



الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire

وزارة التعليم العالي والبحث العلمي
Ministère de L'Enseignement Supérieur et de la Recherche Scientifique

جامعة سعد دحلب البليدة-1
Université Saad Dahlab Blida -1-



Mémoire de fin D'études

En vue de l'obtention du diplôme Master

Faculté des Sciences
Département : Mathématiques

Spécialité : Modélisation stochastique et statistique

THEME

Caractérisation des distributions de Dagum et ses applications pour modéliser les distributions des revenus

Présenté par :
SIDI-MOUSSA Houria
RAMOUL Imene
Devant le Jury :

Soutenu le : 10/07/2021

Président : O.TAMI MCB **Université de Blida 1**

Promoteur: A. RASSOUL Prof. **ENSH de Blida**

Examineur : R.FRIHI MAA **Université de Blida 1**

Promotion : 2020/2021



Remerciement

Nous tenons avant toute chose, à remercier DIEU et à lui rendre grâce pour la force et le courage qu'il nous a donné.

Ainsi nous tenons également à exprimer nos vifs remerciements à notre encadreur monsieur RASSOUL.

Pour avoir d'abord proposé ce thème, pour le suivi continué tout long de la réalisation de ce mémoire et qui n'a pas cessé de nous donner ses conseils.

Nous voudrions aussi remercier nos familles, pour le soutien et l'aide inconditionnel qu'elle nous apporte chaque jour.

Et je remercie mon binôme pour son aide continue et son souci de travailler parfaitement.

Nous adressons nos sincères remerciements et gratitude au membre de Jury pour nous avoir consacré son précieux temps.

Nous tenons à remercier vivement toutes les personnes qui nous ont aidés et réaliser ce mémoire, ainsi à tous ceux qui nous ont aidés de près ou de loin à accomplir ce travail.

Merci à tous



Dédicaces

A Papa , merci pour votre amour et votre soutien.

A tous ceux qui m'ont enseigné et surtout à. Monsieur **RASSOUL**.

A mon frère et ma sœur. .

A mes camarades.

A tous ceux qui ont contribué à mon bonheur

HOURIA



Dédicaces

Je dédie cet ouvrage

A mes parents qui m'ont soutenu et encouragé durant ces années d'études, merci pour votre amour et votre soutien.

A mon mari , mes frères et Ceux qui ont partagé avec moi tous les moments d'émotion lors de la réalisation de ce travail.

A tous ceux qui m'ont enseigné et surtout à. Monsieur Rassoul.

IMEN

ملخص

يلعب توزيع الدخل دورًا مركزيًا في الأدبيات , في هذه المذكرة نقدم توصيفًا لبعض توزيعات الدخل المستخدمة في هذا المجال ، من خلال تحديد توزيع Dagum والتوزيعات المشتقة من هذا التوزيع ، ندرس خصائصها والاستدلال الإحصائي، لهذا، نقدم بعض مقاييس عدم المساواة المعتادة (مؤشر جيني ، مؤشر زنقا(Zenga)، معامل الاختلاف ومنحنى لورنز (Lorenz)) ، في الأخير نقوم بمحاكاة بعض النماذج المناسبة وتطبيقها على بيانات حقيقية مأخوذة من البنك الدولي الخاصة بالدولة الجزائرية وتكييفها بتوزيع داغوم ثم تقدير مؤشرات عدم المساواة المختلفة .

Les distributions des revenus jouent un rôle primordial dans la littérature, dans ce mémoire on présente une caractérisation de quelques distributions des revenus d'utilisation importantes, en précisant la distribution de Dagum et les distributions dérivées à partir de cette distribution, on étudie leurs propriétés et l'inférence statistique, pour cela on présente quelques mesures des inégalités usuelles (indice de Gini, indice de Zenga, coefficient de variation et la courbe de Lorenz), simulations de quelques modèles et une application sur des données réelles tirées de la banque mondiale ensuite on estime les différents indices des inégalités.

Abstract

Income distributions play an important role in the literature, in this paper we present a characterization of some important income distributions, specifying the Dagum distribution and the distributions derived from this distribution, we study their properties and statistical inference, for this we present some usual inequality measures (Gini index, Zenga index, coefficient of variation and the Lorenz curve), simulations of some models and an application on real data then estimating the different inequality indices.

TABLE DES MATIÈRES

Introduction Générale	2
1 Caractérisation des distribution des revenus : Distributions de Dagum et ses applications	4
1.1 Généralités sur les variables aléatoires	4
1.1.1 Rappel sur les variables aléatoire	4
1.1.1.1 Fonction de répartition	5
1.1.1.2 La densité	5
1.1.1.3 Fonction quantile	5
1.1.1.4 L'espérance	5
1.1.1.5 La variance :	5
1.1.1.6 L'écart type	6
1.1.1.7 Moyenne	6
1.1.1.8 Moments	6
1.1.1.9 Fonction génératrice	6
1.1.1.10 Fonction caractéristique	7
1.1.2 Estimation	7
1.1.2.1 Qualité d'un estimateur	7
1.1.3 Types de Convergence	8
1.1.3.1 Convergence en probabilité	8
1.1.3.2 Convergence en loi	8
1.1.3.3 Normalité asymptotique	9
1.1.4 Méthode d'estimation des paramètres	10
1.1.4.1 Méthode des moments	10
1.1.4.2 Méthode du maximum de vraisemblance	10
1.1.4.3 Information de Fisher	11
1.1.5 Teste d'ajustement	12

1.1.5.1	Test de Kolmogorov-Smirnov	12
1.1.6	Critères de validation	13
1.1.6.1	Critère d'information d'Akaike(AIC)	13
1.1.6.2	Critère d'information bayésien (BIC)	13
1.1.7	L'asymétrie (skewness)	14
1.1.8	L'aplatissement (kurtosis)	14
1.1.9	Lois de probabilités usuelles	14
1.1.9.1	Loi exponentielle	14
1.1.9.2	Loi de Gamma	15
1.1.9.3	Loi de Weibull	16
1.1.10	Distribution de zenga	16
1.1.11	Distribution de Singh-Maddala	19
2	Distribution de Dagum et applications	20
2.1	Dagum/Burr III et Bêta Généralisée(BG)	23
2.2	Distribution de Dagum de type II et III	25
2.3	Distribution de Dagum log-logistique	29
2.4	Distribution puissance log-Dagum	30
2.5	Distribution de Topp-leone Dagum	31
2.6	Distribution de Exponentielle Dagum	35
2.7	La Distribution de log-Dagum	36
2.8	La Distribution exponentielle de Kumaraswamy-Dagum	38
2.8.1	la Distribution de Kumaraswamy-Dagum	38
2.8.1.1	La Distribution Exponentielle de Kumaraswamy-Dagum (EKD)	39
2.8.2	La distribution MC-Dagum	40
2.8.3	La distribution de Weibull-Dagum	42
3	Mesure des inégalités	46
3.1	Introduction	46
3.1.1	Courbe de Lorenz	47
3.1.2	Indice de Gini	48
3.1.3	Coefficient de variation	49
3.1.4	Indice de Zenga	49
3.2	Simulations et Applications	52
3.2.1	Simulations	52
3.2.2	Applications	56

TABLE DES FIGURES

2.1	Graphe de la fonction de densité de la distribution de Dagum	22
2.2	Graphe de la fonction de répartition de lois de Dagum pour différentes valeurs des paramètres.	23
3.1	La courbe de Lorenz de la distribution de Dagum	53
3.2	La courbe de Lorenz de la distribution inverse paralogistique	54
3.3	La courbe de Lorenz de La distribution inverse Lomax	55
3.4	La courbe de Lorenz de la distribution de Fisk	56
3.5	La courbe de Lorenz de le PIB par habitants en Algérie	57
3.6	Graphe de la fonction de répartition empirique de le PIB par habitants en Algérie	57

LISTE DES TABLEAUX

2.1	Le système logistique généralisé de distributions des revenus de Dagum	22
3.1	Les mesures des inégalités de la distribution de Dagum	53
3.2	Les mesures des inégalités de la distribution inverse paralogistique . .	53
3.3	Les mesures des inégalités de la distribution inverse Lomax	54
3.4	Les mesures des inégalités de la distribution de Fisk	55
3.5	Les indicateurs de position de le PIB par habitants en Algérie	56
3.6	Les mesures des inégalités de le PIB par habitants en Algérie	56
3.7	Résultats de l'ajustement des données PIB par quelques lois usuelles .	58

13 juillet 2021

INTRODUCTION GÉNÉRALE

En économie, la distribution des revenus désigne la manière dont le PIB total d'un pays est réparti entre ses habitants. La théorie économique et la politique économique ont longtemps considéré le revenu et sa distribution comme une préoccupation centrale.

L'objectif de cette mémoire est d'analyser le revenu en utilisant les distributions de Dagum et leurs propriétés et les mesures des inégalités. La distribution de Dagum est une distribution de probabilité continue définie sur des nombres réels positifs. Elle doit son nom à Camilo Dagum, qui l'a proposée dans une série des travaux dans les années 1970. Tel que Il existe de multiples indicateurs de la « distribution des revenus » et des inégalités. La plus fréquente méthode consiste à classer les ménages par « quantiles » (déciles, centiles...) des revenus croissants et à rapprocher les revenus moyens ou médians des quantiles extrêmes (le premier et le dernier décile par exemple) ou les plafonds et planchers des quantiles extrêmes ou encore de les rapporter au revenu moyen ou médian de l'ensemble de la population. Il est également possible d'estimer la part du revenu total perçue par les 10 % ou les 20 % les plus aisés, les 20 % ou les 40 % les plus pauvres etc. Selon l'indicateur retenu, la mesure des inégalités est différente : le rapport entre les revenus du dernier et du premier décile est toujours bien plus élevé que le rapport entre les revenus du dernier et du premier quartile.

Ce document est organisé comme suit :

Le premier chapitre introduit les concepts essentiels des modèles probabilistes afin d'aborder l'inférence statistique, rappel sur les variables aléatoires, les estimations statistiques, les méthodes d'estimation des paramètres, ensuite les tests d'ajustement des lois usuelles, en particulier le Test de Kolmogorov-Smirnov, Critère de validation d'un ajustement, telles que : information d'Akaike (AIC), Critère d'information bayésien (Bic), après avoir les lois plus utilisées sont décrites : Loi exponentielle, Loi de Gamma, Loi de Weibull, Loi de Singh-Maddala enfin la distribution de Zenga.

Le deuxième chapitre est consacré pour l'étude de la distribution de Dagum et ses propriétés statistiques, on présente les différentes distributions liées à la distribution de Dagum, telles que : Distribution de Dagum log-logistique, Distribution puissance log-Dagum, Distribution de Topp-leone Dagum, Distribution de exponentielle Dagum ensuite La Distribution de log-Dagum, La Distribution exponentielle de Kumaraswamy Dagum, La distribution MC-Dagum enfin La distribution de Weibull-Dagum.

Le troisième chapitre traite les outils et les indices des inégalités le plus utilisés au niveau international : La courbe de Lorenz, le coefficient de Gini qui est calculé à partir d'un graphique représentant la distribution des revenus, dit « courbe de Lorenz », plus un nouvel indice d'inégalité a été proposé par Zenga enfin le coefficient de corrélation. Enfin, en présentant quelques résultats de simulations et des applications sur des données réelles, en particulier le PIB de la population algérienne. Dans la partie de simulation, on génère quelques modèles de la familles de Dagum (inverse paralogistique, Fisk et invers Lomax). Dans la partie d'application, en faisant une application sur des données réelles tirée de la banque mondiale (PIB par habitants), enfin ont effectué une comparaison entre cinq modèles avec différents nombres de paramètres : Weibull, Dagum, normale, lognormale et la distribution généralisée des valeurs extrême pour savoir quel modèle peut représenter les données.

CHAPITRE 1

CARACTÉRISATION DES DISTRIBUTION DES REVENUES : DISTRIBUTIONS DE DAGUM ET SES APPLICATIONS

1.1 Généralités sur les variables aléatoires

1.1.1 Rappel sur les variables aléatoires

Définition 1.1 Soient (Ω, \mathcal{F}, p) un espace probabilisé et (E, \mathcal{E}) un espace mesurable. On appelle variable aléatoire de Ω vers E , toute fonction mesurable X de Ω vers E .

Définition 1.2 Une variable aléatoire est une fonction mesurable d'un espace inconnu (l'univers Ω) vers un espace d'état E connu, à chaque élément mesurable B de E , on associe son image réciproque dans Ω , la mesure P_X sur E se déduit donc de la mesure P sur Ω .

On a deux type de variable aléatoire :

Lorsque l'image est finie ou infini dénombrable, cette variable aléatoire est alors appelée une variable aléatoire discrète, et sa distribution peut être décrite par une fonction de masse de probabilité qui assigne une probabilité de chaque valeur à l'image de X . Si l'image est indénombrablement infinie, alors on appellera X une variable aléatoire continue et sa distribution peut être décrite par une fonction de densité de probabilité, qui affecte des probabilités aux intervalles; dans le reste nous intéressons a la variable aléatoire continue.

Définition 1.3 La probabilité P_X est appelée loi de probabilité de la variable aléatoire X . tel que P_X est une application qui a toute partie A de \mathbb{R} associe

$$P_X = P\{w \in \Omega : X(w) \in A\}$$

1.1.1.1 Fonction de répartition

On appelle fonction de répartition (cdf) d'une variable aléatoire X , la fonction F_X tel que :

$$\begin{aligned} F_X &: \mathbb{R} \mapsto [0, 1] \\ t &\mapsto F_X(t) = P(X \leq t) \end{aligned}$$

1.1.1.2 La densité

Une variable aléatoire X a densité est une variable aléatoire réelle ou vectorielle pour la quelle la probabilité d'appartenance à un domaine se calcul à l'aide d'une intégrale sur ce domaine

$$P(a \leq X \leq b) = \int_a^b f(x) dx.$$

La fonction à intégrer est alors appelée fonction de densité où densité de probabilité égale dans le cas réel à la dérivée de la fonction de répartition.

1.1.1.3 Fonction quantile

La fonction quantile d'une variable aléatoire est l'inverse de sa fonction de répartition. on appelle fonction quantile de X la fonction $F^{-1}(u)$, de $]0, 1[$ dans \mathbb{R} , qui à $u \in]0, 1[$ associe :

$$F^{-1}(u) = \inf \{x \in \mathbb{R} : F(x) \geq u\}, \quad 0 < u < 1$$

1.1.1.4 L'espérance

L'espérance d'une variable aléatoire X (si elle existe) est :

Cas discret

$$E(X) = \sum_x x P(X = x)$$

Cas continue

$$E(X) = \int x f(x) dx$$

1.1.1.5 La variance :

La variance d'une variable aléatoire X est :

$$\text{var}(X) = E \{X - E(X)\}^2 = E(X^2) - (E(X))^2 = \sigma^2$$

1.1.1.6 L'écart type

L'écart type est une mesure de la dispersion des valeurs d'un échantillon statistique ou d'une distribution de probabilité. il est défini comme la racine carrée de la variance ou de manière équivalente, comme la moyenne quadratique des écarts par rapport à la moyenne :

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

ou bien

$$\sigma = \sqrt{\text{var}(X)}$$

σ = l'écart type de la population

N = effectif de la population

x_i = chaque valeur de la population

μ = moyenne de la population

1.1.1.7 Moyenne

La moyenne est un outil de calcul permettant de résumer une liste de valeurs numériques en un seul nombre réel, tel que on a la moyenne empirique d'un échantillon de variables aléatoires réelles (X_1, \dots, X_n) est défini par la moyenne arithmétique des valeurs :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

1.1.1.8 Moments

Si X une variable aléatoire, on appelle moment d'ordre K , s'il existe le nombre $E(X^K)$ tel que on a :

Le moment centré d'ordre K

$$\mu_K = E[(X - E(X))^k]$$

Le moment factoriel d'ordre K

$$E[X(X-1)\dots(X-k+1)]$$

1.1.1.9 Fonction génératrice

On appelle fonction génératrice des moments d'une variable aléatoire X :

$$g(t) = E(e^{Xt}) = \int e^{Xt} f(x) dx$$

Cette relation permet de calculer très aisément les moments d'une loi dont on connaît la fonction génératrice, par exemple :

$$\begin{aligned} E(X) &= g'(0) \\ \text{var}(X) &= E(X^2) - E^2(X) = g''(0) - [g'(0)]^2 \end{aligned}$$

Le coefficient d'asymétrie (γ_1), et le coefficient d'aplatissement (β_2) sont :

$$\begin{aligned} \gamma_1 &= \frac{E(X^3)}{(\text{var}(X))^{\frac{3}{2}}} \\ \beta_2 &= \frac{E(X^4)}{(\text{var}(X))^2} \end{aligned}$$

1.1.1.10 Fonction caractéristique

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

$$U_X(t) = E(e^{iXt}) = \int e^{iXt} f(x) dx$$

1.1.2 Estimation

Définition 1.4 *Un estimateur du paramètre inconnu θ d'un modèle ou loi de probabilité est une fonction qui fait correspondre à une suite d'observations x_1, x_2, \dots, x_n issues du modèle ou de la loi de probabilité, la valeur $\hat{\theta}$ que l'on nomme estimateur ou estimation*

$$\hat{\theta} = f(x_1; x_2; \dots; x_n)$$

1.1.2.1 Qualité d'un estimateur

Biais

Une variable aléatoire fluctue autour de son espérance. On peut donc souhaiter que l'espérance de $\hat{\theta}$ soit égale à θ

$$\text{Biais}(\hat{\theta}) = E(\hat{\theta}) - \theta$$

Lorsque $\text{Biais}(\hat{\theta}) = 0$, l'estimateur est dit sans biais, et si $\text{Biais}(\hat{\theta}) > 0$; l'estimateur est dit positivement biaisé.

1.1.3 Types de Convergence

1.1.3.1 Convergence en probabilité

L'estimateur $\hat{\theta}_n$ est convergent s'il converge en probabilité vers θ , soit :

$$\begin{aligned}\hat{\theta} \xrightarrow{P} \theta &\iff \lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| < \varepsilon) \rightarrow 1, \forall \varepsilon > 0 \\ &\iff \lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| < \varepsilon) \rightarrow 0\end{aligned}$$

Théorème 1.1 *Tout estimateur sans biais dont la variance tend vers 0 est convergent :*

$$\left(E(\hat{\theta}_n) = \theta \text{ et } \text{Var}(\hat{\theta}_n) \rightarrow 0 \right) \implies \hat{\theta}_n \xrightarrow{P} \theta.$$

1.1.3.2 Convergence en loi

Soient F_1, F_2, \dots la suite des fonctions de répartition associées aux variables aléatoires réelles X_1, X_2, \dots , et F la fonction de répartition de la variable aléatoire réelle X . Autrement dit, F_n est définie par $F_n(x) = P(X_n < x)$, et F par

$$F(x) = P(X < x).$$

La suite X_n converge vers X en loi, ou en distribution, si

$$\lim_{n \rightarrow \infty} F_n(a) = F(a)$$

pour tout réel a où F est continue.

