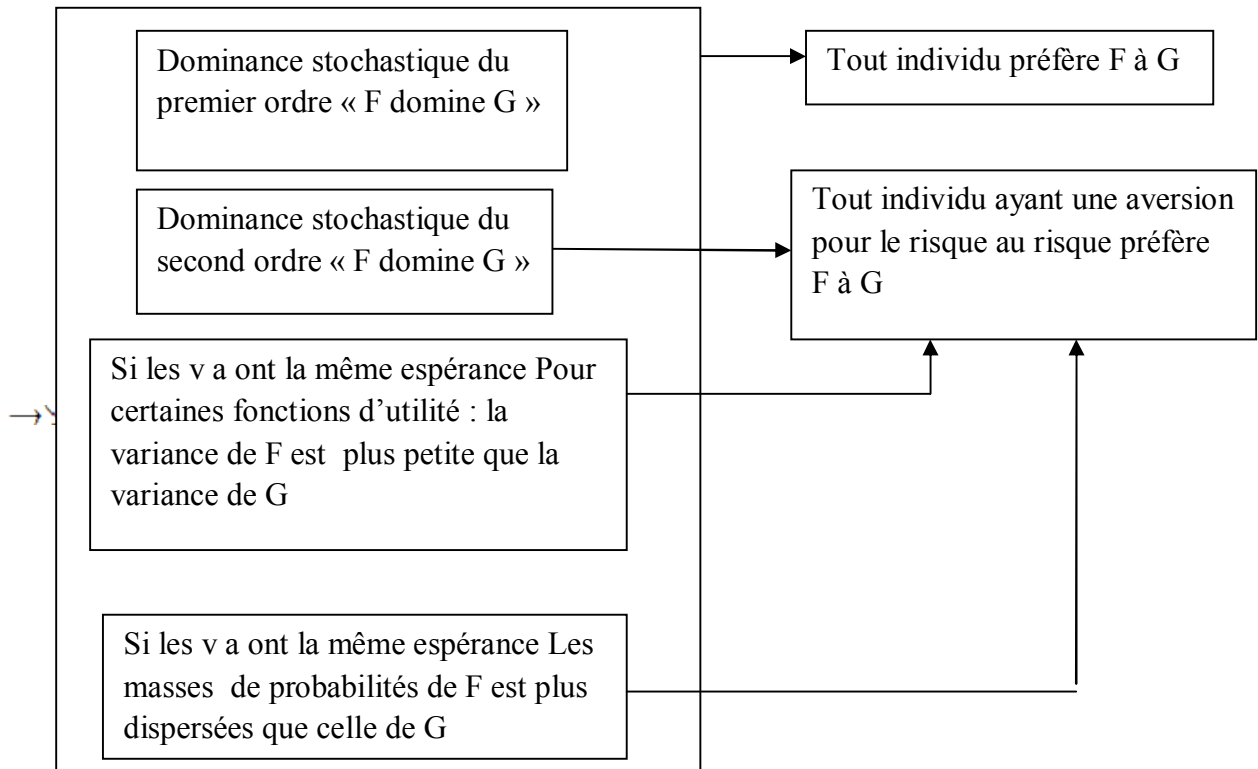
Donc $\int_{xa}^{xz} u'(x)(F(x) - G(x))dx \geq 0$ et $\int_{xz}^{xb} u'(x)(F(x) - G(x))dx \leq 0$.

Si on préfère la relation (1), on peut écrire :

$$-\int_{xa}^{xz} u'(x)(F(x) - G(x))dx - \int_{xz}^{xb} u'(x)(F(x) - G(x))dx$$

On ne peut rien dire sur le signe de cette dernière expression. par ailleurs, il est évident que le terme $u'(x)$ intervient comme un coefficient de pondération (variable selon x) et qu'il est de ce fait assez gênant. C'est pourquoi nous allons nous en débarrasser. Donc on passe à la dominance stochastique de second ordre.

Diagramme de dominance stochastique



Quelques démonstrations

$$F \text{ est HNBUE} \Leftrightarrow \int_x^{+\infty} \bar{F}(u) du \leq \mu \text{Exp}(\mu) \Leftrightarrow \int_x^{+\infty} \bar{F}(u) du \leq \int_x^{+\infty} \text{Exp}(-u/\mu) du$$

$$\text{On a pour } \int_0^{+\infty} \bar{F}(u) du = \mu$$

$$\mu - \int_0^t \bar{F}(u) du \leq \mu - \int_0^t \text{Exp}\left(-\frac{u}{\mu}\right) du$$

$$-\int_0^t 1 - F(u) du \leq -\int_0^t 1 - (1 - \text{Exp}\left(-\frac{u}{\mu}\right)) du$$

$$-t + \int_0^t F(u) du \leq -t + \int_0^t \text{Exp}\left(-\frac{u}{\mu}\right) du \Leftrightarrow \int_0^t F(u) du \leq \int_0^t (1 - \text{Exp}(-u/\mu)) du$$

$$F \text{ est HNBUE} \Leftrightarrow \int_x^{+\infty} \bar{F}(u) du \leq \mu \text{Exp}(\mu) \Leftrightarrow \int_0^t F(u) du \leq \int_0^t (1 - \text{Exp}(-u/\mu)) du$$

Définition: Soit $p \in [0,1]$, $q_X(p) \in \mathfrak{R}$ est un p quantile de la variable aléatoire X si :

$$P\{X < q_X(p)\} \leq P \leq P\{X < q_X(p)\}.$$

Définition : On définit

$$F_X^{-1}:]0,1] \rightarrow \overline{\mathfrak{R}}$$

$$P \rightarrow \inf \{F_X^{-1}(x) \geq p\}$$

L'inverse de F_X^{-1} est croissant et continu à gauche.

Programme de test de KRS

```

m; n;
X=-1/2*log(1-rand(m,1));
x=X.^1/3; s1=sort(x); a1=s1.^2;
y=(-7/10)*log(1-rand(n,1));
s2=sort(y); a2=s2.^2; w=[s1;s2]; z=sort(w);
for p=1:m+n; l(p)=0;
    for i=1:m; if s1(i)<=z(p); l(p)=l(p)+1; else l(p)=l(p);
        end end end
for p=1:m+n; k(p)=0; for j=1:n;
    if s2(j)<=z(p); k(p)=k(p)+1; else k(p)=k(p);
        end end end
if l(1)=0; s(1)=0;else s(1)=s1(l(1));
end
for r=2:m+n; if l(r)==l(r-1); s(r)=s(r-1); else s(r)=s(r-1)+s1(l(r));
    end end
if k(1)==0; d(1)=0; else d(1)=s2(k(1));
end
for q=2:m+n; if k(q)==k(q-1); d(q)=d(q-1); else d(q)=d(q-1)+s2(k(q));
    end end
if l(1)==0; t(1)=0; else t(1)=a1(l(1));
end
for r1=2:m+n; if l(r1)==l(r1-1); t(r1)=t(r1-1); else t(r1)=t(r1-1)+a1(l(r1));
    end end
if k(1)==0; f(1)=0; else f(1)=a2(k(1));
end
end

```

```
for q1=2:m+n; if k(q1)==k(q1-1); f(q1)=f(q1-1); else f(q1)=f(q1-1)+a2(k(q1));
    end end
m2=(m-1)/(n*m^3);
m3=(n-k)/(m*n^3);
m4=m2+m3;
m5=k.*s;
m6=l.*d;
m7=m5-m6;
m8=m4.*m7;
c1=l/m-k/n;
c2=1/m^2.*t+1/n^2.*f;
c3=(m-1)/m^2;
c4=(n-k)/n^2;
c5=(1/m*n)*(c3-c4);
c6=c5.*s.*d;
c7=m2.*d.^2-m3.*s.^2;
e=-c7./m8;
for p=1:m+n; for i=1:m; if s1(i)<e(p); e1(p,i)=e(p)-s1(i); else e1(p,i)=0;
    end end end
for p=1:m+n; for j=1:n; if s2(j)<e(p);
    e2(p,j)=e(p)-s2(j); else e2(p,j)=0;
    end end end
for p=1:m+n;
    Va1(p)=sum(e1(p,1:l(p)));
end
for p=1:m+n;
    Va2(p)=sum(e2(p,1:k(p)));
end
e11=e1.^2; e22=e2.^2;
for p=1:m+n;
    Va11(p)=(1/m).*sum(e11(p,1:l(p)));
for p=1:m+n;
    Va22(p)=(1/n).*sum(e22(p,1:k(p)));
End end
```

```

Va112=1/m.*(Va11-Va1).^2;
%Va222=(1/n).*(Va22-n.*Va2);
Va221=1/n.*(Va22-Va2).^2;
% Va33=(Va111+Va222).^1/2;
%T=(Va1-Va2)./Va33
T1=(Va11-Va22)./(Va112+Va221).^1/2;
T2=min(T1)