Puisque $F(a) = P(X \leq a)$, cela signifie que la probabilité que X appartienne à un certain intervalle est très proche de la probabilité que X_n soit dans cet intervalle pour n suffisamment grand. La convergence en loi est souvent notée

$$X_n \xrightarrow{L} X$$

ou encore

$$X_n \xrightarrow{D} X$$

La convergence en loi est la forme la plus faible au sens où, en général, elle n'implique pas les autres formes de convergence définies ci-dessous, alors que ces autres formes de convergence impliquent la convergence en loi. C'est ce type de convergence qui est utilisé dans le théorème central limite.

De manière équivalente, la suite (X_n) converge en loi vers X si et seulement si pour toute fonction continue bornée

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

Erreur quadratique moyenne

L'erreur quadratique moyenne (Mean Squared Error en anglais) appelée aussi risque quadratique est l'espérance du carré de l'erreur entre la vraie valeur et sa valeur estimée.

$$EQM(\hat{\theta}) = E\left((\hat{\theta}_n - \theta)^2\right)$$

Si le risque est faible, l'estimateur $\hat{\theta}_n$ est proche de θ .

1.1.3.3 Normalité asymptotique

On admettra le théorème central limite et la méthode Delta suivants

Théorème 1.2 (Théorème central limite). Soient une suite de variables aléatoires $X_1; X_2; \dots; X_n$ indépendantes et de même loi (donc de même espérance m et de même σ écart-type).

$$Z_n = \sqrt{n} \frac{\bar{X}_n - m}{\sigma} \rightsquigarrow \mathcal{N}(0, 1)$$

La variable aléatoire Z_n converge en loi vers la loi normale centrée réduite. C'est une conséquence du TCL qui assure que :

$$\sqrt{n}(\bar{X}_n - m) \xrightarrow{D} \mathcal{N}(0, \sigma^2)$$

Méthode Delta

Théorème 1.3 Soient une suite de variables aléatoires $X_1; X_2; \dots; X_n$ d'espérance θ et d'écart-type σ si

$$\sqrt{n}(X_n - \theta) \xrightarrow{L} \mathcal{N}(0, \sigma^2);$$

et une fonction g dérivable telle que $g'(\theta) \neq 0$. Dans ce cas la méthode delta donne :

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{D} \mathcal{N}(0, \sigma^2 [g'(\theta)]^2)$$

Définition 1.5 Soit $\hat{\theta}$ un estimateur du paramètre θ de la loi P_θ d'une v.a. observée X . On suppose qu'il existe deux suites de fonctions réelles strictement positives, $a = a_n(\theta)$ et $b = b_n(\theta)$ telles que :

$$\frac{\hat{\theta} - a}{b} \xrightarrow{D} \mathcal{N}(0, 1)$$

On dit alors que $\hat{\theta}$ est un estimateur asymptotiquement normal.

1.1.4 Méthode d'estimation des paramètres

Il existe plusieurs méthodes pour estimer les paramètres d'une distribution statistique. nous verrons tout d'abord les méthodes bien connues du maximum de vraisemblance et des moments

1.1.4.1 Méthode des moments

L'idée de base est d'estimer une espérance mathématique par une moyenne empirique, une variance par une variance empirique, etc...

Autrement dit, si $\theta = E(X)$, alors l'estimateur $\hat{\theta}$ par la méthode des moments est

$$\hat{\theta}_n = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Généralement, l'estimateur du moment d'ordre r , est donné par la formule :

$$E(X^r) = \frac{1}{n} \sum_{i=1}^n X_i^r, r = 1, 2, \dots,$$

Alors, pour $\theta \in \bar{O}$, si $E(X) = \varphi(\theta)$, où φ est une fonction inversible, l'estimateur de θ par la méthode des moments est :

$$\hat{\theta}_n = \varphi^{-1}(\hat{\delta}_n).$$

De la même manière, on estime la variance de la loi des X_i par la variance empirique de l'échantillon

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

1.1.4.2 Méthode du maximum de vraisemblance

Définition 1.6 Soient $(X_i)_{i=1..n}$ une suite de n variables aléatoires indépendantes et de même loi. La fonction de vraisemblance s'écrit :

$$L(\theta; x_1, \dots, x_n) = \begin{cases} \prod_{i=1}^n P(X = x_i; \theta) & \text{si les } X_i \text{ sont discrètes} \\ \prod_{i=1}^n f_X(X; \theta) & \text{si les } X_i \text{ sont continues} \end{cases}$$

Définition 1.7 L'estimation de maximum de vraisemblance de θ est la valeur $\hat{\theta}$ de θ qui rend maximale la fonction de vraisemblance $L(\theta; x_1, \dots, x_n)$. L'estimateur de maximum de vraisemblance (MV) de θ est la variable aléatoire correspondante. Donc $\hat{\theta}$ sera en général calculé en maximisant la log-vraisemblance :

$$\hat{\theta} = \arg \max \log L(\theta; x_1, \dots, x_n)$$

Quand $\theta = (\theta_1, \dots, \theta_d) \in O$ et que toutes les dérivées partielles ci-dessous existent, $\hat{\theta}$ est solution du système d'équations appelées équations de vraisemblance :

$$\forall j \in \{1, \dots, d\}, \frac{\partial}{\partial \theta_j} \ln L(\theta; x_1, \dots, x_n) = 0$$

où

$$\frac{\partial^2}{\partial \theta_j^2} \ln L(\theta; x_1, \dots, x_n) < 0$$

Dans ce cas, on le résout par des méthodes numériques, comme la méthode de Fisher Scoring.

1.1.4.3 Information de Fisher

Soit $f(x; \theta)$ la distribution de vraisemblance d'une variable aléatoire X (qui peut être multidimensionnelle), paramétrée par θ . Le score est défini comme la dérivée partielle de la log-vraisemblance par rapport au paramètre θ :

$$\frac{\partial}{\partial \theta} \log f(X; \theta) = \frac{1}{f(X; \theta)} \frac{\partial f(X; \theta)}{\partial \theta}$$

L'information de Fisher est alors définie comme le moment d'ordre deux de la fonction de score :

$$I(\theta) = E \left[\left(\frac{\partial}{\partial \theta} \log f(X; \theta) \right)^2 \middle| \theta \right]$$

Il est possible de montrer que la fonction de score a une espérance nulle. L'information de Fisher correspond par conséquent également à la variance de la fonction de score.

Théorème 1.4 Soit $\hat{\theta}_n$ l'MV du paramètre θ calculé sur n échantillon i.i.d $X_i \sim f_\theta$, Sous des conditions de régularité, $\hat{\theta}_n$ est consistant, et si $I_n(\theta)$ est inversible, alors pour tout $\theta \in O$

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{L} \mathcal{N}(0, I_n(\theta)^{-1})$$

où

$$I_n(\theta) = -E_\theta [\log L(\theta, X_1)].$$

Formulation multi-paramétrique

Dans le cas où la distribution de probabilité $f(X)$ dépend de plusieurs paramètres, θ n'est plus un scalaire mais un vecteur $\vec{\theta} = (\theta_1, \theta_2, \dots)$. La recherche du maximum de vraisemblance ne se résume donc non pas à une seule équation mais à un système :

$$E \left[\frac{\partial}{\partial \theta_i} \log f(X; \vec{\theta}) \right] = 0 \quad \forall i$$

On dérive vis-à-vis des différentes composantes de $\vec{\theta}$. Enfin, l'information de Fisher n'est plus définie comme une variance scalaire mais comme une matrice de covariance :

$$I(\theta_i, \theta_j) = E \left[\left(\frac{\partial}{\partial \theta_i} \log f(X; \vec{\theta}) \right) \left(\frac{\partial}{\partial \theta_j} \log f(X; \vec{\theta}) \right) \right]$$

Ecriture sous forme courbure

$$I(\theta_i, \theta_j) = -E \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(X; \vec{\theta}) \right)$$

La matrice de Fisher est couramment appelée la métrique d'information de Fisher

$$g_{i,j} = I(\theta_i, \theta_j)$$

1.1.5 Teste d'ajustement

1.1.5.1 Test de Kolmogorov-Smirnov

Le test de Kolmogorov-Smirnov est un test d'hypothèse utilisé pour déterminer si un échantillon suit bien une loi donnée connue par sa fonction de répartition continue, ou bien si deux échantillons suivent la même loi.

Ce test repose sur les propriétés des fonctions de répartition empiriques. Soit X_1, \dots, X_n n variables i.i.d définies sur un espace de probabilité (Ω, A, P) , à valeurs dans \mathbb{R} avec pour fonction de répartition F . La fonction de répartition empirique F_n de l'échantillon X_1, \dots, X_n est définie par :

$$\forall x \in \mathbb{R}, \forall \omega \in \Omega, F_n(x, \omega) = \frac{\text{Nombre d'éléments} \leq x}{n} = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i(\omega) \leq x\}}$$

où 1_A est la fonction indicatrice sur l'ensemble A .

Notons $F_n(x, \cdot)$ la variable aléatoire $\omega \mapsto F_n(x, \omega)$. On a la convergence suivante :

$$\mathbf{P} \left(\sup_x |F_n(x, \cdot) - F(x)| > \frac{c}{\sqrt{n}} \right) \xrightarrow{n \rightarrow \infty} \alpha(c) = 2 \sum_{r=1}^{\infty} (-1)^{r-1} \exp(-2c^2 r^2)$$

pour toute constante $c > 0$. Le terme $\alpha(c)$ vaut 0,05 pour $c = 1,36$.

Remarquons que la limite à droite ne dépend pas de F . Cela découle du fait que $\sqrt{n}(F_n(x, \cdot) - F(x))$ converge en loi vers un pont brownien changé de temps par l'inverse F^{-1} de F .

La série $\alpha(c)$ se déduit des propriétés de ce dernier processus.

Il est ainsi facile de proposer un test d'hypothèse pour décider si un échantillon provient bien d'une loi donnée, ou si deux échantillons ont la même loi, lorsque leurs fonctions de répartition sont continues.

On peut aussi considérer $\max_x (F_n(x, \cdot) - F(x))$ et $\max_x (F(x, \cdot) - F_n(x))$.

1.1.6 Critères de validation

1.1.6.1 Critère d'information d'Akaike (AIC)

Le critère d'information d'Akaike est une mesure de la qualité d'un modèle statistique proposée par Hirotugu Akaike en 1973.

Lorsque l'on estime un modèle statistique, il est possible d'augmenter la vraisemblance du modèle en ajoutant un paramètre. Le critère d'information d'Akaike, tout comme le critère d'information bayésien, permet de pénaliser les modèles en fonction du nombre de paramètres afin de satisfaire le critère de parcimonie. On choisit alors le modèle avec le critère d'information d'Akaike le plus faible.

Le critère d'information d'Akaike s'écrit comme suit :

$$AIC = 2k - 2\ln(L)$$

où k est le nombre de paramètres à estimer du modèle et L est le maximum de la fonction de vraisemblance du modèle.

Si l'on considère un ensemble de modèles candidats, le modèle choisi est celui qui aura la plus faible valeur d'AIC. Ce critère repose donc sur un compromis entre la qualité de l'ajustement et la complexité du modèle, en pénalisant les modèles ayant un grand nombre de paramètres, ce qui limite les effets de sur-ajustement (augmenter le nombre de paramètre améliore nécessairement la qualité de l'ajustement).

L'AIC est basé sur la théorie de l'information : il propose une estimation de la perte d'information lorsqu'on utilise le modèle considéré pour représenter le processus qui génère les données. L'AIC ne fournit pas un test de modèle dans le sens d'une hypothèse nulle, c'est-à-dire que ce test ne dit rien de la qualité absolue du modèle. Il ne rendrait ainsi pas compte du fait que tous les modèles candidats ne produisent pas de bons ajustements.

1.1.6.2 Critère d'information bayésien (BIC)

Le critère d'information bayésien est un critère d'information dérivé du critère d'information d'Akaike proposé par Gideon Schwarz en 1978.

A la différence du critère d'information d'Akaike, la pénalité dépend de la taille de l'échantillon et pas seulement du nombre de paramètres.

Il s'écrit :

$$BIC = -2\ln(L) + k \cdot \ln(N)$$

avec L la vraisemblance du modèle estimée, N le nombre d'observations dans l'échantillon et k le nombre de paramètres libres du modèle.

Le modèle qui sera sélectionné est celui qui minimise le critère BIC, soit :

$$M_{BIC} = \arg \min_k BIC(M)$$

1.1.7 L'asymétrie (skewness)

Etant donnée une variable aléatoire réelle X de moyenne μ et d'écart type σ , on définit son coefficient d'asymétrie comme le moment d'ordre trois de la variable centrée réduite :

$$\gamma_1 = E \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right]$$

1.1.8 L'aplatissement (kurtosis)

Cette indice est une mesure indirecte de l'aplatissement de la distribution d'une variable aléatoire réelle.

kurtosis non normalisée (coefficient d'aplatissement)

Etant donnée une variable aléatoire réelle X d'espérance μ et d'écart type σ , on définit son kurtosis non normalisé comme le moment d'ordre quatre de la variable centrée réduite :

$$\beta_2 = E \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right]$$

1.1.9 Lois de probabilités usuelles

Dans cette section, on procède à la présentation de quelques lois de probabilités qu'utiles pour la suite du travail.

1.1.9.1 Loi exponentielle

Une loi exponentielle modélise la durée de vie d'un phénomène sans mémoire, où sans vieillissement, ou sans usure. En d'autres termes, le fait que le phénomène ait duré pendant t heures ne change rien à son espérance de vie à partir du temps t .

Une variable aléatoire continue X suit une loi exponentielle de paramètre (d'intensité ou inverse de l'échelle) $\lambda > 0$ si elle admet pour densité de probabilité la fonction

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \forall x \in \mathbb{R}^+ \\ 0 & \text{ailleurs} \end{cases}$$

On note $X \rightsquigarrow \text{Exp}(\lambda)$

La fonction de répartition de X

$$F_X(x) = \begin{cases} 1 - \exp(-\lambda x) & \forall x \in \mathbb{R}^+ \\ 0 & \text{ailleurs} \end{cases}$$

L'espérance mathématique, la varaince et l'écart type de X

$$\begin{aligned} E(X) &= \frac{1}{\lambda} \\ \text{var}(X) &= \frac{1}{\lambda^2} \\ \sigma(X) &= \frac{1}{\lambda} \end{aligned}$$

Les coefficients d'asymétrie et les coefficients d'aplatissement sont

$$\gamma_1 = 2 \text{ et } \beta_2 = 6$$

La fonction génératrice des moments et la fonction caractéristique sont

$$\begin{aligned} g(t) &= \left(1 - \frac{t}{\lambda}\right)^{-1} \\ U_X(t) &= \left(1 - \frac{it}{\lambda}\right)^{-1} \end{aligned}$$

1.1.9.2 Loi de Gamma

Une variable aléatoire continue suit une loi Gamma, de paramètres $\alpha, \beta \in \mathbb{R}_+$, le premier est appelé paramètre d'échelle alors β que est le paramètre de forme, si elle admet pour densité de probabilité la fonction

$$f_X(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & \forall x > 0 \\ 0 & \text{ailleurs} \end{cases}$$

où

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$$

On note $X \rightsquigarrow \text{Gamma}(\alpha, \beta)$.

L'espérance mathématique et la variance de X

$$E(X) = \frac{\alpha}{\beta}; \text{var}(X) = \frac{\alpha}{\beta^2}$$

Les coefficients d'asymétrie et les coefficients d'aplatissement sont

$$\gamma_1 = \frac{2}{\sqrt[3]{\alpha}}; \beta_2 = \frac{6}{\alpha}$$

La fonction génératrice des moments et la fonction caractéristique

$$\begin{aligned} g(t) &= (1 - t\beta)^{-\alpha} \\ U_X(t) &= (1 - it\beta)^{-\alpha} \end{aligned}$$

1.1.9.3 Loi de Weibull

La loi de Weibull est mentionnée en premier lieu par Fréchet puis étudiée par Weibull dont elle prit le nom. Cette loi est très utile en contrôle de qualité, elle peut également modéliser le temps d'attente de la première panne, ou encore le temps écoulé entre deux pannes consécutives.

Une variable aléatoire continue suit une loi de Weibull de paramètres $\beta, \eta \in \mathbb{R}_+^*$ qui sont les paramètres de forme et d'échelle respectivement, si elle admet pour densité de probabilité la fonction

$$f_X(x) = \begin{cases} \frac{\beta}{\eta^\beta} x^{\beta-1} e^{-\left(\frac{x}{\eta}\right)^\beta} & \forall x > 0 \\ 0 & \text{ailleurs} \end{cases}$$

La fonction de répartition de X

$$F_X(x) = 1 - \exp\left(-\left(\frac{x}{\eta}\right)^\beta\right) \quad \forall x \in \mathbb{R}_+^*$$

L'espérance mathématique et la variance de X

$$\begin{cases} E(X) = \eta \Gamma\left(1 + \frac{1}{\beta}\right) \\ \text{var}(X) = \eta^2 \Gamma\left(1 + \frac{2}{\beta}\right) - E^2(X) \end{cases}$$

Fonction génératrice des moments est

$$g(t) = \eta^t \Gamma\left(1 + \frac{t}{\beta}\right)$$

1.1.10 Distribution de zenga

C'est une distribution par taille a trois paramètres proposé par [Zenga(2010)], analysé avec une attention particulière sur les paramètres; le modèle est un mélange Bêta

défini pour une distribution non négative, qui a une asymétrie positive et queue droite Parétienne et il est donc particulièrement indiqué pour décrire les revenus, la richesse, la répartition financière et actuarielle, de plus ses paramètres contrôlent séparément la localisation et l'inégalité.

La distribution de Zenga est obtenue sous la forme d'un mélange de distribution de Pareto tronquée de Poliscchio avec des poids Bêta, les densités conditionnelles ont une fonction de densité de probabilité donnée par :

$$V(x; \mu, k) = \begin{cases} \sqrt{\frac{2\mu k}{x}} (1-k)x^{-\frac{3}{2}}, & x \in [\mu k, \frac{\mu}{k}] \\ 0, & \text{sinon} \end{cases} \quad (1.1)$$

La fonction mixte sur $k \in]0, 1[$ est une distribution Bêta (en fonction des deux paramètres positifs α et θ) avec fonction de densité de probabilité donnée par

$$g(k; \alpha, \theta) = \begin{cases} \frac{k^{\alpha-1}(1-k)^{\theta-1}}{\beta(\alpha, \theta)}, & k \in]0, 1[\\ 0, & \text{sinon} \end{cases} \quad (1.2)$$

avec $\beta(a, b)$ est une fonction de Bêta

$$\beta(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt \quad a > 0, b > 0$$

Par conséquent la fonction de densité de probabilité de la distribution de Zenga est :

$$\begin{aligned} f(x; \mu, \alpha, \theta) &= \int_0^1 V(x; \mu, k)g(k; \alpha, \theta)dk \\ &= \begin{cases} \frac{1}{2\mu\beta(\alpha, \theta)} \left(\frac{x}{\mu}\right)^{-\frac{3}{2}} \int_0^{\frac{x}{\mu}} k^{\alpha-\frac{1}{2}}(1-k)^{\theta-2} dk & \text{si } 0 < x < \mu \\ \frac{1}{2\mu\beta(\alpha, \theta)} \left(\frac{\mu}{x}\right)^{\frac{3}{2}} \int_0^{\frac{\mu}{x}} k^{\alpha-\frac{1}{2}}(1-k)^{\theta-2} dk & \text{si } x > \mu \end{cases} \end{aligned} \quad (1.3)$$

Sa fonction de répartition est

$$F(x; \mu, \alpha, \theta) = \begin{cases} \int_0^{\frac{x}{\mu}} \left[1 - \sqrt{\frac{\mu k}{x}}\right] \frac{k^{\alpha-1}(1-k)^{\theta-2}}{\beta(\alpha, \theta)} dk & \text{si } \alpha < x < \mu \\ 1 - \int_0^{\frac{\mu}{x}} \left[\sqrt{\frac{\mu}{xk}} - 1\right] \frac{k^{\alpha}(1-k)^{\theta-2}}{\beta(\alpha, \theta)} dk & \text{si } x > \mu \end{cases} \quad (1.4)$$

Le moment centré de la distribution de Zenga est

$$E(X - \mu)^r = \int_0^{+\infty} \int_0^1 (x - \mu)V(x; \mu, k)g(k; \alpha, \theta)dk dx$$

à chaque fois que $E(x - \mu)^r$ est fini, par le théorème de Fubini on a

$$E(X - \mu)^r = \int_0^1 \left[\int_0^{+\infty} (x - \mu)^r V(x; \mu, k) dx \right] g(k; \alpha, \theta) dk \quad (1.5)$$

Donc on a le moment de la distribution de Zenga peut être calculée par le théorème de moments central de la densité conditionnelle égal à

$$\frac{\mu^3(1-k)^4}{5k^2}$$

En suite le troisième moment central de la distribution de Zenga est

$$E(X - \mu)^3 = E(X - \mu)^2 = \frac{3\mu(\theta + 2)(\theta + 3)}{5(\alpha - 2)(\alpha + \theta + 1)} \quad (1.6)$$

En particulier

$$E(X^r) = \frac{\mu^r}{(2r-1)\beta(\alpha, \theta)} \sum_{i=1}^{2r-1} \beta(\alpha - r + i, \theta) \quad \text{si } r < \alpha + 1$$

L'espérance

$$E(X) = \mu \quad \text{si } \alpha > 1$$

On a aussi

$$E(X^2) = \frac{\mu^2}{3} \left(\frac{\theta}{\alpha - 1} + \frac{\alpha}{\alpha + \theta} + 2 \right)$$

Donc la variance est

$$E(X - \mu)^2 = \frac{\mu^2 \theta (\theta + 1)}{3(\alpha - 1)(\alpha + \theta)} \quad (1.7)$$

Pour la distribution de Zenga l'écart moyen est

$$E(|X - \mu|) = 2\mu[2F(\mu; \mu, \alpha, \theta) - 1]$$

Avec

$$F(\mu; \mu, \alpha, \theta) = \frac{1}{\theta - 1} \left[(\alpha + \theta - 1) - \frac{(\alpha + \theta - \frac{1}{2})\Gamma(\alpha + \frac{1}{2})\Gamma(\alpha + \theta)}{\Gamma(\alpha + \theta + 1)\Gamma(\alpha)} \right]$$

Il peut être approuvé que pour la distribution de Zenga

$$F(\mu; \mu, \alpha, \theta) \geq \frac{1}{2}$$

Cela implique que la moyenne est supérieurs a la médiane et par conséquent la distribution présente une asymétrie positive.

Soit (X_1, \dots, X_n) un échantillon aléatoire d'une distribution de Zenga alors la moyenne

obtenu par la méthode des moments égale

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \mu$$

1.1.11 Distribution de Singh-Maddala

La distribution Singh-Maddala, introduite par Singh et Maddala en 1975 et sous une forme plus élaborée en 1976, a fait l'objet d'une attention particulière dans la littérature sur le revenus

Sa densité est

$$f(x) = \frac{aqx^{a-1}}{b^a \left[1 + \left(\frac{x}{b}\right)^a\right]^{1+q}}, x > 0$$

où les trois paramètres a, b, q sont positifs. Ici, b est un paramètre d'échelle et a, q sont des paramètres de forme.

La fonction de distribution cumulative de la distribution de Singh -Maddala est disponible sous forme fermée; elle est donnée par l'expression agréablement simple suivante

$$F(x) = 1 - \left[1 + \left(\frac{x}{b}\right)^a\right]^{-q}, x > 0$$

La fonction quantile est également

$$F^{-1}(x) = b \left[(1-u)^{-\frac{1}{q}} - 1 \right]^{\frac{1}{a}}$$

Pour $0 < u < 1$

Le $k^{i\grave{e}me}$ moment existe pour $-a < k < aq$, il est égal à

$$E(X^r) = \frac{b^k B\left(1 + \frac{k}{a}, q - \frac{k}{a}\right)}{B(1, q)} = \frac{b^k \Gamma\left(1 + \frac{k}{a}\right) \Gamma\left(q - \frac{k}{a}\right)}{\Gamma(q)}$$

En particulier

$$E(X) = \frac{bB\left(1 + \frac{k}{a}, q - \frac{k}{a}\right)}{B(1, q)} = \frac{b\Gamma\left(1 + \frac{k}{a}\right) \Gamma\left(q - \frac{k}{a}\right)}{\Gamma(q)}$$

et

$$Var(X) = \frac{b^2 \left\{ \Gamma(q) \Gamma\left(1 + \frac{2}{a}\right) \Gamma\left(q - \frac{2}{a}\right) - \Gamma^2\left(1 + \frac{1}{a}\right) \Gamma^2\left(q - \frac{1}{a}\right) \right\}}{\Gamma^2(q)}$$

Où $B(\cdot)$ et $\Gamma(\cdot)$ sont les fonctions Bêta et Gamma respectivement.

CHAPITRE 2

DISTRIBUTION DE DAGUM ET APPLICATIONS

Introduction

Dans les années 1970, Camilo Dagum s'est lancé dans la recherche d'une distribution statistique étroite des distributions empiriques des revenus et des richesses. Ne correspond pas aux distributions classiques utilisées pour résumer de telles données { la distribution de Pareto (développée par l'économiste et sociologue italien Vilfredo Pareto à la fin du 19^{ème} siècle) et la distribution lognormale (popularisée par l'ingénieur français [Gibrat, (1931).] { il a cherché un modèle accommodant les lourdes queues présentes dans empirique la répartition des revenus et de la richesse ainsi que la possibilité d'un mode intérieur. Le premier aspect est bien saisi par le Pareto mais pas par la distribution lognormale, le second par la distribution lognormale mais pas par la distribution Pareto. Expérimentant une distribution logistique décalée [Dagum,c.1975], une généralisation d'une distribution précédemment envisagée par [Fisk,P.R.1961], il se rendit vite compte qu'un autre paramètre était nécessaire. Cela a conduit à la distribution de type I de Dagum, à une distribution à trois paramètres et à deux généralisations à quatre paramètres .

Il a fallu plus d'une décennie jusqu'à ce que la proposition de Dagum a commencé à apparaître dans la littérature économique et économétrique de langue anglaise. Le premier article d'une importante revue économétrique utilisant la distribution Dagum semble être par [Majumed, (1990)]. Dans la littérature statistique, la situation est plus favorable, en ce que la célèbre Encyclopédie des sciences statistiques contient, dans Volume. 4 ([Kotz et al.(1983)], une entrée sur les modèles de distribution des revenus, sans surprise, écrit par Camilo Dagum [Dagum,C.(1983)]. En rétrospective, la raison de ce long retard est assez évidente : [Dagum,C.(1977)] document a été publié dans Economie Appliquee, une revue française avec seulement des contributions occasionnelles en langue anglaise et assez peu de circulation dans les pays de langue

anglaise. En revanche, l'article introduisant la distribution plus largement connue de [Singh, (1976)] a été publié dans *Econometrica*, juste un an avant la contribution de Dagum. Il est lentement apparu que la distribution Dagum est, néanmoins, souvent préférable à la distribution Singh-Maddala dans les applications aux données sur le revenu.

Ce chapitre fournit un bref aperçu des distributions de Dagum, y compris les interrelations avec plusieurs distributions plus largement connues ainsi que les propriétés statistiques de base et les aspects inférentiels. Il revisite également l'un des premiers ensembles de données considérés par Dagum et présente une enquête sur les applications en économie.

[Dagum,C.(1977)] motive son modèle à partir de l'observation empirique que l'élasticité du revenu $\eta(G;x)$ de la fonction de distribution cumulative G du revenu est une fonction décroissante et bornée de G . A partir de l'équation différentiel :

$$\eta(G,x) = \frac{d \log G(x)}{d \log(x)} = ac \left\{ 1 - [G(x)]^{\frac{1}{c}} \right\}, \quad x \geq 0 \quad (2.1)$$

avec $c > 0$ et $ac > 0$, on obtient

$$G_D(x) = [1 + (x/b)^{-a}]^{-c}, \quad x > 0 \quad (2.2)$$

où b est un paramètre d'échelle, a et c sont paramètres de forme.

Cette approche a été développée dans une série d'articles sur les systèmes de génération pour la distribution des revenus ([Dagum,C.(1983)]; [Dagum,C.(1980b)], [Dagum(1980c)]). Rappelons que le système bien connu de Pearson est un système d'usage général non dérivé des régularités stables observées dans un domaine d'application donnée. Le système de [D'Addario,R.(1949)] est un système de traduction avec des fonctions de génération et de transformation flexibles conçues pour englober autant de distributions de revenus que possible. En revanche, le système spécifié par Dagum part des propriétés caractéristiques des distributions empiriques de revenus et de richesses et conduit à un système de génération spécifier en termes de

$$\frac{d \log \{F(x) - \gamma\}}{d \log x} = v(x)\phi(F) \leq k, \quad 0 \leq x_0 < x < \infty \quad (2.3)$$

où $k > 0$, $v(x) > 0$, $\Phi(x) > 0$, $\gamma < 1$ et $d\{v(x)\phi(F)\}/dx < 0$.

ces contraintes garantissent que l'élasticité-revenu de la fonction de distribution cumulative est une fonction positive, décroissante et bornée de G , et donc de x

Le tableau 1 fournit une sélection de modèles qui peuvent être déduits de système de Dagum pour certaines spécification des fonction v et ϕ , des versions plus étendues.

Exemple 2.1 *La présentation de la densité et la fonction de répartition de la distribution de*

TABLE 2.1 – Le système logistique généralisé de distributions des revenus de Dagum

Distribution	$\nu(x)$	$\phi(F)$	(γ, β)	support
Pareto(I)	α	$\frac{(1-F)}{F}$	$(0, 0)$	$0 < x_0 \leq x < \infty$
Fisk	α	$1 - F$	$(0, 0)$	$0 \leq x < \infty$
Singh-Maddala	α	$\frac{1-(1-F)^\beta}{F(1-F)^{-1}}$	$(0, +)$	$0 \leq x < \infty$
Dagum(I)	α	$1 - F^{\frac{1}{\beta}}$	$(0, +)$	$0 \leq x < \infty$
Dagum(II)	α	$1 - \left(\frac{F-\gamma}{1-\gamma}\right)^{\frac{1}{\beta}}$	$(+, +)$	$0 \leq x < \infty$
Dagum(III)	α	$1 - \left(\frac{F-\gamma}{1-\gamma}\right)^{\frac{1}{\beta}}$	$(-, +)$	$0 < x_0 \leq x < \infty$

Dagum pour différentes valeurs des paramètres sont présentées dans les deux figures suivants

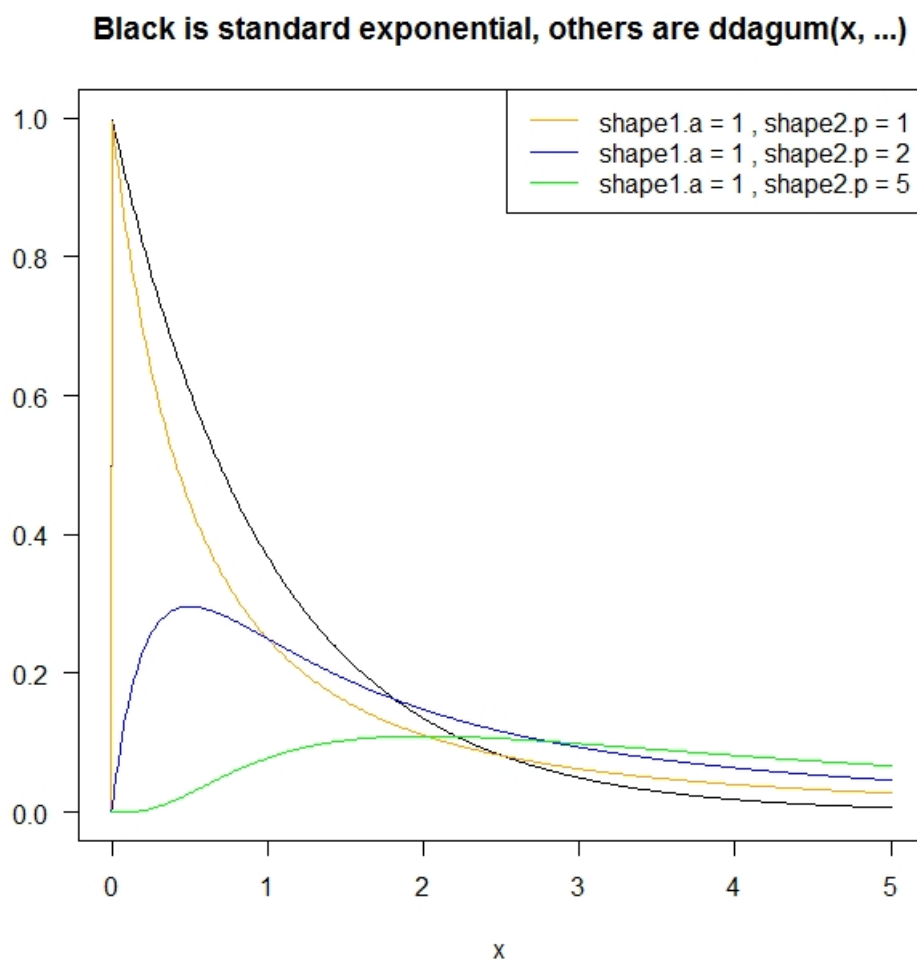


FIGURE 2.1 – Graphe de la fonction de densité de la distribution de Dagum

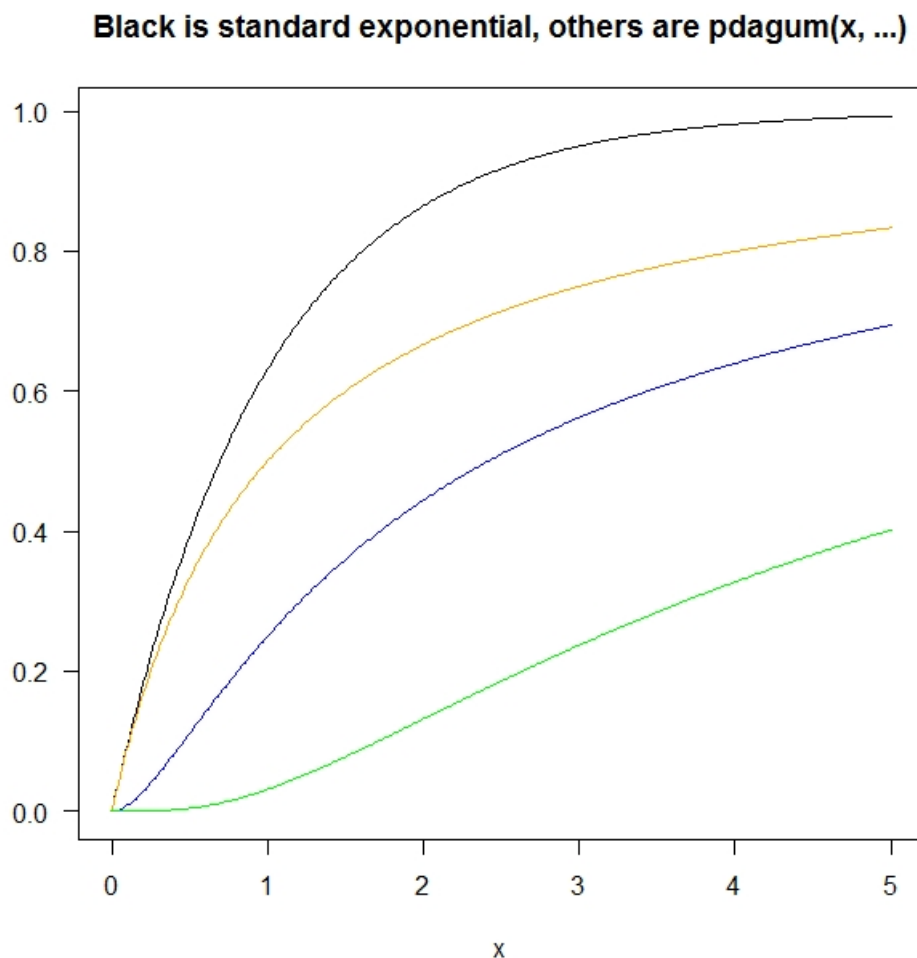


FIGURE 2.2 – Graphe de la fonction de répartition de lois de Dagum pour différentes valeurs des paramètres.