```

Quelque classe des lois non paramétriques

$F(x) = 1 - \exp(-\alpha x^\beta)$ (la fonction de répartition de la loi de Weibull)

$f(x) = \beta^\alpha x^{\alpha-1} \exp(-\beta x) / \Gamma(\alpha)$ (la densité de la loi Gamma)

Le tableau ci dessous présente quelque loi non paramétrique

Les lois	$W(\alpha, \beta)$	$Gamma(\alpha, \beta)$
HNBUE	$\alpha > 0 \quad \beta > 1$	$\beta > 0 \quad \alpha > 1$ $\beta > 0 \quad \alpha < 1$
IMRL	$\alpha > 0 \quad \beta < 1$	
HNBUE		
DFR		

La réassurance

Définition

La réassurance est une opération par la quelle l'assureur partage avec une autre personne (le réassureur), selon des modalités variables, les conséquences pécuniaires des engagements assumés envers ses propres assurés.

La réassurance est une technique permettant à l'assureur direct de céder à une autre société, le réassureur, une partie de risques qu'il a souscrits.

Donc le principe de la prime ajustée est utilisé pour déterminer un estimateur de prime de réassurance. Ce principe se présente comme suit

1- Principe de la prime ajustée :

- Principe de Wang :

Définition. Une fonction de distorsion g est une fonction $g : [0,1] \rightarrow [0,1]$ avec $g(0) = 0$ et $g(1) = 1$.

Comme tout principe de chargement de prime, l'idée des opérateurs de distorsion est d'imposer une marge de sécurité en augmentant la valeur attendue des sinistres.

L'originalité du processus réside dans le traitement mathématique : il s'agit d'effectuer une distorsion de la loi de probabilité sous-jacente des sinistres à l'aide d'une fonction g . En assurance dommage, un risque est une v.a. positive X modélisant un montant cumulé de sinistres sur une période donnée. Elle est définie par sa fonction de répartition $F_X(x) = P(X \leq x)$.

Alors, la prime pure du contrat associé à cette v.a. vaut l'espérance de cette dernière, c'est-à-dire la valeur attendue des sinistres, soit :

$$E(X) = \int_0^{+\infty} \bar{F}_X(t) dt.$$

En s'inspirant de cette formule, S.Wang(1996) propose un nouveau procédé de calcul de prime en appliquant une transformation à la fonction de survie. Il obtient alors la prime chargée (ou prime ajustée au risque) suivante concept :

$$\Pi_g(X) = \int_0^{\infty} g(\bar{F}(x)) dx.$$

Wang propose d'utiliser la fonction puissance, $g(u) = u^{1/p}$ ($p \geq 1$), il en résulte le principe de PH-transform (propositional Hazards) définit par :

$$\Pi_u(X) = \int_0^{\infty} (1 - F(x))^{1/p} dx.$$

Le paramètre $p \geq 1$ est appelé coefficient de distorsion ou le coefficient d'aversion au risque. La prime ajustée pour un seuil élevé $u > 0$ est donnée par :

$$\Pi_u(X) = \int_u^{\infty} (1 - F(x))^{1/p} dx.$$

Notons que le seuil u doit être grand et dépend de la taille de l'échantillon ($n \geq 1$). Pour cette raison on va supposer que $u = u_n$, $n \geq 1$. Ce qui conduit à écrire $\Pi_u(X)$ comme :

$$\Pi_{u_n}(X) = \int_{u_n}^{\infty} (1 - F(x))^{1/p} dx.$$

Analytiquement il est souvent difficile d'exprimer cet indice par une forme exacte,

Necir et Boukhetala (2004) ont proposé un estimateur asymptotiquement convergeant en loi de probabilité, permettant ainsi de définir un intervalle de confiance pour l'encadrement de cette prime. Ce qui donne un outil souple et interactif pour la conception d'un système d'aide à la tarification en réassurance.

D'autre part, cet estimateur constitue un schéma numérique facile à implémenter sur ordinateur, avec lequel l'utilisateur peut réaliser des scénarios en simulation, qui pourront lui servir comme étape intermédiaire avant de fixer son système tarifaire.

Cet estimateur tel qu'il a été conçu par A. Necir et K. Boukhetala, se présente comme suit :

2- Estimation de la prime Ajustée:

Dans cette section on va présenter la forme finale de l'estimateur de la prime ajustée développer par Necir et Boukhetala.