2.1 Dagum/Burr III et Bêta Généralisée(BG)

Alors que [Dagum,C.(1977)] utilise $(\beta; \delta; \lambda) = (c; a; b^a)$. La paramétrisation utilisée ici suit [McDonald.(1984)], car les distributions Dagum/Burr III et Singh-Maddala/Burr XII peuvent être imbriquées dans une distribution Bêta généralisée à quatre paramètres du second type (ci-après :GB2) avec densité

$$f(x) = \frac{ax^{ac-1}}{b^{ac}B(c, q)[1 + (x/b)^a]^{c+1}}, x > 0$$

où $a; b; c; q > 0$. tandis que la distribution de Dagum est un GB2 avec $q = 1$ et donc sa densité est :

$$g_D(x) = \frac{apx^{ac-1}}{b^{ac}[1 + (x/b)^a]^{c+1}}, x > 0 \tag{2.4}$$

Remarque 2.1 Il convient également de noter que la distribution Dagum (D) et la distribution Singh-Maddala (SM) sont intimement liées, tel que :

$$X \sim D(a, b, c) \Leftrightarrow \frac{1}{X} \sim SM(a, \frac{1}{b}, c). \quad (2.5)$$

Fonction quantile : On a aussi la fonction quantile :

$$G^{-1}(u) = b \left[u^{\frac{-1}{c}} - 1 \right]^{\frac{-1}{a}} \text{ pour } 0 < u < 1 \quad (2.6)$$

Moment d'ordre k : Le $k^{\text{ème}}$ moment existe pour $ac < k < a$ et est égal

$$E(X^k) = \frac{b^k B(c + \frac{k}{a}, 1 - \frac{k}{a})}{B(c, 1)} = \frac{b^k \Gamma(c + \frac{k}{a}) \Gamma(1 - \frac{k}{a})}{\Gamma(c)} \quad (2.7)$$

Où $\Gamma()$ et $B()$ désignent les fonctions Gamma et Bêta. Alors

$$E(x) = \frac{b^k B(c + \frac{k}{a}, 1 - \frac{k}{a})}{B(c, 1)}$$

et

$$Var(X) = \frac{b^2 \left\{ \Gamma(c) \Gamma(c + \frac{2}{a}) \Gamma(1 - \frac{2}{a}) - \Gamma^2(c + \frac{1}{a}) \Gamma^2(1 - \frac{1}{a}) \right\}}{\Gamma^2(c)}$$

Le coefficient de l'asymétrie (CS) et l'aplatissement (kurtosis) (CK) sont, respectivement, donnés par

$$CS = \frac{\mu^3}{\sigma^3} = \frac{\Gamma_3 - 3p\Gamma_1\Gamma_2 + 2p^2\Gamma_1^3}{\sqrt[3]{p}(\Gamma_2 - p\Gamma_1^2)}$$

$$CK = \frac{\mu^4}{\sigma^4} = \frac{\Gamma_4 - p\Gamma_1\Gamma_3 + 6p^2\Gamma_1\Gamma_2 - 3p^3\Gamma_1^4}{p^3[\Gamma_2 - p\Gamma_1^2]^4}$$

où $p = \frac{1}{\Gamma(c)}$ et $\Gamma_j = b^j \Gamma\left(1 - \frac{j}{a}\right) \Gamma\left(c + \frac{j}{a}\right)$

Les écarts moyens autour de la moyenne et de la médiane peuvent être utilisés comme mesures de la dispersion dans une population. Soit $\mu = E(T)$ et M la moyenne et la médiane de la distribution de Dagum, respectivement. Les écarts moyens autour de la moyenne et de la médiane peuvent être calculés comme suit

$$\delta_1(T) = \int_0^{\infty} |t - \mu| f(t) dt$$

et

$$\delta_2(T) = \int_0^{\infty} |t - M|f(t) dt$$

respectivement.

$$\begin{aligned} E(|T - m|) &= \int_0^{\infty} |t - m|f(t) dt \\ &= \int_0^m (m - t)f(t) dt + \int_m^{\infty} (t - m)f(t) dt \\ &= 2m \int_0^m f(t) dt - m - E(T) + 2 \int_m^{\infty} tf(t) dt \\ &= 2mF(m) + 2 \int_m^{\infty} tf(t) dt - E(T) - m \end{aligned}$$

Par conséquent, puisque nous avons $\mu = \frac{c}{b}B\left(1 - \frac{1}{a}, c + \frac{1}{a}\right)$ et $M = \left[b^{-a} \left[(0.5)^{\frac{-1}{c}} - 1\right]\right]^{\frac{-1}{a}}$

$$\begin{aligned} \delta_1(T) &= 2\mu F(\mu) - 2\mu + 2 \int_{\mu}^{\infty} tf(t) dt \\ &= 2\mu \left(\left(1 + \left(\frac{\mu}{b}\right)^{-a}\right)^{-c} - 1 \right) + 2cb \left(\left(\frac{\mu}{b}\right)^{-a} \right)^{\frac{-1}{a}} \sum_{n=0}^{\infty} (-1)^n \binom{c+n+1}{n} \frac{\left(\left(\frac{\mu}{b}\right)^{-a}\right)^{n+1}}{\left(n - \frac{1}{a} + 1\right)} \end{aligned}$$

$$\begin{aligned} \delta_2(T) &= 2MF(M) + 2 \int_M^{\infty} tf(t) dt - \mu \\ &= 2M \left(1 + \left(\frac{M}{b}\right)^{-a} \right)^{-c} + 2cb \left(\left(\frac{M}{b}\right)^{-a} \right)^{\frac{-1}{a}} \sum_{n=0}^{\infty} (-1)^n \binom{c+n+1}{n} \frac{\left(\left(\frac{M}{b}\right)^{-a}\right)^{n+1}}{\left(n - \frac{1}{a} + 1\right)} \end{aligned}$$

2.2 Distribution de Dagum de type II et III

Dagum (1977, 1980) introduit deux autres variantes de sa distribution, La version standard précédemment discutée sera appelée distribution Dagum type I dans ce qui suit.

La distribution Dagum de type II de fonction de distribution cumulative G :

$$G(x) = \delta + (1 - \delta)[1 + (x/b)^{-a}]^{-c}, \quad x \geq 0$$

où comme avant $a; b; c > 0$ et $\delta \in [0; 1]$. Il s'agit clairement d'un mélange d'une masse ponctuelle à l'origine avec une distribution Dagum (type I) sur la demi-ligne positive.

La distribution de type II a été proposée comme modèle pour les distributions de revenus à revenus nuls et négatifs, mais plus particulièrement pour les données sur la richesse, qui présentent souvent un grand nombre d'unités économiques à actifs bruts nuls et à actifs nets nuls et négatifs.

Sa fonction de densité est :

$$g(x) = \frac{acx^{ac-1}}{b^{ac} \left[1 + \left(\frac{x}{b}\right)^a\right]^{c+1}}, \quad x > 0$$

Il existe également une distribution de Dagum de type III, comme le type II

$$G(x) = \delta + (1 - \delta)[1 + (x/b)^{-a}]^{-c}$$

avec $a; b; p > 0$. Cependant, ici $\delta < 0$. Par conséquent, le support de cette variante est maintenant $[x_0; 1)$, $x_0 > 0$, où $x_0 = \left\{b \left[\left(1 - \frac{1}{a}\right)^{\frac{1}{c}} - 1 \right]\right\}^{\frac{-1}{a}}$ est déterminé implicitement à partir de la contrainte $G(x) \geq 0$.

Le Dagum type II et le type III sont membres du système logistique généralisé de Dagum-Burr.

[Dagum, C. (1977)], dans une période où les données individuelles étaient rarement disponibles, à minimisé

$$\sum_{i=1}^n \left\{ G_n(x_i) - \left[1 + \left(\frac{x_i}{b}\right)^{-a}\right]^{-c} \right\}^2$$

Une critère des moindres carrés non linéaires basé sur la distance entre la fonction de distribution cumulative empirique G_n et la fonction de distribution cumulative de Dagum,

la plupart des chercheurs utilisent aujourd'hui l'estimation du maximum de vraisemblance, deux cas distingués, les données groupées et les données individuelles.

Dans le cas des données groupées la vraisemblance $L(\theta)$, où $\theta = (a, b, c)^T$ est une vraisemblance multinomiale (en supposant qu'elles soient indépendantes)

$$L(\theta) = \prod_{j=1}^m \left\{ G(x_j) - G(x_{j-1}) \right\}$$

de 2.4

$$L(a, b, c) = (\alpha ca)^n \prod_{i=1}^n t_i^{-a-1} \left(1 + \left(\frac{t_i}{b}\right)^{-a}\right)^{-c-1}$$

L'estimation de vraisemblance à partir d'observations individuelles attire de plus en plus l'attention, le log-vraisemblance $\ell(\theta) = \log L(\theta)$ pour un échantillon de taille n est

$$\ell(a, b, c) = n \log a + n \log c + (ac - 1) \sum_{i=1}^n \log x_i - nac \log b - (c + 1) \sum_{i=1}^n \log \left\{1 + \left(\frac{x_i}{b}\right)^a\right\} \quad (2.8)$$

Ce qui donne les équations de vraisemblance

$$\frac{n}{a} + c \sum_{i=1}^n \log \left(\frac{x_i}{b}\right) = (c + 1) \sum_{i=1}^n \frac{\log \left(\frac{x_i}{b}\right)}{1 + \left(\frac{x_i}{b}\right)^a} \quad (2.9)$$

$$nc = (c + 1) \sum_{i=1}^n \frac{1}{1 + \left(\frac{x_i}{b}\right)^a} \quad (2.10)$$

$$\frac{n}{c} + a \sum_{i=1}^n \log \left(\frac{x_i}{b}\right) = \sum_{i=1}^n \log \left\{1 + \left(\frac{x_i}{b}\right)^a\right\} \quad (2.11)$$

Qui doit être résolu numériquement. Cependant, l'estimation de la vraisemblance dans cette famille n'est pas sans problème : si l'on considère la distribution $\log X$, une distribution logistique généralisée, [Shao.(2002)] montre que la MLE peut ne pas exister, et si ce n'est pas le cas, le problème de modèle intégré se produit.

C'est-à-dire, en laissant certains valeurs limites, une distribution avec moins de paramètres émerge.

Les implications sont que le comportement de la vraisemblance doit être soigneusement vérifié dans le travail empirique. Il serait intéressant de déterminer dans quelle mesure cette complication aux données sur le revenu où toute la flexibilité de la famille Dagum n'est pas nécessaire.

Apparemment peut conscients de ces problèmes, [Domanski.(1998)] fournissent une étude de simulation de la performance des MLE.

Il s'avère que des échantillons assez importants sont nécessaires pour que les estimations des paramètres de forme a, c puissent être considérées comme non biaisées, tandis que l'estimation fiable du paramètre d'échelle semble nécessiter des échantillons encore plus grands.

La matrice d'information de Fisher

$$I(\theta) = \left[-E \left(\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j} \right)_{i,j} \right] = \begin{pmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{pmatrix}.$$

Prend la forme

$$I_{11} = \frac{1}{a^2(2+c)} \left[c \left[\{\Psi(c) - \Psi(1) - 1\}^2 + \Psi'(c) + \Psi'(1) \right] + 2 \{\Psi(c) - \Psi(1)\} \right]$$

$$I_{21} = I_{12} = \frac{c - 1 - c \{\Psi(c) - \Psi(1)\}}{b(2+c)}$$

$$I_{22} = \frac{a^2 c}{b^2(2+c)}$$

$$I_{23} = I_{32} = \frac{a}{b(1+c)}$$

$$I_{31} = I_{13} = \frac{\Psi(2) - \Psi(c)}{a(1+c)}$$

$$I_{33} = \frac{1}{c^2}$$

où Ψ est la fonction Diagamma.

la fonction Diagamma est

$$\Psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$$

avec Γ est la fonction Gamma

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$$

Il faut noter qu'il existe plusieurs dérivations de l'information de Fisher dans la littérature statistique, une détaillée utilisant la paramétrisation de Dagum due à [Latorre.(1988)] et une autre due à Zeltermann (1987). Ce dernier considère la distribution de $\log X$, une distribution logistique généralisée, en utilisant la paramétrisation $(\theta, \sigma, \alpha) = (\log b, \frac{1}{a}, c)$.

En ce qui concerne les estimateurs alternatifs, une inspection des résultats 2.9 et 2.11 révèle que $\sup_x \left\| \frac{\partial \ell}{\partial \theta} \right\| = \infty$, où $\|\cdot\|$ représente la norme euclidienne, et donc la fonction de score non bornée dans le cas de Dagum. Cela implique que l'ELM est plutôt sensible aux

observations uniques situées suffisamment loin de la majorité des données. Les distributions de revenus ont toujours été populaires auprès des auteurs italiens, et la

distribution de Dagum ne fait pas exception. [Cheli,(1995)] étudient des mélanges de distributions de Dagum et leur estimation via l'algorithme EM. Les distributions de la médiane de l'échantillon et de l'étendue de l'échantillon ont été obtenues par [Domma.(1997)]. En plus, [Latorre.(1988)] fournit erreurs standard de la méthode delta pour plusieurs mesures d'inégalité dérivées des MLE pour le modèle de Dagum.

2.3 Distribution de Dagum log-logistique

[Gleaton,(2006)] ont annoncé une nouvelle transformation de la fonction de distribution appelée famille log-logistic-G (OLL-G) avec une forme supplémentaire paramètre $\alpha > 0$ par la fonction de distribution cumulative

$$F(x; \alpha, \varphi) = \frac{G(x, \varphi)^\alpha}{G(x; \varphi)^\alpha + \bar{G}(x; \varphi)^\alpha} \quad (2.12)$$

Où $\bar{G}(x; \varphi) = 1 - G(x; \varphi)$, la fonction de densité de probabilité correspondante de 2.12 est donnée par

$$f(x; \alpha, \varphi) = \frac{\alpha g(x; \varphi) G(x; \varphi)^{\alpha-1} \bar{G}(x; \varphi)^{\alpha-1}}{[G(x; \varphi)^\alpha + \bar{G}(x; \varphi)^\alpha]^2} \quad (2.13)$$

En général, une variable aléatoire X avec fonction de densité 2.13 est désignée $X \sim OLL - G(\alpha, \varphi)$

En utilisant 2.2 et 2.4, on obtient la fonction de distribution cumulative et la fonction de densité de probabilité de la distribution de Dagum logistique (OLLDa) pour $x > 0$ on a

$$F(x) = \frac{[1 + (\frac{x}{a})^{-b}]^{-\alpha c}}{[1 + (\frac{x}{a})^{-b}]^{-\alpha c} \{1 - [1 + (\frac{x}{a})^{-b}]^{-c}\}^\alpha} \quad (2.14)$$

et

$$f(x) = \frac{\alpha b c a^b x^{-b-1} [1 + (\frac{x}{a})^{-b}]^{-\alpha c-1} \{1 - [1 + (\frac{x}{a})^{-b}]^{-c}\}^{\alpha-1}}{([1 + (\frac{x}{a})^{-b}]^{-\alpha c} + \{1 - [1 + (\frac{x}{a})^{-b}]^{-c}\}^\alpha)^2} \quad (2.15)$$

Respectivement, où b, c et α sont des paramètres de forme positifs et $a > 0$ est un paramètre d'échelle.

Si X est un variable aléatoire avec fonction de densité de probabilité 2.15, nous désignons $X \sim OLLDa(\alpha; a, b, c)$, pour $\alpha = 1$ nous obtenons une distribution de Dagum.

En peut obtenir la fonction quantile de la distribution des OLLDa en inversant 2.14 comme suit

$$Q(u) = a \left\{ -1 + \left[\frac{u^{\frac{1}{\alpha}}}{u^{\frac{1}{\alpha}} + (1-u)^{\frac{1}{\alpha}}} \right]^{\frac{-1}{c}} \right\}^{\frac{-1}{b}} \quad (2.16)$$

Le n^{ieme} moment ordinaire de X est donnée par

$$\mu'_n = E(X^n) = \sum_{k=0}^{\infty} d_k \int_0^{\infty} x^n g(x; a, b, ck) dx$$

Pour $n \leq b$, on obtient

$$\mu'_n = E(X^n) = \sum_{k=0}^{\infty} d_k c k a^n B\left(1 - \frac{n}{b}, ck + \frac{n}{b}\right) \quad (2.17)$$

En fixant $n = 1$ dans 2.17, nous avons la moyenne de X .

Le moment central de X est :

$$M_s = E(X - \mu'_1)^s = \sum_{i=0}^s (-1)^i \binom{s}{i} (\mu'_1)^s u'_{s-i}$$

En utilisant la définition de la fonction génératrice de moment que nous avons

$$M(t) = \sum_{k=0}^{\infty} (k+1) d_{k+1} \int_0^1 u^k \exp[tQ_X(u)] du$$

Soit X a L'OLLDa nous pouvons calculer les écarts moyens par rapport a la moyenne

$$v = E(x) = q$$

Plusieurs méthode d'estimation de paramètre ont été proposées, mais la méthode la plus populaire est la méthode du maximum de vraisemblance.

Soit x_1, x_2, \dots, x_n un échantillon aléatoire de taille n de la distribution de OLLDa donnée par 2.15, la fonction de log-vraisemblance des paramètres de vecteur $\theta = (\alpha, a, b, c)$ est le suivante

$$l(\theta) = n \ln(\alpha) + \sum_{i=1}^n \left\{ \ln g(x_i) + (\alpha - 1) [\ln G(x_i) + \ln \bar{G}(x_i)] - 2 \ln [G(x_i)^\alpha + \bar{G}(x_i)^\alpha] \right\}$$

2.4 Distribution puissance log-Dagum

Nous introduisons un nouveau modèle appelé distribution Puissance Log-Dagum (PLD) avec la fonction de distribution cumulative

$$F(x) = \left\{ 1 + e^{-(vx + \text{sign}(x) \frac{\rho}{v}(x)^v)} \right\}^{-\varsigma}; \quad x, v \in \mathbb{R}, \varsigma > 0, \rho > 0 \quad (2.18)$$

où

$$\text{sign} = \begin{cases} 1 & \text{si } x > 0 \\ 0 & \text{si } x = 0 \\ -1 & \text{si } x < 0 \end{cases}$$

La fonction de densité de probabilité est

$$f(x) = \varsigma (v + \rho|x|^{v-1}) e^{-(vx+\text{sign}(x)\frac{\rho}{v}|x|^v)} \left\{ 1 + e^{-(vx+\text{sign}(x)\frac{\rho}{v}|x|^v)} \right\}^{-(\varsigma+1)}. \quad (2.19)$$

Soit X une variable aléatoire PLD , le moment d'ordre r de X est

$$\mu'_r = \varsigma \int_{-\infty}^{\infty} x^r (v + \rho|x|^{v-1}) e^{-(vx+\text{sign}(x)\frac{\rho}{v}|x|^v)} \left\{ 1 + e^{-(vx+\text{sign}(x)\frac{\rho}{v}|x|^v)} \right\}^{-\varsigma-1} dx$$

La fonction génératrice de moment de la distribution PLD est :

$$M(t) = \varsigma \int_{-\infty}^{\infty} e^{tx} (v + \rho|x|^{v-1}) \left\{ 1 + e^{-(vx+\text{sign}(x)\frac{\rho}{v}|x|^v)} \right\}^{-(\varsigma+1)} e^{-(vx+\text{sign}(x)\frac{\rho}{v}|x|^v)} dx \quad (2.20)$$

Soit $x_1, x_2, x_3, \dots, x_n$ un échantillon aléatoire de la distribution PLD avec le vecteur de paramètre $O = (v, \varsigma, \rho)$ et $x_1, x_2, x_3, \dots, x_n$ sont les valeurs observées correspondantes, alors la fonction de probabilité conjointe de $x_1, x_2, x_3, \dots, x_n$ en tant que fonction de log-vraisemblance peut être exprimée comme suit :

$$\begin{aligned} l(O) = & n \log(\varsigma) + \sum_{i=1}^n \log(v + \rho|x_i|^{v-1}) - \sum_{i=1}^n (vx_i + \text{sign}(x_i)\frac{\rho}{v}|x_i|^v) \\ & - (\varsigma + 1) \sum_{i=1}^n \log(1 + e^{-vx_i - \text{sign}(x_i)\frac{\rho}{v}|x_i|^v}) \end{aligned}$$

2.5 Distribution de Topp-leone Dagum

Utilisation du fonction de distribution cumulative et la fonction de densité de probabilité du Topp-L ou TL distribution [[Alshomrani,\(2016\)](#)] ont recommandé une nouvelle famille généralisée de distribution, la fonction de distribution cumulative et la fonction de densité de probabilité sont

$$f_{TLG}(x) = 2\alpha g(x)\bar{G}(x) \left\{ 1 - (\bar{G}(x))^2 \right\}^{\alpha-1} \quad (2.21)$$

et

$$F(x) = \left[1 - (\bar{G}(x))^2 \right]^{\alpha} \quad (2.22)$$

Tel que $G(x)$ et $g(x)$ sont la fonction de distribution cumulative et la fonction de densité de probabilité de Dagum respectivement.

Pour obtenir la densité et la fonction de distribution cumulative avec quatre paramètres de la distribution de Topp-leone Dagum ou $TLDa$ on utilisant 2.21 et 2.22

$$f_{TLDa}(x) = 2\alpha \frac{ca}{b} x^{-a-1} \left(1 + \left(\frac{x}{b}\right)^{-a-1}\right)^{-c-1} \left\{1 - \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-c}\right\} \times \left[1 - \left\{1 - \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-c}\right\}^2\right]^{\alpha-1} \quad (2.23)$$

Alors que

$$F_{TLDa}(x) = \left[1 - \left\{1 - \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-c}\right\}^2\right]^{\alpha} \quad (2.24)$$

$x, b, c, a, \alpha > 0$

De 2.23 en expansion par série binomiale

$$\left[1 - \left\{1 - \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-c}\right\}^2\right]^{\alpha-1} = \sum_{j=0}^{\infty} \binom{\alpha-1}{j} (-1)^j \left\{1 - \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-c}\right\}^{2j}$$

En ramplacant les résultat ci-dessus par 2.23, on peut obtenir

$$f_{TLDa}(x) = \sum_{j=0}^{\infty} \binom{\alpha-1}{j} (-1)^j 2\alpha ca \left(\frac{x}{b}\right)^{-a} \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-c-1} \left\{1 - \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-c}\right\}^{2j+1}$$

On a aussi

$$\left\{1 - \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-c}\right\}^{2j+1} = \sum_{k=0}^{2j+1} \binom{2j+1}{k} (-1)^k \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-ck}$$

Insertion cette relation en 2.23 on obtient

$$f_{TLDa} = \sum_{j=0}^{\infty} \sum_{k=0}^{2j+1} \binom{\alpha-1}{j} \binom{2j+1}{k} (-1)^{j+k} 2\alpha c \left(\frac{1}{b}\right)^{-a} ax^{-a-1} \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-[c(k+1)+1]}$$

Par conséquent la densité des $TLDa$ est constituée de la somme de séries infinies et densité du Dagum, enfin 2.23 prend la forme

$$f_{TLDa} = \sum_{j=0}^{\infty} \sum_{k=0}^{2j+1} W_{j,k} \left(\frac{1}{b}\right)^{-a} ax^{-a-1} \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-[c(k+1)+1]} \quad (2.25)$$

En inversant 2.24, on obtient

$$F^{-1}(u) = \left[\left(\frac{1}{b} \right)^a \left(\left\{ 1 - \left(1 - u^{\frac{1}{a}} \right)^{\frac{1}{2}} \right\}^{\frac{-1}{c}} - 1 \right) \right]^{\frac{-1}{a}}$$

$u \sim$ distribution uniforme allant de 0 à 1.

Les 1^{er}, 2^{eme}, 3^{eme} quartiles sont

$$Q_1 = x_{0.25} = \left[\left(\frac{1}{b} \right)^a \left(\left\{ 1 - \left(1 - 0.25^{\frac{1}{a}} \right)^{\frac{1}{2}} \right\}^{\frac{-1}{c}} - 1 \right) \right]^{\frac{-1}{a}}$$

$$Q_2 = x_{0.5} = \left[\left(\frac{1}{b} \right)^a \left(\left\{ 1 - \left(1 - 0.5^{\frac{1}{a}} \right)^{\frac{1}{2}} \right\}^{\frac{-1}{c}} - 1 \right) \right]^{\frac{-1}{a}}$$

$$Q_3 = x_{0.75} = \left[\left(\frac{1}{b} \right)^a \left(\left\{ 1 - \left(1 - 0.75^{\frac{1}{a}} \right)^{\frac{1}{2}} \right\}^{\frac{-1}{c}} - 1 \right) \right]^{\frac{-1}{a}}$$

Coefficient d'asymétrie (sk) et kurtosis (k) sont

$$sk = \frac{x_{0.75} + x_{0.25} - 2x_{0.5}}{x_{0.75} - x_{0.25}}$$

$$k = \frac{x_{\frac{3}{8}} - x_{\frac{1}{8}} + x_{\frac{7}{8}} - x_{\frac{5}{8}}}{x_{\frac{6}{8}} - x_{\frac{2}{8}}}$$

Le r^{ieme} moments sont des mesures utiles pour trouver la moyenne, la variance, asymétrie et l'aplatissement par rapport aux moments moyens pour la distribution de TLDa 2.25 elle est obtenu comme

$$\mu'_r = \sum_{j=0}^{\infty} \sum_{k=0}^{2j+1} W_{j,k} \left(\frac{1}{b} \right)^{-a} a \int_0^{\infty} x^{r-a-1} \left(1 + \left(\frac{x}{b} \right)^{-a} \right)^{-[c(k+1)+1]} dx \quad (2.26)$$

$$\text{Soit } y = \left(\frac{x}{b} \right)^{-a} \Rightarrow dx = \frac{dy}{-\left(\frac{1}{b} \right)^{-a} a x^{-a-1}}$$

Après quelques simplification de base de 2.26 on obtient

$$\mu'_r = \sum_{j=0}^{\infty} \sum_{k=0}^{2j+1} W_{j,k} \left(\frac{1}{b} \right)^{-r} \int_0^{\infty} y^{\frac{r}{a}} (1+y)^{-[c(k+1)+1]} dy$$

$$\mu'_r = \sum_{j=0}^{\infty} \sum_{k=0}^{2j+1} W_{j,k} \left(\frac{1}{b} \right)^{-r} B\left(1 - \frac{r}{a}, c(k+1) + \frac{r}{a} \right) \quad (2.27)$$

$r = 1, 2, 3, \dots$

avec

$$B(a, b) = \int_0^{\infty} \frac{x^{a-1}}{(1+x)^{a+b}} dx,$$

est la fonction de Bêta de 2^{eme} type.

On prend $r = 1$, on obtient la moyenne de la distribution

$$\mu'_1 = \sum_{j=0}^{\infty} \sum_{k=0}^{2j+1} W_{j,k} \left(\frac{1}{b}\right) B\left(1 - \frac{1}{a}, c(k+1) + \frac{1}{a}\right) \quad (2.28)$$

Par la relation entre le r^{ieme} moments et la moyen

$$variance = \mu_2 = \mu'_2 - (\mu'_1)^2$$

L'écart moyen est obtenu comme :

$$\pi(x) = \int_0^{\infty} |x - \mu| f(x) dx$$

Tel que $\mu = E(x)$ après la simplification

$$\pi(x) = 2 \left\{ \mu F(\mu) - \int_0^{\mu} x f(x) dx \right\} \quad (2.29)$$

avec

$$\int_0^{\mu} x f(x) dx = \sum_{j=0}^{\infty} \sum_{k=0}^{2j+1} W_{j,k} \left(\frac{1}{b}\right)^r B\left(1 - \frac{r}{a}, c(k+1) + \frac{r}{a}, y\right)$$

Mais $y = \frac{(\frac{1}{b})^{-a}}{\mu^a}$

Par conséquent

$$\pi(x) = 2 \left\{ \mu f(\mu) - \sum_{j=0}^{\infty} \sum_{k=0}^{2j+1} W_{j,k} \left(\frac{1}{b}\right)^r B\left(1 - \frac{r}{a}, c(k+1) + \frac{r}{a}, y\right) \right\} \quad (2.30)$$

Supposons que la variable aléatoire $X \sim TLDA(\alpha, b, p, a)$ et que nous tirons un échantillon aléatoire de taille n , soit $x_1, x_2, x_3, \dots, x_n$ sont n observation alors la fonction log-vraisemblance

$$L(\theta) = \log \prod_{i=1}^n f(x)$$

est :

$$\begin{aligned}
 L(\theta) = & n \log(2) + n \log(\alpha) + n \log(b^a) + n \log(c) + n \log(a) - (a+1) \sum_{i=1}^n \log(x_i) \\
 & - (c+1) \sum_{i=1}^n \log\left(1 + \left(\frac{x_i}{b}\right)^{-a}\right) + \sum_{i=1}^n \log\left\{1 - \left(1 + \left(\frac{x_i}{b}\right)^{-a}\right)^{-c}\right\} \\
 & + (\alpha-1) \sum_{i=1}^n \log\left[1 - \left\{1 - \left(1 + \left(\frac{x_i}{b}\right)^{-a}\right)^{-c}\right\}^2\right]
 \end{aligned}$$

2.6 Distribution de Exponentielle Dagum

Soit T une variable aléatoire avec une fonction de densité de probabilité $\lambda e^{-\lambda t}$, $t > 0$, $\lambda > 0$. Alors, la fonction de distribution cumulative de la famille exponentielle généralisée (EGE) est défini comme :

$$G(x) = \int_0^{-\log[1-(1-\bar{F}^d(x))^c]} \lambda e^{-\lambda t} dt = 1 - \left\{1 - \left[1 - (1 - F(x))^d\right]^c\right\}^\lambda \quad (2.31)$$

Alors la fonction de distribution cumulative de la distribution exponentielle généralisée Dagum (EGEDD) est donnée

$$G(x) = 1 - \left\{1 - \left[1 - \left(1 - \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-c}\right)^d\right]^p\right\}^\lambda \quad (2.32)$$

Où les paramètres b, p, a, λ, c et d sont non négatifs, avec p, a, λ, c et d étant paramètres de forme et b un paramètre d'échelle.

La fonction de densité de probabilité correspondant de l'EGEDD est donnée par :

$$g(x) = \frac{\left(\frac{1}{b}\right)^{-a} p \lambda a c d \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-c-1} \left(1 - \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-c}\right)^{d-1} \left[1 - \left(1 - \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-c}\right)^d\right]^{p-1}}{x^{a+1} \left\{\left[1 - \left(1 - \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-c}\right)^d\right]^p\right\}^{1-\lambda}} \quad (2.33)$$

$x > 0$,

La fonction de densité de probabilité de l'EGEDD peut être exprimé en termes de la fonction de densité de la distribution de Dagum comme

$$g(x) = \lambda p d \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} W_{ijk} f_D(x; b, a, \beta_{k+1}) \quad (2.34)$$

$\lambda > 0, b > 0, a > 0, \beta_{k+1} > 0, p > 0, d > 0, x > 0$

où $f_D(x; b, a, \beta_{k+1})$ est la fonction de densité de probabilité de la distribution de Dagum avec paramètre b, a et $\beta_{k+1} = \beta(k+1)$ et

$$W_{ijk} = \frac{(-1)^{i+j+k} \Gamma(\lambda) \Gamma(p(i+1)) \Gamma(d(j+1))}{i! j! (k+1)! \Gamma(\lambda-i) \Gamma(p(i+1)-j) \Gamma(d(j+1)-k)}, \Gamma(a+1) = a!.$$

La fonction quantile de l'EGEDD pour $u \in (0, 1)$ est donnée par

$$Q_X(u) = \left\{ b^a \left[\left(1 - \left(1 - \left(1 - u \right)^{\frac{1}{\lambda}} \right)^{\frac{1}{p}} \right)^{\frac{1}{d}} - 1 \right]^{\frac{-1}{c}} - 1 \right\}^{\frac{-1}{a}} \quad (2.35)$$

Le moment central de l'EGEDD est donnée par

$$\mu'_r = \lambda p d b^r \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} W_{ijk} \beta_{k+1} B\left(\beta_{k+1} + \frac{r}{a}, 1 - \frac{r}{a}\right) \quad (2.36)$$

$r < 0$, ou $B(.,.)$ est la fonction Bêta et $r = 1, 2, \dots$

Soit $x_1, x_2, x_3, \dots, x_n$ un échantillon aléatoire de taille n de l'EGEDD, $Z_i = \left(1 + \left(\frac{x_i}{b}\right)^{-a}\right)^{-c}$ puis la fonction log-vraisemblance est

$$\begin{aligned} l = & n \log(b^a \lambda p a c d) - (a+1) \sum_{i=1}^n \log(x_i) - (c+1) \sum_{i=1}^n \log(Z_i) + (d-1) \sum_{i=1}^n \log(1 - Z_i^{-c}) \\ & + (p-1) \sum_{i=1}^n \log[1 - (1 - Z_i^{-c})^d] + (\lambda-1) \sum_{i=1}^n \log\{1 - [1 - (1 - Z_i^{-c})^d]^p\} \end{aligned}$$

2.7 La Distribution de log-Dagum

Une variable aléatoire positive Y est distribuée par Dagum si sa fonction de distribution cumulative est donnée par 2.2

La transformation logarithmique de Y , $X = \ln Y$, a la fonction de distribution cumulative suivante

$$F_X(x; a, b, c) = F_Y(e^x; a, b, c) = \left(1 + \left(\frac{e^x}{b} \right)^{-a} \right)^{-c} \quad (2.37)$$

où, contrairement au modèle de Dagum, $x \in \mathbb{R}, c > 0$ est un paramètre de forme, $b > 0$ influence uniquement l'emplacement tandis que $a > 0$ est un paramètre d'échelle. La fonction de densité de probabilité est

$$f_X(x; a, b, c) = a b^a c e^{-ax} \left(1 + \left(\frac{e^x}{b} \right)^{-a} \right)^{-c-1}$$

Le modèle Log-Dagum sera désigné par $LDa(a, b, c)$

Nous obtenons l'expression simple et fermée du p^{ieme} quantile de $LDa(a, b, c)$

$$x_p = \frac{1}{a} \ln \left(\frac{b^a}{p^{\frac{1}{c}} - 1} \right) \quad (2.38)$$

La fonction génératrice de moments de $LDa(a, b, c)$ est égale au moment d'ordre t de la variable aléatoire de Dagum

$$m_X(t) = E[e^{tX}] = E[e^Y] = cbB\left(c + \frac{t}{a}, 1 - \frac{t}{a}\right), a > t$$

où $B(.,.)$ est la fonction mathématique de Bêta. Néanmoins, afin de calculer les moments de $LDa(a, b, c)$, il est plus pratique d'utiliser la fonction génératrice du cumulants, $\ln[m_X(t)]$. dont le r^{ieme} cumulants est donné par

$$K_r(t) = \left\{ \frac{\partial^r \ln[m_X(t)]}{\partial t^r} \right\}_{t=0}$$

En utilisant la fonction génératrice du cumulants, il est possible de prouver que pour tout $a, b, c > 0$, les deux premiers moments de $LDa(a, b, c)$ sont

$$E(X) = a^{-1} [\ln b^a + \Psi(c) - \Psi(1)]$$

$$E(X^2) = a^{-2} \left\{ [\Psi'(c) + \Psi'(1)] + [\ln b^a + \Psi(c) - \Psi(1)]^2 \right\}$$

où $\Psi(.), \Psi'(.)$ sont respectivement digamma et trigamma.

Soit $x = (x_1, x_2, \dots, x_n)$ une réalisation de l'échantillon aléatoire $X = (X_1, X_2, \dots, X_n)$, où X_1, X_2, \dots, X_n sont des variables aléatoires i.i.d. selon $LDa(a, b, c)$. Alors, la fonction de log-vraisemblance de x est

$$\ell(a, b, c; X) = n \ln(abc) - a \sum_{i=1}^n x_i - (c+1) \sum_{i=1}^n \ln \left(1 + \left(\frac{e^{x_i}}{b} \right)^{-a} \right)$$

La maximisation de $\ell(a, b, c; X)$ n'admet aucune solution explicite. Par conséquent, les ML estimations $\hat{\theta}_n = (\hat{a}_n, \hat{b}_n, \hat{c}_n)$ ne peuvent être obtenues qu'au moyen de procédures numériques telles que la méthode des scores de Fisher.

Sous les conditions de régularité habituelles, les propriétés asymptotiques bien connues de la méthode ML garantissent que $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} \mathcal{N}(0, \Sigma_\theta)$ où $\Sigma_\theta = [I(\theta)]^{-1}$ est la variance-correspondance asymptotique de variance-covariance et $I(\theta)$ est la matrice d'information de Fisher dans une observation unique, dont les éléments sont

$$i_{cc} = \frac{1}{c^2}, i_{cb} = \frac{1}{b^a(c+1)}, i_{ca} = -\frac{\ln b^a + \Psi(c) - \Psi(2)}{a(c+1)}, i_{bb} = \frac{c}{b^{2a}(c+2)},$$

$$i_{ba} = -c \frac{\ln b^a + \Psi(c+1) - \Psi(2)}{b^a a(c+2)}, i_{aa} = \frac{1}{a^2} (1 + A_{1,a} + A_{2,a})$$

où

$$A_{1,a} = \frac{c}{c+2} \left\{ \begin{array}{l} [\ln b^a + \Psi(c+1) - \Psi(2)]^2 + \Psi'(c+1) - 2\Psi'(c+3) + \Psi'(2) \\ + 2[\Psi(c+1) - \Psi(c+3)][\Psi(2) - \Psi(c+3)] \end{array} \right\}$$

$$A_{2,a} = 2c(c+1) \left[\frac{\Psi(c+3) - \Psi(1)}{(c+2)^2} - \frac{\Psi'(c+3)}{c+2} - \frac{\Psi(c+2) - \Psi(1)}{(c+1)^2} + \frac{\Psi'(c+2)}{c+1} \right]$$

2.8 La Distribution exponentielle de Kumaraswamy-Dagum

2.8.1 la Distribution de Kumaraswamy-Dagum

[Kumaraswamy,P.(1980)] a introduit une distribution a deux paramètres sur $(0, 1)$ son fonction de distribution cumulative est donnée

$$G(x) = 1 - (1 - x^\Psi)^\Phi$$

$$x \in (0, 1), \Psi > 0 \text{ et } \Phi > 0$$

La famille des distribution de kumaraswamy-G avec la fonction de distribution cumulative est donnée par

$$G_K(x) = 1 - (1 - F^\Psi(x))^\Phi$$

Pour $\Psi > 0$ et $\Phi > 0$, en laissant $F(x) = G_K(x)$ nous obtenons la distribution Kumaraswamy-Dagum (KD) avec fonction de distribution cumulative

$$G_{KD} = 1 - (1 - G_D^\Psi(x))^\Phi$$

La fonction de densité de probabilité de la distribution de Kumaraswamy est donnée par :

$$G_K(x) = \Psi \Phi F^{\Psi-1}(x) (1 - F^\Psi(x))^\Phi$$

Pour $\Psi > 0$ et $\Phi > 0$

2.8.1.1 La Distribution Exponentielle de Kumaraswamy-Dagum (EKD)

En général la distribution EKD est $G_{EKD}(x) = [F_{KD}(x)]^\theta$ où $F_{KD}(x)$ est la fonction de distribution de base (Kum-Dagum) avec la fonction de densité de probabilité correspondante donnée par

$$g_{EKD}(x) = \theta [F_{KD}(x)]^{\theta-1} f_{KD}(x)$$

Alors la fonction de distribution cumulative et la fonction de densité de probabilité de la distribution EKD sont donnée par

$$G_{EKD}(x; c, b, a, \Phi, \theta) = \left\{ 1 - \left[1 - \left(1 + \left(\frac{x}{b} \right)^{-a} \right)^{-c} \right]^\Phi \right\}^\theta \quad (2.39)$$

et

$$g_{EKD}(x; c, b, a, \Phi, \theta) = pb^a a \Phi \theta x^{-a-1} \left(1 + \left(\frac{x}{b} \right)^{-a} \right)^{-c-1} \quad (2.40)$$

$$* \left[1 - \left(1 + \left(\frac{x}{b} \right)^{-a} \right)^{-c} \right]^{\Phi-1} \quad (2.41)$$

$$* \left\{ 1 - \left[1 - \left(1 + \left(\frac{x}{b} \right)^{-a} \right)^{-c} \right]^\Phi \right\}^{\theta-1}$$

Pour p, a, b, Φ, θ et $x > 0$ respectivement.