Soit $k = k_n$ une suite d'entier satisfaisant $1 \leq k \leq n$, $k \rightarrow \infty$ et $k/n \rightarrow 0$ tel que $u_n = Q(1 - k/n)$ où $Q(s) = \inf\{t \in R, F(t) \geq s\}$, $0 \leq s < 1$,

est la fonction quantile ou l'inverse généralisée de F, alors $\Pi_{u_n}(X)$ s'écrira :

$$\Pi_{u_n}(X) = \int_{Q(1-k/n)}^{\infty} (1 - F(x))^{1/p} dx.$$

Pour tout $n \geq 1$, soit $X_{(1)}, \dots, X_{(n)}$ la statistique d'ordre de l'échantillon $X_1, \dots, X_n$.

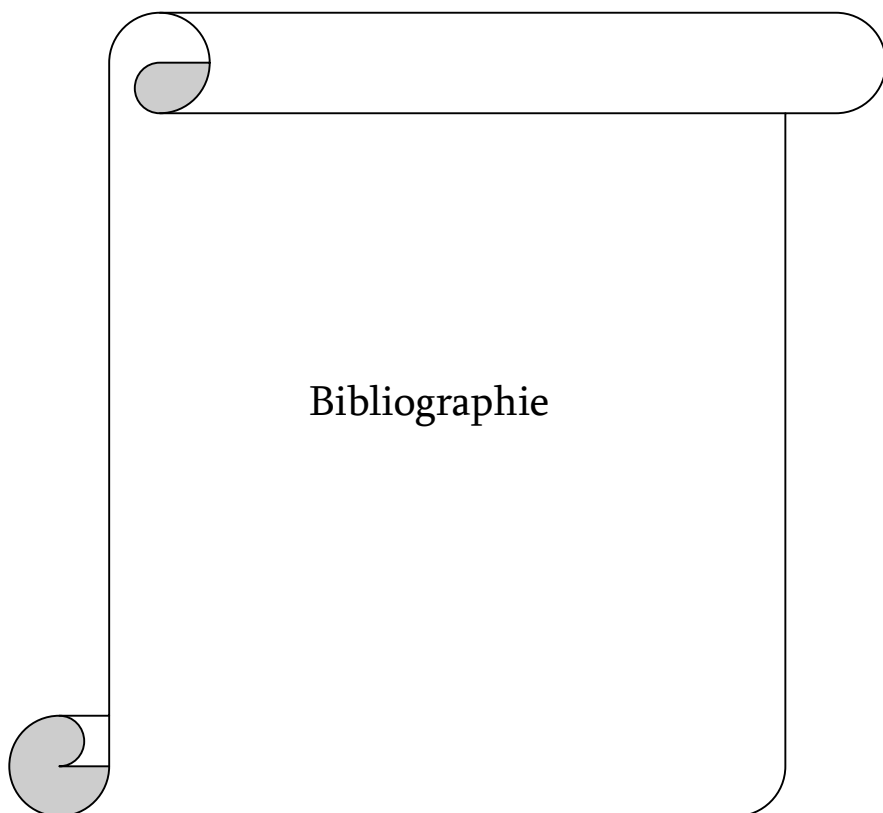
La forme finale de l'estimateur sera

$$\hat{\Pi}_{u_n}(X) = \sum_{i=1}^k (i/n)^{1/p} \{X_{(n-i+1)} - X_{(n-i)}\}, \quad p \geq 1.$$

Annexe

Pour plus de détaille voir A. Necir et K. Boukhetala.

Cet estimateur a été implémenté, sous forme d'un logiciel interactif, pour application au problème de réassurance concernant des données de sinistralité observées au sud de l'Algérie. (Projet de fin d'étude réalisé par T.Kessi et A.Tata sous la direction de M. K.Boukhetala).