La fonction quantile de la distribution EKD est sous forme

$$G_{EKD}^{-1}(x) = x_q = b \left\{ \left[1 - \left(1 - q^{\frac{1}{\theta}} \right)^{\frac{1}{\Phi}} \right] - 1 \right\}^{\frac{-1}{a}} \quad (2.42)$$

Lorsque $\theta = 1$ on obtient la distribution de Kumaraswamy-Dagum

Lorsque $\Phi = \theta = 1$ on obtient la distribution de Dagum

Nous appliquons l'expansion en série

$$(1 - Z)^{p-1} = \sum_{j=0}^{\infty} \frac{(-1)^j \Gamma(p)}{\Gamma(p-j) j!} Z^j \quad (2.43)$$

Pour $p > 0$ et $|Z| < 1$ pour obtenir le développement en série de la distribution EKD, en utilisant 2.43

$$g_{EKD}(x) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} W(i, j) x^{-a-1} \left(1 + \left(\frac{x}{b} \right)^{-a} \right)^{-c(j+1)-1} \quad (2.44)$$

où

$$W(i, j) = cb^a a \Phi \theta \frac{(-1)^{i+j} \Gamma(\theta) \Gamma(\Phi i + \Phi)}{\Gamma(\theta - i) \Gamma(\Phi i + \Phi - j) i! j!}$$

Soit $t = (1 + (\frac{x}{b})^{-a})^{-1}$

La s^{ieme} moment non centraux de la distribution EKD est donnée par :

$$E(X^s) = \int_0^{\infty} x^s g_{EKD}(x) dx = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} W(i, j, s) B(c(j+1) + \frac{s}{a}, 1 - \frac{s}{a}) \quad (2.45)$$

où

$$W(i, j, s) = cb^s \Phi \theta \frac{(-1)^{i+j} \Gamma(\theta) \Gamma(\Phi i + \Phi)}{\Gamma(\theta - i) \Gamma(\Phi i + \Phi - j) i! j!}$$

et $s < a$.

La fonction gènératrice des moments de la distribution EKD est donnée par

$$M(t) = \sum_{r=0}^{\infty} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} W(i, j, r) \frac{t^r}{r!} B(c(j+1) + \frac{r}{a}, 1 - \frac{r}{a})$$

Pour $r < a$.

Soit $x = (x_1, \dots, x_n)^T$ un échantillon aléatoire de la distribution EKD et on a la vecteur de paramètre inconnu $\theta = (p, b, a, \Phi, \theta)$. la fonction de log-vraisemblance est

$$\begin{aligned} l(\theta) = & n(\ln(c) + \ln(b) + \ln(a) + \ln(\Phi) + \ln(\theta)) - (a+1) \sum_{i=1}^n \ln(x_i) \\ & - (c+1) \sum_{i=1}^n \ln(1 + (\frac{x_i}{b})^{-a}) + (\Phi-1) \sum_{i=1}^n \ln[1 - (1 + (\frac{x_i}{b})^{-a})^{-c}] \\ & + (\theta-1) \sum_{i=1}^n \ln\{1 - [1 - (1 + (\frac{x_i}{b})^{-a})^{-c}]^{\Phi}\} \end{aligned}$$

2.8.2 La distribution MC-Dagum

Certains fonction utiles qui sont employées dans ce qui suit sont donnée ci-dessous. les fonction Gamma et Diagamma sont donnée par

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt, \Psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$$

Respectivement

où

$$\Gamma'(x) = \int_0^{\infty} t^{x-1} (\log t) e^{-t} dt$$

Est la dérivée première de la fonction Gamma. la formule de la dérivée d'ordre n de la fonction Gamma est donnée par :

$$\Gamma^{(n)}(s) = \int_0^{\infty} z^{s-1} (\log z) \exp(-z) dz$$

Considérons une fonction de distribution cumulative arbitraire $G(x)$. la fonction de densité de probabilité $f(x)$ de la distribution de *MC – Donald* est donnée par :

$$f(x; \alpha, \beta, \gamma) = \frac{\gamma g(x)}{B(\alpha, \beta)} G^{\alpha\gamma-1}(x) (1 - G^\gamma(x))^{\beta-1} \quad (2.46)$$

Pour $\alpha > 0$, $\beta > 0$ et $\gamma > 0$

Notez que $g(x)$ est la fonction de densité de probabilité de la distribution $g(x) = dG(x)/dx$

nous écrivons $X \sim MC - G(\alpha, \beta, \gamma)$, la fonction de distribution cumulative de cette distribution généralisée est donnée par

$$F(x; \alpha, \beta, \gamma) = I_{G(x)^\gamma}(\alpha, \beta) = \frac{1}{B(\alpha, \beta)} \int_0^{G(x)^\gamma} W^{\alpha-1} (1 - W)^{\beta-1} dW \quad (2.47)$$

où

$$I_{G(x)^\gamma}(\alpha, \beta) = B(\alpha, \beta)^{-1} \int_0^{G(x)^\gamma} W^{\alpha-1} (1 - W)^{\beta-1} dW$$

Désigne le rapport de la fonction Bêta incomplète. la même équation peut être exprimé comme suit

$$F(x; \alpha, \beta, \gamma) = \frac{G(x)^{\alpha\gamma}}{\alpha B(\alpha, \beta)} [{}_2F_1(\alpha, 1 - \beta; \alpha + 1; G(x)^\gamma)] \quad (2.48)$$

où

$${}_2F_1(\alpha, \beta, \gamma; x) = B(\beta, \gamma - \beta)^{-1} \int_0^1 \frac{t^{\beta-1} (1 - t)^{\gamma-\beta-1}}{(1 - tz)^\alpha} dt \quad (2.49)$$

est la fonction hypergéométrique et

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

Alors la fonction de densité de probabilité de la distribution *MC – Dagum* est :

$$f(x; a, b, c, \alpha, \beta, \gamma) = \frac{\gamma c b^a a x^{-a-1}}{B(\alpha, \beta)} \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-c\alpha\gamma-1} \left[1 - \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-\gamma c}\right]^{\beta-1} \quad (2.50)$$

pour $a, b, c, \alpha, \beta, \gamma > 0$, la fonction de distribution cumulative de cette distribution

est donnée par

$$F(x) = I_{(1+(\frac{x}{b})^{-a})^{-c\gamma}}(\alpha, \beta) \quad (2.51)$$

où

$$I_y(\alpha, \beta) = \frac{1}{B(\alpha, \beta)} \int_0^y W^{\alpha-1} (1-W)^{\beta-1} dW$$

Est la fonction Bêta incomplet. la fonction de distribution cumulative être écrit comme suit

$$F(x) = \frac{(1+(\frac{x}{b})^{-a})^{-c\alpha\beta}}{\alpha B(\alpha, \beta)} \left[{}_2F_1(\alpha, 1-\beta; \alpha+1; (1+(\frac{x}{b})^{-a})^{-c\gamma}) \right] \quad (2.52)$$

où

$${}_2F_1(\alpha, \beta, \gamma; x) = B(\beta, \gamma - \beta)^{-1} \int_0^1 \frac{y^{\beta-1} (1-y)^{\gamma-\beta-1}}{(1-yz)^\alpha} dy \quad (2.53)$$

Si $\alpha = \beta = \gamma = 1$, nous avons la distribution de Dagum

Le K^{ieme} moment non centraux de la distribution de MC-Dagum est donnée par

$$E(X^K) = \int_0^\infty x^K \frac{\gamma^c}{aB(\alpha, \beta)} (1+(\frac{x}{b})^{-a})^{-c\alpha\gamma-1} (1-(1+(\frac{x}{b})^{-a})^{c\gamma})^{\beta-1} dx \quad (2.54)$$

Soit $O = (a, b, c, \alpha, \beta, \gamma)^T$ pour estimer les paramètres $a, b, c, \alpha, \beta, \gamma$ de la distribution de MC-Dagum nous utilisons la méthode d'estimation du maximum de vraisemblance

Soit x_1, x_2, \dots, x_n un échantillon aléatoire de $f(x; b, c, a, \alpha, \beta, \gamma)$ la fonction log-vraisemblance $L(b, c, a, \alpha, \beta, \gamma)$ est :

$$\begin{aligned} L(b, c, a, \alpha, \beta, \gamma) &= n \log(\gamma) + n \log(c) + n \log(b^a) + n \log(a) - n \log B(\alpha, \beta) \\ &\quad - (a+1) \sum_{i=1}^n \log x_i - (c\alpha\gamma + 1) \sum_{i=1}^n \log \left[1 + \left(\frac{x_i}{b} \right)^{-a} \right] \\ &\quad + (\beta - 1) \sum_{i=1}^n \log \left[1 - \left(1 + \left(\frac{x_i}{b} \right)^{-a} \right)^{-\gamma c} \right] \end{aligned}$$

2.8.3 La distribution de Weibull-Dagum

Tout d'abord, [Domma et al.(2013)] utilisent la classe Bêta-G ([Eugene et al, (2002)];[Jones,(2004)] ont étendu le modèle de Dagum en introduisant deux paramètres de forme positifs supplémentaires α et β dont le rôle est de gouverner l'asymétrie et le poids de la queue. La fonction de distribution cumulative et la fonction de densité de probabilité du modèle de la distribution Bêta-Dagum (BD) à cinq paramètres sont données par

$$F_1(x; \alpha, \beta, a, b, c) = \frac{1}{B(\alpha, \beta)} \int_0^{G_{a,b,c}(x)} w^{\alpha-1} (1-w)^{\beta-1} dw = I_{G_{a,b,c}(x)}(\alpha, \beta)$$

et

$$f_1(x; \alpha, \beta, a, b, c) = \frac{cb^a ax^{-a-1}}{B(\alpha, \beta)} \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-\alpha c-1} \left[1 - \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-c}\right]^{\beta-1}$$

où $I_w(\alpha, \beta)$ est le rapport de la fonction Bêta incomplète et $G_{a,b,c}(x)$ est donnée par [2.2](#)

tel que $B(p, q) = \int_0^1 w^{p-1} (1-w)^{q-1} dw$ est la fonction Bêta incomplète

Deuxièmement, [[Oluyede et al, \(2013\)](#)] ont adopté la classe McDonald-G ([[Alexander,\(2012\)](#)]) pour définir la distribution McDonald-Dagum (McD) avec six paramètres positifs. Sa fonction de distribution cumulative et la fonction de densité de probabilité sont donnés par [2.47](#) et [2.50](#) respectivement.

[[Zografos et al, \(2009\)](#)] ont été les premiers à créer une classe de distributions gamma-G polyvalente et flexible basée sur la distribution gamma généralisée de Stacy et la théorie de la valeur record. Plus récemment, [[Bourguignon,\(2014\)](#)] ont proposé la classe de distributions Weibull-G, influencée par la classe gamma-Classe G. Soit $G(X; O)$ et $g(X; O)$ désignent les fonctions cumulatives et de densité d'un modèle de base.

avec le vecteur de paramètres O et la fonction de distribution cumulative de Weibull $\pi_W(t) = 1 - e^{-t^\beta}$ (pour $t > 0$) avec le paramètre d'échelle 1 et le paramètre de forme $\beta > 0$. [[Bourguignon,\(2014\)](#)] ont remplacé l'argument t par $G(t; O)/\bar{G}(t; O)$ où $\bar{G}(t; O) = 1 - G(t; O)$, et ont défini la fonction de distribution cumulative de leur classe, dite Weibull-G($t; O$), par

$$F(x) = F(x; \beta, O) = \beta \int_0^{\frac{G(x; O)}{\bar{G}(x; O)}} t^{\beta-1} e^{-t^\beta} dt = 1 - e^{-\left[\frac{G(x; O)}{\bar{G}(x; O)}\right]^\beta}, x \in \mathbb{R}, \beta > 0 \quad (2.55)$$

L'utilisation des cotes $G(x; O)/\bar{G}(x; O)$ est une fonction croissante de X . Alors, la fonction de densité de probabilité de Weibull-G est donnée par

$$f(x) = f(x; \beta, O) = \beta g(x; O) \left| \frac{G(x; O)}{\bar{G}(x; O)} \right|^{\beta-1} e^{-\left[\frac{G(x; O)}{\bar{G}(x; O)}\right]^\beta}, x \in \mathbb{R}, \beta > 0 \quad (2.56)$$

alors la distribution de Weibull-Dagum (WD) est

En insérant [2.2](#) dans l'équation [2.55](#), on obtient la fonction de distribution cumu-

lative de WD à quatre paramètres (pour $x > 0$)

$$F(x) = F(x; \beta, a, b, c) = 1 - e^{-\left[\left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-c} - 1\right]^{-\beta}} \quad (2.57)$$

La fonction de densité de probabilité correspondant à 2.57 est donnée par

$$f(x) = f(x; \beta, a, b, c) = \beta b^a a c x^{-a-1} \frac{\left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-\beta c - 1}}{\left[1 - \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^c\right]} e^{-\left[\left(1 + \left(\frac{x}{b}\right)^{-a}\right)^c - 1\right]^{-\beta}} \quad (2.58)$$

La fonction quantile est donné par

$$Q_X(u) = b \left\{ \left[\left[1 + (-\log(1-u))^{\frac{-1}{\beta}} \right]^{\frac{1}{c}} - 1 \right]^{\frac{-1}{a}} \right\} \quad (2.59)$$

tel que $u \sim U[0, 1]$

L'analyse de la variabilité de l'asymétrie et de l'aplatissement sur les paramètres de forme de X peut être étudiée sur la base de mesures quantiles. être étudiée sur la base de mesures quantiles. L'asymétrie de Bowley ([Kenney et al, (1962)]) basée sur les quartiles est donnée par

$$B = \frac{Q\left(\frac{3}{4}\right) + Q\left(\frac{1}{4}\right) - 2Q\left(\frac{2}{4}\right)}{Q\left(\frac{3}{4}\right) - Q\left(\frac{1}{4}\right)}$$

Les défauts de la mesure classique du kurtosis sont bien connus. Le kurtosis de Moors ([Moors, J.J.A. (1998)]) basée sur les octiles est donnée par

$$M = \frac{Q\left(\frac{3}{8}\right) - Q\left(\frac{1}{8}\right) + Q\left(\frac{7}{8}\right) - Q\left(\frac{5}{8}\right)}{Q\left(\frac{6}{8}\right) - Q\left(\frac{2}{8}\right)}$$

le r^{ieme} moment de X est donnée par

$$\mu_r = c b^r \sum_{k=0, j=0}^{\infty} (k\beta + j) \frac{(-1)^{k+1} \Gamma(k\beta + j)}{k! j! \Gamma(k\beta)} B\left(1 - \frac{r}{a}, (k\beta + j)\beta + \frac{r}{a}\right) \quad (2.60)$$

Nous considérons l'estimation des paramètres inconnus de la distribution de WD par la méthode du maximum de vraisemblance. Soit x_1, \dots, x_n est un échantillon aléatoire de taille n provenant de la distribution WD donnée par 2.56. La fonction de log-vraisemblance pour le vecteur de paramètres $O = (\beta, a, b, c)^T$ peut être exprimée comme suit

$$\begin{aligned} \ell &= \ell(O) = n \log(\beta abc) - (a+1) \sum_{i=1}^n \log x_i - (\beta c + 1) \sum_{i=1}^n \log \left(1 + \left(\frac{x_i}{b} \right)^{-a} \right) \\ &\quad - (\beta + 1) \sum_{i=1}^n \log \left[1 - \left(1 + \left(\frac{x_i}{b} \right)^{-a} \right)^{-c} \right] - \sum_{i=1}^n \left[\left(1 + \left(\frac{x_i}{b} \right)^{-a} \right)^c - 1 \right]^{-\beta} \end{aligned}$$

3.1 Introduction

Au cours de plus d'un siècle de littérature économique, plusieurs approches ont été proposées pour étudier l'inégalité des variables quantitatives, principalement les distributions de revenus. Au sein de ce courant de recherche, de nombreux indices synthétiques sont apparus pour résumer et comparer l'inégalité des distributions à l'aide d'un seul scalaire. Parmi eux, l'indice d'inégalité le plus célèbre est, sans aucun doute, le coefficient de [Gini(1914) .], qui a également une explication graphique à travers la courbe de [Lorenz (1905)]. En outre, une méthodologie basée sur le coefficient de Gini a été proposée dans la littérature fiscale pour mesurer le degré de progressivité de l'impôt, donnant lieu à l'indice de [Kakwani (1977b).] (basé sur le coefficient de Gini), considérant que les écarts de revenus relatifs se compriment lors de la transition entre la distribution avant et après impôt. De même, l'effet redistributif produit par l'impôt est mesuré par l'indice (basé sur le Gini) de [Reynolds et al]. Bien que la courbe de Lorenz soit un outil fondamental pour les comparaisons de bien-être ([Atkinson, (1970).]; [Shorrocks,(1983)]; [Atkinson et al (1987)]), nous n'en avons trouvé que quelques applications dans la littérature fiscale, car il n'est pas facile de tirer des conclusions spécifiques en examinant et en comparant les courbes de Lorenz et de concentration de différentes distributions (c'est-à-dire la distribution avant impôt, la distribution après impôt et la distribution fiscale). En effet, en raison de leur nature cumulative inhérente, les différentes courbes de Lorenz sont difficilement distinguables. Par conséquent, dans la plupart des recherches empiriques existantes, l'effet global de la fiscalité et de la politique de transfert est principalement dérivé des coefficients de Gini et de concentration (pour une application récente, voir [Guillaud (2019)]). Il y a quelques années, [Zenga (2007)] a proposé une nouvelle méthodologie pour tracer et mesurer les inégalités, dans laquelle la nouvelle courbe et le

nouvel indice sont basés sur des comparaisons entre le revenu moyen des personnes les plus pauvres et le revenu moyen de la partie restante la plus riche de la population. Plusieurs études récentes ont souligné les différentes caractéristiques de l'approche de Zenga par rapport au point de vue standard, basé sur la courbe de Lorenz. Par conséquent, nous allons plus loin dans ce courant de la littérature pour explorer l'efficacité de la première méthode dans l'étude des effets d'un impôt sur le revenu des personnes physiques et les interprétations et implications de la progressivité de l'impôt en termes de bien-être social ([Son (2013)]; [Kakwani et al, (2020)]) selon cette nouvelle procédure. En particulier, nous montrons principalement que la représentation graphique et certains outils analytiques basés sur la courbe d'inégalité de Zenga fournissent un instrument précis pour comprendre quelle partie d'une distribution avant impôt est principalement affectée par le système fiscal ou par une réforme fiscale.

3.1.1 Courbe de Lorenz

La courbe de Lorenz est une mesure d'inégalité la plus largement utilisée dans plusieurs domaines (épidémiologie, traitement du signal, psychologie expérimentale...etc).

Elle peut être facilement transposée, notamment la répartition d'une donnée statistique quelconque, comme les inégalités de répartition d'un actif ou de toute autre distribution de richesse, l'état de la répartition des clients au sein d'une clientèle, le revenu et la répartition des richesses. Dans le cas de l'analyse des revenus, la courbe de Lorenz $P(L)$ représente la part du revenu total détenue par la proportion $P \in [0;1]$ d'individus les plus pauvres :

$$L(P) = \frac{\text{revenu total des plus pauvres}}{\text{revenu total}}$$

La courbe de Lorenz pour une variable aléatoire X positive est définie comme le graphe du rapport

$$L(p) = \frac{\int_0^x t f(t)}{\int_0^{\infty} t f(t)}, 0 \leq p \leq 1$$

où

$$p = F(x)$$

$$L(p) = \frac{E(X|X \leq x)P(X \leq x)}{E(X)}$$

Avec les propriétés $L(p) \leq p$, $L(0) = 0$ et $L(1) = 1$: Si X représente le revenu annuel, $L(p)$ est la proportion du revenu total qui revient aux personnes ayant les revenus les

plus faibles de $100p\%$.

Si tous les individus gagnent le même revenu alors $L(p) = p$ pour tout p : La zone située entre la ligne $L(p) = p$ et la courbe de Lorenz peut être considérée comme une mesure de l'inégalité des revenus, ou plus généralement de la variabilité de X .

3.1.2 Indice de Gini

Indice de Gini, est une mesure statistique permettant de rendre compte de la répartition d'une variable (salaire, revenus, patrimoine) au sein d'une population. Autrement dit, il mesure le niveau d'inégalité de la répartition d'une variable dans la population.

Ce coefficient est typiquement utilisé pour mesurer l'inégalité des revenus dans un pays 1. Il a été développé par le statisticien italien Corrado Gini. Le coefficient de Gini est un nombre variant de 0 à 1, où 0 signifie l'égalité parfaite et 1, qui ne peut être atteint, signifierait une inégalité parfaite (une seule personne dispose de tous les revenus et une infinité d'autres n'ont aucun revenu) L'indice de Gini est calculé grâce à la courbe de Lorenz. Cette courbe affecte à chaque part de population, organisée par revenu croissant, la part que représentent ses revenus.

Alternativement, l'indice de Gini peut être défini comme la moitié de la différence moyenne relative de Gini de la série des revenus, c'est-à-dire comme la valeur :

$$G = \frac{E}{2M}$$

où M la moyenne des revenus et E représente la différence moyenne de Gini des revenus, c'est-à-dire la moyenne de tous les écarts en valeur absolue pour tous les couples de la variable statistique étudiée (cette différence moyenne mesure l'écart espéré entre les revenus de deux individus pris au hasard avec remise dans la population étudiée). Cela donne, si les (x_i) $1 \leq i \leq n$ sont les revenus des n individus :

$$E = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|$$

$$M = \frac{1}{n} \sum_{k=1}^n x_k$$

En pratique, on ne dispose pas de cette fonction, mais du revenu par « tranches » de la population. Pour n tranches, le coefficient s'obtient par la formule de Brown :

$$G = 1 - \sum_{k=0}^{n-1} (X_{k+1} - X_k)(Y_{k+1} + Y_k)$$

où X est la part cumulée de la population, et Y la part cumulée du revenu.

Pour n personnes ayant des revenus Y_i , pour i allant de 1 à n , indicés par ordre croissant ($Y_i \leq Y_{i+1}$) :

$$G = \frac{2 \sum_{i=1}^n i Y_i}{n \sum_{i=1}^n Y_i} - \frac{n+1}{n}$$

L'indice de Gini ne permet pas de tenir compte de la répartition des revenus. Des courbes de Lorenz différentes peuvent correspondre à un même indice de Gini. Si 50% de la population n'a pas de revenu et l'autre moitié a les mêmes revenus, l'indice de Gini sera de 0,5. On trouvera le même résultat de 0,5 avec la répartition suivante, pourtant moins inégalitaire 75% de la population se partage de manière identique 25% du revenu global d'une part, et d'autre part le 25% restant se partage de manière identique le 75% restant du revenu global.

L'indice de Gini ne fait pas de différence entre une inégalité dans les bas revenus et une inégalité dans les hauts revenus.

Le coefficient de Gini est principalement utilisé pour mesurer les inégalités de revenu, mais peut aussi servir à mesurer les inégalités de richesse ou de patrimoine.

Le coefficient de Gini en économie est souvent combiné avec d'autres données. Se situant dans le cadre de l'étude des inégalités, il va de pair avec la politique.

Il est aussi utilisé par les logisticiens en entrepôts pour étudier l'implantation des références en fonction des statistiques de sorties des articles. En informatique, le coefficient de Gini est employé dans le cadre de certaines méthodes d'apprentissage supervisé, comme les arbres de décision.

3.1.3 Coefficient de variation

Le coefficient de variation également nommé écart type relatif, est une mesure de dispersion relative. Le RSD (relative standard deviation en anglais) est défini comme la valeur absolue du coefficient de variation et est souvent exprimé en pourcentage.

Le coefficient de variation est défini comme le rapport entre l'écart-type et la moyenne :

$$c_v = \frac{\gamma}{\mu}$$

3.1.4 Indice de Zenga

L'écart croissant observé entre les individus les moins fortunés et les plus fortunés (voir, parmi beaucoup d'autres, [Piketty, T.]), a motivé une nouvelle réflexion sur l'inégalité et a donné lieu à de nombreuses propositions dans la littérature (voir [Zenga (2007)]; [Gastwirth, (2014).], [Gastwirth, (2016).]; [Davydov (2018)], et les références dans lesquelles). Il existe un consensus sur le fait qu'aucune mesure ne peut être considérée comme supérieure aux autres ([Osberg (2017).]); par conséquent, le choix d'une me-

sure d'inégalité doit reposer sur son adéquation à des problèmes de fond spécifiques [Jasso (1982.)]. En particulier, il y a quelques années, pour saisir les changements récents dans les parties extrêmes de la distribution des revenus, [Zenga (2007)] a proposé une nouvelle courbe d'inégalité $I_F(p)$ basée sur le contraste entre le revenu moyen des p pour cent de personnes les plus pauvres qui gagnent le moins $\mu_F^-(p)$ défini dans l'équation 3.1 et le montant qui est détenu, en moyenne, par les personnes les plus riches qui gagnent le plus, à savoir les $(1 - p)$ pour cent restants de la population :

$$\mu_F^-(p) = \frac{1}{p} \int_0^p F^{-1}(s) ds \quad (3.1)$$

$$\mu_F^+(p) = \frac{1}{1-p} \int_0^p F^{-1}(s) ds$$

Par conséquent, [Zenga (2007)] a défini la courbe $(p, I_F(p))$, où

$$I_F(p) = \frac{\mu_F^+(p) - \mu_F^-(p)}{\mu_F^+(p)}$$

pour $0 < p < 1$. Lorsque la variable aléatoire Z est égale à une constante, le quantile correspondant $F^{-1}(p)$ est également égal à la constante, ainsi que les moyennes inférieure et supérieure $\mu_F^-(p)$ et $\mu_F^+(p)$; ainsi, $I_F(p) = 0, \forall p \in (0, 1)$, indiquant une égalité parfaite ou une société égalitaire. L'autre scénario extrême est celui où, grosso modo, il n'y a qu'un seul membre de la société qui obtient la totalité du revenu de la population; ainsi, $I_F(p) = 1, \forall p \in (0, 1)$. Comme illustré par [?], cette approche considère que les notions de pauvres et de riches sont relatives les unes aux autres et résume, en une seule mesure, le degré d'inégalité au sein de la population par l'indice suivant :

$$I_F = \int_0^1 \frac{\mu_F^+(p) - \mu_F^-(p)}{\mu_F^+(p)} dp$$

L'indice de Zenga obéit à un certain nombre de propriétés qui peuvent être considérées comme intrinsèques au concept d'inégalité. Dans ce qui suit, nous examinons ces propriétés en utilisant la notation I_Z ou I_F pour l'indice, chaque fois que cela permet de simplifier la présentation.

Invariance d'échelle. L'indice d'inégalité de Zenga est une mesure d'inégalité relative, car les changements proportionnels de tous les revenus, où la variation de Z dans c_Z avec $c > 0$, ne modifient pas le niveau d'inégalité.

$$I_Z = I_{c_Z}. \quad (3.2)$$

Techniquement, on dit qu'il est homogène de degré zéro dans les revenus. De toute évidence, un indice satisfaisant à l'équation 3.2 gère l'illusion monétaire; à savoir que

si les revenus sont mesurés en livres au lieu de dollars, l'inégalité ne change pas.

Sensibilité à la traduction. En ajoutant toute constante $c > 0$ au revenu Z , transformant Z en $Z + c$, la mesure relative de l'inégalité diminue :

$$I_{Z+c} \leq I_Z.$$

En d'autres termes, une mesure relative de l'inégalité doit considérer que les distances relatives au sein des revenus sont réduites en ajoutant un montant positif constant.

Ordre de Lorenz. Suivant [Aaberge (2001)], l'ordonnancement de Lorenz $Z \leq L_Y$ indique la borne $L_Z(p) \geq L_Y(p)$ pour tout $p \in [0, 1]$. Si les variables aléatoires Z et Y suivent l'ordre de Lorenz, alors

$$I_Z \leq I_Y$$

Principe de transfert de Pigou-Dalton. Le principe de transfert de Pigou-Dalton stipule que les transferts progressifs (c'est-à-dire des riches vers les pauvres) par ordre de rang et préservant la moyenne devraient diminuer la valeur des mesures d'inégalité, transformant Z en Y , ce qui donne

$$I_Z \geq I_Y.$$

Mesure de l'inégalité à l'aide des distributions de Dagum

L'outil le plus largement utilisé pour analyser et visualiser les inégalités de revenus est la courbe de Lorenz ([Lorenz,(1905)]), et plusieurs indices d'inégalité de revenus sont directement liés à cette courbe, notamment l'indice de Gini ([Gini.(1914)]). Puisque la fonction quantile de la distribution de Dagum est disponible sous forme fermée, son intégrale normalisée, la courbe de Lorenz

$$L(u) = \frac{1}{E(X)} \int_0^u F^{-1}(u) dt; u \in [0; 1]$$

est également d'une forme relativement simple, à savoir ([Dagum,C.(1977)])

$$L(u) = I_Z \left(c + \frac{1}{a}, 1 - \frac{1}{a} \right); 0 \leq u \leq 1 \quad (3.3)$$

Où $Z = u^{\frac{1}{c}}$, $I_Z(x, y)$ désigne le rapport de la fonction Bêta incomplète. Clairement, la courbe existe si $a > 1$.

En ce qui concerne les mesures scalaires de l'inégalité, le plus utilisé de tous ces indices, le coefficient de Gini, prend la forme ([Dagum,C.(1977)]) :

$$G = \frac{\Gamma(c)\Gamma\left(2c + \frac{1}{a}\right)}{\Gamma(2c)\Gamma\left(c + \frac{1}{a}\right)} - 1 \quad (3.4)$$

le coefficient de variation (CV) est de :

$$CV = \sqrt{\frac{\Gamma(c)\Gamma\left(c + \frac{2}{a}\right)\Gamma\left(1 - \frac{2}{a}\right)}{\Gamma^2\left(c + \frac{1}{a}\right)\Gamma^2\left(1 - \frac{1}{a}\right)}} - 1 \quad (3.5)$$

Rappelons que le coefficient de variation est une transformation monotone d'une mesure contenue dans la classe d'entropie généralisée des mesures d'inégalité (par exemple, [Kleiber.(2003)]).

La courbe de Zenga est maintenant dénommée en termes de quantiles $F^{-1}(u)$ de la distribution des revenus elle-même et de ceux de la distribution du premier moment correspondante, $F_{(1)}^{-1}(u)$ pour

$$Z(u) = \frac{F_{(1)}^{-1}(u) - F^{-1}(u)}{F_{(1)}^{-1}(u)} = 1 - \frac{F^{-1}(u)}{F_{(1)}^{-1}(u)}, 0 < u < 1 \quad (3.6)$$

l'ensemble $\{(u, Z(u)) | u \in (0, 1)\}$ est la courbe de concentration de Zenga. Notons que $F_{(1)} \leq F$ implique $F^{-1} \leq F_{(1)}^{-1}$, donc la courbe de Zenga appartient au carré unitaire. Il découle de que la courbe est sans échelle.

3.2 Simulations et Applications

3.2.1 Simulations

Dans cette section, on considère des simulations des distribution de classe Dagum avec différentes paramètres, et on calcul les différentes valeurs des mesures des inégalités, en particulier : cv :coefficient de variation

I.Gini :indice de Gini

I.Zenga :Indice de Zenga

Min :Minimum

Max :Maximum

n^{ieme} qua : n^{ieme} quartile

Distribution de Dagum

Pour $b = 1$, $a = 2$ et $c = 4$

TABLE 3.1 – Les mesures des inégalités de la distribution de Dagum

n	I.Gini	cv	I.Zenga
100	0.3083687	0.622423	0.880731
250	0.4052653	0.9214593	0.8593354
500	0.4027615	1.012073	0.850323
1000	0.4590984	1.842875	0.8462574

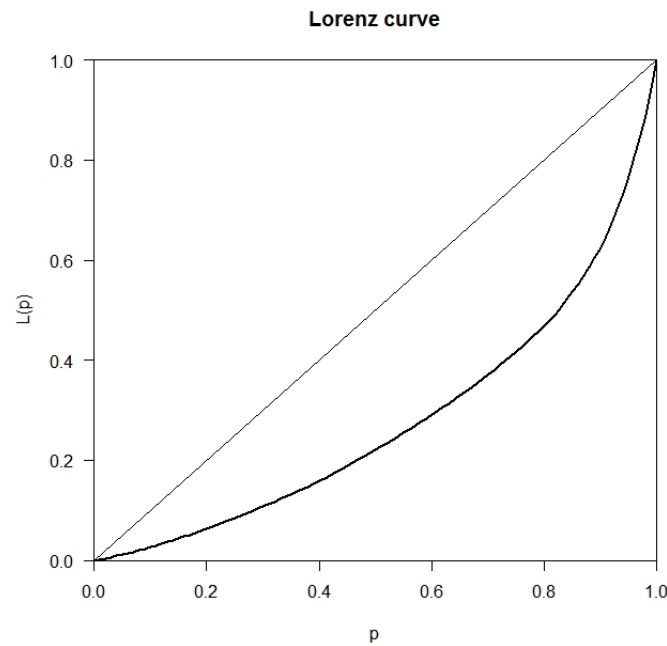


FIGURE 3.1 – La courbe de Lorenz de la distribution de Dagum

Certaines distributions qui sont des cas particuliers de la distribution de Dagum à trois paramètres sont

La distribution inverse paralogistique($a=c$)

Pour $b = 1$ et $a = c = 2$

TABLE 3.2 – Les mesures des inégalités de la distribution inverse paralogistique

n	I.Gini	cv	I.Zenga
100	0.4521175	1.082085	0.880924
250	0.4747236	1.324584	0.8605501
500	0.4592959	1.301037	0.8523369
1000	0.4450254	1.383419	0.8471686

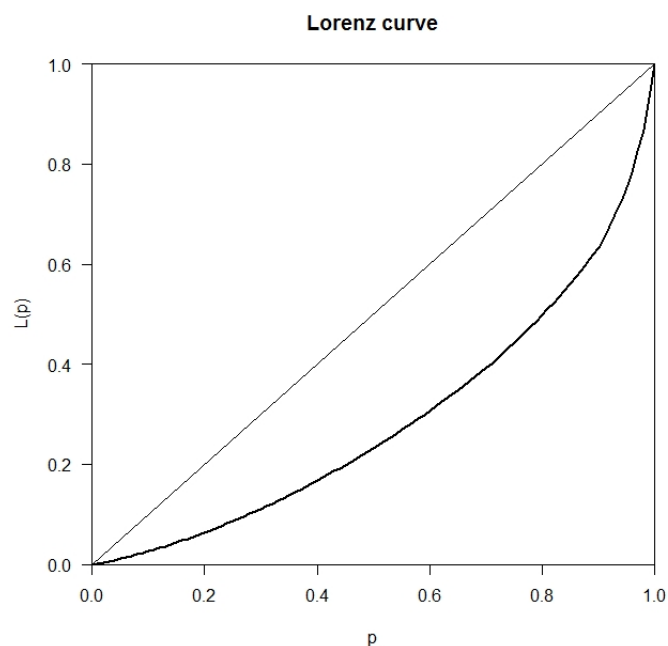


FIGURE 3.2 – La courbe de Lorenz de la distribution inverse paralogistique

La distribution inverse Lomax(a=1)

Pour $b=1, a=1$ et $c=3$

TABLE 3.3 – Les mesures des inégalités de la distribution inverse Lomax

n	I.Gini	cv	I.Zenga
100	0.7278327	2.227435	0.910590
250	0.8088214	5.200716	0.878697
500	0.8098182	4.880973	0.8575303
1000	0.7584805	4.266261	0.8508234