Bibliographie

- [1] Smail Adjabi, Bouraine and Amre aissani (2002) "On statistical tests for non parametric life distributions" CiteSeer.ist.psu.edu/529905.
- [2] J, L- Bon(1995) : « Fiabilité des système : Méthodes Mathématiques » Ed. Masse, Paris.
- [3] Yann Braouezec (2005): Microéconomie: «décision en univers risqué» Ecole Supérieure d'Ingénieurs Léonard de Vinci (ESILV) Cours de S5 (troisième année)
- [4] Jordi Caballe · Joan Esteban (2002): «Stochastic dominance and absolute risk aversion». UPF economics and businen working paper N° =643.
- [5] Philippe Capéraà Bernard Van Cutsem, (1988) : «méthodes et modèles en statistique non paramétrique» -Bordas, XUI 353.
- [6] Jérôme Collet (2000) : «La modélisation des durées de réparation pour les études de sûreté de fonctionnement» thèse des Applications des Mathématiques
- [7] Russell Davidson. April 2007: «Testing for Restricted Stochastic Dominance Some Further Results» working paper
- [8] Olivier Decourt : « Tests non paramétriques" www.faq.sas.com
- [9] Isha Dewan Rohini Somanathan June 2004: «Poverty targeting in public programs: Acomparison of alternative nonparametric» Discussion Paper 04-16
- [10]Anderson, G. (1996), "Nonparametric tests of stochastic dominance in income distributions", *Econometrica*, 64, 1183 — 1193.
- [11] Hodges J.L. Jr (1958): «The significance probability of the Smirnov two-sample test. Arkiv for Mthematik, 3, 469-486
- [12] Gleser, L.j On a theory of intersection-union tests. Institute of mathematical Statistics Review 59 (1969):25-34.
- [13] Kaur, A., Rao, B.L.S.P., Singh, H. (1994): «Testing for Second Order Stochastic Dominance of Two Distributions ». *Econometric Theory* 10, 849–866.
- [14] Klar, B(2000).: "A class of tests for exponentiality against HNBUE alternatives", *statistics and probability letters*, 47,199-2.7.
- [15] Singh, H and Kochar, s.c. (1986), "A tests for exponentiality against HNBUE alternatives", *Communications in statistics theory and methods*, 15, 2295-2304.

Bibliographie

- [16] J.P Lecoutre P Tassi (1987): «statistique non paramétrique et robustesse» Collection « Economie et statistique avancées », Economica.
- [17] Arnaud Lefranc Nicolas Pistolesi et Alain Trannoy (2004): Le revenu selon l'origine sociale, économie et statistique N° 371, 44-88.
- [18] Arnaud Lefranc Nicolas, Pistolesi, Alain Trannoy (Janvier 2004): « Une réduction de l'inégalité des chances dans n revenu en France? »Revue Economie publique, publication de l'IDEP.
- [19] Chin-Diew Lai , Min Xie (2006) : «Stochastic Ageig and dependence for reability» Springer.
- [20] Ogryczak, W., Ruszczyński, A. (1999): «From Stochastic Dominance to Mean-Risk Models: Semideviations as Risk Measures». European Journal of Op-erational Research 116, 33–50.
- [21] W. Ogryczak and A. Ruszczyński (2001): Dual Stochastic Dominance and Related Mean-Risk Models, RUTCOR Research Report, 2001.
- [22] Ramousse R., Le Berre M. & Le Guelte L. (1996) : « Introduction aux statistiques » www.com-dev.org/elearning/stat/.
- [23] V.K.ROHATGI (1976): “An Introduction to Probability theory and Mathematical Statistics” Wiley professional software.
- [24] Shorrocks, A.F. (1983): Ranking Income Distributions. Economica 50, 3–17.
- [25] M. Shaked , J.G Shanthikumar(1994): « Stochastic order and their appliquations» Academic Press, San Diego.
- [26] Langenberg ,P et Srinivasan, R (1979), “Null distribution of the Hollander Proschan statistic for decreasing, mean residual life”, biometric, 66(3), p679-680.
- [27] Philippe Tassi (2004): «méthodes statistiques» Economica, (3edition) Economie et statistique avancées
- [28] Y.K. Tse and X.B. Zhang (2004): «A Monte Carlo Investigation of Some Tests for Stochastic Dominance» , journal of statistical computation and simulation 74 p 361-378.

Bibliographie

- [29] B. Weaver (2002): « Non parametric Tests », www.bmj.com/collection/statsbk.
- [30] JL CAYATTE (1971) “ note pédagogique sur la dominance stochastique” *Economie theory* volume 3 p 66-84(11).
- [31] A Necir K.Boukhetala, M Cherfi, (2004): “Estimating the risk-Adjusted Premium the Largest Claims Reinsurance Covers” *COMSTAT*” Edition Physica Verlag Heidelberg 2004/ Springer, Volumel. PP. 1577-1585.
- [32] S.Wang(1996) » Premium Calculation by transforming the layer Premium Density” *ASTIM Bulletin*, 26 (1996) pp.71-92.