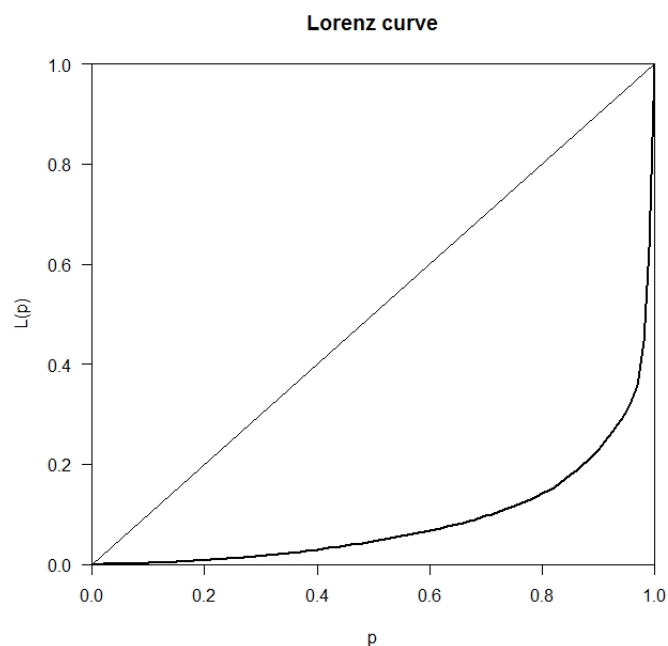


FIGURE 3.3 – La courbe de Lorenz de La distribution inverse Lomax

La distribution de Fisk($c=1$)

Pour $b=1, a=4$ et $c=1$

TABLE 3.4 – Les mesures des inégalités de la distribution de Fisk

n	I.Gini	cv	I.Zenga
100	0.2455303	0.4828938	0.8700124
250	0.2691379	0.6032144	0.8528438
500	0.2457758	0.5352745	0.8476799
1000	0.2569182	0.5382518	0.845571

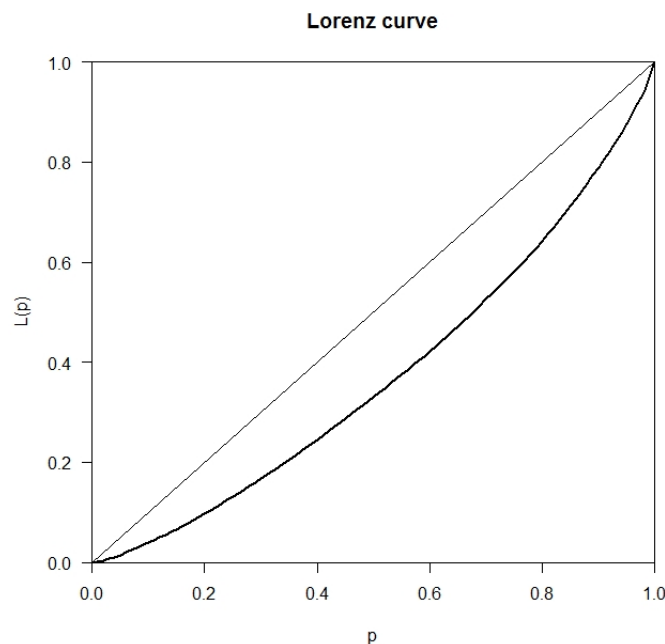


FIGURE 3.4 – La courbe de Lorenz de la distribution de Fisk

3.2.2 Applications

Produit intérieur brut par habitant

Le produit intérieur brut par habitant, ou par tête (PIB par habitant ou par tête) est un indicateur du niveau d'activité économique. C'est la valeur du PIB divisée par le nombre d'habitants d'un pays. Il est plus efficace que le PIB pour mesurer le développement d'un pays, cependant, il n'est qu'une moyenne donc il ne permet pas de rendre compte des inégalités de revenu et de richesse au sein d'une population

On a la population suivant tirée de la banque mondiale (PIB par habitant en Algérie) et consiste en 58 observations

(<https://databank.worldbank.org/reports.aspx?source=2&series=NY.GDP.PCAP.CD&country=>)

GDP.PCAP.CD&country=)

Min	1 ^{er} quartile	médiane	moyenne	3 ^{eme} quartile	Max	écart type
172.2	1076.4	1795.0	2213.9	3036.7	5591.2	1551.057

TABLE 3.5 – Les indicateurs de position de le PIB par habitants en Algérie

I.Gini	CV	I.Zenga
0.388179	0.7005935	0.9897309

TABLE 3.6 – Les mesures des inégalités de le PIB par habitants en Algérie

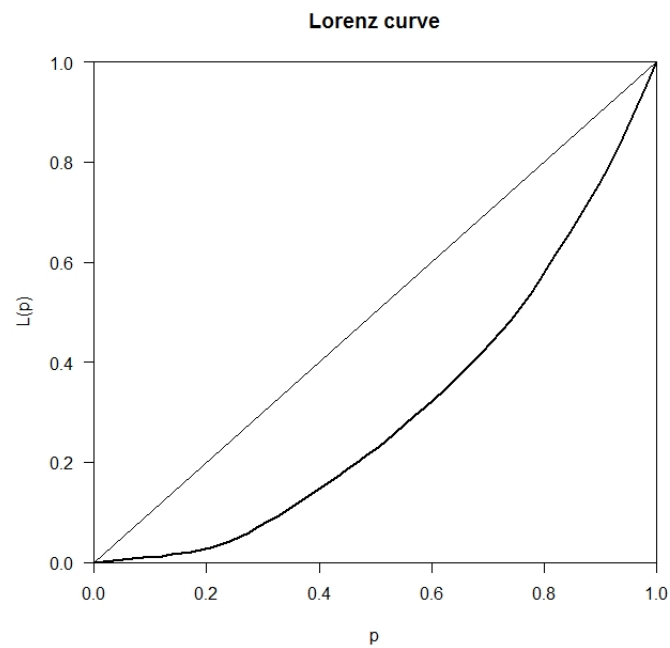


FIGURE 3.5 – La courbe de Lorenz de le PIB par habitants en Algérie

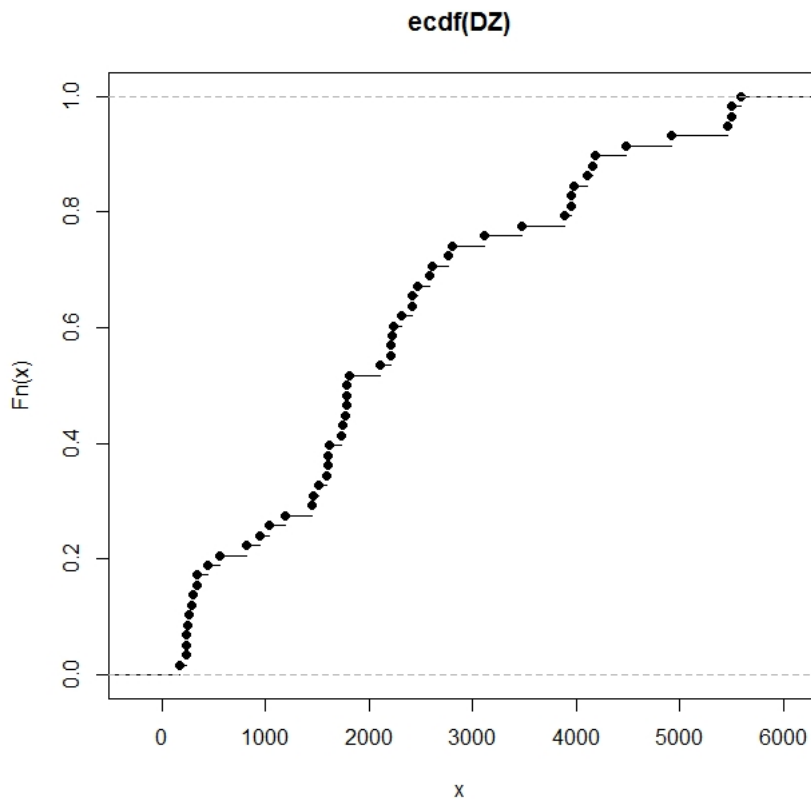


FIGURE 3.6 – Graphe de la fonction de répartition empirique de le PIB par habitants en Algérie

TABLE 3.7 – Résultats de l’ajustement des données PIB par quelques lois usuelles

Lois	paramètres	AIC	BIC
Weibull	forme 1.221051 échelle 2480.228	1005.676	1009.797
Normal	moyenne 2213.918 écart type 1537.628	1019.804	1023.925
lognormal	moyenne 7.347657 écart type 0.9742509	1017.899	1022.02
GEV	moyenne 1507.66802680 échelle 1272.66869937 forme -0.01669533	1015.26	1021.442
Dagum	forme.a 0.9159081 échelle 4.4114292 forme.c 1.4165255	387.8322	394.0135

d’après les résultats d’ajustement la distribution de Dagum est la meilleur pour ajuster les données des revenus.

CONCLUSION

Dans ce travail, nous avons étudié et analysé la distribution de Dagum et ses propriétés, ensuite, les distribution émanant à partir de cette distribution. L'analyse des revenus dans l'économie qu'est considéré comme une préoccupation centrale, la modélisation des revenus par la distribution de dagum est montré par l'ajustement des données réelles par cette distribution, grâce à sa flexibilité totale, en identifiant quelques mesures d'inégalités (indice de Gini, indice de Zenga, coefficient de variation et la courbe de Lorenz) pour résumer et comparer l'inégalité des distributions qu'ont été proposées pour étudier l'inégalité des variables quantitatives, principalement les distributions de revenus. D'autre part, tous les membres appartenant inégalité GEM ont des fonctions d'utilité et de dés utilité strictement utilitaires, ce qui exclut toute comparaison interpersonnelle des utilités et des dés-utilités. Par la suite on a ajusté et d'après les résultats qu'on a obtenue la distribution de Dagum est la meilleure.

BIBLIOGRAPHIE

- [Zenga(2010)] Zenga, M. (2010). Mixture of Poliscchio's truncated Pareto distributions with beta weights. Mixture of Poliscchio's truncated Pareto distributions with beta weights.
- [Gibrat, (1931).] Gibrat, R. (1931). Les Inegalit es Economiques. Paris : Librairie du Recueil Sirey.
- [Fisk,P.R.1961] Fisk, P. R. (1961). The graduation of income distributions. *Econometrica : journal of the Econometric Society*, 171-185..
- [D'Addario,R.(1949)] D'Addario, R. (1949). Ricerche sulla curva dei redditi. *Giornale degli economisti e Annali di economia*, 91-114.
- [Dagum,c.1975] Dagum, C. (1975). A model of income distribution and the conditions of existence of moments of nite order. *Bulletin of the International Statistical Institute*, 46 (Proceedings of the 40th Session of the ISI, Warsaw, Contributed Papers), 199-205.
- [Dagum,C.(1977)] Dagum, C. (1977). New model of personal income-distribution-specification and estimation. *Economie appliquée*, 30(3), 413-437.
- [Dagum,C.(1980b)] Dagum, C. (1980). Generating systems and properties of income distribution models.
- [Dagum(1980c)] Dagum, C. (1980). Sistemas generadores de distribución del ingreso y la Ley de Pareto. *El trimestre económico*, 47(188 (4), 877-917.
- [Dagum,C.(1983)] Dagum, C. (1983).. Income distribution models. In : S.Kotz, N.L. Johnson, and C.Read (eds.) : *Encyclopedia of Statistical Sciences*, Vol. 4. New York : John Wiley, pp.27-34.
- [Majumed, (1990)] Majumder, A., and Chakravarty, S.R. (1990). Distribution of personal income : Development of a new model and its ap-

- plication to US income data. *Journal of Applied Econometrics*, 5, 189-196.
- [Kotz et al.(1983)] Kotz, S., Johnson, N.L., and Read, C. (eds.) (1983). *Encyclopedia of Statistical Sciences*, Vol. 4, New York : John Wiley.
- [Singh, (1976)] Singh, S. K., & Maddala, G. S. (2008). A function for size distribution of incomes. In *Modeling income distributions and Lorenz curves* (pp. 27-35). Springer, New York, NY.
- [Kleiber.(2003)] Kleiber, C., & Kotz, S. (2003). *Statistical size distributions in economics and actuarial sciences* (Vol. 470). John Wiley & Sons.
- [McDonald.(1984)] McDonald, J.B. (1984). Some generalized functions for the size distribution of income. *Econometrica*, 52, 647-663.
- [Lorenz,(1905)] Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. *Publications of the American statistical association*, 9(70), 209-219.
- [Shao.(2002)] Shao, Q. (2002). Maximum likelihood estimation for generalised logistic distributions.
- [Gini.(1914)] Gini, C. (1914). Sulla misura della concentrazione e della variabilità dei caratteri. *Atti del Reale Istituto veneto di scienze, lettere ed arti*, 73, 1203-1248.
- [Domanski.(1998)] Domański, C., & Jędrzejczak, A. (1998). Maximum likelihood estimators of the dagum model parameters. *International Advances in Economic Research*, 4(3), 243-252.
- [Domma.(1997)] Domma, F. (1997). Mediana e range campionario per il modello di Dagum. *Quaderni di Statistica e Matematica Applicata alle Scienze Economico-Sociali*, 19, 195-204.
- [Latorre.(1988)] Latorre, G. (1988). Proprietà campionarie del modello di Dagum per la distribuzione dei redditi. *Statistica*, 48, 15-27.
- [Cheli,(1995)] Cheli, B., Lemmi, A., & Spera, C. (1995). An EM algorithm for estimating mixtures of Dagum's models. *Income distribution, social welfare, inequality and poverty*, 6, 131-142.
- [Gleaton,(2006)] Gleaton, J. & Lynch, J. (2006), Properties of generalized log-logistic families of lifetime distributions', *Journal of Probability and Statistical Science* 4(1), 51-64.
- [Kumaraswamy,P.(1980)] Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of hydrology*, 46(1-2), 79-88.

- [Alshomrani,(2016)] Al-Shomrani, A., Arif, O., Shawky, A., Hanif, S., & Shahbaz, M. Q. (2016). Topp–Leone family of distributions : some properties and application. *Pakistan Journal of Statistics and Operation Research*, 443-451.
- [Domma et al.(2013)] Domma, F., & Condino, F. (2013). The beta-dagum distribution : definition and properties. *Communications in Statistics-Theory and Methods*, 42(22), 4070-4090..
- [Eugene et al, (2002)] Eugene, N., Lee, C., & Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics-Theory and methods*, 31(4), 497-512.
- [Jones,(2004)] Jones, M. C. (2004). Families of distributions arising from distributions of order statistics. *Test*, 13(1), 1-43.
- [Oluyede et al, (2013)] Oluyede, B. O., & Rajasooriya, S. (2013). The Mc-Dagum distribution and its statistical properties with applications. *Asian Journal of Mathematics and Applications*, 2013.
- [Alexander,(2012)] Alexander, C., Cordeiro, G. M., Ortega, E. M., & Sarabia, J. M. (2012). Generalized beta-generated distributions. *Computational Statistics & Data Analysis*, 56(6), 1880-1897.
- [Zografos et al, (2009)] Zografos, K., & Balakrishnan, N. (2009). On families of beta-and generalized gamma-generated distributions and associated inference. *Statistical methodology*, 6(4), 344-362.
- [Bourguignon,(2014)] Bourguignon, M., Silva, R.B., Cordeiro, G.M. (2014). The Weibull-G family of probability distributions. *Journal of data science*, 12(1), 53-68.
- [Kenney et al, (1962)] Kenney, J., Keeping, E. (1962). *Mathematics of Statistics*. Volume 1. Princeton.
- [Moors,J.J.A.(1998)] Moors, J. J. A. (1988). A quantile alternative for kurtosis. *Journal of the Royal Statistical Society : Series D (The Statistician)*, 37(1), 25-32.
- [Lorenz (1905)] Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. *Publications of the American statistical association*, 9(70), 209-219..
- [Gini(1914) .] Gini, C. (1914). Sulla misura della concentrazione e della variabilità dei caratteri. *Atti del Reale Istituto veneto di scienze, lettere ed arti*, 73, 1203-1248.

- [Kakwani (1977b).] Kakwani, N. C. (1977). Measurement of tax progressivity : an international comparison. *The Economic Journal*, 87(345), 71-80.
- [Kakwani et al, (2020)] Kakwani, N., & Son, H. H. (2021). Normative measures of tax progressivity : An international comparison. *The Journal of Economic Inequality*, 19(1), 185-212.
- [Reynolds et al] Reynolds, M. and E. Smolensky, *Public Expenditures, (1977). Taxes and the Distribution of Income : the United States, 1950, 1961, 1970*, Academic Press, New York, NY.
- [Atkinson, (1970).] Atkinson, A. B. (1970). On the measurement of inequality. *Journal of economic theory*, 2(3), 244-263.
- [Atkinson et al (1987)] Atkinson, A. B., & Bourguignon, F. (1987). Income distribution and differences in needs. In *Arrow and the Foundations of the Theory of Economic Policy* (pp. 350-370). Palgrave Macmillan, London..
- [Shorrocks,(1983)] Shorrocks, A. F. (1983). Ranking income distributions. *Economica*, 50(197), 3-17.
- [Guillaud (2019)] Guillaud, E., Olckers, M., & Zemmour, M. (2020). Four levers of redistribution : The impact of tax and transfer systems on inequality reduction. *Review of Income and Wealth*, 66(2), 444-466..
- [Zenga (2007)] Zenga, M. (2007). Inequality curve and inequality index based on the ratios between lower and upper arithmetic means. *Inequality curve and inequality index based on the ratios between lower and upper arithmetic means..*
- [Son (2013)] Son, H. H. (2013). *Equity and well-being : measurement and policy practice*. Routledge.
- [Piketty,T.] Piketty, T. (2013). *Le capital au XXIe siècle*. Média Diffusion.
- [Gastwirth,(2014).] Gastwirth, J. L. (2014). Median-based measures of inequality : Reassessing the increase in income inequality in the US and Sweden. *Statistical Journal of the IAOS*, 30(4), 311-320.
- [Gastwirth,(2016).] Gastwirth, J. L. (2016). Measures of economic inequality focusing on the status of the lower and middle income groups. *Statistics and Public Policy*, 3(1), 1-9.
- [Davydov (2018)] Davydov, Y., & Greselin, F. (2020). Comparisons between poorest and richest to measure inequality. *Sociological Methods & Research*, 49(2), 526-561..

- [Osberg (2017).] Osberg, L. (2017). On the limitations of some current usages of the Gini Index. *Review of Income and Wealth*, 63(3), 574-584.
- [Jasso (1982).] Jasso, G. (1982). Measuring inequality : using the geometric mean/arithmetic mean ratio. *Sociological Methods & Research*, 10(3), 303-326.
- [Aaberge (2001).] Aaberge, R. (2001). Axiomatic characterization of the Gini coefficient and Lorenz curve orderings. *Journal of Economic Theory*, 101(1), 115-132.
- [1] Kleiber, C., & Kotz, S. (2003). *Statistical size distributions in economics and actuarial sciences* (Vol. 470). John Wiley & Sons.
- [2] Betti, G., & Lemmi, A. (Eds.). (2008). *Advances on income inequality and concentration measures*. Routledge.